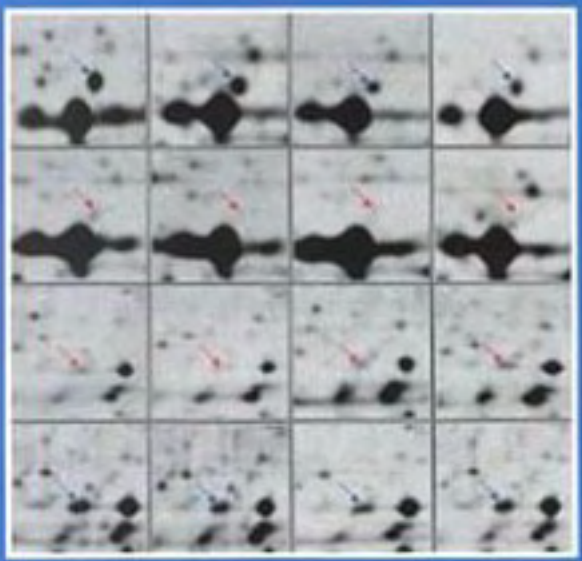
WILEY-VCH

From Genome to Proteome

Advances in the Practice and Application of Proteomics

Edited by Michael J. Dunn



From Genome to Proteome

Advances in the Practice and Application of Proteomics

Edited by Michael J. Dunn

WILEY-VCH

From Genome to Proteome

Advances in the Practice and Application of Proteomics

Edited by Michael J. Dunn

WILEY-VCH

Weinheim · New York · Chichester · Brisbane · Singapore · Toronto

Editor: Prof. Michael J. Dunn
National Heart and Lung Institute
Heart Science Center, Harefield Hospital
Harefield, Middlesex UB9 6JH
UK

Cover illustration: see Fig. 2 on page 323 of this book.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data:
A catalogue record for this book
is available from the British Library

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
A catalogue record for this book is available from Der Deutschen Bibliothek

Printed on acid-free and chlorine-free paper

Composition and Printing: Rheinhessische Druckwerkstätte, E. Dietl GmbH & Co.KG, Alzey.
Bookbinding: J. Schäffer GmbH & Co.KG, Grünstadt.

Preface

It is only five years since the concept of the "proteome" was introduced. However, in this short time proteomics has developed into a powerful approach for exploitation of the constantly increasing wealth of genomic data for a wide variety of organisms. In the post-genome era, proteomics is now providing new insights into how cells, tissues and even whole organisms function at the molecular level.

Two-dimensional electrophoresis (2-DE), using immobilised pH gradients (IPG) in the first dimension, remains the core technique of proteomics, with its almost unique ability to separate reproducibly several thousand proteins simultaneously. Recent advances have been made in the separation of very acidic and basic proteins, while sample preparation and solubilization remains a problem for certain classes of proteins. There is also a need for automation and miniaturization of protein separations, neither of which can be easily achieved using classical 2-DE. These issues are addressed in the papers and review articles dealing with methodological aspects. Mass spectrometry is now accepted as the standard approach to the high sensitivity identification and characterization of proteins in proteomic projects. However, the detailed analysis of co- and post-translational events is an emerging area of interest in proteomics. That proteomics has been widely accepted is evidenced by the large number of papers in this book devoted to its application to a diverse range of biological problems.

The articles collected in this book have been published in the scientific journals *Electrophoresis* and *Angewandte Chemie (International Edition)* in 1999. Most of the papers were presented at the meeting *From Genome to Proteome* which was held in Siena, Italy from 31st August to 3rd September 1998. Therefore, it is also due to the excellent work of the organizers of the meeting Denis Hochstrasser, Vitaliano Pallini and Luca Bini that this bookproject could be realized. In addition, further important review articles in the field of proteomics have been included. I would like to thank all the authors for allowing their contributions to be republished in this volume.

Harefield, September 1999

Michael J. Dunn
Editor

Contents

Eukaryotic cells and tissue

Oncology

Plants

Index

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Angew. Chem. Int. Ed.* 1999, *38*, 2476–2492

Proteome Analysis: A Pathway to the Functional Analysis of Proteins

Friedrich Lottspeich*

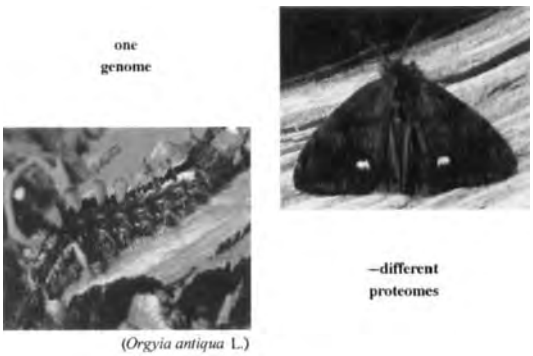
Genomics, the sequencing of whole genomes, is progressing at a constantly increasing rate. Many projects have started, 20 of which have already been concluded, and within a few years the human genome will have also been completely sequenced. Just recently a new milestone was achieved with the decoding of the genome of the helminth *Caenorhabditis elegans*, with 97 million base pairs the largest yet. Its analysis offers promise of information on the approximately 19000 genes that have been found. However, a "decoding" of these data in the true sense of the word is still far removed. What can we learn from genome data? What is the value of these data? Without doubt, genome data and their analysis have for the first time revealed the immense complexity of nature. In place of the dogma "one gene, one protein, one function", an understanding of interwoven regulatory networks has developed. We have learned how to handle large amounts of data. Data bank structures have been constructed, and highly specialized programs for data mining have been or are being developed. Classical protein chemistry has also profited from the genome projects. No longer must every individual amino acid of a large protein be analyzed—in many cases a hopeless task—rather it is sufficient to determine small regions of a protein. However, not everything can be derived from the DNA sequence. What is the function of the gene product, what do the active proteins look like? The answers to these questions are often not set out in the primary sequence. What is the situation at the mRNA level? The most recent developments with cDNA chips have given rise to much hope that changes in mRNA can be rapidly and cost effectively analyzed. Nonetheless, can the abundance of mRNA also characterize the complex relationships which constitute a certain metabolic situation or a pathological status? In part, yes; in a few cases it has been possible to correlate an observation with an altered mRNA pattern. But whenever posttranslational modifications, interactions, or degradation and transport phenomena determine the function of a protein, mRNA can no longer provide any information. Here the only recourse is to turn to the level of the proteins and to investigate directly their type, modifications, and above all their abundance. Proteomics, the quantitative analysis of proteins present in an organism at a certain time and under certain conditions, is a key to functional analysis. In the near future proteomics will determine target search and target selection for both basic research, for example in the unraveling of reaction and regulation networks, as well as for applied research, as in the development of medicines.

Keywords: analytical methods · electrophoresis · mass spectrometry · proteins · proteomes

1. Introduction

The genome projects that are underway worldwide have as their target the sequencing and subsequent analysis of complete genomes. The sequence analyses of many genomes of simple organisms, such as bacteria (*Escherichia coli*, *Bacillus subtilis*, *Helicobacter pylori*, *Haemophilus influenzae*, etc.) and the yeast *Saccharomyces cerevisiae* as a representative of eukaryots, have already been determined.[1] Large genomes, such as those of the human being or of the plant *Arabidopsis thaliana* are also part of a concerted worldwide action; the human genome project will probably be finished within a few years. However, the complete genome of an organism gives only a relatively static overview of the functional potential of an organism and does not describe the immense dynamic process which occurs in a living organism. For example, every somatic cell of the butterfly illustrated in Figure 1 and its caterpillar contains identical genetic information. The conversion of this genetic information—that is, the expression of the different genes into proteins—takes place, however, during the different development stages of an organism as well as in different cell types

[*] Dr. F. Lottspeich
Max-Planck-Insitut für Biochemie
D-82152 Martinsried (Germany)
Fax: (+49)89-8578-2802
E-mail: lottspei@biochem.mpg.de



(*Orgyia antiqua* L.)

Figure 1. Caterpillar and butterfly of *Orgyia antiqua L.* (reprinted with permission of TOPLAB GmbH, Martinsried, Germany).

and under different environmental conditions. This leads to an enormous individual phenotypic diversity in nature.

During the conversion of genetic information into proteins, the tools and raw materials of a cell, there are regulatory mechanisms that adjust the relative amounts of individual proteins with the utmost precision. The smallest of disturbances of this finely tuned protein expression can lead to considerable biological consequences. Nature has therefore developed a complex (and very robust) regulation network which at all levels of information conversion—from transcription, through translation, to the level of the protein—is equipped with numerous feedback and reference points. Thus, the expression of proteins is regulated by transcription and translation factors, which are themselves proteins and therefore subject to the same regulation mechanisms of synthesis and protein degradation. Moreover, the regulation network becomes more complex since proteins can be structurally altered by posttranslational modifications (e.g. phosphorylation, glycosylation, processing), and hence their biological activity and function can be modulated. Knowledge of posttranslational modifications of a protein is therefore an important factor in the understanding of its mode of action and is thus essential for a functional analysis. These endogenous molecular relationships are themselves affected to a considerable extent by exogenous parameters such as temperature, culture conditions, and stress.

All these complex, interconnected processes are under precise spatio-temporal regulation. It has been demonstrated that even the most lavish molecular biological methods for the functional analysis of individual genes ("gene disruption", "knock-out techniques", etc.) frequently yield results that are either ambiguous or are of no practical value at the molecular level. Furthermore, in general no conclusions can be drawn on the amount of a corresponding active protein from the amount of an mRNA. There is therefore a considerable need for complementary strategies directed at proteins which, together with molecular biological investigations, can overcome the problem of the comprehensive functional analysis of genes and gene products.

2. From Protein Analysis to Proteome Analysis

Up to 1950 it was still unclear whether a certain protein possessed a defined and unequivocal covalent structure, or consisted of a very heterogeneous mixture of amino acid polymer chains. The work of Pehr Edman and Frederic Sanger on the sequence analysis of proteins and peptides demonstrated that a particular protein has a clear and unified structure.[2] The following three decades were characterized by efforts to assign certain biological functions to individual proteins. A protein was almost always isolated and purified on the basis of its biochemical activity before its covalent structure (amino acid sequence and modifications) could be elucidated in detail. Once this information was known, possible interaction partners were sought and analyzed in detail, which in turn served as the starting point in the search for further interacting molecules. The disadvantage of this reasonable, function-based, and very successfully strategy was mainly the relatively large material consumption and the tedious isolation, during which consideration always had to be paid to retention of biological activity.

The methods of protein development were under constant development, and their sensitivity increased considerably. An important milestone in this methodical advance was the development of two-dimensional (2D) gel electrophoresis in 1975.[3] Even then the great potential of this high-resolution separation method—which could separate a large number of individual proteins from even such complex mixtures as tissues, cells, or body fluids—was recognized. Protein patterns from normal and pathological states were compared, but the differences observed had solely diagnostic value since the

Friedrich Lottspeich, born in 1947, studied chemistry at the Universität Wien. In 1978 he completed his PhD with Pehr Edman at the Max-Planck-Institut für Biochemie in Martinsried, Germany, on the primary structure of fibrinogen. He worked at this same institute until 1984, when he was appointed as the leader of the independent research group "Mikrosequenzierung" at the genetics center of the Universität München. After habilitations at the Universität München and the Universität Innsbruck he returned in 1990 to the Max-Planck-Institut für Biochemie in Martinsried as leader of protein analysis. He is the author of over 550 original publications and the editor of several scientific books. His research interests center on methodical and practical approaches for protein structure elucidation and proteome analysis.

proteins identified could not be further characterized; at that time the analytical methods were either lacking or of inadequate sensitivity. It was only the appearance of special sample preparation techniques[4] and the further development of protein sequencing[5] that from the middle of the 1980s permitted analysis of proteins that had been separated by 2D gel electrophoresis. On the other hand, as a result of the spectacular success and the high development potential of molecular biology, protein chemistry had slipped from the focus of scientific interest at the beginning of the 1980s. Molecular biological investigations for solving biological questions were preferred and which were also successful, often in an astonishingly short time. New gene technology methods were being constantly developed and improved, and their application achieved an all-time high point in the sequencing projects of complete genomes.[1]

An awareness of the complexity of natural regulation networks was obtained, and the handling of a large number of complex samples in microbiology gave rise to considerable pressures for the development of automated and more rapid analysis techniques. These high-throughput methods produced a large amount of data which in turn stimulated the development of data banks and the associated software tools, and hence the birth of bioinformatics.

Towards the end of the 1980s, increasingly loud voices proclaimed that microbiological methods alone could not unravel the multiplicity and complexity of biological processes. It became clear that although the mRNA pattern can in many cases give information on switched-on or switched-off genes and gene families, the amount of mRNA allowed no conclusions to be drawn on the amount of the corresponding active protein. Incalculable processes (mRNA degradation, translation control, protein degradation, and posttranslational modifications) destroy the strict correlation between RNA and protein amount, at the mRNA level as well as at the protein level.[6] The view that proteins must also be included for a more complete picture of biological events continued to assert itself since proteins are the active players in the cell. It was precisely at this point in time that two new mass spectrometric techniques, electrospray mass spectrometry[7] (ESI-MS) and matrix assisted laser desorption/ionization mass spectrometry[8] (MALDI-MS) were first used successfully in protein analysis. With the improvement in these MS techniques and the enormously fast growing information on sequence data from the genome projects, the basis of a completely new, sensitive, and comparatively rapid protein analysis was formed. This analysis make it possible to process a large number of proteins. The old idea of differential analysis of complex protein mixtures could now be started afresh, and it now finds use in proteome analysis.

The word proteome was first used by Marc Williams in 1994 at the 2D gel electrophoresis meeting in Sienna as "the protein equivalent of a genome", and it was rapidly accepted by reason of its "convenience" and analogy to the expression "genome". This term was subsequently more closely defined as it became clear that the mere detection and compilation of all possible protein and protein variants which could be produced from a genome had very little informative value. Even the mere positioning of each possible protein within a given separation space—that is, pure mapping—will for many reasons bring only very little practical use despite enormous expenditure:

- The amount and the posttranslational modifications of a protein are decisively important for its biological activity and action.
- No biological state exists in which all possible proteins of an organism are expressed.
- Even with very good separation systems, several proteins are often located in one position. Therefore the identity of the protein of interest must be repeatedly checked analytically in the relevant experiment.
- Technical problems of proteome analysis and small differences in sample preparation and analytical protocols obstruct the comparability of "proteome maps" from different laboratories.

For these reasons the following views on proteome experiments are increasingly gaining ground: A proteome represents the protein pattern of an organism, a cell, an organelle, or even a body fluid determined quantitatively at a certain moment and under precisely defined limiting conditions. Furthermore, the proteome reflects a current metabolic status of the corresponding cell (or organism) which is determined by manifold interactions of the different molecules that are currently still difficult to analyze and by innumerable environmental parameters. Unlike the genome, the proteome is thus a highly dynamic system which is characteristically altered by changes in "environmental conditions" (for example, the change in culture conditions for a production strain or the addition of a drug to a cell culture).

Proteome analysis are only meaningful in combination with a subtractive procedure (Figure 2) in which two or more well-defined states can be compared. Changes in individual proteins for the protein patterns of these different states (e.g. a cell with and without a drug, or by comparison of cells from normal and pathological states) are observed and quantitatively evaluated. Therefore the proteome can be regarded as a unique, highly sensitive monitor for complex

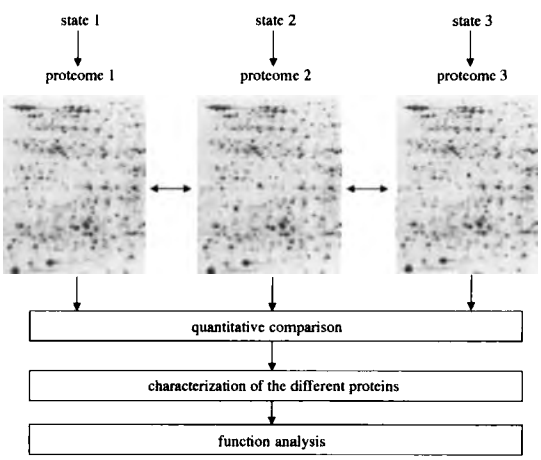


Figure 2. Schematic representation of the subtractive procedure. Differences that are characteristic of the individual starting states are recognized by the comparison of two protein patterns.

metabolic and regulatory relationships of an organism. It should be possible to draw conclusions about the participating functional networks (metabolic pathways, regulatory cascades) from the observed changes which usually involve many proteins. A functional statement can be reached by means of a correlation with the phenotype. The situation can be likened to a movie which, although it consists of a sequence of individual static pictures, obtains a significant information content from the dynamics of the changes between the individual pictures. The stills of a movie are, however, still able to give an impression of the events of the movie, and this all the better the more carefully and the more characteristically the scenes are selected for the stills. Similarly, a meaningful statement can also be achieved in proteome analysis through an investigation of the changes between different (per se static) snap shots of quantitative protein composition. However, we are only at the threshold of an understanding of the "pictures" of the proteome movie. We must learn to correlate the observed differences in protein patterns with biological effects. The better the choice of individual proteome states, the clearer the interpretation of sequence and type of changes.

Proteome analysis can thus give an new quality of answers to biological questions that currently cannot be obtained by any other technique. Only the interaction of proteome analysis with the molecular biological gene technology methods of genome analysis will give a better insight into the complex functional regulation and metabolic networks of nature.

3. Methods of Proteome Analysis

In contrast to classical protein analysis, proteome analysis has the distinct advantage that a biologically relevant statement can be obtained independent of the biological activity of the individual protein. As a result, no consideration needs to be given to protein activity during separation, which allows the use of denaturing separation conditions and hence significantly faster and more efficient analyses. However, the systematic and comprehensive analysis is much more difficult than the analysis of genes since, unlike with RNA or DNA, no amplification possibilities are available for proteins (a limited sensitivity arises from this), and the physical-chemical and biochemical properties of proteins are extremely varied. Therefore, new methods must be developed to be able to handle efficiently protein mixtures that can contain more than 10000 different protein species. Methods of proteome analysis are being developed worldwide, particularly in university research laboratories. However, because of limited resources, only single aspects of proteome analysis are being investigated in individual laboratories. This has given rise to the current situation that, although in principle all methods for the analysis of proteomes exist, the expertise for the totality of the necessary techniques (e.g. sample preparation, 2D gel electrophoresis, enzymatic and chemical cleavage, sequence analysis, mass spectrometry, bioinformatics) within a group is available in only very few places.

Proteome analysis can be divided into individual steps:

- formulation of the question and determination of the starting conditions for the proteome analysis (or for the individual states of a subtractive procedure),
- sample preparation,
- separation of the proteins,
- quantification of the proteins,
- data bank analysis of the protein pattern with bioinformatic methods,
- protein chemical characterization of the altered proteins and analysis of posttranslational modifications.

3.1 Formulation of the Question and Determination of the Starting Conditions for the Proteome Analysis

The methodology and thus the effort and costs of a proteome analysis are significantly influenced by the formulation of the question itself: A proteome analysis in the "puristic" sense should be able to provide statements on functional networks and/or involvement of certain proteins in individual reaction and control mechanisms by means of the differences between two or more snap shots of the protein patterns. The procedural possibilities are very numerous so that only a few representative and random examples will be described:

- Which proteins are involved in reaction networks or biological mechanisms? This information can serve as a basis for the selection of target proteins or marker proteins.[6b, 9, 10]
- How do chemical compounds and environmental conditions affect protein expression, and how do they alter differentiation, proliferation, and metabolism through their pharmacological and toxicological actions? What differentiates here, for example, between drugs with and without side effects? Which proteins or which protein constellations are characteristic of side effects (see also Figure 6)?[6b, 10]
- What are the molecular foundations for efficient production strains in microbiology? Can the expression of proteins be influenced on a rational basis such that higher production yields or product purities can be achieved?[10b]

In general, the difficulty in drawing functional conclusions from changes increases greatly with the number of changed proteins. Therefore, to be able to make meaningful statements, the differences between the individual states of a proteome analysis should not be too widely chosen.

A thorough consideration of the effort needed should be established before a proteome analysis is undertaken. For technical reasons—both during separation as well as during detection—a complete quantitative observation of all expressed proteins is (still?) not possible. Hydrophobic proteins and proteins that are very large, very small, very acidic, or very basic give rise to serious separation problems which prevent the preparation of a complete proteome even under otherwise optimal conditions. It is currently assumed that significantly more than 50% of expressed proteins can be quantitatively detected and analyzed in organisms with a small genome, for example yeast. Since a few milligrams of

protein material can be loaded onto a 2D gel,[11] proteins which occur in copy numbers of more than 100000 per cell can easily be directly recognized, quantified, and also characterized by protein chemistry. However, most proteins are far less well expressed and must be enriched in additional steps (see Section 3.2), which means considerably more effort. In setting the objectives of the experiment consideration must therefore be given to how far an (approximately) complete proteome analysis is at all necessary or meaningful (or even feasible), or whether the investigations can be limited to a certain group of proteins. This can be achieved by biological fractionation (e.g. of certain organelles), but it can also include a physical fractionation (e.g. precipitation). In all cases it is important that the starting material for the actual proteome analysis can be prepared reproducibly.

An important area of application for proteome analysis is the recognition of proteins which are correlated (e.g. diagnostic and therapeutic markers) with a certain state (e.g. a disease). For this purpose the protein patterns of, for example, healthy and pathological cells, tissues, or body fluids are compared. Even if here the individual states of the proteome analysis are biologically widely separated, and there are no expectations that complex functional relationships can be elucidated directly from the changes in protein patterns, the aim to recognize important diagnostic and therapeutic proteins is often achieved. These generally serve as a starting point for continuing analyses. The following aspects must be taken into consideration, however, in the evaluation of results:

- There are almost always problems in the sampling of ex vivo material which can rarely be carried out with complete reproducibility. Consequently, under certain circumstances several cell types with different proliferation and differentiation states are present.
- The "normal values" for individual proteins can differ considerably in different test subjects and must be statistically validated.
- The protein pattern will also differ significantly, even under identical limiting conditions, owing to the relatively frequent occurrence of "polymorphism". It is estimated that the genomes of two people differ in about one million base pairs; many of these differences are manifested at the protein level as amino acid exchanges. The resulting isofunctional, but physically slightly different proteins are found at different positions during proteome analysis and thus alter the protein pattern.

All these difficulties with sampling and the sample material can be ameliorated by a large statistical basis. However, the material for this is sometimes not available.

A further group of experiments is also frequently ascribed to proteome analysis since in terms of methodology it uses the same tools. This is the investigation of protein mixtures which are present, for example, in the search for interaction partners of a protein by immunoprecipitation or in the analysis of the composition of protein complexes. Here the objective is to characterize all proteins that bind to a given protein or are the components of a protein complex. Even if here too all high-throughput methods of proteome analysis are used, the significant difference—and the enormous simplification—to a real proteome analysis is that on one hand the limiting conditions for the analysis are far less complex and on the other the accurate, quantitative determination of the amount of protein present plays only a subordinate role (the absence of an absolute quantity determination prevents a simple statement on the stoichiometric composition of the complexes).

3.2. Sample Preparation

Sample preparation is the first important step of every proteome analysis. If a good description of individual proteome states is to be guaranteed, all possible experimental parameters must be registered and held constant since protein expression of a cell is highly sensitive to changes in external parameters. To be able to keep these parameters under the best possible control, cultivated cells must normally be used for proteome analysis in order to limit biological variability and to have the necessary amount of protein available. All, including unintended, changes during sampling will result in quantitative and qualitative variations in the protein pattern, which again must be compensated for by multiple analysis and complex statistical safeguards.

To be able to assess quantitative relationships correctly, care must be taken that the proteins remain completely intact during every type of sample preparation, that is, that they are neither modified nor degraded. Proteases are frequently released during cell disintegration, which can degrade proteins very rapidly and thus artificially increase the heterogeneity of the protein mixture and at the same time alter the natural quantitative relationships of the proteins to each other. This means that sample preparation must be specifically optimized for each starting material.[12] It is possible that protease inhibitors and/or high concentrations of chaotropic agents for denaturing proteolytic enzymes must be added or a low pH must be used during disintegration. In every case workup must be rapid and at low temperatures. Relatively simple samples, for example body fluids or cells without extremely stable cell membranes, are best dissolved directly in the application buffer for 2D gel electrophoresis. A specific sample preparation protocol must be individually worked out for very demanding samples. The main aim must be to make sample preparation as simple and complete as possible, but above all it must be reproducible; for the latter standardized and essentially automated procedures should be worked out and applied.

A good sample preparation can also include fractionation of the sample, which leads to a reduction in sample complexity. This is particularly advantageous for the subsequent separation of the protein mixture and quantification of the separated proteins. Biological preseparations—such as isolation of certain organelle fractions (e.g. mitochondria, membranes, cell nuclei) or physical-chemical methods such as precipitation, preparative electrophoretic procedures (e.g. free flow electrophoresis), chromatographic prepurification steps—can be used.

Sample preparation for proteome analysis is extremely difficult, mainly because of the broad spectrum of physical-

chemical properties of the individual proteins. Therefore the proteins in such a complex protein mixture, as represented by the proteome, will inevitably differ quite dramatically in their solubility properties. Consequently they behave quite differently and should ideally also be handled differently:

- Proteins that are readily soluble in water or dilute buffer cause few problems. These are mostly cytosolic proteins or proteins from body fluids.
- Proteins that have very stable secondary and tertiary structures and are poorly soluble in water can be first solubilized by the addition of chaotropic substances such as urea or guanidine hydrochloride. The danger of undesired partial and uncontrollable modifications of individual proteins can arise (e.g. by carbamylation), which causes an originally uniform protein to occur in many forms and the sample mixture to become even more heterogeneous. This can cause considerable problems during separation and quantification.
- Proteins which, when brought out of their natural environment, form large insoluble complexes are first solublized by chemical reactions, for example reduction of disulfide bridges.
- Membrane proteins, whose natural environment is lipid membranes and which aggregate very readily during isolation from the membrane and consequently become insoluble, are particularly difficult to handle. These hydrophobic proteins can only be held in solution by the action of detergents, which, however, frequently disrupt the subsequent stages of efficient protein separation, or even make them impossible.
- The high protein concentrations that are necessary for most separation and analytical procedures often involve the risk of aggregation and thus the precipitation of certain proteins. In contrast, low protein concentrations, which are generally advantageous for protein solubility, require additional steps prior to the separation and analytical procedures; this again involves the risk of protein loss.

3.3. Protein Separation

The best possible high-resolution separating techniques that are able to separate all proteins with the most diverse properties at the same time must be used for the quantitative analysis of a protein. The only high-resolution methods available in the molar mass region of proteins are electrophoretic and chromatographic techniques. Yet in a given analysis, neither of these techniques can separate much more than about 100 components. However, since a simple cell probably contains at least 10000 protein species, and moreover the amounts of proteins present in a sample can differ a factor of 10^6 or more, an sufficiently large separation space must be available. This can only be achieved by coupled, multidimensional separation procedures.

3.3.1. Electrophoretic Techniques

Proteins have a zwitterionic character and, depending upon the buffer conditions, can be positively or negatively charged. Electrophoretic procedures separate compounds according to their mobility in an electric field; the electrophoretic mobility of each protein is a characteristic value. Two high-resolution electrophoretic techniques are used in proteome analysis, isoelectric focusing (IEF) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).[3]

In IEF, the individual proteins of a mixture move to their isoelectric point in a pH gradient where they lose their net charge and thus their electrophoretic mobility. If because of thermal diffusion a protein moves from the pH region of its isoelectric charge, it gains a charge and in this moment returns electrophoretically to its "correct" position, which corresponds to the isoelectric point. Thus isoelectric focusing is a concentration end point method which forms sharp, highly concentrated protein bands. It is also compatible with (non-ionic) surfactants, which is an important prerequisite for its use in proteome analysis.

In SDS-PAGE, all protein molecules are loaded with SDS and move as negative SDS-protein complexes in the direction of the anode in an electric field and are separated in the polyacrylamide matrix according to size.

In the combination of IEF and SDS-PAGE, the 2D gel electrophoresis developed by Klose and O'Farrell in 1975,[3] the IEF gel with its highly concentrated protein bands is placed on an SDS-polyacrylamide gel and the samples are further separated according to molecular size in a second separation step. Highly resolved, two-dimensional protein patterns are formed which can be visualized by coloration and evaluated quantitatively (Figure 3). The 2D electrophoresis is

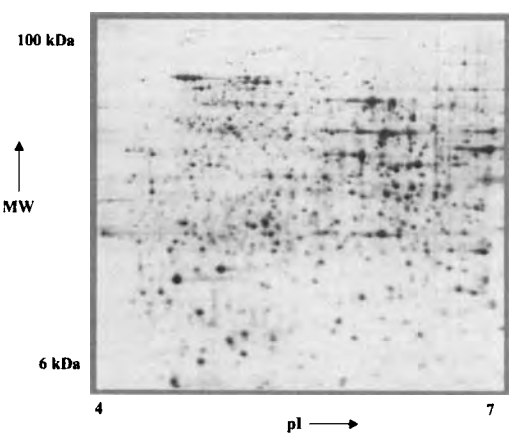


Figure 3. View of a typical 2D gel from *Saccharomyces cerevisiae* (immobiline technique, pH 4-7).

currently the only method for proteins which is able to make available a separating space of several thousand components and can separate complex protein mixtures within a few hours. However, only with the introduction of immobilized pH gradients (IPG)[13] for the isoelectric focusing dimension and through improved sample application techniques[11] over the last few years has 2D electrophoresis reached a level that leads to reproducible protein patterns. With the use of IPG gels, which are also obtainable commercially as ready-to-use

gels, many of the inadequacies of isoelectric focusing with ampholytes (cathode drift, unstable ampholyte mixtures of variable composition) could be overcome. A further, immense advantage of the immobiline technique is that gels with very narrow pH ranges (total pH gradient, for example, 1 pH unit for a gel width of 18 cm) can be prepared which can separate proteins that differ by less than 0.01 pH unit in their isoelectric point (Figure 4).[14] These gels also have the advantage that they can be loaded with a large amount of protein (up to ca. 15 mg) so that poorly expressed proteins can also be visualized directly.[11]

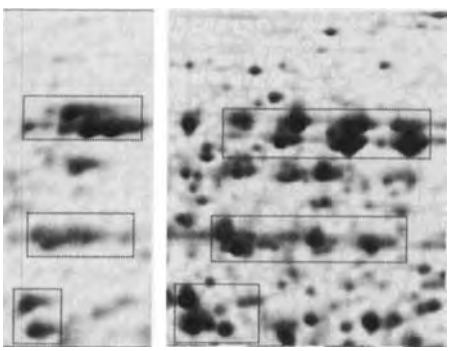


Figure 4. Split pH gradients of IPG gels improve the resolution and loadability. The proteins in the marked areas of the gel shown on the left (pH range 3 – 10) are much better separated on the gel shown on the right (pH range 5 – 7).

Several special properties make 2D gel electrophoresis such an excellent separating technique for proteome analysis:

- There are two complementary and efficient separating principles (IEF separates by charge, SDS-PAGE by molar mass).
- The technique may in principle be used for all proteins (detergent compatibility).
- The technique is parallel and therefore rapid.
- Partial regions of the proteome can be analyzed with very high resolution and high protein loading over a narrow pH gradient.

However, serious disadvantages confront the advantages of high universality and rapidity, which makes further development of this procedure imperative:

- Skilled sample preparation is necessary.
- Sample loading is quantity-limited, so that very poorly expressed proteins cannot be directly recorded.
- Protein transfer from the first to the second dimension is not quantitative and poorly reproducible.
- There is no simple quantification. Proteins must be colored with dyes which bind differently to different proteins.
- Proteins are located in a polyacrylamide gel matrix which is chemically not inert and which therefore can lead to modifications and, moreover, which prevents a direct analysis of the separated proteins.
- Automation is not available.
- The processing of the 2D gel analysis is technically demanding, and it is extremely difficult to achieve good comparability and reproducibility of the gels (at least between individual laboratories).
- Not all proteins are equally well detected. No useful information is obtained for very small (MW < 10000) and very large (MW > 100000) proteins.

In summary, an excellent resolution and an excellent compatibility with almost all protein classes conflict with an only partially adequate reproducibility and robustness. In spite of this, however, 2D gel electrophoresis is currently the only method which is used for the proteome project.

3.3.2. Chromatographic Techniques

Selectivity is achieved in high-resolution chromatographic methods (e.g. ion exchange, hydrophobic interactions, reverse-phase) by the individual interactions of protein molecules with the chromatographic surface, the stationary phase. Strongly interacting proteins can be irreversibly adsorbed onto the chromatographic surface. Unfortunately these adsorption losses are hard to predict since they depend upon the total protein concentration, the individual protein and its concentration as well as other, poorly controllable factors. The adsorbed protein changes the properties of the stationary phase so that a reproducible separation is extremely difficult.

A serious problem of all chromatographic methods is peak broadening. Every protein, even if applied to a column in very small volumes (high concentration), undergoes dilution during the chromatographic separation and elutes from the column in a concentration distribution which corresponds to a Gaussian curve. Since extremely complex protein mixtures are involved, fraction change occurs at well-defined time-points, but is unpredictable with respect to individual proteins. Consequently, many proteins are found in several fractions, and the finally desired quantification of these proteins is made much more difficult.

A further serious disadvantage of the chromatographic techniques is that they are generally carried in series. After an initial fractionation into, for example, 100 fractions, each individual fraction must be subjected to a further separation dimension (e.g. a further chromatographic step). Even with very rapid chromatography (e.g. the total duration of an analysis 30 min including regeneration of the column), the total workup of the fractions from the first separation can require several days. Moreover, a large number of fractions are produced which have to be analyzed further. Proteins are, however, relatively labile compounds which can be rapidly modified, denatured, or degraded in solution, particularly in mixtures with other proteins. Workup periods become unacceptably long if a third chromatographic dimension must be added.

For these reasons multidimensional chromatographic separations have currently found no application in proteome analysis. However, because of the indisputable advantages of chromatographic techniques in quantification (UV detection), and automation, and the numerous possibilities for modulating separation selectivities, there is still the chance that one day multidimensional chromatography will be used as a complementary technique to electrophoresis for proteome analysis

3.4. Quantitative Recording of Separated Proteins

Since most physiological and pathophysiological processes are associated with quantitative changes to individual protein species, the central theme of proteome analysis is the most accurate determination possible of the amounts of the separated proteins. For this to be realized, the proteins must be visualized in an evaluable form.

3.4.1. Detection

The methods most frequently used are different stainings. The problem associated with all stainings is that they give a specific and unpredictable color intensity for each protein (according to its amino acid composition and the modifications that are present). The result is that no absolute statement on the amount of single proteins can be made, and protein patterns with different types of stainings cannot be compared. Furthermore, stainings are linear over a very limited range (maximum two orders of magnitude) because of the saturation effects. To cover a large dynamic region several gels with different amounts of protein must be prepared, so that an approximate calibration curve for each protein can be prepared. Only proteins which lie outside the saturation region of this calibration curve are suitable for quantification. Consequently, computer-supported image evaluation coupled with high-performance data processing is absolutely necessary. If different gels must be compared with each other and if at the same time the total protein amounts or color intensities are different, the question of standardization arises. Occasionally every protein spot is compared with the intensity of a known and constitutively expressed protein, which is itself problematic because constitutively expressed proteins can also change significantly in their amount. The better way is to relate the intensity of each protein to the total intensity of all proteins.

The most frequently used staining for proteome analysis is currently silver staining, in which the proteins are fixed in the gel with trichloroacetic acid and the gels are then placed in a silver nitrate solution. A number of silver ions are bound by the proteins and are precipitated in the form of elemental silver by reduction. The proteins darken very rapidly as a result of the high silver concentrations. The reaction is stopped by a large pH change. The silver staining can be used to detect proteins with relative sensitivity and exhibits a linear range of about $0.5-20\,ng\,mm^{-2}$. Unfortunately this staining is very difficult to reproduce since the blackening is highly dependent upon the duration of development and the temperature. A number of different protocols exist which differ from each other in duration, sensitivity, and simplicity, and which are being continually improved.[15]

Staining with the triphenylmethane dye Coomassie blue is somewhat more robust than silver staining.[16] It is, however, also significantly less sensitive (linear range ca. $50\,ng\,mm^{-2}$ to $1\,\mu g\,mm^{-2}$). Here too different staining protocols exist. The method of Neuhoff et al.,[16b] which achieves color penetration of the protein, gives the best results with respect to a quantitative evaluation.

Protein detection methods with fluorescent dyes such as SYPRO orange or SYPRO red have advantages over conventional staining with silver or Coomassie blue.[17] There are about as sensitive as silver staining but are much faster to carry out (ca. 30 min) and require no fixing of the proteins in the gel. Thus subsequent cleavage or transfer of the protein to a chemically inert membrane should be simplified. A further principle advantage of fluorescence detection is that over measurement times of different duration the fluorescence can often be detected in the presence of the fluorescence of rare proteins in the same experiment and quantified. The absolute amount of a protein cannot be derived from fluorescent staining as amino acid concentration and the type and number of translational modifications affect color intensity. Since the proteins are only visible under UV light and must be cut out, automated spot recognition and automated preparation of protein spots for the subsequent analysis is especially important in fluorescence staining.

In principle it should also be possible to allow all proteins of a proteome state to react covalently with a fluorescence reagent before their separation. In addition to an increased detection sensitivity this would bring the advantage that different states could also be treated with different fluorescent dyes. Then the proteins of the two states could be separated in the mixture under identical conditions ("multiplexing"), and the respective proteins could be assigned to the two states by detection of the different emission wave lengths. Differences would then be simpler to recognize since the technical problems of reproducible separation need not be considered.[18] Fluorescence labels with different optical but identical electrophoretic properties would need to be used. The problem with this procedure lies in the uniform covalent labeling of the proteins with the fluorescent dyes. The conversion is a chemical reaction at the functional groups of a protein. This reaction almost certainly does not go to completion in complex mixtures. Even if a reaction yield of 99% with every single functional group were achievable, a uniform protein would provide a very heterogeneous mixture, which would also have to be separated with highly selective separation methods (primarily in isoelectric focusing). Each of the artificially generated by-products may indeed be present in a small fraction of a few thousandths, but even these minimal amounts almost always lie at the levels of naturally occurring proteins since the concentration range of different proteins in a cell spans at least six orders of magnitude.

Attempts can also be made to prevent the proteins from reacting fully, but to treat them instead with very small amounts of a fluorescence reagent. In this way only a small fraction of the protein is labeled and the main fraction of the protein remains unmodified. With a suitable choice of reagent it is possible to arrange that the separation behavior of the modified and the unmodified protein do not differ significantly.[18]

A fluorescence reaction after the first dimension of 2D electrophoresis could also be in part a way out. On the one hand, high-resolution separation is effected by isoelectric focusing, and on the other the mass heterogeneity from a few

fluorescence molecules lies below the resolution capabilities of the second dimension, SDS-PAGE.[19]

Immunological colorations with antibodies are very sensitive and can also detect very rare proteins directly. Unfortunately there are no specific antibodies for the peptide bond, so it cannot be used as a general detection tool. However, antibodies are very valuable in the detection of certain specific proteins or protein groups (e.g. tyrosine-phosphorylated proteins with a phosphotyrosine antibody). This applies similarly to lectin staining, which can be used to recognize glycosylated proteins.

Radiolabeling can be used very profitably for specific questions which allow an incorporation of radioactivity during the culturing of cells. This normally means the handling of high quantities of radioactive material, for which laboratories must be specially equipped. Thus, for example, specific proteins that are synthesized at certain time points can be detected by metabolic labeling with [^{35}S]Met or [^{35}S]Cys in pulse – chase experiments. Alternatively, a selective analysis of phosphorylated proteins (i.e., phosphoprotein partial proteome) can be achieved by labeling with [γ-^{32}P]ATP.

Even though radiolabeling is a highly sensitive detection method and quantification has been improved in its linearity by the development of photoimaging techniques, it does not reflect the absolute amount of a protein since the incorporation of the labeling is dependent upon the amino acid composition of the individual proteins.

3.4.2. Quantification

The next step after the detection of the separated proteins is their quantification. Except for amino acid analysis, no method exists to determine the absolute amount of individual proteins separated in a 2D gel. However, amino acid analysis is very sensitive towards contamination and work-intensive (even if readily automated), and it gives reliable results only with more than about 0.4 µg of protein. Thus, it is too insensitive for most proteins in a 2D gel.

Stained gels are normally measured by computer-supported laser densitometry. It has been demonstrated that only laser scanners have the required high dynamic range to provide good results. Special software packages for automatic spot recognition, quantification, and presentation of the results are used for data collection and further evaluation. The raw data and results are deposited in structured form in large data banks which can be accessed by special data bank programs for "data mining".

3.5 Data Evaluation by Bioinformatics

The large amount of data that is obtained during proteome analysis can no longer be evaluated without the massive support of specialist data bank software. Not only is rapid access to data concerning gel position, quantity, and identity of a protein important, but also the in part totally unstructured data on the origin, preparation, and work-up history of the sample needs to be readily available. Moreover, it must also be possible to include clinical data banks and literature data. Usually the desired information is deposited worldwide in different data banks which are linked by on-line connections. Part of the information is available from publicly accessible data banks by means of the World Wide Web.[1b] Other information, however, is so sensitive that it cannot be accessed publicly. The data bank software must also allow complex interrogation as, for example, which drug positively influences the expression of which proteins in which patient group, or which compounds influence the expression of certain proteins in the same way. The programs used for the quantitative evaluation of 2D electrophoresis are hopelessly overwhelmed by such questions. Special data bank programs for extensive data mining of RNA expression data and proteome data are currently a priority in the development in bioinformatics. The clear presentation of results of proteome analysis data is also receiving increased attention. These can be simple bar charts in which the protein amounts for individual proteins or protein groups from different experiments are illustrated (Figure 5). The progression of the amount of protein present at any point in time can be readily recognized, and proteins with a similar quantitative progression can be further collated by cluster and correlation analyses

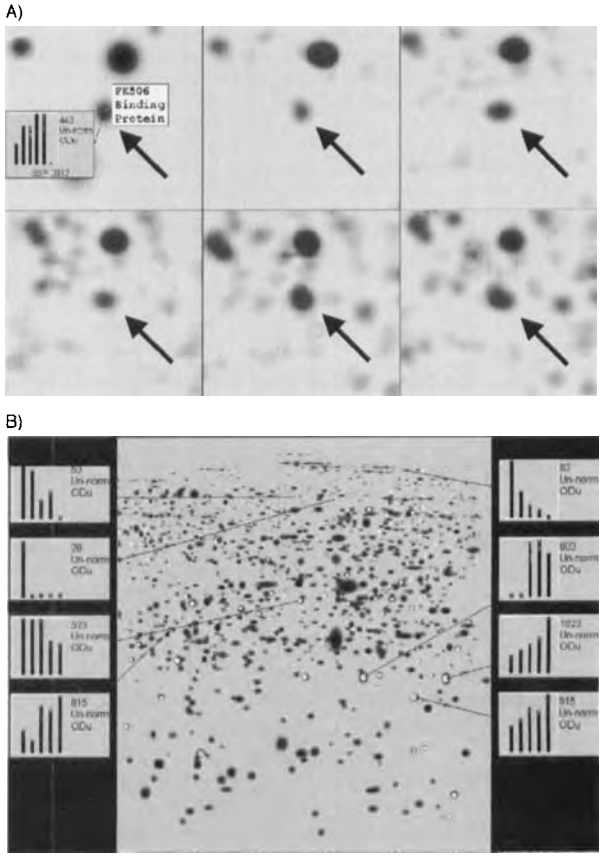


Figure 5. Illustration of proteome data. A) Change in the amount of a protein in different experiments. B) Illustration of the changes of several proteins in different experiments.

and compared.[20] Representations which depict complex relationships are also illustrative, and significance parameters can also be clearly considered (Figure 6).

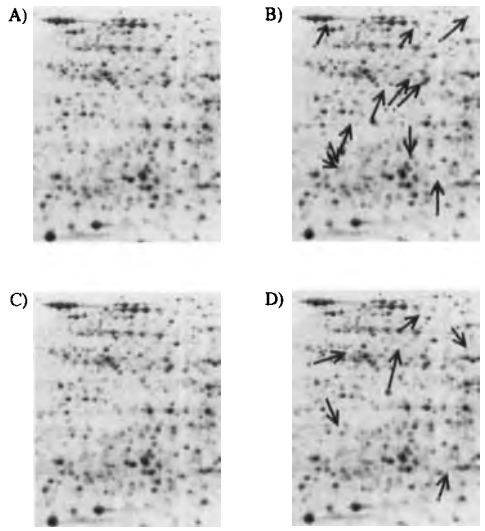


Figure 6. Illustration of a proteome analysis with consideration of statistical parameters, and schematic example for estimation of the action of a drug. A) Proteome of a normal cell. B) Proteome of a pathologically changed cell. The arrows show the changed proteins characteristic of the disease state. C) Pathologically changed cell under the influence of the ideal drug. The pathological changes are reversed. The cell appears normal. D) Pathologically modified cell under the influence of a real drug. A few pathological changes are reversed, but new changes to the proteome may also be induced, an indication of side effects. The upward or downward slope of the arrow is a measure of the increase or decrease in the respective amount of protein. The length of the arrow shows the statistical significance of these changes (unchanged proteins would all have horizontal arrows and are omitted for clarity).

3.6. Protein Chemical Analysis of Proteins Separated by Gel Electrophoresis

Methods for the analysis of proteins that have been separated by gel electrophoresis have been the subject of intensive efforts over the last 15 years. Since the proteins in a gel matrix are practically inaccessible to normal protein chemical methodology—such as sequence analysis, amino acid analysis, and mass spectrometry—the first success was to transfer the proteins from the gel onto chemically inert membranes and to immobilize them there.[4a-d] The intact, immobilized proteins can be analyzed by amino acid sequence analysis, by amino acid analysis, or more recently by mass spectrometric methods.[21, 22] However, since the analysis of intact proteins is work-intensive and rarely produces sufficient information for unambiguous protein identification, the analysis of internal fragments after enzymatic or chemical cleavage of the protein has proved to be the more efficient strategy. The procedure most generally used today within the context of proteome analysis is reproduced in Figure 7 and will be discussed in the following sections.

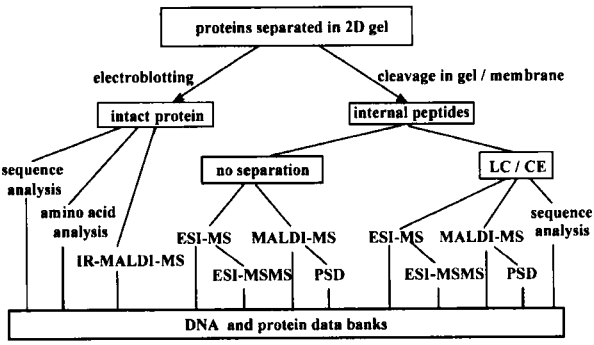


Figure 7. Schematic representation of the setup for high-throughput protein identification, and strategies for the characterization of proteins separated by gel electrophoresis.

3.6.1 Analysis of Intact Proteins

Amino Acid Sequence Analysis

The amino acid sequence analysis of blotted proteins is an automated, relatively work intensive, expensive, and slow standard procedure, and can be used without serious problems for protein amounts in the picomole range.[4a-c] The protein is frequently identified, or at least clearly characterized, by its N-terminal sequence. The achievable sequence length (maximum 40 amino acids) is sufficient for identification or homology searches in data banks, but it covers only a small part of the whole protein sequence. Therefore, closely related isoenzymes, splice variants, modifications, or point mutations outside the investigated amino terminal region of the protein cannot be recognized. The main problem is, however, that more than half of all proteins are N-terminally modified and are thus not accessible to sequence analysis.

Amino Acid Analysis

Every protein has a characteristic amino acid composition. Therefore, almost every protein can be identified with high probability in a protein data bank simply from the relationship of its amino acids to each other.[4e-h] In practice, however, the accuracy of the determination of individual amino acids is limited. Because of the drastic reaction conditions used in the hydrolysis of peptide bonds, every amino acid analysis is a compromise between total cleavage of the peptide bonds and the least possible destruction of sensitive amino acids. Therefore, only the most stable amino acids with relatively large error margins are used for a data bank search. The methods may be readily automated and is almost as sensitive as sequence analysis, but has the advantage that it can produce results even with N-terminally blocked proteins. At the same time amino acid analysis is the only method which can deliver information on the absolute amount of a protein. Unfortunately in practice these advantages are subject to a number of severe limitations which greatly restrict the general value of amino acid analysis in proteome analysis. The technique can only identify proteins whose sequence has been deposited in a data bank. With most proteins, however, the protein sequence is available only as the translation of the DNA sequence,

which often differs significantly from the sequence of the naturally occurring form of the protein. After translation, signal sequences are cleaved and proteins further processed and modified. The amino acid composition of the experimentally accessible protein thus differs from the theoretical composition deposited in the data bank. The former is normally not included in the data banks at all, or only marginally so. Because of the described technical and intrinsic problems, and because of the ubiquitous and unavoidable contamination (free amino acids from buffers, keratins, etc.), the margin of error in the determination of individual amino acids is so large that an identification of particularly poorly expressed proteins is frequently ambiguous. In practice, single amino acid exchanges or modification of an identified protein cannot be recognized.

IR-MALDI Mass Spectrometry

A significant characteristic of a protein is its molar mass. Even if the current methods are unable to determine the masses of proteins with adequate accuracy (errors in mass determination are greater than 100 ppm) to identify a protein in a data bank on the basis of its mass alone, mass information of the whole protein in association with other data is extremely valuable. For smaller proteins at least, the quality of mass determination is sufficient to recognize posttranslational modifications by a comparison of the mass calculated theoretically from the DNA sequence and the mass determined experimentally. However, the identity of a protein must be determined by means of sequence analysis, amino acid analysis, or the methods described in Section 3.6.2.

MALDI-MS is suitable for the analysis of immobilized proteins (Figure 8).[8, 21, 22] In a standard MALDI-MS preparation the proteins are embedded in a crystalline matrix of small organic molecules in which they are then vaporized and ionized by laser bombardment. The task of the matrix is to separate the individual protein molecules, absorb the laser light, and relax the energy in the solid lattice within a short period of time. Thus, an explosionlike dissipation of a small region of the solid surface and a transfer of the matrix and the protein molecules into the gas phase is achieved. With correct selection of laser energy this process is so gentle that the large, thermally labile protein molecules also remain intact. It is probable that the matrix also plays a role in the ionization of the protein molecules. Different wave lengths of laser light can be used when the interaction of the laser type (UV or IR) and the matrix (e.g. 2,5-dihydroxybenzoic acid for UV-MALDI or succinic acid for IR-MALDI) is important. The resulting protein ions are then accelerated in a time-of-flight (TOF) analyzer and the precise time of flight from ionization to detection is measured (Figure 8a).[7a] Since the time of flight at a given acceleration potential and flight path is dependent only upon the root of mass/charge (m/z), an accurate mass determination of the analyte molecule can be carried out by means of a calibration.

To analyze electrophoretically separated proteins after a transfer onto a chemically inert membrane directly with MALDI, the membrane must be incubated with the matrix immediately after electroblotting. UV-MALDI-MS usually

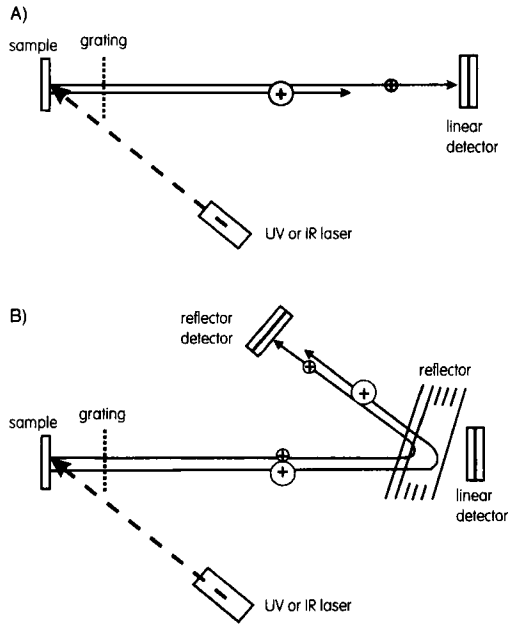


Figure 8. Schematic representation of the setup for MALDI mass spectrometry with a A) linear and B) reflector TOF detector. A) A laser pulse evaporates and ionizes the sample molecule. These are accelerated by the potential applied to the grating of the TOF analyzer. The speed of flight of the ions is proportional to $1/\sqrt{m}$ (m = mass of the ion). B) Ions can disintegrate during passage through the field-free drift path (PSD). These fragments, which carry structure information, travel with equal speed and reach the linear detector at the same time. For separation they are braked by the potential wall of the reflector, forced into reverse, and accelerated to the reflector detector. Since larger masses penetrate the reflector field more deeply, they travel along a longer path to the detector than small ions.

requires relatively hydrophobic molecules as matrix which are soluble only in organic solvents. During incubation of a membrane with such a hydrophobic matrix the protein molecules are partly dissolved and the geometric arrangement of the protein spots is lost by diffusion.[22a]

In contrast, it has been shown that during incubation of a membrane with a hydrophilic matrix the local arrangement of the blotted protein spots on the membrane and thus the resolution of the gel electrophoresis and even the intensity distribution within the protein spots is maintained.[21] Such hydrophilic matrices are used for IR-MALDI-MS, and therefore electroblotted proteins can be analyzed with high sensitivity (down to the attomole range).[21b] The limits and the intensities of the protein spots can be readily recognized, and protein mixtures within a 2D gel spot can also be recognized. Unfortunately the very different signal intensities for the individual proteins do not allow an estimation of the respective protein amounts. The development of automation and the improvements in MALDI-MS methodology with respect to ion yield and mass accuracy of larger proteins suggest that in the near future detection of electroblotted proteins will be carried out by MALDI-MS with automatic scanning of the blot membrane; a spot will be characterized by its mass as well as by its electrophoretic position.

3.6.2 Analysis of Internal Fragments

Enzymatic Cleavage and Elution of Peptides

Since the analysis of intact proteins is time-consuming and rarely delivers sufficient information, the analysis of internal protein fragments has proved to be the most efficient strategy for the characterization and identification of electrophoretically separated proteins. After enzymatic cleavage of the separated proteins, which can be carried out directly in the polyacrylamide gel matrix,[4i–o] the peptides obtained are eluted and analyzed by mass spectrometric and/or other protein chemical methods. A plethora of protocols exist, each of which show slight changes but principally fall back on one piece of work.[4i] The proteins separated by gel electrophoresis are punched out as tightly as possible, washed, and either dried or dehydrated by the addition of acetonitrile. A small volume of buffered enzyme solution is pipetted onto the shrunken gel fragments. Mainly trypsin, endoproteinase Lys-C, endoproteinase Glu-C, or endoproteinase Asp-N are used, which cut all proteins very specifically and completely. Trypsin is used very frequently, particularly if mass spectrometric analyses are to follow since it is itself relatively rapidly cleaved, and the autoproteolysis fragments serve as internal reference peptides for mass calibration. After incubation for a few hours at elevated temperature the reaction mixture, which contains the cleaved peptides partly in the supernatant and partly in the gel, is worked up differently for the subsequent analysis methods. The supernatant of the cleavage solution can be analyzed directly; the yields of larger or more hydrophobic peptides are often poor. Usually these peptides are eluted from the gel pieces with volatile acids containing organic solvents, for example trifluoracetic acid/acetonitrile (0.1/0.99 – 1/99). The further workup of the eluted peptides is guided by whether the proteome under investigation originates from an organism whose genome has already been fully elucidated. The large number of samples from proteome analysis has already led to automation of cleavage and subsequent elution of the proteins separated gel electrophoretically.[23] Commercial apparatus is available which can process automatically 50 samples per day (with sample amounts of less than 1 pmol).

Internal Sequences of Proteins from an Organism with Completely Sequenced Genome

A very simple and rapid identification of the proteins of an organism with known genome can be achieved with solely mass spectrometric methods. The eluted peptides are analyzed by MALDI-MS or by nano-electrospray MS without further separation.[7d, 24]

In general, the peptide mixture from cleavage on the gel is subjected directly to MALDI-MS, where robotic sampling systems may also be used. MALDI-MS is now so far developed that the samples can be analyzed automatically with very high mass accuracy (error less than 20 ppm). The data can be evaluated on-line so that the protein can be identified with very high probability in sequence data banks from the peptide mass pattern obtained. By this method more than 90% of the analyzed proteins can be identified directly with certainty and with high sequence coverage (Figure 9). Identification is not unambiguous with some proteins so that further analyses must be carried out. This can be achieved with MALDI-MS with post source decay (PSD) spectra.[25] The spontaneous disintegration of proteins and peptides during the flight in the field-free drift path of the TOF analyzer (metastable decomposition) is utilized here. The fragments of an ion all continue to travel with the same speed and reach the detector of a linear TOF analyzer at the same time. If the fragments in a TOF analyzer with reflector are allowed to travel up instead of against a uniformly charged

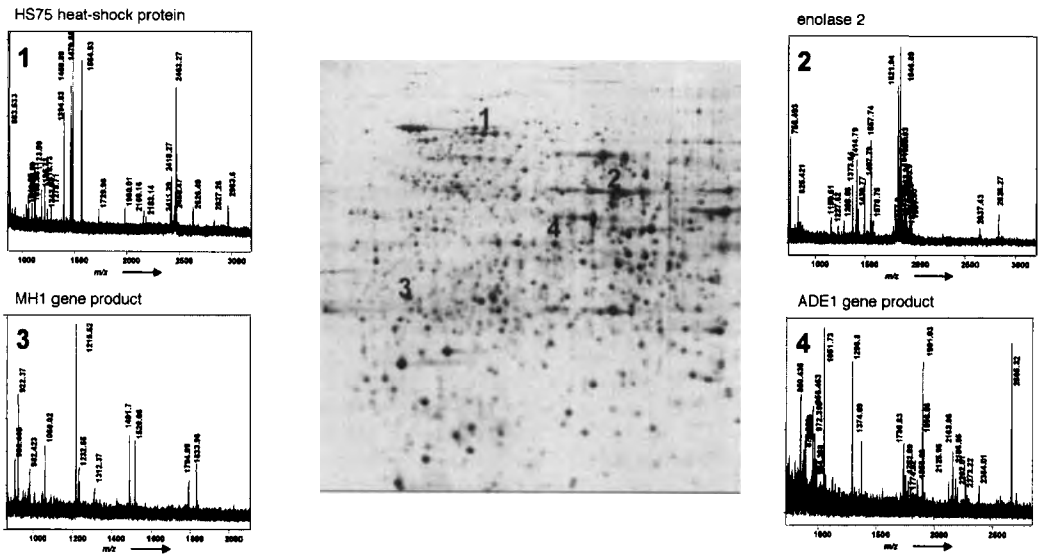


Figure 9. MALDI spectra that identify proteins separated by 2D electrophoresis. Enzymatic cleavage of a protein gives a mixture of proteins whose masses are analyzed in a data bank with computer support.

field (the reflector), the fragments come to a stop and reverse (Figure 8b). The different masses penetrate into the reflector field to different depths and so have to travel different paths, consequently reaching the detector of the reflector instrument at different times. The PSD fragments are thus separated, and their mass can be determined by calibration with reference compounds. The interpretation of the spectra is relatively complicated and tedious so that an exact evaluation of these spectra cannot be made in a proteome experiment where many samples must be analyzed. The uninterpreted spectra obtained are therefore compared automatically and on-line with a data bank of calculated spectra which were generated in silico from all theoretically possible peptides of an organism. The software then gives an identity suggestion based on similarity.[26]

The alternative to MALDI-MS analysis of the peptide mixture from cleavage in the gel is electrospray (ESI) MS (Figure 10).[7b-c] In ESI-MS the sample in liquid form is continuously sprayed in an electrostatic field to form small

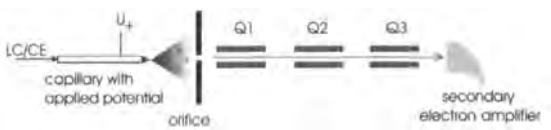


Figure 10. Schematic representation of the setup for ESI mass spectrometry. The samples in liquid form (HPLC, CE, injection) are passed continuously through the capillary, to which a high potential is applied. The resulting spray contains the (multiply) ionized molecules which are introduced into the high vacuum field of the quadrupole mass spectrometer through the orifice. For mass determination the quadrupoles Q1 and Q2 are so set that all ions can pass. The actual mass determination occurs in Q3. For structure investigations the mass filter Q1 allows only ions of a single mass to pass into Q2. Fragmentation of the sample ions occurs in Q2, which is filled with argon gas, and the masses of the resulting fragments are analyzed in Q3.

droplets, which are rapidly desolvated and the charge density on the surface of the droplets becomes even larger. After repeated, spontaneous disintegration of the droplets (Coulombic explosion), the desolvated, multiply charged molecule ions are directed to the mass spectrometer, where they are normally detected with a quadrupole mass analyzer. This type of mass analyzer is a mass filter which under preset physical conditions only allows ions with a totally defined mass/charge ratio to pass through. All other ions cannot pass the analyzer and are lost. Through continuous change of the potential at the quadrupole, ions of different masses are allowed to pass sequentially (scanning), and the intensity of the ion flow is recorded in relation to the m/z ratio. The accuracy of the mass determination allows the charge state of each mass signal to be determined from the isotope distribution, and hence multiply charged ions to be recognized and, with computerized support, the mass of singly charged ion to be calculated.

Since the observed ion flow correlates with the concentration of the sprayed sample, attempts are being made to spray highly concentrated solutions at very low flow rates of $nL\,min^{-1}$ (nanospray ESI[7d]). In addition to increased sensitivity, this has the advantage that very long measurement times are available with a sample of a few microliters.

ESI-MS also offers the opportunity to obtain at least partial structure (sequence) information from the fragmentation of individual peptides.[27] A triple quadrupole apparatus is used in which the first quadrupole is used for the selection of a peptide ion. These selected ions are directed to a second quadrupole where they collide with argon gas and are fragmented. The resulting fragments are then analyzed in a third quadruple (Figure 10). The ESI-tandem-MS fragment spectra are somewhat clearer and more convenient to interpret than MALDI-PSD spectra, but here too automated and software-controlled interpretation programs are increasingly used for proteome analysis. Labeling of the C-terminal end of the peptide with ^{18}O, by the enzymatic cleavage of the protein in the presence of $H_2{}^{18}O$, simplifies the interpretation of the fragment spectra considerably.[28] An alternative to the quadrupole apparatus are the ion traps, whereby ions are trapped in a suitable electric field and can be kept on stable paths.[29] The individual ions can be catapulted from the trap by changes in the electric conditions and then recorded at a detector. The scan speeds are up to ten times faster than with a quadrupole detector, and the ion trap is particularly suitable as a rapid detector for coupling with HPLC, where a limited amount of time is available for each substance peak. Like the triple quadrupole apparatus, the ion trap is also suitable for structural determination of peptides since individual ions can be selected in the trap and fragmented by collision with inert gas atoms. From the results of the mass spectrometric sequence analysis—which usually does not give complete peptide sequences, but only "sequence tags"—and in combination with peptide mass finger print, each protein can usually be unambiguously identified. Recent developments also allow the coupling of electron spray ion sources with TOF analyzers (orthogonal TOF apparatus, Q-TOF, Figure 11). The whole ion flow is passed through quadrupoles and hexapoles to an orthogonal acceleration system (pusher), which directs the ions to a reflector TOF analyzer. This apparatus also permits structure identification when the in-line quadrupole is used to select the ion, and the selected ion is fragmented in a subsequent collision cell and then directed into the TOF by the pusher and analyzed.

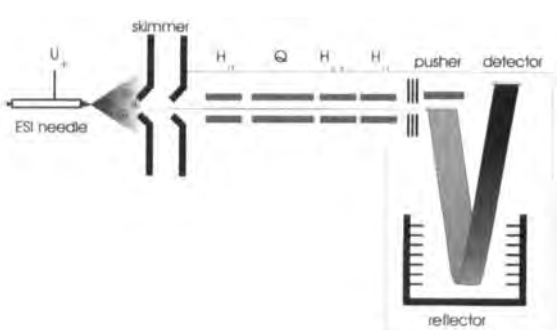


Figure 11. Schematic representation of the setup for ESI-Q-TOF mass spectrometry. The ESI ion flow is directed to the pusher through the quadrupole (Q) and the hexapole (H_n) then deflected to a reflector TOF analyzer. Structure information is obtained when Q is used for the selection of one ion which is fragmented in the collision cell H_{col}. The fragments are analyzed in the TOF unit.

Internal Sequences of Proteins from an Organism with Unsequenced Genome

The strategy of identifying proteins by mass spectrometry solely by means of protein mass patterns currently has two main limitations:

- The genome sequence of the organism must be essentially known. This still has a practical significance today if one reflects, for example, that the human genome project is still not concluded, although it is anticipated that the DNA sequence of the most important organisms will be known in the near future.
- For a number of reasons, posttranslational modifications are only recognized to a limited extent.

The methods of rapid protein identification by means of peptide mass pattern fail if the sequence of the protein under investigation is not present in a data bank, if the protein is extensively modified, or if several proteins are present in one protein spot. Even if large parts of a genome are accessible only in partial sequences, EST data banks (EST = expressed sequence tags), analysis by peptide mass pattern is unsuccessful. Furthermore, difficulties in the cleavage and elution of larger peptides from the gel lead to a poor sequence coverage, and in mass spectrometry itself there are also inherent difficulties. Thus, suppression effects prevent quantification of individual peptides,[30] signals from ubiquitously present contamination (e.g. keratines) complicate the spectra, and artificial modifications during sample workup (e.g. oxidation or modifications of cystein residues) or during the measurement itself (e.g. oxidation, fragmentation) frustrate a simple assignment of the signals detected.

Additional time-consuming and slow protein-chemical microtechniques must be used for detailed analyses, such as the determination of posttranslational modifications of a protein or the characterization of proteins from organism with unsequenced genome (Figure 7). Capillary HPLC or capillary electrophoresis with coupled on-line mass spectrometry procedures are used for the separation of peptides after enzymatic cleavage in polyacrylamide gels. Such analyses, with which the protein is not only identified but also investigated, should be carried out starting from at least two different enzymatic cleavages in order to cover the total protein as far as possible with the sequence data. The end effect is that methods that produce de novo sequence information—that is, either mass spectrometric sequencing, which is cumbersome and is often associated with high uncertainty, or the classical Edman degradation, which gives unequivocal results but has a sensitivity limit of about 1 pmol of peptide starting material—must almost always be used. Both sequencing techniques are slow and currently cannot handle the large number of samples from a proteome analysis.

3.6.3. Analysis of Posttranslational Modifications

An important area of proteome analysis is the analysis of posttranslational modifications, which have a considerable effect on the functions and properties of a protein. Since this detailed protein analysis is cumbersome and tedious, only proteins for which there is an indication of posttranslational modifications should be investigated in detail. This information can be obtained from the mass spectrometric analysis of the whole protein with IR-MALDI (see Section 3.6.1). A deviation of the observed isoelectric point of a protein from that calculated from the DNA sequence is also a good indication of a modification.[31] Special mass spectrometric techniques such as precursor scan or neutral loss scan and the mass spectrometric sequencing methods MALDI-PSD and nanospray ESI MSMS are in particular used for the exact determination of the type and the position of the posttranslational modification.[32] They are supplemented by the classical structure determination procedures, mainly by the Edman sequence analysis. Partial modifications at several sites of a protein, which often occurs during phosphorylations and glycosylations, are especially difficult to analyze. In these cases separation of the peptide mixture by nano-HPLC with on-line mass spectrometric analysis must almost always be carried out. In summary, the characterization of posttranslational modifications is still demanding for protein chemists and, in spite of the enormous advances in recent years, still cannot by solved with high-throughput methods.[33]

4. Summary and Outlook

Unlike the static genome, a proteome—the quantitative protein pattern of an organism, a cell, or a body fluid under quite precisely defined limiting conditions—is highly dynamic. Several proteomes, each of which reflects a current development and metabolic state at a certain timepoint, exist for a single genome. Through appropriate selection of different states, functional conclusions can be made from the different protein patterns of the corresponding proteome with the help of bioinformatics.[13] Proteome analysis can only be used to its full potential in association with other areas of bioscience such as molecular biology, genetics, immunology, and medicine (Figure 12). At the present time proteome analysis is in statu nascendi; the methods are still immature, but are being developed at astonishing speed at all levels. Attempts are being made to improve or replace the only currently successful method, 2D gel electrophoresis, and

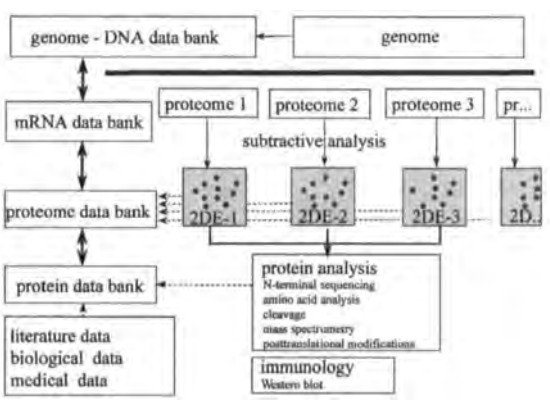


Figure 12. Proteome analysis in the context of different areas of bioscience. 2DE = two-dimensional electrophoresis.

quantification is gaining increasing significance with automation, new detection procedures, and the entry of bioinformatics. The identification and characterization of the separated proteins is most advanced. It is almost routine for proteins from organisms with totally sequenced genomes, although there is still room for improvement in terms of throughput and sensitivity. The recognition and localization of posttranslational modifications in high-throughput procedures will be the next stage in the development. Proteome analysis of organisms with incompletely established genomes is currently so difficult that in the immediate future it will remain restricted to special problems.

Since proteome analysis can be applied to a large number of complex questions, it is surely only a matter of time and financial investment before the first large practice-oriented proteome projects are started. The first realization of a comprehensive proteome project will certainly only be possible by a combination of the different and only partly available techniques with simultaneous massive automation and further development at all levels of proteome analysis; the development of new technologies is also foreseen. The large expenditure for instrumentation and the expertise required for a proteome analysis makes it likely that large proteome projects will only be carried out in specialized centers, the first of which are currently being developed.

The correlation of the proteome with mRNA expression (transcriptome) will be especially interesting since the two processes are regulated at quite different levels. The flow and changes in metabolic products and small molecules (the metabolome/fluxosome/physiome—the name has not yet been finally assigned), which have diverse feedback mechanisms to the transcriptome and the proteome, will also be more closely investigated in the future. Finally, insights into the interrelationship of genome, transcriptome, proteome, and metabolome will surely bring new knowledge of biologically relevant and very complex interconnected mechanisms of living organisms.

Received: February 16, 1999 [A 326 IE]
German edition: *Angew. Chem.* **1999**, *111*, 2630 – 2647
Translated by Dr. David Le Count, Congleton, Cheshire (UK)

[1] a) http://pedant.mips.biochem.mpg.de/; b) http://www.expasy.ch/.

[2] a) P. Edman, *Acta Chem. Scand.* **1950**, *4*, 283 – 290; b) F. Sanger, H. Tuppy, *Biochem. J.* **1951**, *49*, 481 – 490.

[3] a) P. H. O'Farrel, *J. Biol. Chem.* **1975**, *250*, 4007 – 4021; b) J. Klose, *Humangenetik* **1975**, *26*, 231 – 243.

[4] a) J. Vandeckerkhove, G. Bauw, M. Puype, J. Van Damme, M. Van Montagu, *Eur. J. Biochem.* **1985**, *152*, 9 – 19; b) R. Aebersold, D. B. Teplow, L. E. Hood, S. B. Kent, *J. Biol. Chem.* **1986**, *261*, 4229 – 4239; c) P. Madsudaira, *J. Biol. Chem.* **1987**, *262*, 10035 – 10038; d) C. Eckerskorn, W. Mewes, H. W. Goretzki, F. Lottspeich, *Eur. J. Biochem.* **1988**, *176*, 509 – 519; e) C. Eckerskorn, P. Jungblut, W. Mewes, J. Klose, F. Lottspeich, *Electrophoresis* **1988**, *9*, 830 – 838; f) M. Ploug, A. L. Jensen, V. Berkholt, *Anal. Biochem.* **1989**, *181*, 33 – 39; g) G. I. Tous, J. L. Fausnaught, O. Akinyosoye, H. Lachland, P. Winter-Cash, F. J. Vitoria, S. Stein, *Anal. Biochem.* **1989**, *179*, 50 – 55; h) S. Nakagawa, T. Fukuda, *Anal. Biochem.* **1989**, *181*, 75 – 78; i) C. Eckerskorn, F. Lottspeich, *Chromatographia* **1989**, *28*, 92 – 94; j) R. Aebersold, J. Leavitt, L. E. Hood, S. H. Kent, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6970 – 6974; k) G. Bauw, M. Van Den Bulcke, J. Van Damme, M. Puype, M. Van Montagu, J. Vandeckerkhove, *Electrophoresis* **1990**, *11*, 528 – 536; l) M. J. Walsh, J. McDougall, B. Wittmann-Liebold, *Biochemistry* **1988**, *27*, 6867 – 6876; m) P. Tempst, A. J. Link, L. R. Riviere, M. Fleming, C. Elicone, *Electrophoresis* **1990**, *11*, 537 – 543; n) S. D. Patterson, D. Hess, T. Youngwirth, R. Aebersold, *Anal. Biochem.* **1992**, *202*, 193 – 203; o) J. Fernandez, M. DeMott, D. Atherton, S. M. Mische, *Anal. Biochem.* **1992**, *201*, 255 – 264; p) F. Lottspeich, C. Eckerskorn, R. Grimm in *Cell Biology: A Laboratory Handbook*, *Vol. 3* (Ed.: J. E. Celis), Academic Press, Orlando, **1994**, pp. 417 – 421.

[5] R. M. Hewick, M. W. Hunkapiller, L. E. Hood, J. Dreyer, *J. Biol. Chem.* **1981**, *256*, 7990 – 7997.

[6] a) L. Anderson, J. Seilhamer, *Electrophoresis* **1997**, *18*, 533 – 537; b) N. L. Anderson, N. G. Anderson, *Electrophoresis* **1998**, *19*, 1853 – 1861.

[7] a) K. Biemann, S. A. Martin, *Mass Spectrom. Rev.* **1987**, *6*, 1 – 76; b) J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **1989**, *246*, 64 – 67; c) K. Biemann, *Annu. Rev. Biochem.* **1992**, *61*, 977 – 1010; d) J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Mass Spectrom. Rev.* **1990**, *9*, 37 – 70; e) M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, *Nature* **1996**, *379*, 466 – 469.

[8] M. Karas, F. Hillenkamp, *Anal. Chem.* **1988**, *60*, 2299 – 2301.

[9] S. Müllner, T. Neumann, F. Lottspeich, *Arzneim. Forsch./Drug Res.* **1998**, *48*, 93 – 95.

[10] a) D. F. Hochstrasser in *Proteome Research* (Eds.: M. R. Wilkins, K. L. Williams, R. D. Appel, D. F. Hochstrasser), Springer, Berlin, **1997**, pp. 187 – 219; b) K. L. Wilkins, V. Pallini in *Proteome Research* (Eds.: M. R. Wilkins, K. L. Williams, R. D. Appel, D. F. Hochstrasser), Springer, Berlin, **1997**, pp. 221 – 232.

[11] a) B. Bjellqvist, J. C. Sanches, C. Pasquali, F. Ravier, N. Paquet, S. Frutiger, G. J. Hughes, D. Hochstrasser, *Electrophoresis* **1993**, *14*, 1375 – 1378; b) T. Rabilloud, C. Valette, J. J. Lawrence, *Electrophoresis* **1994**, *15*, 1552 – 1558; c) J. C. Sanchez, V. Rouge, M. Pisteur, F. Ravier, L. Tonella, M. Moosmayer, M. R. Wilkins, D. F. Hochstrasser, *Electrophoresis* **1997**, *18*, 324 – 327.

[12] T. Rabilloud, *Electrophoresis* **1996**, *17*, 813 – 829.

[13] a) B. Bjellqvist, P.-G. Rhighetti, E. Gianazza, A. Görg, R. Westermeier, W. Postel, *J. Biochem. Biophys. Methods* **1982**, *6*, 317 – 139; b) A. Görg, W. Postel, S. Gunther, *Electrophoresis* **1988**, *9*, 351 – 346.

[14] A. Görg, W. Postel, J. Weser, W. Patutschnig, H. Cleve, *Am. J. Hum. Genet.* **1985**, *37*, 922 – 930.

[15] a) C. R. Merril, J. E. Joy, G. J. Creed in *Cell Biology: A Laboratory Handbook*, *Vol. 3* (Ed.: J. E. Celis), Academic Press, Orlando, **1994**, pp. 281 – 287; b) A. Wallace, H. P. Saluz in *Cell Biology: A Laboratory Handbook*, *Vol. 3* (Ed.: J. E. Celis), Academic Press, Orlando, **1994**, pp. 289 – 298; c) H. Blum, H. Beier, H. J. Gross, *Electrophoresis* **1987**, *8*, 93 – 99; d) H. M. Poehling, V. Neuhoff, *Electrophoresis* **1981**, *2*, 141 – 147.

[16] a) W. Dietzel, G. Kopperschläger, E. Hofmann, *Anal. Biochem.* **1973**, *48*, 617 – 620; b) V. Neuhoff, R. Stamm, H. Eibl, *Electrophoresis* **1985**, *6*, 427 – 448.

[17] T. H. Steinberg, R. P. Haugland, V. L. Singer, *Anal. Biochem.* **1996**, *239*, 238 – 245.

[18] M. Ünlü, M. Morgan, J. S. Minden, *Electrophoresis* **1997**, *18*, 2071 – 1077.

[19] P. Jackson, V. E. Urwin, C. D. Mackay, *Electrophoresis* **1988**, *9*, 330 – 339.

[20] J. N. Weinstein, T. G. Myers, P. M. O'Connor, S. H. Friend, A. J. Fornace, Jr., K. W. Kohn, T. Fojo, S. E. Bates, L. V. Rubinstein, N. L. Anderson, J. K. Buolamwini, W. W. van Osdol, A. P. Monks, D. A. Scudiero, V. N. Viswanadhan, G. S. Johnson, R. E. Wittes, K. D. Paull, *Science* **1997**, *275*, 343 – 349.

[21] a) C. Eckerskorn, K. Strupat, F. Hillenkamp, F. Lottspeich, *Electrophoresis* **1992**, *13*, 664 – 665; b) C. Eckerskorn, K. Strupat, D. Schleuder, D. F. Hochstrasser, J. C. Sanchez, F. Lottspeich, F. Hillenkamp, *Anal. Chem.* **1997**, *69*, 2888 – 2892; c) K. Strupat, M. Karas, F. Hillenkamp, C. Eckerskorn, F. Lottspeich, *Anal. Chem.* **1994**, *66*, 464 – 470; C. W. Sutton, C. H. Wheeler, J. M. Corbett, J. S. Cottrell, M. J. Dunn, *Electrophoresis* **1997**, *18*, 424 – 431.

[22] a) M. Schreiner, K. Strupat, F. Lottspeich, C. Eckerskorn, *Electrophoresis* **1996**, *17*, 954 – 961; b) M. M. Vestling, C. Fenselau, *Anal. Chem.* **1994**, *66*, 47 – 477; c) J. C. Blais, P. Nagnan-Le-Meillour, G.

Bolbach, J. C. Tablet, *Rapid Commun. Mass Spectrom.* **1994**, *5*, 230 – 237; d) S. D. Patterson, *Electrophoresis* **1993**, *16*, 1104 – 1114.

[23] a) F. Hsieh, H. Wang, C. Elicone, J. Mark, S. Martin, F. Regnier, *Anal. Chem.* **1996**, *68*, 455 – 462; b) T. Houthaeve, H. Gausepohl, M. Mann, K. Ashman, *FEBS Lett.* **1995**, *376*, 91 – 94.

[24] S. D. Patterson, R. H. Aebersold, *Electrophoresis* **1995**, *16*, 1791 – 1814.

[25] B. Spengler, D. Kirsch, R. Kaufmann, E. Jaeger, *Rapid Commun. Mass Spectrom.* **1992**, *6*, 105 – 108.

[26] a) K. J. Eng, A. L. McCormac, J. R. Yates III, *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976 – 989; b) J. R. Yates III, *Electrophoresis* **1998**, *19*, 893 – 900.

[27] M. Mann, M. Wilm, *Anal. Chem.* **1994**, *66*, 4390 – 4399.

[28] M. Schnölzer, P. Jedrzejewski, W. D. Lehmann, *Electrophoresis* **1996**, *17*, 945 – 953.

[29] a) K. R. Jonscher, J. R. Yates, *Anal. Biochem.* **1997**, *244*, 1 – 15; b) R. E. March, *J. Mass Spectrom.* **1997**, *32*, 351 – 369.

[30] R. Kratzer, C. Eckerskorn, M. Karas, F. Lottspeich, *Electrophoresis* **1998**, *19*, 1910 – 1919.

[31] a) B. Bjellqvist, G. Hughes, C. Pasquali, N. Paquet, F. Ravier, J. C. Sanchez, S. Frutiger, D. Hochstrasser, *Electrophoresis* **1993**, *14*, 1023 – 1031; b) B. Bjellqvist, B. Basse, E. Olsen, J. E. Celis, *Electrophoresis* **1994**, *14*, 1023 – 1031.

[32] a) C. Eckerskorn in *Bioanalytik* (Eds.: F. Lottspeich, H. Zorbas), Spektrum, Heidelberg, **1998**, pp. 323 – 368; b) J. W. Metzger, C. Eckerskorn, C. Kempter, B. Behnke in *Microcharacterization of Proteins* (Eds.: R. Kellner, F. Lottspeich, H. E. Meyer), 2nd ed., WILEY-VCH, Weinheim, **1999**, pp. 213 – 234.

[33] a) H. E. Meyer in *Microcharacterization of Proteins* (Eds.: R. Kellner, F. Lottspeich, H. E. Meyer), 2nd ed., WILEY-VCH, Weinheim, **1999**, pp. 159 – 175; b) A. A. Gooley, N. H. Packer, in *Proteome Research* (Eds.: M. R. Wilkins, K. L. Williams, R. D. Appel, D. F. Hochstrasser), Springer, Berlin, **1997**, pp. 65 – 91.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 2149–2159

Review

Ruth A. VanBogelen[1]
Erin E. Schiller[1]
Jeffrey D. Thomas[1]
Frederick C. Neidhardt[2]

[1]Molecular Biology Department, Parke-Davis Pharmaceutical Research, Division of Warner-Lambert Company, Ann Arbor, MI, USA
[2]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA

Diagnosis of cellular states of microbial organisms using proteomics

Two-dimensional (2-D) polyacrylamide gel electrophoresis has much to contribute to experimental analysis of the proteomes of microbial organisms, since this method separates most cellular proteins and allows synthesis rates to be determined quantitatively. Databases generated using 2-D gels can grow to be very large from even just a few experiments, since each sample provides the data for a field (or column) in the database for several hundreds to even thousands of records (or rows), each of which represents a single polypeptide species. The value of such databases for generating an encyclopedia of how each of the cell's proteins behave in different conditions (protein phenotypes) has been recognized for some time. The potential exists, however, to glean even more valuable information from such databases. Because the measurements of each protein are made in the context of all other proteins, a comprehensive glimpse of the cell's physiological state is theoretically achievable with each 2-D gel. By examining enough conditions (and 2-D gels), expression patterns of subsets of proteins (proteomic signatures) can be found that correlate with the cell's state. This type of information can provide a unique contribution to proteomic analysis, and should be a major focus of such analyses.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Physiology / *Escherichia coli* / Bacteria / Review
EL 3466

Contents

Correspondence: Dr. Ruth A. VanBogelen, Parke-Davis Pharmaceutical Research, Division of Warner-Lambert Company, 2800 Plymouth Rd., Ann Arbor, MI 48105, USA
E-mail: vanbogr@aa.wl.com
Fax: +734-622-5970

1 Introduction

For many microorganisms, research studies are now in the postgenomic era. Their DNA sequences are known and have been analyzed with several computational tools to extract theoretical information. Beyond this rather static description of the cell are the dynamic "transcriptome" and "proteomic" analyses, often referred to as functional genomic studies. Transcriptome refers to the transcript (mRNA) profile in the cell [1]. Proteomics deals with the variations in translation (protein) activity in the cell. Prior to 1975 cellular proteins were studied as individual proteins with biochemical methods, or as bulk cellular protein (*e.g.*, the total quantity of cell protein, or its total rate of synthesis) by chemical or radiochemical methods. In 1975 two-dimensional (2-D) polyacrylamide gel electrophoresis was introduced as a method to separate complex mixtures of cellular protein into individual polypeptides [2]. This method is useful for studying a cell's protein architecture, namely, the amount and subcellular location of individual proteins [3], but has also been used in many studies to examine changes in protein expression in re-

sponse to a stimulus [4], or to identify target genes of regulatory proteins [5]. In all of these applications, 2-D gels are useful both to find proteins of interest and to isolate them. These applications have made, and will continue to make, contributions to our understanding of how proteins function in the cell. It should be recognized that these applications of 2-D gels, while employing a global approach, yield information that is basically reductionist in nature. In contrast, this article describes how proteomics can be used to provide information of a highly integrative nature, specifically, to depict the physiological state of the cell.

The outlook of this application of proteomics is so global that the identity and function of the proteins that provide the diagnostic clues, while helpful, need not be known. The underlying concept is to find proteins that respond specifically and uniquely to particular states of the cell, that is, to certain fundamental metabolic conditions such as energy sufficiency, redox state, building block supply, translational capacity, envelope integrity, protonmotive force, and chromosome copy. Each such set of proteins, termed a proteomic signature, relates to the current working status of a particular core metabolic process. The physiological state of the cell in a particular circumstance is read from the ensemble of the proteomic signatures displayed. Naturally, the trick is to discern what specific signatures are represented within the complete protein expression profile, and to have a table that relates these signatures to their cognate cellular functions. Progress in this field will rely on the construction of this table of correspondences, linking signatures to core cellular functions. This task will require much data gathering on the responses of cells to different stimuli, and the thoughtful analysis of these data in the light of current biochemical and genetic information about the cell and its growth response. The goal of this publication is to illustrate some correlations between protein expression and physiology that have already been established in *Escherichia coli*. Although the examples used here have been recognized by manual image searching methods, the hope is that as more investigators use proteomics to find these correlations, better tools for discovery and analysis will be developed.

2 Definitions

In 1994, nearly twenty years after 2-D gel methodology was first introduced, the term proteome was introduced to align the work being done on complex mixtures of proteins with genomics [6]. Many derivations of the word proteome have been suggested and are useful to describe the analysis of proteomes. The following words and phrases will be used throughout this presentation and are defined below.

The proteome of an organism is the complete set of proteins that it can produce. To date no proteome (other than viral) has been unambiguously determined, but many have been predicted from the genome sequence. For example, a proteome predicted by Blattner [7] for *E. coli* MG1655 consists of 4405 proteins. This proteome can be displayed in the manner shown in Fig. 1. This view shows the range of p*I* and M_r values for the proteins of an organism and therefore can help in deciding what the appropriate 2-D gel composition is to optimally separate the organism's proteins. Further, this manner of displaying an organism's proteome can reveal interesting characteristics. For example, proteomes appear to have a bimodal isoelectric distribution, probably related to a relationship between the p*I* of proteins and the intracellular pH of the cell (A. Persemlidis, R. VanBogelen, unpublished observations).

Proteomics is the study of the proteome of cells, tissues, or organisms including the interaction of the proteome with the environment.

Protein expression profile is the quantitative catalog of proteins made by cells under a given circumstance. For example, each 2-D gel or set of 2-D gels run from a single sample reveals most of a particular protein expression profile. Obtaining a complete protein expression profile is

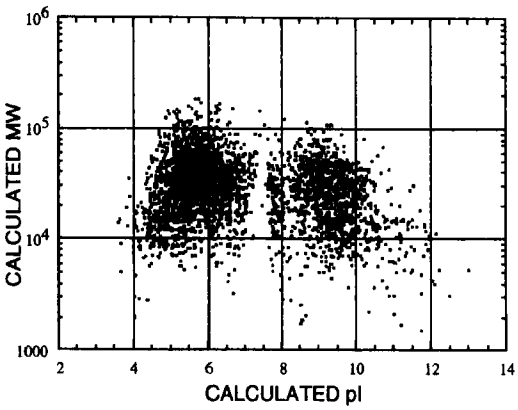


Figure 1. Graphical representation of a proteome. This figure plots the predicted isoelectric point (p*I*) *versus* the predicted molecular mass (M_r) of predicted open reading frames for *E. coli*. The p*I* and M_r of 4405 ORFs were predicted using the Compute p*I*/M_r in ExPAsy (http://www.expasy.ch/ch2-D/pi_tool.html). Each dot represents a protein. No post-translational modifications are represented in this figure. This display of *E. coli, H. influenzae, M. genitalium, M. pneumoniase, M. jannaschii* and *Synechocystis* sp. was previously published [20, 24–26].

not technically possible at this time due to difficulties associated with protein solubility, extreme p*I* and low abundance.

Protein phenotype is the character or state of a specific protein in a defined condition and point in time. This term was introduced by J. Klose [8]. Information in the phenotype of a protein includes: the amount, the rate of synthesis, the rate of degradation, the extent of post-translational modification, and the activity of the protein. For example, the protein phenotype of alkaline phosphatase in *E. coli* includes all of this information about the protein for each growth condition of the cell. Protein phenotypes can include data from many sources. The most comprehensive collection of protein phenotypic data for one organism has been compiled by Garrels *et al.* [9] for *Saccharomyces cerevisiae*.

Regulon is a set of proteins whose synthesis is regulated by the same regulatory protein [10]. Genetic analysis has been the classical way to find and study regulons, but proteomics can be used to help find members of regulons. For example, comparing the protein expression profiles of wild-type and mutant strains can reveal proteins with the expression characteristics of a regulon member [5, 11].

Stimulon is a set of proteins whose amount or synthesis rate changes in response to a single stimulus [10]. Several stimulons have been described. They can range in size from a handful of proteins to nearly half of the proteins found in a cell. A stimulon may consist of several regulons.

Proteomic signature is the subset of proteins whose alteration in expression is characteristic of a response to a defined condition or genetic change. Signatures often relate to specific pathways or functions. There can be several signatures within a protein expression profile and, therefore, proteomic signatures cannot be recognized simply by comparing two protein expression profiles. Signatures are recognized after analyzing numerous profiles of samples from related and unrelated conditions. Signatures can be a single protein or a large set of proteins. Several examples of proteomic signatures will be described in Section 4.

3 The use of protein expression profiles

Correlating protein expression with cell physiology entails the comparison of protein expression profiles. Table 1 lists steps in this process. Completion of these steps provides the information helpful in finding protein phenotypes, stimulons, regulons, and proteomic signatures. Much of the assessment in finding these characteristics deals with the changes in the quantity or synthesis rate of each protein. For example, a protein could be considered responsive based on its qualitative change (absence or presence) or its quantitative change. Another type of assessment involves relationship changes. The abundance ratio between two proteins in a single protein expression profile or between two subunits of a single protein might change in a characteristic manner (see Section 4.1 and 4.2).

Protein phenotypes, stimulons, regulons, and proteomic signatures can and have been found using simple image analysis methods (gel gazing) as well as by using comprehensive computer-assisted image analysis systems. The depth of the analysis should be consistent with the goals of the project. A comprehensive analysis of every protein induced or repressed 1.5-fold in one condition compared to another may be excessive if the goal is to find a protein marker for a survey or screen. On the other hand, such a comprehensive analysis would be justified if the goals were to identify all the proteins controlled by a certain regulatory protein. The goals also help to define what information is important for the project. In some cases the goal is to define additional phenotypes of a particular protein. In other cases, stimulons and regulons need to be identified to help reveal the signature(s) that will provide the necessary information to understand a cell's reaction to a stimulus.

Additional phenotypes of a protein are unveiled each time new observations are made. Even if a protein has not yet been linked to a gene, the protein's phenotype can reveal information about the protein. For example, in a pair of studies done in the late 1970s, proteins were grouped by their phenotypes according to their expression in different growth media and at different growth temperatures [12, 13]. At the time of publication few of these proteins were identified, but a generalized function could be assigned based on their phenotypes. (Note that the spot names used in these publications are still in use in the *E. coli* gene-protein database [14]). Examples of a few interesting protein phenotypes found using proteomics are described in Section 4.

Stimulons are directly identified using protein expression profiles. Comparing the protein expression profile of a control condition with that seen in a test condition will reveal the stimulon for the test condition. Inclusion of proteins in the stimulon will depend on the sophistication of the analysis. In a series of qualitative studies done on *Vibrio* in the 1980s, several stimulons were identified and related to potential function for the cell [4]. In this study not a single protein was identified as the product of a specific gene, and yet much physiological data and information was obtained.

Table 1. Steps in Proteome Analysis

	Steps	Description
1	Identify project goals	
2	Establish knowledge base of physiology and cell biology	Examine variables such as required nutrients, optimal temperature, and growth rate
3	Predict the theoretical proteome	Using DNA sequence data, predict open reading frames (ORFs), predict p*I* and M_r of the expected gene products of the genome, and predict possible regulons or stimulons
4	Prepare samples	Release, fractionate and/or solubilize proteins from the cells into buffers compatible with 2-D gels
5	Produce high quality 2-D gels	Run the first dimension to separate proteins by their p*I*, run the second dimension to separate proteins by their M_r
6	Detect proteins	Use stains (coomassie blue, silver or fluorescent) to visualize non-radioactive proteins, use films or phosphorimager screens to visualize radioactively labeled proteins
7	Acquire images	Transfer gel images to a computer *via* a scanner or camera
8	Perform visual exam of gels "gel gazing"	Use the human eye to look for changes between conditions, determine if there are enough significant changes to proceed with computer-assisted analysis
9	Analyze images	Find spots, edit spot finding, assign anchors between images, match images, edit match, transfer data to tables for further editing
10	Determine data statistics (average %SE, fold changes, distribution curves, pie charts)	Determine % standard error between spots on duplicate gels, look at proteins that change 2-fold or more, use distribution curves and pie charts to look at general trends in the entire population of proteins
11	Verify data (reexamine 2-D gels)	Verify that the data in tables agrees with the spots seen on the 2-D gels
12	Match to reference set (links to pervious experiments)	Match 2-D gels to a reference image containing other data and protein identifications
13	Identify proteins	Excise protein spots from a gel, digest them with trypsin and run them on MALDI-MS (to determine a mass fingerprint) and ES-MS/MS (to determine a fragmentation sequence)
14	Compare stimulons (clustering, neural networks, decision trees, etc.)	Study populations of data
15	Identify proteomic signatures (results of above and human input)	Use knowledge of the effect of the treatment on the cells to link protein phenotypes, stimulons, and regulons
16	Visualize data	Devise simple visual presentation of complex data sets
17	Integrate information into a systems analysis model	Build and refine mathematical models of cell function using data from protein expression profiles

Regulons are usually identified in one of two ways: (i) using strains in which the gene for the regulatory protein is mutated; and (ii) examining the response to multiple conditions if a regulon is thought to be the sole intersection of the responses. The heat shock regulon was first identified using protein expression profiles [5]. In this case, a mutant strain revealed the regulon set by comparing the protein expression profile of the parent and mutant strain after the mutant had been shifted to a nonpermissive temperature. In a recent study, potential members of the Pho regulon in *E. coli* were quantitatively identified by comparing three stimulons: early phosphate starvation, late phosphate starvation, and phosphorus restriction [15]. The intersection of these three stimulons, which contains about 150 proteins, is predicted to include members of the Pho regulon. For eight of the proteins, a promoter sequence similar to the consensus sequence for binding of the regulator, PhoB, was found upstream of the gene that encoded the protein. Most of the other 150 proteins have not yet been identified.

Proteomic signature is a new term presented in this publication. Although the term is new, we have been identifying proteomic signatures for over twenty years. Section 4 describes some of the *E. coli* proteomic signatures we have found. Much of the current effort with the ECO2DBASE database is designed around generating the data to identify proteomic signatures related to as many physiological states of the cell as possible. The goal is to use these proteomic signatures to diagnose the physiological

state of the cell when the target of a stimulus or function of a mutant protein is unknown. To date, we have used manual methods for finding signatures, but it is expected that more signatures can be found within these same data sets through the use of more comprehensive, machine-based tools. Computational tools such as heuristic clustering and neural networks have been used by other disciplines to scan large data sets for information, and hopefully some of these same tools will be used to study the large data sets that are generated by proteomics.

4 Examples of protein phenotypes and proteomic signatures

Studies of the protein expression profiles of *E. coli* over the past twenty-five years have defined fairly comprehensive phenotypes of many proteins. The phenotypes of RplL, MetE, and PhoE are described in this section. The same work has identified many important signatures of physiological states; some involve only a single protein, others are more complex. Although in many of the examples described below, the identity and often a function of the protein is known, in almost all cases, this information does not contribute to the use of the protein to diagnose cellular states and in most cases does not help us to understand the cell's response.

4.1 Protein phenotype for RplL

The protein phenotype of the ribosomal protein RplL is similar to that of other ribosomal proteins in that the amount of the protein (the modified and unmodified forms combined) varies coordinately with the other ribosomal proteins as the growth rate varies, but is more complex due to a post-translational modification that occurs. The amounts of the modified, L7, and unmodified, L12, forms of this protein do not follow coordinately with other ribosomal proteins [12]. The protein phenotypes of RplL provide useful information about this protein and about the cell's ribosome, but perhaps the phenotypes are most useful as part of the proteomic signature described in the next paragraph.

4.2 Proteomic signature for growth rate

The protein phenotype of ribosomal protein RplL described above has been recognized as a proteomic signature for growth rate. The ratio of L7 to L12 varies with growth rate as shown in Fig. 2. This figure shows the results of a study in which the amount of these two proteins (A013.0 and B013.0 in ECO2DBASE [14]) was determined in cultures growing with different carbon sources [12]. No biological function for controlling the degree of modification level of this ribosomal protein as a function of growth rate has yet been determined. These two spots on a 2-D gel are identified, but their identity and even their function (known to date) do not help in the understanding of why they can be used as a proteomic phenotype for growth rate. Nonetheless, the L7/L12 ratio is so highly correlated with growth rate that it can serve as a surrogate measure to obtain a crude indication of the growth rate.

This proteomic signature is not useful for general assessment of the growth rate in laboratory studies where growth rate can be measured by more traditional methods (spectrophotometer). Two obvious applications are, first, as an internal monitor, and second, in cases where traditional methods cannot be applied. An example of the first application is in a bioreactor or chemostat. Under these conditions, overall growth of the culture can be monitored with a spectrophotometer; however, the growth rate of the active population is not known. An example of the second application has been previously published to estimate the growth rate of *Salmonella* growing in macrophages [16]. The *Salmonella typhimurium* homolog of *rplL* has a very similar protein phenotype and also provides a proteomic signature for growth rate [16]. In this study, macrophage cells were infected with *Salmonella*, and cellular proteins were radioactively labeled with [^{35}S]methionine (only metabolically active cells take up the [^{35}S]methionine and incorporate it into newly synthesized protein). The cellular extracts were then run on 2-D gels and the ratio of L7/L12 was determined. This ratio predicted a much faster

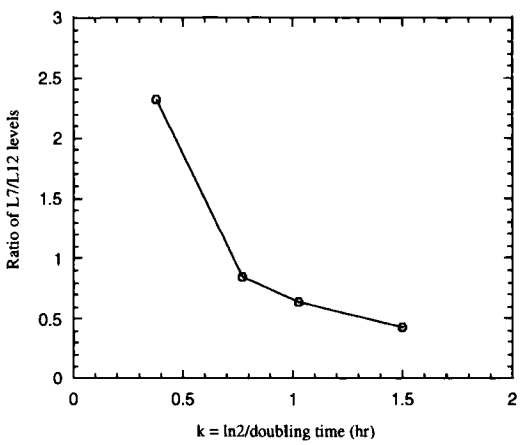


Figure 2. The proteomic signature for growth rate. This figure plots the growth rate of *E. coli versus* the ratio of the amount of ribosomal protein L7 and ribosomal protein L12. The growth rate was varied due to growth in different carbon sources. The data are published [12].

growth rate than a measurement based on the number of viable *Salmonella* extracted from the macrophage cells. Which measurement is the accurate reflection of the growth of *Salmonella* in macrophage cells? Another study using ampicillin and chloramphenicol to determine the number of bacteria actively growing suggested that there were two populations of cells in the macrophage, one static and the other actively growing, thereby confirming the prediction based on the L7/L12 ratio.

4.3 Protein phenotype for MetE

The product of the *metE* gene, F088.0 in ECO2DBASE [14], is the enzyme 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase. It has an unusual phenotype. Although all of the enzymes in the methionine biosynthetic pathway are expressed during growth in the absence of methionine and repressed in its presence, the amount of MetE in media without methionine is extraordinarily high in comparison to the other enzymes in this biosynthetic pathway. The phenotype of MetE is not understood but one possibility is that it is a bifunctional enzyme.

4.4 Proteomic signature for the availability of methionine

The MetE protein phenotype has been found to be a useful proteomic signature. This example also illustrates that the identity and known function of the protein does not provide further information into the cell's physiology. How is this high expression level of MetE useful as a diagnostic indicator? If an experiment is designed to have a long-term labeling with [^{35}S]methionine, then the specific activity of the [^{35}S]methionine is lowered (by supplementing with unlabeled methionine) to ensure that incorporation continues throughout the labeling period. If the MetE protein in the resulting 2-D gels was undetectable, then the labeling protocol was successful (methionine was available throughout the labeling). However, if MetE expression was high, it was apparent that the culture had run out of methionine during the labeling period. The resulting protein expression profile, therefore, would not be an accurate reflection of steady state growth.

4.5 Proteins with distinctive temperature phenotypes

There are three known distinctive sets of proteins that show unique temperature phenotypes. The first set is made up of three *E. coli* proteins which seem to behave as cellular thermometers (G027.2, D074.0 and G044.0; G044.0 has been identified as tyrosinyl-tRNA synthetase). The amounts of these proteins change in cultures grown at different temperatures (except in the range from 23°C to 37°C) [13]. The second set contains proteins that have been identified as being highly inducted at the extremes of growth temperature. G029.7 is induced greater than 10-fold at 13.5°C relative to 37°C; and A165.0 is induced greater than 10-fold at 46°C compared to 37°C. A third set of proteins with distinctive temperature phenotypes consists of the heat shock and cold shock proteins. Their synthesis rate is known to dramatically increase after shifts to temperatures outside the linear portion of the Arrhenius plot [5]. Their amounts also increase to varying degrees at the temperature extremes. However, both the synthesis rate and amount of these proteins respond to stimuli other than just heat and cold. In fact, these other phenotypic characteristics are used as signatures described in Section 4.7 and 4.8.

4.6 Proteomic signature for growth temperature

The temperature range for growth of *E. coli* is from 8°C to 48°C. An Arrhenius plot of the growth temperature *versus* the growth rate shows that in the range 23°C to 37°C the logarithm of growth rate varies linearly with the reciprocal of the absolute temperature as happens for some simple chemical reactions [13]. Over this linear range the quantities of most proteins stay constant, providing no clear proteomic signature to distinguish between 23 and 37°C. However, there are proteomic signatures for the temperature extremes. These consist of the first two sets of proteins (G027.2, D074.0, G044.0, G029.7 and A165.0) described in the previous section. The amounts of some heat and cold shock proteins do increase at the temperature extremes. However, these proteins are also known to be induced by other stress conditions that interfere with their usefulness as indicators of the growth temperature (see Section 4.7 and 4.8). Nonetheless, because most organisms have HSP60 and HSP70 homologies, they can be used to predict whether an organism is growing above its "normal" range of growth. For example, examination of 2-D gels of cultures of *Legionella* that were grown at 37°C revealed two proteins with approximate masses of 60 and 70 KDa that were among the most abundant proteins detected (VanBogelen, unpublished observation). Based on gel location these proteins are most likely the heat shock proteins HSP60 and HSP70. Because the normal host for *Legionella* is the amoebae, the normal temperature range for *Legionella* is probably much lower than for *E. coli*. The high quantity of the HSP60 and HSP70 is consistent with 37°C being a high temperature for *Legionella*.

4.7 Proteomic signature for ribosome functionality

The heat and cold shock proteins also provide a proteomic signature for ribosome function. Antibiotics that target

Table 2. Proteins that contribute to the proteomic signatures for growth temperature, ribosome function and protein secretion

Protein[a]		High temperature	Low temperature	Heat shock-like ribosome dysfunction	Cold shock-like ribosome dysfunction	Protein secretion dysfunction
G027.2		Decrease[b]	Increase[b]			
D074.0		Decrease	Increase			
G044.0	TyrS		Decrease			
G029.7			Increase			
A165.0		Increase				
B025.3	GrpE			Increase	Decrease	
B056.5	GroL	Increase		Increase	Decrease	Increase
B066.0	DnaK	Increase		Increase	Decrease	Increase
C014.7	IbpB			Increase	Decrease	Increase
C015.4	GroS			Increase	Decrease	Increase
C062.5	HtpG			Increase	Decrease	Increase
D033.4	HtpH			Increase	Decrease	Increase
D048.5	HtpI			Increase	Decrease	
D060.5	LysU			Increase	Decrease	
E072.0	ClpB			Increase	Decrease	
F010.1	HtpK			Increase	Decrease	Increase
F021.5	ClpP			Increase	Decrease	
F084.1	ClpB			Increase	Decrease	Increase
G013.5	IbpA			Increase	Decrease	
G021.0	HtpO			Increase	Decrease	
H094.0	Lon			Increase	Decrease	
B065.0	RpsA			Decrease	Increase	
A013.0	RplL			Decrease	Increase	
B013.0	RplL			Decrease	Increase	
B046.5					Increase	
B061.0	NusA				Increase	
C039.3	RecA				Increase	
C070.0	AceF				Increase	
C078.0	Pnp				Increase	
F010.6	CspA				Increase	
F014.7	H-NS				Increase	
F084.0	Ppc				Increase	
F099.0	AceE				Increase	
F119.0	InfB				Increase	
G041.2					Increase	
G055.0					Increase	
G074.0	PflB				Increase	
G117.0	InfB				Increase	
C015.3	RpsF			Decrease	Increase	
E042.0	TufA			Decrease	Increase	
D084.0	FusA			Decrease	Increase	
C031.6	Tsf			Decrease	Increase	
B038.1	PhoE-pre					Increase
B040.4	OmpF-pre					Increase
A040.0	OmpC-pre					Increase
F040.9	Male-pre					Increase

a) The ECO2DBASE spot name (alpha-numeric) is listed in the first column and the name of the gene is used as the protein name in the second column.

b) Increase and decrease indicate the direction of change in the level or synthesis rate of protein by stimuli that have the indicated proteomic signature(s).

the ribosome induce either the heat shock proteins (specifically the sigma-32 regulated proteins) or the cold shock proteins [17]. The antibiotics streptomycin, puromycin, and kanamycin induce the heat shock proteins (HSPs) in a concentration-dependent manner. The higher the concentration of antibiotics used, the greater the induction of the HSPs. Some proteins induced by heat shock that are not regulated by sigma-32 were not induced by the antibiotics. Thus, it appears that the sigma-32 regulon contributes to the heat shock signatures, but not exclusively (see Table 2). In wild-type strains (for the *relA* gene), a heat shock also causes the cells to go into a stringent mode of gene regulation. Chemically, the stringent response is triggered by an increase in the amount of ppGpp [18]. Transcriptionally, it is seen as a downregulation of numerous genes, including the genes encoding ribosomal RNAs and translation proteins, that leads to a decrease in the amount of ribosomal proteins and translation factors. This proteomic signature for the stringent response is seen also with this set of antibiotics. Therefore, two signatures are evoked by these antibiotics: the sigma-32 regulon signature, and the stringent signature.

Other antibiotics, namely tetracycline, chloramphenicol, erythromycin, fusidic acid, and spiramycin, elicit a cellular response similar to the response induced by a shift to a cold temperature (below 23°C for *E. coli*; see Table 2). This response includes: (i) induction of a set of proteins called cold shock proteins; (ii) repression of the heat shock proteins; (iii) induction of ribosomal proteins (the proteomic signature for the "relaxed" state); and (iv) repression of many other proteins (with clustering algorithms this general set of repressed proteins might become recognized as a defined signature). Although the specific regulons for this cold shock response have not been determined, the current descriptions of the signatures for these antibiotics and for cold shock include: (i) the cold shock protein signature; (ii) the repressed heat shock protein signature; and (iii) the relaxed signature.

4.8 Proteomic signatures for dysfunction of protein secretion

Blockage of protein secretion generates three proteomic signatures (see Table 2) (unpublished observations R. A. VanBogelen and E. R. Olson, manuscript in preparation). First, a subset of the heat shock proteins are induced (see Table 2); second, PspA is induced (see Fig. 3); and third, the precursor forms of secreted proteins can be detected by pulse-chase labeling of the culture after secretion is inhibited (see Fig. 3). The presence of these three signatures in the same protein expression profile predicts that protein secretion is dysfunctional.

4.9 Proteomic signature for phosphorus source

The preferred phosphorus source for *E. coli* is inorganic phosphate, but *E. coli* can use alternative phosphorus sources, largely due to the induction of proteins in the Pho regulon [19]. Using genetic methods, about 30 genes have been identified as members of this regulon [19]. PhoA (alkaline phosphatase) has been the classic gene product used for genetic studies for two reasons: (i) transcription of *phoA* is induced greater than 1000-fold in response to phosphorus limitation and (ii) alkaline phosphatase activity can be easily measured with a colorimetric assay. Using proteomics, about 150 proteins appear to belong to this regulon [15]. This includes 19 proteins that are repressed (no repressed genes had previously been identified as members of this regulon). A complete signature for phosphorus restriction includes all of these proteins, yet PhoE alone appears to be a sufficient indicator that the Pho regulon is either activated or deregulated. This protein is not detectable when inorganic phosphate is present in the medium and when the cells have functioning PhoR/PhoB proteins [19]. If the medium is depleted of inorganic phosphate, if the cells are using an alternate phosphorus source, or if the PhoR protein is inactivated [20], then the PhoE protein is the most visually obvious indicator.

The phosphorus restriction signature was used in a study designed to find chemical inhibitors of histidine kinases [20]. In partnership with other proteins, histidine kinases are found in two-component regulatory systems. Bacterial genomes encode many of these two-component regulators which control systems for adaptation, virulence, *etc.* [21]. PhoR is the histidine kinase involved in regulating the Pho regulon, which includes *phoE* (PhoB is the second component, the response regulator, that works with PhoR). Compounds were initially identified using a purified histidine kinase (NRII) and the activity of the histidine kinase was monitored by its ability to autophosphorylate. Validation of the active compounds was difficult. For example, alkaline phosphatase assays were not useful because looking for loss of activity in a cell that already contained alkaline phosphatase prior to the addition of the compound could at most reveal a 50% decrease in activity per generation. However, a cell that can no longer maintain an induced quantity of the Pho regulon will not grow in media containing a phosphorus source other than inorganic phosphate. Growth inhibition alone was also insufficient to validate the target because of the possibility of inactivating some other target unrelated to the Pho regulon. The Pho regulon proteomic signature, however, was an effective validation test. By measuring rates of synthesis of proteins after addition of the compound, both the inactivation of the phosphatase and histidine kinase activity of PhoR were monitored *via* the synthesis of PhoE [20].

4.10 Predictions of chromosome functionality

Chromosomal DNA serves two major functions in cells: it serves as both the template for replication and the template for transcription. There are many cellular activities (such as DNA repair and chromosome segregation) that ensure that the chromosome is properly maintained so that it can provide these basic functions [22]. The protein phenotype for one cellular protein, RecA, includes an induced state in all conditions that are known to have a compromised chromosome structure or function [23]. Figure 4 shows the induction of RecA protein in cultures treated with six different chemicals. While this protein can be used as an initial indicator of loss of chromosome function, closer examination of these six conditions have revealed specific signatures (E. E. Schiller and VanBogelen, manuscript in preparation). Twenty-two proteins were found (by visual examination of 2-D gels) that can serve as indicators to separate these six agents into four groups. A more comprehensive analysis of all proteins detectable on these images is in progress and will further define these different signatures. These signatures have been successfully used to distinguish the mechanism of action of novel compounds that are all known to induce the SOS regulon (E. E. Schiller and R. A. VanBogelen). Because some of these targets are unique to microbial organisms while some are found in all living things, these signatures help ensure that compounds that could have detrimental effects on mammals are recognized early on in pharmaceutical development.

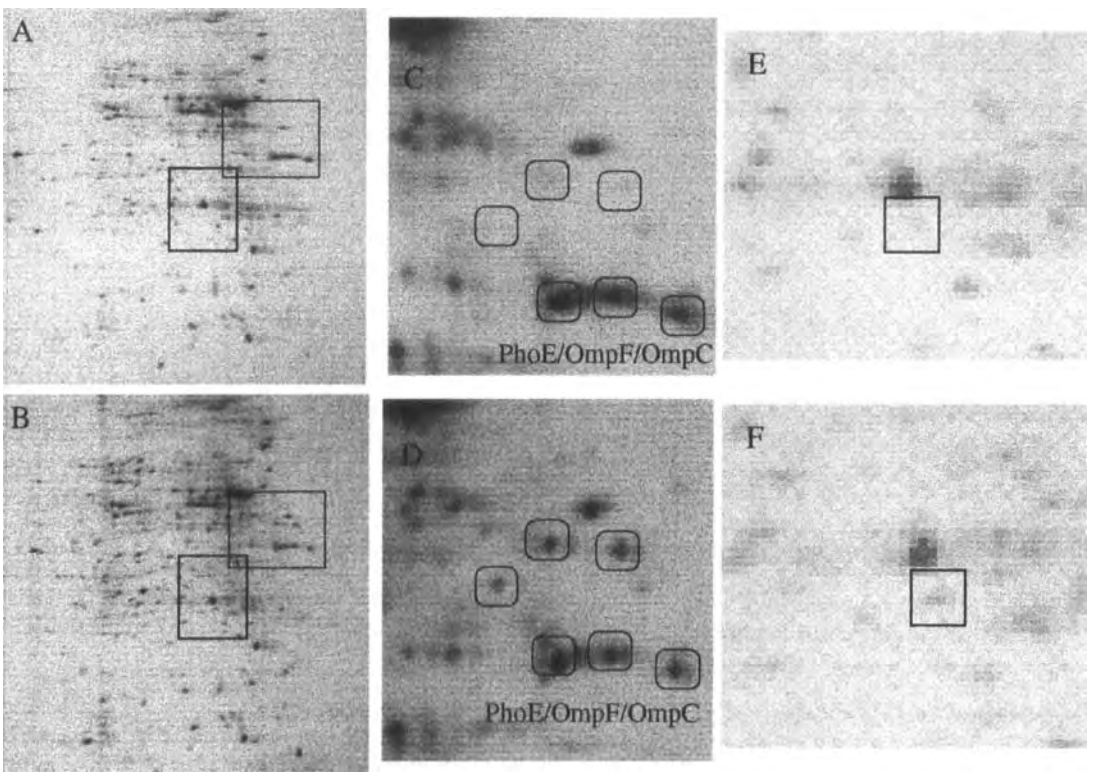


Figure 3. Proteomic signatures for protein secretion. The protein expression profiles (2-D gels) shown in this figure represent the profiles seen with agents or conditions known to block protein secretion (*e.g.*, sodium azide, carbonyl cyanide *m*-chlorophenylhydrazone). (A) and (B) display the entire 2-D gels for both (A) the control and (B) the treated cells. (C) Control and (D) treated cells show the precursor and mature forms of three outer membrane porin proteins (OmpF, OmpC and PhoE). (E) Control and (F) treated cells display the induction of the PspA protein. These 2-D gels were run on the Investigator system (Genomic Solutions) with pH 4–8 Ampholines in the first dimension and 11.5% duracryl in the second dimension. The samples were from ^{35}S-labeled proteins of *E. coli*.

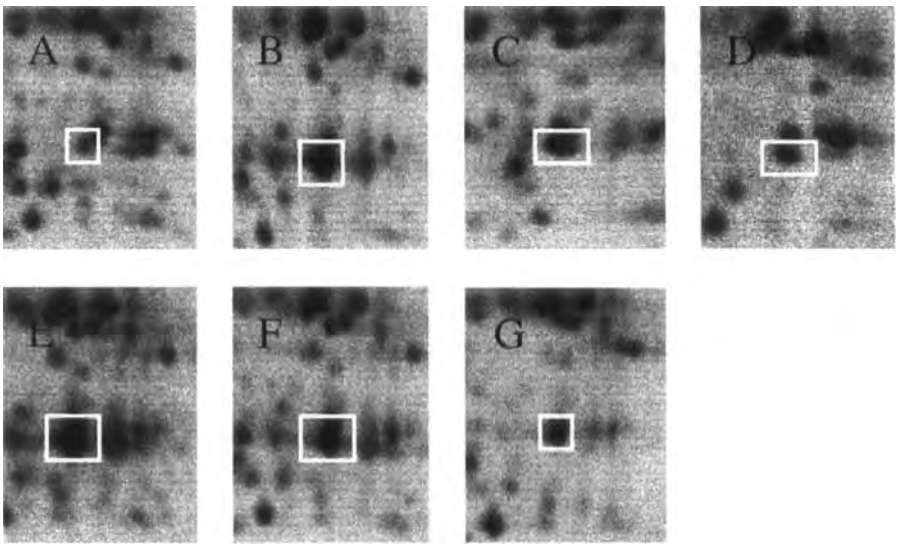


Figure 4. Induction of RecA protein by agents that affect chromosome function or structure. This figure displays a small section of 2-D gels (run as described in Fig. 3) showing the RecA protein. (A) Control sample; (B) sample treated with nalidixic acid; (C) sample treated with novobiocin; (D) sample treated with 9-aminoacridine; (E) sample treated with ciprofloxacin; (F) sample treated with mitomycin *c*; and (G) sample treated with coumermycin. The RecA protein is indicated in the white box.

5 Discussion

The examples presented in the previous section illustrate how one can organize the extensive information contained in the protein expression profiles of microbial (and other) cells under various conditions. Protein phenotypes display the key physiological, biochemical, regulatory, and architectural features of each cellular protein. These features include not only the individual properties and activities of each protein taken singly, but also the manner in which its function is integrated with that of other proteins in the cell. Proteomic signatures specify the key protein components of unique physiological states brought on by specified environmental agents or genetic states. Thus, signatures provide useful information for diagnosing the physiological state (*e.g.*, hunger for nitrogen, oxidative toxicity, DNA damage, *etc.*) of cells in a given environmental and genetic circumstance. Together, phenotypes and signatures provide the microbial physiologist with powerful proteomic tools with which to solve the biological function of unknown proteins, to discover the mode of action of unstudied physical and chemical agents, and to approach modeling the integrated workings of the intact cell.

Received January 7, 1999

6 References

[1] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, Jr., D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[2] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[3] Fountoulakis, M., Lagen, H., Evers, S., Gray, C., Takacs, B., *Electrophoresis* 1997, *18*, 1193–1202.

[4] Nystrom, T., Albertson, N. H., Flardh, K., Kjellegerg, S., *FEMS Microbiol. Ecol.* 1990, *74*, 129–140.

[5] Neidhardt, F. C., VanBogelen, R. A., *Biochem. Biophys. Res. Comm.* 1981, *100*, 894–900.

[6] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A. Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[7] Blattner, F. R., Plunkett, III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y., *Science* 1997, *277*, 1453–1474.

[8] Klose, J., *Electrophoresis* 1999, *20*, 643–652.

[9] Hodges, P. E., Payne, W. E., Garrels, J. I., *Nucleic Acids Res.* 1998, *26*, 68–72.

[10] Neidhardt, F. C., Ingraham, J. L., Schaechter, M., *Physiology of the Bacterial Cell: A Molecular Approach*, Sinauer Publishing, Sunderland, MA 1990.

[11] Ernsting, B. R., Atkinson, M. R., Ninfa, A. J., Matthews, R. G., *J. Bacteriol.* 1992, *174*, 1109–1118.

[12] Pedersen, S., Bloch, P. L., Reeh, S., Neidhardt, F. C., *Cell* 1978, *14*, 179–190.

[13] Herendeen, S. L., VanBogelen, R. A., Neidhardt, F. C., *J. Bacteriol.* 1979, *139*, 185–194.

[14] VanBogelen, R. A., Abshire, K. Z., Pertsemlidis, A., Clark, R. L., Neidhardt, F. C., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC 1996, pp. 1458–1496.

[15] VanBogelen, R. A., Olson, E. R., Wanner, B. L., Neidhardt, F. C., *J. Bacteriol.* 1996, *178*, 4344–4366.

[16] Abshire, K. Z., Neidhardt, F. C., *J. Bacteriol.* 1993, *175*, 3744–3748.

[17] VanBogelen, R. A., Neidhardt, F. C., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 5589–5595.

[18] Cashel, M., Gentry, D. R., Hernandez, V. J., Vinella, D., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC 1996, pp. 1458–1496.

[19] Wanner, B. L., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC 1996, pp. 1357–1381.

[20] VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olson, E. R., Neidhardt, F. C., *Electrophoresis* 1997, *18*, 1243–1251.

[21] Parkinson, J. S., in: Hoch, J. A., Silhavy, T. J. (Eds.), *Two-Component Signal Transduction*, ASM Press, Washington DC 1995, pp. 9–24.

[22] Walker, G. C., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC 1996, pp. 1400–1416.

[23] VanBogelen, R. A., Kelley, P. M., Neidhardt, F. C., *J. Bacteriol.* 1987, *169*, 26–32.

[24] Link, A. J., Robinson, K., Church, G. M., *Electrophoresis* 1997, *18*, 1259–1313.

[25] Link, A. J., Hays, L. G., Carmack, E. B., Yates, III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[26] Wasinger, V. C., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 2299–2308

Review

Masakazu Niimi
Richard D. Cannon
Brian C. Monk

Molecular Microbiology Laboratory, Department of Oral Sciences and Orthodontics, School of Dentistry, University of Otago, Dunedin, New Zealand

Candida albicans pathogenicity: A proteomic perspective

Candida albicans is an opportunistic fungus which causes both superficial infections and life-threatening systemic candidiasis in immunocompromised hosts such as AIDS patients, people with cancer, or other immunosuppressed individuals. Virulence factors for this fungus include the ability to adhere to host tissues, production of tissue damaging secreted enzymes, and changes in morphological form that may enhance tissue penetration and avoidance of immune surveillance. Treatment of candidiasis patients is hampered by a limited choice of antifungal agents and the appearance of clinical isolates resistant to azole drugs. Proteome analysis involves the separation and isolation of proteins by two-dimensional gel electrophoresis and their identification and characterization by mass spectrometry. The systematic application of this methodology to *C. albicans* is in its infancy, but is progressing rapidly. Comparing protein profiles between avirulent and virulent *C. albicans* strains, between drug-sensitive and -resistant strains, or between different morphological forms, could identify key control and effector proteins. There are difficulties, however, associated with the display of low abundance and cell envelope-associated proteins and the choice of conditions for obtaining suitable *C. albicans* cells. This article describes the potential of applying proteome analysis to *C. albicans* in order to better understand pathogenicity and identify new antifungal targets.

Keywords: *Candida albicans* / Pathogenicity / Proteome / Two-dimensional polyacrylamide gel electrophoresis / Cell envelope / Review EL 3536

Contents

1 Introduction

Candida albicans is a commensal fungus that can be a component of the microbial floras of the oral cavity, gastrointestinal tract or vagina. Not all people carry *C. albicans*, and colonization is not always continuous; the carriage rate varies from 1.0–80.6%, depending on the population surveyed [1]. Multiplication of the organism is normally kept in check through physical barriers such as the skin, by competition with the endogenous microflora, and through host defense mechanisms. Under conditions of host immune suppression, however, or when natural barriers to infection are degraded, colonizing *C. albicans* can give rise to opportunistic infections [2, 3]. These infections can be superficial, involving the mucous membranes, or can be hematogenously disseminated, resulting in systemic disease which has a high associated mortality rate [1]. Oropharyngeal and esophageal candidiasis are infections that frequently affect HIV positive individuals and AIDS patients [4]. Other people with severely compromised immune systems, such as cancer patients or organ transplant recipients, represent additional groups at risk of serious *Candida* infections [5]. Debilitated surgery patients are the most likely candidates for disseminated disease, especially when indwelling catheters are used. *C. albicans* still accounts for about 60% of disseminated infections, but there is a rising tide of strictly opportunistic infections caused by other fungi, including non-*albicans Candida* species, *Aspergillus, Cryptococcus* and *Pneumocystis*.

Correspondence: Dr. Masakazu Niimi, Molecular Microbiology Laboratory, Department of Oral Sciences and Orthodontics, School of Dentistry, University of Otago, PO Box 647, Dunedin, New Zealand
E-mail: masa.niimi@stonebow.otago.ac.nz
Fax: +64-3-479-0673

The status of *C. albicans* as a commensal and opportunist, in contrast with other fungi which are strictly opportunistic, suggests that it must be able to express organism-specific determinants that allow it to compete on the mucosal surface. Such factors have only begun to be identified. These are clearly important because most instances of serious disease involve strains carried by the host [6]. It also follows that a critical factor in pathogenesis is a changed interaction of the *C. albicans* with the host defences. It is likely that multiple factors contribute to *C. albicans* virulence, and that different factors are involved at different stages of pathogenesis of each type of infection [7]. Several *C. albicans* features have been suggested to contribute to pathogenicity [8–10]. These include secretion of hydrolytic enzymes (such as aspartic proteinases, phospholipases and *N*-acetylglucosaminidase), the ability to adhere to host tissues and phenotypic switching, a phenomenon that involves changing several growth and morphological characteristics simultaneously. *C. albicans* is a polymorphic fungus which grows predominantly in either a yeast-like or mycelial form. The ability to convert between yeast and mycelial growth is also considered to be a virulence factor. The mycelial form may assist *C. albicans* to invade host tissues [11] and help avoid immune surveillance by macrophages [12]. Other strictly opportunistic fungi, such as *Cryptococcus* and other non-*albicans Candida* species, use alternative routes to colonize selected sites, and without necessarily requiring morphological change. Thus, certain *C. albicans* virulence factors may be species-specific.

Fungal infections of the immunocompromised are often difficult to treat because of adverse drug interactions, side effects, the need for prophylactic treatment over long periods and a limited repertoire of effective fungicidal drugs [13]. The increased incidence of azole drug resistance among clinical *Candida* isolates is causing further problems in the treatment of candidiasis patients, while superinfection by inherently resistant *Candida* species such as *C. glabrata* and *C. krusei* is increasingly common among prophylactically treated patients with chronic vaginitis and AIDS, respectively [14, 15]. A better understanding of the biology, pathogenesis and drug resistance mechanisms of *Candida* species and the identification of new broad spectrum antifungal drug targets are urgently needed to combat opportunistic fungal infections.

Classical genetic studies of *C. albicans* biology have proved difficult because the organism is a diploid eukaryotic microbe that has no known sexual cycle [16]. For some strains certain chromosomes may also be polyploid. A further complication is that *C. albicans* decodes CUG as serine rather than leucine [17]. Despite these limitations, the application of molecular biological methodologies such as gene cloning, genetic transformation with plasmids and gene disruption by homologous recombination, has allowed the genetic description of many interesting biological and pathological properties of *C. albicans*. An internet server on *C. albicans* research, compiled by Dr. Scherer (Acadia Biosciences, Richmond, CA, USA) and Dr. Magee (University of Minnesota, MN, USA), contains a curated *Candida* database that provides regularly updated sequence data and other useful information on this fungus [18]. The *C. albicans* haploid genome size is 16 Mb; by the middle of 1998 more than 2 000 genes had been tentatively identified [19]. The sequencing of the *Saccharomyces cerevisiae* genome (14 Mb, 6340 open readingframes) [20] has greatly assisted the study of *C. albicans* because *Candida* sequences can, in many cases, be compared to *Saccharomyces* counterparts. While *S. cerevisiae* is a relatively avirulent facultative anaerobe and *C. albicans* is a pathogenic obligate aerobe, the comparison is often biologically relevant because many of the genes identified in *C. albicans* have been isolated by functional complementation in *S. cerevisiae* or cloned by screening libraries through hybridization with homologous *S. cerevisiae* DNA probes. It is significant, however, that approximately 5% of *C. albicans* genes analyzed to date have no clear *S. cerevisiae* homologue [19].

High sensitivity proteomic characterization of *C. albicans* is still in its infancy. Although protein sequence information has been obtained in order to characterize the products of several *Candida* genes [21] and many two-dimensional gel electrophoresis (2-DE) separations have been undertaken [22–36] (Fig. 1), the high costs associated with the technology required for systematic detailed proteomic analyses mean that such studies are presently restricted to large research groups and companies. In this paper we discuss the possible utility of proteome analysis of *C. albicans* in relation to this organism's pathogenicity and its virulence factors.

2 Insights and oversights from proteomic analyses

Since proteins carry out or affect most cellular functions in biological systems, measurement of the extent and timing of the expression of individual coding sequences in physiologically relevant contexts can provide a valuable but preliminary insight into biological processes such as the cell cycle, growth and pathogenesis. Pathogenesis can be simplistically defined as a multistep process which requires the controlled expression and regulation of particular proteins, including virulence factors, that are needed for functions such as colonization, cell sustenance and growth in specific host environments, and avoidance of host defense mechanisms. Since protein expression in

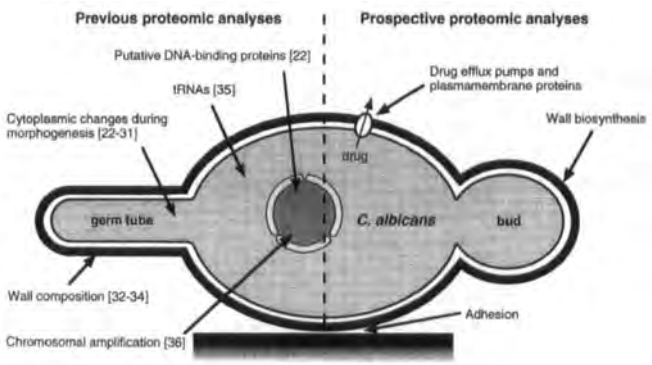


Figure 1. Previous and prospective applications of proteomic analyses to *C. albicans*.

cells cannot necessarily be predicted from mRNA levels, more appropriate methods of measurement are required. In addition to the amounts of proteins expressed, protein modification, processing, turnover and macromolecular associations affect protein function. An understanding of normal and pathogenic states therefore requires a comprehensive analysis of the identity, quantity, state of modification and the association of proteins [37]. Proteomics is beginning to give access to such data. Improved protein separation, sensitive peptide detection and process automation mean that it is now possible to resolve reproducibly thousands of proteins from complex samples, and identify proteins directly from 2-DE gels. Up to 10 000 proteins can be resolved on large-format 2-DE gels and the reproducibility of patterns is enhanced by employing first-dimensional gels with covalently immobilized pH gradients (IPGs) [37]. The use of fluorescent dyes to label proteins allows detection of less than 1 pg protein, which is more sensitive than silver staining [38]. The fluorescent images of labeled gels can be recorded electronically and used to generate libraries of expressed protein profiles. The identity of proteins can be discovered by excising them from 2-DE gels and subjecting them to peptide mapping [37]. Proteins can be cleaved into peptides by enzymatic digestion, chemical cleavage or collision with argon, and the parental protein and its peptide fragments characterized by mass spectrometry after matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI) [37, 39]. The peptide mass fingerprints generated can be compared with databases of peptides predicted from protein sequences to identify the parent protein. Often protein databases are linked to genomic databases allowing immediate access to coding sequences. As these databases grow, it is becoming possible to detect post-translational modifications of proteins, such as acetylation and phosphorylation [38].

Correct identification of proteins from peptide fingerprints requires databases containing information from complete coding sequences. Much of the currently available *C. albicans* genome sequence has been obtained by shotgun sequencing the ends of DNA fragments. This means that the number of complete protein sequences is relatively low, making it difficult to identify unknown proteins. If a protein cannot be identified from existing databases it can often be cloned from genomic DNA by PCR using primers based on the amino acid sequences of peptides derived from the protein. Tandem mass spectrometry (MS/MS) uses parent peptide ions that are identified by MS, isolated, and randomly fragmented by collision with an inert gas such as argon. The resultant fragments are then characterized by MS. This process can be used to obtain complete amino acid sequences for sample proteins [39]. The most sensitive analytical procedure is capillary electrophoresis of peptides followed by MS/MS. This powerful combination can identify femtomole amounts of protein [37].

Despite advances in the sensitivity of 2-DE, suitable model systems are required for *C. albicans* proteome analysis because it is not yet possible to obtain sufficient quantities of useful fungal material from disease states. The use of *in vitro* systems to study virulence factors requires rigorous controls to ensure that detectable protein differences are not artefacts. Based on genome size relative to *S. cerevisiae*, *C. albicans* can be expected to encode about 7000 proteins. Inherent limitations in 2-DE resolution of proteins, including both related and unrelated proteins having overlapping physiochemical properties, protein microheterogeneity and limitations on protein sample loading, as well as widely disparate levels of expression, mean that still only the most abundant fraction of gene products are likely to be successfully analyzed. One approach which circumvents some of these limitations is the use of a very narrow range IPG for the first-dimensional separation of 2-DE. This enhances sample loading within the desired pH range [39] and should lead to the visualization and identification of less abun-

dant proteins. Fractionation of cellular components into soluble and organellar fractions, provided that adequate solubilization of protein components can be achieved, will also enhance the potential for the separation of the protein repertoire. *C. albicans* behaves similarly to *S. cerevisiae* in terms of organellar purification for fractions such as nuclei, mitochondria, plasma membranes and cell walls. Likewise, about 40% of its protein components can be expected to be in, or associated with, membranes while a further significant percentage of proteins is in the less tractable cell wall fraction.

3 Application of proteomics to *C. albicans* pathogenicity

The multifactorial nature of *C. albicans* virulence indicates that a proteomic approach, which monitors the expression of many proteins simultaneously, may be a useful way to compare the commensal status of the organism with disease progression or the consequences of manipulation in model systems. In addition, several putative virulence factors occur in the *C. albicans* genome within gene families. For example, there are at least nine genes encoding secreted aspartic proteinases [40, 41] and multiple genes encode putative drug efflux pumps [15]. Proteomics could be used to examine the expression patterns of members of these gene families, particularly when coupled with sensitive detection methods. An interaction of the fungus with the host is central to pathogenesis. In addition to proteins involved in tissue attachment, signaling of attachment, and tissue penetration, part of the proteome may also be involved in avoidance of peptide-based defense mechanisms and immune surveillance. So far these factors have eluded discovery.

A combination of proteomics with genomics could identify gene products expressed during carriage or pathogenesis that are expressed only in *C. albicans* and not in nonpathogenic fungi such as *S. cerevisiae*. Alternatively, it should be possible to identify proteins which are structurally conserved and that are expressed in a range of fungi. For example, biochemical, immunochemical and molecular genetic studies show that the plasma membrane H^+-ATPase is the dominant protein in the plasma membrane of *Candida* species and *S. cerevisiae* ([21]; Monk, Mason and Fuge, unpublished). The core of this essential P-type ATPase appears to be strongly structurally conserved, although interesting kinetic and regulatory features can be demonstrated that correlate with morphogenetic capability [42]. Similar broad-based protein analytical approaches could identify possible drug targets for either specific or broad-spectrum drug design. The following sections examine in more detail the application of proteomics to the analysis of potential *C. albicans* virulence factors and novel drug targets.

3.1 Morphogenesis and DNA binding proteins

Germ tube formation is an early stage of the yeast-to-mycelial (dimorphic) transition and may enhance *Candida* cell adhesion to host cell surfaces, mediate tissue penetration and help the fungus evade host defences. The dimorphic transition is also induced *in vitro* in response to changes in a wide variety of environmental factors [1] which have been used to model morphogenesis. Factors known to induce the dimorphic transition include carbon source (*N*-acetylglucosamine (GlcNAc), glucose, galactose, proline [22, 43]), incubation temperature [44], pH [45], and zinc [46]. Several genes involved in the signal transduction pathways for the dimorphic transition in *C. albicans* have been tentatively identified by functional complementation in *S. cerevisiae* mutants [9, 12, 47–52] and there appears to be more than one signaling pathway that can induce mycelial growth in *C. albicans*. However, neither the receptors for morphological signals nor the effectors of morphological change have yet been discovered. This may be because they are minor cellular components. There are also distinct differences between the induction and expression of morphological change in *C. albicans* and *S. cerevisiae*. The latter can be induced for pseudomycelial growth but not mycelial growth after nitrogen starvation, while germ tube formation in *C. albicans* is induced after carbon starvation.

We have evaluated the protein profiles of the cytosolic fractions (supernatants obtained after 100 000 × *g* centrifugation) from cells during an early stage of the transition between yeast and mycelia using 2-DE [22]. Carbon-starved *C. albicans* ATCC 10261 cells supplemented with glucose or galactose at 37°C were induced to undergo pH-dependent morphogenesis, with yeast growth occurring at pH 4.5 and efficient germ tube formation at pH 6.7. Morphogenesis of cells induced with GlcNAc under the same conditions was pH-independent. The use of multiple morphogenesis-inducing conditions enabled us to distinguish between specific morphogenetic changes and nonspecific effects due to carbon metabolism or the changed pH.

In an attempt to identify candidate polypeptides which regulate the early events of morphogenesis, we have focused on polypeptides which are differentially synthesized in the period prior to budding and significant germ tube emergence, *i.e.*, within the first hour of glucose-, galactose- or GlcNAc-induced germ tube induction. Final commitment to budding or germ tube formation occurs a significant time after germ tube or bud emergence, as

shown by the ability to reverse either phenotype [22]. Similar to previous reports which examined ^{35}S-labeled extracts from whole cells [23], almost all of the 400 polypeptides visualized in the cytosolic fractions of *C. albicans* changed little in their electrophoretic properties and relative abundance when germ tube formation was compared with yeast growth (Fig. 2). This 2-DE analysis combined with silver staining detected some minor changes in the levels of a small group of polypeptides whose behavior seemed to reflect the resumption of growth and carbon source metabolism. This result emphasizes the utility of even a simple and low technology proteomic approach in detecting the responses of cells to environmental and metabolic change. There were no polypeptides, however, whose properties indicated a specific association with the early phase of morphogenesis. Our results do not exclude the possibility of additional regulatory polypeptides that have remained undetected because of analytical limitations, their low abundance, or their transient appearance in the cells [23]. As proteomic technology develops it should be possible to detect other underlying changes of significance. A limitation of this analysis is the somewhat artificial conditions used to obtain sufficient material for 2-DE. Suitable animal models and much more sensitive detection methods would be required to analyze material from *in vivo*-induced morphogenesis.

DNA-binding proteins are key regulators of cellular response to environmental and metabolic changes [53] and are therefore likely to be involved in some of the early events in morphogenesis. Unfortunately, the lack of information about the activity of specific genes during *C. albicans* morphogenesis means that conventional foot-printing assays cannot be used to detect regulatory proteins.

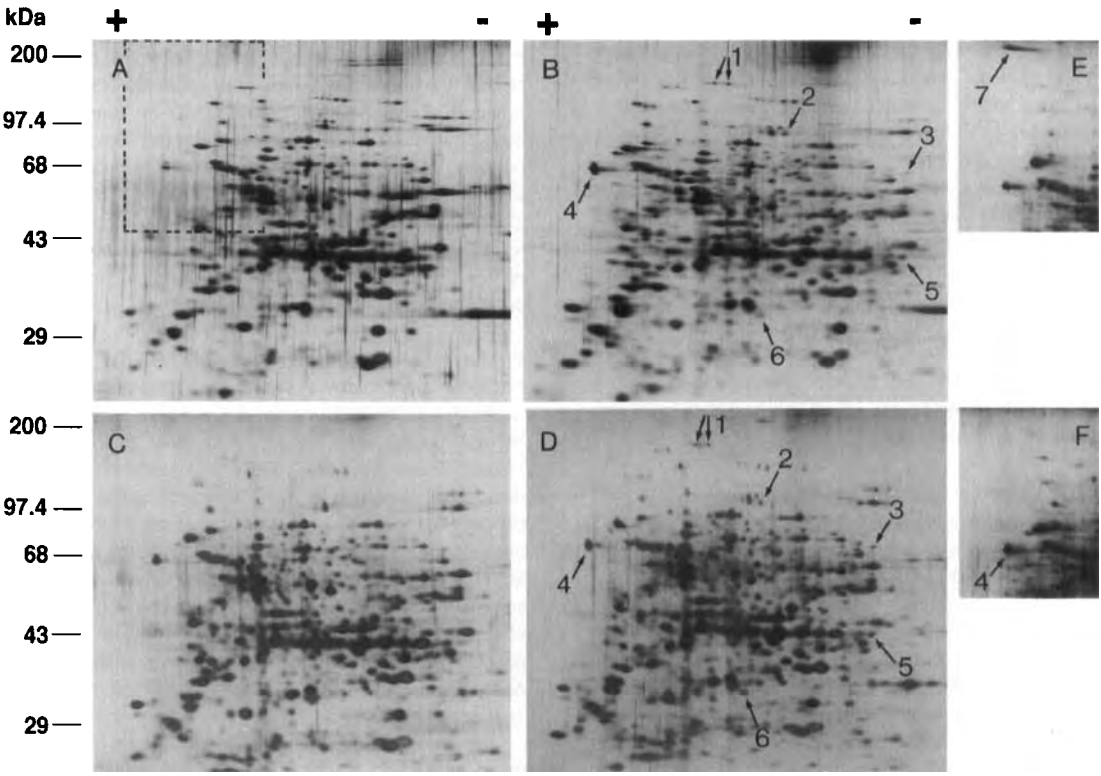


Figure 2. Two-dimensional PAGE separation of cytosolic polypeptides (S 100) extracted from glucose-induced *C. albicans* cells programmed for germ tube formation and yeast growth. Glucose (16 mM) was added to carbon-starved cells to induce germ tubes at pH 6.7 (A, B and E) or to grow as budding yeast at pH 4.5 (C, D and F); (A) and (C) samples prepared immediately after glucose addition; (B) and (D) samples prepared 1 h after glucose induction; (E) and (F) (small panels, with the equivalent portion of the gel indicated in A) samples prepared 2 h after glucose addition. S100 fractions (5 µg) were electrophoresed and gels were stained with silver. Arrows indicate differentially synthesised polypeptides. Reproduced from [22], with permission.

We have therefore adopted the less specific affinity-based polyethylenimine precipitation [54] or heparin-agarose chromatography approaches [55, 56]. These exploit the properties of DNA-binding proteins so that candidate regulatory polypeptides are concentrated from crude extracts of cells which are becoming committed to either yeast growth or germ tube formation. Both polyethylenimine and heparin-agarose chromatography indicated the presence of a small group of polypeptides whose abundance was affected by physiological parameters such as pH difference, response to carbon source or release from starvation. Heparin-agarose chromatography concentrated a further minor group of low abundance polypeptides (molecular mass range of 130–200 kDa) that was differentially synthesized during glucose-induced morphogenesis (Fig. 3). These encouraging results demonstrate how a simple fractionation technique can extend the sensitivity of proteomic analysis so that potential developmental changes can be identified. Further tests are needed to determine whether the polypeptides are specifically induced during germ tube formation or by exposure of cells to metabolisable inducer at high pH; pH-dependent galactose- or pH-independent GlcNAc induction should help to answer this question. Given the broad specificity of heparin-agarose, it will also be necessary to determine if they are indeed DNA-binding proteins or part of a DNA-binding complex. If these polypeptides exhibit appropriate properties, heparin-agarose chromatography may allow the partial purification of sufficient material to facilitate microscale sequence analysis [57], or a sensitive 2-DE analysis as a prelude to gene cloning.

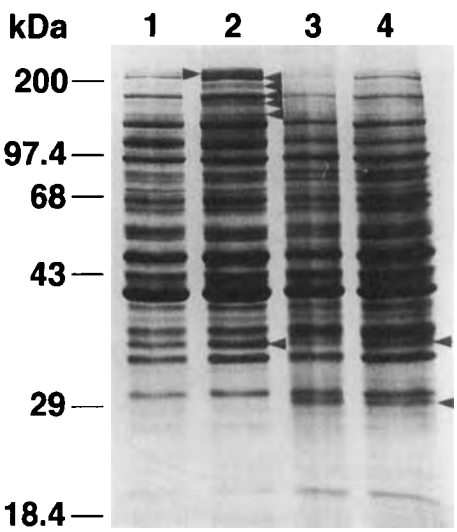


Figure 3. Silver-stained SDS-polyacrylamide gel electrophoretic analysis of fractions of S100 polypeptides from heparin-agarose affinity chromatography. Samples are as follows: glucose-induced germ tube forming cells at pH 6.7 (lanes 1 and 2) and budding yeast cells at pH 4.5 (3 and 4); (1) and (3) S100 fractions prepared immediately after glucose addition; (2) and (4) S100 fractions prepared 1 h after glucose addition; 0.25 mM ammonium sulfate eluates. Arrows indicate differentially synthesised polypeptides.

We have measured *PMA1* (plasma membrane H^+-ATPase) mRNA during morphogenesis relative to expression of the constitutive *ADE2* gene [42]. Although the expression of both messages increased by between 2- and 6-fold on release from starvation, with a 3-fold lower level of expression during germ tube formation compared with yeast growth, a constant ratio of the two messages was maintained. Interestingly, the 97 kDa ATPase band reached and maintained its normal expression levels in the membrane (rising from about 10% to about 20% of plasma membrane protein) after release from starvation and with either the induction of yeast growth or germ tube formation. Kinetic analysis indicated that although the ATPase was rapidly activated after addition of metabolisable sugar, as expected from earlier studies which showed extensive cytoplasmic alkalization during germ tube formation [58], the enzyme remained preferentially activated in the germ-tube-forming cells. These data imply that although many genes and gene products required for essential cellular functions are regulated in complex patterns during germ tube formation and yeast growth, gene expression appears to be similarly coordinated in both contexts, and additional regulatory factors may exert profound influences at the functional level. These results are consistent with the limited changes that were seen in our analysis of protein profiles obtained with the cytosolic fraction and the considerable delay between initiation of morphogenesis and the commitment to morphogenesis [22]. Thus, factors not readily visualized during low sensitivity proteomic analysis may be key regulators of the physiological, cell-cycle-related and morphogenic changes required for germ tube formation. These include far-reaching events such as intracellular pH modification mediated by the activated ATPase, or general transcriptional modification effected by a small fraction of low abundance polypeptides, possibly DNA binding proteins or regulatory complexes. Furthermore, it still cannot be excluded that an appropriate environmental manipulation of the signaling pathways extant at the end of carbon starvation is all that is required to induce a commitment to morphological change.

3.2 Adhesins

Adhesins are the fungal surface moieties that mediate the binding of *C. albicans* to other cells (host or microbial) or

other surfaces [3, 59]. Inhibition of the interaction between adhesin and receptor molecule may prevent colonization of the human host and so preclude candidiasis. Most *C. albicans* adhesins are mannoproteins and both their protein and carbohydrate portions have been implicated in adherence [59]. Whether due to differences in experimental approach, reagents, strains, or analysis protocols, there is disagreement among studies as to the identity and number of candidal receptors for various ligands. Nevertheless, *C. albicans* clearly possesses multiple adhesins [59, 60]. Traditionally, adhesins have been examined individually, but proteomics may allow the simultaneous study of changes in expression of multiple adhesins.

Hydrophobic proteins in the cell wall of *C. albicans* are involved in adhesion of this organism to host tissues and medical prostheses and thus play a role in its pathogenicity. The cell wall is a complex structure primarily consisting of cross-linked β-1,3 and β-1,6 glucan, chitin and mannoproteins [59]. Proteins can be extracted from cell walls with 2-mercaptoethanol or glucanase treatments and some 2-DE analyses have been undertaken [32]. Wall proteins, however, are difficult to purify and analyze for several reasons. They are highly heterogeneous, with variable glycosylation, and they are often covalently cross-linked to other wall components and the plasma membrane. The hydrophobic nature of these cell wall proteins results in their loss during purification due to adsorption to apparatus surfaces. Masuoka *et al.* [34] described a system that combines preparative isoelectric focusing with continuous elution preparative electrophoresis. The system provides a two-dimensional protein separation while maintaining protein solubility and minimizing protein loss due to adsorption.

The biosynthesis of cell wall components and the remodeling enzymes resident in this region of the cell are attractive drug targets, since the cell wall is essential for survival and is unique to fungal cells. The biosynthesis of cell wall glucan polymers provides the target for the echinocandins and pneumocandins that are now approaching clinical trials [61], while the nikkomycins target fungal chitin synthase [62]. In contrast to the intracellular action of these drugs, demethyl-allosidamine targets the chitinases required for cell wall plasticity during growth and proliferation while the pradamycins initially target the phosphomannan of cell walls and then destroy plasma membrane integrity [62]. The relative paucity of drugs targeting fungal as compared to bacterial cell walls is at least partially due to difficulties in analyzing fungal cell wall proteins and identifying essential targets. The development of improved methods to solubilise and display the repertoire of cell wall components would be a major breakthrough.

Drugs which target extracytoplasmic regions of membrane proteins are also desirable since they do not need to enter the cell, thus avoiding intracellular pump-based detoxification resistance mechanisms. They can be fungicidal rather than fungistatic agents, as has been demonstrated by using omeprazole as a surface active inhibitor of the essential plasma membrane H^+-ATPase [63–65]. Successful proteomic analysis of the full complement of plasma membrane proteins would be a valuable resource for antifungal development. This will require the discovery and application of suitable nonionic detergents. In the interim, SDS-PAGE can provide a useful way to display the plasma membrane proteins. Reaction of concanavalin A (ConA) with SDS-PAGE profiles of *C. albicans* plasma membranes showed that a poorly Coomassie-stained glycoconjugate of approximately 80 kDa was found in the membranes of germ tube-forming cells and not budding cells, with slightly increased amounts displayed after germ tube emergence [42]. These results suggest that the biosynthesis, membrane integration, or unmasking of ConA-binding sites in the 80 kDa glycoconjugate may be associated with germ tube emergence and germ tube elongation. The identity and function of the glycoconjugate remains to be established.

Several technical limitations to proteomic analyses remain and these are particularly relevant for the cell envelope of *C. albicans*, a region that dominates interactions with host cells and is the most likely resource of potential fungicidal targets. The low solubility of many hydrophobic membrane and cell wall proteins plus the effects of variable covalent association can prevent clear separation by 2-DE. These features, combined with the low abundance and extensive post-translational modification of many components, has made it difficult to purify such proteins in quantities sufficient for sequencing or biochemical analysis. Despite efforts to improve the solubility of membrane proteins [66] and cell wall components [34], the 2-DE display and identification of these proteins are far from ideal.

3.3 Drug resistance

By 1997, about 90% of AIDS patients had required antifungal therapy for oropharyngeal or esophageal candidiasis at some time during their illness. At present the antifungal agents most frequently used to treat such patients are imidazole (miconazole and ketoconazole) or triazole drugs (fluconazole and itraconazole). The triazole drugs have advantages over the polyenes of oral delivery and lower host toxicity. However, a significant proportion of patients have experienced azole treatment failure due to the development of drug resistance in *C. albicans*. There are numerous ways in which *Candida* cells could become

resistant to the azole drugs but in many clinical strains drug efflux pumps and altered 14 α-sterol demethylase (an enzyme from the ergosterol biosynthetic pathway and a target of azole drugs), are major factors that contribute to *C. albicans* azole antifungal resistance [15].

There are two main families of drug efflux pumps: the ABC (ATP-binding cassette) transporters and the MFS (major facilitator superfamily) proteins [67]. Currently, at least eight members of the *CDR* family have been identified by sequence homolgy [18]. Several of these genes have been cloned by functional complementation of hypersensitive *S. cerevisiae* mutants to give a fluconazole-resistant phenotype [68, 69]. However, many additional *S. cerevisiae* transformants from these experiments have not been fully characterized [68, 69]. Indeed, there are a total of 55 genes encoding ABC and MFS transporters in the *S. cerevisiae* genome [70, 71] and so it is reasonable to predict a similar number of these transporters in *C. albicans*. For some of the genes already identified, transcript levels in fluconazole-sensitive and fluconazole-resistant clinical isolates have been compared. The amounts of *CDR1, CDR2* and/or *CaMDR1* mRNAs have been shown to be significantly higher in azole-resistant strains than in azole-sensitive strains [68, 69, 72, 73]. Important questions remain. Have the key gene products been identified, and do mRNA concentrations reflect plasma membrane efflux pump expression levels? A proteomic examination of sensitive and resistant strains could help answer these questions. One-dimensional electrophoresis is unlikely to be suitable as many of these proteins have similar molecular masses. However, their different calculated isoelectric points (Cdr1p 170 kDa, p*I* 6.93; Cdr2p 169 kDa, p*I* 6.85; Cdr3p 170 kDa, p*I* 8.09) suggest that separation could be achieved by 2-DE. Analysis of drug efflux pumps by 2-DE would also be facilitated by comparison of isogenic mutants which over- or under-express specific efflux pumps and by immunodetection of reference proteins. We have immunodetected overexpression of Cdr1p or CaMdr1p proteins on 1-D Western blots of plasma membrane fractions isolated from azole-resistant strains. We have also constructed a fluconazole-resistant mutant (FL3) from a fluconazole-sensitive and *CDR1*-deleted strain derived from *C. albicans* CAI14. This was achieved by transforming the *CDR1*-deleted strain with a multicopy *CDR1*-containing plasmid. FL3 is 500-fold more resistant to fluconazole than *CDR1*-deleted CAI14 and is 250-fold more resistant than CAI14. The amount of *CDR1* mRNA in FL3 was at least 8 times higher than CAI14. These strains provide the necessary controls that will facilitate the examination of other resistant variants derived from CAI14 and of fluconazole-resistant clinical isolates. Although the separation of plasma membrane proteins by 2-DE may prove problematic, the solubility of the closely related *S. cerevisiae* Pdr5p multidrug efflux pump in the nonionic detergent dodecylmaltoside [74] holds much promise for the 2-DE display of *C. albicans* drug efflux pumps.

Marichal *et al.* [36] reported that the amount of *CYP51* mRNA transcript (encoding 14 α-sterol demethylase) from azole-resistant *C. glabrata* clinical isolates was greater than that in susceptible isolates. The resistance was correlated with an amplified chromosome which was gradually lost in successive subcultures of the resistant isolate in drug-free medium. The amplified chromosome gave major differences in the 2-DE protein profiles of the susceptible and revertant isolates compared to that of the resistant isolate. They reported that the expression of 25 proteins increased and that the expression of 76 proteins decreased in the resistant isolate. The alteration in gene expression by changing chromosome number or by chromosomal recombination could also be a mechanism by which the asexual yeast *C. albicans* adapts to drug challenge. As more genes are linked to the chromosomal physical map, proteomic analyses may be used to more precisely display and analyze the effect of changes in chromosomal ploidy.

4 Conclusion

It is important to define which parts of the genome are being expressed at any one time since, in the case of both procaryotes and eukaryotes, large portions of the genome are only expressed at specific periods, such as during host infection or morphogenesis [39]. Two recently described techniques, the microarray assay for gene expression [75] and the serial analysis of gene expression (SAGE) [76], allow the expression levels of thousands of genes to be monitored simultaneously. These new methods of gene expression analysis are quantitative and provide systematic approaches to the investigation of gene function and regulation. The future availability of complete genome sequences, developments in microchip technology, and new mass spectrometric identification methods will allow comprehensive analysis of gene expression at the genomic level as well as the mapping of a significant proportion of the approximately 7 000 components of the *C. albicans* proteome. It is certain that investigation of *Candida* will continue to be expedited by insights gained from the sequence and functional analysis of the *Saccharomyces* genome and proteome. However, a full exposition of the biology and pathogenicity of *Candida* cannot rely solely on information borrowed from this and other sources. In this paper we have described some preliminary proteomic studies and have indicated aspects of *C. albicans* biology of relevance to pathogenesis which will benefit from proteomic analysis. The mapping and

sequencing of the *C. albicans* genome will help to focus attention onto those genes whose importance as virulence factors will need to be confirmed by proteomics-based analysis and functional studies. Similarly, differentially expressed proteins detected by proteomic analysis will help target studies towards complimentary genetic analysis of specific genes and gene families. While it could be suggested that the most interesting properties of *C. albicans* and other pathogenic *Candida* species reside in the differences between these yeast species and *Saccharomyces*, this is a simplistic view that is inconsistent with the pragmatic need of the pharmaceutical industry to obtain broad spectrum fungicides. Essential gene products shared among fungi, and preferably those which offer targets specific to fungi, provide the best candidates for antifungal intervention. However, experience with existing antifungal agents, plus the problems that they have generated, suggest that the plasma membrane and cell wall are likely to offer preferred antifungal targets. While these regions present major technical challenges, the successful application of proteomic analysis to this area should yield important practical dividends.

We wish to acknowledge research funding from the Lottery Board of New Zealand, the Health Research Council of New Zealand, and a University of Otago research grant.

Received February 3, 1999

5 References

[1] Odds, F. C., *Candida and Candidosis*, 2nd Ed., Bailliere Tindall, London 1988.

[2] Cannon, R. D., Holmes, A. R., Mason, A. B., Monk, B. C., *J. Dent. Res.* 1995, *74*, 1152–1161.

[3] Cannon, R. D., Chaffin, W. L., *Crit. Rev. Oral Biol. Med.* 1999, *9*, in press.

[4] Alexander, B. D., Perfect, J. R., *Drugs* 1997, *54*, 657–678.

[5] Darouiche, R. O., *Clin. Infect. Dis.* 1998, *26*, 259–274.

[6] Voss, A., Hollis, R. J., Pfaller, M. A., Wenzel, R. P., Doebbeling, B. N., *J. Clin. Microbiol.* 1994, *320*, 975–980.

[7] Perfect, J. R., *Antimicrob. Agents Chemother.* 1996, *40*, 1577–1583.

[8] Cutler, J. E., *Annu. Rev. Microbiol.* 1991, *45*, 187–218.

[9] Kobayashi, S. C., Cutler, J. E., *Trends Microbiol.* 1998, *6*, 92–94.

[10] Odds, F. C., *ASM News* 1994, *60*, 313–318.

[11] Ryley, J. F., Ryley, N. G., *J. Med. Vet. Mycol.* 1990, *28*, 225–239.

[12] Lo, H. J., Kohler, J. R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A., Fink, G. R., *Cell* 1997, *90*, 939–949.

[13] Kauffman, C. A., Carver, P. L., *Drugs* 1997, *53*, 539–549.

[14] Klepser, M. E., Ernst, E. J., Pfaller, M. A., *Trends Microbiol.* 1997, *5*, 372–375.

[15] White, T. C., Marr, K. A., Bowden, R. A., *Clin. Microbiol. Rev.* 1998, *11*, 382–402.

[16] Pla, J., Gil, C., Monteoliva, L., Navarro-Garcia, F., Sanchez, M., Nombela, C., *Yeast* 1996, *12*, 1677–1702.

[17] Santos, M. A., Tuite, M. F., *Nucleic Acids Res.* 1995, *23*, 1481–1486.

[18] Scherer, S., http://alces.med.umn.edu/Candida.html

[19] Magee, P. T., Scherer, S., *ASM News* 1998, *64*, 505–511.

[20] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, *274*, 546–567.

[21] Monk, B. C., Kurtz, M. B., Marrinan, J. A., Perlin, D. S., *J. Bacteriol.* 1991, *173*, 6826–6836.

[22] Niimi, M., Shepherd, M. G., Monk, B. C., *Arch. Microbiol.* 1996, *166*, 260–268.

[23] Finney, R., Langtimm, C. J., Soll, D. R., *Mycopathologia* 1985, *91*, 3–15.

[24] Sevilla, M.-J., Odds, F. C., *J. Med. Vet. Mycol.* 1986, *24*, 419–422.

[25] Manning, M., Mitchell, T. G., *J. Bacteriol.* 1980, *144*, 258–273.

[26] Manning, M., Mitchell, T. G., *Infect. Immun.* 1980, *30*, 484–495.

[27] Brown, L. A., Chaffin, W. L., *Can. J. Microbiol.* 1981, *27*, 580–585.

[28] Brummel, M., Soll, D. R., *Devel. Biol.* 1982, *89*, 211–224.

[29] Ahrens, J. C., Daneo-Moore, L., Buckley, H. R., *J. Gen. Microbiol.* 1983, *129*, 1133–1139.

[30] Paranjape, V., Gupta Roy, B., Datta, A., *J. Gen. Microbiol.* 1990, *136*, 2149–2154.

[31] Vespa, M. N., Lebecq, J. C., Simonetti, N., Bastide, J. M., *Comp. Immunol. Microbiol. Infect. Dis.* 1993, *16*, 163–172.

[32] Sepulveda, P., Lopez-Ribot, J. L, Gozalbo, D., Cervera, A., Martinez, J. P., Chaffin, W. L., *Infect. Immun.* 1996, *64*, 4406–4408.

[33] Calderone, R. A., Wadsworth, E., *Rev. Infect. Dis.* 1988, *10*, Suppl. 2, 423–427.

[34] Masuoka, J., Glee, P. M., Hazen, K. C., *Electrophoresis* 1998, *19*, 675–678.

[35] Tuite, M. F., Bower, P. A., McLaughlin, C. S., *Biochim. Biophys. Acta* 1986, *866*, 26–31.

[36] Marichal, P., Vanden Bossche, H., Odds, F. C., Nobels, G., Warnock, D. W., Timmerman, V., Van Broeckhoven, C., Fay, S., Mose-Larsen, P., *Antimicrob. Agents Chemother.* 1997, *41*, 2229–2237.

[37] Haynes, P. A., Gygi, S. P., Figeys, D., Aebersold, R., *Electrophoresis* 1998, *19*, 1862–1871.

[38] Fey, S. J., Nawrocki, A., Larsen, M. R., Görg, A., Roepstorff, P., Skews, G. N., Williams, R., Larse, P. M., *Electrophoresis* 1997, *18*, 1361–1372.

[39] James, P., *Biochem. Biophys. Res. Commun.* 1997, *231*, 1–6.

[40] Hube, B., *Curr. Top. Med. Mycol.* 1996, *7*, 55–69.

[41] Monod, M., Hube, B., Hess, D., Sanglard, D., *Microbiology* 1998, *144*, 2731–2737.

[42] Monk, B. C., Niimi, M., Shepherd, M. G., *J. Bacteriol.* 1993, *175*, 5566–5574.

[43] Holmes, A. R., Shepherd, M. G., *J. Gen. Microbiol.* 1987, *133*, 3219–3228.

[44] Evans, E. G. V., Odds, F. C., Richardson, M. D., Holland, K. T., *Can. J. Microbiol.* 1975, *21*, 338–342.

[45] Buffo, J., Herman, M. A., Soll, D. R., *Mycopathologia* 1984, *85*, 21–30.

[46] Soll, D. R., *Curr. Topics Med. Mycol.* 1985, *1*, 258–285.

[47] Liu, H., Kohler, J., Fink, G. R., *Science* 1994, *266*, 1723–1726.

[48] Leberer, E., Harcus, D., Broadbent, I. D., Clark, K. L., Dignard, D., Ziegelbauer, K., Schmidt, A., Gow, N. A. R., Brown, A. J. P., Thomas, D. Y., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 13217–13222.

[49] Kohler, J. R., Fink, G. R., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 13223–13228.

[50] Braun, B. R., Johnson, A. D., *Science* 1997, *277*, 105–109.

[51] Gadd, G. M., Foster, S. A., *Microbiology* 1997, *143*, 437–448.

[52] Csank, C., Schroppel, K., Leberer, E., Harcus, D., Mohamed, O., Meloche, S., Thomas, D. Y., Whiteway, M., *Infect. Immun.* 1998, *66*, 2713–2721.

[53] Verdier, J.-M., *Yeast* 1990, *6*, 271–297.

[54] Burgess, R. R., *Methods Enzymol.* 1991, *208*, 3–10.

[55] Farooqui, A. A., *J. Chromatogr.* 1980, *184*, 335–345.

[56] Sorger, P. K., Ammerer, G., Shore, D., *Protein Function: A Practical Approach*, IRL Press, Oxford 1989, pp. 199–223.

[57] LeGendra, N., Matsudaira, P., *A Practical Guide to Protein and Peptide Purification for Microsequencing*, Academic Press, New York 1989, pp. 49–69.

[58] Stewart, E., Gow, N. A. R., Bowen, D. V., *J. Gen. Microbiol.* 1988, *134*, 1079–1087.

[59] Chaffin, W. L., López-Ribot, J. L., Casanova, M., Gozalbo, D., Martínez, J. P., *Microbiol. Mol. Biol. Rev.* 1998, *62*, 130–180.

[60] Holmes, A. R., McNab, R., Jenkinson, H. F., *Infect. Immun.* 1996, *64*, 4680–4685.

[61] Kurtz, M. B., *ASM News* 1998, *64*, 31–39.

[62] Groll, A. H., De Lucca, A. J., Walsh, T. J., *Trends Microbiol.* 1998, *6*, 117–124.

[63] Monk, B. C., Perlin, D. S., *Crit. Rev. Microbiol.* 1994, *20*, 209–223.

[64] Monk, B. C., Mason, A. B., Abramochkin, G., Haber, J. E., Seto-Young, D., Perlin, D. S., *Biochim. Biophys. Acta* 1995, *1239*, 81–90.

[65] Seto-Young, D., Monk, B. C., Mason, A. B., Perlin, D. S., *Biochim. Biophys. Acta* 1997, *1326*, 249–256.

[66] Pasquali, C., Fialka, I., Huber, L. A., *Electrophoresis* 1997, *18*, 2573–2581.

[67] Balzi, E., Goffeau, A., *Biochim. Biophys. Acta* 1994, *1187*, 152–162.

[68] Prasad, R., De Wergifosse, P., Goffeau, A., Balzi, E., *Curr. Genet.* 1995, *27*, 320–329.

[69] Sanglard, D., Ischer, F., Monod, M., Bille, J., *Microbiology* 1997, *143*, 405–416.

[70] Decottignies, A., Goffeau, A., *Nature Genet.* 1997, *15*, 137–145.

[71] Goffeau, A., Park, J., Paulsen, I. T., Jonniaux, J.-L., Dinh, T., Mordant, P., Saier Jr., M. H., *Yeast* 1997, *13*, 43–54.

[72] Albertson, G. D., Niimi, M., Cannon, R. D., Jenkinson, H. F., *Antimicrob. Agents Chemother.* 1996, *40*, 2835–2841.

[73] Sanglard, D., Kuchler, K., Ischer, F., Pagani, J.-L., Monod, M., Bille, J., *Antimicrob. Agents Chemother.* 1995, *39*, 2378–2386.

[74] Decottignies, A., Kolaczkowski, M., Balzi, E., Goffeau, A., *J. Biol. Chem.* 1994, *269*, 12797–12803.

[75] Schena, M., Shalon, D., Davis, R. W., Brown, P. O., *Science* 1995, *270*, 467–470.

[76] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, *270*, 484–487.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

 When citing this article, please refer to: *Electrophoresis* 1999, *20*, 2013–2026

Review

Proteomics for genetic and physiological studies in plants

Hervé Thiellement[1,2]
Nasser Bahrman[2]
Catherine Damerval[3]
Christophe Plomion[4]
Michel Rossignol[5]
Véronique Santoni[5]
Dominique de Vienne[3]
Michel Zivy[3]

[1]Laboratoire de Biochimie et Physiologie Végétales, Département de Botanique et Biologie Végétale, Université de Genève, Switzerland
[2]Laboratoire de Génétique et Amélioration des Plantes, INRA, Estrées-Mons, France
[3]Laboratoire de Génétique des Systèmes Végétaux, Station de Génétique Végétale, INRA/CNRS/INA-PG/U Paris-Sud, Gif-sur-Yvette, France
[4]Laboratoire de Génétique et Amélioration des Arbres Forestiers, INRA, Pierroton, France
[5]Biochimie et Physiologie Moléculaire des Plantes, INRA/ENSA-M/CNRS URA 2133, Montpellier, France

Proteomics is becoming a necessity in plant biology, as it is in medicine, zoology and microbiology, for deciphering the function and role of the genes that are or will be sequenced. In this review we focus on the various, mainly genetic, applications of the proteomic tools that have been developed in recent years: characterization of individuals or lines, estimation of genetic variability within and between populations, establishment of genetic distances that can be used in phylogenetic studies, characterization of mutants and localization of the genes encoding the revealed proteins. Improvements in specifically devoted software have permitted precise quantification of the variation in amounts of proteins, leading to the concept of "protein quantity loci" which, combined with the "quantitative trait loci" approach, results in testable hypotheses regarding the role of "candidate proteins" in the metabolism or phenotype under study. This new development is exemplified by the reaction of plants to drought, a trait of major agronomic interest. The accumulation of data regarding genomic and cDNA sequencing will be connected to the protein databases currently developed in plants.

Keywords: Proteomics / Plant genetics / Plant physiology / Genetic localization / Protein quantity loci / Candidate genes / Plant phylogeny / Review EL 3539

Contents

Correspondence: Dr. Hervé Thiellement, Départment de Biologie Végétale, Université de Genève, 3 place de l'Université, CH-1211 Genève 4, Switzerland
E-mail: herve.thiellement@bota.unige.ch
Fax: +41-22-329-77-95

Abbreviations: P/A, presence/absence; **PQL**, protein quantity loci; **QTL**, quantitative trait loci

1 Introduction

By 2000, the genome of *Arabidopsis thaliana* will be the first plant genome to be entirely sequenced. To date, functions can be hypothesized for only one third of the genes already sequenced in this species [1]. Of course new technologies are emerging, such as high-density filters, DNA chips and other "transcriptome" tools [2, 3], that permit simultaneous examination of thousands of transcripts. However, the complex regulatory routes, from post-translational modifications to protein turnover cannot be studied at the cDNA level. Thus the proteome approach is necessary to help answer questions of functional analysis. The proteome approach (the protein complement expressed by a genome [4]) still relies on one technique: two-dimensional gel electrophoresis (2-DE) of

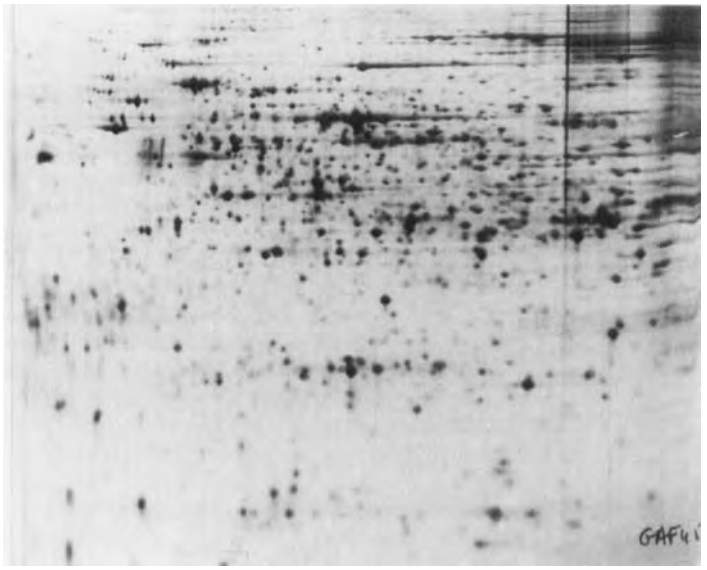


Figure 1. 2-D gel of proteins extracted from 8-day-old etiolated maize coleoptile.

proteins. Since its description, more than twenty years ago, by O'Farrell [5] and Klose [6], improvements have been added constantly (Fig. 1). It is easier today to perform 2-D gels, to examine the expression of several hundreds of genes and to compare the patterns obtained from different genotypes, conditions or developmental stages using specifically dedicated software. These several hundreds of protein spots are *de facto* genetic and physiological markers (see [7, 8] and below for reviews). Protein markers can be useful for assessing genetic variability and for establishing genetic distances and phylogenetic relationships between lines, species and genus. In addition, their encoding genes may be mapped whenever they exhibit positional genetic variability (charge and/or molecular mass; Fig. 2). Thus the genetic maps, usually constructed with nucleic acid markers such as RFLPs, RAPDs, microsatellites or other single tag sequences (STS), can be usefully completed with expressed genes.

Innumerable proteins have been described, whose relative abundance depends on the conditions (light, heat, cold, hormones, pathogens, *etc.*), and on the development stages or organs. As far as these proteins can be sequenced, the corresponding genes that may exist in the databases as "orphan" genes or ESTs may thus be further characterized and tentative functions may be proposed. Also in progress is a new approach that considers a protein amount as a quantitative trait to which the quantitative trait loci (QTL) mapping methodology can be applied, leading to localization of the loci controlling the abundance of the protein, and to a strategy to pinpoint "candidate proteins" involved in the realization of phenotypic traits under study. In this review we will focus on results obtained using proteomics in plant genetics and physiology, and on current developments which combine these approaches to progress in the cataloguing of the plant genes and their functions.

2 Genetic diversity

The 2-DE technique was used, from the beginning, to detect genetic diversity. Zivy *et al.* [9], in 1983, investigated the genetic variability of 18 wheat (*Triticum aesti-*

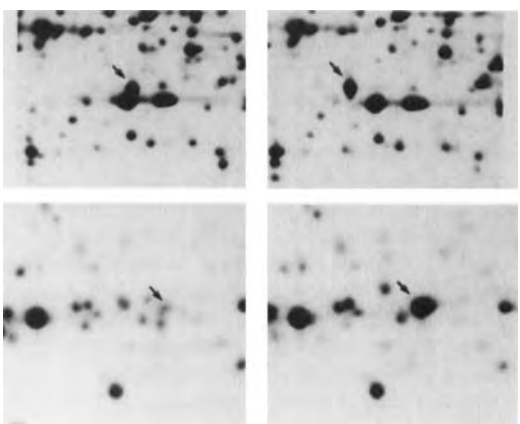


Figure 2. Examples of genetic differences between two maize lines. Top: the arrow indicates a position shift variation (two allelic spots differeing by their isoelectric point); bottom: the arrow indicates a variation in amount.

vum) alloplasmic lines. Genetic differences between two wheat lines and two groups of cytoplasms were observed on proteins extracted from five developmental stages or organs. Each stage or organ was unambiguously characterized. Twenty allelic variations and 20 quantitatively variable proteins [10] characterized nuclear variations between Chinese Spring and Selkirk varieties. In some cases, the post-translational modifications were found to be stage-specific. Technical improvements allowed an increase in the percentage of revealed polymorphisms [11]. Anderson *et al.* [12] analyzed storage proteins of wheat seeds including glutenins and gliadins from 14 varieties. From the analysis of protein patterns, it was deduced that charge modifications were more frequent than mass modifications. A comparison of two maize lines showed that the range of variability revealed (Fig. 2) depends upon the organ analyzed (from 7.5% for the blade to 12.6% for the mesocotyl and 13.2% for the sheath [13, 14]). An experiment involving five lines, seven of their hybrids and six organs or physiological stages of maize, allowed examination of the inheritance of about 2000 situations of between-line differences in protein amount [15]. Additive inheritance was the most frequent (96% of the situations), but nonadditivity concerned different proteins in the different stages/organs. On the other hand, positive nonadditive inheritance (hybrid similar to parental line displaying high amount) is more frequent than the reverse [13–15]. Zivy [17] found higher polymorphism for heat shock proteins than for other wheat seedling proteins. Branlard [18] used 2-DE to improve the study of the determination of wheat gliadins routinely separated by 1-D electrophoresis.

Several successful attempts were made, using 2-DE to characterize closely related lines of durum wheat [19], flour allergens in bread wheat cultivars [20], different lines of barley [21] differing in malting quality [22], cultivars of cultivated and wild rice [23, 24], genotypes of maize [25, 26], of sugar cane [27], of *Vigna angularis* [28] or of carrots [29]. Van Telgen and van Loon [30] examined the chromatin-associated proteins from leaves of different species and varieties of Nicotiana differing in their sensitivity towards tobacco mosaic virus (TMV). Over 90% of the polypeptides were similar in all Nicotiana while each of the species or cultivars showed different patterns. Unfortunately, no specific polypeptide spot could be associated with the presence of one hypersensitivity gene towards TMV. Bahrman *et al.* [31] tested different methods of protein extraction from needles of Douglas fir (*Pseudostsuga menziesii*) and compared three different genotypes. Among 225 expressed loci, 22 displayed variations in amount and 7 of them showed presence/absence (P/A) variations. Sieffert [32] analyzed three provenances of Norway spruce and found significant differences.

In several studies, 2-DE was used as a means for describing the structure of genetic diversity. Two studies on limited samples of maize lines showed that the quantitative variability of proteins gives a pattern of relationships between genotypes that is different from the qualitative and protein structure variability. Moreover, this kind of variability was found to be correlated with some agronomic performances [16, 33]. A larger study on 21 inbred lines used isozymes, RFLPs, and 2-DE. Quantitative protein data still showed a picture of relationships between lines clearly different from the other markers, but was no longer clearly associated with the phenotypic variation [34]. Posch *et al.* [35] studied ten pepper (*Capsicum annuum* L.) inbred lines. The polymorphism of water-soluble and urea/detergent-soluble seed proteins was investigated. In both cases, dendrograms, plotted from maximum-parsimony analysis on the basis of qualitative variations, clearly separated the lines. In a second study [36], seed protein polymorphism was investigated in ten inbred lines of the same species. Computer analysis was carried out on the protein patterns and 68 out of 184 water-soluble as well as 34 out of 419 urea/detergent-soluble proteins were found to be variable. Both fractions contained common proteins, while the variable proteins found in both fractions were nonidentical. Genetic distances between all pairs of inbred pepper lines were calculated from each protein fraction, allowing two distinct genetic groups of five inbred lines to be distinguished. Hirano [37] characterized 17 varieties of mulberries (*Morus* sp.) which were classified on the basis fo 23 variable leaf proteins using principal component analysis.

Bon [38] applied the 2-DE technique to separate the proteins of single meristems from *Sequoiadendron giganteum, Sequoia sempervirens, Pseudostsuga menziesii, Picea abies, Pinus pinaster, Eucalyptus gunii* and *Populus nigra*. The tool was successfully used for the study of inter- and intraclonal differences in meristem cultures of forest tree species. In maritime pine (*Pinus pinaster*), the protein patterns of 42 megagametophytes from seven genetic origins were first compared. Only qualitative variations (presence or absence of the spots) were taken into account and considered as spot characters. A total of 968 polypeptide spots were scored and showed that more than 84% of the polypeptides were variable. The intra- and interorigin variability levels were of similar magnitude. Correspondence analysis and dendrograms plotted on the distance index between individuals permitted classification of the seven origins of maritime pine into three main groups, namely Atlantic, Mediterranean and North African [39]. In another study [40], six populations of mari-

time pines were screened. The 2-DE was performed with proteins extracted from 32 haploid megagametophytes of each population. The allelic frequencies of 16 protein loci for which the segregation was known were scored. The same populations were also analyzed for eight isozymes and six terpene loci. Mean diversity and differentiation were computed and the data were compared with the isozymes and terpenic loci data. Each class of loci displayed a different level of average diversity. A high level of differentiation (G_{ST} = 0.17) was found, which is typical of a species having a highly fragmented range.

Barreneche *et al.* [41] compared 23 oaks from six European countries, covering partly the natural geographic range of white oaks in Europe. Total proteins from seedlings were analyzed and 530 polypeptide spots scored, among which 101 were polymorphic. Interspecific and intraspecific distances were found to be very close. The results confirmed the low level of genetic differentiation between *Quercus petraea* and *Quercus robur*, as was also observed with other markers such as isozymes, RAPDs or chloroplastic DNA. David *et al.* [42] used 2-DE for the comparison of wheat populations, derived from a single original population, that were cultivated under different environmental conditions. They showed that all populations differed from the initial one, and that most differences between them were not due to random drift, but to adaptation to different climatic conditions. All these studies demonstrate the power and the usefulness of the proteome approach for distinguishing lines, populations, varieties and even species, and for addressing important questions in population genetics. The interest of the proteome tool in phylogenetic studies will now be illustrated.

3 Studies of phylogenetic relationships

Several series of experiments were performed in the Triticeae tribe in order to understand the relationships between homologous genomes and to decipher their controversial phylogeny. Comparison of alloplasmic lines (the same nuclear genome introduced by backcrosses in different cytoplasms) led to the detection of three cytoplasmically encoded proteins [43]: the large subunit of Rubisco and two forms of the beta subunit of chloroplast ATPsynthase, *i.e.* three products of two chloroplast genes. Looking at the variability of these three proteins in more than fifty different diploid, tetraploid and hexaploid Triticum (including Aegilops) genotypes has permitted definition of only two cytoplasmic types. The representatives of the A, M and D genomes and several species of the Sitopsis section share one type. The other cytoplasmic type is found in *T. aestivum* (AABBDD), *T. turgidum* (AABB), *T. timopheevi* (AAGG) and only two species of the Sitopsis section, *T. speltoïdes* and *T. aucheri*. This finding is consistent with the hypothesis on the origin of the cultivated wheats where the cytoplasmic and B genome donor of AB and ABD wheats is closer to *T. speltoïdes* and *T. aucheri* than to any other Sitopsis species [43].

The whole 2-D patterns from representatives of the Sitopsis section were then compared and compared to the Chinese Spring (CS) standard wheat line. For each of the 1231 spots found in one or the other genotype, the presence or absence was recorded in each of the genotypes examined. Similarity indices were computed and phenograms constructed from the similarity matrix. The wheat line was found to be more similar to *T. speltoïdes* and *T. aucheri* (these two being so close that they must be classified as one species) than to any other Sitopsis. Moreover, when the spots present only in CS and in one other Sitopsis are considered, it is also with *T. speltoïdes* and *T. aucheri* that CS shared the maximum number of spots. Thus, taking into account the cytoplasmic type, the phenogram computed from all the spots, and the rare species-specific spots, *T. speltoïdes* (and its variety *T. aucheri*) appears as actual species the closest of the B genome donor of the cultivated wheats [44].

Three representatives of the A genome, three representatives of the D genome, and ten accessions from the Sitopsis section (S genome) were then compared. Since different series of 2-D gels were run, 457 spots of CS, run as a standard in every batch, were used as references. As previously, but on this restricted set of spots, similarity indices were computed. As expected, the dendrogram constructed from the similarity matrix was less precise for the Sitopsis species but the main features were still obvious, in particular the proximity between *T. speltoïdes* and *T. aucheri* and the fact that they both appear far from the other Sitopsis. The A representatives are clustered together, as are the D ones. The D cluster merged with the S cluster before the two merged with the A cluster, as expected according to classical systematics [45]. To test further the proteomic approach as a phylogenetic tool, more distantly related species were studied, belonging to different genera: diploids from Triticum (A and D genome representatives), Secale (R genome) and Hordeum (H genome). It was found that barleys are relatively far from wheats (30% of common spots) whereas ryes were closer to wheats, with *circa* 45% of spots found in common [46], which is in accordance with classical taxonomy (Fig. 3).

Compared to sequence analysis, which usually provides neutral markers and describes relationships between species in terms of time divergence (the "molecular clock"), the proteome approach, which is restricted to closely related species, probably gives another kind of informa-

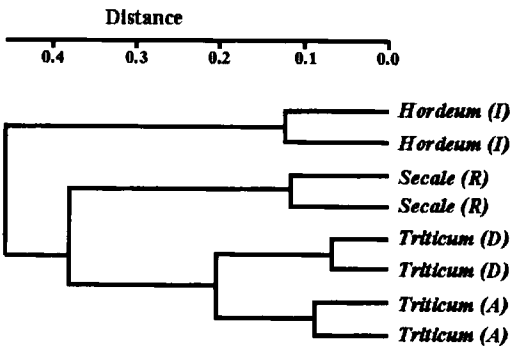


Figure 3. Dendrogram computed from the matrix of genetic distances between species of the Triticeae tribe, estimated from the numbers of common and noncommon spots.

tion: several hundreds of expressed genes are taken into account, among which nonneutral polymorphisms are expected to be more frequent, because genetic variability of the amounts of proteins may have physiological consequences. In addition, the definition of identical products in different species may be of help in sorting out genes in agronomic plants, *e.g.* rape, from those described in a model plant from the same family, *e.g.* Arabidopsis, also a Crucifereae.

4 Expression of homologous genomes in polyploids

The expression of a genome at different ploidy levels and in combination with homologous genomes was also examined in the Triticeae. Using the Chinese Spring (CS) ditelosomic (DT) lines developed by Sears [47], the structural genes encoding the proteins revealed on 2-D gels of the CS hexaploid wheat line can be located on the different chromosome arms. A spot present in the euploid and absent in a given DT line should have its structural gene located on the missing chromosome arm. But only 8% of the CS spots were located this way, although 30 of the 42 possible DT lines were examined [48]. The hexaploid nature of bread wheat with its three homologous genomes led to the hypothesis that most gene products are synthesized by 2–3 indistinguishable homoalleles. They are thus two-dose or three-dose spots when only the one-dose spots can have their strucutral genes located, *i.e.*, disappearing in one or the other DT line [48].

The experiments reported above, where diploids with A, S and D genomes were compared on the basis of 457 CS spots, permitted testing of this hypothesis. More than 82% of the CS spots can be found among the representatives of each of the A, S and D genomes, the B genome of the hexaploid wheat being represented by its closest relative the S genome of the Sitopsis section. Because the genes encoding electrophoretically identical products are very likely homologous structural genes located on the three homologous genomes, it can be deduced that the same allelic forms exist in the three genomes. The hexaploid pattern can be considered as the sum of one A, one B (or S) and one D pattern. Ninety ($3 \times 10 \times 3$) combinations can theoretically be constructed from 3 A, 10 S and 3 D patterns. The number of multiple-dose (two- or three-dose) spots in such "theoretical" hexaploids varied between 72% and 86% [45], explaining why so few genes were located in the DT analysis.

The relationships between homologous genes can raise another question about the extent of a phenomenon described as "intergenomic suppression" [49], *i.e.*, the nonexpression of one or two of the three genomes for a given set of homologous genes. To study the expression of a genome at different ploidy level, the 2-D patterns of a D diploid, an AB tetraploid and the synthetic amphiploid hexaploid wheat, resulting from a colchicin-treated hybrid between the two, were compared. With only few exceptions (among them are the cytoplasmically encoded proteins) all the proteins synthesized by D or AB alone were also found in ABD. Moreover, the 2-D pattern of the amphiploid was indistinguishable from the one obtained by running a 2-D gel with a 1/3 D extract and a 2/3 AB extract. Thus it has be concluded that for most gene products the three genomes are expressed independently, intergenomic suppression being a rare phenomenon [50].

5 Mutant characterization

The 2-DE technique is also a valuable approach for evaluating modifications of protein expression caused by a mutation [7]. The strategy is to compare the 2-D patterns of the mutant and of the wild-type plant cultivated under same conditions. The identification of the affected proteins, by chemical sequencing, mass spectrometry or Western blot, can then provide valuable information to understand the biochemical processes that underlie the mutant phenotype, whether the function of the mutated gene is known or not.

Total proteins of developmental mutants of the model plant *Arabidopsis thaliana* were analyzed [52]. It was shown that these mutants had specific 2-D patterns that could be distinguished from the pattern of the wild-type plant [52]. This study allowed the identification of an isoform of actin whose amount was correlated with the length of hypocotyl. The comparative 2-D protein pattern analysis of a tomato wild-type plant sufficient in Fe as well

as the Fe-deficient mutant allowed identification of several enzymes (glyceraldehyde-3-phosphate dehydrogenase, formate dehydrogenase, ascorbate peroxidase, superoxide dismutase, plastocyanin) that are involved in anaerobic metabolism and in stress defense [53]. The function of these enzymes was discussed in terms of their involvement in Fe acquistion [53]. In another study, four membrane polypeptides involved in Fe transport across the root plasma membrane of maize were characterized by comparison of the 2-D patterns of an Fe-efficient maize cultivar and a mutant defective in Fe uptake [54]. In experiments performed with moss, the 2-D patterns of the wild-type and of a chloroplast mutant were compared [55]. Selection of proteins to be sequenced was based on their altered abundances upon cytokinin treatment in the mutant and between the two genotypes. It was shown that cytokinin affects nuclear- as well as plastome-encoded plastid proteins, which suggested that both compartments are involved in cytokinin action [55].

Another strategy to gain insight into particular physiological processes is the simultaneous comparison of 2-D patterns of mutants differentially affected in the process under study. As such, a set of mutants of *Arabidopsis thaliana*, affected in early development, was analyzed by 2-DE in order to estimate biochemical relationships between them [56]. A previously characterized hormonal mutant and hormone-treated wild-type plants were included as controls for biochemical comparisons. The differences in the 2-D patterns were exploited in order to establish a phenogram of biochemical distances between mutants, using statistical procedures based on cluster analysis. A positive correlation was obtained between this classification and the genetic classification of these mutants. Also a good correlation was shown between physiological traits and biochemical classification since a mutant that overproduced auxin was closely related to the wild-type cultivated in the presence of auxin. In addition, this biochemical classification led to the hypothesis that a mutant, closely related to the wild-type plant cultivated in the presence of cytokinin, was a cytokinin over-producer. Cytokinin dosage confirmed this hypothesis [56]. The same kind of analysis contributed to show cytokinins to be involved in the phenotype of the Pasticcino mutant [57].

When looking at leaf proteins in a series of late flowering Arabidopsis mutants and comparing their patterns to the wild-type, several differences were detected ([59]; Fink, unpublished). But, when examining the cotransmission of protein variation and the flowering phenotype in the selfed progeny of the hybrid between mutant and wild-type, not one cosegregated. These experiments reported negative results but emphasized (i) the analytical power of the method, which can detect the genetic variability still present in the lines or ecotypes used in mutagenesis experiments, and (ii) that, as discussed below, only a fraction of the more abundant expressed genes was examined.

The 2-DE approach is also relevant when the mutated gene is known since it can provide information on the proteins under the control of the gene. The *Opaque 2* (*O2*) gene in maize that encodes a transcriptional activator of the basic leucine-zipper family [60] exerts pleiotropic effects on protein expression [61, 62] (Fig. 4). Recently, comparative analysis of 2-D patterns between wild-type and mutants in several unrelated backgrounds was carried out to identify specific targets of *O2*, marginal *O2* effects being excluded with this approach. Enzymes belonging to various metabolic pathways were identified, suggesting that the *O2* gene could act as a connecting regulatory gene for different pathways of grain metabolism. In a similar approach, Gottlieb and de Viene [63] compared near-isogenic lines of the *r*-gene in pea, which determined round (*RR*), *vs.* wrinkled (*rr*) seed. Mature seeds of the two lines showed differences for about 10% of the proteins. This high percentage of pleiotropic effects at the proteome level is consistent with the many known physiological differences of the two seed types, and could give us an access to their molecular analysis.

The results above demonstrate that a mutation at a single locus can have pleiotropic effects on the protein pattern, visible using the 2-DE approach. This pleiotropy can then be used to provide a working hypothesis to interpret the physiological process altered by the muation. Note that, in these studies, only partial proteomes (400–600 polypeptides) were studied. These proteomes correspond to the most abundant proteins, soluble ones, ranging between pH 4–8 and 20–100 kDa. A more extensive proteomic approach should be relevant for the analysis of mutant phenotypes. In that view, some groups recently described the use of extensive immobilized pH gradients and acrylamide concentration gels [64, 65] for the recovery of the maximum number of sptos possible. Alternatively, access to minor proteins in 2-DE can be improved by the analysis of proteins from different subcellular compartments. As such, the development of a proteome database of the plasma membrane from *Arabidopsis thaliana* was recently undertaken [66]; this includes protein expression data (see paragraph "plant protein databases") in addition to protein identification data. The extensive proteomic approach also relies on the possibility of handling hydrophobic proteins, which are rarely detectable in 2-D gels [67]. Notwithstanding, recent results obtained with new detergents in combination with thiourea seem promising for the recovery of such proteins in 2-D gels [68].

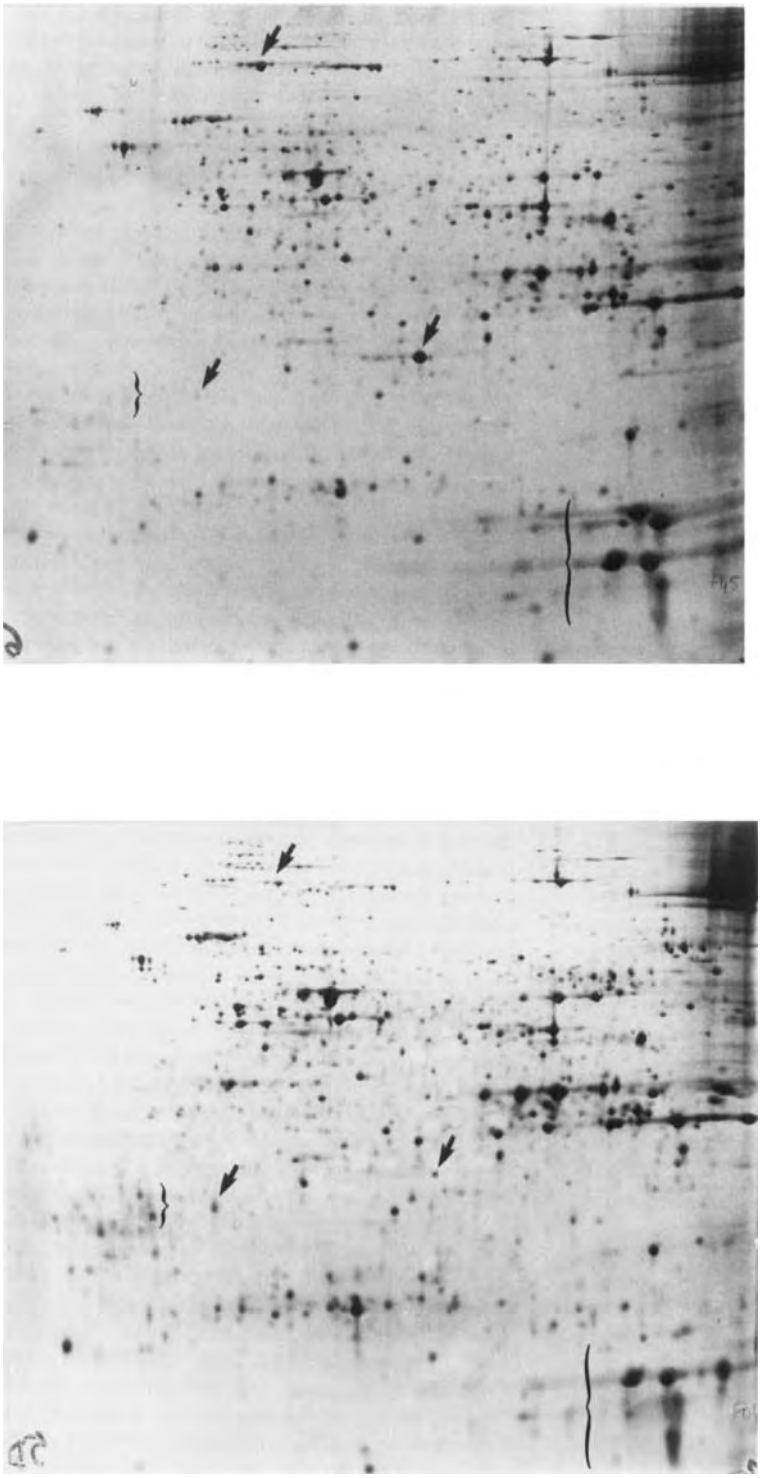


Figure 4. Two maize lines differeing by the *O2* mutation. 2-D gels of proteins from albumen, 45 days after pollination. Top: wild-type; bottom: *O2* mutant. The arrows indicate the variations in gene expression due to a single mutation.

6 Genetic localization of protein loci

Over the last decade, genetic mapping of plant species has undergone explosive growth. The development of new techniques (initially RFLP and now PCR-based) generating many different types of highly polymorpic and abundant DNA-based markers [69], has made possible the construction of genetic linkage maps for virtually any species, allowing geneticists to unravel some of the underlying genetic factors (*i.e.*, QTL) controlling quantitative trait variation (see reviews [70, 71]). While they can be considered as the techniques of choice for quickly saturating a genome, RFLPs and PCR-based markers appear to have a limited value for defining the relationship between phenotypic variation and genes of known function. Alternatively, proteins revealed by 2-DE not only provide useful molecular markers for mapping the expressed genome, but also have the great advantage of providing candidate genes to characterize QTL (see below).

In 2-D patterns, spots exhibiting qualitative variations (position shifts: PS; see Fig. 2, and presence/absence: P/A) were shown to be under monogenic control ([8, 72, 73]; Fig. 2). However the same gene may be represented in 2-D gels by several spots resulting from post-translational modifications or degradation products [9, 74]. PS are those cases where two spots with similar aspects and staining intensity are characterized by the following criteria: (i) They are located relatively close to each other on 2-D gels (*i.e.*, they have close p*I* and/or apparent molecular masses); (ii) they present a monogenic and codominant Mendelian mode of inheritance in a segregating pedigree (1:1 segregation ratio in haploid and double haploid lines; 1:2:1 segregation ratio in F2 selfed families; see [8] for a review), and (iii) last but not least, they correspond to the same protein. This was shown by microsequencing of such allelic pairs in maize and pine [75, 76]. These three criteria demonstrate that PS variations correspond to allelic differences in the primary structure of a protein, most probably due to mutations at the structural locus.

If it is clear that a PS locus corresponds to the structural gene of a protein, this parallelism is less straightforward for spots showing P/A variations. In 2-D gels obtained from F2 offsprings, such proteins revealed a clear monogenic dominant mode of inheritance (3:1 segregation ratio for presence: absence of a spot), showing that a single gene is responsible for this polymorphism. If a P/A variant corresponds to a PS variant, where a member of the pair is not detected, then the P/A locus corresponds to the structural gene of the protein. Conversely, if a P/A variant corresponds to monogenic quantitative differences with a spot below the level of detection, then the P/A locus corresponds to a regulatory gene controlling the expression level of the protein.

In order to give new insight regarding the genetic control of P/A variations, the map location of P/A loci was compared from an F2 progeny of maritime pine using a qualitative and a quantitative approach based on QTL mapping. In the first case, P/A loci were mapped by cosegregation analysis with other P/A and PS variants as well as with PCR-based markers [76, 77]. In the second case [77], the integrated intensity of P/A variants was quantified using an image analysis system, and a QTL analysis was performed for each segregating spot, taking into account the whole data set, *i.e.* $\approx$ 1/4 of F2 individuals characterized by a null intensity (absence of the spot and $\approx$ 3/4 of F2 individuals with nonnull intensity (presence of the spot). The intensity score was used as the trait value when the spot was present. The composite interval mapping method, implemented in the QTL-cartographer software [78], was used to localize QTLs accounting for spot intensity variation. In all cases, a single QTL was detected and its most probable map location was the same as the P/A loci localized by cosegregation analysis. When the genome was rescanned after this major QTL was fixed, no additional significant genomic region was identified. Furthermore, when both members of a PS variant were considered independently (3:1 instead of 1:2:1 segregation ration in the F2 progeny) they both mapped in the same location using either the categorical or quantitative approach. These results support the existence of a single P/A locus corresponding either to a major regulatory gene, or to the structural gene of the protein. Under the latter hypothesis, the alternative allele either exists and is hidden by other spots or does not exist and corresponds to a silent allele.

Genetic localization of protein markers has mainly been reported for wheat, maize and maritime pine, although a few PS loci were mapped in others crops including pea [8] and barley [73]. In wheat, Colas des Francs and Thiellement [48] reported chromosomal localization of 35 proteins using ditelosomic lines. In maritime pine, Bahrman and Damerval [72] reported linkage analysis for 119 protein loci and Gerber *et al.* [79] reported a 65 loci linkage map covering 530 cM of the genome. A three-generation inbred pedigree of this species was recently used to map 61 proteins from haploid [80] and diploid [76, 77] tissues. In maize, a composite linkage map showing the distribution of 65 PS loci was presented by de Vienne *et al.* [8]. In pine and maize, protein loci were found on each chromosome, interspersed with other markers (RFLPs in maize, RAPDs and amplified fragment length polymorphism, AFLP).

RFLP markers developed from cDNA libraries generally reveal several unlinked loci [81–83] making it difficult to identify the gene that is actually expressed. 2-DE of pro-

teins may prove to be an interesting technique for constructing maps of expressed genes and for sorting out which cDNA locus is expressed in which organ/developmental stage. For example, a protein showing a PS variation in leaf, kernel and coleoptile of maize was identified using microsequencing and amino acid composition as a phosphoglycerate mutase. The cDNA probe identified four loci, on three chromosomes. The PS locus in each organ colocalized with one of these loci on chromosome 3, indicating that this locus codes for the expressed protein [84].

To date, no more than about a hundred protein loci have been mapped in a single plant species. The number of such markers could be substantially increased by analyzing some physiologically contrasted organs or tissues (root, leaves, bud, pollen, xylem, *etc.*) of the same individuals, until full coverage of the genome is obtained. Several techniques are now available for protein identification (amino acid sequencing, mass spectrometry, *etc.*). Such maps of expressed and known function genes are of invaluable help for the candidate gene/protein strategy of QTL characterization (see below).

7 Protein quantity loci

The quantity of any given protein in a cell or tissue results from the combined effect of the various factors controlling gene expression, from transcription to post-translational modifications and processing, and protein turnover. Automatic systems now allow quantification of intensity of polypeptide spots revealed by 2-DE. It has been shown that for a large proportion of proteins, the integrated optical density was linearly related to the protein amount [85, 86]. The 2-DE approach thus affords a unique opportunity to investigate the genetic determinination of protein amount for several gene products at a time, whether their function is known or not.

A pioneering study for the detection of regulators *sensu lato* was conducted by comparing the intensity of more than a hundred proteins from etiolated shoots of euploid and ditelosomic lines of the Chinese Spring wheat variety [48]. It was found that several chromosome arms could modulate the expression of a given protein. When comparing two different proteomes (from etiolated shoot and from leaf) it was further shown that they are not the same "regulators" that are acting, according to the developmental stage, on the same protein [87]. In maritime pine, the analysis of 56 endosperms from the same individual not only allowed a genetic map to be constructed, but also the mapping of a few loci controlling discrete variation of protein amount. In this study only this kind of variation was accessible due to visual inspection of the electropherograms [72].

Considering individual protein quantity as any quantitative trait, genetic determinination was investigated in segregating progenies using a QTL methodology [88]. Using a genetic map consisting of 109 RFLP and 2-D protein markers, the PQLs (protein quantity loci) of 72 coleoptile proteins were searched in an F2 segregating progeny of maize [89]. Seventy PQLs were detected for 42 proteins, 20 of them having more than one PQL. The highest number of PQLs was 5 for two proteins. For 69% of the proteins, the variability explained was below 60%, suggesting that other PQLs with low effects remained undetected. PQLs were found to be distributed all over the genome. Dominance was frequently observed, the dominance of the high quantity alleles being three times as frequent as the reverse, which gave genetic support to the heterosis of protein quantity observed in F1 hybrids [15]. Epistatic interactions were involved in the control of several proteins. In maritime pine, investigation of seedling protein quantities in an F2 descent allowed the detection of 1–4 PQLs for half of the ten studied proteins [86]. It was already known that the control of a given gene expression involved numerous factors. The various studies of PQL detection showed that several of these factors can be polymorphic for a given gene, and that they are distributed throughout the genome.

8 Candidate proteins for drought tolerance

Such analyses have also been designed to investigate the genetic control of protein expression under water deficit. Water availability is a major limiting factor for plant growth. Limited water availability leads to several responses that have been described at different levels: reduced growth of aerial parts, stomatal closure, leaf rolling, leaf senescence, synthesis of osmolytes for osmotic adjustment (*e.g.*, betaine, proline, hexoses, inorganic ions). These responses are at least partially controlled by abscisic acid (ABA), a phytohormone whose concentration increases in plants subjected to water deficit. As yield of cultivated plants is limited by drought, there is a special interest in identifying proteins whose expression is modified by the constraint, and whose genetic variation could cause variations of plant tolerance. The study of proteome response to drought, saline stress, or ABA showed that a large number of proteins are quantitatively or qualitatively affected by this stress: Ramani and Apte [90] detected 35 salinity stress-induced and 17 repressed polypeptides in rice seedlings. Costa *et al.* [91] found 38 proteins affected by drought in needles of maritime pine, 24 of them being induced. Riccardi *et al.* [92] found 78 proteins affected in growing parts of maize leaves, 50 of which were being upregulated.

Proteome analysis allowed the identification of proteins already known as typically water-stress-induced: late embryogenesis abundant (LEA) proteins, dehydrins [92, 93], but also of other proteins: some are related to protection against possible damage caused by stress (protease inhibitors, heat shock proteins, enzymes related to oxidative stress), others are related to other metabolic functions (*e.g.*, glycolysis, lignin synthesis) or are of still unknown function [91, 92, 94–97].

Protein response depends on the type of stress: Leone *et al.* [98] showed that the protein response fo *Solanum tuberosum* cell suspensions abruptly submitted to low water potential was different from the response to gradual decrease of the water potential. ABA also induced most of proteins induced by abrupt treatment. The interaction of hormones with response to drought has also been studied: Leymarie *et al.* [58] compared the response of Arabidopsis auxin-insensitive mutants faced with drought, and were able to identify polypeptides whose regulation upon drought was modified by the mutations. Of course, response to water stress is also organ-dependent; roots and leaves show different responses ([92]; Zivy, unpublished results).

A large genetic variability of protein response to drought has been found in maritime pine and maize [91, 92]. For these species, the "candidate protein" strategy is being used in order to identify physiological, possibly causal, relationships between protein expression and tolerance to drought. The starting point of this strategy is that if a quantitative phenotypic trait is at least partially controlled by the quantity of a protein, then the genetic variation of the protein quantity should cause variation of the phenotypic trait. Thus, a locus that affects protein quantity should also affect the phenotypic trait: PQLs and QTLs detected on the same segregating population should be found at the same location on genetic maps. By looking for colocalization of QTLs (in this case, QTLs of traits related to plant response to drought) with PQLs of proteins involved in the studied trait (proteins induced by drought), one adds a genetic argument to a physiological argument, to support the hypothesis of a causal relationship between proteic and phenotypic variation. A recent study on maritime pine showed that for a given protein, the genetic control involved largely different elements in standard as compared to water stress conditions [77]. In maize, three PQLs were detected for a drought-induced protein, which were mapped to the same chromosomal region as QTLs for growth under water deficit, on three chromosomes. The alleles of PQLs responsible for high amounts were associated with PQL alleles for restricted growth. Further experiments are needed to investigate a possible role of this protein in the trait variation. Another example is given by the maize ASR1, a protein of unknown function, that was first described by Iusem *et al.* [99] as inducible by ABA, water stress, and fruit ripening. This protein was found to be induced by drought in only one of the two parental lines of the studied cross. The PQL regulating the expression/nonexpression of this protein was localized to chromosome 10. The cDNA was isolated by M. Hoefer and P. Westhoff (unpublished data), and by using it as a probe, it was localized to the same location as the PQL on chromosome 10. Very likely the PQL is the gene itself, *i.e.*, there is a polymorphism within or close to the coding region that affects the expression upon drought. In the same region of chromosome 10, QTLs for leaf senescence and anther-silking interval (a symptomatic trait of the effect of drought on maize) were detected. Work is in progress to confirm the role of this protein in the degree of leaf senescence [84].

9 Plant protein databases

The availability of standardized and reproducible procedures for the analysis of proteins by 2-DE made large-scale investigations possible. In combination with protein identification with the help of various approaches (*e.g.*, protein immunodetection, microsequencing, mass spectrometry of peptides, *etc.*), this allows one to compile extensive information about protein expression. Such information can then be stored in protein databases, together with links to other kinds of databases. Table 1 summarizes the main features of plant protein databases presently accessible through the Internet. Available databases concern the plant model *Arabidopsis thaliana*, important crop species such as rice or maize, and the woody plant *Pinus pinaster.* All offer clickable images of 2-D gels with links for annotated proteins, and most of the latter are identified (Fig. 5). In terms of analysis of genome expression, all these databases include comparisons of polypeptide patterns in various plant organs (root, leaf, stem, bud, seed) or tissues (callus, xylem, phloem), or between genotypes. In addition, physiological data are also available in some cases, such as effects of seasonal variations or water deficit on maritime pine needle proteins, as well as genetic data such as the localization of the corresponding genes on a linkage map.

Among genome-wide approaches, one unique advantage of proteomic investigations is that they give access to the analysis of genome expression at the subcellular level. This opportunity was first examined in tobacco [51, 100]. A striking result was the observation of an unexpectedly high plasticity of the plasma membrane proteome during plant development, even under constant environmental conditions. In a more comprehensive work on Arabidopsis, a similar approach led to the demonstration that

Table 1. Progress with plant proteomes

Species	Proteins characterized (% identified)	Reference	WWW address	Genome project
Arabidopsis thaliana	57 (63%)	[102]	www.rs.noda.sut.ac.jp/~kamom/2de/2d.html	[103]
	62 (56.5%)	[104]		
	150 (50%)	[66]	sphinx.rug.ac.be:8080/ppmdb/index.html	
Oriza sativa	50 (58%)	[105]	wwr.rs.noda.sut.ac.jp/~kamom/2dc/2d.html	[106]
	59 (56%)	[108]		[107]
	51 (53%)	[104]		[109]
Pinus pinaster	65 (90%)	[110]	www.pierroton.inra.fr/genetics/2D/	[111]
				[112]
Zea mays	74 (48.5%)	[113]	moulon.inra.fr/imgd	[114]
	80 (67.5%)	Zivy, M. (unpublished)		[115]

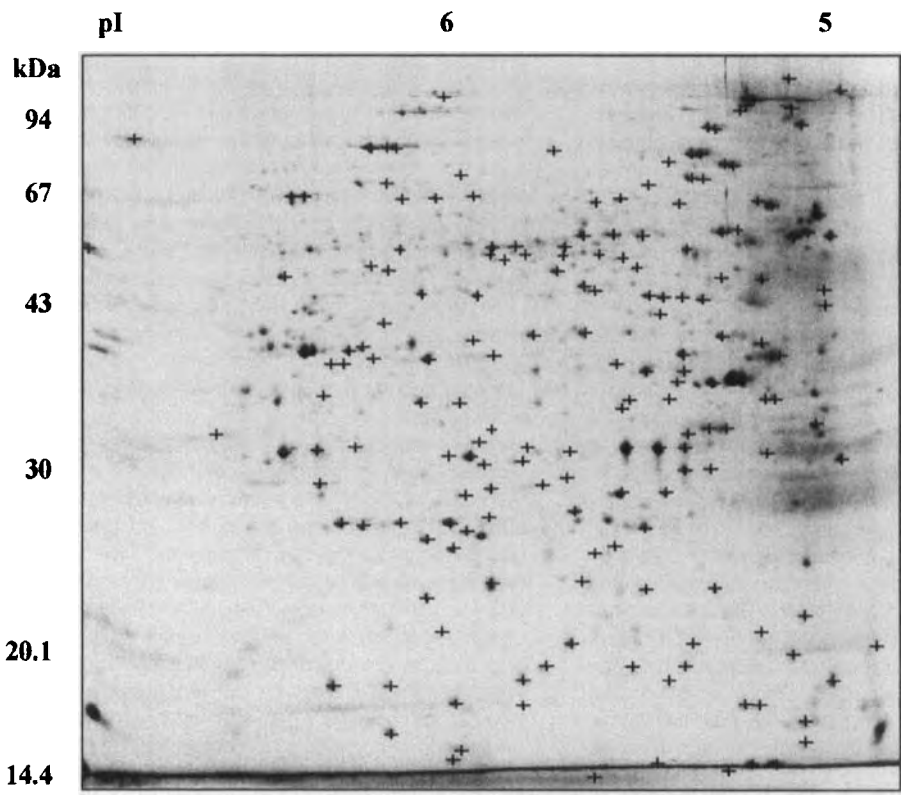


Figure 5. Arabidopsis plasma membrane annotated map, an example of available plant proteome resources (http://sphinx.rug.ac.bc:8080/ppmbd/index.html.)

numerous membrane peripheral proteins display various subcellular locations in the cell [66]. On the other hand, as for other biological materials, the access to very hydrophobic proteins appears to constitute a limitation of proteomic approaches in plants. However, recent improvements were reported with the use of new detergents [68, 101].

Another important feature of annotated protein databases is that they create links with genome projects. In fact, all

of the available databases concern plant species for which systematic sequencing was undertaken (Table 1), either at the genomic level (Arabidopsis) or at the expressed genome level (ESTs: Arabidopsis, corn, rice, pine tree). Interestingly, in the case of the Arabidopsis plasma membrane database, numerous proteins were found to correspond to unknown ESTs, thus giving the first information on the subcellular expresion of the encoding genes. Furthermore, using both 2-D gel data and high-density filters immobilizing Arabidopsis ESTs, this allowed the first comparison in plants between the accumulation of transcripts and that of corresponding proteins (Scheideler *et al.*, unpublished results).

10 Conclusions

As can be deduced from the examples given above, proteomics has become an essential field of research in plant biology where the combined approaches of genetics, physiology and molecular biology will, in the coming years, provide essential tools to understand the mechanisms underlying plant growth and development. These breakthroughs will permit further improvement of crops and crop management by furnishing new tools to plant breeders. As in medicine and, more generally, in the "mainstream" of biology [116, 117] proteome approaches are of the foremost importance in plant sciences.

The large genetic variability of gene expression seems to be a major biological fact. With the advent of chip/microarray methods for analyzing the expression of thousands of genes simultaneously [118, 119], proteome analysis will represent an essential complement for studying the physiological consequences of this variability. Thus proteomics appears to be a relevant field for the study of the relationships between genotype and phenotype.

Received March 5, 1999

11 References

[1] Cooke, R., Raynal, M., Laudié, M., Grellet, F., Delseny, M., Morris, P. C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y. F., Mache, R., Krivitzky, M., Gy, I. J. J., Kreis, M., Lecharny, A., Parmentier, Y., Marbach, J., Fleck, J., Clément, B., Phillips, G., Hervé, C., Bardet, C., Tremousaygue, D., Lescure, B., Lacomme, C., Roby, D., Jourjon, M. F., Chabrier, P., Charpenteau, J. L., Desprez, T., Amselem, J., Chiapello, H., Höfte, H., *Plant J.* 1996, *9*, 101–124.

[2] Novak, R., *Science* 1995, *270*, 368–371.

[3] Pennisi, E., *Science* 1996, *272*, 1736–1738.

[4] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Biotechnol. Genet. Engineer. Rev.* 1996, *13*, 19–50.

[5] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[6] Klose, J., *Humangenetik* 1975, *26*, 231–243.

[7] Damerval, C., Zivy, M., Granier, F., de Vienne, D., in: Chrambach, A., Dunn, M. J., Radola, B. J. (Eds.), *Advances in Electrophoresis*, VCH Weinheim 1988, pp. 263–340.

[8] de Vienne, D., Burstin, J., Gerber, S., Leonardi, A., Le Guilloux, M., Murigneux, A., Beckert, M., Bahrman, N., Damerval, C., Zivy, M., *Heredity* 1996, *76*, 166–177.

[9] Zivy, M., Thiellement, H., de Vienne, D., Hofmann, J.-P., *Theor. Appl. Genet.* 1983, *66*, 1–7.

[10] Zivy, M., Thiellement, H., de Vienne, D., Hofmann, J.-P., *Theor. Appl. Genet.* 1984, *68*, 335–345.

[11] Damerval, C., de Vienne, D., Zivy, M., Thiellement, H., *Electrophoresis* 1986, *7*, 52–54.

[12] Anderson, N. G., Tollaksen, S. L., Pascoe, F. H., Anderson, L., *Crop Sci.* 1985, *25*, 667–674.

[13] Leonardi, A., Damerval, C., de Vienne, D., *Genet. Res. Camb.* 1987, *50*, 1–5.

[14] Leonardi, A., Damerval, C., de Vienne, D., *Genet. Res. Camb.* 1988, *52*, 97–103.

[15] de Vienne, D., Leonardi, A., Damerval, C., *Electrophoresis* 1988, *9*, 742–750.

[16] Damerval, C., Hébert, Y., de Vienne, D., *Theor. Appl. Genet.* 1987, *74*, 194–202.

[17] Zivy, M., *Theor. Appl. Genet.* 1987, *74*, 209–213.

[18] Branlard, G., *Theor. Appl. Genet.* 1983, *64*, 155–162.

[19] Picard, P., Bourgoin-Grenèche, M., Zivy, M., *Electrophoresis* 1997, *18*, 174–181.

[20] Weiss, W., Huber, G., Engel, K.-H., Pethran, A., Dunn, M. J., Gooley, A. A., Görg, A., *Electrophoresis* 1997, *18*, 826–833.

[21] Görg, A., Postel, W., Domscheit, A., Günther, S., *Electrophoresis* 1988, *9*, 681–692.

[22] Görg, A., Postel, W., Baumer, M., Weiss, W., *Electrophoresis* 1992, *13*, 192–203.

[23] Saruyama, H., Shinbashi, N., *Theor. Appl. Genet.* 1992, *84*, 947–951.

[24] Abe, T., Gusti, R. S., Ono, M., Sasahara, T., *Genes Genet. Syste.* 1996, *71*, 63–68.

[25] Hagen, G., Rubenstein, I., *Plant. Sci. Lett.* 1980, *19*, 217–223.

[26] Wall, J. S., Fey, D. A., Paulis, J. W., *Cereal Chem.* 1984, *61*, 141–146.

[27] Ramagopal, S., *Theor. Appl. Genet.* 1990, *79*, 297–304.

[28] Kajiwara, H., Hirano, H., *Japan J. Breed.* 1992, *42*, 727–735.

[29] Posch, A., van den Berg, B. M., Burg, H. C., Görg, A., *Electrophoresis* 1995, *16*, 1312–1316.

[30] van Telgen, H. J., van Loon, L. C., *J. Plant. Physiol.* 1985, *118*, 285–296.

[31] Bahrman, N., de Vienne, D., Thiellement, H., Hofmann, J.-P., *Biochem. Genet.* 1985, *23*, 247–255.

[32] Sieffert, A., *Trees* 1988, *2*, 188–193.

[33] Leonardi, A., Damerval, C., Hébert, Y., Gallais, A., de Vienne, D., *Theor. Appl. Genet.* 1991, *82*, 552–560.

[34] Burstin, J., de Vienne, D., Dubreuil, P., Damerval, C., *Theor. Appl. Genet.* 1994, *89*, 943–950.

[35] Posch, A., van den Berg, B. M., Postel, W., Görg, A., *Electrophoresis* 1992, *13*, 774–777.

[36] Posch, A., van den Berg, B. M., Duranton, C., Görg, A., *Electrophoresis* 1994, *15*, 297–304.

[37] Hirano, H., *Phytochemistry* 1982, *21*, 1513–1518.

[38] Bon, M. C., *Electrophoresis* 1989, *10*, 530–532.

[39] Bahrman, N., Zivy, M., Damerval, C., Baradat, P., *Theor. Appl. Genet.* 1994, *88*, 407–411.

[40] Petit, R. J., Bahrman, N., Baradat, P., *Heredity* 1995, *75*, 382–389.

[41] Barreneche, T., Bahrman, N., Kremer, A., *For. Genet.* 1996, *3*, 89–92.

[42] David, J. L., Zivy, M., Cardin, M.-L., Brabant, P., *Theor. Appl. Genet.* 1997, *95*, 932–941.

[43] Bahrman, N., Cardin, M.-L., Seguin, M., Zivy, M., Thiellement, H., *Heredity* 1988, *60*, 87–90.

[44] Bahrman, N., Zivy, M., Thiellement, H., *Heredity* 1988, *61*, 473–480.

[45] Thiellement, H., Seguin, M., Bahrman, N., Zivy, M., *J. Mol. Evol.* 1989, *28*, 89–94.

[46] Zivy, M., El Madidi, S., Thiellement, H., *Electrophoresis* 1995, *16*, 1295–1300.

[47] Sears, E. R., *Univ. Mo. Agric. Exp. Stn. Res. Bull.* 1954, *572*, 3–58.

[48] Colas des Francs, C., Thiellement, H., *Theor. Appl. Genet.* 1985, *71*, 31–38.

[49] Galili, G., Geldman, M., *Can. J. Genet. cytol.* 1984, *26*, 651–656.

[50] Bahrman, N., Thiellement, H., *Theor. Appl. Genet.* 1987, *74*, 218–223.

[51] Masson, F., Rossignol, M., *Plant J.* 1995, *8*, 77–85.

[52] Santoni, V., Bellini, C., Caboche, M., *Planta* 1994, *192*, 557–566.

[53] Herbik, A., Giritch, A., Horstmann, C., Becker, R., Balzer, H.-J., Bäulmein, H., Stephan, U. W., *Plant Physiol.* 1996, *111*, 533–540.

[54] von Wiren, N., Peltier, J. B., Rouquié, D., Rossignol, M., Briat, J. F., *Plant Physiol. Biochem.* 1997, *35*, 945–950.

[55] Kasten, B., Buck, F., Nuske, J., Reski, R., *Planta* 1997, *201*, 261–272.

[56] Santoni, V., Delarue, M., Caboche, M., Bellini, C., *Planta* 1997, *202*, 62–69.

[57] Faure, J.-D., Vittorioso, P., Santoni, V., Fraisier, V., Prinsen, E., Barlier, I., Van Onckelen, H., Caboche, M., Bellini, C., *Development* 1998, *125*, 909–918.

[58] Leymarie, J., Damerval, C., Marcotte, L., Combes, V., Vartaniana, N., *Plant Cell Physiol.* 1996, *37*, 966–975.

[59] Tacchini, P., Fink, A., Thiellement, H., Greppin, H., *Physiol. Plant.* 1995, *93*, 312–316.

[60] Schmidt, R. J., Burr, F. A., Aukerman, M. J., Burr, B., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 46–50.

[61] Damerval, C., de Vienne, D., *Heredity* 1993, *70*, 38–51.

[62] Damerval, C., Le Guilloux, M., *Mol. Gen. Genet.* 1998, *257*, 354–361.

[63] Gottleib, L. D., de Vienne, D., *Genetics* 1988, *119*, 705–710.

[64] Wasinger, V. C., Bjellqvist, B. J., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[65] Urquhart, b. L., Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B. J., Britton, W. L., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1384–1392.

[66] Santoni, V., Doumas, P., Rouquié, D., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouzé, P., Dehais, P., Sahoun, I., Barlier, I., Rossignol, M., *Plant J.* 1998, *16*, 633–641.

[67] Adessi, C., Miege, C., Albrieux, C., Rabilloud, T., *Electrophoresis* 1997, *18*, 127–135.

[68] Chevallet, M., Santoni, V., Poinas, A., Rouquié, D., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1909.

[69] Karp, A., Isaac, P. G., Ingram, D. S., *Molecular Tools for Screening Biodiversity*, Chapman & Hall, London 1998, pp. 71–212.

[70] Kearsey, M. J., Farquhar, A. G. L., *Heredity* 1998, *80*, 137–142.

[71] Tanksley, S. D., *Ann. Rev. Genet.* 1993, *27*, 205–233.

[72] Bahrman, N., Damerval, C., *Heredity* 1989, *63*, 267–274.

[73] Zivy, M., Devaux, P., Blaisonneau, J., Jean, R., Thiellement, H., *Theor. Appl. Genet.* 1992, *83*, 919–924.

[74] Colas des Francs, C., Thiellement, H., de Vienne, D., *Plant Physiol.* 1985, *78*, 178–182.

[75] Touzet, P., Morin, C., Damerval, C., Le Guilloux, M., Zivy, M., de Vienne, D., *Electrophoresis* 1995, *16*, 1289–1294.

[76] Plomion, C., Costa, P., Bahrman, N., *Silvae Genet.* 1997, *46*, 161–165.

[77] Costa, P., *PhD Thesis*, Nancy 1999.

[78] Basten, C. J., Weir, B. S., Zeng, Z. B., *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*, NCSU, Raleigh, USA 1997.

[79] Gerber, S., Rodolphe, F., Bahrman, N., Baradat, P., *Theor. Appl. Genet.* 1993, *85*, 521–528.

[80] Plomion, C., Bahrman, N., Durel, C.-E., O'Malley, D. M., *Heredity* 1995, *74*, 661–668.

[81] Helentjaris, T., Weber, D., Wright, S., *Genetics* 1988, *118*, 353–363.

[82] Song, K. M., Suzuki, J. Y., Slocum, M. K., Williams, P. H., Osborn, T. C., *Theor. Appl. Genet.* 1991, *82*, 296–304.

[83] Devey, M. E., Fiddler, T. A., Liu, B.-H., Knapp, S. J., Neale, B. D., *Theor. Appl. Genet.* 1994, *88*, 273–278.

[84] de Vienne, D., Leonardi, A., Damerval, C., Zivy, M., *J. Exp. Bot.* 1999, in press.

[85] Damerval, C., *Electrophoresis* 1994, *15*, 1573–1579.

[86] Costa, P., Plomion, C., *Silvae Genet.* 1999, in press.

[87] Thiellement, H., Bahrman, N., Colas des Francs, C., *Theor. Appl. Genet.* 1986, *73*, 246–251.

[88] Lander, E. S., Botstein, D., *Genetics* 1989, *121*, 185–199.

[89] Damerval, C., Maurice, A., Josse, J. M., de Vienne, D., *Genetics* 1994, *137*, 289–301.

[90] Ramani, S., Apte, S. K., *Biochem. Biophys. Res. Comm. 1997, 223, 663–667.*

[91] Costa, P., Bahrman, N., Frigerio, J.-M., Kremer, A., Plomion, C., *Plant Mol. Biol.* 1998, *38*, 587–596.

[92] Riccardi, F., Gazeau, P., de Vienne, D., Zivy, M., *Plant Physiol.* 1998, *117*, 1253–1263.

[93] Moons, A., Bauw, G., Prinsen, E., Van Montagu, M., Van der Straten, D., *Plant Physiol.* 1995, *107*, 177–186.

[94] Reviron, M. P., Vartanian, N., Sallantin, M., Huet, J. C., Pernollet, J. V., de Vienne, D., *Plant Physiol.* 1992, *100*, 1486–1493.

[95] Downing, W. L., Mauxion, F., Fauvarque, M. O., Reviron, M. P., de Vienne, D., Vartanian, N., Giraudat, J., *Plant J.* 1992, *2*, 685–693.

[96] Pruvot, G., Cuine, S., Peltier, G., Rey, P., *Planta* 1996, *198*, 471–479.

[97] Rey, P., Pruvot, G., Becuwe, N., Eymery, F., Rumeau, D., Peltier, G., *Plant J.* 1998, *13*, 97–107.

[98] Leone, A., Costa, A., Tucci, M., Grillo, S., *Plant Physiol.* 1994, *106*, 703–712.

[99] Iusem, N. D., Bartholomew, D. M., Hitz, W. D., Scolnik, P. A., *Plant Physiol.* 1993, *102*, 1353–1354.

[100] Rouquié, D., Peltier, J.-B., Marquis-Mansion, M., Tournaire, C., Doumas, P., Rossignol, M., *Electrophoresis* 1997, *18*, 654–660.

[101] Santoni, V., Rabilloud, T., Doumas, P., Rouquié, D., Mansion, M., Kieffer, S., Garin, J., Rossignol, M., *Electrophoresis* 1999, *20*, in press.

[102] Kamo, M., Kawakami, T., Miyatake, N., Tsugita, A., *Electrophoresis* 1995, *16*, 423–430.

[103] Cooke, R., Mache, R., Höfte, H., in: Foster, G. D., Twell, D. (Eds.), *Plant Gene Isolation: Principles and Practice*, J. Wiley & Sons, London 1996, pp. 401–419.

[104] Tsugita, A., Kamo, M., Kawakami, T., Ohki, Y., *Electrophoresis* 1996, *17*, 855–865.

[105] Komatsu, S., Kajiwara, H., Hirano, H., *Theor. Appl. Genet.* 1993, *86*, 935–942.

[106] Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, H., Sue, N., Wu, J., Antonio, B. A., Shomura, A., Shimizu, T., Lin, S. Y., Inoue, T., Fukuda, A., Shimano, T., Kuboki, Y., Toyama, T., Miyamoto, Y., Kirihara, T., Hayasaka, K., Miyao, A., Monna, L., Zhong, H. S., Tamura, Y., Wang, Z. X., Momma, T., Umehara, Y., Yano, M., Sasaki, T., Minobe, Y., *Nature Genet.* 1994, *8*, 365–372.

[107] Sasaki, T., Minobe, Y., *Plant J.* 1994, *6*, 615–624.

[108] Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N., Nozu, Y., *Electrophoresis* 1994, *15*, 708–720.

[109] Yamamoto, K., Sasaki, T., *Plant. Mol. Biol.* 1997, *35*, 135–144.

[110] Costa, P., Pionneau, C., Bauw, G., Dubos, C., Bahrman, N., Kremer, A., Frigerio, J.-M., Plomion, C., *Electrophoresis* 1999, *20*, in press.

[111] Neale, D. B., Kinlaw, C. S., Sewell, M. M., *Forest Genet.* 1994, *1*, 197–206.

[112] Allona, I., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J., Riedl, J., Retzel, E., Campbell, M., Sederoff, R., Whetten, R. W., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 9693–9698.

[113] Touzet, P., Riccardi, F., Morin, C., Damerval, C., Huet, J.-C., Pernollet, J.-C., Zivy, M., de Vienne, D., *Theor. Appl. Genet.* 1996, *93*, 997–1005.

[114] Keith, C. S., Hoang, D. O., Barrett, B. M., Feigelman, B., Nelson, M. C., Thai, H., Baysdorfer, C., *Plant Physiol.* 1993, *101*, 329–332.

[115] Shen, B., Carneiro, N., Torres-Jerez, I., Stevenson, R., McGreery, T., Helentjaris, T., Baysdorfer, C., Almira, E., Ferl, R., Habben, J., Larkins, B., *Plant Mol. Biol.* 1994, *26*, 1085–1101.

[116] Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P., *Electrophoresis* 1997, *18*, 1217–1242.

[117] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., *Proteome Research: New Frontiers in Functional Genomics*, Springer-Verlag, Berlin 1997, 243 p.

[118] Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., Fodor, S. P. A., *Science* 1996, *274*, 610–614.

[119] Yershov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S., Mirzabekov, A., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 4913–4918.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 2100–2110

Review

Proteomics in human disease: Cancer, heart and infectious diseases

Peter R. Jungblut[1,7]
Ursula Zimny-Arndt[1]
Evelyn Zeindl-Eberhart[2]
Jiri Stulik[3]
Kamila Koupilova[3]
Klaus-Peter Pleißner[4]
Albrecht Otto[5]
Eva-Christina Müller[5]
Wanda Sokolowska-Köhler[6]
Gertrud Grabher[7]
Georg Stöffler[7]

[1]Max-Planck-Institut für Infektionsbiologie, Protein Analyse Einheit, Berlin, Germany
[2]Ludwig-Maximilians-Universität München, Pathologisches Institut, München, Germany
[3]Purkyne Military Medical Academy, Institute for Immunology, Hradec Kralove, Czech Republic
[4]Deutsches Herzzentrum Berlin, Germany
[5]Max-Delbrück-Center, Proteinchemie, Berlin, Germany
[6]Institut für Mikrobiologie und Hygiene, Humboldt-Universität, Charité, Berlin, Germany
[7]Institut für Mikrobiologie, Medizinische Fakultät, Universität Innsbruck, Austria

In recent years, genomics has increased the understanding of many diseases. Proteomics is a rapidly growing research area that encompasses both genetic and environmental factors. The protein composition represents the functional status of a biological compartment. The five approaches presented here resulted in the detection of disease-associated proteins. Calgranulin B was upregulated in colorectal cancer, and hepatoma-derived aldose reductase-like protein was reexpressed in a rat model during hepatocarcinogenesis. In these two investigations, attention was focused on one protein, obviously differing in amount, directly after two-dimensional electrophoresis (2-DE). Additional methods, such as enzyme activity measurements and immunohistochemistry, confirmed the disease association of the two candidates resulting from 2-DE subtractive analysis. The following three investigations take advantage of the holistic potential of the 2-DE approach. The comparison of 2-DE patterns from dilated cardiomyopathy patients with those of controls revealed 25 statistically significant intensity differences, from which 12 were identified by amino acid analysis, Edman degradation or matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS). A human myocardial 2-DE database was constructed, containing 3300 protein spots and 150 identified protein species. The number of identified proteins was limited by the capacity of our group, rather than by the principle of feasibility. Another field where proteomics proves to be a valuable tool in identifying proteins of importance for diagnosis is proteome analysis of pathogenic microorganisms such as *Borrelia burgdorferi* (Lyme disease) and *Toxoplasma gondii* (toxoplasmosis). Sera from patients with early or late symptoms of Lyme borreliosis contained antibodies of various classes against about 80 antigens each, containing the already described antigens OspA, B and C, flagellin, p83/100, and p39. Similarly, antibody reactivity to seven different marker antigens of *T. gondii* allowed differentiation between acute and latent toxoplasmosis, an important diagnostic tool in both pregnancy and immunosuppressed patients.

Keywords: Borrelia / Colorectal cancer / Dilated cardiomyopathy / Hepatocellular carcinomas / Toxoplasma / Review EL 3500

Contents

Correspondence: Dr. Peter R. Jungblut, Max Planck Institute for Infection Biology, Protein Analysis Unit, Monbijoustr. 2, D-10117 Berlin, Germany
E-mail: jungblut@mpiib-berlin.mpg.de
Fax: +49-30-2846-0174

Abbreviation: DCM, dilated cardiomyopathy

1 Introduction

Disease may be caused by an exchange of single base pairs within the genome. The resulting exchange of Glu to Val in position 6 of β-hemoglobin leads to sickle-cell ane-

mia if both alleles are affected. However, most diseases are more complex and the elucidation of pathomechanisms is more complicated. Methods adequate to deal with this complexity are necessary for understanding the molecular mechanisms involved during disease development. As the acting macromolecules in cells, proteins are the main candidates for disease targets. Today, the only method for resolving proteins from complex mixtures like cells or tissues is two-dimensional electrophoresis (2-DE). The technique with the highest resolution separates 10 000 protein species in one experiment [1]. The term "protein species" describes a chemically clearly defined molecule [2].

Different strategies for the identification of proteins from gels have been developed and are summarized by Jungblut *et al.* [2]. Two mass spectrometric methods have been established for protein identifications from gels: matrix-assisted laser desorption/ionization (MALDI-MS) [3] and electrospray ionization (ESI-MS) [4] mass spectrometry. Both strategies have the potential to identify even low intensity spots on 2-DE gels, and post-translational modifications may also be analyzed. The sensitivity of 2-DE-separated protein identification was improved in the last ten years from about 50 pmol to 100 fmol. For identification of proteins from organisms with completely sequenced genomes, peptide mass fingerprints are sufficient for identification (Jungblut *et al.*, submitted). These fingerprints may be obtained by MALDI-MS or ESI-MS. If there is no corresponding gene sequence of the protein under investigation in the sequence databases, protein sequence information is necessary to find the gene within the organism or to identify the protein by comparison with proteins of other organisms in the sequence databases. Sequence information is obtained by Edman degradation [5], post-source-decay MALDI-MS [6], or fragmentation in an ESI mass spectrometer [7] with extreme sensitivity using nanospray ESI-MS [8]. The comprehensive investigation of the proteins of a biological compartment has been named systematic analysis of proteins [9], proteins analyzed on a genomic scale [10] or proteome analysis [11]. An overview is shown in Fig. 1.

The potential of 2-DE in clinical investigations was recognized with the first publications of the combination of isoelectric focusing with SDS-PAGE [12–15]. The state of the art of the beginning 1980s was summarized in the two special issues of *Clinical Chemistry* from 1982 [16] and 1984 [17]. The journal *Electrophoresis* has continued this tradition [18, 19]. Hochstrasser [20] and Hochstrasser and Tissot [21] reviewed clinical applications extensively. 2-DE databases contain disease-associated proteins resulting from, for example, skin diseases [22], bladder cancer [22] and heart diseases [23–26]. One excellent

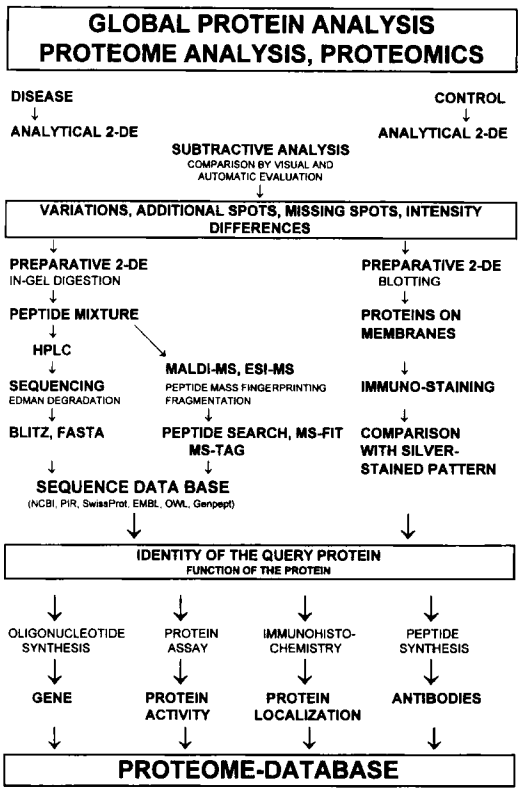


Figure 1. Proteome analysis. Strategies to detect and analyze disease-associated proteins.

example is the detection of the disease-associated spots 130/131, selectively present in the cerebrospinal fluid of patients with Creutzfeldt-Jakob disease, and identified as members of the 14-3-3 protein family [27].

Here we review some aspects of our approaches concerning the detection and characterization of disease-associated proteins. We used 2-DE analysis for two cancer investigations with the aim to reduce our view for the moment to only one tumor-associated protein, each. The tumor association was confirmed by enzyme activity measurements or by immunohistochemistry. The dilated cardiomyopathy approach has shown that series of disease-associated proteins may be elucidated and identified. A myocardial 2-DE database has been established. Proteomics of microorganisms has the advantage of a reduced number of genes, a clear definition of surrounding parameters, and if the genome is sequenced, an easy identification of the protein species. It helps to detect antigens, virulence factors and vaccine candidates.

2 Proteome analysis for the detection of tumor-associated proteins

2.1 Colorectal cancer

Colorectal tumorigenesis is a multistep process involving numerous gene mutations leading to the loss of function of tumor suppressor genes, as well as the activation of oncogenes [28]. Yet, the precise mechanism by which tumors develops remains elusive. Recently, we have started a project directed towards the analysis of polypeptide changes associated with malignant transformation of colon mucosa. We ran 2-DE of 15 colorectal carcinomas and 13 samples of normal colonic epithelia. The polypeptide pattern in each sample was computer-evaluated using Melanie I 2-DE analysis software. In this way, we formed master gels of tumor tissue and normal colonic mucosa containing 882 and 861 spots, respectively. By comparing these master gels we detected a protein with a molecular mass of 13 kDa and a p*I* value of 5.6, whose expression was restricted to tumor tissue only (Fig. 2 A, B). The results show that from the group of 15 carcinomas the 13/5.6 protein was upregulated in 13 samples (87%). Furthermore, this protein was detected in all cases of early localized adenocarcinomas in B stage and left-sided adenocarcinomas, which were predominant in our collection.

To verify the presence of 13/5.6 protein in precancerous lesions we performed 2-DE analyses of 7 adenomoatous polyps differing in the degree of dysplasia. We found an overexpression of 13/5.6 protein in both cases of polyps with a moderate degree of dysplasia (Fig. 2C) and also in polyps of low-degree dysplasia collected from patients suffering either from colorectal carcinoma or ulcerative colitis for 24 years. In contrast, the remaining polyps of low-degree dysplasia expressed only barely detectable amounts of the 13/5.6 protein (Fig. 2D). For the identification of 13/5.6 protein we first tried to compare our 2-DE maps with the reference protein map of liver tissue located on the ExPASy molecular biology server [29] or with the 2-DE protein map of colon carcinoma cell lines [30], but a distinct assignment was not possible. Therefore we used 16 spots directly excised from 2-DE gels for internal sequencing. After trypsin digestion in the gel and separation of the resulting peptides by μ-HPLC, internal sequencing of two peptides was successful. We obtained two sequences, LGHPDTLNQ and VIEHMEDLDTNADK, that matched perfectly the sequences of two peptides from calgranulin B. The identity of calgranulin B was further verified by MALDI-MS analysis of several HPLC-separated peptide fractions [31]. This finding of high specificity of calgranulin B for preneoplastic and neoplastic tissues extends the earlier results describing the elevated levels of heterodimeric protein calprotectin, composed of calgranulin B and A, in samples of stool collected from patients suffering from gastrointestinal neoplasia [32]. The exact function of calgranulin B or calprotectin in the progress of malignant transformation needs to be further clarified.

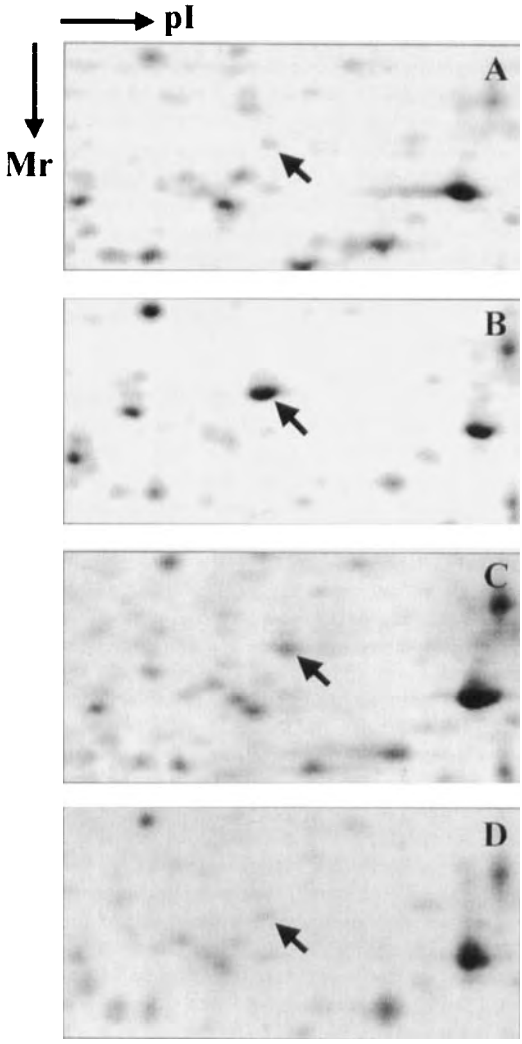


Figure 2. Comparison of 13/5.6 protein expression in normal, preneoplastic and neoplastic colonic mucosa. Only the 13/5.6 protein regions from originally silver-stained 2-DE protein patterns are shown. For 2-DE immobilized pH gradient IEF was combined with SDS-PAGE. (A) Control colonic tissue; (B) colon adenocarcinoma; (C) adenomatous polyp with moderate dysplasia; (D) tubulovillous polyp with low grade dysplasia.

2.2 Hepatocellular carcinomas

Misprogramming of genetic information in cancer leads to quantitative and/or qualitative protein alterations, which can be studied by high resolution 2-DE. 2-DE has been used successfully to detect tumor-associated proteins in chemically induced rat hepatomas. 2-DE in combination with protein-chemical methods enabled us to further characterize these tumor-associated proteins. As described previously we detected several protein variants in *N*-methyl-*N*-nitrosourea-induced rat hepatomas by 2-DE and identified one of these variants by internal amino acid microsequencing as hepatoma-derived aldose reductase-like protein (35 kDa/p*I* 7.4) [33, 34]. A related aldose reductase isoform discovered in rat lens by 2-DE (37 kDa/p*I* 6.8) disclosed after amino acid microsequencing 80% sequence identity to hepatoma-derived aldose reductase-like protein but showed 98.5% homology to the known rat lens aldose reductase sequence. This suggests that rat hepatoma-derived aldose reductase-like protein and lens aldose reductase are encoded by two related genes. The two aldose reductase isoforms are expressed differently in various rat organs. Lens aldose reductase is expressed in heart, brain, muscle, lung, duodenum, spleen, kidney, bone marrow and erythrocytes, while hepatoma-derived aldose reductase-like protein is preferentially expressed in hepatomas and in embryonic liver (Fig. 3) [34].

The expression of the two aldose reductase isoforms is modulated by growth factors. An identical increment of both types was found in transformed rat liver cell lines after fibroblast growth factor 1 (FGF-1) stimulation *in vitro*. This indicates that the two proteins belong to a family of growth factor responsive enzymes, but are present in different isoforms depending on the physiological or neoplastic state of an organ [34]. Immunohistochemistry using an FR-1 antibody directed against hepatoma-derived aldose reductase-like protein, but not against lens aldose reductase, revealed that hepatoma-derived aldose reductase-like protein is strongly expressed already in preneoplastic and in early neoplastic stages of chemically induced hepatocarcinogenesis, but not in normal surrounding liver tissue (Fig. 4).

Aldose reductase (E.C.1.1.1.21) is a member of the aldo-keto reductase superfamily, which catalyses the conversion of glucose to sorbitol in the sorbitol pathway and is involved in the pathogenesis of various diabetic complications [35]. Aldose reductase was described as an enzyme catalyzing the reduction of various xenobiotic and endogenous compounds involved in detoxification of a broad range of substrates [35]. During hepatocarcinogenesis an increased level and/or activity of various detoxifying enzymes was described, *e.g.*, placental type of glutathione-S-transferase (GST-P) [36], or a carcinogen

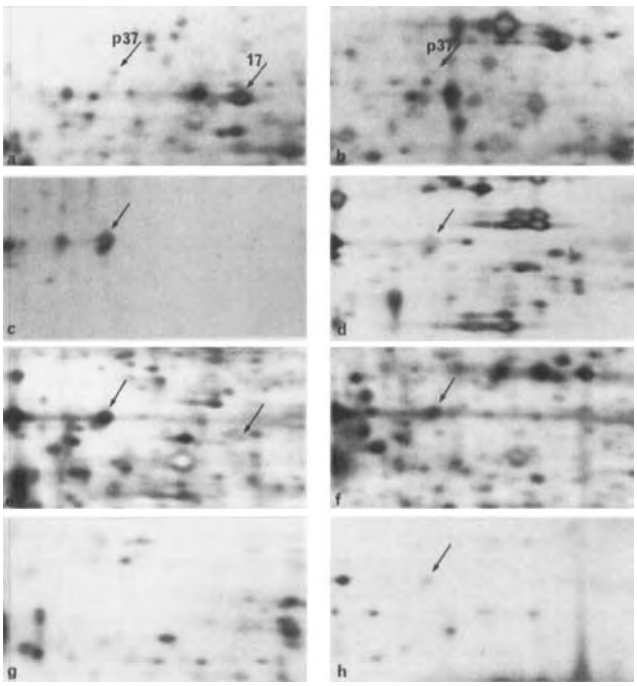


Figure 3. Presence of hepatoma-derived aldose reductase-like protein (spot 17) and lens aldose reductase (p 37) in various rat organs, erythrocytes, and serum. Parts of 2-DE gels prepared with the ISO-DALT system (16 × 16 × 0.15 cm) are shown for (a) hepatoma, (b) normal liver, (c) eye lens, (d) muscle, (e) heart, (f) brain, (g) serum, and (h) erythrocytes [34]. One hundred µg of soluble proteins were separated by 2-DE. Proteins were stained by polychromatic silver staining. Reproduced from [34], with permission.

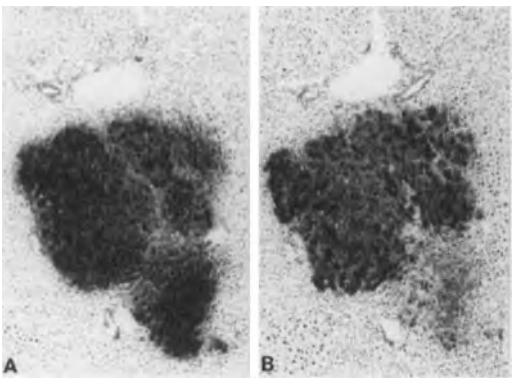


Figure 4. Immunohistochemistry of serial sections of a neoplastic region in rat liver. Immunostaining of Carnoy-fixed sections with (A) GST-P antibody as positive control (1:1000), and (B) FR-1 antibody recognizing hepatoma-derived aldose reductase-like protein (1:75) for 17 h at 4°C using the avidin-biotin-peroxidase method. Reproduced from [34], with permission.

metabolizing aldehyde reductase subtype [37] besides aldose reductase [38]. Here it was shown that different aldose reductase types exist: lens aldose reductase and the hepatoma-derived aldose reductase-like protein. However, only the hepatoma-derived aldose reductase-like protein, primarily expressed as an embryonal enzyme but not expressed in adult liver, was reexpressed in rat hepatomas. From these results it was concluded that the reexpression of rat hepatoma-derived aldose reductase-like protein is connected with detoxification processes in the liver and might reflect the putative resistance of initiated liver cells against the toxic action of various carcinogens. Now in human hepatocarcinomas a protein was identified as a homolog to rat hepatoma-derived aldose reductase-like protein [39]. This human aldose reductase-like protein appears to be selectively expressed (as well as aldose reductase) in human tissue [39].

3 Proteome analysis for the detection of dilated cardiomyopathy-associated proteins

Dilated cardiomyopathy (DCM) is a severe heart disease leading to heart insufficiency. Most heart transplantations are indicated by DCM. Pathomechanisms and molecular causes of this disease are unknown. It cannot be expected that there is only one pathomechanism leading to DCM. Therefore high sample numbers are necessary for investigation. International cooperation and standardization of sample preparation and the 2-DE technique will be prerequisites in the analysis of several hundred samples. Multiple cluster analysis [40–42] will help to distinguish several types of the disease. In comparison to other organs the heart is a relatively homogeneous organ, mainly consisting of myocardial muscle cells, and is therefore well-suited for a proteomic investigation by 2-DE. Two groups started in the beginning of the 1990s with a proteomic investigation of DCM [43, 44]. Myocardial 2-DE databases were constructed [45, 46] and introduced into the Internet [46–49]. Despite the fact that different sample preparations, different isoelectric focusing conditions, and different gel sizes were used, the interlaboratory comparison has shown that spots identified by protein chemical methods appeared at the same positions and identification by pattern comparison was successful in many cases [45].

Our high performance 2-DE procedure resulted in the resolution of 3300 myocardial protein species, from which 150 were identified by amino acid analysis, *N*-terminal and internal Edman degradation, and MALDI-MS. Comparison of 2-DE patterns from biopsies and explanted hearts between DCM and control patients have shown that the patterns are comparable. The assignment of most of the spots is possible. Nevertheless, there are many nonreproducible spot intensity variations, probably caused by different forms of the disease or by parameters such as disease stage, medication used, age, *etc.* However, a comparison of DCM atria with controls revealed 25 statistically significant differences [50]. We were able to identify 12 of these protein species (Fig. 5). Protein expression characteristics of DCM-associated proteins are also presented *via* Internet [51]. These results demonstrate the principal practicability of the approach. Biomedical significance will be obtained only in large-scale investigations including several hundred samples. Because these large numbers of samples were not available, we reduced our investigation to one protein family. Affinity-purified polyclonal antibodies against heat-shock protein 27 (Hsp27) recognized 59 protein species of the large gel myocardial 2-DE pattern [52]. Nine of these protein species were confirmed to belong to the Hsp27 protein by MALDI-MS peptide mass fingerprinting and post-source-decay sequencing. Differences of spot intensity within the Hsp27 protein family between DCM and controls have been described [53]. A detailed study is in progress (to be published).

The preliminary biomedically relevant results of these investigations are 12 identified proteins with their functions, whose amounts are either increased or decreased in DCM patients. These proteins can now be investigated by other methods, such as immunohistochemistry, enzyme activity testing, *etc.*, to confirm the disease-association, but more demanding is a large-scale subtractive

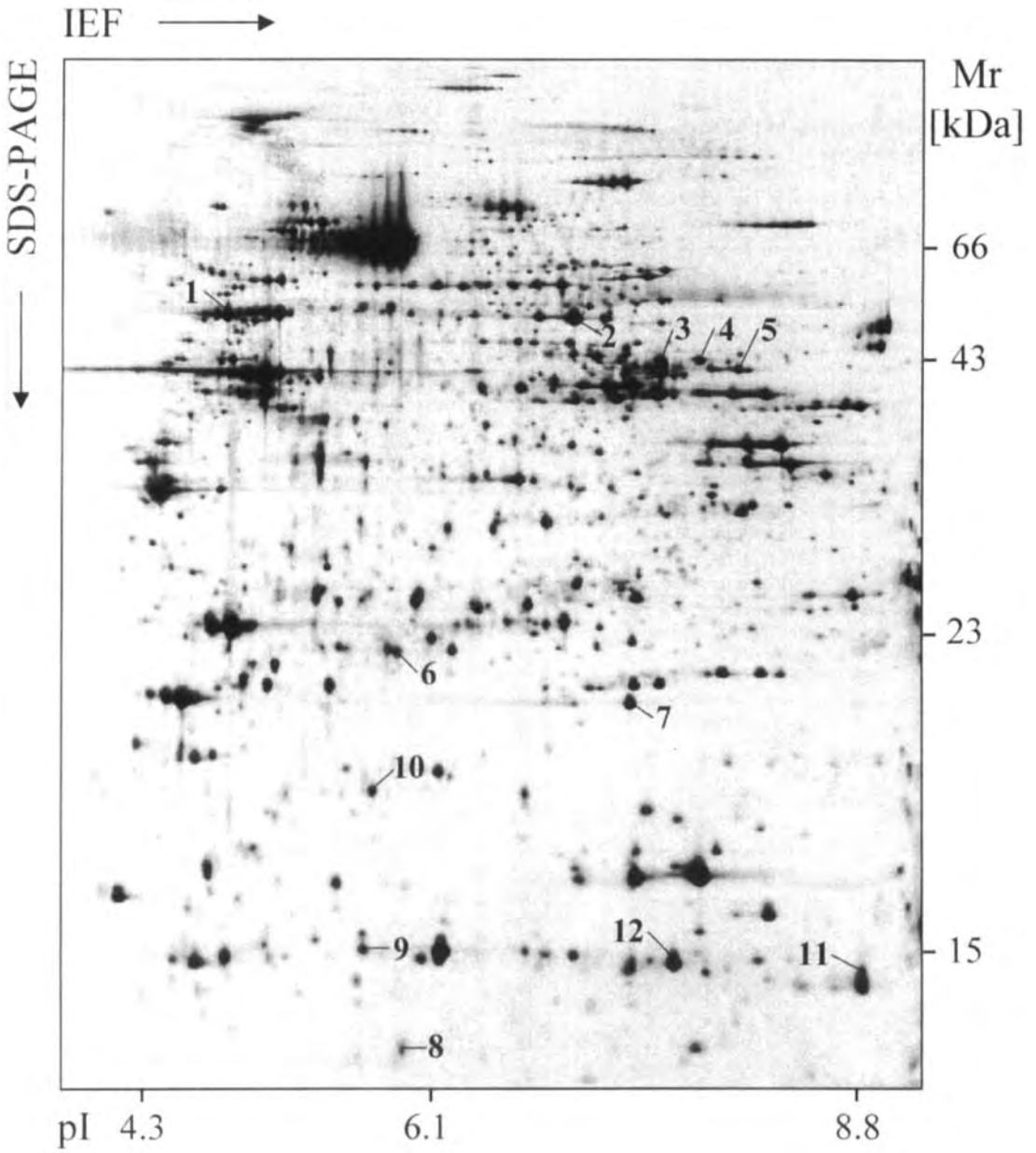


Figure 5. Protein composition of human myocardium from atrium. One hundred and fifty μg of myocardial protein were separated by NEPHGE combined with SDS-PAGE with a gel size of 23 cm × 30 cm × 1.5 mm. DCM patterns were compared with control patterns and spots marked with an arrow were found to show statistically significant differences in intensity between disease and control samples. Spots marked with an arrow and a number were identified: 1, ATP synthase, β-chain; 2, ATP synthase; 3, creatine kinase, sarcomeric mitochondrial; 4, 3-oxoacyl-CoA thiolase; 5, isocitrate dehydrogenase; 6, serum albumin fragment; 7, α-crystallin, B-chain; 8, cytochrome *c* oxidase polypeptide VIB; 9, fatty acid-binding protein; 10, haptoglobin fragment; 11, hemoglobin α-chain; 12, hemoglobin β-chain.

analysis of myocardium including several hundred well-defined samples of DCM and controls to realize the potential of the methods improved so far.

4 Proteome analysis of pathogenic microorganisms

The increasing importance of the analysis of infectious diseases for the health of the world's population has been stressed by the World Health Organization (WHO) for several years [54]. New infectious agents such as *Borrelia burgdorferi*, HIV, or the Ebola virus appeared in addition to old diseases long thought to be controlled, such as tuberculosis, multidrug restistant *Streptococcus* infections or the opportunistic pathogens, which achieved new importance with the current AIDS epidemic such as *Toxoplasma*. Therefore, the analysis of the proteomes of virulent microorganisms is a demanding task for the elucidation of virulence factors, antigens and vaccines, all important for diagnosis, therapy and protection. The genomes of eighteen microorganisms have been completely sequenced [55] and more than 60 further microorganism genomes are under investigation. The complete gene sequence information and the restricted number of genes in contrast to eukaryotic tissues are ideal prerequisites for proteomic investigations.

4.1 Detection of immune-relevant proteins in *Borrelia garinii*

The spirochete *Borrelia burgdorferi* is the causative agent of a multisystemic disease termed Lyme disease, characterized by an initial erythematous annular rash and flu-like symptoms, which proceed towards neurological complications and arthritis in about 50% of untreated patients. Three species have been delineated for *Borrelia burgdorferi sensu lato: B. burgdorferi sensu stricto, B. garinii, and B. afzelii. Borrelia burgdorferi* is transmitted by ticks of the genus *Ixodes*. Although the diagnosis is primarily based on clinical findings, it may be assisted by the results of serological tests like ELISA or immunoblot. These assays are not standardized, resulting in tests with various levels of sensitivity and specificity. In the United States, the Centers for Disease Control and Prevention (CDC) recommend a two-step approach to serological diagnosis. For screening, an ELISA should be performed. Positive results should then be confirmed by immunoblotting [56]. The immunoblot interpretation advanced the following cri-

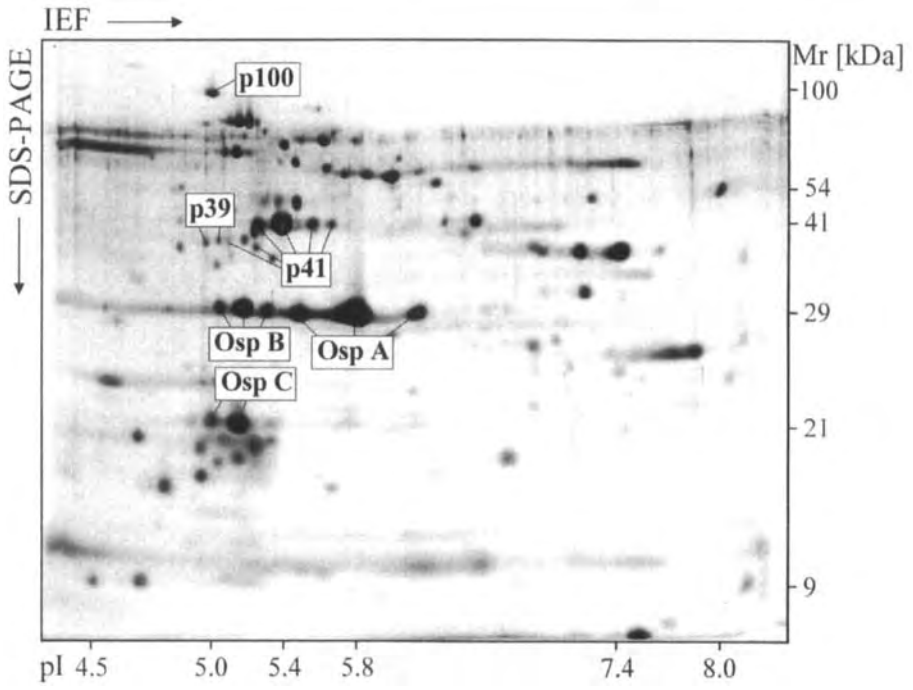


Figure 6. Protein pattern of *Borrelia garinii*. *B. garinii* proteins were separated by 2-DE, combining NEPHGE-IEF with SDS-PAGE, gel size 7 × 8 cm, and silver stained [57]. Six proteins were identified by immunostaining. Only one of them, p83/100, occurred as a single spot on the pattern.

teria for positive immunoblots: for IgM immunoblots at least two of the following three bands, OspC, 39, 41 [57] and for IgG immunoblots at least five of the ten following bands, 18, OspC, 28, 30, 39, 41, 45, 58, 66, 83/100 [58]. Standardization of immunoblots has not been established so far in Europe. Hauser *et al.* [59] tried to define interpretation criteria for all strains of *Borrelia burgdorferi*. We found that the immunoblot criteria recommended by the CDC were both very sensitive and specific.

B. burgdorferi contains 853 genes in its chromosome and additionally 430 genes on 11 plasmids [60]. Our preliminary 2-DE gels of *B. garinii* and *B. burgdorferi sensu stricto* (strain B31) resolved about 300 spots. A comparison of silver-stained small gels (7 × 8 cm) [61] with large gels (23 × 30 cm) [1] resulted in nearly the same spot numbers. Therefore, we used the small-gel technique for the detection of immune-relevant antibodies. We numbered 217 spots of silver-stained small gels of *B. garinii* (Fig. 6) and identified six known antigens by immunoblotting using polyclonal antibodies derived from rabbits: outer surface protein (Osp) A (BBA15), OspB (BBA16), OspC (BBB19), p83/100, p39, and flagellin p41 (BB147). With one exception, p83/100, all of these antigens occurred with more than one spot on the 2-DE pattern of *B. garinii*. Sera of patients with different manifestations of Lyme borreliosis were tested with *B. garinii* 2-DE blots. Sera of ten patients with erythema migrans stained 60 and 88 antigens on the *B. garinii* 2-DE pattern using anti-IgM and anti-IgG secondary antibodies, respectively. Sera of arthritis patients contained IgM antibodies against 15 antigens and IgG antibodies against 76 different antigens. Sera of late neuroborreliosis patients had IgM-antibodies against 33 antigens and IgG-antibodies against 76 antigens. The sera of patients from the three disease stages contained antibodies against the known antigens OspA, OspB, OspC, flagellin, p83/100 and p39, confirming results from SDS-PAGE blots. Additionally, many previously not described antigens were resolved, all of which are potential diagnostic markers. A detailed description of these results is in preparation.

4.2 Detection of antigens in *Toxoplasma gondii*

The parasitic infection toxoplasmosis is caused by the protozoa *Toxoplasma gondii*. Approximately 30% of the global population carries this parasite and in Europe toxoplasmosis is one of the most frequent infectious diseases. The obligate intracellular organism *T. gondii* exists in two forms in humans, the actively proliferating trophozoites (or tachyzoites) and the slower growing bradyzoites. Phylogenetic and statistical analysis indicated a highly unusual population structure consisting of three widespread clonal lineages [62]. Furthermore, the studies of Saavedra *et al.* [63] and Suggs *et al.* [64] showed that antigenic differences exist among the three *Toxoplasma* groups. Regardless of the host or geographic origin, the virulent strains of *T. gondii* comprise one of these three lineages [65]. In healthy persons, infections with *T. gondii* are generally mild or asymptomatic, but if acquired for the first time during pregnancy, the parasites may cross the placenta and often cause fatal infection of the fetus. Since the probability of transmission increases with time of pregnancy and is highest at the end of pregnancy, it is essential to assess the time of acquisition. On the other hand, the consequences of infection are more severe during organogenesis in the early phases of pregnancy. During later stages of pregnancy, fetal infection often results in persistent infections with a high risk of late complications such as retinochorioiditis. Approximately half of the acute maternal infections result in congenital diseases in the newborn. If the mother is adequately treated during pregnancy, fetal infections are less frequent and the sequels of infection are less severe.

Up to 90% of primary infections in pregnant women are not recognized [66]. At present, detection is exclusively based on serological screening and PCR. The use of the present serological methods for the detection of IgG, IgM and IgA antibodies alone is insufficient to assess the risk of the active disease, especially during pregnancy and in immunocompromised patients. Reactivation of a latent infection occurs frequently in immunocompromised patients. In AIDS patients, *T. gondii* is the main reason for intracerebral lesions which are often lethal. Therefore early detection of the disease is crucial for an effective therapy. *T. gondii* RH strain (kindly provided by Dr. Janitschke, Robert Koch Institute, Berlin, Germany) was maintained by serial passages in the human amnion cell line FL521 (kindly provided by Prof. Krüger, Institute of Virology, Charité, Berlin, Germany). 2-DE [61] resolved about 300 spots in silver-stained small gels (7 × 8 cm; Fig. 7). By immunoblotting we compared sera from (i) pregnant women with acute toxoplasmosis (n = 11), (ii) nonpregnant patients with acute toxoplasmosis (n = 6), and (iii) patients with latent toxoplasmosis (n = 9). Secondary antibodies to IgG, IgM, IgA, and IgE were used to differentiate the immunoglobulin classes. Nine spots reacted with all classes of immunoglobulins independently of the state of the infection. These spots are markers for infection with *T. gondii*. Seven markers may be valuable to distinguish between different stages of disease (see Fig. 7). Spot 1 reacted only with IgA and IgE, spots 2–4 with IgA and IgM, spot 5 with IgA, IgM, and IgE, and spot 6 with IgE in sera from patients in the acute state. Spot 7 reacted with all antibody classes in sera from patients in latent and acute disease states. Whereas the reaction of spot 7 with all antibody classes was found in

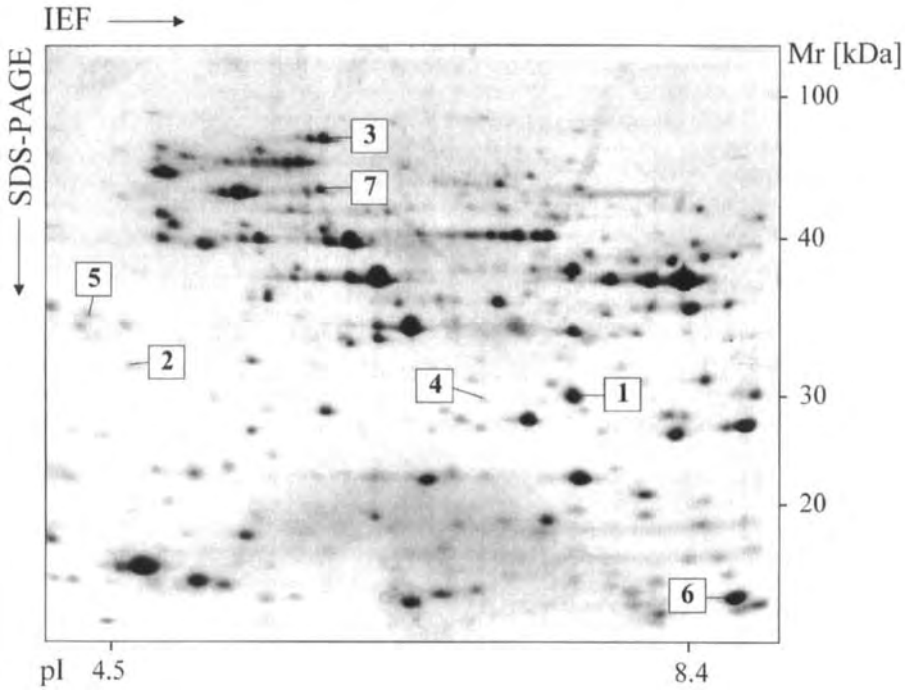


Figure 7. Protein pattern of *Toxoplasma gondii* RH. *T. gondii* proteins were separated by 2-DE, combining NEPHGE-IEF with SDS-PAGE, gel size 7 × 8 cm, and silver stained [61]. M_r and p*I* were calibrated with 2-DE marker proteins (Bio-Rad, Munich, Germany). Spots marked with numbers are potential diagnostic markers with the following p*I* and M_r characteristics: 1, 32.5/7.4; 2, 33.1/4.5; 3, 73.0/5.7; 4, 32.4/6.4; 5, 34.0/4.3; 6, 17.3/8.6; and 7, 57.1/5.7.

sera from all patients with latent toxoplasmosis, not all antibody reactivities specific for the acute state were present in all acute-phase patients. Therefore only a combination of several antigens will help to distinguish between acute and latent phase. The seven antigens found are candidates for diagnostic markers.

5 Concluding remarks

Two-dimensional electrophoresis acts like a molecular microscope. Complex protein mixtures are separated into distinct protein species. Subtractive analyses comparing disease with controls elucidate disease-associated proteins. At the stage of subtractive analysis the approach has the potential to unravel complex networks of protein interactions. This ultimate goal will be reached after automation of 2-DE and identification of proteins from 2-DE spots. At the moment an early reduction to obvious differences between disease and control 2-DE patterns allows the detection of single disease-associated proteins. Calgranulin B and hepatoma-derived aldose reductase-like protein have been confirmed by biochemical and immunological methods to be cancer-associated proteins. The investigation of DCM showed that, in principle, a proteomic approach is successful, but international cooperation in necessary for the collection of the necessary sample amount and a large-scale investigation promises better understanding of the molecular mechanisms of DCM. The proteomic investigations on microorganisms show the value of the proceeding for the detection of antigens and even the possibility to distinguish between different disease stages. These are prerequisites for the development of effective diagnostic tests.

Dr. Stulik is supported by a grant from the Ministry of Health, 4075-3, Czech Republic, K. Koupilova by a grant from the Agency of the Czech Republic, project No. 310/98/P238, and Dr. Zeindl-Eberhart by a grant to H. M. Rabes from the Dr. Mildred Scheel Stiftung für Krebsforschung, Bonn, Germany.

Received November 25, 1998

6 References

[1] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[2] Jungblut, P., Thiede, B., Zimny-Arndt, U., Müller, E. C., Scheler, C., Wittmann-Liebold, B., Otto, A., *Electrophoresis* 1996, *17*, 839–847.

[3] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., *Science* 1989, *246*, 64–71.

[4] Karas, M., Hillenkamp, F., *Anal. Chem.* 1988, *60*, 2299–2301.

[5] Edman, P., *Acta. Chem. Scand.* 1950, *4*, 283–293.

[6] Spengler, B., Kirsch, D., Kaufmann, R., *Rapid Commun. Mass Spectrom.* 1991, *5*, 198–202.

[7] Biemann, K., Scoble, H. A., *Science* 1987, *237*, 992–998.

[8] Wilm, M., Mann, M., *Anal. Chem.* 1996, *68*, 1–8.

[9] Klose, J., *Electrophoresis* 1989, *10*, 140–152.

[10] Jungblut, P., Wittmann-Liebold, B., *J. Biotechnol.* 1995, *41*, 111–120.

[11] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[12] Klose, J., *Humangenetik* 1975, *26*, 211–234.

[13] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[14] McGillivray, A. J., Rickwood, D., *Eur. J. Biochem.* 1974, *41*, 181–190.

[15] Scheele, G., *J. Biol. Chem.* 1975, *250*, 5375–5385.

[16] Young, D. S., Anderson, N. G. (Eds.), *Clin. Chem.* 1982, *28*, 737–1092.

[17] Anderson, L., Anderson, N. G. (Eds.), *Clin. Chem.* 1984, *30*, 1897–2108.

[18] Celis, J. E. (Ed.), *Electrophoresis* 1994, *15*, 307–556.

[19] Appel, R. D., Dunn, M. J., Hochstrasser, D. F. (Eds.), *Electrophoresis* 1997, *18*, 2701–2848.

[20] Hochstrasser, D. F., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 187–220.

[21] Hochstrasser, D. F., Tissot, J.-D., *Adv. Electrophor.* 1993, *6*, 267–375.

[22] http://biosun.biobase.dk/~pdi/jecelis/human_data_select.html

[23] http://www.harefield.nthames.nhs.uk/nhli/protein/

[24] http://userpage.chemie.fu-berlin.de/~pleiss/dhzb.html

[25] http://www.mdc-berlin.de/~emu/heart/heart.html

[26] Arnott, D., O'Connell, K. L., King, K. L., Stults, J. T., *Anal. Biochem.* 1998, *258*, 1–18.

[27] Lee, K. H., Harrington, M. G., *Electrophoresis* 1997, *18*, 502–506.

[28] Williams, A. C., Browne, S. J., Manning, A. M., Hague, A., van der Stappen, W. J., *Cancer Biol.* 1993, *4*, 153–159.

[29] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 131–151.

[30] Ji, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D., Simpson, R. J., *Electrophoresis* 1994, *15*, 391–405.

[31] Stulik, J., Kovarova, H., Macela, A., Bures, J., Jandik, P., Langr, F., Otto, A., Thiede, B., Jungblut, P., *Clin. Chim. Acta* 1997, *265*, 41–55.

[32] Roseth, A. G., Christinsson, J., Aadland, E., Schonsby, H., Nygaard, K., Fagerhol, M. K., *Gastroenterology* 1993, *104*, A445.

[33] Zeindl-Eberhart, E., Jungblut, P. R., Otto, A., Rabes, H. M., *J. Biol. Chem.* 1994, *269*, 14589–14594.

[34] Zeindl-Eberhart, E., Jungblut, P. R., Otto, A., Kerler, R., Rabes, H. M., *Eur. J. Biochem.* 1997, *247*, 792–800.

[35] Maser, E., *Biochem. Pharmakol.* 1995, *49*, 421–440.

[36] Tsuchida, S., Sato, K., *Crit. Rev. Biochem. Mol. Biol.* 1992, *27*, 337–384.

[37] Ellis, E. M., Judah, D. J., Neal, G. E., Hayes, J. D., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 10350–10354.

[38] Takahashi, M., Hoshi, A., Fujii, J., Mijoshi, E., Kasahara, T., Suzuki, K., Aozasa, K., Tanaguchi, N., *Jpn. J. Cancer Res.* 1996, *87*, 337–341.

[39] Cao, D., Sheung, T. F., Chung, S. S., *J. Biol. Chem.* 1998, *2773*, 11429–11435.

[40] Garrels, J. I., Franza, B. R., Chang, C., Latter, G., *Electrophoresis* 1990, *11*, 1114–1130.

[41] Appel, R. D., Hochstrasser, D. F., Funk, M., Vargas, J. R., Pellegrini, C., Muller, A. F., Scherrer, J. R., *Electrophoresis* 1991, *12*, 722–735.

[42] Vohradsky, J., *Electrophoresis* 1997, *18*, 2749–2754.

[43] Baker, C. S., Corbett, J. M., May, A. J., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1992, *13*, 723–726.

[44] Jungblut, P., Otto, A., Regitz, V., Fleck, E., Wittmann-Liebold, B., *Electrophoresis* 1992, *13*, 739–741.

[45] Jungblut, P., Otto, A., Zeindl-Eberhart, E., Pleißner, K.-P., Knecht, M., Regitz-Zagrosek, V., Fleck, E., Wittmann-Liebold, B., *Electrophoresis* 1994, *15*, 685–707.

[46] Corbett, J. M., Wheeler, C. H., Baker, C. S., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1994, *15*, 1459–1465.

[47] Müller, E.-C., Thiede, B., Zimny-Arndt, U., Scheler, C., Prehm, J., Müller-Werdan, U., Wittmann-Liebold, B., Otto, A., Jungblut, P., *Electrophoresis* 1996, *17*, 1700–1712.

[48] Evans, G., Wheeler, C. H., Corbett, J. M., Dunn, M. J., *Electrophoresis* 1997, *18*, 471–479.

[49] Pleißner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1996, *17*, 1386–1392.

[50] Knecht, M., Regitz-Zagrosek, V., Pleißner, K.-P., Jungblut, P., Hildebrandt, A., Fleck, E., *Eur. Heart J.* 1994, *15*, 37–44.

[51] Pleißner, K.-P., Söding, P., Sander, S., Oswald, H., Neuß, M., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1997, *18*, 802–808.

[52] Scheler, C., Müller, E. C., Stahl, J., Müller-Werdan, U., Salnikow, J., Jungblut, P., *Electrophoresis* 1997, *18*, 2823–2831.

[53] Otto, A., Benndorf, R., Wittmann-Liebold, B., Jungblut, P., *J. Prot. Chem.* 1994, *13*, 478–480.

[54] The World Health Report 1998, World Health Organization, Geneva 1998, ISBN 9241561890.

[55] http://www.tigr.org/tdb/mdb/mdb.html

[56] Centers for Disease Control and Prevention, *Morbid. Mortal. Weekly Rep.* 1995, *44*, 590–591.

[57] Engstrom, S. M., Shoop, E., Johnson, R. C., *J. Clin. Microbiol.* 1995, *33*, 419–427.

[58] Dressler, F., Whalen, J., Reinhardt, B. N., Steere, A. C., *J. Infect. Dis.* 1993, *167*, 392–400.

[59] Hauser, U., Lehnert, G., Lobentanzer, R., Wilske, B., *J. Clin. Microbiol.* 1997, *35*, 1433–1444.

[60] Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., Venter, J. C., *Nature* 1997, *390*, 580–586.

[61] Jungblut, P., Seifert, R., *J. Biochem. Biophys. Methods* 1990, *21*, 47–58.

[62] Howe, D. K., Sibley, L. D., *Infect. Dis.* 1995, *172*, 1561–1566.

[63] Saavedra, R., de Meuter, F., Herion, P., *Hybridoma* 1990, *9*, 453–461.

[64] Suggs, M., Walls, K. W., Kagan, I. G., *J. Immunol.* 1968, *101*, 166–175.

[65] Sibley, L. D., Boothroyd, J. C., *Nature* 1992, *359*, 82–85.

[66] Zygmunt, D. J., *Infect. Control. Hosp. Epidemiol.* 1990, *11*, 207–211.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 643–652

Review

Joachim Klose

Humboldt-Universität,
Charité, Campus Virchow-Klinikum, Institut für Humangenetik,
Berlin, Germany

Genotypes and phenotypes

Within the framework of a pilot project on the analysis of the mouse proteome, we investigated C57BL/6 mice (*Mus musculus*), a standard inbred strain of the mouse, starting with the analysis of brain, liver and heart proteins. Tissue extraction and the separation of proteins were performed with techniques offering a maximum of resolution. Proteins separated were analyzed by mass spectrometry. Gene-protein identification was performed by genetic analyses using the European Collaborative Interspecific Backcross (EUCIB), established from the two mouse species *Mus musculus and Mus spretus*. On the basis of protein polymorphisms we mapped hundreds of genes on the mouse chromosomes, allowing us new insight into the relationship between genotype and phenotype of proteins. In particular, the results showed that protein modifications can be genetically determined, therefore representing their own class of protein phenotypes. In this context, results are discussed suggesting that phenotypes of single protein species may result from several genes. Accordingly, proteins are considered as polygenic traits. In contrast, one example demonstrates that proteins may also have pleiotropic effects: a single gene mutation (a single altered protein) may affect several other proteins. From these studies we conclude that gene-related functional proteomics will show in the future that genetic diseases, defined today by clinical symptoms and considered as etiological entireties, can be subdivided into different diseases according to different affected genes.

Keywords: Mouse / Two-dimensional polyacrylamide gel electrophoresis / Functional proteomics / Gene function / Polygenic diseases EL 3413

Contents

Correspondence: Prof. Dr. Dr. Joachim Klose, Virchow-Klinikum, Institut für Humangenetik, Augustenburger Platz 1, D-13353 Berlin, Germany
E-mail: joachim.klose@charite.de
Fax: +49-30-45066904

1 Genotypes and phenotypes

It is well known what a gene is, but what is a genotype? We always obtain one half of our genes from the mother, the other half from the father. Consequently, we have each gene twice – the two alleles of every gene, and this is brought about by the two homologous chromosomes. Frequently, however, the two alleles of a gene differ slightly from one another, due to mutations. These mutations are usually spontaneous point mutations which lead to amino acid substitutions. In principle, a point mutation may occur in the maternal allele, in the paternal allele, or in both alleles of a gene. These three possibilities are described by the term "genotype": a gene *a* and its mutant allele *a'* may create the genotypes *aa* (homozygous, wild type), *aa'* (heterozygous), and *a'a'* (homozygous, mutant type). Considering two different natural populations, the chance that the two alleles of a gene differ is higher the greater the genetic distance is between the two populations. For comprehensive genetic investigations in an

organism it is an advantage if many genes show an allelic variation in this organism. Working with mice, instead of with human populations, one can take the mother from one strain and the father from another strain, or even from another species, to reach a maximum in the genetic distance between the parents. Moreover, taking inbred strains, all the genes show the homozygous genotype, which facilitates genetic studies considerably. In our investigations, for example, we use the two mouse species *Mus musculus* (strain C57BL/6; B6) and *Mus spretus* (SPR). Because these two strains belong to different species, the genetic distance is relatively large. On the other hand, the genetic distance in this case is within a range where cross breeding (at least in the direction B6 ♀ × SPR ♂) is still possible.

1.1 Definition of phenotype

A gene creates a phene, *i.e.*, a "visible" character, and because a mutation in a gene may alter its phene, the three genotypes of a gene usually create three different phenotypes. According to King and Stansfield [1], the term phenotype is defined as the observable properties of an organism produced by the genotype (in conjunction with the environment, see below). In the classical sense of genetics, observable properties are external traits of an organism, such as the hair color of the mouse, morphological characteristics of an animal, or clinical symptoms in humans. Nowadays, however, many different instruments and techniques are available (*e.g.*, microscopes, physiological tests, electrophoresis, molecular analytical techniques) that allow us to observe properties of an organism on many different levels of gene expression. Therefore, one may distinguish between morphological, physiological, biochemical, and molecular phenotypes, the latter including phenotypes of proteins and mRNAs (Fig. 1).

1.2 Relationship between genotype and phenotype

Figure 1 may imply that the relationship between genotype and phenotype is a linear one. This, however, is not the case. On the contrary, any phenotype may be the result of the genotype of more than one gene, most likely of many genes, which may, however, contribute to a phenotype to a different extent (major and minor genes). Moreover, environmental factors may modify a phenotype. This leads us into a dilemma if we try to determine the precise and specific functions of a particular gene. The function of a gene is reflected by its phene. But where is the place on the long road from the genes to the external traits of an organism that most directly and specifically reveals the function of a gene? This question is addressed in Fig. 2. The DNA sequence of a gene tells us nothing about its function. The mRNA is somewhat more informative in this respect. If a distinct mRNA species occurs, for example, in the brain, but in no other tissue, we may conclude that the function of this mRNA species has something to do with brain functions. The cellular concentration of the different mRNA species reflects the degree of activity of the corresponding genes, but does not necessarily correlate with the concentration of the proteins translated from these mRNAs. Therefore, quantitatively, mRNAs are not very informative with regard to gene function. The next level in gene expression, the protein level, reflects gene function to a much higher degree.

The protein of a gene offers all the molecular structures and properties needed to fulfill the functions of a gene.

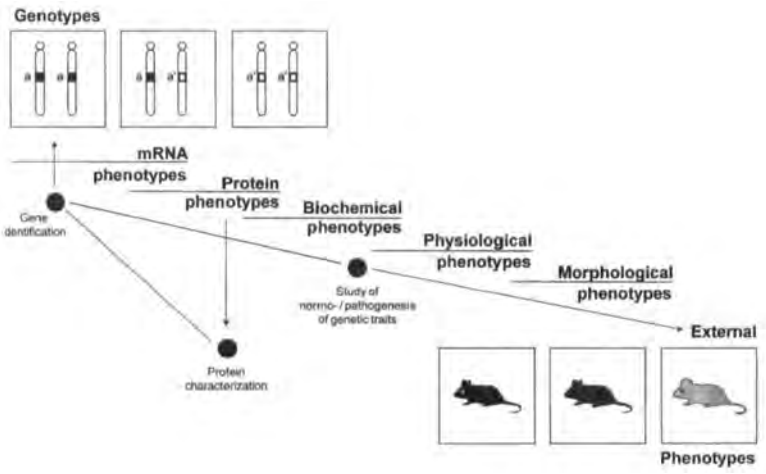


Figure 1. Genotype-phenotype relationships and a strategy to analyze normal genetic traits and genetic diseases.

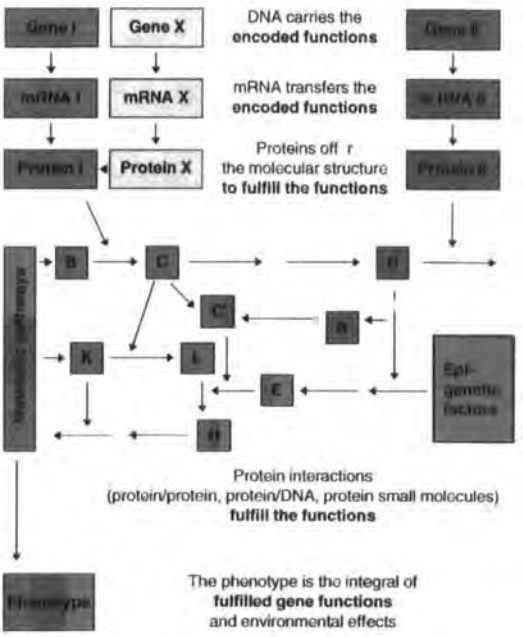


Figure 2. The progression from genotypes to phenotypes is shown in some detail to illustrate the problem of determining the specific function of a gene. On the way the proteins are in a particular position. On the one hand they are still directly related to the individual genes, and, on the other hand, they offer all the molecular properties necessary to interact with other molecules to fulfill the functions of the individual genes. At higher levels of gene expression other genes and epigenetic factors become involved in creating distinct phenotypes so that the specific function of genes, *i.e.*, their specific contribution to a distinct phenotype, becomes more and more obscure. The special case is shown in which even the function of a single protein depends on two genes: the molecular reaction from B to C needs the presence of protein I, but protein I can fulfill its function only in connection with protein X (see Section 1.2).

For example, protein X of gene X (Fig. 2) may occur specifically in the cell nuclei and show a sequence motive for DNA binding. We would assume that the function of this gene concerns the regulation of the transcription of a particular structural gene I. This would be the most direct and specific information about function obtainable from gene X. This information, however, is soon obscured if other proteins (transcription factors XI, XII) are necessary to activate target gene, gene I by interacting with protein X. "Activation of gene I" is then no longer the function of gene X, but the combined function of gene X+XI+XII. Here, the path from genotypes to phenotypes enters the network of gene regulation, and, in a broader sense, the network of metabolic pathways. The metabolic pathways further obscure the specific function of a gene. Many genes (II, III, ...) contribute to the cascades of metabolic reactions which lead to phenotypes of higher levels, and finally to the external genetic traits of an organism. In this complex process of gene expression, the proteins offer the most suitable target for gaining information about specific functions of individual genes. Elucidating gene functions therefore means determining the chemical, biochemical and biological characteristics of proteins. These characteristics include the molecular structure of the individual proteins, the co- and post-translational modifications, the binding properties of the various protein species, the quantitative properties (such as synthesis rate, cellular concentration and degradation rate), and all the biological characteristics of proteins: tissue specificity, cell structure and organelle specificity, sex specificity, specificity to the various stages of embryonic and postnatal development, and specificity to the stages of aging.

1.3 Two-dimensional electrophoresis

Two-dimensional electrophoresis (2-DE) is a unique method for large-scale protein characterization. By comparing 2-DE protein patterns from different tissues, cell fractions, and developmental stages, proteins can be characterized according to different biological parameters. Western blotting followed by immunologically based procedures for glyco- or phospho-staining allows the detection of post-translationally modified proteins. The structure of proteins can be investigated by extracting protein spots from 2-DE gels and employing analytical techniques such as mass spectrometry and partial sequencing. Using such a global strategy, individual proteins, whether known or unknown, become characterized according to many different parameters. Taking all the features attributed to a distinct protein spot, conclusions about the function of that protein – and, consequently, of its gene – can be drawn. One may learn, for example, that protein spot No. *xy* is brain-specific, occurs in the membrane fraction of neural cells late in life, shows increasing phosphorylation in the course of aging, and reaches higher levels in cellular concentration in males than in females. One may conclude that this protein plays a role in the process of aging.

After proteins have been characterized in several respects, the genes of these proteins must be identified, if discovering the functions for individual genes is the aim. There are, in principle, two ways to detect the gene of a particular protein: (i) genetic linkage studies and gene mapping on the basis of protein polymorphisms, and (ii) mapping genes on a physical map of chromosomes on the basis of the sequence homologies between proteins

and their corresponding genes. Protein polymorphisms indicate that the gene of this protein exists in different alleles. Protein polymorphisms represent different phenotypes of a gene existing in different genotypes. Two-dimensional protein patterns offer a unique opportunity to detect protein polymorphisms on a large scale, and to observe various protein phenotypes. Working with distantly related mice, many proteins can be genetically mapped. However, in terms of total genomes, one has to realize that the vast majority of proteins does not reveal polymorphisms in 2-DE patterns. Additional strategies are necessary for gene-protein identification, as mentioned above in item (ii) and explained elsewhere [2].

2 Functional genomics and functional proteomics

The term "genomics" covers the whole genome of a single organism, and "genome analysis" means sequencing of the total DNA and mapping of all genes of a genome (structural genomics [3]). At present, genome analysis is performed worldwide in human as well as model organisms. Genomes of several microorganisms [4] and the first genome of a multicell organism (*C. elegans*) [5] have already been completely sequenced. As a consequence of the rapidly proceeding genome projects, subject and aim of the post-genome (post-sequence) era are problems of current interest. It is, however, already commonly agreed that the topic of the coming era will be what is called "functional genomics". "Functional genomics is the attachment of information about function to knowledge of DNA sequence" [6]. But what should be attached to the sequences that offers this information? According to considerations mentioned in Chapter I, genome-wide analysis of the proteins of an organism and genome-wide gene/protein identification would be the most basic (*i.e.*, the most single-gene-related) approach towards discovering gene functions.

Genome-wide analysis of the proteins of an organism is an idea first introduced 20 years ago, shortly after 2-DE had been introduced. In particular Leigh and Norman Anderson presented the idea to separate and catalogue all the human proteins [7–9], a concept today called proteome analysis. The term proteome* was introduced to describe the entire protein complement of an organism [10]. According to the terminology used in genomics, one should distinguish between structural and functional proteome analysis. Structural proteome analysis would mean isolation and sequencing of all the proteins encoded in the genome of an organism (the "primary proteins"), and functional proteome analysis would mean determining all the chemical, biochemical, and biological characteristics of the different primary proteins. In other words, identification of functionally significant sequence motives in primary proteins would be a matter of functional proteome analysis. However, functional proteome analysis would not be restricted to the amino acid sequence of proteins, but would include the broad spectrum of structural modifications and quantitative changes to which the proteins are subjected in different tissues, cell organelles and developmental stages, *i.e.*, in the various spacial and temporal dimensions of an organism.

The structural and quantitative heterogeneity that the proteins create to fulfill their functions is the central subject of functional proteomics. At present, many laboratories perform studies using 2-DE protein patterns to detect proteins that may be involved in a biological or pathological process of particular interest. This, however, is not what has been called proteome analysis, just as genome analysis does not mean searching for a distinct gene. Also of interest in this field is the detection of known proteins in complex 2-DE patterns of tissue proteins. Protein spots from the gels are analyzed by mass spectrometry, and the data obtained are used to screen sequence databases to find matches with known proteins. Studies of this kind cover an important part of the work that has to be done in analyzing proteomes structurally and functionally. In this way, proteins known with respect to their amino acid sequences and some functional properties are sorted out from the bulk of unknown proteins. However, proteome analysis, in its real sense, aims at the analysis of all proteins of a cell type, tissue or organism, and this also include the yet unknown proteins which may, to date, be considered to be the vast majority of proteins of an organism.

In conclusion, proteome analysis should include the following features: (i) the use of techniques (protein extraction, 2-DE) which offer the chance to detect the vast majority, if not all of the proteins of a tissue; (ii) the inclusion of the unknown as well as the known proteins in structural proteome analysis; (iii) characterization of the separated proteins (both the known and the unknown proteins) on the basis of a broad spectrum of biochemical and biological parameters, *i.e.*, performing functional proteome analysis; finally, (iv) the genes corresponding to the separated and characterized proteins should be identified and mapped on the chromosomes. Proteome analysis done this way will result in functional genomics.

* The term genome, first used by H. Winkler in 1920, was created by elision of the words GENes and chromosOMEs [3]. Therefore, the word GENOME is artificial, but signifies: the complete set of chromosomes and their genes [3]. The word proteome, consequently, is artificial as well, and signifies, according to the term genome, the complete set of chromosomes and their encoded proteins.

3 Analysis of the mouse proteome

We started a systematic analysis of the mouse proteome. The whole procedure consists of four steps: (i) extraction of proteins from selected tissues and cell fractions, (ii) separation of proteins by 2-D electrophoresis, (iii) image analysis of protein patterns, and establishing protein standard patterns as the basis for a mouse protein database, and (iv) spot identification or, in case of unknown spots, spot characterization by mass spectrometry or partial sequencing. This procedure is followed by mapping genes of polymorphic proteins on the mouse chromosomes. At the same time, the proteins registered in our protein database are characterized on the basis of a broad spectrum of biochemical and biological parameters.

3.1 Fractionation of total tissue proteins

For the analysis of the mouse proteome we selected an inbred strain, the strain C57BL/6, which is one of the most commonly used mouse strains in research. In a first approach we analyze the proteins of the brain, liver and heart, which represent the three germ layers ectoderm, endoderm and mesoderm, respectively. These organs were collected from both males and females, from different developmental stages, and from postnatal and adult stages, the latter including the final stages of aging. In order to reveal as many proteins as possible from a particular tissue, we fractionate the total tissue proteins into three fractions: (i) the buffer-soluble proteins (supernatant I + II), which may represent the cytoplasmic proteins, (ii) the urea/CHAPS-soluble proteins (pellet extract), which may consist of proteins normally bound to the cell structures, and (iii) a DNase-digested rest pellet suspension that reveals chromosomal proteins such as histones. The fractionation procedure (described in detail elsewhere [11]) was based on a concept that avoids any loss of particular groups or classes of proteins. The 2-DE patterns of these three fractions may represent the vast majority of the total proteins of a tissue. In addition to these basic fractions, we prepared highly concentrated protein extracts from purified cell organelles, primarily from cell nuclei. Protein patterns from these extracts reveal many minor proteins, not detectable in the three basic patterns.

3.2 Large gel 2-DE

In order to reach maximum resolution of the proteins extracted, we developed a 2-DE technique for large gels [12], a modification of our original 2-DE technique [13]. Isoelectric focusing is performed in capillary tube gels, 40 cm in length (46 cm tubes). The separation distance in the second dimension, the SDS flat gel, is 30 cm. Carrier ampholytes were used for isoelectric focusing. Immobilines compared to carrier ampholytes were found to have a lower resolving power in large-distance gels [14]. Protein detection in 2-DE gels was performed by silver staining [12].

Liver and brain protein patterns are shown elsewhere [11, 14]. A protein pattern from the mouse heart muscle is

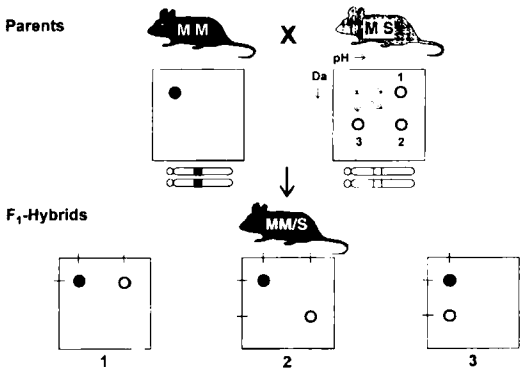


Figure 3. The three protein phenotypes are shown electrophoretic mobility variants may reveal two-dimensional protein patterns. The two parental mouse species *Mus musculus* (MM) and *Mus spretus* (MS) differ in the electrophoretic position of a protein spot. The difference can be caused by changes in the isoelectric point, the molecular weight, or in both parameters of a protein. Consequently, among the hybrids (MM/S) three different protein phenotypes may occur. The homozygous genotypes of the parental strains are shown by schematic chromosomes.

Table 1. Number of protein spots as revealed by large-gel 2-DE of proteins from three different organs of the mouse

Tissue fractions	Number of protein spots		
	Liver	Brain	Heart
(A) Supernatant[a] (buffer)	9 204	8 458	4 790
(B) Pellet extract[b] (urea, CHAPS)	1 975	1 692	1 470
(C) Pellet suspension[c] (DNA digestion)	73	50	40
Total No. of spots/Organ[d]	11 252	10 200	6 300
Total No. of spots/Mouse	27 752 protein spots		

a) Spots/pattern
b) Spots not present in A or C
c) Spots not present in A or B
d) Spots which may occur in two or three organs were not identified to bring the total number of spots completely down to the level of unique spots.

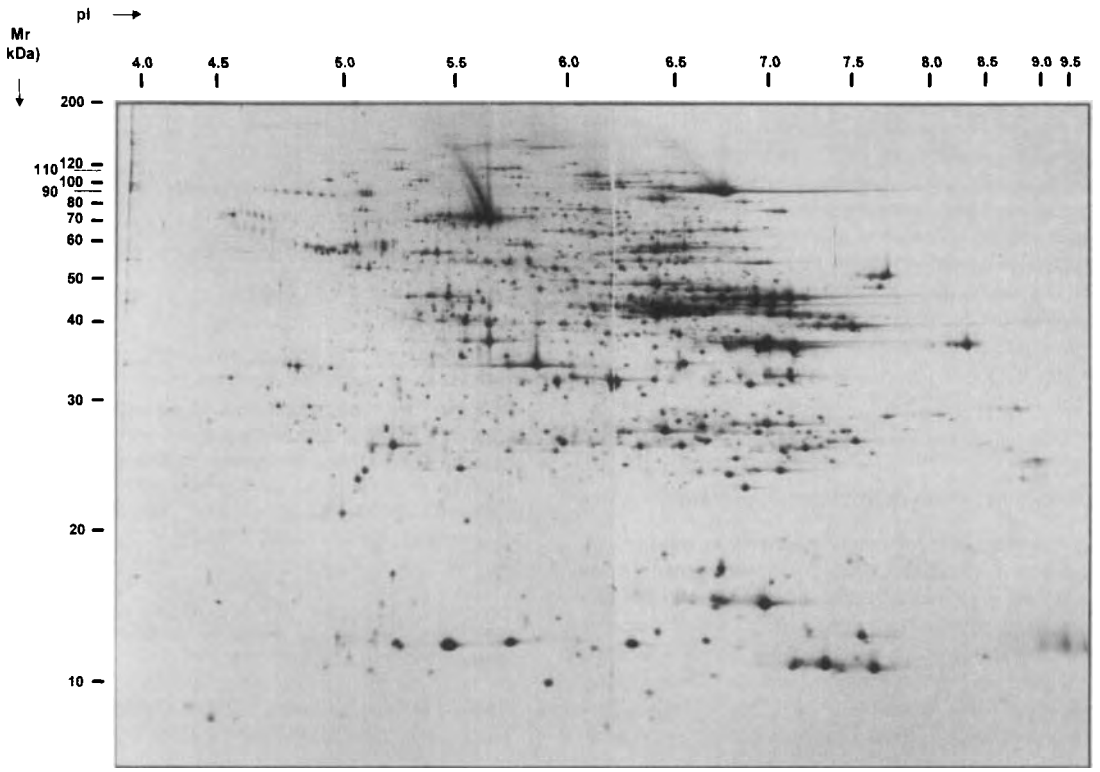


Figure 4. Two-dimensional protein pattern from the mouse heart supernatant fraction. The proteins were extracted from the heart in three fractions, the supernatant, the pellet extract and the remaining pellet suspension [11]. The proteins were separated by large-gel two-dimensional electrophoresis and revealed in the gel by silver staining [12].

shown in Fig. 4. A rough estimation of the total number of protein spots detectable in the 2-DE patterns of the three organs and the three fractions mentioned is shown in Table 1. In comparing 2-DE patterns from different tissues and cell fractions, it is actually impossible to avoid that in some cases a distinct protein registered in one pattern is registered again in another pattern that also reveals this protein. The same protein may occur in the patterns of different tissues at different places, if they are modified differently. On the other hand, a protein spot that occurs in different patterns in the same position is not necessarily the same protein. In Table 1, the figures given for the supernatant fractions indicate the total number of spots of corresponding 2-DE patterns. For the other two fractions the attempt was made to count only spots which were not already registered in one of the other two patterns. The three organs, however, were not compared, *i.e.*, redundant spots were not subtracted from the total number of spots found per pattern and per organ. To date, the highest number of spots revealed in one pattern was found in the supernatant fraction of mouse testis. Here, more than 10 000 protein spots were detected per pattern [14]. In spite of the high resolution reached with our technique, and even if protein patterns from cell organelles are taken into account, we cannot say that all the different proteins of a tissue are presented in our patterns. Certain protein species may exist only in a few copies per cell, or not even in all cells of a tissue. These proteins would not be detected in a 2-DE pattern. Three important questions cannot be answered at the present stage of our studies: How many proteins are expressed in a certain tissue (cell type)? How many of these are specific for this tissue? How many proteins arise by modifications of the primary proteins?

3.3 2-DE standard patterns

From each tissue and protein fraction we establish a 2-DE standard pattern. This is a synthetic pattern produced from a stained 2-DE gel by scanning, digitizing, and analyzing the image with a computer program for spot detection [15]. The pattern generated by the comput-

er is carefully compared, spot by spot, with the original gel pattern and with several other patterns produced from the same kind of tissue. The computer pattern is then interactively corrected on the screen by searching for spots which were not recognized perfectly by the program. The final pattern is divided into 40 sections, and the spots in each section are provided with numbers. The standard patterns constructed from the different mouse tissues constitute the basis for our mouse protein database. Any information we obtain for a distinct protein spot is stored in the database with reference to the corresponding spot number. The standard pattern of the mouse brain supernatant fraction together with a first set of data, concerning the protein spots identified so far, was recently presented by our homepage http://www.charite.de/humangenetik [16]. We analyzed 560 protein spots by using mass spectrometry in combination with a genetic approach [16]. Of these spots, 331 have been identified, and out of these, 90 indicated different proteins.

In the past several years, 2-DE protein patterns from many different cell types and tissues of various organisms, including man, have been published [17, 18]. Federated 2-DE databases were established on the Internet network, allowing laboratories worldwide to share 2-DE data [19]. In practice, however, matching 2-DE patterns from different laboratories was difficult or impossible, due to the different techniques used (carrier ampholytes, IPGs, gel format, staining procedure, sample preparation). In the future, this problem will be overcome by the increasing improvements in analyzing 2-DE patterns by mass spectrometry. This will allow the laboratories to compare 2-DE spots on the level of mass spectrometry data rather than by matching 2-DE patterns. Consequently, establishing a 2-DE technique which has to be used precisely in all laboratories to allow sharing of data will no longer be an indispensable aim.

3.4 Gene-protein identification

Following protein extraction of selected mouse tissues, 2-D electrophoresis, image analysis of 2-DE patterns, and chemical analysis of protein spots, detecting the genes of the separated proteins is the next step in our pilot study on the mouse proteome. Gene-protein identification was started by genetic linkage analysis of genes revealing protein polymorphisms between the two mouse species *Mus musculus* (B6) and *Mus spretus* (SPR). Among the ~ 8700 protein spots revealed in 2-DE patterns of brain supernatant proteins, more than 1000 genetically variant spots were found by comparing B6 and SPR. About one half of these variants showed electrophoretic mobility changes, and the other half showed changes in spot volume (protein amount). A European collaborative project has produced a comprehensive mouse backcross (the European Collaborative Interspecific Backcross, EUCIB) using B6 and SPR as the parental strains. About 1000 animals were generated in the backcross generation. We used 64 of these animals to study the segregation patterns of polymorphic proteins. By genetic linkage studies and gene mapping procedures, we mapped the genes of several hundred protein spots on the mouse chromosomes (publication in preparation).

4 Protein phenotypes

4.1 Polymorphic proteins

Two-dimensional electrophoresis is a unique tool to study the effect of gene mutations on properties – or, in terms of genetics, on phenes – of proteins. Applying large-gel 2-DE to a genetic mouse system that thereby reveals more than one thousand polymorphic proteins solely in one organ (brain), protein phenotypes can be investigated on a large scale and all under the same conditions. The protein phenes visible in 2-DE gels include the electrophoretic position, the spot volume (spot area × optical density) and the heterogeneity of proteins (spot series, spot families). The investigation of genetic changes in proteins leads to interesting questions; for example: Does a variant protein that occurs in several tissues show the variation in each of these tissues, and, if so, is this variation then always of the same type? Does the occurrence of a certain protein alteration in an individual depend on its age? Is a quantitative deviation in the early developmental profile of a protein stable throughout the whole embryonic development and even in postnatal life? Does an amino acid substitution in a protein, due to a point mutation, affect the post-translational modification, the conformation, the turnover rate, or some of the binding properties of this protein? All these questions point toward problems of fundamental significance for human genetic diseases. Investigations of these questions may explain why genetic diseases usually show tissue specificity, why diseases often set in at a certain age of the persons affected, or why the expression of a particular disease depends on certain environmental factors (food, drugs). Moreover, with respect to the heterogeneity frequently observed in genetic diseases, it is of interest to search for genes which act on a protein apart from the structural gene. Findings like this would explain why the same disease, *i.e.*, the same clinical symptoms, may result from mutations in different genes.

4.2 Mutations

Mutations alter the position of protein spots in 2-DE gels by affecting the charge, the molecular weight or (and) the conformation of proteins (positional variants = electrophoretic mobility variants, mV; Fig. 3). Mutations may also

have consequences on the synthesis rate or degradation rate of proteins. In 2-DE patterns, this is revealed by changes of protein spots in size and intensity (variation in spot volume = variants of protein amount, aV). The three phenotypes of a quantitatively variant protein are composed of the two homozygous parental spots, one with a high, the other one with a low spot volume, and the heterozygous spot of the F_1 generation with a spot volume in-between. In an extreme situation, a protein may completely disappear in a mouse strain or in a human individual (presence/absence variants, paV). The amount of a protein, *i.e.*, its cellular concentration, is regulated by transcription factors and factors involved in the process of translation and protein processing. Therefore, quantitative protein variants most frequently may reflect mutations not in the structural gene, but in genes or DNA sequences that are components of the regulatory system of proteins.

When we compared hundreds of polymorphic proteins from B6 and SPR mice in the hybrid patterns, we frequently observed that the two spots of the heterozygous positional variants differed not only in the horizontal position, but also, or only, in the vertical position (Fig. 3). This points to differences in the molecular weight between the two variants of a protein. The maximum effect of an amino acid substitution on the molecular weight of a protein would be given if tryptophan (204 Da) were replaced by glycine (75 Da); the difference would then be 129 Da. In many cases, however, we observed differences much higher than that, the maximum ranging at 2500 Da (unpublished results). This indicates that mutations may frequently affect the structure of a protein much more extensively than just by amino acid substitution. These alterations may include changes in co- or post-translational modifications, truncations, or altered conformations of protein molecules. While this observation is currently being investigated in more detail, other findings support this assumption.

We frequently observed spot families in 2-DE patterns, another indication for protein modifications. The spot families were detected by mass spectrometry in combination with genetic criteria: variant protein spots which showed in the hybrid pattern exactly the same distance (mm), the same relative position, the same positional orientation with regard to the parental positions, and, moreover, which mapped to the same locus on the mouse chromosomes, were considered as spots which originate from the same protein. Usually, we identified or characterized the most prominent spots of a spot family by mass spectrometry. In this way we confirmed to some extent the family character of these spots, and, at the same time, tentatively identified many minor spots of the pattern which may be difficult to analyze directly by mass spectrometry. As a result, we found in the brain protein patterns, for example, that gamma enolase, synapsin, and L-lactate dehydrogenase H chain (LDH-H) revealed 23 spots, 38 spots, and 23 spots, respectively [16]. The protein tau, a protein involved in Alzheimer's disease, showed more than 100 spots in 2-DE patterns from human brain proteins. By analyzing the complexity of these spots, we found that alternative splicing and phosphorylation was one of the protein modifying mechanisms [20]. Other proteins, *e.g.* LDH-H, formed spot family patterns interpretable as protein degradation patterns [16]. Some of these spot families were found to be extremely reproducible with regard to spot composition. The degradation products were apparently stable in the cells and seemed to be the result of an ordered cleavage process rather than of random degradation. Limited and ordered degradation is known to be a significant mechanism for certain cell functions [21, 22].

4.3 Size of spot families

Positional variants as shown in Fig. 3 and quantitative variants as mentioned above were the most frequently occurring protein phenotypes in 2-DE patterns obtained from B6 and SPR mice. Moreover, however, we observed that the size of a spot family, *i.e.*, the number of spots found to belong to a certain family, may also vary between B6 and SPR. Of the 14 degradation spots found for the LDH-H family in the SPR pattern, five did not occur in B6. Additionally, the degradation spots showed higher intensities in SPR than in B6. This can be interpreted as a higher degradation rate occurring in the LDH-H of SPR than in the LDH-H of B6. Synapsin, as another example, revealed two extended horizontal spot series in 2-DE patterns, due to a protein modification not clarified so far. The number of spots of the series differed between B6 and SPR. In both LDH-H and synapsin, the variation in the phenotype "size of spot families" segregated in the backcross progeny of B6 and SPR. Preliminary results show that these phenotypes mapped to the locus of the structural gene of these proteins. Apparently, a mutation in the structural gene in one case led to an altered degradation rate of the protein, and in the other case to an alteration in the degree of modification of the protein.

Genetic variation of the complexity of spot families was also found in "one-spot families". Proteins were observed that create one spot in one mouse species, but, by splitting, two spots are created in the other species. This suggests that a protein can be modified in one species but not in the other. Protein variants of this type were found to be even more interesting when compared in different tissues. A protein was detected that split into two spots in one organ (liver), but not in other organs (brain, heart;

Kaindl *et al.*, in preparation). In this case protein modification was not only genetically determined, but also tissue-specifically regulated. This may be an example of why genetic diseases in some cases (*e.g.*, Huntington's chorea) affect one organ, but not others.

5 The protein, a polygenic trait

5.1 Understanding the network of gene activity

As mentioned, proteins in 2-DE patterns show different phenes and phenotypes, and the different phenotypes may result from changes in molecular weight or charge of proteins, from variations in the amount of proteins, from the degree of degradation, and from the degree of post-translational modifications, or they may result from alterations in tissue specificity or developmental stage specificity of proteins. For an understanding of multifactorial diseases, and more basically, for an understanding of the network of gene activity, it is of fundamental significance whether the various phenes of a protein depend on different genes. A mutation in the structural gene of a protein certainly may affect several phenes of this protein at the same time; for example, charge, molecular weight and prosthetic groups attached to the amino acids substituted by the mutation. However, quantitative changes of proteins most likely result from mutations in regulatory sequences. Furthermore, the degree to which a protein is modified by phosphorylation or glycosylation, for example, may depend on the concentration and structure of certain enzymes, and, therefore, on other genes than the structural gene of the protein. If different phenes of a protein were affected by different genes, this could be detected by genetic linkage studies. In this case the different phenes of the protein would segregate differently in the progeny and map to different loci on the chromosomes. A protein that is present in a high amount in a 2-DE pattern, but in a low amount in another mouse species, frequently shows an intermediary level in the hybrid pattern. However, in most cases, quantitative protein variants show various levels of concentrations in the progeny. This indicates that the cellular concentration of such a protein depends on several genes. Segregation studies of these loci (quantitative trait loci, QTL), however, require precise measuring spot volumes from a large number of animals.

A genetic analysis of QTL of proteins, quantitatively variant in 2-DE patterns, has been performed by Damerval *et al.* [23] in maize. This interesting and important investigation showed that the cellular concentration of a single protein species can depend on several chromosomal loci. Up to five, or even up to 12 loci were mapped for single proteins. At least some of these loci were located on different chromosomes. This finding demonstrates that proteins represent polygenic traits. In our studies on mice, we found that protein modifications, as revealed by protein spots which split into two spots in another mouse species, can be caused by genes which do not map to the locus of the structural gene of the protein. If confirmed, observations like this would show that not only the amount of a protein but also its structure can depend on several genes.

5.2 Pleiotropy

Another genetic phenomenon worth considering under the aspect of proteins is pleiotropy. Contrary to polygeny, the situation in which several genes act on the same genetic trait, pleiotopy is the phenomenon where a single gene is responsible for a number of distinct and seeming-

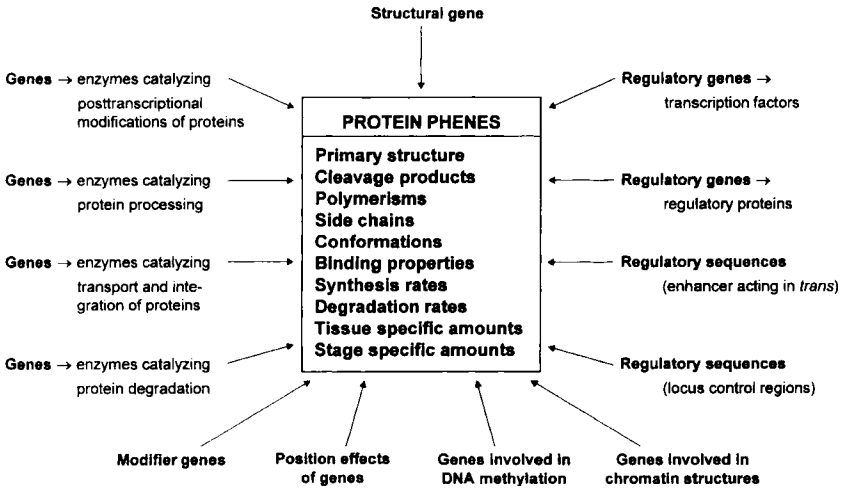


Figure 5. The polygenic nature of proteins.

ly unrelated phenotypic effects [24]. We recently described a pleiotropic effect observed in the crystallins of the mouse eye lens [25]. A mouse strain carrying a cataract mutation in the gene for γB-crystallin was investigated by 2-DE. First, the lens proteins of normal mice were separated and analyzed by mass spectrometry and partial sequencing. All the various crystallins of the crystallin family, encoded by different genes, were identified. Then, analyzing the proteins of the mutant strain, the unexpected observation was made that not only the amount of the γB-crystallin was drastically reduced, but also all the other γ-crystallins (subfamily γA–γE). In principle, one may assume that the gene of γB-crystallin had a pleiotropic effect (*e.g.*, *via* gene regulation, frameshift) on the other γ-crystallin genes, which form a cluster on chromosome No. 1, or that the protein γB-crystallin had a pleiotropic effect on the other γ-crystallins by affecting the normal development of the whole lens. In any case, this is an example that leads to a general conclusion in the analysis of genetic diseases. When trying to elucidate a genetic defect, a useful strategy might be, at least in model organisms, to start from the protein level instead of the DNA level (Fig. 1). Looking at an enormous number and at a broad spectrum of proteins offers the chance to detect not only the primarily defective protein, or several proteins of this type, but also coaffected proteins. The next step would be to identify the genes of the abnormal proteins, and then to return to the proteins and to higher levels of abnormal phenotypes. This strategy starts with a survey of the complex level of gene expression and does not postulate the rather unrealistic situation that only one gene is responsible for a genetic disease. Moreover, the proteins found to be affected may give us some hints towards the pathogenesis induced by the genetic defect.

5.3 Multifunctional diseases

Considering multifactorial diseases, which include polygenic factors, findings mentioned in this article suggest that even a single protein involved in such a disease might be a polygenetic trait (Fig. 5). Human diseases are usually defined in terms of clinical symptoms such as high blood pressure, heart malformations or mental retardation. There is no doubt that several genes are involved in regulating blood pressure, and that many genes contribute to the normal development of heart and brain. Consequently, one can assume that any genetically based disease can be caused by a defect in one of the numerous genes involved in a particular disease. Many different genes, if mutated, may be able to induce, for example, microcephaly. A distinct patient, however, showing the symptoms of microcephaly, usually carries a mutation in only one of these genes, *i.e.*, whereas any genetic disease, as defined by clinical symptoms, might be polygenic in nature, with respect to an individual patient, each genetic disease is monogenic. This implies that the phenotype of a certain disease, the symptoms, may differ more or less among patients as far as different genes are involved. When, far in the future, the specific function of each of the genes is known, as well as their role in the regulatory and metabolic network, one will realize that each patient has his own disease. Genetic diseases will then be defined by the special genes affected rather than by the abnormal phenotypes described by the physicians.

Received January 28, 1999

6 References

[1] King, R. C., Stansfield, W. D., *A Dictionary of Genetics*, Oxford University Press, New York, Oxford 1990, p. 239.

[2] Nock, C., Gauss, C., Schalkwyk, L. C., Klose, J., Lehrach, H., Himmelbauer, H., *Electrophoresis* 1999, *20*, 1027–1032.

[3] Editorial, *Genomics* 1997, *45*, 244–249.

[4] URL:http://www.tigr.org/tdb/mdb/mdb.html

[5] The *C. elegans* Sequencing Consortium, *Science* 1998, *282*, 2012–2018.

[6] Goodfellow, P., *Nature Genet.* 1997, *16*, 209–210.

[7] Anderson, N., *Nature* 1979, *278*, 122–123.

[8] Anderson, N., Anderson, L., *Clin. Chem.* 1982, *28*, 739–748.

[9] Wade, N., *Science* 1981, *211*, 33–35.

[10] Swinbanks, D., *Nature*, 1995, *378*, 653.

[11] Klose, J., *Methods Mol. Biol.* 1999, *112*, 67–86.

[12] Klose, J., *Methods Mol. Biol.* 1999, *112*, 147–172.

[13] Klose, J., *Humangenetik* 1975, *26*, 211–243.

[14] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[15] Prehm, J., Jungblut, P., Klose, J., *Electrophoresis* 1987, *8*, 562–572.

[16] Gauss, C., Kalkum, M., Löwe, M., Lehrach, H., Klose, J., *Electrophoresis* 1999, *20*, 575–600.

[17] WORLD-2DPAGE URL: http://www.expasy.ch/ch2d/2d-index.html

[18] 2DWG metadatabase URL: http://www.lecb.ncifcrf.gov/2dwgDB

[19] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[20] Janke, C., Holzer, M., Goedert, M., Arendt, T., *FEBS Lett.* 1996, *379*, 222–226.

[21] Glotzer, M., Murray, A. W., Kirschner, M. W., *Nature* 1991, *349*, 132–138.

[22] Stuart, D. I., Jones, E.Y., *Nature* 1997, *386*, 437–438.

[23] Damerval, C., Maurice, A., Josse, J. M., deVienne, D., *Genetics* 1994, *137*, 289–301.

[24] King, R. C., Stansfield, W. D., *A Dictionary of Genetics*, Oxford University Press, New York, Oxford 1990, p. 244.

[25] Jungblut, P., Otto, A., Favor, J., Löwe, M., Müller, E.-C., Kastner, M., Sperling, K., Klose, J., *FEBS Lett.* 1998, *435*, 131–137.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 653–659

Review

Ian Humphery-Smith

The University of Sydney,
Centre for Proteome
Research and Gene-Product
Mapping,
National Innovation Centre,
Australian Technology Park,
Eveleigh, Australia

Replication-induced protein synthesis and its importance to proteomics

Replication-induced protein synthesis (RIPS) can occur following the passage of the replisome due to transcription initiated by RNA polymerase in association with: (i) negative supercoiling trailing the replisome / replication fork, (ii) hemimethylation prior to the action of *dam* methylase, (iii) transient derepression following passage of the replisome / replication fork and prior to renewed synthesis of the repressor gene-product, and (iv) 'sliding clamp' accessory DNA-binding proteins binding to the lagging strand DNA duplex to retard rotational upstream propagation of supercoils. The latter include subunits of DNA polymerase III in *Escherichia coli* and gp45 in T4 bacteriophage. By far the most convincing evidence for the existence of RIPS comes from the pulse of protein synthesis which follows the passage of the replisome in late T4 bacteriophage, the dynamics of replication in *Escherichia coli*, recent results from cDNA high-density expression arrays in yeast and the workings of the *lac*-operon. More circumstantial evidence is provided by 'leaky' or 'aberrant' protein expression in genetic systems where attempts have been made to turn off protein synthesis by molecular means. In higher vertebrates, RIPS may have a potentially important role in explaining the mechanisms by which thymic and peripheral immune self-tolerance is established, either directly through antigen presentation on dendritic cells or through the presentation of peptides derived from T-cells. The latter model is preferred, as young T-cells will have recently divided and will be dying in large numbers near the antigen-presenting dendritic cells in the thymus. The functional utility of RIPS would appear to be linked to both facilitating cellular metabolism and an improved survival during stress. RIPS, as a potentially universal molecular phenomenon, presents proteomics with numerous challenges and opportunities, both technical and commercial.

Keywords: Replication-induced protein synthesis / Proteomics / Functional genomics / Gene products / Arrays / Expression profiling / Self tolerance / Bioholonics / Review EL 3443

Contents

Correspondence: Dr. Ian Humphery-Smith, The University of Sydney, Centre for Proteome Research and Gene-Product Mapping, National Innovation Centre, Australian Technology Park, Eveleigh, Australia, 1430.
E-mail: ian@proteome.usyd.edu.au
Fax: +61-2-9319-1081

Abbreviations: **RIPS**, replication-induced protein synthesis; **SGP**, second generation proteomics

1 Introduction

Proteomics is thought to have a viable future in the expression profiling of biological systems [1]. However, this long term future will depend upon its ability to provide quantitative and qualitative analyses concerning the near-to-total proteome of biological systems in health and disease [2]. When Wasinger *et al.* [3] first defined the term 'proteome' as the "total protein complement of a genome", they were aware of the unlikely event that: "the totality of

this potential for protein expression will be realised at any given instant". However, there is a significant body of evidence to suggest that, in all biological systems, all open reading frames (ORF) are transcribed and translated to provide a 'background soup' of at least a few protein molecules (≥ 20–200) representative of every gene product. Although representative of each ORF within a genome, it is nonetheless improbable that all possible splice variants and the co- and post-translational modifications will be synthesised as part of this 'molecular soup'. The latter is thought to arise from replication-induced protein synthesis (RIPS). Superimposed upon this molecular soup is the protein expression patterns that characterise different cell, organ and tissue types, and growth stages and physiological states of an organism. In turn, the intracellular abundance of particular gene products derived from RIPS is determined by the relative influences of the length of cell cycle or replication time and the molecular half-lives of both mRNA and protein gene products, *i.e.*, it is more likely to see replication-induced proteins well represented in a rapidly dividing cell type, such as activated lymphocytes, as opposed to neurons, which divided rarely.

2 Mechanisms responsible for RIPS

The likely universality of RIPS can be linked to the high degree of conservatism in biological systems for both possession of DNA and the manner in which it is duplicated in order to transfer information from one generation to the next. In dividing cells, there exists a small window of time during which RIPS can occur and before genomic DNA is again protected from the enzymes necessary for transcription. The RNA polymerases and associated molecular machinery exist in the immediate vicinity of replicating DNA. Given the availability of mRNA transcripts, it is also most probable that translation would follow these transcription events. RIPS can occur following the passage of the replisome due to transcription initiated by RNA polymerase [4, 5] in association with: (i) negative supercoiling trailing the replisome / replication fork, (ii) hemimethylation prior to *dam* or other methylase activity, (iii) transient derepression following passage of the replisome / replication fork and prior to renewed synthesis of the repressor gene-product, and (iv) 'sliding clamp' accessory DNA-binding proteins binding to the lagging strand DNA duplex to retard rotational upstream propagation of supercoils. The latter include subunits of DNA polymerase III in *Escherichia coli* and gp45 in T4 bacteriophage.

2.1 Evidence

By far the most convincing evidence for the existence of RIPS comes from the pulse of protein synthesis which follows the passage of the replisome in late T4 bacteriophage [6, 7]. Other evidence, predating the complete genomic sequence of *Escherichia coli* [8], can be derived from the dynamics of replication in this organism, as determined by calculations based upon complexity measures and DNA / RNA hybridization [9]. Although the number of genes was significantly underestimated, the mathematics and experimental observations would appear to hold true. Indeed, this work [9] demonstrated that the near totality of the one strand coding capacity of *E. coli* underwent transcription. Estimates based on ORFs detected in the *E. coli* genome suggest that the coding capacity is in the order of 87.8% of 4 639 221 nucleotides [8], not at all incongruous with the conclusions derived by radiolabeling and hybridization [9]. More recently, data have become available from high-density arrays containing complementary DNA for every ORF found in the yeast *Saccharomyces cerevisiae* [10, 11]. In this simplest of eukaryotes with multiple chromosomes, the authors were surprised to observe transcripts corresponding to some 87% of ORFs. The absence of a transcript being detected for the remaining genes could be explained by the difficulty of detecting the least abundant transcripts and of optimising the stringency of all cDNA templates contained within the array. Thus, these results are interpreted as being consistent with the existence of RIPS. Indeed, the authors were surprised by such a high level of transcript generation. Previously, such levels of overall genomic expression were thought to be likely only in association with a variety of physiological and stress conditions. The *lac* operon was for many years a major enigma in our understanding of gene regulation, it is now among the best known and better-studied systems in molecular biology. From Nobel laureate Jacques Monod comes the following quotation concerning enzymatic adaptation: "A bacterium with one molecule of enzyme could metabolise lactose to make more enzyme and therefore had a great advantage" [12]. Uncertainty remains, however, only as to the origin of that 'one molecule' of β-galactosidase required to engender others. Much involved in the final demonstration of the *lac* repressor [13], Walter Gilbert (personal communication) recalled that the origin of these molecules was 'linked to growth'. Thus, the tenet here would be that the source of the initial β-galactosidase molecules is most likely explained by RIPS. The functional utility of possessing these few molecules will be discussed below.

2.2 Circumstantial evidence

Circumstantial evidence for the existence of RIPS can be found in molecular systems where researchers have endeavoured to turn off gene expression completely in organisms as diverse as transgenic mice and *Escherichia coli*. Here, the phenomenon of 'aberrant' or leaky' expres-

sion is well known to molecular biologists [14], again most likely due to RIPS. In higher vertebrates, RIPS may help explain the mechanisms by which both central and peripheral immune self-tolerance is established, either directly through antigen presentation on dendritic cells or through the presentation of peptides derived from T-cells. The latter model is preferred, as young T-cells will have recently divided and will be dying in large numbers near the antigen-presenting dendritic cells. Modern immunology and the recognition of non-self in mammals is dependent upon T-cells being exposed to all self-antigens during ontogeny and/or self-reactive T-cells being selectively excluded from the body. Without this exposure, the organism runs the risk of developing autoimmune responses. Thus, RIPS may have a pivotal role in explaining how self-tolerance is established in higher vertebrates. The levelof antigen expression is critical in this priming of developing T-cells [15, 16], while the origin of 'ignorant' T-cells leaving the thymus is most likely to be due to low abundance and/or low affinity antigens [17]. Thus, antigens not sufficiently well expressed in the thymus may need to be reinforced peripherally. Parijs and Abbas [18] further qualified this situation by the suggestion that: "different self-antigens may induce peripheral tolerance by different mechanisms. For instance, tissue-restricted antigens present at low concentrations may induce anergy, and widely disseminated and abundant self-antigens may trigger activation-induced cell death". The existence of redundancy and secondary back-up is considered logical, especially in a mechanism as essential to the well-being of an organism as the recognition of nonself by the immune system. In such cases, having escaped thymic stimulation and ensuring apoptosis, self-reactive T-cells would then be eliminated peripherally. Although the potential for exposure to a variety of self-antigens may be greatest peripherally, there exists presumably a good number of rarely expressed self-antigens (*e.g.*, those expressed during embryogenesis or wound healing). If these antigens have not been previously exposed to developing T-cells, then self-reactivity is possible. RIPS provides a mechanism whereby gene products for all genes in a genome could be present in multiple tissues at all times, thus allowing effective formation of T-cells.

3 Functional utility

If RIPS is indeed occurring in most biological systems, what then could be its biological significance? The most likely role of replication-induced proteins intra-cellularly would be to provide a short-term or 'stop gap' solution to cells while they await the relatively slow response time prior to synthesis of new gene products required by a cell when exposed, for example, to potentially lethal stress. The functional utility of RIPS would appear to be linked to both facilitating cellular metabolism and improved survival during stress. If a cell must synthesize large proteins while under severe stress, this synthesis is likely to occur less efficiently, as a result of the stress, and probably more slowly than under ideal conditions. In bacterial systems growing exponentially, amino acid residues can be synthesized at the rate of 20–21 residues per second [19]. Thus, even under optimal growth conditions, a 70 kDa heat shock protein would necessitate a lag phase > 30 s. Under less favourable conditions, synthesis of such proteins could require a lag phase of 1–2 min and be synthesized poorly due to chaperonin inefficiencies. Given that reaction times for biochemical pathways can occur over nanoseconds, there may exist considerable survival value and evolutionary significance in possessing low concentrations of all metabolites likely to be required by an organism at some time in the future, *i.e.*, to allow the cell to respond rapidly to a physiological constraint. This could also be an energetically cost-effective strategy, if this low-level protein synthesis were linked to cellular reproduction. Note that the biochemical relevance of RIPS may be restricted in cells during stationary phase or possessing extended cell cycles, except for proteins with extended molecular half-lives. Perhaps, therefore, variable half-lives of gene products provide a means whereby nature can selectively preserve those elements engendered by RIPS that are most essential to cellular well-being, even in non-rapidly-dividing cell types, *i.e.*, in order to maximize the corresponding selective advantage to the organism. The cellular utility of RIPS can again be demonstrated from our knowledge of the workings of the *lac* operon. The final induction of β-galactosidase is not possible without the prior existence of low concentrations of both *lac* permease and β-galactosidase. Indeed, the transformation of lactose into allo-lactose (the specific inducer) by β-galactosidase is the means whereby, in the presence of lactose, the system is capable of responding by a 1000- to 10 000-fold up-regulation in the abundance of the enzyme required to metabolise lactose [20, 21].

4 Intracellular abundance of proteins derived from RIPS

The *lac* repressor can be considered as a housekeeping gene, yet it is in molecular excess at a concentration of 20–30 molecules per cell [13]. If the *lac* repressor gene product is in fact derived from RIPS, then a number of molecules in the order of 20–200 could be considered as reasonable for having been synthesised by this method of transcription/translation. Other examples are known of biologically relevant gene products that occur at similarly low intracellular concentrations [4]. More importantly, however, large mRNA templates with extended half-lives

could synthesize significant quantities of reaction components for cellular processes.

5 Consequences for proteomics

RIPS, as a potentially universal molecular phenomenon, presents proteomics with numerous challenges and opportunities, both technical and commercial. The appreciation of opportunities in this sector would appear to have become more evident in recent times to both the biotech and pharmaceutical industry and financial commentators [22–24]. As recently as 16 April 1998, Oxford GlycoSciences underwent a successful Initial Public Offer on the London Stock Exchange for some $51 million (USD), while the biotechnology investment advisers Frost & Sullivan presented the following assessment in *Healthcare Market Engineering News*, March 1998: 'Strategically speaking, it is a bad time to enter into the genomics race. A more lucrative and relatively unexplored field is "proteomics" and ..... 'shift the drug development emphasis from genomics to the study of its higher-value products, proteomics – in effect, climbing up the value chain'. Although encouraging sentiments, some major hurdles still wait to be overcome.

5.1 Proteomic contigs

Through the use of 'proteomic contigs' (stitching together windows of protein expression as displayed by two-dimensional electrophoresis and based upon overlapping M_r and p*I* in a manner analogous to DNA sequence contigs) [25], Wasinger *et al.* (in preparation) have recently been able to characterize some 136 proteins encoded by the genome of *Mycoplasma genitalium*, the simplest self-replicating organism on the planet and containing just 470 ORFs. Within eight 'proteomic contigs', 427 distinct protein spots were visualized, while the gene products characterized were encoded by just 110 genes. Thus, although representing the most advanced display of a total proteome globally, no more than 20% of the expected protein diversity was characterized and no more than 73% of the expected gene products were visualized. This is inadequate if one is to move to more complex systems such as humans, where a few hundred thousand proteins are expected. The latter figure includes isoforms derived from differential splicing and co- and post-translation modifications of these gene products. Current technologies based on principally two-dimensional gel electrophoresis have not been able to resolve more than 11 000 proteins in mammalian systems [26], less than at very best 5–7% of the expected human proteome. If proteomics is to become the mainstay of functional genomics, then it must make the transformation to even higher throughput in conjunction with high sensitivity. However, with current-generation proteomics, there exists an inverse relationship between the level of sensitivity and the extent of automation [27].

5.2 Second-generation proteomics

For this reason, a paradigm shift is required in the manner in which proteomics is conducted, if it is to deliver competitiveness with high density cDNA expression arrays and other technologies employed to follow the mRNA gene products [28]. Progenitors of this paradigm shift are becoming apparent and include such approaches as: (i) green fluorescent protein Genome Reporter Matrixes [29], (ii) transposon tagging (Snyder, personal communication), (iii) protein / ligand arrays (*cf.* Ciphergen's SELDI ProteinChip arrays), (iv) yeast two-hybrid system [30–32], (v) mRNA-peptide fusion libraries [33], PROfusion, and (vi) fusion library arrays [34]. As such, 'second generation proteomics' (SGP) is defined as array technologies used to detect the total protein complement of a genome without calling upon the separation sciences (for example, two-dimensional gel electrophoresis, mass spectrometry, column chromatography or capillary electrophoresis), but rather employing more traditional molecular biological approaches to conduct holistic analysis of cellular proteins. However, such procedures must maintain a similar focus to that of first or current generation proteomics, for example, the detection of the levels of protein expression in a variety of virulent and/or drug-resistant bacterial strains, or the differences associated with health and disease progression in a multitude of individuals. This can then be supplemented by information pertaining to protein-protein and protein-nucleic acid interactions and co- and post-translational modifications, the objective being the detection of potential drug targets or gene products of relevance to precocious diagnosis or intervention strategies for disease prevention.

5.3 Current limitations

Although encouraging, the technologies detailed above have yet to surpass the resolution of current generation proteomics when applied to the human genome / proteome and remain unable to dissect the multitude of isoforms associated with mammalian gene products. Indeed, these approaches are poorly amenable to a human setting, where mutant strains positioned on arrays will be unacceptable. However, the primary objective of proteomics must remain the detection, in parallel, of the levels of protein expression within complex biological systems and the relevance of these levels of protein expression to experimental procedures or health and disease. Nonetheless, the application of array technologies to proteomics will be

accompanied by advantages similar to those seen in other areas of genomic science [35], and concomitantly, potentially provide solutions and improvements with respect to procedures currently employed within the discipline. The current technological limitations associated with proteomics are: (i) a dependence upon the highly demanding technologies of two-dimensional gel electrophoresis and mass spectrometry, (ii) a lack of comparable sensitivity to that achieved by the polymerase chain reaction (PCR) used in the analysis of nucleic acids, (iii) an inability to visualize the total proteome, and (iv) the need for truly international reproducibility readily accessible by all. The technologically demanding nature of the techniques currently employed within proteomics has meant that proteomic analysis is far from being implemented in every molecular laboratory globally. In all cases, proteomics requires a far greater investment in both personnel and hardware than do the majority of kit-based procedures currently playing a dominant role within the molecular laboratory. If the above obstacles can be overcome by either current or SGP, several major challenges and opportunities are offered to proteomics by the phenomenon of RIPS and its potential universality.

6 Challenges

The challenges facing proteomics can be largely summarized by the need to deliver quantifiable information on the 'near-to-total' proteome. To achieve the display and quantification of the 'near-to-total' proteome, inadequacies in the following areas must be redressed specifically. These include: (i) small gene products which comprise, for example, approximately 20% of the *Escherichia coli* genome [36, 37], (ii) high molecular mass basic proteins which comprise approximately 25–30% of all bacterial genomes sequenced thus far, (iii) hydrophobic proteins, which are poorly resolved in two-dimensional electrophoresis gels, and (iv) low abundance protein analytes, which may comprise as much as 50–80% of all cellular proteins in mammalian systems, as few as 20–30 molecules per cell, and possibly be involved in important 'housekeeping' and response mechanisms intracellularly. The inability to overcome these shortcomings has meant that the discipline is currently capable of complementing, rather than delivering competitiveness within the genomic sciences, in particular, 'biochips' and high-density cDNA expression arrays. An excellent example of the latter technology and the insights provided into physiological analysis is seen in the recent work of Iyer *et al.* [38].

7 Opportunities

The following combine to suggest that, if current technological inadequacies can be overcome, proteomics is likely to become the mainstay of functional genomics: (i) A transcript and probably a translated protein product are likely to occur in association with all ORFs in rapidly reproducing cells. (ii) In higher eukaryotes, less than 10% of the genome is thought to correspond to genes. As such, approaches directed towards the analysis of gene products can effectively overcome some of the difficulties associated with sequencing entire genomes and the inherent problems of low complexity regions and genomic duplications. (iii) Proteomics is capable of delivering far more reliable information than that currently being accessed by either small or large high-density cDNA expression arrays. The latter is best substantiated by the difficulties presented by the highly variable half-lives of mRNA transcripts in living systems (Fig. 1). This variability can translate into dramatically different quantities of intracellular proteins with respect to that predicted from mRNA alone [39, 40]. During exponential growth phase, *Escherichia coli* synthesizes approximately 20–21 amino acid residues per second, yet mRNA half-lives can vary between a few seconds and several hours [41–46]. This phenomenon further underscores the utility of addressing directly the activity of the 'cellular workhorses', namely, proteins. In addition, the corresponding half-lives of both the mRNA template and the translated protein allow the levels of gene products to be regulated intracellularly, where cellular analytes are not all required in equal amounts. Obviously, mechanisms such as promotion and suppression of transcription also play a dominant role in regulating expression levels of gene products.

8 Bioholonics

Both first- and second-generation proteomics are currently able to deliver reliable information on a significant portion of the total expected proteome in microbial systems, thereby permitting parallel analysis of cellular mRNA and proteins in fully sequenced genomes. This technological progress is readying itself to herald in a new era of holistic cellular analysis, namely 'bioholonics'. Such information should allow the current preoccupation with the absolute quantity of gene product (RNA and/or protein) to move backstage with respect to more molecularly relevant parameters, such as molecular half-life, synthesis rate, functional competence (presence or absence of mutations), reaction kinetics, the influence of individual gene products on biochemical flux, the influence of the environment, cell-cycle, stress, drug administration and disease on gene products, and the collective roles of multigenic and epigenetic phenomena governing cellular processes. The term 'bioholonics' is used here for the first time in the scientific literature, but owes its origins to an internal report for the Japanese Government prepared by Teruhisa Noguchi in 1980.

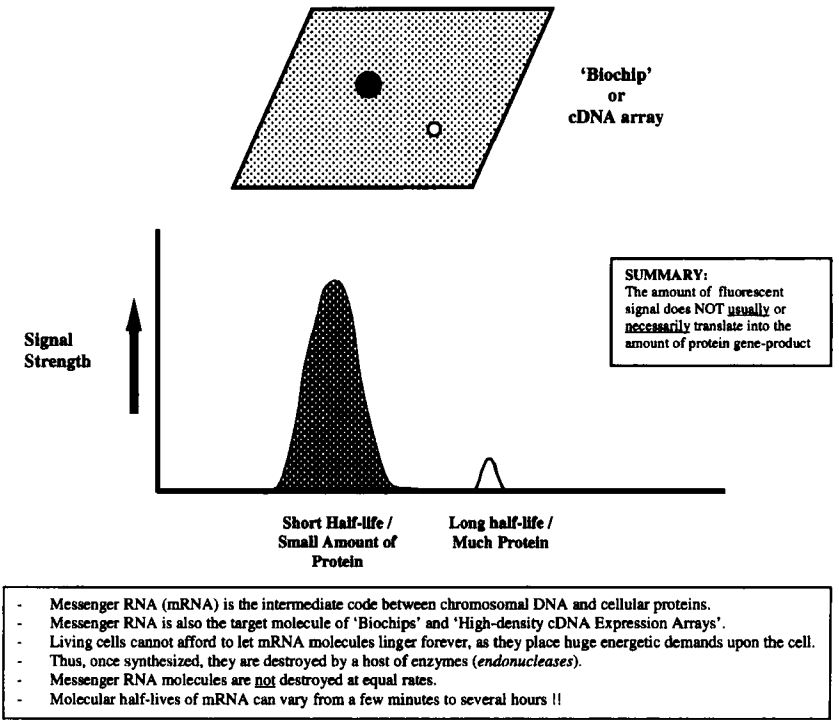


Figure 1. The variability of mRNA half-lives and preliminary results [23, 36, 37] would suggest that high-density cDNA expression arrays are poor predictors of intracellular protein abundance.

9 Conclusions

The mechanisms suggested for the occurrence of RIPS and the evidence for its existence would contest that a protein product corresponding to all ORFs is likely to be found in rapidly dividing cells, as also proposed in a recent study detailing mRNA expression analysis of a single human lymphocyte [47]. Many of these gene products will occur at very low intracellular abundance. Thus, proteomics must evolve if it is to contribute an ever-increasing knowledge base of intellectual property of relevance to the biotechnological and pharmaceutical industries. This contribution need not always be subservient to approaches based upon analysis of nucleic acids. The significant advances achieved in protein chemistry over the last few years, particularly the characterization of proteins by mass spectrometry, will continue to complement both genomics and proteomics. SGP is likely to significantly increase the demand for a variety of analytical procedures in protein chemistry, providing value-added information for individual gene products. Although protein-protein and protein-nucleic acid interactions and the nature of co- and post-translational modifications will continue to be of interest, the initial focus of proteomics must remain the relative levels of protein expression in biological systems. Still further technological advances are required to face the major hurdle of coming to grips with the total protein expression of human cells, tissues and organs. However, cellular molecular biology dictates that the quest for novel drug targets or intervention strategies for disease prevention must incorporate information concerning proteins. Thus, the challenge for proteome analysis into the next century lies clearly with the task of achieving a combination of high-throughput screening, while maintaining high sensitivity for the detection of low-copy-number proteins. The latter may constitute the bulk of proteins in humans and higher enkaryotes. As in genomics, array technologies offer the greatest hope of achieving this. Current-generation, separation-dependent technologies are delivering valuable information in many sectors of biology, but particularly with respect to entire microbial proteomes (see earlier) and protein complexes [48–50]. Far from reducing the importance of these approaches, array-based proteomics is likely to become the primary discovery platform and subsequently increase demand for procedures calling upon the significant advances witnessed in protein chemistry over the last 4–5 years. Lessons learnt from the analysis of nucleic acids indicate that proteomics has reason to diversify the approaches currently employed during problem solving and analyte detection.

I wish to thank Valerie Wasinger for initially drawing my attention to replication-induced protein synthesis and the

following for valuable discussion and comments regarding this phenomenon: Michael Panaccio, Staffan Kjelleberg, Graham Mitchell, John Sprent, Walter Gilbert, George Church, Frederick Neidhardt, Barbara Fazekas de St. Groth and members of the University of Sydney, Centre for Proteome Research and Gene-Product Mapping.

Received February 23, 1999

10 References

[1] Kozian, D. H., Kirschbaum, B. J. *Trends Biotech.* 1999, *17*, 73–78.

[2] Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P., *Electrophoresis* 1997, *18*, 1217–1242.

[3] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[4] Guptasarma, P., *BioEssays* 1995, *17*, 987–997.

[5] Guptasarma, P., *BioEssays* 1996, *18*, 325–332.

[6] Herendeen, D. R., Kassavetis, G. A., Barry, J., Alberts, B. M., Geiduschek, E. P., *Science* 1989, *245*, 952–958.

[7] Brody, E. N., Kassavetis, G. A., Ouhammouch, M., Sanders, G. M., Tinker, R. L., Geiduschek, E. P., *FEMS Microbiol. Lett.* 1995, *128*, 1–8.

[8] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y., *Science* 1997, *277*, 1453–1462.

[9] Hahn, W. E., Pettijohn, D. E., Van Ness, J., *Science* 1977, *197*, 582–:585.

[10] Wodicka, L., Dong, H., Mittmann, M., Ho, M.-H., Lockhart, D. J., *Nature Biotech.* 1997, *15*, 1359–1367.

[11] DeRisi, J. L., Iyer, V. R., Brown, P. O., *Science* 1997, *278*, 680–686.

[12] Cohn, M., in: Miller, J. H., Reznikoff, W. S. (Eds.), *In Memoriam to Jacques Monod* in *The Operon*, pp 1–10, Cold Spring Harbor Laboratory, 1978.

[13] Gilbert, W., Muller-Hill, B., *Proc. Natl. Acad. Sci. USA* 1996, *56*, 1891–1898.

[14] Yarranton, G. T., *Curr. Opinion Biotech.* 1992, *3*, 506–511.

[15] Hogquist, K. A., Jameson, S. C., Heath, W. R., Howard, J. L., Bevan, M. J., Carbone, F. R., *Cell* 1994, *76*, 17–27.

[16] Sebzda, E., Wallace, V. A., Mayer, J., Yeung, R. S. M., Mak, T. W., Ohashi, P. S., *Science* 1994, *263*, 1615–1618.

[17] Sprent, J., Kishimoto, H., *Ann. N.Y. Acad. Sci.* 1998, *284*, 236–245.

[18] Van Parijs, L., Abbas, A. K., *Science* 1998, *280*, 243–248.

[19] Bremer, H., Dennis, P. P., in: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, Jr., K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology* 1996, ASM Press, Washington DC, pp. 1553–1569.

[20] Miller, J. H., Reznikoff, W. S., *The Operon* 1978, Cold Spring Harbor Laboratory, 1978.

[21] Müller-Hill, B., *The* Lac *Operon: A Short History of a Genetic Paradigm*, Walter de Gruyter, Berlin 1996.

[22] Persidis, A., *Nature Biotech.* 1998, *16*, 393–394.

[23] Blackstock, W. P., Weir, W. P., *Trends Biotech.* 1999, in press.

[24] Wang, J., Hewick, R., *DDT* 1999, *4*, 129–133.

[25] Wasinger, V., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[26] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[27] Haynes, P. A., Gygi, S. P., Figeys, D., Aebersold, R., *Electrophoresis* 1998, *19*, 1862–1871.

[28] Fraser, C. M., Fleischmann, R. D., *Electrophoresis* 1997, *18*, 1207–1216.

[29] Glaser, V., *Genetic Engin. News* 1997, *17*, September 15.

[30] Bartel, P. L., Fields, S., *Methods Enzymol.* 1995, *254*, 241–263.

[31] Phizicky, E. M., Fields, S., *Microbiol. Revs.* 1995, *59*, 94–123.

[32] Bai, C., Elledge, S. J., *Methods Enzymol.* 1996, *273*, 331–347.

[33] Roberts, R. W., Szostak, J. W., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 12297–12302.

[34] Büssow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H., Walter, G., *Nucleic Acids Res.* 1998, *26*, 5007–5008.

[35] Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E., Davis, R. W., *Trends Biotech.* 1998, *16*, 301–306.

[36] Rudd, K. E., Humphery-Smith, I., Wasinger, V. C., Bairoch, A., *Electrophoresis* 1998, *19*, 536–544.

[37] Wasinger, V. C., Humphery-Smith, I., *FEMS Microbiol. Lett.* 1998, *169*, 375–382.

[38] Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, Jr., J., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., Brown, P. O., *Science* 1999, *283*, 83–87.

[39] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[40] Anderson, N. L., Anderson, N. G., *Electrophoresis* 1998, *19*, 1853–1861.

[41] Hargrove, J. L., Schmidt, F. H., *FASEB J.* 1989, *3*, 2360–2369.

[42] Higgins, C. F., *Curr. Opin. Cell Biol.* 1991, *3*, 1013–1018.

[43] Flärdh, K., Cohen, P., Kjelleberg, S., *J. Bacteriol.* 1992, *174*, 6780–6788.

[44] Sørensen, M. A., Vogel, U., Jensen, K. F., Pedersen, S., *Antonie van Leeuwenhoek* 1993, *63*, 323–331.

[45] McAdams, H. H., Arkin, A., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 814–819.

[46] Miller, C. G., in: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, Jr., K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Protein degradation and proteolytic modification in Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC 1996, pp. 938–954.

[47] Kimoto, Y., *Mol. Gen. Genet.* 1998, *258*, 233–239.

[48] Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, W., Ajuh, P., Sleeman, J., Angus Lamond, A., Mann, M., *Nature Genet.* 1998, *20*, 46–50.

[49] Grant, P. A., Schieltz, D., Pray-Grant, M. G., Yates III, J.R., Workman, J. L., *Mol. Cell* 1998, *2*, 863–867.

[50] Zhang, J.-G., Farley, A., Nicholson, S. E., Willson, T. A., Zugaro, L. M., Simpson, R. J., Moritz, R. L., Cary, D., Richardson, R., Hausmann, G., Kile, B. J., Kent. S. B. H., Alexander, W. S., Metcalf, D., Hilton, D. J., Nicos, A., Nicola, N. A., Baca, M., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 2071–2076.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 660–663

Review

Ben Herbert

Proteome Systems Ltd.,
North Ryde, Sydney,
Australia

Advances in protein solubilisation for two-dimensional electrophoresis

Two-dimensional (2-D) electrophoresis remains the highest resolution technique for protein separation and is the method of choice when complex samples need to be arrayed for characterisation, as in proteomics. However, in current proteome projects the total number of proteins identified from 2-D gels is often only a small percentage of the predicted proteome. In addition, there is an almost complete lack of hydrophobic proteins on 2-D gels, especially those using immobilised pH gradients. Recently there have been a number of publications reporting reagents which improve protein solubilisation prior to isoelectric focusing. The improved solubilization possible with these reagents has increased the total number of proteins able to be visualised on 2-D gels and also allowed the separation of hydrophobic proteins, such as integral membrane proteins.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Membrane proteins / Solubility / Proteome / Review
EL 3428

Contents

1 Introduction

As we reflect on the third Siena meeting it is appropriate to look at what progress has been made in the various proteomic technologies towards achieving the goals of proteomics, for example, to identify and characterise all proteins expressed by an organism or tissue [1]. The scope of this review is sample preparation for 2-DE and I would like to focus on the advances that relate to enhanced protein solubility and increased numbers of hydrophobic proteins on 2-D gels. For a comprehensive review of many other aspects of sample preparation for electrophoresis, see Rabilloud [2]. Two-dimensional electrophoresis (2-DE) remains the highest resolution method for arraying proteins prior to their characterisation by mass spectrometry; however, the number of proteins actually identified on 2-D gels, even from species where the entire genome is sequenced, is very low. Wilkins *et al.* [3] reported that for three species, *Bacillus subtilis, Escherichia coli* and *Saccharomyces cerevisiae,* with completely sequenced genomes, less than 5% of the proteins in SWISS-PROT had been identified on 2-D gels. In addition, within the 5% of identified proteins, only between 5% for *S. cerevisiae* and 15% for *E. coli* were hydrophobic, compared to the theoretical hydrophobic contents of 16% and 29%, respectively. An initial appraisal of these figures may lead to the conclusion that 2-DE and proteomics is failing to deliver; however, there have been a number of recent advances to sample preparation and 2-D gel methodology [4–7] which are already allowing more proteins to be arrayed in micropreparative quantities.

Pretreatment of samples for isoelectric focusing (IEF) involves solubilisation, denaturation and reduction to completely break the interactions between the proteins and to remove nonprotein sample components such as nucleic acids [2]. Ideally, to avoid protein losses, one would achieve complete sample solubilisation in a single step and thus eliminate unnecessary handling, as is the case for soluble protein samples which can be readily taken up in the most commonly used IEF sample solution of 8 M urea, 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 50–100 mM dithiothreitol

Correspondence: Dr. Ben Herbert, Proteome Systems Ltd., Locked Bag 2073, North Ryde, Sydney, NSW 1670, Australia
E-mail: ben.herbert@proteomesystems.com
Fax: +61-2-9889-1805

Abbreviations: ASB 14, amidosulfobetaine 14; **GRAVY**, grand average of hydropathy; **SB 3–10**, *N*-decyl-*N,N*-dimethyl-3-ammonio-1-propane sulfonate; **TBP**, tributyl phosphine

(DTT) and 40 mM Tris. However, the 'standard' IEF sample solution is not ideal for many proteins and the challenge for 2-D PAGE, particularly in the context of proteomics, is the solubilisation and separation of insoluble samples such as membrane and membrane-associated proteins and proteins from highly resistant tissues like hair and skin. Enhanced protein solubility has been reported by introducing new reagents such as thiourea, sulfobetaine surfactants and tributyl phosphine into the IEF sample solution [4, 5].

To obtain a broad picture of a sample, the unfractionated proteins can be applied to a single 2-D gel with a wide pH gradient, but with complex samples this approach only reveals a small percentage of the proteome. Organelle and plasma membrane fractions can be used to considerably reduce the complexity of cellular samples; however, they require considerable expertise and access to expensive equipment such as ultracentrifuges. In contrast, the simple concept of sequentially extracting proteins by exploiting differential solubility has been used on many different tissues, usually with aqueous, organic solvent, and surfactant-based extraction solutions [8–10]. Recently Molloy *et al.* [6] have expanded on the sequential extraction approach by incorporating highly solubilising conditions in the IEF sample solutions [4, 5]. Denaturing IEF sample solutions are made up of three main types of reagents, chaotropes, surfactants and reducing agents, and in the main part of this review I will summarise the advantages of some of the new reagents which constitute the enhanced, highly solubilising IEF sample solution.

2 Chaotropes

Chaotropic agents such as urea allow proteins to unfold and thus expose their hydrophobic cores. This is achieved by changing the hydrogen bond structure in the solution, thus decreasing the energy penalty for contact of the hydrophobic residues with the solution [2, 11]. In practice, because the hydrophobic residues of proteins are exposed by urea denaturation it is normal to have surfactants, such as CHAPS, present to aid in protein solubilisation. At the 1996 Siena meeting, and subsequently in a number of publications [4, 12], T. Rabilloud has shown that proteins can be lost, by adsorption to the gel matrix, when IEF is conducted in immobilised pH gradients (IPGs). Rabilloud *et al.* [4] introduced the use of thiourea in combination with urea to increase the solubility of proteins in IPGs and showed that thiourea had a major positive effect on the number of proteins visualised in the second dimension with four different subcellular fractions, including integral membrane proteins. Thiourea is an efficient chaotrope, although it is poorly soluble in water and requires high concentrations of urea for solubility, the optimal conditions being solutions of 2 M thiourea in 5–7 M urea.

The enhanced protein solubility obtained by using thiourea has also been demonstrated by Pasquali *et al.* [13], who showed improved 2-DE of up to 2 mg of total membrane preparations and integral membrane preparations from murine mammary epithelial cells. Fialka *et al.* [14] combined subcellular fractionation of murine mammary epithelial cells with enhanced solubility using thiourea to create a range of organelle maps. They were able to separate and identify some transmembrane proteins such as E-cadherin and caveolin. However, neither of these proteins would be classified as highly hydrophobic by the grand average of hydropathy (GRAVY) measurement used by Wilkins *et al.* [3] to classify hydrophobic proteins in *B. subtilis, E. coli* and *S. cerevisiae.* Although the use of thiourea/urea mixtures alone is a major advance in the solubility of proteins for 2-DE, the combination of these chaotrope mixtures with new sulfobetaine surfactants has provided a whole range of powerful sample solutions for 2-DE.

3 Surfactants

As stated in Section 2 it is normal to have at least one surfactant present in the IEF sample solution to solubilise the hydrophobic residues that are exposed as a result of denaturation in chaotropes. Large amounts of ionic substances are not compatible with steady-state IEF; therefore, the use of SDS is not recommended and we are restricted to nonionic or zwitterionic surfactants. It is possible to use SDS in the initial sample solubilisation and then dilute it out with a large excess of a nonionic surfactant [2]; however, obtaining sufficient dilution of the SDS can become impossible when micropreparative protein loads are required. Traditionally, surfactants such as the Triton X-100 and Nonidet P-40 have been used, as well as sugar-based surfactants such as octyl glucoside [2, 15]. In recent years the sulfobetaine CHAPS has become the surfactant of choice and is generally used at between 2% and 5% in 8 M urea. These commonly used surfactants are soluble in high concentrations of urea, and thus it would appear appropriate to add thiourea to the sample solution and combine the enhanced chaotropic power with urea-soluble surfactants. However, while soluble, these surfactants are not efficient at protein solubilisation in high concentrations of chaotropes and their solubilising power is further minimised in the presence of highly chaotropic thiourea [4, 16]. In contrast, sulfobetaines with long linear alkyl tails such as *N*-decyl-*N,N*-dimethyl-3-ammonio-1-propane sulfonate (SB 3–10) are more efficient than CHAPS although they suffer from poor solubility in high concentrations of urea [17–19].

As a compromise between chaotropic power and surfactant efficiency, Rabilloud *et al.* [4] suggested two IEF sample solution formulations which take advantage of the increased chaotropic power of thiourea. The first solution, 5 M urea, 2 M thiourea, 2% CHAPS and 2% SB 3–10 is for proteins which require strong surfactants for solubility. The sulfobetaine SB 3–10 is not soluble in concentrations of urea greater than 5 M and thus the overall chaotrope concentration is relatively low. The second solution, 7 M urea, 2 M thiourea and 4% CHAPS is for proteins which require a high concentration of chaotropes.

To address the problem of surfactant-chaotrope compatibility, Chevallet *et al.* [20] synthesised a range of novel sulfobetaines with more polar, but zwitterionic, head groups and long alkyl tails of more than 12 carbon atoms. The new surfactants were tested on membrane preparations from bovine neutrophils and also on plasma membrane preparations from *Arabidopsis thaliana*. The most efficient surfactants were of the amidosulfobetaine type with either a linear alkyl tail, such as amidosulfobetaine 14 (ASB 14), or a mixed alkyl-aryl tail, such as C8Φ. The number refers to the number of carbon atoms in the tail. Because these surfactants have more polar head groups than sulfobetaines such as SB 3–10 they are soluble in high concentrations of urea, typically 7–8 M. When compared to the conventional IEF sample solution containing CHAPS, the new surfactants allowed the separation and detection of many more membrane proteins. Specifically, C8Φ allowed the separation of some, quite hydrophobic proteins such as the PIP 1 isoform of a plasma membrane water channel protein from *A. thaliana* [20]. PIP 1 has a GRAVY score of 0.37, which rates it as one of the most hydrophobic proteins ever detected on 2-D gels. In the case of the bovine neutrophil membranes the ASB 14 surfactant was highly efficient and resulted in many more proteins being visualised, including stomatin, a transmembrane protein [20].

4 Reducing agents

As discussed above, reagents such as urea and surfactants are used to denature proteins and expose their hydrophobic cores; however, to allow complete unfolding of many proteins it is necessary to reduce disulfide bonds. Reduction is usually achieved with a free-thiol-containing reducing agent such as β-mercaptoethanol or dithiothreitol (DTT) [2]. However, free-thiol-containing reagents such as DTT are charged, especially at alkaline pH, and thus migrate out of the pH gradient during the IEF, which results in a loss of solubility for some proteins, especially those which are prone to interation by disulfide bonding, such as the keratins and keratin-associated proteins from hair and wool. Herbert *et al.* [5] replaced the thiol-containing reducing agent DTT with an uncharged reducing agent, tributyl phosphine (TBP), which greatly enhanced protein solubility during the IEF and resulted in increased transfer to the second dimension with a range of samples including wool proteins. A further advantage of the mechanism of phosphine reduction is that because phosphines do not contain a thiol they cannot be alkylated, which leads to a simplified IPG equilibration protocol incorporating reduction and alkylation in a single step. Using an improved single-step IPG equilibration protocol [5], it is possible to combine TBP and an alkylating agent and obtain complete alkylation of cysteine. Because thiourea, sulfobetaine surfactants and TBP increase protein solubility through different routes, they can be seen as complementary reagents. Thiourea increases the chaotropic power of the sample solution, sulfobetaines such as SB 3–10, ASB 14 and C8Φ are efficient surfactants, and TBP ensures complete reducing conditions during the IEF; thus it is advantageous to combine all three in an enhanced IEF sample solution.

5 Sequential extraction

By combining thiourea, sulfobetaines such as SB 3–10 and the uncharged reducing agent TBP, it is possible to create a powerful IEF sample solution and thus solubilise proteins which would remain insoluble in conventional IEF sample solutions. However, the drawback of this enhanced solubility is that for complex samples the 2-D gels become swamped with overlapping proteins, especially when high loads are applied for micropreparative purposes. Molloy *et al.* [6] have adapted the concept of differential solubility [8–10] by incorporating the enhanced solubilising conditions as the final step of a sequential extraction of *E. coli*. The first step is cell lysis and protein extraction using Tris base. The resulting pellet is extracted using conventional IEF sample solution, 8 M urea, 4% CHAPS and DTT and the pellet remaining after this extraction is rich in membrane proteins and represents only 11% w/w of the starting material. The final extraction is with 5 M urea, 2 M thiourea, 2% CHAPS, 2% SB 3–10 and 2 mM TBP. Eleven membrane proteins were identified on the 2-D gel from the final extraction, representing many of the outer membrane proteins of *E. coli*. Five of the identified proteins had not previously been identified from 2-D gels and two of these were only previously known as open reading frames. It appears that many of the outer membrane proteins (OMPs) from *E. coli* were partitioned into the final pellet in a highly enriched state and were solubilised in the enhanced solubilising solution.

Note that while the proteins in the final extract are clearly insoluble in conventional IEF solutions they are not rated as hydrophobic according to the GRAVY scale [3, 6], as

they all have negative GRAVY scores. It seems that at the extremes of the GRAVY scale the values are predictive of protein solubility because the protein is mainly composed of either hydrophilic or hydrophobic residues. However, in the mid ranges of the GRAVY scale it is quite possible to have hydrophilic proteins, as measured by GRAVY, which are insoluble in the conventional IEF sample solutions [6]. The insolubility of a 'hydrophilic' protein may be due to the presence of hydrophobic, possibly transmembrane, domains that are, on average, outweighed by the majority of hydrophilic residues.

Given the efficiency of some of the novel sulfobetaine surfactants synthesised by T. Rabilloud *et al.* [20] it will be interesting to repeat the sequential extraction of Molloy *et al.* [6] using the new surfactants in the final step and look for more hydrophobic proteins. In another study, Molloy *et al.* [21] have used the novel sulfobetaine ASB 14 [20] as part of an enhanced solubilising solution to separate proteins extracted from *E. coli* with mixtures of chloroform/methanol. The ASB 14 solution gave greater protein solubility, compared to a solution using SB 3–10 and CHAPS, probably because of the increased efficiency of the ASB 14 and the fact that 7 M urea can be used with ASB 14, compared with only 5 M urea with SB 3–10. Out of the thirteen proteins identified, eight had not been previously identified on 2-D gels and five of these new proteins had positive GRAVY scores and are thus rated as hydrophobic.

6 Concluding remarks

The past few years have seen 2-DE undergo a revitalisation which is strongly linked to the interest in proteomics. The high throughput nature of mass spectrometry, especially MALDI-MS for peptide mass fingerprinting, has placed a large demand on 2-DE to deliver large numbers of micropreparative gels of 'unknown' proteins. It is clear that conventional technology, using urea, CHAPS, and DTT solubilisation on a single broad-range pH gradient is becoming exhausted. The same proteins are solubilised and identified many times, leaving behind the majority of the proteome, either insoluble or hidden behind the abundant proteins on a broad pH range 2-D gel. The work summarised in this review has made significant progress on the protein solubility issue and has allowed many previously insoluble proteins to be separated by 2-DE. In addition, the sequential extraction approach has proved to be a rapid and effective method of partitioning and concentrating insoluble proteins such as membrane proteins. A combination of organelle fractionation with sequential extraction will prove to be very powerful. For example, sequential extraction of purified organelles will provide cellular localisation data as well as simplifying the 2-D patterns by separating the soluble fraction from the membrane fraction. The next steps in the development of 2-DE for proteomics will be continued refinements to prefractionation and enhanced protein solubility, combined with 2-DE using multiple pH gradients to cover the expected p*I* range of most proteomes.

Received February 17, 1999

7 References

[1] Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., Williams, K. L., *Biotechnol. Genet. Eng. Rev.* 1995, *13*, 19–50.

[2] Rabilloud, T., *Electrophoresis* 1996, *17*, 813–829.

[3] Wilkins, M. R., Gasteiger, E., Sanchez, J.-C., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 1501–1505.

[4] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[5] Herbert, B. R., Molloy, M. P., Walsh, B. J., Gooley, A. A., Bryson, W. G., Williams, K. L., *Electrophoresis* 1998, *19*, 845–851.

[6] Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J.-C., Hochstrasser, D. F., Williams, K. L., Gooley, A. A., *Electrophoresis* 1998, *19*, 837–844.

[7] Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J.-J., Madjar, J.-J., *Electrophoresis* 1997, *18*, 328–337.

[8] Weiss, W., Postel, W., Görg, A., *Electrophoresis* 1992, *13*, 770–773.

[9] Gurd, J. W., Evans, W. H., Perkins, H. R., *Biochem. J.* 1972, *126*, 459–466.

[10] Ramsby, M. L., Makowski, G. S., Khairallah, E. A., *Electrophoresis* 1994, *15*, 265–277.

[11] Herskovitis, T. T., Jaillet, H., Gadegbeku, B., *J. Biol. Chem.* 1970, *245*, 4544–4550.

[12] Adessi, C., Miege, C., Albrieux, C., Rabilloud, T., *Electrophoresis* 1997, *18*, 127–135.

[13] Pasquali, C., Fialka, I., Huber, L. A., *Electrophoresis* 1997, *18*, 2573–2581.

[14] Fialka, I., Pasquali, C., Lottspeich, F., Ahorn, H., Huber, L. A., *Electrophoresis* 1997, *18*, 2582–2590.

[15] Righetti, P. G., *Isoelectric Focusing: Theory, Methodology and Applications*, Elsevier Biomedical Press, Amsterdam 1983, pp. 1–386.

[16] Navarette, R., Serrano, R., *Biochim. Biophys. Acta* 1983, *728*, 403–408.

[17] Dunn, M. J., Burghes, A. H. M., *Electrophoresis* 1983, *4*, 97–116.

[18] Gianazza, E., Rabilloud, T., Quaglia, L., Caccia, P., Astrua-Testori, S., Osio, L., Grazioli, G., Righetti, P. G., *Anal. Biochem.* 1987, *165*, 247–257.

[19] Rabilloud, T., Gianazza, E., Catto, N., Righetti, P. G., *Anal. Biochem.* 1990, *185*, 94–102.

[20] Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Keiffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1909.

[21] Molloy, M. P., Herbert, B. R., Williams, K. L., Gooley, A. A., *Electrophoresis* 1999, *20*, 701–704.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 664–677

Review

Proteomics and automation

Manfredo Quadroni
Peter James

Protein Chemistry,
ETH Zürich, Switzerland

Proteome analysis is concerned with the global changes in protein expression as visualized most commonly by two-dimensional gel electrophoresis and analyzed by mass spectrometry. A drastic increase in the rapidity and reproducibility of protein isolation and identification is needed for proteome analysis to become a useful complement to global mRNA analysis. Simplification and standardization, based on innovation in both hard- and software, are prerequisites to the creation of automated proteomics platforms that are both robust and user-friendly, and will allow many more laboratories access to this technique. In this review we highlight the weak points in the chain of analysis (such as sample handling, protein separation and digestion) and summarize recent trends toward automation in instrumentation and software and offer our own personal view of future developments in the field.

Keywords: Proteomics / Automation / High throughput / Mass spectrometry / Two-dimensional polyacrylamide gel electrophoresis / Review EL 3361

Contents

Correspondence: Dr. Peter James, Protein Chemistry Laboratory, Universitätsstrasse 16, ETH-Zentrum, CH-8092 Zürich, Switzerland
E-mail: peter.james@bc.biol.ethz.ch
Fax: +41-1632-1213

Abbreviations: CAD, collisionally activated dissociation; **EST**, expressed sequence tag; **MS/MS**, tandem mass spectrometry; **PSD**, post-source decay; **PTM**, post-translational modification

1 Introduction

The availability of complete genome sequences [1] or extensive expressed sequence tag (EST, [2]) libraries now allows the entire potential protein complement of organisms to be defined (the proteome). The focus of biological problem-solving must now move from a reductionist to a global approach. Instead of dissecting a process to identify a putative single effector, more subtle analyses based on monitoring the entire genome expression can be carried out. The description of a biological system by a qualitative and quantitative analysis of mRNA expression has now been made possible with the development of DNA microchips and arrays [3], differential display PCR [4] and

serial analysis of gene expression [5]. Genome-wide studies [6, 7] are now possible and true global experiments such as systematic analyses of perturbations of a system and following the return to equilibrium can be carried out. A prerequisite to these large-scale types of mRNA expression studies and to the sequencing of genomic DNA is a high degree of automation. Reproducibility is a prerequisite for a statistical analysis of these complex data sets and automation and high throughput data accumulation are essential for such large undertakings. The onus of extracting the salient features of a process from this flood of data is firmly in the hands of the bioinformatics experts. Data must be processed and displayed in such a way as to show the complex relationships between objects previously thought to be unrelated.

DNA and mRNA are physico-chemically homogeneous and 'easy to handle' and can be amplified by polymerase chain reaction methods, hence are amenable to automation. However, there are several key objections to the reduction of biological studies to the following changes in mRNA: (i) the level of mRNA does not allow one to predict the level of protein expression [8, 9], (ii) protein function is controlled by many post-translational modifications, and (iii) protein maturation and degradation are dynamic processes which dramatically alter the final amount of active protein independent of mRNA level. In order to perform proteome analysis on a large scale in an automated fashion, several requirements must be met which allow: (i) the extraction and high-resolution separation of all protein components, including membrane, extreme p*I* and low copy number proteins; (ii) the identification and quantitation of each component; and (iii) the comparison, analysis, and visualization of complex changes in expression patterns.

Proteins, in contrast to nucleic acids, are vastly more diverse and a universal handling method is unlikely to be found. Currently the only systematic methods for analyzing the state of expression of the majority of proteins in a cell are those based on two-dimensional gel electrophoresis [10–12]. In order for proteome analysis to become a viable and widely used method, a reasonable degree of automation must be achieved to: (i) increase reproducibility to facilitate data comparison between laboratories, and (ii) make the process less labor-intensive and increase throughput. In this manuscript we will outline the approaches and pitfalls in trying to automate protein identification and quantification methods for comprehensive proteome analysis. We have divided the process into four parts: sample preparation to gel imaging (Fig. 1, Section 2, 3), image to spot digestion (Fig. 2, Section 3, 4), digest to database searching (Fig. 3, Section 4, 5), and search results to data analysis and archiving (Fig. 4, Section 6, 7). Finally we discuss the trends in analytical methods developments that will shape the future of proteomics.

2 Sample preparation and prefractionation

Perhaps the single most critical point in obtaining reproducible 2-D gels is sample preparation. The cells or tissue must be efficiently disrupted and the contents of the cells solubilized completely in order to obtain a representative protein population sample. Physical disruption methods such as sonication, mechanically driven rapid pressure changes, homogenization and shearing-based techniques, are often used to open the cells prior to protein extraction with a urea-based solution containing nonionic detergents, reducing agents and a protease inhibitor cocktail. For each cell or tissue type, a specific method must be developed, though recent developments including the use of thiourea [13], tributylphosphine [14], and novel zwittergents [15] have improved sample solubilization greatly. The extent of recovery of membrane and cytoskeletal proteins is variable, with some proteins being completely solubilized whilst up to 10% of the cell protein remains in the pellet after extraction.

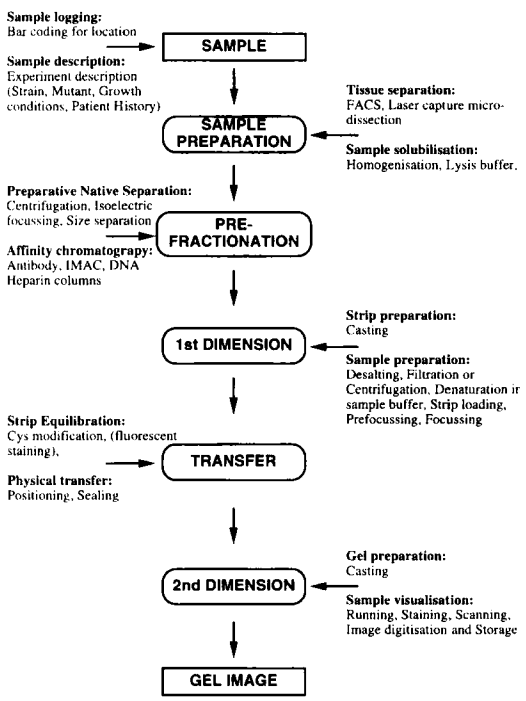


Figure 1. Module 1: Sample processing to image acquisition.

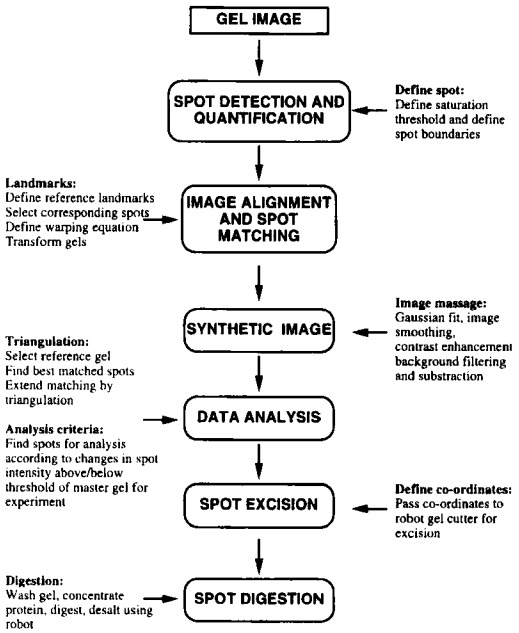


Figure 2. Module 2: Spot selection to protein digestion.

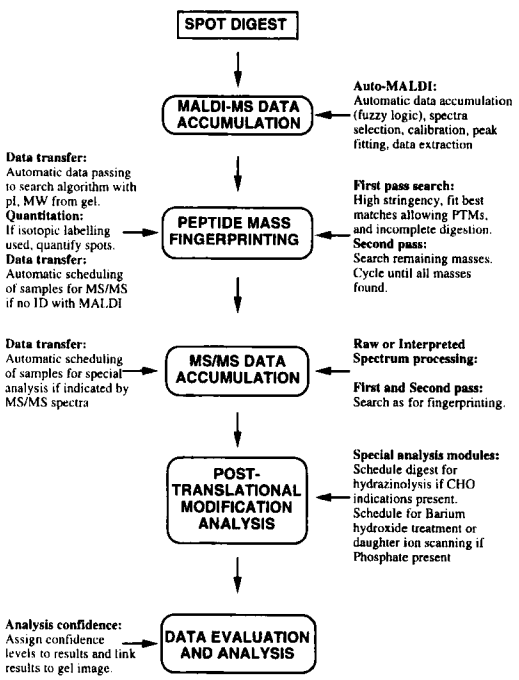


Figure 3. Module 3: Digest to data evaluation.

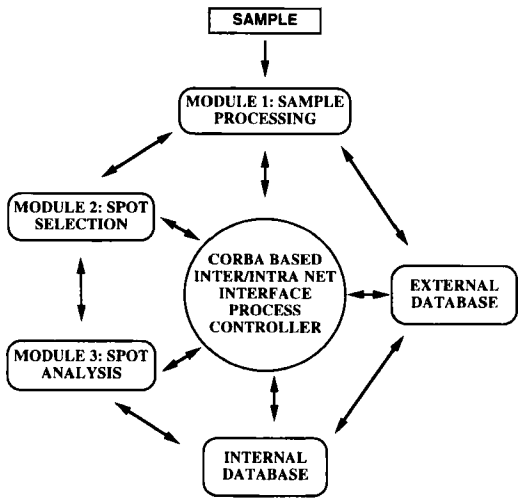


Figure 4. Module 4: Process integration and data inspection. The module is responsible for data accumulation, data analysis and data linkage. The control module must coordinate all other submodules, allowing sample logging and tracking, data passing between disimilar programs and integrating the data analysis output with the databases. The module should act as a user interface facilitating data viewing and evaluation and presenting significant links between data sets in an intuitive fashion; it should allow cross-correlation of MS data and 2-D patterns with a wide variety of other data such as patient files, genome database annotation, toxicity databases, gene knockout, and two-hybrid experiments, *etc.*

The extremely high degree of complexity of eukaryotic tissues often requires that a prefractionation step be carried out in order to reduce the complexity and allow the resolution and analysis of minor components. When dealing with a mixed cell population such as a tissue, prefractionation of cells using a fluorescence-activated cell sorter (FACS), [16] can allow small subpopulations to be specifically isolated, greatly increasing the sensitivity of the analysis. Similarly, preconcentration of the proteins to be analyzed can be carried out using methods orthogonal to 2-D gel separation, such as native PAGE or IEF [17], or by an affinity pre-enrichment, such as heparin chromatography for DNA-binding proteins [18], immobilized metal ion affinity chromatography for phosphoproteins [19] or by antibody precipitation to select for a specific protein complex. Alternatively, a series of increasingly powerful solubilizing buffers may be used to obtain a series of protein fractions [20], or the various cell compartments and/or organelles may be isolated [21]. All of the above methods can be automated and before any large-scale study is started, a systematic study of sample preparation reproducibility should be carried out.

3 2-D gel separation and analysis

3.1 Automating 2-D PAGE

The standard denaturing 2-D gel system has remained an essentially manual operation since its establishment over twenty years ago and as yet there is no valid alternative with an equivalent dynamic range and resolving power. A detailed study using a prototype semi-automated instrument clearly showed the advantages of mechanical reproducibility [22]. However, the only developments in large-scale automated production of 2-D gels has been carried out at commercial enterprises such as Large Scale Biology and Oxford Glycosciences. The largest commercially available 2-D gel system (Iso-Dalt, Amersham-Pharmacia) can run ten 2-D gels simultaneously. Steps are being taken to automate at least the first dimension. The IPG-Phor instrument from Amersham-Pharmacia automates sample loading and focusing. The optimization of reagents is another necessary step although the availability of well-defined IPG strips, in which the pH gradient is covalently immobilized and the gel is attached to a stable plastic support, has improved the reproducibility of 2-D patterns greatly, facilitating inter-laboratory comparisons. The use of tributylphosphine as the reducing agent in the first dimension has decreased the amount of streaking observed with mercaptoethanol and dithiothreitol, both of which become charged and migrate out of the gel. The transfer of the first-dimension strip to the second-dimension SDS-PAGE gel is much easier with the IPG strips mounted on plastic compared to the cylindrical IEF tube gels, since the former are mechanically much more stable. A critical step is the change from the first-dimensional to the second-dimensional running conditions. During this equilibration, proteins are prone to precipitation and extreme sample losses can be incurred, especially with membrane proteins. Ideally, an instrument could be designed in which no physical movement of the gels would be necessary between dimensions and the pore size of the first dimension could be large to allow rapid equilibration.

Another hurdle to be overcome in order to fully automate spot identification is that of the weak mechanical properties of the normal polyacrylamide gel matrix of the second dimension. The ideal matrix must be mechanically stable to allow the gel to be handled without breaking as well as to give a clear background during protein visualization. Mechanical stability can be increased by including prepolymerized polyacrylamide in the polymerization mixture, although this can produce a turbid background when scanning the gel. A gel mixture with enhanced properties has been partially described [23] and is commercially available. Alternatively, the gel can be covalently attached to a plastic support, which has the advantage of preventing alterations in gel size during the staining procedure (due to shrinkage in organic solvents and expansion upon rehydration or dehydration upon exposure to air), thus precluding computer-driven spot cutting. There are potentially many other polymer matrices which could be used with desirable properties such as solvent and pH resistance (essential for first-dimensional separations at extreme pH) and strength [24, 25] although only the cross-linker piperazine diacrylamide [26] has received any commercial attention. Since electroblotting large format 2-D gels is difficult and often nonquantitative due to the different transfer properties of the proteins and their binding affinity to the membrane, this will not be considered here, especially as it brings another handling step which must be automated.

3.2 Protein detection

Since polyacrylamide absorbs in the same UV range as proteins, proteins must be stained in order to be visualized. The staining method should be effective for all proteins in order to allow quantitation and should be compatible with further analysis steps. By far the most commonly used protocols are Coomassie blue and silver staining of proteins. Although both methods are sensitive to protein sequence (especially silver), they are both linear over a wide dynamic range for most proteins. Standard Coomassie blue-stained gels can detect proteins down into the subpicomole range while silver staining extends this to the low femtomole region. In our experience, colloidal Coomassie staining is the easiest to use, although silver staining, if closely monitored, can be very reproducible [27]. Staining of individual gels is easy to automate and commercial apparatuses are available. Staining a large number of gels is much more difficult and spot intensities vary by as much as 20% from batch to batch. Both staining methods require protein fixation (usually by precipitation with a high concentration of organic solvent or acids) which simultaneously removes SDS. This is convenient for subsequent analysis though it does render the proteins somewhat refractory to enzymatic digestion.

An alternative is the use of fluorescence labeling, either covalently introduced into the proteins during equilibration between the first and second dimensions or by using fluorescently labeled SDS substitutes or fluorescent dyes [28]. Again, linearity of response and general applicability are the limiting factors for most fluorescent dyes, although the extended sensitivity range down to the mid-attomole level is exceptional. None of the methods suffer from side reactions which disturb further analysis; even silver staining [29] shows little chemical modification. Gels are most commonly scanned by laser densitometers in the absorbance mode, which gives a linear response from 0 to

3.5 OD units. Detection methods based on labeling with radioactive isotopes are extremely sensitive (subattomole) but are not well-suited for automation due to the difficulty of correlating the positions of the spots on a film (or the computer printout of an image from a $BaFBr:Eu^{2+}$ storage screen) and on the corresponding gel matrix.

3.3 2-D gel analysis programs and automation

If all the steps from sample preparation to scanning were to be automated and the gels were mechanically stable, then gel analysis would become trivial. This not being the case, several generations of software packages have evolved from those primordial programs running on the DEC PDP 11 family of minicomputers (such as Elsie, LIPS, Gellab and Tycho), through the second generation of Unix-based programs (Gellab-II, Elsie-4, Kepler, Melanie, and Quest) to the third generation of Unix, NT and Mac programs (2D, BioImage, Melanie-II and Phoretix 2D). Currently, gel distortions must be mathematically treated and gel images warped and transformed to allow gel matching. The analysis is a multistage process involving spot detection, quantification, fitting and modeling, background subtraction, contrast enhancement and artifact removal, image alignment, and finally gel comparison. Although these processes can all be carried out automatically in batch mode (Melanie-II and BioImage), approximately 10% of the spots always remain and must be manually corrected and fitted due to the quality of the gel. Thus, although extremely complex mathematical corrections can be made, the limiting factor is the reproducibility of sample preparation, 2-D gel running and staining, and this should be the main focus of attention of automation rather than computational 'face-lifting'.

3.4 Limitations of 2-D gel separations

The advent of ultrasensitive MS techniques for protein identification has highlighted one of the shortcomings of wide pH range 2-D gel separations. Whilst this single gel format offers an immediate overview of the state of the proteome, many of the spots are inadequately separated. Partially overlapping spots can be resolved by computational approaches but now it is becoming clear that many symmetrical spots contain two or more proteins. For prokaryotes, about 20% of all spots contain one or more proteins [30]; for eukaryotes, the number approaches 40%. A collaborative study by Proteome Inc. and PerSeptive Biosystems [31] has shown in detail the extent and nature of spot cross-contamination in 2-D gels of yeast extracts. This was based solely on mass matching algorithms. Many of the spots contained three or more proteins. Although MS techniques are powerful enough to resolve the protein mixtures [32], multiple proteins per spot make interpretation of many experiments extremely difficult. This limitation can be eliminated almost totally, either by using very large format gels [33], or with a series of overlapping narrow pH range gels, a technique which has been termed 'Proteome Contigs' [34] by analogy to genome mapping.

4 Protein digestion

There are two approaches to analyzing the spots obtained from the 2-D separation: either a brute force analysis of all proteins can be undertaken, or a subtractive approach in which only proteins which show altered expression levels between the control and experimental state are chosen for analysis. Several approaches have been put forward for the total analysis approach. The 2-D gel may be gridded, cut into a series of 1 mm^3 cubes and each one digested and analyzed (H. Langen, unpublished); or the proteins may be electroblotted through a membrane which has a protease covalently bound to it and the resulting peptides are trapped on a second hydrophobic membrane (D. F. Hochstrasser, unpublished) for direct analysis in a scanning MALDI mass spectrometer. Alternatively, once the series of gels has been analyzed, the spots selected for further analysis may be excised by a robotics system for individual processing [35]. The Australian Proteome Analysis Facility (APAF) has developed such a system, in collaboration with a robotics company, ARRM (Advanced Rapid Robotics Manufacturing, Kent Town, South Australia), which is now commercially available. The robot excises spots, either from gels or PVDF membranes, and places them in a 96-well plate for automated proteolysis by a second robotics system (marketed by Canberra Packard, Downer's Grove, IL, USA) which subsequently loads the digests together with matrix onto a multi-sample MALDI target plate. The analysis of 288 protein spots from a single 2-D membrane blot was achieved within ten working days. Ideally, the 2-D analysis software should select the spots for analysis and download the spot coordinates to the robot for subsequent excision. This is simple to implement for PVDF membranes, but gels are more difficult since they deform easily, especially during spot excision, and they are prone to drying out. A prerequisite therefore is a mechanically stable gel as discussed previously. Other automatic digestion devices have been described [36, 37]. The performance of all such apparatuses are limited by sample loss to surfaces and the concentration limit imposed by the use of proteases.

4.1 Enzymatic digestion

The most commonly used exo- and endoproteases have a K_m in the 5–50 mM range, which means that they are

operating at 50% maximum velocity with a protein concentration of 5 pmol/μL, which is equivalent to a strong Coomassie-stained spot. A medium intensity silver-stained spot may contain 200 femtomoles of protein in a volume of 20 μL, 500 times less than the K_m; thus the proteolytic activity toward the substrate will be minimal, while autolysis rates will be relatively high. The correlation between the decrease in the number of peptides produced and protein concentration is well documented [38, 39] but little attention has been paid to the root cause. The critical concentration appears to be around 100 femtomoles/20 μL; below this, virtually no peptides are recoverd. The extreme sensitivity of MS compensates for this to a certain extent but also serves to mask the problem. A digest of 50 pmol/20 μL BSA produces around 60 main peptides and the % mole recovery is around 95%. In contrast, for 100 fmol/20 μL, only six peptides are seen, with a recovery of around 5%.

In order to obtain maximal yield and sequence coverage, the protein must be concentrated from the 2-D gel spot into a volume *ca.* 500 times smaller in order to return to the K_m range of endoproteases. Several methods of concentrating proteins from 2-D gel spots (originally intended to concentrate protein from multiple bands) have been described [40–42]. Spots from multiple gels can be concentrated into a final volume of 1–5 μL for digestion. The other problem of increasing sample loss at lower concentrations has been analyzed; a simple but elegant solution was proposed in which the peptides are collected by placing C18 or Poros RR beads into the digest medium, preventing loss and allowing sample cleanup [43]. We have recently constructed a device to concentrate protein together with protease into a final volume of 50 nL before trapping the peptides onto a hydrophobic membrane [39]. We are adapting this method to accomodate 100 spots excised from a 2-D gel in the format of a standard MALDI multi-sample target, thus eliminating all sample handling. Total automation, especially when carried out in the environment of a clean room, is an important step toward avoiding contamination, by keratins for example, especially as the sensitivity levels are increasing rapidly.

4.2 Chemical digestions

An alternative to enzymatic digestion is residue-specific chemical cleavage which does not suffer the drawback of a K_m-type limitation. The most useful and least prone to side reactions is cleavage by cyanogen bromide in acid solution, which is specific for methionine. The reaction can be carried out in the gas phase and does not produce contaminating peaks like those found with autolysis using proteases. Other reactions such as cysteine cleavage by 2-nitro-5-thiocyano-benzoic acid, tryptophan cleavage by BNPS-skatole (2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine), the cleavage of Asn-Gly bonds by hydroxylamine and acid cleavage of Asp-Pro bonds, are well documented but of little use due to the scarcity of the amino acids or bands and to side ractions. Recently the group of Tsugita [44] proposed several new site-specific chemical cleavage methods which can be carried out in the gas phase using proteins immobilized on PVDF membranes. Treatment with 0.2% pentafluoropropionic acid (PFPA) aqueous vapor at 90°C for 4–16 h lead to ε specific cleavage *C*-terminal to aspartic acid whereas exposure to S-ethyl trifluorothioacetate vapor at 50°C for 6–24 h leads to a cleavage *N*-terminal to serine and threonine [45]. Both methods showed little evidence of side reactions and could be carried out in the gas phase. These reactions are easy to automate and could be performed directly on a MALDI target.

5 Mass spectrometry: protein identification using MS data

In order to achieve high throughput protein identification, a hierarchical approach must be taken. The first level is very rapid, low-information content data accumulation to identify the bulk of the proteins, followed by a second level of lower throughput, high-information content data generation to define the rest. Finally, a third level of labor-intensive data accumulation can be carried out to define post-translational modifications. In practical terms, MALDI-based peptide mass fingerprinting is the ideal first step, followed by MS/MS peptide fragment fingerprinting as the second.

5.1 Peptide mass fingerprinting

In 1993, five groups independently proposed the idea of peptide mass fingerprinting [46–50]. The concept is that the set of peptide masses obtained by mass spectrometric analysis of a digestion of a protein with a specific protease can act as a fingerprint, and this property is unique to that protein. Therefore the set of masses can be used to search a protein database in which the sequences have been replaced by the calculated fingerprints to find a similar pattern.

5.1.1 Data accumulation

Peptide mass fingerprinting can easily be adapted to a high-throughput format. Protein spots automatically excised from 2-D gels can be destained (50 mM potassium ferricyanide and 50 mM sodium thiosulphate for silver), washed, dehydrated and then perfused with trypsin using a standard fluid delivery robot station. Furthermore, the resultant digest can be automatically loaded onto a

MALDI sample target together with matrix. Protocols for high throughput analysis have been described [51] for in-gel digestion in a low-salt, nonvolatile buffer allowing the analysis of hundreds of proteins per day. Most MALDI mass spectrometers now come with sample plates capable of holding 100–10 000 samples. In order to process such a high sample volume, data accumulation must be monitored on-line to ensure a reasonable quality level. Fuzzy logic feedback control [52] and other algorithms have been implemented which allow sequential sample measurement. The spot is searched until a reasonable signal is found and then the laser power is optimized (reduced to the minimum possible) to obtain high resolution spectra before a spectrum is accumulated and written to disk and the process moves on to the next sample position.

5.1.2 Data extraction

The next stage is the extraction of the peptide fingerprint data from the mass spectrum. The first step is to identify all the peaks in the spectrum and to calibrate the spectrum using predefined internal standards (either added externally or, more often, using autolytic fragments of the protease used). The spectrum must contain enough data points for a reasonable curve fitting to allow peak identification, although for weak peaks and low numbers of data points, a Maximum Entropy™ algorithm can be used to match a theoretical isotope pattern to the experimental data. The ideal MALDI-TOF-MS for high throughput peptide fingerprinting should be capable of very rapid > 4 Ghz data capture with at least 0.1 ns resolution and possess an ion gate which removes all slow moving ions > 5000 *m/z* from the flight tube to allow a very high laser pulse rate. High S/N ratios could be achieved with more data points over the peaks. Once peaks are fitted, the monoisotopic peak must be identified and extracted for the fingerprint file. Several MS firms have now implemented such data extraction procedures which are carried out automatically after data accumulation and the fingerprint file can be submitted to the database searching program.

5.1.3 Mass accuracy

The effect of mass accuracy was shown early on to be a critical factor for the confidence level of peptide mass searches [47]. Mass errors can now be kept at 30 ppm or lower using 'natural' internal standards (such as known tryptic autolytic fragments) over the range of 800–3000 *m/z*. The use of nitrocellulose [53] as a substratum onto which the matrix can be added allows much more uniform crystallization and peptide signals are located over the entire target spot. The sample can also be washed with ice-cold water to lower the effect of salt contamination. The development of delayed extraction MALDI-MS, in which the ions are not extracted immediately after the laser impulse but are allowed 100–300 nanoseconds to equilibrate before the ions ae accelerated, allows measurement to within 5 ppm. If an initial search is carried out, keeping the mass error search parameter fixed at 30 ppm or less and partial digest products are ignored, a single high scoring match is usually obtained. After matching as many of the remaining masses as possible to the top scoring sequence (allowing incomplete digestion, defined modifications and lowering the mass accuracy to 80 ppm for example, using the program PepIdent and FindMod, http://expasy.hcuge.ch [54]), a "second pass search" can automatically be carried out to identify the next component in the mixture [55]. A related approach applies Bayes' Theorem to test the hypothesis that a protein matches a database entry (W. Profound *et al.*, Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics, Atlanta, Georgia, 1995). The probability of a match being correct is calculated taking into account the MS data and background information (p*I* intact M_r, *etc.*) and the prior probability for the hypothesis given the background information only, to calculate the likelihood probability for the data given the hypothesis and the background information. One of the limiting factors when using peptide mass searching to identify proteins in a sequence database is the presence of very large proteins. These tend to dominate the high scoring positions when the mass accuracy of the search or the number of masses used is low.

5.1.4 Orthogonal data

One of the fundamental problems of database searching is how to determine the confidence level of a search result. In order to resolve this, additional parameters can be included, such as a molecular weight estimate, limiting the number of mismatches allowed, or using a scoring system weighted according to the frequency of occurrence of a mass in the protein of a given mass range. These restrictions are not readily applicable to sequences derived from genomic or EST data since only partial or fragmented sequences (due to introns or reading-frame shifts) are represented and the accuracy of EST sequences is lower since they are obtained by single-pass DNA sequencing. The random noise element in the database search can be minimized by using two or more orthogonal data sets [56]. Data from two protease digests with differing specificities (*e.g.*, AspN and LysC) can be combined. Alternatively, a single digest can be measured in the native state and again after carrying out a chemical modification on the sample plate (such as methylation, acetyla-

tion or deuterium exchange). This greatly increases the confidence level but doubles the analysis time.

5.2 Peptide tagging

5.2.1 Ragged termini

An alternative to using dual digestions or native and chemically modified digests is to exploit the sequence information that is already present in a simple peptide fingerprint. Trypsin cuts at Lys and Arg, two of the most abundant amino acids in proteins. Consequently there are a comparatively large number of repetitive sequences such as KK, RR, KKR, *etc.*, in proteins. Since trypsin cleaves randomly at these sites and has an intrinsically low exoprotease activity, ragged *C*-termini are produced. On average, 2.5 such sequence tags are found per peptide mass spectrum [57]. These sequence patterns are also found with other proteases such as AspN, LysC, Glu(Asp)C and can easily be extracted and used for protein database searches. Two *N*- or *C*-terminal sequence tags were reported to be enough to unambiguously identify a protein in most cases. Ragged *N*- or *C*-termini can be efficiently generated by sequential endo- and exopeptidase digestions and the extracted data used to search sequence databases [58]. The confidence level of protein identification is much higher than simple fingerprinting alone; however, the method suffers the same drawback as enzymatic digestion, namely the effectiveness of the exoprotease activity drops rapidly with decreasing peptide concentrations and increasing numbers of peptides in a digest.

5.2.2 Chemical tag generation

As with enzymatic digestion, the alternative to exo- or endopeptidase ragged termini generation is the use of sequential chemical degradation steps. The use of a single-step manual Edman degradation of peptide mixtures to generate a single *N*-terminal amino acid tag has been shown to be effective [57] but the derivatization causes a loss in sensitivity by a factor of approximately 10 and the washing step can cause extensive sample loss. Essentially, the method is an abbreviated form of 'ladder sequencing' in which automated Edman degradation using phenylisothiocyanate was carried out in the presence of a low amount of sequence terminator (phenylthiocyanate) and the resulting ladder of peptide fragments was analyzed by MALDI-MS [59]. A modified form of this procedure without a chain terminator generates the peptide ladder by adding equal aliquots of starting peptide each cycle and driving both the coupling and cleavage reactions to completion [60]. The main drawback is the extra number of handling steps though the retention of the peptide terminal amine allows for subsequent modification with quaternary ammonium alkyl NHS esters to improve sensitivity. We have recently developed a thioester-based degradation that can be carried out using aqueous reagents on peptides immobilized on C-18 reverse-phase beads immobilized in a Teflon membrane. The procedure can be carried out on one hundred protein digests simultaneously on a single membrane fixed in a MALDI target which can be inserted directly into a mass spectrometer for analysis.

6 Mass spectrometry: protein identification using MS data

6.1 MS/MS data acquisition

Conceptually similar approaches to protein identification by peptide mass fingerprinting using the MS/MS fragmentation pattern from a single peptide have been proposed [61, 62]. Tandem mass spectrometry of peptides was pioneered in the 1980s by D. Hunt *et al.* [63] for low energy, triple quadrupole instrumentation and by K. Biemann [64] for high-energy, four-sector magnetic instruments. The first mass scanning stage is used to isolate a single peptide before acceleration of the ions through a region of higher pressure containing a collison gas such as argon. The second mass scanning stage is used to measure the masses of the fragments arising from the peptide as a result of collisionally activated dissociation (CAD). Post-source decay (PSD) spectra can be obtained using TOF instruments equipped with ion gates and mirrors by increasing the laser power by a factor of two over that needed to obtain a normal spectrum. A large fraction of the desorbed ions obtained by MALDI undergo 'delayed' fragmentation before reaching the detector as a result of multiple collisions of the peptides with the matrix during plume expansion and ion acceleration. The set of ions produced by either technique can be used to determine the sequence of the peptide. The set of fragment masses acts as a fingerprint for the peptide and can be used to search sequence databases for similar peptides. PSD analysis should be amenable to automation, and curved field reflectrons which allow an entire fragmentation spectrum to be accumulated in one scan have made it simpler to carry out. However, it suffers from the disadvantage that PSD spectra are difficult to obtain from peptide mixtures and some form of fractionation should be carried out prior to analysis [65].

6.1.1 Static *vs.* dynamic nanospray

The accumulation of CAD-MS/MS fragmentation data is somewhat more time-consuming than simple peptide fingerprinting, though the data obtained has a much higher information content. Essentially two approaches to obtaining high sensitivity MS/MS data can be taken using an

electrospray interface to a triple quadrupole, ion trap or quadrupole-TOF mass spectrometer: either the digest is sprayed as a mixture directly from a capillary with a 1–10 μM exit tip at very low flow rates (< 50 nL/min) [66–70], or *via* a nanospray on-line separation method such as CZE or HPLC. The peptide mixture can then be analyzed automatically using a program to pick out the ions for MS/MS, which has the advantage that certain impurities such as autolytic fragments and gel-derived contaminants can be filtered out using a look-up table [71]. This approach has been successfully used to sequence entire proteins [72]. The dynamic CZE/HPLC approach is much more sensitive than the static nanospray method since the peptides are more concentrated, by a factor of 50 or more [73]. Both methods have the advantage that they can be combined with parent ion scanning. Scans for the parents of common fragmentation products of peptides, such as immonium or phosphate ions, allow determination of peptide ion masses even when these ions have signal-to-noise ratios of one in the mass spectrum [74]. However, the greatest advantage of the dynamic methods is that they can easily be interfaced with an autosampler allowing automated sample clean-up.

6.1.2 CZE- and HPLC-MS

Two approaches are being taken to combining an on-line peptide separation with automatic MS/MS data accumulation, capillary zone electrophoresis and high performance liquid chromatography. A microscale, variable flow LC-MS setup has been described which allows complex peptide mixtures to be analyzed in a fully automated way. The HPLC is run at 200 nL/min and the eluting peptides are monitored by the MS. Once a peak of interest elutes, the flow is dropped immediately to approximately 2 nL/min, allowing the MS/MS accumulation to be extended from 15 s (the peak width from the column at high flow) to 10 min [73]. During this time, the MS spectra can be automatically optimized by varying the collision energy and averaging until the S/N is above a specified level. The second advantage is that the S/N increases by a factor of 30 or more as the flow rate decreases. This can be combined with a series of scans known as a 'triple-play' on the Finnigan LCQ ion trap, during which a full MS scan is followed by a high resolution scan over the parent ion of interest before the MS/MS scan. After optimization [75] the peak parking can give sensitivities in the < 10 femtomole range. The disadvantage of this method is that a normal 30 min HPLC run can be extended to over 6 h, though the amount and quality of the data obtained is excellent.

The alternative approach of CZE-MS was limited for a long time by the volume of liquid that could be loaded onto a column for separation. Two groups have described the development of on-line hydrophobic membrane [76] or a miniature C-18 reverse-phase column [77] as a preconcentration device for CZE-MS analysis of peptide mixtures. Large sample volumes (> 100 μL) can be loaded onto the membrane and washed before being eluted with a small (< 10 nL) plug of organic solvent at the start of the CZE separation. Detection limits are in the mid-attomole range for peptide mass measurement, and < 10 femtomole for collision-induced dissociation. An analogous technique to HPLC 'peak parking' can be carried out using CZE, which has been called 'reduced elution speed CE' [78]. In this case, the voltage can be dropped, or even reversed, to slow the osmotic flow direction and adjust the speed of elution. Currently, CZE- and HPLC-auto-MS/MS offer the most advanced automated protein identification system possible, and have been combined with autosamplers and automatic database searching.

6.2 MS/MS database searching

6.2.1 Raw data searches – Sequest

Algorithms for searching sequence databases using uninterpreted MS/MS spectra have been developed over the past four years. The most widely used of these is Sequest, developed by J. Yates and J. Eng at the University of Washington, Seattle [61]. The program was originally intended for searching protein databases using MS/MS fragmentation spectra of unmodified peptides but has been subsequently extended to allow searching with post-translationally modified peptides [79], DNA database searching [80], PSD and high energy CAD or PSD data [81, 82]. Essentially, the program searches for all peptides in the database (a protease can, but does not have to be defined) which have the same mass as the parent ion (within a defined mass window) and then matches the predicted MS/MS spectrum with the experimentally determined one, and finally carries out a cross-correlation analysis of the best scoring peptides in order to determine the best match. The program can automatically strip all the MS/MS data from an HPLC-auto-MS/MS data file, search the databases and then produce a summary of all the individual search results. The power of this approach lies in the ability to deal with proteins present as mixtures and molar ratios of 30:1 even at the low femtomole levels [83]. Several other programs have now been described, which are available on the web, that can use uninterpreted spectra, such Fragfit – part of the Prowl software suite (http://prowl.rockefeller.edu/) – at the Rockefeller Institute, New York, and MS-Tag – part of the ProteinProspector suite (http://prospector.ucsf.edu/) – at the University of California, San Francisco. These programs take advantage of another source of information present in fragmentation

spectra besides the sequence ions, the presence or absence of immonium ions which are indicative of certain amino acids.

6.2.2 Partial MS/MS interpretation – tag searching

Another algorithm for protein identification in sequence databases, PeptideSearch, was developed by M. Mann [62]. The MS/MS spectrum must be manually inspected to find a group of ions which form a series from which a small sequence (the tag) can be read and used with the intact peptide mass, the mass from the *N*-terminal to the start of the tag sequence, and the mass from the end to the tag to the *C*-terminal, to search the databases. Since the search can be carried out using only the tag and the *N*-terminal mass, for example, the algorithm can identify peptides carrying undefined post-translational modifications. The program can also carry out mixed searches, for example, a tag search with a peptide mass search, and then combine the results. Other programs have been developed which combine various elements of Sequest and PeptideSearch such as MassFrag [65], No-name [84] and Fragfit [32].

6.2.3 Automated MS/MS interpretation and database searching

The original approach to protein identification, the manual interpretation of an MS/MS spectrum to obtain a peptide sequence for use in a homology search such as BLASTA or TFASTA [85, 86] is receiving renewed attention. Early attempts at automated MS/MS spectra interpretation used a partial correlation method to fit increasingly longer sequences to a spectrum (such as SEQPEP [87]) or used a pattern matching approach such as that described by Hines [88]. Recently a program was developed by Lutkefisk [89] which uses the set of possible sequence interpretations of an MS/MS spectrum generated by SEQPEP to search sequence databases using a modified FASTA approach. An alternative to this method is to generate sequence ions in the MS/MS spectrum that contain an isotopic signature that allows them to be immediately identified as *N*-terminally derived *b* or *C*-terminal *y* series ions. Digestion of proteins in 50:50 $^{18}O/^{16}O$-labeled water produces peptides which appear as doublets in the mass spectrum with a spacing of two mass units between the peaks. This is due to the introduction of a water molecule from the solvent during the hydrolysis of the peptide bond [90]. The *b* ions in the MS/MS spectrum thus appear as a singlet whilst the *y* ions appear as doublets. This has been used as the base for automated *de novo* sequence tag extraction for database searching using a high-resolution quadrupole/time-of-flight mass spectrometer [91]. Several alternative chemical approaches have been used to specifically label the *N*-terminus of a peptide with an isotopic label though this has been less successful due to a certain lack of discrimination between alpha (*N*-terminal) and epsilon (lysine) amino groups and to the drop in detection sensitivity incurred by removing a basic site.

7 Data evaluation and databases

7.1 Large-scale protein analysis

Recent advances in protein identification using mass spectrometric data to search sequence databases have allowed a broader approach to be taken to biological problem solving. The 'proof of principle', that MS data can be successfully combined with large-scale 2-D-based proteome analysis, was provided by the group of M. Mann [92] and J. Yates [30] who reported the analysis of 150 spots from two-dimensional gels of yeast and 303 spots from *Haemophilus influenzae*, respectively. Up to 90% of the yeast proteins could be identified by high accuracy peptide mass fingerprinting. However, this method tends to underestimate the number of proteins present in a 2-D gel spot. Automated MS/MS and database searching with uninterpreted peptide fragmentation spectra showed that at least 10% of the prokaryotic spots contained two or more proteins using a broad pH range gel (4–8), increasing to as much as 50% for eukaryotic organisms. Currently 'only' fifty spots can be analyzed in one day using an autosampler and reverse-phase chromatography/automated electrospray tandem mass spectrometry [93]. The MS-based methods can be useful even when a complete genome sequence is not available as seen by the characterization of the human multi-protein spliceosome complex using EST-database searching [94].

7.2 Data interpretation tools

The automation process does not stop once the database searches have been carried out. High throughput protein analysis requires the development of software tools to aid data interpretation to present meaningful trends and conclusions to the experimentalist. In the case of HPLC or CZE-autoMS/MS runs, the resulting data files are enormous and not much can be obtained from manual viewing. In order to compare various files, powerful data analysis software packages are required. One of the earliest programs, which was widely used and user-friendly, was MacProMass [95] developed in T. Lee's laboratory for interpreting mass spectral data obtained from MS protein analysis. A second generation program, Sherpa, was developed at the University of Washington, Seattle by A. Taylor [96]. The program can search concurrently against multiple protein sequences, compare two LC/MS files concurrently, search for glycosylated and phosphory-

lated peptides, and give a simple evaluation of the quality of a match between the data and a prediction. Such tools are invaluable aids to analysis. A variant of the Sequest program was recently described by J. Yates [97] to compare collision-induced dissociation spectra of peptides for library searching and subtractive analysis of tandem mass spectra obtained during LC/MS/MS experiments.

Since only few laboratories have the resources to develop such tools, the trend now is to make software available *via* the WWW, either as downloadable programs, transient JAVA scripts or as on-line services. Database searches can be carried out using HTTL data input forms from a distant client and small JAVA scripts can be downloaded to clients for data analysis. For more demanding data analysis, helper programs can be made available and downloaded from the website. An interesting example of such a project has been published [98] describing the construction of the Prowl protein analysis tool kit. Access to sequence database and search machines such as BLAST are now common over the web. Slowly the next generation of tools for large-scale data interpretation are appearing, for example, the Swiss Institute of Bioinformatics Expasy site in Geneva which provides many services for proteome analysis as well as being the home of annotated databases such as SWISS-PROT and ProSite (http://expasy.hcuge.ch/). In order to be useful, such data must be able to be shared between laboratories. Software must be developed which can access databases and interact with nonrelated query programs in a highly heterogeneous computer network environment. CORBA (Common Object Request Broker Architecture) applications are being applied in the bioinformatics area to enhance utilization, management, and interoperation between biological resources [99]. In the case of proteomics this means the processing and comparison of 2-D images should be made possible over the web. Both JAVA and CORBA tools are now being developed, for example, Make2ddb to generate a web-based Proteome map [100], Flicker, a program to allow the comparison of gels over the web [101], *etc.* A highly laudable data integration interface has been developed at the Institute for Chemical Research, Kyoto University, Japan, the Kyoto Encyclopaedia of Genes and Genomes (KEGG; http://www.genome.ad.jp/kegg/). The website provides a graphical interface with which one can explore metabolic and regulatory pathways that consist of interacting molecules or genes and it provides links from the gene catalogues produced by genome sequencing projects.

7.3 Database management

Genomes are arriving in the databases almost monthly now and forty are scheduled to be finished by the end of the decade. The large amount of information brought by such analyses brings with it new problems: how to bring some order to the data, especially with the large number of genes of unknown function. Computer programs can be used to extract possible open reading frames (ORFs) and match them against current sequence databases to find homologous proteins with known functions. However, this is a perilous undertaking; for example, let us assume that a gene XYZ in rat was cloned after its isolation and characterization as a sodium pumping ATPase. The computer then finds ORF 1 from organism A is 70% homologousto this gene and annotates it as a sodium pump homolog. This annotation is propagated by computer analysis to ORF 2,311 of organism B, which is to be 65% homologous to an ORF1 from organism A. As the number of genome sequences increases, this eventually leads to the annotation of ORF X in organism J as a sodium pump although it has no homology to the only characterized gene product, that of rat gene XYZ. Thus, databases annotation must be carefully monitored, and projects like SWISS-PROT, run by A. Bairoch in Geneva, Switzerland, are becoming increasingly important in preventing databases from drowning in computer speculation.

7.4 Overall data management and process control

Most of the basic components of automated high throughput proteomics already exist. What is needed is a standardized format for data acquistion, storage, and querying (interpretation). When dealing with such large-scale projects, tasks such as tracking samples, linking samples to gels, gels to spots and spots to sequence database entries become overwhelming and software is urgently needed to prevent chaos. The data generated in such projects is vast; 2-D gel images occupy around 4 MB each, an HPLC-MS/MS run occupies up to 40 MB, and sequence databases are expanding continuously and require around 400 MB. In order to deal with such vast amounts, parallel computing must become much cheaper and user-friendly for programing [102]. Data compression without losses must be implemented to speed retrieval and searching [103]. Many of these tools are being developed but, unfortunately, only within companies, making the platforms inaccessible to the academic researcher. A commercial platform which allows integration of home-built tools and data sharing on a common format between labs would be highly appreciated.

8 Future trends

As yet, no viable alternative has been put forward to 2-D PAGE. Many new developments in the technology were described at the Siena 2-D electrophoresis meeting in

1998. The research firm Large Scale Biology (Rockville, MD, USA) has developed instrumentation to allow one hundred gels a week to be run in an upscaled version of the famous Iso-Dalt format. Proteome Inc. (Beverley, MA, USA) have developed a fully automated instrument, the Proteomatron, on which six large-scale (40 × 40 cm) gels can be cast, run, and stained without handling. The sensitivity of protein detection is also increasing: multiphoto detection (MPD) imagers capable of detecting 10^{-22} molecules using radiolabeling have been developed by BioTraces (Fairfax, VA, USA). Moreover, two experimental conditions can be run on the same gel if they are labeled with different isotopes such as ^{125}I and ^{131}I. Since the isotopes can be distinguished, two images can be obtained and the differential protein expression visualized. This level of sensitivity, which allows tiny amounts of material to be analyzed, could be combined with the recently developed technique of laser capture microdissection (commercially available from Arcturus Engineering) which can excise selected populations of cells or areas from tissue slices to produce proteome maps of specific anatomical regions. Another application of isotopic labeling is the use of metabolic labeling using amino acids containing a specific isotope of an element such as ^{15}N. This can be used to accurately quantitate the rate of protein synthesis or degradation of a protein by comparing the ratio of normal and isotopically labeled peptide peaks, or quantification of the absolute amount if a standard amount of unlabeled material is mixed with the sample.

Mass spectrometric analysis of proteins directly from 2-D PAGE is being actively explored. Intact protein mass measurements have been reported using an infrared laser to desorb and ionize samples directly from PVDF blots [104]. The direct UV laser desorption of proteins directly from the first-dimension IPG strip has been described and suggested as an alternative second dimension to mass separation by 2-D PAGE. The main disadvantage is that proteins can rarely be confidently identified on the basis of mass and estimated *p*I, especially those from eucaryotic organisms. Indications are appearing, however, that fragmentation of intact proteins in the MS can be obtained by either delayed extraction MALDI-TOF or by Fourier transform ion cyclotron MS to give enough 'sequence tag' information to identify the protein [105–107]. Transblotting, the electroblotting of an entire 2-D PAGE gel through a membrane covalently modified with a protease onto a PVDF membrane, has been proposed as part of a molecular scanner project (D. F. Hochstrasser, unpublished). The PVDF membrane can then be coated with matrix and scanned with a MALDI-MS to reconstruct a 2-D gel image from the peptide ion current intensity. The peptide masses found within a spot contour can then be automatically sent for a mass-fingerprinting search and a 2-D image with protein identification markers is generated.

A new field of research that is beginning to have an impact on proteome research is nanotechnology. The field is advancing so fast that realistic alternatives to 2-D PAGE could emerge; the combination of CZE with a hydrophobicity-based separation, for example, is under development. Chip technology is already making its presence felt and a solid-phase extraction-CZE device has been constructed and interfaced to a standard mass spectrometer, allowing extremely sensitive autoMS/MS analyses [108]. The advantage of these devices is that they are cheap to manufacture in bulk and can be made disposable, ensuring no cross-contamination of samples. Also, multiple samples can be loaded by a robotic pipeter to wells which can be individually addressed. Karger's group in Boston has developed a microdevice with an integrated array of channels and electrospray tips for high throughput ESI/MS analysis. The sample wells on the microdevice are arranged in the standard 96-well plate format, and each well is connected by a microchannel to an independent ESI tip. A computer-controlled mounting stage is used to position each electrospray tip in front of the sample orifice of the mass spectrometer. Analysis times of less than 10 s per sample can be achieved.

The ultimate aim of proteomics, however, must be the integration with other related fields. The combination of gene expression and proteome studies (including post-translational modifications), lipid and oligosaccharide compositions and fluxes through metabolic pathways can ultimately be combined into complex databases which will allow whole organisms to be compared (which is especially useful for targeting differences between pathogenic and nonpathogenic bacteria for drug development). The uses of these databases are unlimited, ranging from molecular scanners for diagnostics, drug discovery, dissecting signaling and metabolic pathways, as well as in gene function studies. The bottleneck for many of the future developments in proteomics will be bioinformatics and the need for parallel computers to deal with such huge amounts of data.

Received January 6, 1999

9 References

[1] Devine, K. M., Wolfe, K., *Trends Genet.* 1995, *11*, 429–431.

[2] Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Fields, C., Venter, J. C., *Nature* 1992, *355*, 632–634.

[3] Schena, M., Shalon, D., Davis, R. W., Brown, P. O., *Science* 1995, *270*, 467–470.

[4] Liang, P., Pardee, A. B., *Science* 1992, *257*, 967–971.

[5] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, *270*, 484–487.

[6] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett Jr., D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[7] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., Davis, R. W., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 13057–13062.

[8] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[9] Haynes, P. A., Gygi, S. P., Figeys, D., Aebersold, R., *Electrophoresis* 1998, *19*, 1862–1871.

[10] Klose, J., *Humangenetik* 1975, *26*, 231–243.

[11] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250* 4007–4021.

[12] O'Farrell, P. Z., Goodman, H. M., O'Farrell, P. H., *Cell* 1977, *12*, 1133–1142.

[13] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[14] Herbert, B. R., Molloy, M. P., Gooley, A. A., Walsh, B. J., Bryson, W. G., Williams, K. L., *Electrophoresis* 1998, *19*, 845–851.

[15] Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1906.

[16] Madsen, P. S., Hokland, M., Ellegaard, J., Hokland, P., Ratz, G. P., Celis, A., Celis, J. E., *Leukemia* 1988, *2*, 602–615.

[17] Corthals, G. L., Molloy, M. P., Herbert, B. R., Williams, K. L., Gooley, A. A., *Electrophoresis* 1997, *18*, 317–323.

[18] Fountoulakis, M., Langen, H., Evers, S., Gray, C., Takacs, B., *Electrophoresis* 1997, *18*, 1193–1202.

[19] Porath, J., Carlsson, J., Olsson, I., Belfrage, G., *Nature* 1975, *258*, 598–599.

[20] Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J.-C., Hochstrasser, D. F., Williams, K. L., Gooley, A. A., *Electrophoresis* 1998, *19*, 837–844.

[21] Anderson, N. G., *Methods Biochem. Anal.* 1967, *15*, 271–310.

[22] Harrington, M. G., Lee, K. H., Yun, M., Zewert, T., Bailey, J. E., Hood, L., *Appl. Theor. Electrophor.* 1993, *3*, 347–353.

[23] Patton, W. F., Lopez, M. F., Barry, P., Skea, W. M., *BioTechniques* 1992, *12*, 580–585.

[24] Righetti, P. G., Chiari, M., Casale, E., Chiesa, C., Jain, T., Shorr, R., *J. Biochem. Biophys. Methods* 1989, *19*, 37–49.

[25] Harrington, M. G., Lee, K. H., Bailey, J. E., Hood, L. E., *Electrophoresis* 1994, *15*, 187–194.

[26] Hochstrasser, D. F., Harrington, M. G., Hochstrasser, A. C., Miller, M. J., Merril, C. R., *Anal. Biochem.* 1988, *173*, 424–435.

[27] Rodriguez, L. V., Gersten, D. M., Ramagli, L. S., Johnston, D. A., *Electrophoresis* 1993, *14*, 628–637.

[28] Steinberg, T. H., Haugland, R. P., Singer, V. L., *Anal. Biochem.* 1996, *239*, 238–245.

[29] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[30] Link, A. J., Hays, L. G., Carmack, E. B., Yates III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[31] Parker, K. C., Garrels, J. I., Hines, W., Butler, E. M., McKee, A. H., Patterson, D., Martin, S., *Electrophoresis* 1998, *19*, 1920–1932.

[32] Arnott, D., Henzel, W. J., Stults, J. T., *Electrophoresis* 1998, *19*, 968–980.

[33] Voris, B. P., Young, D. A., *Anal. Biochem.* 1980, *104*, 478–484.

[34] Wasinger, V. C., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[35] Walsh, B. J., Molloy, M. P., Williams, K. L., *Electrophoresis* 1998, *19*, 1883–1890.

[36] Davis, M. T., Lee, T. D., Ronk, M., Hefta, S. A., *Anal. Biochem.* 1995, *224*, 235–244.

[37] Houthaeve, T., Gausepohl, H., Ashman, K., Nillson, T., Mann, M., *J. Protein Chem.* 1997, *16*, 343–348.

[38] Courchesne, P. L., Luethy, R., Patterson, S. D., *Electrophoresis* 1997, *18*, 369–381.

[39] Staudenmann, W., Hatt, P. D., Hoving, S., Lehmann, A., Kertesz, M., James, P., *Electrophoresis* 1998, *19*, 901–908.

[40] Lombard-Platet, G., Jalinot, P., *BioTechniques* 1993, *15*, 668–670, 672.

[41] Rider, M. H., Puype, M., Van Damme, J., Gevaert, K., De Boeck, S., D'Alayer, J., Rasmussen, H. H., Celis, J. E., Vandekerckhove, J., *Eur. J. Biochem.* 1995, *230*, 258–265.

[42] Dainese Hatt, P., Quadroni, M., Staudenmann, W., James, P., *Eur. J. Biochem.* 1997, *246*, 336–343.

[43] Gevaert, K., Demol, H., Sklyarova, T., Vandekerckhove, J., Houthaeve, T., Electrophoresis 1998, *19*, 909–917.

[44] Kawakami, T., Kamo, M., Takamoto, K., Miyazaki, K., Chow, L. P., Ueno, Y., Tsugita, A., *J. Biochem. (Tokyo)* 1997, *121*, 68–76.

[45] Kamo, M., Tsugita, A., *Eur. J. Biochem.* 1998, *255*, 162–171.

[46] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

[47] James, P., Quadroni, M., Carafoli, E., Gonnet, G., *Biochem. Biophsy. Res. Commun.* 1993, *195*, 58–64.

[48] Mann, M., Hojrup, P., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 338–345.

[49] Pappin, D. J. C., Hojrup, P., Bleasby, A. J., *Curr. Biol.* 1993, *3*, 327–332.

[50] Yates III, J. R., Speicher, S., Griffin, P. R., Hunkapiller, T., *Anal. Biochem.* 1993, *214*, 397–408.

[51] Fountoulakis, M., Langen, H., *Anal. Biochem.* 1997, *250*, 153–156.

[52] Jensen, O. N., Mortensen, P., Vorm, O., Mann, M., *Anal. Chem.* 1997, *69*, 1706–1714.

[53] Vorm, O., Mann, M., *J. Am. Soc. Mass Spectrom.* 1994, *5*, 955–958.

[54] Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Hochstrasser, D. F., Appel, R. D., *Electrophoresis* 1997, *18*, 403–408.

[55] Jensen, O. N., Podtelejnikov, A. V., Mann, M., *Anal. Chem.* 1997, *69*, 4741–4750.

[56] James, P., Quadroni, M., Carafoli, E., Gonnet, G., *Protein Sci.* 1994, *3*, 1347–1350.

[57] Jensen, O. N., Vorm, O., Mann, M., *Electrophoresis* 1996, *17*, 938–944.

[58] Korostensky, C., Staudenmann, W., Dainese, P., Hoving, S., Gonnet, G., James, P., *Electrophoresis* 1998, *19*, 1933–1940.

[59] Chait, B. T., Wang, R., Beavis, R. C., Kent, S. B., *Science* 1993, *262*, 89–92.

[60] Bartlet-Jones, M., Jeffery, W. A., Hansen, H. F., Pappin, D. J., *Rapid Commun. Mass Spectrom.* 1994, *8*, 737–742.

[61] Eng, J. K., McCormack, A. L., Yates III, J. R., *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989.

[62] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.

[63] Hunt, D. F., Yates III, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., *Proc. Natl. Acad. Sci. USA* 1986, *83*, 6233–6237.

[64] Biemann, K., *Methods Enzymol.* 1990, *193*, 455–479.

[65] Gevaert, K., Verschelde, J. L., Puype, M., Van Damme, J., Goethals, M., De Boeck, S., Vandekerckhove, J., *Electrophoresis* 1996, *17*, 918–924.

[66] Wahl, J. H., Goodlett, D. R., Udseth, H. R., Smith, R. D., *Electrophoresis* 1993, *14*, 448–457.

[67] Anderegg, R. J., Wagner, P. E., Emmet, M. R., Caprioli, R. M., *J. Am. Mass Spectrom.* 1994, *5*, 867–869.

[68] Gale, D. C., Smith, R. D., *Rapid Comm. Mass Spectrom.* 1993, *7*, 1017–1021.

[69] Wilm, M., Mann, M., *Anal. Chem.* 1996, *68*, 1–8.

[70] Valaskovic, G. A., Kelleher, N. L., McLafferty, F. W., *Science* 1996, *273*, 1199–1202.

[71] Davis, M. T., Stahl, D. C., Hefta, S. A., Lee, T. D., *Anal. Chem.* 1995, *67*, 4549–4556.

[72] Piccinni, E., Staudenmann, W., Albergoni, V., De Gabrieli, R., James, P., *Eur. J. Biochem.* 1994, *226*, 853–859.

[73] Davis, M. T., Lee, T. D., *J. Am. Soc. Mass Spectrom.* 1997, 8, 1059–1069.

[74] Wilm, M., Neubauer, G., Mann, M., *Anal. Chem.* 1996, *68*, 527–533.

[75] Courchesne, P. L., Jones, M. D., Robinson, J. H., Spahr, C. S., McCracken, S., Bentley, D. L., Luethy, R., Patterson, S. D., *Electrophoresis* 1998, *19*, 956–967.

[76] Tomlinson, A. J., Naylor, S., *J. Capil. Electrophor.* 1995, *2*, 225–233.

[77] Figeys, D., Ducret, A., Yates III, J. R., Aebersold, R., *Nat. Biotechnol.* 1996, *14*, 1579–1583.

[78] Goodlett, D. R., Wahl, J. H., Udseth, H. R., Smith, R. D., *J. Microcol. Sep.* 1993, *5*, 57–62.

[79] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1426–1436.

[80] Yates III, J. R., Eng, J. K., McCormack, A. L., *Anal. Chem.* 1995, *67*, 3202–3210.

[81] Griffin, P. R., MacCoss, M. J., Eng, J. K., Blevins, R. A., Aaronson, J. S., Yates III, J. R., *Rapid Commun. Mass Spectrom.* 1995, *9*, 1546–1551.

[82] Yates III, J. R., Eng, J., Clauser, K. R., Burlingame, A. L., *J. Am. Soc. Mass Spectrom.* 1996, *7*, 1089–1098.

[83] McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., Yates III, J. R., *Anal. Chem.* 1997, *69*, 767–776.

[84] Patterson, S. D., Thomas, D., Bradshaw, R. A., *Electrophoresis* 1996, *17*, 877–891.

[85] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* 1990, *215*, 403–410.

[86] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., *Nucleic Acids Res.* 1997, *25*, 3389–3402.

[87] Johnson, R. S., Biemann, K., *Biomed. Environ. Mass Spectrom.* 1989, *18*, 945–957.

[88] Hines, W. M., Falick, A. L., Burlingame, A. L., Gibson, B. W., *J. Am. Soc. Mass Spectrom. 1992, 3, 326–336.*

[89] Taylor, J. A., Johnson, R. S., *Rapid Commun. Mass Spectrom.* 1997, *11*, 1067–1075.

[90] Schnolzer, M., Jedrzejewski, P., Lehmann, W. D., *Electrophoresis* 1996, *17*, 945–953.

[91] Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomson, B., Wilm, M., Mann, M., *Rapid Commun. Mass Spectrom. 1997, 11, 1015–1024.*

[92] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[93] Ducret, A., Van Oostveen, I., Eng, J. K., Yates III, J. F., Aebersold, R., *Protein Sci.* 1998, *7*, 706–719.

[94] Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., Mann, M., *Nature Genet.* 1998, *20*, 46–50.

[95] Lee, T. D., Vemuri, S., *Biomed. Environ. Mass Spectrom.* 1990, *19*, 639–645.

[96] Taylor, J. A., Walsh, K. A., Johnson, R. S., Rapid Commun. Mass Spectrom. 1996, *10*, 679–687.

[97] Yates III, J. R., Morgan, S. F., Gatlin, C. L., Griffin, P. R., Eng, J. K., *Anal. Chem.* 1998, *70*, 3557–3565.

[98] Fenyö, D., Zhang, W., Chait, B. T., Beavis, R. C., *Anal. Chem.* 1996, *68*, 721A–726A.

[99] Hu, J., Mungall, C., Nicholson, D., Archibald, A. L., *BioInformatics* 1998, *14*, 112–120.

[100] Hoogland, C., Baujard, V., Sanchez, J.-C., Hochstrasser, D. F., Appel, R. D., *Electrophoresis* 1997, *18*, 2755–2758.

[101] Lemkin, P. F., *Electrophoresis* 1997, *18*, 461–470.

[102] Martino, R. L., Johnson, C. A., Suh, E. B., Trus, B. L., Yap, T. K., *Science* 1994, *265*, 902–908.

[103] Williams, H., Zobel, J., *Comput. Appl. Biosci.* 1997, *13*, 549–554.

[104] Eckerskorn, C., Strupat, K., Schleuder, D., Hochstrasser, D. F., Sanchez, J.-C., Lottspeich, F., Hillenkamp, F., *Anal. Chem.* 1997, *69*, 2888–2892.

[105] Mortz, E., O'Connor, P. B., Roepstorff, P., Kelleher, N. L., Wood, T. D., McLafferty, F. W., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 8264–8267.

[106] Katta, V., Chow, D. T., Rohde, M. F., *Anal. Chem.* 1998, *70*, 4410–4416.

[107] Lennon, J. J., Walsh, K. A., *Protein Sci.* 1997, *6*, 2446–2453.

[108] Figeys, D., Ning, Y., Aebersold, R., *Anal. Chem.* 1997, *69*, 3153–3160.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 678–688

Review

Keith L. Williams

Proteome Systems, North Ryde, Sydney, Australia, Australian Proteome Analysis Facility, Department of Biological Sciences, Macquarie University, Sydney, Australia

Genomes and proteomes: Towards a multidimensional view of biology

The third Siena proteomics conference held August 31–September 4, 1998, heralded a change in emphasis from technology development to using proteomics to assist in resolving biological questions. In this review, proteomics is placed in context with other major influences in the way discovery research is conducted in biology. The current status of genomics is examined in its broadest sense, including how such studies may influence the development of proteomics. It is suggested that we are entering a new phase in biology where information is no longer limiting and integration of different technologies is required to attack the big problems of biology. While much of the focus of funding bodies, both in the public and private sector, is on practical outcomes (new drugs, *etc.*), the new technologies are equally amenable to attacking long-standing fundamental challenges, such as cell division, cell patterning and morphogenesis.

Keywords: Genomics / Proteomics / Pharmacogenomics / Review EL 3427

Contents

Correspondence: Professor Keith L. Williams, Proteome Systems Ltd, Locked Bag 2073, North Ryde, Sydney, NSW Australia 1670
E-mail: keith.williams@proteomesystems.com
Fax: +61-2-9889-1805

1 Holistic biology

Until very recently, biologists have studied molecules one at a time. There have been some quite heroic efforts in learning about biochemical pathways using this approach, whereby the individual components are identified step by step. The classical biochemical pathways were assembled in this manner using protein biochemistry and genetic analysis [1]. A combination of insensitive analytical techniques and the absence of tools for DNA sequence analysis made for slow progress. The molecular biology revolution changed the approach towards biological systems. Although the functional molecules of biology are proteins, genes represent the information banks for protein specification. In the last decade the discovery route has increasingly been through the genes made accessible through gene cloning. Even with the advent of gene cloning, assembling pathways is a slow and challenging endeavour, especially when one considers branch points and cross-talk between pathways [2]. Biology is so complex that one needs tools that allow a multiplicity of proteins to be studied simultaneously. In other words, there

is a need to introduce elements of parallel processing to tissue or even organism-wide studies. In this review, the technologies enabling such an approach will be discussed. For the first time, the complexity of a biological system can begin to be approached in its entirety.

1.1 The technologies enabling holistic biology

There are two technologies now reaching maturity that herald the holistic revolution. These are genomics and combinatorial chemistry. I suggest that while genomics (largely DNA sequencing, and preparation of informational maps of organisms) is a discipline in its own right [3], combinatorial chemistry [4] fits intellectually as a sub-set of the new technology of proteomics, a description of which forms much of this review. An outcome of genomics is the discipline of functional genomics whereby one uses DNA-based technologies to make inferences about organism structure and behaviour. Functional genomics is not clearly defined, although many researchers use the term to refer to large-scale analysis of gene expression patterns [5]. Here I shall summarise briefly the technologies.

2 Genomics

Genomics is a word that encompasses many different technologies, all of which are related in some way to the information content of a cell, in other words, its DNA or RNA. Until recently, the field was primarily involved with information collection, organisation and mining. The function of a gene is, of necessity, inferred rather than demonstrated in genomics as the field is primarily concerned with information rather than the functional products (proteins). There are a number of levels at which genomics can be considered.

2.1 Chromosome maps and comparative gene mapping

The big picture involves the generation of chromosome maps, whereby genes are identified and ordered along the chromosomes [6, 7]. This information can be used comparatively both between species and also amongst individuals within a species. Gross rearrangements, gene deletions and polymorphisms within genes are all revealed in such studies. It is interesting that such basic information has practical applications in, for example, human medicine. Traditionally, karyotyping has been used to identify gross abnormalities and recently the new field of pharmacogenomics has been spawned from studies on gene polymorphisms [8].

2.2 DNA sequencing

The automation of high throughput DNA sequencing has completely changed our view of biology. In the past, a fragmentary approach based on characterization of individual genes led to a highly dispersed view of the makeup of an organism. Now it is possible to systematically identify the complete information content of an organism. The nucleotide sequence of the extremely simple bacteriophage Φ-X 174 was completed in the late 1970s, at the time a *tour de force* [9]. Today it is becoming routine to sequence the genomes of whole microorganisms. Hence, the first free-living organism. *Haemophilus influenzae*, was fully sequenced in 1995 [10], *Escherichia coli* followed in 1997 [11], and now many prokaryotes are fully characterized [12]. Amongst these organisms are a number of human pathogens such as *Helicobacter pylori* [13] and *Rickettsia prowazekii*, which causes typhus [14]. Interestingly, these studies are now being extended to compare individual organisms within a species in order to seek, for example, pathogenicity genes. The first of such studies has recently been completed where two isolates of *H. pylori*, one from a patient with a duodenal ulcer, the other from a patient with gastritis, were compared and a small number of candidate genes responsible for the different pathogenicity were identified [15]. This opens up the integration of DNA sequencing with comparative mapping studies.

DNA sequencing initiatives are not only important for studies on simple organisms (prokaryotes and viruses), hence the first unicellular eukaryote, the yeast *Saccharomyces cerevisiae*, was fully sequenced in 1996 [16] and a multicellular eukaryote, the nematode worm *Caenorhabditis elegans*, was completed at the end of 1998 [17]. We can expect very soon the complete sequences of the fruit fly, *Drosophilia melanogaster*, the plant *Arabidopsis thaliana* and several other organisms. And the complete sequence of the human genome is now within view as a result of the entry of private sector groups into the race [18].

It is likely that complete sequencing efforts will become the norm for prokaryotes and indeed more than 70 such sequencing projects were reported in mid-1998 [19]. There are also more than 20 model eukaryote genomic sequencing programs underway [19]. Nevertheless, because eukaryotes often have large amounts of noncoding DNA, most DNA sequencing efforts for more complex organisms are likely to focus, at least in the first instance, on sequencing just the genes (and not the DNA between the genes). This is achieved by making a DNA copy of messenger RNA to generate expressed sequence tags (ESTs) [20]. Such databases are under construction for most organisms of interest to laboratory biologists (*e.g.*,

Dictyostelium discoideum [21]). Of course, this only provides information on genes being expressed at the time of sampling; the complete information about an organism is assembled by examining each tissue where different genes are turned on. ESTs are interesting in that they provide tissue-specific information, which gives an indication of function as a result of specific sets of genes being turned on. Databases of ESTs are much cheaper to construct than a full DNA sequencing program [22].

2.3 Bioinformatics for genomics programs

The above DNA sequencing initiatives generate massive amounts of genetic code, which must be reduced to useful information. The field of bioinformatics encompasses the different elements of such information ordering. It involves finding the start, intervening sequences, and end of a coding DNA sequence as well as the upstream regulatory elements, which give clues as to where and when the gene might be expressed. It is also possible to highlight sites for possible post-translational modification based on specific protein sequence motifs. Bioinformatics has been crucial to the microbial sequencing initiatives as much of the DNA sequencing involves studies on random fragments that are assembled using powerful informatic tools [12].

2.4 Functional genomics

While most of the above discussion concering genomics involves informational rather than functional data, inferences can be made about function. Expressed sequence tags give indications of transcriptional (but not necessarily translational) gene activity, and, when this is related to a specific tissue or a disease state, it becomes a discovery tool to find proteins likely to be relevant to the question under study. Several companies are developing the ability to organise large numbers of oligonucleotides on chips so that expression of specific proteins can be monitored [23]. In the case of some microbes (*e.g.*, the yeast *S. cerevisiae*), the expression state (*i.e.*, mRNA levels) of the whole organism can be studied and the response to drugs investigated [24]. There are still some technical hurdles, but the new chip-based technologies give confidence that this will become a robust experimental tool in the near future.

2.5 Pharmacogenomics

People are not all the same and there are an increasing number of genetic polymorphisms being identified which are relevant to disease state. It is generally accepted that few of the major human diseases have single gene defects as their cause [25]. Cystic fibrosis is one exception where mutation in a chloride channel gene, CFTR, is sufficient to cause the disease [26], but even in this case the genetic background has some influence on its severity. Nevertheless, it is clear that an individual's genotype contributes to his/her predisposition to specific diseases or likelihood of infection. By screening for such differences, it will be possible to build up genetic profiles on an individual basis. There are also gene polymorphisms that affect how an individual metabolises drugs. Clearly, if a new drug is to be registered, the side effects must be acceptable and it is not appropriate if a large number of people react to the drug adversely. The concept behind pharmacogenomics is to identify groups within the community who can be treated with a particular drug or, alternatively, to be able to screen out those patients who would react adversely to a given drug. Hence, the power of pharmacogenomics may well be to allow the introduction of some drugs which would otherwise not be registered because of their adverse affects on specific groups in the community [27].

Pharmacogenomics points to differences between individuals at specific gene loci, and where such loci are associated with disease, potential differences in susceptibility can be inferred. For pharmacogenomics to be feasible, it must be possible to rapidly and cheaply screen for the status of particular gene variants in individuals [28]. Although the discovery side of pharmacogenomics, *i.e.*, the existence of different alleles, is still largely the province of academic researchers, there is now a major corporate interest in commercialising this technology. In a recent short perspective, twenty-two corporate projects were outlined [29]. The initial effort is to develop diagnostic tests as the forerunner to use of pharmacogenomics in product developments. In fact, a number of pharmacogenomics partnerships that were announced in 1998 are directed towards new products or revisiting old products. For example, there is intense interest in the role of polymorphism in the cytochrome p450 3A4 gene in metabolising various classes of currently marketed drugs [30]. The outcomes of such studies will give information on the toxicity and efficacy of particular drugs on people with different genetic makeup.

2.6 Structural genomics

The success in functional genomic screening described above has led to extending the concept of inferring function from information. For a numer of years now, molecular and computational biologists have sought to infer the three-dimensional structure of a protein from its primary sequence. In the case of proteins that are members of a gene family, where at least one of the members has had

its structure determined by X-ray crystallography or NMR structural techniques, it is possible to infer structure of a new family member by reference to a known structure [31]. Such a study points to interesting areas of the structure that might require further examination. An extension of this approach is to computationally fold up a protein based on assembling its primary sequence into a three-dimensional shape [32]. This is still a challenging but very active area of current research. Because there are now several thousand structures determined, it is possible to conceive that within the not-too-distant future all protein folds will effectively be known [33]. The above-mentioned bioinformatics groups have developed automated routines for predicting structures of literally thousands of proteins sequenced in microbial genomics initiatives.

An extension of this is to attempt to experimentally determine the complete tertiary structures of all proteins in a given organism [34]. International consortia have been formed to conduct pilot projects. The concept at this stage is to express proteins of an organism of choice in *E. coli* to determine the three-dimensional structure of all proteins. Pilot projects concern *H. influenzae*, and two thermophiles: *Pyrobaculum aerophilum* and *Methanococcus jannaschii* [35]. This in itself is a daunting task as many proteins are not soluble and do not form crystals. Such proposals take little notice of post-translational modifications, which are found on most eukaryote proteins. However, a few years ago many thought that sequencing even a prokaryote genome was an impossibly difficult project! A follow-up of such a huge effort would be an investigation on how the proteins in an organism interact with each other. In fact, such experiments are already underway without necessarily knowing the structure of the interacting proteins (see Section 4.7.3).

3 Transgenics: gene manipulation *in vivo*

Genomics gives us information, but we as biologists really wish to know about function. One way to approach the activity of particular genes in the context of a whole organism is to manipulate a particular gene *in vivo*. There are various ways that this can be done and for some organisms gene manipulation *in vivo* is becoming relatively routine. Broadly speaking, interference with gene activity can be done at the DNA level by actually disrupting, or deleting, the gene using homologous recombination techniques [36]. A corollary of such an approach is reinsertion of the gene (and potentially analyse at different copy number) or a modified form of the gene. This has been done in a number of functional studies. For example, myosin function in *D. discoideum* has been probed by inserting various segments of the gene in a mutant strain or by site-directed mutagenesis [37]. Recently, this approach has been extended to study location of the relevant protein or domain of the protein by making green fluorescent protein fusions with the relevant gene [38]. Alternatively, the activity of a particular gene can be interfered with indirectly using antisense or ribozyme technology [39]. The outcome of both of these approaches is to prevent messenger RNA transcription or prevent the translation of the messenger RNA and, hence, the functional molecule (protein) is not synthesised.

Effectively, this is the study of function, one gene at a time. It is difficult to implement with essential genes as lack of an essential protein leads to death of the organism. There are various "tricks" that can be used, such as production of temperature-sensitive forms of the protein, to overcome this difficulty. Obviously, it is easier to implement such projects in haploid rather than diploid organisms. To disrupt or alter a particular gene is, in general, a time-consuming process as the native gene must be targeted and inactivated in some way. However, in some organisms, this is now becoming quite straightforward [40]. Again, this is easier to do if the organism is haploid and can be easily transformed. This applies to the yeast *S. cerevisiae* and the slime mould *D. discoideum*. These organisms can be readily stored as spores; strain collections comprising thousands of strains are thus easy to maintain. Even complex organisms such as the mouse are now relatively easy to transform and, hence, gene disruptions are now commonplace, although the challenge of maintaining transgenic mouse strains is costly [41].

How does one use a transgenic microorganism or animal? The first question is whether deleting the gene affects mortality and other gross aspects of morphology, physiology or behaviour. If there are such gross effects, it is possible to study this a little more closely by, for example, over-expressing the gene or perhaps, expressing portions of the gene, to determine functional domains of the protein product [42]. A second approach is to use genomics to probe in a more holistic fashion the outcome of the disrupting of a particular gene. By using messenger RNA display techniques, which are now being developed using chip technology, one can get an understanding of the changes in gene expression caused by the absence of a particular protein. It is crucial to understand that the absence of a particular protein almost never means that the only change in the organism is that the deleted protein is absent. An organism is a network and the disappearance of one protein leads to some proteins increasing in amount and others disappearing. There may also be changes in post-translational modifications. This is best studied at the protein level because, after all, organisms use proteins as their functional molecules. This leads us

to the field of proteomics, the study of protein expression on a global basis.

4 Proteomics

Protein science was once a slow and difficult art form. Since proteins can not be easily amplified (unlike DNA where PCR has revolutionised isolation and amplification of minute amounts of nucleic acids), sufficient material to characterize a protein must be extracted from the tissue under study and separated from contaminating proteins. Since eukaryotic tissue samples may comprise several thousand proteins [43], as well as numerous post-translationally modified forms of many of these proteins [44], it is not surprising that proteins have traditionally been purified and characterized only after a long and tedious process involving a number of steps. The development of a purification protocol for a particular protein may take several years [45]. While monoclonal antibody technology and its use for affinity purification of proteins from complex mixtures has simplified the purification process, there is still the requirement for the establishment of a hybridoma line for each protein to be purified. And the process remains a serial one (*i.e.*, only one protein is purified at a time) [46].

For many years it has been possible to array proteins by two-dimensional gel electrophoresis, but the amount of each protein separated has been at least one order of magnitude below the amount needed for chemical characterization [47]. Before protein analysis could approach the power of DNA characterization, two things were needed. First, it was necessary to develop semipreparative methods for purifying proteins in parallel. Second, more sensitive techniques for protein characterization were needed [48]. The simultaneous solution to both of these problems has led to a huge burst of activity in protein science. To reflect this paradigm shift, a new name has been given to describe the field that has opened up: "proteomics" the study of proteins expressed by a genome (simple organism) or tissue [49]. It is now possible to conceive of a complete description at the protein level of an organism, or tissue under a given set of conditions. Unlike the genome of an organism, which is the essentially fixed information base underpinning the organism, the proteome is a varying feature, subject to changes due to developmental stage, disease state, or environmental conditions.

4.1 Proteomics technology – lessons from genomics

For proteomics to be widely adopted, a robust technology must be established that allows the large-scale discovery research needed for a holistic approach to protein science. There are various ways that this could be implemented and the pace of change is such that it is likely that today's technology will become dated very quickly. It is instructive to consider what has happened in the development of genomics to seek parallels in the emergence of proteomics. Genomics is based primarily on sequencing of DNA, or using such information to assist in constructing ancillary features of genomics, such as chromosome maps, and pharmacogenomics. The ability to amplify DNA is central to essentially all of the technologies used in genomics. This involves preparing sufficient numbers of copies of the DNA of interest. PCR has revolutionised this activity [50]. Such amplified DNA is either sequenced directly or cloned into an *E. coli* vector for further amplification and sequencing. Large pieces of DNA (chromosome fragments obtained using rare base cutting restriction enzymes) may be inserted into a yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC) for amplification [51]. In this way a complex organism can be broken into smaller DNA fragments, which in turn can be subdivided into smaller pieces (in bacterial plasmids) of a size suitable for direct DNA sequencing. Genomic sequencing can be ordered using such progressive techniques, where the piece of DNA sequenced can be located on a chromosome. Alternatively, a "shotgun" approach can be taken where the genome is randomly cleaved and small DNA fragments sequenced. In this second approach the genomic sequence is assembled using informatics. This works well for prokaryotes and has recently been proposed for use in sequencing the human genome [52]. It is unlikely that protein amplification will be possible in proteomics. Instead, amplification will be achieved by an enrichment process, where proteins of interest are differentially extracted (*e.g.*, on the basis of solubility, organelle location, or physical characteristics such as ability to bind specific ligands [53]). Unlike genomics, if two-dimensional gels are used to array the protein samples, proteins are systematically presented so that one can minimise the amount of repetition in analysis.

Once the DNA sample is available for sequencing in genomics, or the protein arrayed for proteome analysis, the question is what technology is to be used, remembering that one wants the highest possible throughput and fidelity in analysis. In genomics there have been a number of quite different technologies explored over the last decade. These include: (i) traditional gel-based separation of DNA to obtain the sequence by tracking the changes as each base is added to an extending DNA copy [54]; (ii) determination of sequence based on hybridisation of a target piece of DNA with a chip that contains many combinations of DNA oligomers [55]; (iii) mass spectrometry (MS), proposed as a rapid means of sequence polymorphism identification, based on measuring the mass of different DNA fragments. An advantage of MS is that it requires no prior

labeling as do gel-based methods [28]; and (iv) scanning tunneling microscopy, proposed as a means of directly visualising a DNA sequence [56]. Gel-based techniques are probably the most primitive of the above possibilities, but they have been perfected and they work. Virtually all genomic sequencing to date has used gel-based instruments and while DNA chips and mass spectrometry are coming of age for special applications, it is certain that the human genome will be completed with gel-based techniques. Hybridisation on chips is coming of age for screening expression of genes in complex tissues and there are now chips that display the *S. cerevisiae* genome, or a specific human tissue [57]. This has a big future. Scanning tunneling microscopy is perhaps the most sophisticated of all of the methods, but it is both too late and probably too complex to be of practical utility in genomics.

What are the lessons for proteomics? Some researchers have expressed reservations about two-dimensional gel technology as a practical possibility for protein array in large-scale proteomics, instead preferring to use a combination of liquid chromatography steps. It is true that there are remarkable developments in microfluidics and applying proteins to chips (especially in the biosensor area [58]), but fundamentally one requires sufficient molecules to study and so there are practical limits to miniaturisation. It is also possible to study complex mixtures of proteins with mass spectrometric techniques, and so some argue that it is not necessary to purify proteins [59]. My personal view is that simplicity always has its attractions. Two-dimensional gel technology works, *albeit* it requires skill to prepare gels with mg quantities of protein loaded [60]. Hundreds of proteins can be separated and characterized in a single gel. Just as the core technologies in genomics have consolidated around gel technology, proteomics is likely to consolidate around the two-dimensional gel for protein array.

4.1.1 Automation: turning technologies into fast throughput systems

Scientists in mid-career in the 1990s have grown up with a pipette or a micropipette in their hand. They are accustomed to doing things one at a time in a manual fashion. The instruments of the 90s on the other hand are moving towards automation where human intervention is minimised. This is done for several reasons, probably the most important of which is to ensure accuracy of execution of very large numbers of repetitive tasks.

4.2 Protein array using two-dimensional gels

For many years two-dimensional gels were neglected as a purification technology as it was difficult to characterize the small amounts of proteins separated. A small group of dedicated researchers persisted with two-dimensional gel separations in combination with sensitive techniques such as Western blotting. Celis and co-workers [61] have used such an approach over many years to produce the first detailed proteome maps involving diseased human tissue. Another group of researchers sought to develop the original superb separations achieved with two-dimensional gels into a reliable preparative technique. The key innovation was to adopt the immobilised pH gradient technology developed by Righetti [62] and apply it to high-load two-dimensional gels. Several groups in Europe (Dunn, Görg, Hochstrasser, Pallini) have perfected highly reproducible two-dimensional gels that contained milligram loading of proteins [63–65]. Rabilloud [66] and Herbert [53] have developed differential extraction techniques to display a complex tissue over several two-dimensional gels on the basis of differential solubility. By seeking appropriate detergents to keep protein samples soluble through both the first- and second-dimensional separations, not only soluble proteins, but also highly insoluble membrane proteins can now be arrayed in preparative loads. A further refinement of the two-dimensional electrophoresis technique is to use narrow-range pH gradients to stretch out the sample in the first-dimensional separation. Of course the two-dimensional gel step can be prefaced by a range of sample preparation steps (isolation of organelles, *etc.*). The outcome of combining the above is likely to be a series of two-dimensional gels for a specific sample that are reconstructed using informatic techniques. This is the future for parallel protein array prior to characterization.

4.3 Protein characterization

Proteins are characterized using attribute matching techniques such as describing their size, isoelectric point, presence of modifications, solubility, start and finish (*N*- and *C*-terminus, respectively). The act of arraying proteins using two-dimensional gels already provides information about size and p*I*. Traditionally, the start of a protein was determined by Edman sequencing, which is slow and expensive, but providing the protein is not *N*-terminally blocked, gives sufficient information to identify the protein based on a short sequence of 4–8 amino acids. With the development of blotting technologies by Matsudaira [67] it became possible to sequence proteins from preparative two-dimensional gel blots. More recently, it has become possible to determine one or two amino acids at the *C* terminus, although sensitivity is still a problem with this technology. Proteins can also be identified from two-dimensional gel spots based on their amino acid composition and matching to the composition of proteins in databases [68, 69].

All of the above technologies are being superseded by mass spectrometric identification of proteins [70]. This has become possible through a combination of improvement in mass spectrometric technologies and the development of databases of DNA (protein sequence) information. Peptides, and their fragments, can be weighed with such accuracy that, by reference to databases, a particular peptide and its fragmentation pattern can lead to the identification of the protein. In effect, mass spectrometry and the weighing of molecules is now the preferred way for doing protein characterization. Not only can proteins be identified, but their modifications can also be characterized by taking into account the mass of modifications such as phosphate, sugars and other protein modifications.

4.4 Proteomics: differential protein display

Many applications of proteomics relate to phenotypic differences between treatments, be it a pathogenic *versus* nonpathogenic strain of a particular microorganism or cancer *versus* normal tissue. Here the focus is on quickly finding differences and then focusing on the nature of these differences. Such studies complement a genomics approach based on differential display [71]. Differential display in both genomics and proteomics is still in its infancy, with method development still a major focus. Note that the absence of a protein or a change in its modification may be as crucial as the presence of a new protein. In cancer biology, it is now well understood that cancer is often a combination of new activities such as making cell division constitutive (*e.g.*, loss of growth factor requirement) and loss of control functions (such as failure of cell death to occur) [72]. Using two-dimensional display technology, an obvious outcome of imaging gels of controls and treatments is the identification of differences. There are a number of software packages for two-dimensional gel analysis such as Melanie [73] and PDQuest [74] where alterations can be automatically highlighted. With complex tissues, it is important to study the differences, as what may appear to be the disappearance of a protein may instead be a change in its modification (see Section 4.5). In order to conduct detailed studies on differential display, it is important to be able to array a complex tissue sample over more than just one two-dimensional gel. This of necessity involves fractionation techniques; to be able to make sense of such fractionated samples, great care must be taken in quantitating the fractionated tissues. There are various other ways that proteins can be differentially distributed in tissues. For example, a gene may be expressed but the protein product may be differentially located under different conditions (*e.g.*, membrane-associated or soluble).

4.5 Post-translational modifications

Perhaps the defining difference between proteomics and genomics is in the area of post-translational modifications. It is now clear that the activity state of a protein often depends on its modification state. While the expression of a gene may be the same in two situations, if the phosphorylation status is different this may signal an active protein under one set of conditions and an inactive protein under the other. For example, it is clear that signal transduction pathways, the cell cycle, and many other crucial pathways in eukaryote development are dependent on phosphorylation/dephosphorylation cycles [75]. Recent studies on cytoplasmic glycosylation indicate that sites that can be phosphorylated can also be glycosylated [76]. A given serine or threonine residue may thus have one of three activity states: unmodified, glycosylated, or phosphorylated. Only studies at the proteome level clarify such diversity. In other cases proteins may be differentially processed so that they occur in different forms. Indeed, the complexity of processing at the mRNA level can be enormous. For the cell adhesion molecule (CAM) a number of differently spliced forms are known to produce both cell surface and secreted forms of the protein from the same gene [77]. Little is known about the abundance of the different forms at the protein level, but such analysis is now possible through proteomics. If there is a charge or size difference as a result of the post-translational modification, the different forms of the protein are easily separated using two-dimensional gels. For the cancer suppressor protein p53, different forms of the protein have been identified by two-dimensional display [78].

4.6 Combinatorial chemistry

An area almost as well-developed as genomics is that of combinatorial chemistry, which describes the technology for finding small molecule targets in drug discovery. Just as high throughput DNA sequencing revolutionised genomics, the development of rapid combinatorial chemistries for manufacturing huge numbers of small molecules has, in combination with rapid screening technologies, opened up new ways of finding, in a randomised fashion, molecules of interest [79]. The technology involved is beyond the scope of this review but the real relevance is that it offers an indirect approach to functional studies on proteins and, in practical terms, drug discovery.

4.7 New opportunities for proteomics

4.7.1 Structural proteomics

Currently there is a massive effort underway in individual laboratories around the world to determine the structure

of all proteins and the structural genomics initiatives have already been discussed. While the genomics approach will yield a huge amount of data for prokaryotes, there will be problems with post-translational modifications (*E. coli* will not necessarily handle the modifications correctly) and some proteins will not fold correctly. This will be a major issue for studies on eukaryotes [80]. A proteomics approach might be to seek to develop techniques for structure determination of native proteins arrayed in a form similar to two-dimensional gels. There are new mass spectrometric techniques for determining compact and accessible regions of proteins, giving "functional" structural information. Since techniques for growing protein crystals are being miniaturised it is not impossible that eventually the amounts of material arrayed on preparative gels will be sufficient for X-ray crystallographic study.

4.7.2 Quantitation of cellular proteins

There is surprisingly little attention paid to the numbers of protein molecules in cells. It is relatively easy to estimate the number of molecules of the most abundant proteins and this is of the order of 100 million molecules for a eukaryotic cell, representing several per cent of the cell's protein content (*e.g.*, actin). The least abundant proteins are more controversial, although it is hard to imagine a rare protein could find its site of action if there were less than 100 molecules made. This means a dynamic range of approximately one million from the least to most abundant proteins. It is not feasible to analyse such a dynamic range without prefractionating the proteins in some way. When no fractionation is carried out, finding the rare proteins in a huge excess of the abundant proteins becomes impossible. If, for example, one is displaying proteins using two-dimensional gel array, many rare proteins would be hidden underneath an abundant protein spot. The answer to this, of course, is to fractionate the sample so that abundant proteins are specifically removed. Affinity techniques are often available for specifically removing abundant proteins, while, conversely, rare proteins can be concentrated using the same techniques.

The new characterization techniques using mass spectrometry are not well suited to quantitation. Currently, it is a difficult process to quantitate cellular proteins because techniques such as amino acid analysis or Edman sequencing must be used. High throughput techniques for such analyses are limited by sensitivity at the level of approximately one picomole. It is possible to quantitate proteins according to staining using, for example, silver stain or one of the new generation fluorescent stains that are becoming available. The problem here is that such stains vary in their intensity depending on the protein; comparative estimates of protein levels can therefore be done but actual quantitation is not easy. When it becomes possible to quantitate the number of protein molecules, it will be feasible to walk around metabolic pathways and understand potential bottlenecks.

The questions that the above discussion ultimately addresses include "how important is the amount of any particular protein?" and also "to what extent can the amount change without affecting the system?" In bacteria, at least, it is becoming possible to address such questions by deleting particular genes and replacing them under tightly regulated promoters so that the amounts of protein produced can be accurately controlled. Of course, this assumes that the amount of protein produced is related to the level of gene expression and it is now well known that this is not always the case [81]. Such a study has been conducted on the chemotaxis system in *E. coli* [82]. The amounts of protein CheR, which is responsible for chemoreceptor methylation, were manipulated. Varying the induction of the gene approximately 100-fold led to an approximately 10- to 50-fold variation in the amount of protein. Adaptation proved to be a robust feature of the system, while other features such as time to adapt and the actual steady-state behaviour were sensitive to levels of protein. Clearly, this is the beginning of a new world where complex phenotypes will be dissected, both in terms of the component parts and also the quantitative amounts of the components.

4.7.3 Protein-protein interactions

Cells are clearly more than bags of molecules and it has long been suspected that there is order in protein interactions. As detailed analysis of signaling pathways leads to precise studies on the components of such pathways, the importance of protein dimerization and interaction is becoming clear. For example, in transmitting signals from a cell surface receptor via the Raf/Mek/ERK pathway which leads to gene activation, key proteins first dimerise and then interact with other proteins [83]. Just as gene analysis was done one gene at a time in the past, so has the study of protein-protein interaction been a single interaction process. The yeast two-hybrid system is a molecular biologist's way of looking at how proteins talk to each other, or, more particularly, which proteins talk to each other [84]. Western blotting has long been a protein chemist's way of looking at protein-protein interactions (in this case the very special interaction between an antibody and its antigen). There is good reason to suspect that proteomics will develop blotting techniques for studying protein-protein interactions between two-dimensional gel-separated proteins, although of course if such interactions require protein oligomerisation, then native two-dimensional gels may be required. Mass spectrometry is emerging as the

new protein interaction tool [85]. It is possible to investigate, with judicious use of proteinases, the interaction regions of proteins.

5 New ways of doing science: public sector/private sector

5.1 The public sector

The public sector represents a huge reservoir of experience in biology. There are thousands upon thousands of university and medical research labs, each investigating a particular aspect of a problem of interest. The development of a cancer cell line, for example, that exhibits particular properties in comparison with an untransformed cell line provides the basis for large-scale interrogation of the changes occurring in cancer. This is also true in the development of databases, where a large number of publicly funded initiatives have produced hundreds of databases. The challenge for the future is whether such databases will continue to be supported solely by public funding or whether other means can be devised to maintain their operations [86]. With difficulties in funding being experienced by laboratory researchers in the university and the medical scene, there has been a huge pressure to seek other sources of funding and this has led to an aggressive assessment of the status of basis research labs by the emerging biotech community. Nowhere is this more apparent than in the USA where 180 biotech companies with $7.5 billion annual research and development expenditure have more than 300 products in clinical trials. To provide a context for this huge number of potential new products, Merck Corporation, a blue-chip company that invests US $1.7 billion in research, has only 20 products in clinical trials. The world is changing and these figures illustrate that change. It is important to remember that the above discussions largely focus on developments that are essentially one-dimensional, *i.e.*, they involve applying a particular technology, be it genomics or combinatorial chemistry, to a particular biological system. There is a new positive attitude towards multidisciplinary approaches and the development of centres of excellence. A new initiative for a combined genomics and proteomics facility at Harvard University announced early in 1999 reflects this trend.

5.2 The private sector

The economic miracle in the USA in the last decade has come about from small technology companies which started on the basis of an interesting idea and constructed a company around it. A decade of such start-ups is now beginning to require integration. The funding of the start-up companies has come largely from the very large Ag-Biotech and Pharma groups. Initially they were content to fund different aspects of technology development. Now they are increasingly seeking integrated solutions. In some cases, they are doing this by seeking out small companies or universities with different aspects of the technology jigsaw and putting the picture together themselves. But the biotech companies have got the message and they are now beginning to integrate in their own right. As the approach to biology becomes more holistic, it is necessary to develop more integrated structures.

6 The future

6.1 Integration of proteomics, genomics and combinatorial chemistry

The purpose of all the technological developments outlined in this review is ultimately to progress our understanding of living organisms. We are at the beginning of an information deluge much greater than even the most optimistic seers would have predicted five years ago. Some would argue that biology will soon become a cyber discipline where all the ingredients are known and the challenge is to make sense of it all. This underestimates the importance of experimental biology, although biologists can no longer confine their interest to a single molecule or even pathway. It is time to think about integration both at the informatic as well as experimental levels. This is the future, and it promises many challenges.

Many people have contributed to sparking our interest in proteomics. I acknowledge specifically Denis Hochstrasser and Thierry Rabilloud who have generously supported visits from young researchers from our Sydney group. The Australian government major national facilities program has supported the development of the Australian Proteome Analysis Facility (APAF), without which the tremendous boost to the development of proteomics in Australia would not have happened. The Australia Research Council and Australia Medical Research Council have supported research in my group that has encompassed proteomics. Particular thanks to Bio-Rad Australia and US groups who have supported our two-dimensional gel technology developments and to the Advanced Rapid Robotic Manufacturing group (Adelaide, Australia) who have made our entry into robotics such fun. Thanks to Marc Wilkins and Andrew Gooley for helpful comments on this manuscript.

Received February 17, 1999

7 References

[1] Krebs, H. A., *Perspect. Biol. Med.* 1970, *14*, 154–170.

[2] Zuker, C. S., Ranganathan, R., *Science*, 1999, *283*, 650–651.

[3] Antonarakis, S. E., *Genomics* 1998, *51*, 1–16.

[4] Myers, P. L., *Curr. Opin. Biotechnol.*, 1997, *8*, 701–707.

[5] Woychik, R. P., Klebig, M. L., Justice, M. J., Magnuson, T. R., Aver, E. D., *Mutat. Res.* 1998, *400*, 3–14.

[6] http://www.hgmp.mrc.ac.uk/Genome Web/nematode-gen.db.html

[7] http://www.hgmp.mrc.ac.uk/Genome Web/human-gen-db-chromosomes.html

[8] Housman, D., Ledley, F. D., *Nature Biotechnol.* 1998, *16*, 492–493.

[9] Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M., Smith, M., *J. Mol. Biol.* 1978, *125*, 225–246.

[10] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., Fitzbugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., *Science* 1995, *269*, 496–512.

[11] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y., *Science* 1997, *277*, 1453–1471.

[12] http://www/tigr.org/tdb/mdb/mdb/html

[13] Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Venter, J. C., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C., *Nature* 1997, *388*, 539–547.

[14] Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H., Kurland, C. G., *Nature* 1998, *396*, 133–140.

[15] Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F., Trust, T. J., *Nature* 1999, *397*, 176–180.

[16] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq., C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, *274*, 563–567.

[17] The *C. elegans* Sequencing Consortium, *Science* 1998, *282*, 2012–2018.

[18] Brower, V., *Nature Biotechnol.* 1998, *16*, 1004.

[19] www.mcs.anl.gov/home/gaasterl/genomes.html

[20] Fields, C., *Curr. Opin. Biotechnol.* 1994, *5*, 595–598.

[21] www.csm.biol.tsukuba.ac.jp/cDNAproject.html

[22] Carulli, J. P., Artinger, M., Swain, P. M., Root, C. D., Chee, L., Tulig, C., Guerin, J, Osborne, M., Stein, G., Lian, J., Lomedico, P. T., *J. Cell Biochem. Suppl.* 1998, *30–31*, 286–296.

[23] Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., Davis, R. W., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 2150–2155.

[24] Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, Jr., D. E., Hartwell, L. H., Brown, P. O., Friend, S. H., *Nature Med.* 1998, *4*, 1293–1301.

[25] Strohman, R., *Bio/Technol.* 1994, *12*, 156–164.

[26] Hung, L.-W., Wang, I. X., Nikaido, K., Liu, P.-Q., Ames, G. F.-L., Kim, S.-H., *Nature* 1998, *396*, 703–707.

[27] Persidis, A., *Nature Biotechnol.* 1998, *16*, 209–210.

[28] Laken, S. J., Jackson, P. E., Kinzler, K. W., Vogelstein, B., Strickland, P. T., Groopman, J. D., Friesen, M. D., *Nature Biotechnol.* 1998, *16*, 1352–1356.

[29] Persidis, A., *Nature Biotechnol.* 1998, *16*, 791–792.

[30] Michalets, E. L., *Pharmacotherapy* 1998, *18*, 84–112.

[31] Peitsch, M. C., Guex, N., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 177–186.

[32] Holm, L., Sander, C., *Nucleic Acids Res.* 1999, *27*, 244–247.

[33] Holm, L., Sander, C., *Proteins*, 1998, *33*, 88–96; http://www.embl-ebi.ac.uk/dali/

[34] Kim, S.-H., *Nature Struct. Biol.* 1998, *5*, 643–645.

[35] Gaasterland, T., *Nature Biotechnol.* 1998, *16*, 625–627.

[36] Shashikant, C. S. L., Carr, J., Bhargava, J., Bentley, K. L., Ruddle, F. H., *Gene* 1998, *223*, 9–20.

[37] Fujita, H., Sugiura, S., Momomura, S., Sugi, H., Sutoh, K., *Adv. Exp. Med. Biol.* 1998, *453*, 131–137.

[38] Zang, J. H., Spudich, J. A., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 13652–13657.

[39] Gomer, R. H., *Genet. Eng. (N.Y.)* 1998, *20*, 135–141.

[40] Kuspa, A., Dingermann, T., Nellen, W., *Experientia* 1995, *51*, 1116–1123.

[41] Cohen-Tannoudji, M., Babinet, C., *Mol. Hum. Reprod.* 1998, *4*, 929–938.

[42] Novak, K. D., Titus, M. A., *Mol. Biol. Cell* 1998, *9*, 75–88.

[43] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[44] Gooley, A. A., Packer, N. H., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 65–91.

[45] Mockus, S. M., Vrana, K. E., *J. Mol. Neurosci.* 1998, *10*, 163–179.

[46] Jones, C., Patel, A., Griffin, S., Martin, J., Young, P., O'Donnell, K., Silverman, C., Porter, T., Chaiken, I., *J. Chromatogr. A* 1995, *707*, 3–22.

[47] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[48] Vorm, O., Mann, M., *J. Am. Mass. Spec.* 1994, *5*, 955–958.

[49] Williams, K. L., Hochstrasser, D. F., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 1–12.

[50] Hill, P. J., Stewart, G. S., *Biotechnol. Genet. Eng. Rev.* 1992, *10*, 343–377.

[51] Schlessinger, D., Nagaraja, R., *Ann. Med.* 1998, *30*, 186–191.

[52] Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., Hunkapiller, M., *Science* 1998, *280*, 1540–1542.

[53] Herbert, B. R., *Electrophoresis* 1999, *20*, 660–663.

[54] Sanger, F., *Science* 1981, *214*, 1205–1210.

[55] Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J. D., Drmanac, R., *Nature Biotechnol.* 1998, *16*, 54–58.

[56] Haggerty, L., Lenhoff, A. M., *Biotechnol. Prog.* 1993, *9*, 1–11.

[57] Watson, A., Mazumder, A., Stewart, M., Balasubramanian, S., *Curr. Opin. Biotechnol.* 1998, *9*, 609–614.

[58] Turner, A. P. F., *Nature Biotechnol.* 1998, *16*, 824.

[59] Aebersold, R., Patterson, S. D., in: Angeletti, R. H. (Ed.), *Protein Structure*, Academic Press, San Diego 1998, pp. 4–77.

[60] Herbert B. R., Sanchez, J.-C., Bini, L., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 13–33.

[61] Celis, J. E., Østergaard, M., Jensen, N. A., Gromova, I., Rasmussen, H. H., Gromov, P., *FEBS Lett.* 1998, *430*, 64–72; http://biobase.dk/cgi-bin/celis

[62] Righetti, P. G., Bossi, A., *J. Chromatogr. B* 1997, *699*, 77–89.

[63] Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J. J., Madjar, J. J., *Electrophoresis* 1997, *18*, 328–337.

[64] Bini, L., Magi, B., Marzocchi, B., Arcuri, F., Tripodi, S., Cintorino, M., Sanchez, J.-C., Frutiger, S., Hughes, G., Pallini, V., Hochstrasser, D. F., Tosi, P., *Electrophoresis* 1997, *18*, 2832–2841.

[65] Corbett, J. M., Dunn, M. J., Posch, A., Görg, A., *Electrophoresis* 1994, *15*, 1205–1211.

[66] Rabilloud, T., *Electrophoresis* 1998, *19*, 758–760.

[67] Matsudaira, P., *J. Biol. Chem.* 1987, *256*, 7990–7997.

[68] Shaw, G., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5138–5142.

[69] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Nature BioTechnol.* 1996, *14*, 61–65.

[70] Roepstorff, P., *Curr. Opin. Biotechnol.* 1997, *8*, 6–13.

[71] Szallasi, Z., *Nature Biotechnol.* 1998, *16*, 1292–1293.

[72] Lopez-Saez, J. F., de la Torre, C., Pincheira, J., Gimenez-Martin, G., *Histol. Histopathol.* 1998, *13*, 1197–1214.

[73] Appel, R. D., Vargas, J. R., Palagi, P. M., Walther, D., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2735–2748.

[74] Sinha, P., Hutter, G., Kottgen, E., Dietel, M., Schadendorf, D., Lage, H., *J. Biochem. Biophys. Methods* 1998, *37*, 105–116.

[75] Meek, D. W., *Cell Signal.* 1998, *10*, 159–166.

[76] Snow, D. M., Hart, G. W., *Int. Rev. Cytol.* 1998, *181*, 43–74.

[77] Gower, H. J., Barton, C. H., Elsom, V. L., Thompson, J., Moore, S. E., Dickson, G., Walsh, F. S., *Cell* 1988, *55*, 955–964.

[78] Sanchez, J.-C., Wirth, P., Jaccoud, S., Appel, R. D., Sarto, C., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 638–641.

[79] Veber, D. F., Drake, F. H., Gowen, M., *Curr. Opin. Chem. Biol.* 1997, *1*, 151–156.

[80] Montelione, G. T., Anderson, S., *Nature Struct. Biol.* 1999, *6*, 11–12.

[81] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[82] Alon, U., Surette, M. G., Barkai, N., Leibler, S., *Nature* 1999, *397*, 168–171.

[83] Marshall, C. J., *Nature* 1996, *383*, 127–128.

[84] Wallach, D., Boldin, M. P., Kovalenko, A. V., Malinin, N. L., Mett, I. L., Camonis, J. H., *Curr. Opin. Immunol.* 1998, *10*, 131–136.

[85] Chait, B. T., *Nature Biotechnol.* 1996, *14*, 1544.

[86] Ellis, L. B. M., Kalumbi, D., *Nature Biotechnol.* 1998, *16*, 1323–1324.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 689–700

Rosamonde E. Banks[1]
Michael J. Dunn[2]
Mary A. Forbes[1]
Anthea Stanley[1]
Darryl Pappin[3]
Tom Naven[3]
Michael Gough[1]
Patricia Harnden[4]
Peter J. Selby[1]

[1]ICRF Cancer Medicine Research Unit, St James's University Hospital, Leeds, UK
[2]Department of Cardiothoracic Surgery, National Heart and Lung Institute, Imperial College School of Medicine, Heart Science Centre, Harefield Hospital, Harefield, UK
[3]ICRF Protein Sequencing Lab, Lincolns Inn Fields, London, UK
[4]Department of Pathology, St James's University Hospital, Leeds, UK

The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis – Preliminary findings

Proteomics-based studies offer a powerful complementary approach to DNA/RNA-based investigations and are now being applied to investigate aspects of many diseases including cancer. However, the heterogeneous nature of tissue samples often makes interpretation difficult. We have undertaken a study into the potential use of a novel laser capture microdissection (LCM) system to isolate cells of interest for subsequent proteomic analysis. Retrieval of selected cells is achieved by activation of a transfer film placed in contact with a tissue section, by a laser beam (30 or 60 μm diameter) which is focused on a selected area of tissue using an inverted microscope. The precise area of film targeted by the laser bonds to the tissue beneath it and these cells are then lifted free of surrounding tissue. Although the technique has been shown to be readily compatible with subsequent analysis of nucleic acids, little information is yet available regarding the application of protein-based analyses to the captured tissue. We report here preliminary data regarding the potential use of the LCM system in combination with two-dimensional electrophoresis to examine protein profiles of selected tissue areas. Electrophoretic profiles of proteins from normal and malignant renal tissue samples showed little change following LCM, nine selected proteins showed identical mass spectrometric sequencing profiles, and two selected proteins retained antigenicity. Dissection of epithelial tissue from a sample of normal human cervix resulted in enrichment of some proteins compared with analysis of the whole tissue. LCM will be a valuable adjunct to proteomic studies although further detailed validation is necessary.

Keywords: Proteomics / Protein / Laser capture microdissection / Two-dimensional electrophoresis / Sequencing
EL 3396

1 Introduction

Major advances in our understanding of the pathogenesis of diseases have been made at the molecular genetic level, for example the identification of oncogenes and tumour suppressor genes in cancer [1], illustrating the power of such technologies. However, investigation of diseases by analysis of DNA/RNA sequence information alone has inherent limitations, amongst these the analysis of the phenotype of multigenic phenomena, and the inability to predict whether gene products are translated or have undergone post-translational modifications [2]. In addition, the correlation between mRNA levels and protein concentration is often poor [3]. The use of proteomics-based strategies, examining the profiles of the "functional units" or proteins of cells or tissues under particular conditions, is expanding rapidly, following relatively recent advances in technology, and offers a powerful complementary approach, overcoming some of these limitations of nucleic acid analysis.

One of the main problems with the analysis of tissue samples, either at the level of proteins or genes, is the heterogeneous nature of the sample. Many different cell types are often present and in the case of diseased tissue, abnormal cells may lie within or adjacent to unaffected areas. Previous approaches to overcome this and allow the examination of pure or enriched populations of cells have included the preparation of cell suspensions and either positive selection for the cells of interest or depletion of contaminating cells, often by immunological means. Often used in conjunction with this initial strategy has been the subsequent generation of short-term selective cultures of particular cell types from the tissue. Such strategies have been used in many studies to produce exciting results but the extent to which the manipulations inherent in these approaches influence the protein expression profile is not known.

Correspondence: Dr. Rosamonde E. Banks, ICRF Cancer Medicine Research Unit, St. James's University Hospital, Leeds LS9 7TF, UK
E-mail: r.banks@leeds.ac.uk
Fax: +0113-2429886

Abbreviation: LCM, laser capture microdissection

Various microdissection techniques have been employed to obtain pure or enriched populations of cells from tissue sections, including manual scraping of an area of tissue with a needle or other probe to positively select cells of interest [4, 5], or ablation of unwanted tissue using a targeted ultraviolet laser microbeam and subsequent collection of circumscribed areas using a probe [6–8]. However, these techniques are relatively time-consuming, require skilled operators due to the manual dexterity required, and in some cases are hampered by poor delineation of tissue and susceptibility to contamination. The use of membrane-based mounts for tissue sections has subsequently facilitated the more rapid retrieval of cells following ablation of unwanted areas with an ultraviolet laser (MOMeNT – microbeam microdissection of membrane-mounted native tissue) and minimised contamination [9].

Recently a new infra-red laser capture microdissection (LCM) technology has been described which was developed at the National Institutes of Health (USA) [10–12] and is now available commercially (Arcturus Engineering Inc., Mountain View, CA, USA). This allows the selective, relatively rapid microdissection of specific areas of tissue using a low maintenance system that is easy to operate. The capture of areas of tissue is achieved by activation of a transfer film by a solid-state near-infrared laser beam which is focused by the operator on a selected area of tissue under the microscope. The precise area of film, targeted by the laser, bonds to the tissue beneath it and these cells can then be lifted free of surrounding tissue and used for analysis. Although the technique has been shown to be readily compatible with subsequent analysis of nucleic acids, little information is yet available regarding the application of protein-based analyses to the captured tissue. Such an approach may allow the more ready identification of differences in protein expression of selected cell types or areas of tissue, for example tumour tissue, with a minimal effect on the protein expression profile present *in vivo*. We report here preliminary data regarding the potential use of the LCM system in combination with 2-D electrophoresis to examine protein profiles of selected tissue areas, specifically examining protein electrophoretic mobilities, protein antigenicity, and protein sequencing following LCM.

2 Materials and methods

2.1 Materials

All chemicals were Analar grade or equivalent and were purchased from BDH (Poole, Dorset, UK) or Sigma (Poole, Dorset, UK) unless listed below. Complete™ protease inhibitor cocktail tablets were from Boehringer Mannheim (Lewes, UK); ultracentrifuge microtubes from Beckman (High Wycombe, UK); urea from ICN Biomedicals (Thame, UK); CHAPS from Calbiochem (Nottingham, UK); Pharmalyte, IPG strips, ECL system, Plus-One™ silver staining kit and Hybond C from Amersham Pharmacia Biotech (Amersham, UK); protein assay kit, ammonium persulphate and TEMED from Bio-Rad (Hemel Hempstead, UK); Protogel polyacrylamide solution from Flowgen (Hull, UK); mini gradient gels and molecular weight markers from Novex (San Diego, CA, USA); HSP-60 antibody from Santa Cruz (Santa Cruz, CA, USA); and β_2-microglobulin antibody and secondary detection reagents from Dako (High Wycombe, UK).

2.2 Tissues

Four samples of renal cortical tissue (two normal and two malignant) and normal tissue from one cervix (hysterectomy specimen) were obtained from tissue specimens routinely sent to the Department of Pathology. Areas of tumour selected by the pathologist were free of necrosis or haemorrhage. Briefly, small blocks of tissue (approximately 5–10 mm cubed) were rinsed in sterile phosphate-buffered saline (PBS), pH 7.2, embedded in OCT compound, wrapped in foil and snap-frozen in liquid nitrogen within 30 min of surgical removal. Sections (5–10 μm) were subsequently cut using a Leica Cryocut 1800 microtome, stained using haematoxylin and eosin, and the nature of the tissue confirmed by a pathologist. Additional sections were also cut for protein analysis as described below.

2.3 LCM

For a comparison of effects of sample preparation and LCM on protein profiles, 8 μm sections of frozen renal tissue from each of the samples above were cut and alternately placed directly in lysis buffer (described below) or on clean alcohol-dipped glass slides. Sections on glass slides were placed on dry ice and then stained in batches of four sections at a time. Staining consisted of sequential incubations in 70% ethanol (1 min), haematoxylin (30 s), Scott's tap water (10 s), eosin (10 s), 70% ethanol (30 s) and 100% ethanol (2 × 30 s) with brief water washes between most steps, followed by two final rinses in xylene (5 min each). Both the haematoxylin and eosin solutions contained Complete protease inhibitor cocktail. After brief air-drying, whole sections were then either scraped into lysis buffer or alternatively subjected to laser capture using a Pixcell™ laser capture microdissection system (Arcturus Engineering Inc.) with image archiving workstation. Whole sections were captured using a 60 μm diameter laser beam typically at 30–80 mW power with pulse duration of 50 ms and machine gun mode with a laser firing

frequency of 1 shot per 500 ms. Typically 1000–3000 shots were taken per cap. Caps containing tissue were used to seal Eppendorf tubes containing 100 μL of lysis buffer (containing Pefabloc) every 30 min, with tissue being solubilised by inversion of the tube between introduction of successive caps. As a preliminary demonstration of the protein profile obtained following dissection of specific areas of tissue, cervix was used because the depth of the epithelial layer was compatible with the area of the minimal laser beam (30 μm). Collection of protein was essentially as described above with processing and subsequent LCM of whole sections or, specifically, the epithelial layer.

2.4 Sample preparation and electrophoresis

Sample lysis buffer was based on the urea/thiourea mix previously described [13] but contained 8 M urea, 2 M thiourea, 2% w/v CHAPS, 1% w/v DTT, 0.8% Pharmalyte, pH 3–10. Samples were centrifuged at 42 000 × *g* for 1 h at 20°C, the supernatant aliquoted and stored at –80°C until analysis. Protein concentration was determined with a Bio-Rad protein assay kit. First-dimensional electrophoresis was carried out using a Multiphor horizontal electrophoresis system with 18 cm immobilised pH gradient (IPG) strips (pH range 3–10 nonlinear or 4–7 linear) and sample being applied overnight using the in-gel rehydration method as previously described [14]. The rehydration solution contained 8 M urea, 2 M thiourea, 4% w/v CHAPS, 0.46% w/v DTT, 20 mM Tris, 0.2% Pharmalyte, pH 3–10, and a total sample volume of 450 μL containing 30–80 μg of protein was loaded per strip. Focusing was carried out at 20°C for a maximum of 50 μA/strip with 150 V for 30 min followed by gradient increases to 300 V during the next hour, 600 V for the next hour and increasing to 3500 V over the next 2 h. Total focusing was for 50–60 kVh. Second-dimensional electrophoresis was carried out with Bio-Rad Protean II xi vertical electrophoresis systems using SDS-PAGE gels of 1 mm thickness (12%T, 2.6%C) with a 4%T stacking gel. IPG strips were equilibrated for 2 × 15 min in 6 M urea, 30% v/v glycerol, 2% w/v SDS in 0.05 M Tris-HCl, pH 6.8, containing 1% w/v DTT and 4% w/v iodoacetamide for the first and second period of equilibration, respectively. Strips placed on the vertical gels were overlayed with 1% w/v agarose and subjected to electrophoresis at 15 mA/gel overnight at 10°C using a Tris-glycine buffer. Following electrophoresis, gels for analysis were fixed and stained using the Plus-One silver staining kit (Pharmacia) and scanned using a GS-700 scanning densitometer (Bio-Rad). Gels for sequencing were stained in 0.1% w/v Coomassie blue in 30% methanol, 10% acetic acid, 60% water for 5 min and destained for 10–20 min, spots were excised using disposable scalpels and placed into sterile microfuge tubes for subsequent sequencing.

2.5 Western blotting

For Western blotting, 7 cm pH 3–10 (NL) IPG strips were equilibrated like the 18 cm strips (*i.e.* Section 2.4) but each was loaded with 10 μg of protein from a renal carcinoma tissue sample. Parallel samples were loaded which had either been sectioned directly into lysis buffer or sectioned, stained and subjected to LCM as described. Focusing was at 3000 V (linear increase over the first hour) for a total of 5.25 kVh. For the second dimension, mini gradient SDS-polyacrylamide gels (8–16%) were electrophoresed using Tris-glycine buffer at 125 V for 1 h 45 min using a Novex XCell™ Mini-Cell system. Rainbow molecular weight markers were coelectrophoresed. Proteins were transferred to Hybond-C nitrocellulose membranes using the Novex system for 1 h at 25 V with a transfer buffer of 12 mM Tris-base, 96 mM glycine, 20% v/v methanol, pH 8.3. Blots were blocked for 1 h in Tris-buffered saline, pH 7.6, containing 0.1 % v/v Tween-20 (TBS-T) and 10% w/v dried milk. Blots were then incubated in TBS-T, 1% dried milk containing goat anti-HSP-60 (1/5000) and rabbit anti-β_2-microglobulin (1/500) for 1 h, washed in TBS-T and incubated sequentially in TBS-T, 1% dried milk containing biotinylated swine anti-rabbit immunoglobulins (1/2000) and biotinylated rabbit anti-goat immunoglobulins (1/2000), each for 1 h. Following further washing and incubation for 45 min in HRP-conjugated streptavidin (1/3000), blots were developed using the ECL system (Amersham) and Hyperfilm™ ECL film.

2.6 Peptide mass fingerprinting and microsequencing

Proteins from a single renal carcinoma sample were examined in parallel gels following either sectioning into lysis buffer or sectioning, staining and LCM as described above. Following SDS-PAGE electrophoresis, the stained protein spots were excised from the gel and digested with trypsin essentially as previously described [15]. Following overnight digestion, 0.5 μL was sampled directly from the digest supernatants for MS analysis using a TofSpec 2E laser-desorption time of flight MS (Micromass, Manchester, UK). Peptide masses were screened against the OWL nonredundant protein database (270 000 entries) using the MOWSE search software [16]. The remaining digested peptides (> 90% of total digest) were extracted with washes of 5% v/v aq. formic acid and acetonitrile, pooled and dried. For sequence analysis, dried peptides were then derivatised with *N*-succinimidyl-2-morpholine acetate (SMA) to improve b-ion abundance and facilitate sequence analysis by tandem mass spectrometry [17].

Dried peptide fractions were treated with 7 μL of freshly prepared, ice-cold 1% w/v *N*-succinimidyl-2-morpholine acetate in 1.0 M HEPES (pH 7.8 with NaOH) containing 2% v/v acetonitrile. Following reaction for 20 min on ice, the reaction was terminated by the addition of 1 μL heptafluorobutyric acid (HFBA) and diluted with an equal volume of water. The solution was then injected in 3 × 5 μL aliquots onto a capillary reverse-phase column (300 μm × 15 cm) packed with POROS R2/H material (PerSeptive Biosystems, Framingham, MA, USA) equilibrated with 2% v/v methanol / 0.05% v/v TFA running at 3 μL/min. The adsorbed peptides were washed isocratically with 15% v/v methanol / 0.05% v/v TFA for 30 min at 3 μL/min to elute the excess reagent and HEPES buffer. Derivatised peptides were eluted with a single-step gradient to 75% v/v methanol / 0.1% v/v formic acid and collected in two 3 μL fractions. Derivatised peptides were then fully sequenced by low-energy collision-activated dissociation (CAD) using a Finnigan MAT LCQ ion-trap MS fitted with a nanoelectrospray source [18, 19]. CAD was typically performed with collisional offset voltages between –17 and –35 V.

3 Results

The LCM system is shown in Fig. 1 together with a diagrammatic representation of the tissue dissection mechanism. Examples of the types of microdissection possible using this equipment are shown in Fig. 2. Importantly, using the specimens and protocols described here, no gross effects were apparent on protein profiles following either the staining procedure or the staining procedure and LCM although some differences could be seen (Fig. 3). Similarly, retention of antigenicity of the two proteins examined in this study, HSP-60 and β_2-microglobulin, was not affected by the tissue preparation or laser treatment (Fig. 4). Nine of the ten protein spots excised from the gel were successfully identified in both the control and laser-captured sample (Fig. 5, Table 1). One protein was not identified in either sample. Mass spectrometric profiles were practically identical between control and LCM samples, as exemplified in Fig. 6 for actin and ATP synthase. For these two proteins selected for more detailed examination, the peptides present and sequenced were

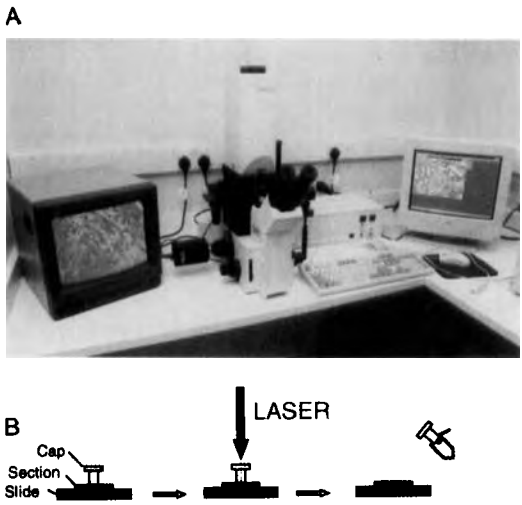


Figure 1. (A) Pixcell laser capture microdissection apparatus showing the inverted microscope with laser attachment attached to a viewing screen and image acquisition computer, and (B) diagrammatic representation of the laser-mediated capture of cells from a tissue section on the microscope stage. A cap containing the UVA polymer film on its undersurface is placed on top of a tissue section on a glass slide and the laser beam is focused on the cells of interest. Following firing of the laser, the cap containing the adherent cells is lifted from the tissue surface using a mechanical arm and placed into a microfuge tube containing lysis buffer.

Table 1. Proteins identified by peptide mass fingerprinting (PMF) and/or mass spectrometric microsequencing in extracts of normal kidney either following direct extraction or LCM as described

Number	PMF	Sequencing	Protein
1	+	+	Albumin
2	+	+ (4)	Albumin
3	+	+ (12)	Beta-actin
4	+	–	Beta-actin
5	+	–	ATP synthase beta-chain
6	+	–	Alpha-tubulin
7	+	+ (5)	HSP-60
8	+	+ (2)	Apolipoprotein A1
9	–	+ (6)	Lactate dehydrogenase H-chain
10	–	–	Not identified

Numbers in parenthesis are the number of peptides fully sequenced from the respective proteins (all from the LCM samples)

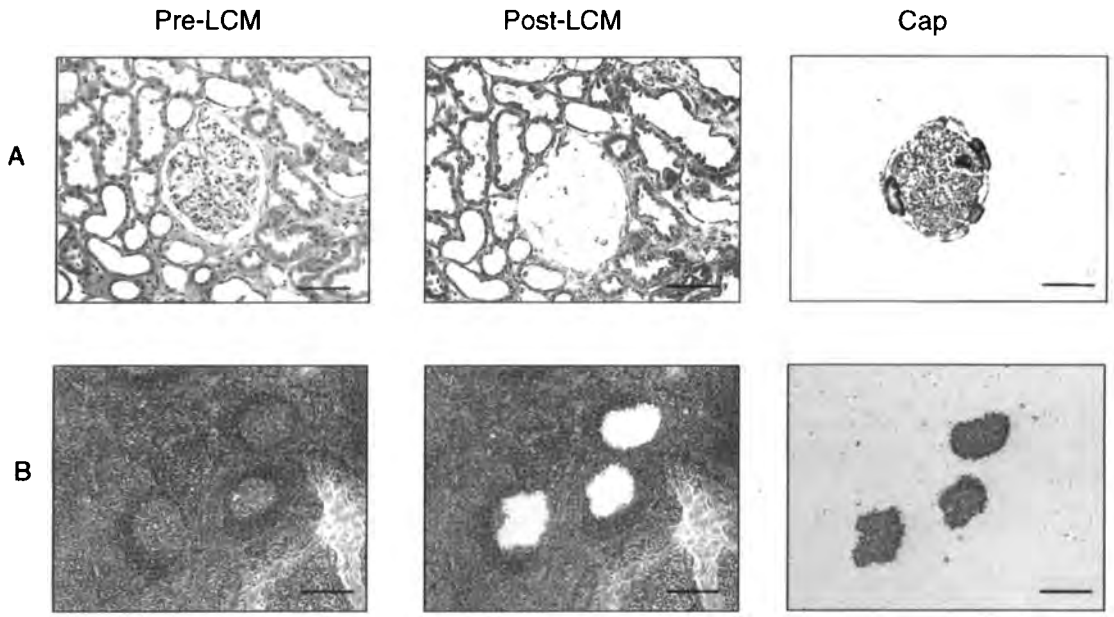


Figure 2. Examples of laser-capture microdissection of (A) a renal glomerulus from kidney and (B) germinal centres from lymphoid tissue. Shown are the appearance of the tissue section prior to and following capture and the captured cells present on the cap. The bar represents 100 μm.

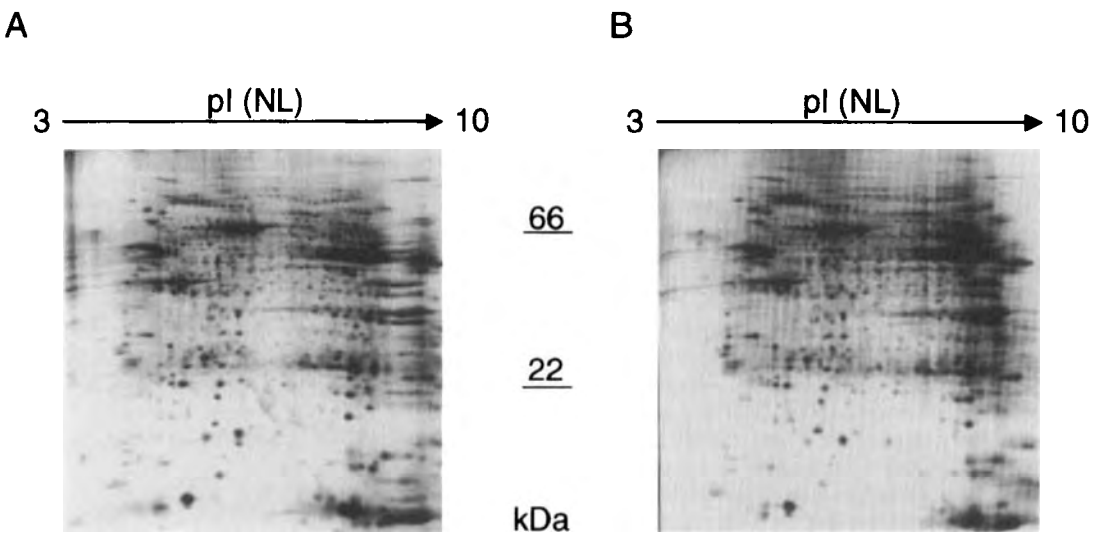


Figure 3. Representative example of a 2-D PAGE comparison between (A) control and (B) laser-treated tissue samples. Specimen shown is normal renal cortical tissue prepared as described with 70 μg protein per gel.

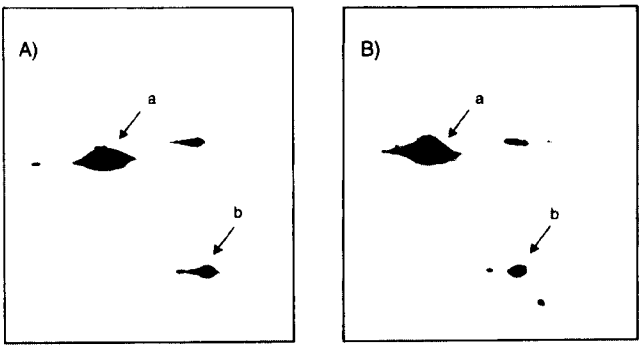


Figure 4. 2-D Western blot of a comparison between (A) control and (B) laser-treated tissue from a normal renal specimen. Arrows indicate a, HSP-60 and b, β_2-microglobulin.

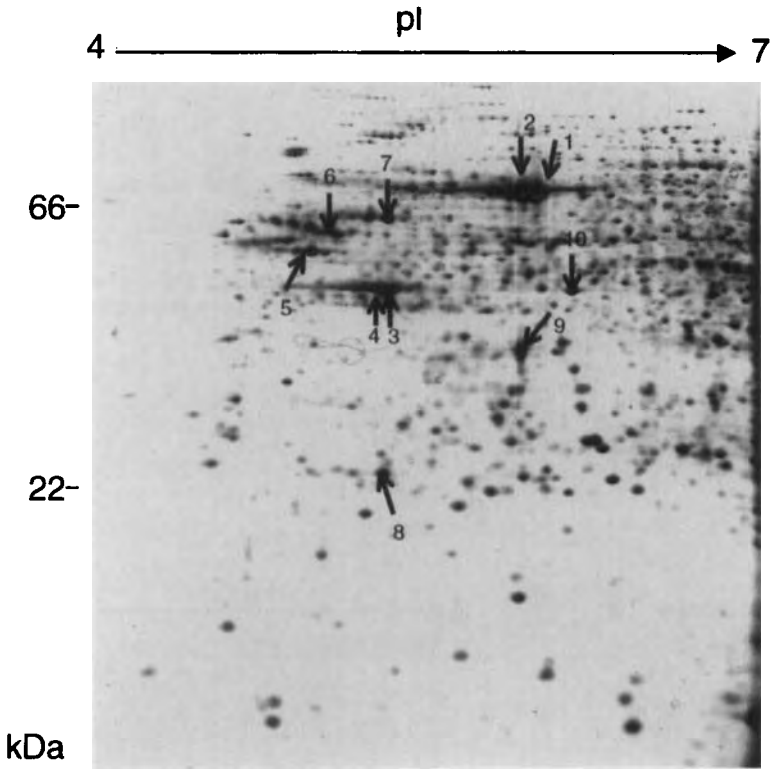


Figure 5. 2-D PAGE gel of normal renal tissue (40 µg protein) with arrows indicating the location of proteins identified by peptide mass fingerprinting/mass spectrometric sequencing following laser capture.

identical, in both the control and laser-treated/processed samples (with the exception of peptide 295–310 of ATP synthase which was not detected in the control sample). The extent of partial oxidation of methionine residues was also unchanged between LCM and control samples (Fig. 6 and Table 2). Fifty sections of cervix were successfully dissected and the epithelium was removed from the underlying stroma and pooled (Fig. 7). A comparison of the protein profiles of whole sections *versus* epithelium clearly shows the comparable presence of many proteins, some of which were tentatively identified on the basis of molecular weight and p*I* as housekeeping proteins such as actin and glyceraldehyde 3-phosphate dehydrogenase. However, enrichment of some proteins in the gels from epithelial tissue, which were either weakly detectable or undetectable when whole tissue sections were examined, can clearly be seen.

4 Discussion

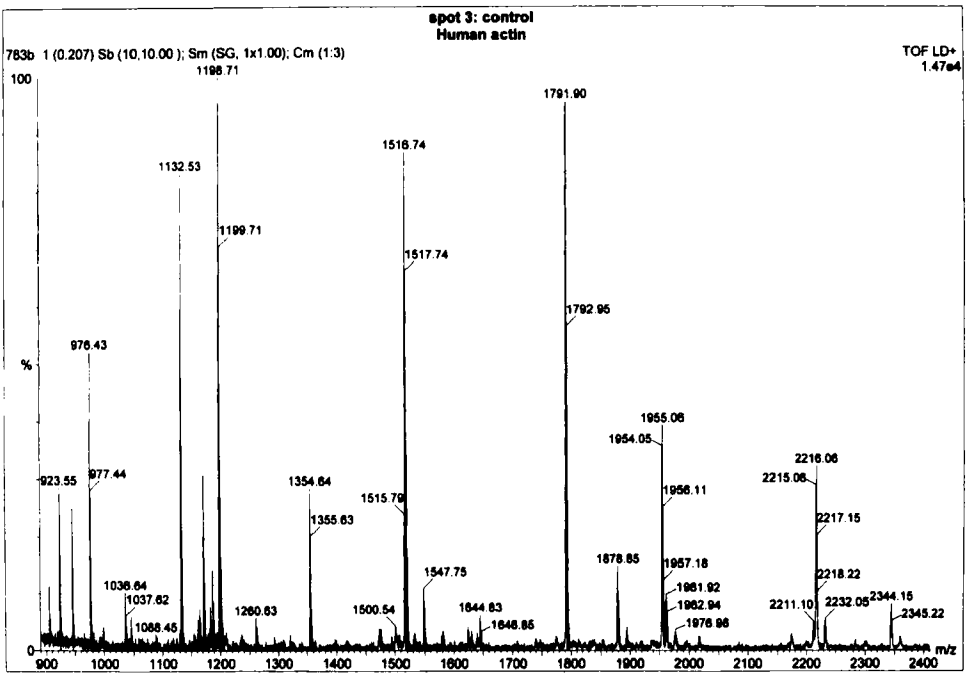
To our knowledge this is the first study to have examined the effects of LCM on the protein profiles and properties

Table 2. Expected tryptic peptides in the MALDI spectrum of the proteins a) actin and b) ATP synthase and their presence or absence in control and laser-treated material from a normal renal tissue specimen

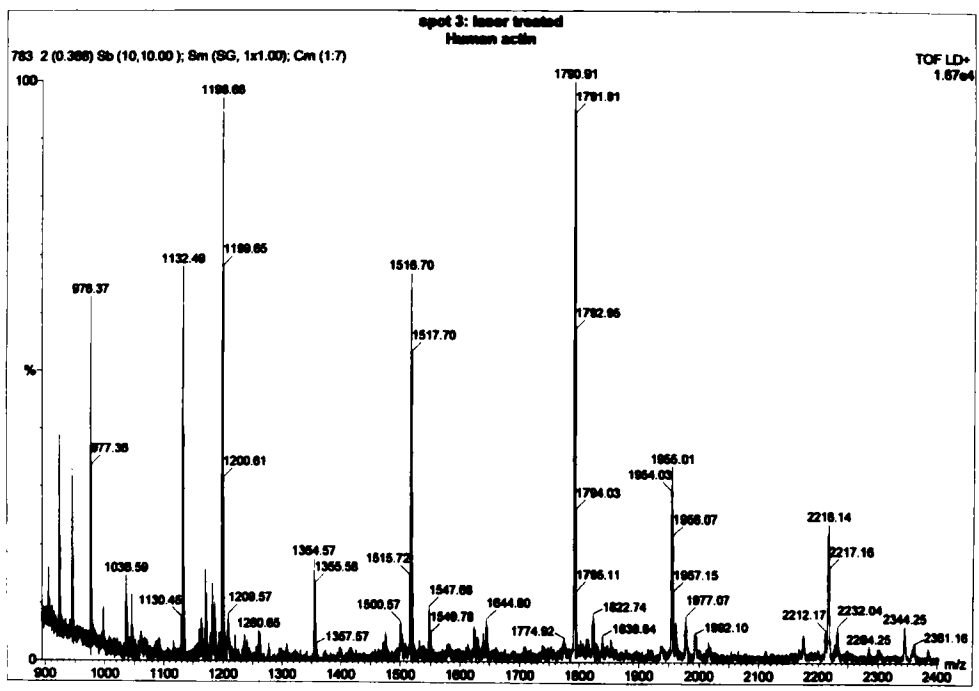
	Expected peptides	Position	Peptide mass [M+H] (Da)	Observed peptides Control	Laser treated
a)	LCYVALDFEQEMATAASSSSLEK	216 – 238	2493.153	+	+
	DLYANTVLSGGTTMYPGIADR	292 – 312	2215.070	+	+
			2231.065 [O]	+	+
	VAPEEHPVLLTEAPLNPK	96 – 113	1954.065	+	+
	YPIEHGIVTNWDDMEK	69 – 84	1946.896	–	–
	MDDDIAALVVDNGSGMCK	1 – 18	1853.808	–	–
	SYELPDGQVITIGNER	239 – 254	1790.82	+	+
	QEYDESGPSIVHR	360 – 372	1516.703	+	+
	IWHHTFYNELR	85 – 95	1515.750	+	+
	AVFPSIVGRPR	29 – 39	1198.523	+	+
	DSYVGDEAQSK	51 – 61	1198.523	+	+
	HQGVMVGMGQK	40 – 50	1171.572	+	+
			1187.573 [O]	+	+
	EITALAPSTMK	316 – 326	1161.619	–	–
	GYSFTTTAER	197 – 206	1132.527	+	+
	DLTDYLMK	184 – 191	998.487	+	+
	AGFAGDDAPR	19 – 28	976.449	+	+
	IIAPPER	329 – 335	795.473	+	+
	CDVDIR	285 – 290	720.335	–	–
	GILTLK	63 – 68	644.435	+	+
	LDLAGR	178 – 183	644.373	+	+
	ILTER	192 – 196	631.378	+	+
b)	IPSAVGYQPTLATDMGTMQER	325 – 345	2266.050	+	+
			2282.070 [O]	+	+
			2298.074 [O]	+	+
	SLQDIIAILGMDELSEEDK	433 – 451	2119.048	–	–
	EGNDLYHEMIESGVINLK	242 – 259	2060.996	+	+
			2076.991 [O]	+	+
	FLSQPFQVAEVFTGHMGK	463 – 480	2023.011	+	+
			2039.006 [O]	+	+
	AIAELGIYPAVDPLDSTSR	388 – 406	1988.034	+	+
	DQEGQDVLLFIDNIFR	295 – 310	1921.996	–	+
	IMDPNIVGSEHYDVAR	407 – 422	1815.870	+	+
			1831.865 [O]	+	+
	LVLEVAQHLGESTVR	95 – 109	1650.918	+	+
	VALVYGQMNEPPGAR	265 – 279	1601.811	+	+
			1617.806 [O]	+	+
	LTPSASLPPAQLLLR	20 – 34	1576.943	–	–
	TVLIMELINNVAK	213 – 225	1457.840	+	+
			1473.835	+	+
	VALTGLTVAEYFR	282 – 294	1439.790	+	+
	FTQAGSEVSALLGR	311 – 324	1435.754	+	+
	AHGGYSVFAGVGER	226 – 239	1406.682	+	+
	IMNVIGEPIDER	144 – 155	1385.710	+	+
			1401.705 [O]	+	+
	TIAMDGTEGLVR	110 – 121	1262.641	+	+
			1278.636 [O]	+	+
	DYAAQTSPSPK	45 – 55	1164.554	–	–
	VVDLLAPYAK	189 – 198	1088.636	+	+
	IPVGPETLGR	134 – 143	1038.595	+	+
	AAPTAVHPVR	35 – 44	1018.580	–	–
	VAAAPASGALR	8 – 18	983.564	–	–
	IGLFGGAGVGK	202–212	975.563	+	+
	VLDSGAPIK	125–133	899.520		

[O] denotes a peptide containing an oxidised methionine residue.

a


spot 3: control
Human actin
783b 1 (0.207) Sb (10,10.00); Sm (SG, 1x1.00); Cm (1:3)
TOF LD+
1.47e4
1198.71
1791.90
1516.74
1132.53
1199.71
1517.74
1792.95
976.43
1955.06
1954.05
2216.06
2215.06
923.55
977.44
1354.64
1515.79
1956.11
1355.63
2217.15
1678.85
1957.18
1547.75
2218.22
1036.64
1037.62
1961.92
1962.94
1260.63
1500.54
1644.63
1646.85
2211.10
2232.05
2344.15
2345.22
1088.45
1976.98
%
m/z

b


spot 3: laser treated
Human actin
783 2 (0.366) Sb (10,10.00); Sm (SG, 1x1.00); Cm (1:7)
TOF LD+
1.67e4
1198.66
1790.91
1791.91
1132.49
1199.65
976.37
1516.70
1517.70
1792.95
977.36
1200.61
1956.01
1964.03
1794.03
1958.07
2218.14
1354.57
1515.72
2217.16
1038.59
1795.11
1957.15
1547.68
1209.57
1130.46
1500.57
1644.80
1822.74
1977.07
1849.78
1280.65
1774.92
1638.84
1982.10
2212.17
2232.04
2344.25
1357.57
2284.25
2381.16
%
m/z

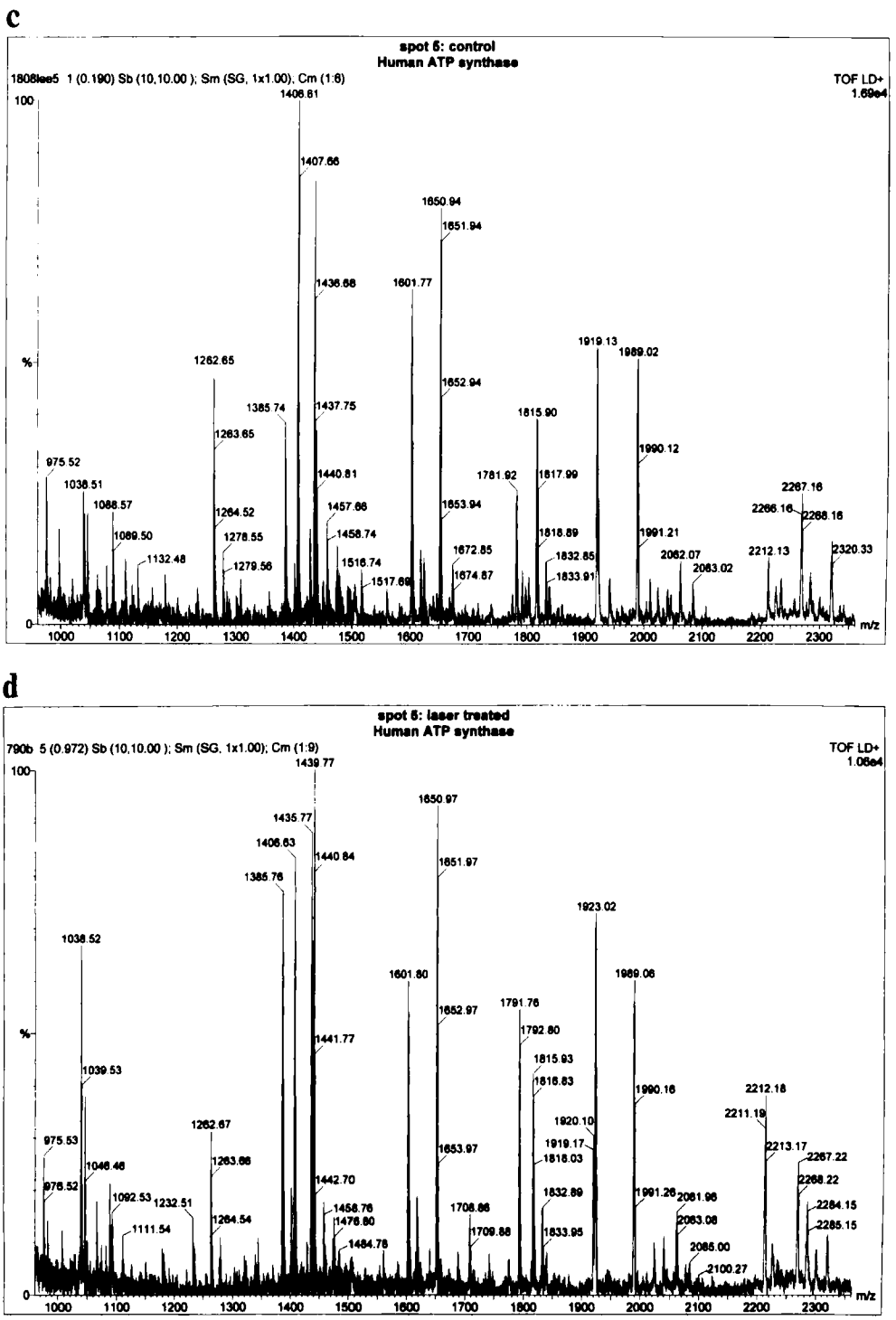


Figure 6. Mass spectrometric profiles of spots 3 and 5 (actin and ATP synthase, respectively). Profiles are shown for normal renal tissue under control conditions (a, c) and following laser capture (b, d).

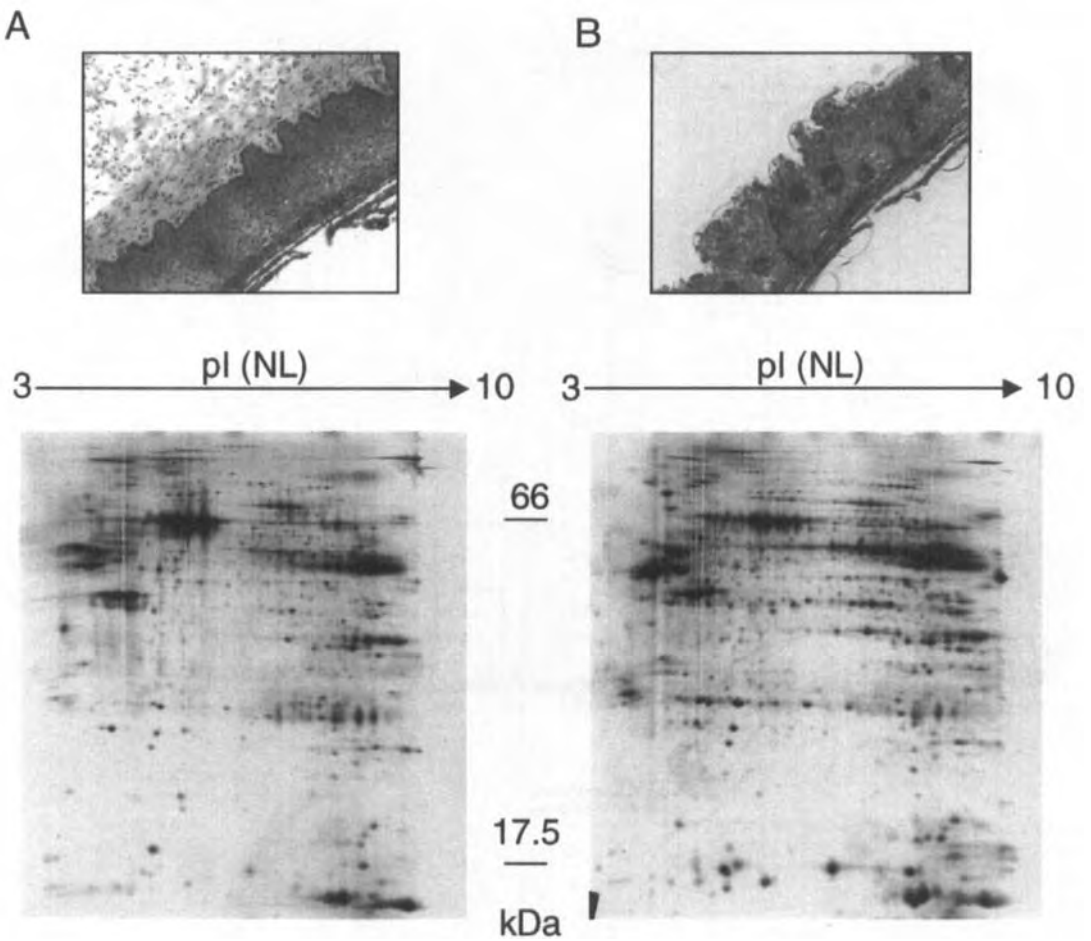


Figure 7. 2-D PAGE gels of normal cervix tissue (40 μg protein) with a representative micrograph of the tissue sample above each gel; (A) represents the pattern obtained with whole tissue compared with (B) where the epithelial tissue has been selectively laser-captured.

which we have examined. The results described here are extremely promising and indicate that laser capture microdissection may offer a viable approach to the enrichment of particular cell types prior to subsequent proteomic analysis. However, it is clearly essential that further detailed validation is carried out to fully establish its suitability. This would include comparison of protein profiles of further control samples with those subjected to staining and laser treatment and detailed analysis of the results using the sophisticated software packages now available for 2-D electrophoresis work, examination of the antigenicity of a wider range of proteins, and rational refinement of processing and dissection times based on effects on proteins. This is necessary to ensure that proteins are not being degraded or modified in other ways and that recovery from the adhesive film is not selective in any way. The need for protease inhibitors should be explored further and ideally over a range of tissues including malignant tissues, where endogenous proteolytic activity may be expected to be more prevalent, as activity has been observed even in denaturing buffers [20]. Strategies should be explored to allow the pooling and concentration of dissected samples to minimise the time between dissection and freezing of the samples. To confirm the selective enrichment of cell-specific proteins, as potentially indicated here in the example of the dissected cervical epithelium, the identity of the enriched proteins needs to be established. In addition, the degree of contamination of microdissected tumour specimens with interstitial cells can be assessed by examining proteins specific to these cells. Studies such as those described above are currently underway.

In further support of a lack of detrimental effect of this tissue preparation and laser treatment on protein integrity, others have briefly indicated that, based on immunohistochemical and soluble immunoassay results, the structure of cellular proteins appears to be preserved following LCM [12]. Clearly, fresh frozen tissue would be the material of choice. Formalin-fixed paraffin-embedded material is unlikely to be of use due to the chemical cross-linking effects on proteins [21]. The effects of the tissue treatment described here appear to be small but clearly need further investigation. In particular, the binding of the stains haematoxylin and eosin may exert effects on the charge of some proteins although dissociation of such complexes may occur during the denaturation step. It is unlikely that the laser treatment itself would damage the proteins directly as the near-infrared wavelengths are sufficiently far removed from the absorption maxima of proteins, unlike UV lasers previously used for ablation of samples and negative selection [6–8]. However, heat-induced damage remains a possibility although the strength of the laser beam and structure of the transfer film is such that thermal changes should be mild and transient [11].

A major consideration in deciding on the potential use of such a technique has to be the time taken to produce a sufficient amount of protein for subsequent electrophoretic analysis. In the example provided here of cervical epithelium, dissection took approximately 13 h. Clearly, dissection of smaller structures present whithin more complex tissue such as renal proximal tubules, which is now possible with the development of the Pixcell 2 LCM system incorporating a wider range of laser diameters ranging from 5 μm to 30 μm, may be expected to take significantly longer. With such time constraints, it is unlikely that LCM would be a routine tool for protein analysis but rather used on a limited number of samples to provide "master maps" to which nonpurified samples, or samples prepared by, for example, enzymatic disaggregation and immunological selection, could be compared. Once the analytical gels prepared from the microdissected material have been used to illustrate the proteins of potential interest, clearly preparative-type gels of nondissected tissue can be used to prepare the proteins identified for subsequent sequencing.

Proteomics-based approaches have now been used in initial studies of several tumour types including renal, breast, bladder (squamous cell) and lung cancers [22–26] using either whole tissue homogenates or fractions enriched for the cells of interest. From these studies, the potential of such approaches in identifying proteins which may either be involved in the pathogenesis of the disease and/or function as markers is already apparent and it is likely that the use of systems such as the one described here may further facilitate such studies in the future. However, it should also be borne in mind that, in addition to the malignant cell type, a variety of other cells including stromal cells and endothelial cells are involved in tumour development and progression and hence a variety of approaches are likely to be needed depending on the question being addressed.

We acknowledge the funding support of the Imperial Cancer Research Fund, the donation of caps by Arcturus Engineering and the help of Bio-Rad in being able to scan gels. Work in MJD's laboratory is supported by the British Heart Foundation.

Received January 13, 1999

5 References

[1] Franks, L. M., Teich, N. M. (Eds.), *Introduction to the Cellular and Molecular Biology of Cancer*, Oxford University Press, Oxford, UK 1997.

[2] Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P., *Electrophoresis* 1997, *18*, 1217–1242.

[3] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[4] Whetsell, L., Maw, G., Nadon, N., Ringer, D. P., Schaefer, F. V., *Oncogene* 1992, *7*, 2355–2361.

[5] Zhuang, Z., Bertheau, P., Emmert-Buck, M. R., Liotta, L. A., Gnarra, J., Linehan, W. M., Lubensky, I. A., *Am. J. Pathol.* 1995, *146*, 620–625.

[6] Meier-Ruge, W., Bielser, W., Remy, E., Hillenkamp, F., Nitsche, R., Unsold, R., *Histochem. J.* 1976, *8*, 387–401.

[7] Shibata, D., Hawes, D., Li, Z. H., Hernandez, A. M., Spruck, C. H., Nichols, P. W., *Am. J. Pathol.* 1992, *141*, 539–543.

[8] Becker, I., Becker, K.-F., Rohrl, M. H., Minkus, G., Schulze, K., Hofler, H., *Lab. Invest.* 1996, *75*, 801–807.

[9] Bohm, M., Wieland, I., Schutze, K., Rubben, H., *Am. J. Path.* 1997, *151*, 63–67.

[10] Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., Liotta, L. A., *Science* 1996, *274*, 998–1001.

[11] Bonner, R. F., Emmert-Buck, M. R., Cole, K., Pohida, T., Chuaqui, R., Goldstein, S., Liotta, L. A., *Science* 1997, *278*, 1481–1483.

[12] Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R., Liotta, L. A., *Trends Genet.* 1998, *14*, 272–276.

[13] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[14] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[15] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[16] Pappin, D. J. C., Hojrup, P., Bleasby, A. J., *Curr. Biol.* 1993, *3*, 327–332.

[17] Sherman, N. E., Yates, N. A., Shabanowitz, J., Hunt, D. F., Jeffery, W. A., Bartlet-Jones, M., Pappin, D. J. C., *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, Georgia 1995, 626.

[18] Hunt, D. F., Yates III, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., *Proc. Natl. Acad. Sci. USA* 1986, *84*, 6233–6237.

[19] Wilm, M., Mann, M., *Anal. Chem.* 1996, *68*, 1–8.

[20] Rabilloud, T., *Electrophoresis* 1996, *17*, 813–829.

[21] Hopwood, D., *Histochem. J.* 1968, *1*, 323–360.

[22] Sarto, C., Marocchi, A., Sanchez, J.-C., Giannone, D., Frutiger, S., Golaz, O., Wilkins, M. R., Doro, G., Cappellano, F., Hughes, G., Hochstrasser, D. F., Mocarelli, P., *Electrophoresis* 1997, *18*, 599–604.

[23] Franzen, B., Linder, S., Alaiya, A. A., Erikson, E., Fujioka, K., Bergman, A. C., Jornvall, H., Auer, G., *Electrophoresis* 1997, *18*, 582–587.

[24] Bini, L., Magi, B., Marzocchi, B., Arcuri, F., Tripodi, S., Cintorino, M., Sanchez, J.-C., Frutiger, S., Hughes, G., Pallini, V., Hochstrasser, D. F., Tosi, P., *Electrophoresis* 1997, *18*, 2832–2841.

[25] Østergaard, M., Rasmussen, H. H., Nielsen, H. V., Vorum, H. Ørntoft, T. F., Wolf, H., Celis, J. E., *Cancer Res.* 1997, *57*, 4111–4117.

[26] Okuzawa, K., Franzén, B., Lindholm, J., Linder, S., Hirano, T., Bergman, T., Ebihara, Y., Kato, H., Auer, G., *Electrophoresis* 1994, *15*, 382–390.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 701–704

Mark P. Molloy
Ben R. Herbert
Keith L. Williams
Andrew A. Gooley

Australian Proteome Analysis Facility, School of Biological Sciences, Macquarie University, Sydney, Australia

Extraction of *Escherichia coli* proteins with organic solvents prior to two-dimensional electrophoresis

Compared to soluble proteins, hydrophobic proteins, in particular membrane proteins, are an underrepresented protein species on two-dimensional (2-D) gels. One possibility is that many hydrophobic proteins are simply not extracted from the sample prior to 2-D gel separation. We attempted to isolate hydrophobic proteins from *Escherichia coli* by extracting with organic solvents, then reconstituting the extracted proteins in highly solubilising sample solution amenable to 2-D electrophoresis using immobilized pH gradients (IPGs). This was conducted by an extraction with a mixture of chloroform and methanol, followed by solubilisation using a combination of urea, thiourea, sulfobetaine detergents and tributyl phosphine. Peptide mass fingerprinting assisted in the identification of 13 proteins, 8 of which have not previously been reported on 2-D gels. Five of these new proteins possess a positive hydropathy plot. These results suggest that organic solvent extractions may be useful for selectively isolating some proteins that have previously been missing from proteome maps.

Keywords: Two-dimensional electrophoresis / Proteome / Solubility / Organic solvent / Extraction
EL 3395

In recent times the resurgence of two-dimensional electrophoresis (2-DE) as a powerful array technology has been driven largely by proteomics [1]. The aim of proteomics is firstly to display in a parallel means, an organism's protein content using an array technology such as 2-DE, then secondly to identify and characterise those proteins of interest. Thus, the need for high resolution and reproducibility in proteomics has stimulated advances in gel technology including widespread use of immobilised pH gradients (IPGs) for isoelectric focusing [2], improvements in methods for enhancing protein solubility [3–5] and sample loading techniques [6], and very recently, reports of alkaline pH IPGs offering reproducible separation of basic proteins up to pH 12 [7]. While such developments have improved the separation of many proteins, it has become apparent that hydrophobic proteins are under-represented on 2-D gels from even the most extensively studied species, *Escherichia coli* and *Saccharomyces cerevisiae* [8]. Explanations include protein precipitation in the IPG matrix [9] and inadequate detection and analysis methods for proteins in low abundance [10]. While these are valid observations they are unlikely to account for the almost complete absence of hydrophobic proteins on 2-D maps. A further possibility to explain the situation is that standard sample preparation techniques prior to array by 2-DE are not optimal for extraction of hydrophobic proteins. A consideration often overlooked is the functional interactions proteins have in the cell. Protein-protein interaction and interactions with other biostructures such as the phospholipid bilayer, extracellular matrix, cell wall and cytoskeletal network all are likely to contribute in some degree to impaired protein extraction. Considering this and the chemical nature of hydrophobic proteins, it is likely that many hydrophobic proteins are simply not soluble in standard denaturing solutions used for 2-DE and thus not extracted from the sample. Recently we have shown that this is the case for *E. coli* porin outer membrane proteins (OMPs) which are absent from most 2-D gel maps, even though they are abundant molecules and possess hydrophilic hydropathy plots [5]. There are reports that indicate some highly hydrophobic proteins are soluble in solutions of organic solvents [11, 12]. We report our search for hydrophobic proteins in *E. coli* by extracting whole lyophilised *E. coli* in a 1:1 v/v solution of chloroform and methanol prior to solubilisation for 2-DE.

Correspondence: Prof. K. L. Williams, APAF, Proteome Systems Ltd, Locked Bag 2073, North Ryde, Sydney, NSW Australia 1670
E-mail: keith.williams@proteomesystems.com
Fax: +61-2-9889-1805

Abbreviations: ASB14, amidosulfobetaine 14; **GRAVY,** grand average hydropathy; **OMPs,** outer membrane proteins; **PMF,** peptide mass fingerprinting; **SB 3–10,** *N*-decyl-*N*,*N*-dimethyl-3-ammonio-1-propane sulfonate; **TBP,** tributyl phosphine

E. coli K-12, strain W3110, was cultured and prepared as previously described [13, 14]. For extraction with organic solvents, 40 mg of lyophilised *E. coli* was incubated in 1.5 mL of a 1:1 v/v solution of chloroform and methanol, mixed thoroughly, then placed in an ultrasonic waterbath for 5 min before incubating on ice for a further 30 min. For phase separation, 600 µL of MilliQ water was added to the sample, mixed and left to stand at 25°C for 5 min. The tube was then centrifuged for 10 min in a standard bench-

top centrifuge. The upper aqueous layer was discarded and the organic layer removed and dried by rotary evaporation. For 2-DE, the dry organic solvent extracted sample was reconstituted with 120 μL of either solubilisation solution A (5 M urea, 2 M thiourea, 2 % w/v CHAPS, 2 % w/v SB 3–10, 2 mM TBP, 40 mM Tris-base, 0.5 % v/v Biolytes 3–10 (Bio-Rad, Richmond, CA) or solubilisation solution B (7 M urea, 2 M thiourea, 1 % w/v ASB14 (a gift from T. Rabilloud), 2 mM TBP, 40 mM Tris-base, 0.5 % v/v Biolytes 3–10 (Bio-Rad). IEF was conducted using 7 cm pH 3–10 NL (nonlinear) IPGs (Amersham Pharmacia, Uppsala, Sweden) loaded with 120 μL of solubilisation solution by rehydration as previously described [6]. IEF was conducted at a maximum of 5000 V for a total of 15 000 Vh. For SDS-PAGE, Bio-Rad 10–20% Ready gels were used. These were run at 5 mA per gel for 30 min, then 12 mA per gel for 2–3 h. All other conditions remained as previously described [5]. Gels were stained overnight with 0.1 % w/v Coomassie Blue G-250 in 17 % w/v ammonium sulfate, 34 % v/v methanol and 3 % v/v *o*-phosphoric acid. Excised protein spots were subjected to MALDI-TOF-MS for identification as described previously [15].

Organic solvents are widely used in biochemistry for protein precipitation. With this in mind we anticipated only a limited number of proteins to be extracted under these nonaqueous conditions. Consequently we used a large amount of starting sample to ensure that sufficient protein was extracted to enable protein identification by MS following electrophoretic separation. The proteins extracted with organic solvents were solubilised for 2-DE using solutions similar to those that had previously shown good results with insoluble samples. Following the introduction of thiourea and sulfobetaines to standard CHAPS-based solutions, improvements were reported for solubilising various membrane proteins [3]. Furthermore, by the use of a novel amidosulfobetaine surfactant (ASB14) in the solubilising solution, the overall chaotrope concentration can be increased to 9 M, which has demonstrated a dramatic improvement in the separation of membrane proteins from bovine neutrophils [16]. In this work, these solutions were amended to include tributyl phosphine as the reducing agent. The separations achieved using solubilising solutions A and B with mini gels are shown in Fig. 1a and 1b. Good separation was achieved in both dimensions, suggesting the maintenance of high protein solubility. Few differences in the spot patterns between the gels were observed, although larger protein spots, indicative of more protein, were detected when ASB14 was included in the solubilisation solution (Fig. 1b). In this case, compared to solubilisation solution A, the combination of a strong detergent compatible with high chaotrope molarity increased protein solubility, probably by decreasing the loss of protein often observed with IEF of preparatively

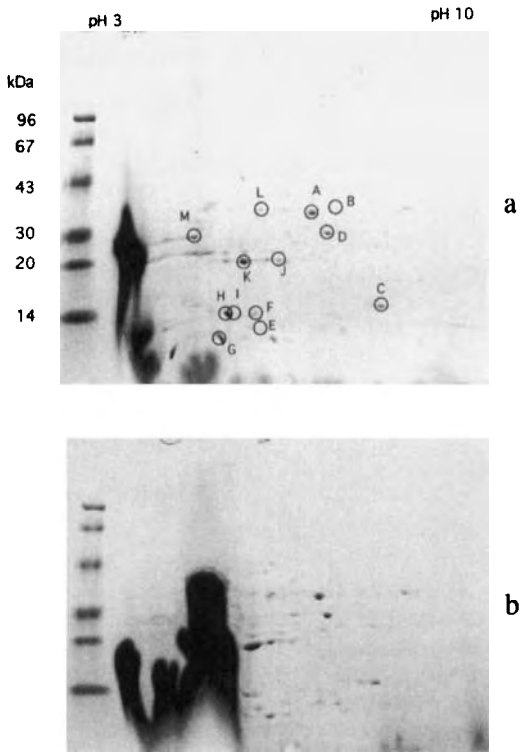


Figure 1. Coomassie Blue G-250 stained mini 2-DE of *E. coli* proteins extracted with organic solvent, then reconstituted with (a) 5 M urea, 2 M thiourea, 2 % CHAPS, 2 % SB 3–10, 2 mM TBP or (b) 7 M urea, 2 M thiourea, 1 % ASB14, 2 mM TBP. Circled spots refer to proteins identified and shown in Table 1.

loaded IPGs. The greater protein solubility with ASB14 (solution B) came at the cost of resolution as there was a concomitant increase in the solubilisation of lipid, visualised with Coomassie as the dark smear in the acidic region of the gel (Fig. 1b). The presence of lipid was not unexpected considering our original extraction protocol using organic solvents.

Excised protein spots were subjected to MALDI-TOF-MS; a list of proteins identified by peptide mass fingerprinting (PMF) is shown in Table 1. Grand average hydropathy (GRAVY) values of proteins are presented to indicate overall protein hydropathy. Proteins with positive GRAVY values are recognised as hydrophobic and are generally lacking from lists of proteins identified on 2-D gels. According to Wilkins *et al.*, [8] the highest positive GRAVY value identified on a 2-D gel from *E. coli* and indexed in SWISS-PROT has a value of +0.30 (PFS_ECOLI). Theo-

Table 1. Organic solvent extracted *E. coli* proteins identified by peptide mass fingerprinting

Spot	SWISS-PROT Accession No.	Description	Subcellular location	GRAVY	Residues of identified peptides	Peptides matched *vs.* expected[a)]	Theoretical *p*I and molecular mass (kDa)	Experimental *p*I and molecular mass (kDa)	Notes
A	P03841	Maltose operon periplasmic protein	Periplasmic	–0.021	157–166, 167–179, 184–198, 252–268	4/4	5.9 30	5.7 36	New ID on 2-D gel
B	P00479	Aspartate Carbamoyl transferase	Cytoplasmic	–0.111	8–17, 43–54, 57–65, 66–83, 152–164, 168–178, 280–296, 297–306	8/12	6.1 34	6.0 38	
C	P23875	Lipopolysaccharide biosynthesis protein	Probably cytoplasmic	+0.006	5–24, 25–42, 43–51, 80–88, 92–107, 108–122, 123–133, 138–157	7/7 (1)	6.5 18	6.7 16	New ID on 2-D gel
D	P02925	Ribose-binding Periplasmic protein	Periplasmic	–0.031	55–70, 71–81, 85–101, 102–115, 151–164, 167–178, 179–191, 232–247, 254–268	9/11	6.0 28	6.0 31	New ID on 2-D gel
E	P45502	Unknown protein from 2-D gel	Unknown	–0.103	45–55, 56–62, 103–116, 122–136, 137–157	3/4 (2)	5.5 14	5.4 12	
F	P11555	Fucose operon FUCU protein	Cytoplasmic	+0.179	4–15, 91–104, 117–126	3/3	5.6 15	5.4 14	New ID on 2-D gel
G	P39330	13.5 kDa protein	Unknown	+0.072	38–50, 118–126	2/2	5.4 13	5.2 11	
H	P52100	Hypothetical 20.9 kDa protein	Unknown	–0.379	32–43, 40–49, 60–72, 95–110	4/8	4.9 21	5.2 14	New ID on 2-D gel
I	P25540	Riboflavin synthase beta chain	Cytoplasmic	+0.294	1–14, 22–39, 73–84, 137–152	4/4	5.2 16	5.2 14	New ID on 2-D gel
J	P19245	ATP-dependent CLP protease	Cytoplasmic	–0.154	30–36, 163–180, 187–206	1/2 (2)	5.5 23	5.5 20	New ID on 2-D gel
K	P09157	Superoxide dismutase	Cytoplasmic	–0.180	1–11, 12–29, 30–43, 51–57, 92–107, 108–116, 117–127	5/5 (2)	5.6 21	5.3 20	
L	P06994	Malate dehydrogenase	Cytoplasmic membrane-associated	+0.194	88–99, 143–153, 218–233, 241–262, 263–272, 280–300, 302–312	5/8 (2)	5.6 32	5.4 38	
M	P28635	Hypothetical lipoprotein	Membrane-associated	–0.238	35–51, 79–97, 103–119, 179–189, 224–234, 246–261	6/7	4.9 27	5.0 30	New ID on 2-D gel

a) Number of tryptic peptides predicted between 1000–2200 Da considering no missed cleavages. Number in parentheses indicates number of additional peptides matched with a mass <1000 Da or >2200 Da.

retically, the *E. coli* proteome contains proteins with GRAVY values from –1.95 to +1.88. In this study we identified 13 proteins, 8 of which have not previously been reported in SWISS-PROT as being identified on 2-D PAGE. Five of these 8 proteins have positive GRAVY values (38 % of the identified proteins) with the lowest GRAVY value of the remaining identified proteins being –0.379. Our highest GRAVY value was +0.294, extremely close to the most hydrophobic *E. coli* protein thus far detected on 2-D gels. The proteins identified here can be considered quite hydrophobic as they rank favourably in terms of hydrophobicity when compared to some classical integral membrane proteins. For example, murine erythrocyte Band 3 is a Type III membrane protein with 12 transmembrane domains and a GRAVY value of +0.250, comparable to values obtained here (Table 1), yet remains refractory to 2-DE separation when extracted with standard urea/CHAPS-based solubilising conditions (T. Rabilloud, personal communication). For other single pass transmembrane murine red blood cell proteins, typical GRAVYs are much lower. For example, Glycophorin and Band 7 have GRAVYs of –0.023 and +0.057, respectively, and could be considered less hydrophobic than some proteins identified here. The *E. coli* porin OMPs that solubilise only under strongly denaturing conditions (Solution A) have far lower GRAVYs (ranging between –0.77 to –0.16) than the proteins recovered with the extraction method reported here [5]. Clearly, the use of an organic

solvent extraction has selectively enriched hydrophobic proteins. Importantly, we show that these proteins can be resolubilised with denaturing conditions to permit separation by 2-DE.

While the organic solvent extraction technique has provided us with new protetins of hydrophobic nature, we did not encounter the highly hydrophobic proteins needed to effectively represent the entire predicted *E. coli* proteome. Interestingly, our list of identified proteins consists of hydrophobic cytoplasmic and periplasmic proteins and a hypothetical lipoprotein, but lacks the highly positive GRAVY proteins predicted of multiple pass transmembrane proteins. This could imply that the organic solvent treatment failed to dissolve the phospholipid bilayer and thus failed to release transmembrane proteins. Ames *et al.* [17] also reported the enrichment of periplasmic proteins from *E. coli* by chloroform treatment due to the chloroform susceptibility of an outer membrane "plug". At first glance, our results appear to support this finding. However, several other factors could account for the absence of integral membrane proteins on our gels: (i) Many of the high GRAVY proteins are predicted to be small (<8000 Da) and alkaline (pH >9) and thus do not favour separation and analysis in standard 2-DE and PMF systems; (ii) these proteins extracted with organic solvents are lost at some point in the procedure due to insolubility, or (iii) these proteins are not extracted from the sample in quantities detectable using current staining protocols that are compatible with PMF. An alternative approach for the separation of the predicted highly hydrophobic proteins may be to perform a prefractionation step to first isolate the membranes using a technique such as density centrifugation prior to solubilisation and 2-DE [18]. Such an approach has claimed to separate (presumably some hydrophobic) inner membrane proteins and offers some encouragement, although in this report the identity of the proteins was not determined. In conclusion, further method development is required for routine 2-DE separation of highly hydrophobic *E. coli* proteins.

This research has been facilitated by acccess to the Australian Proteome Analysis Facility established under the Australian Government's Major National Research Facilities Program. MPM is the recipient of an Australian Proteome Industry Research and Development postgraduate scholarship. We thank Thierry Rabilloud for his generous gift of ASB14 and stimulating input. We acknowledge Marc Wilkins for discussions concerning GRAVY values. Bacterial cultures were prepared by Martin Slade.

Received November 19, 1998

References

[1] Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., Williams, K. L., *Biotech. Gen. Eng. Rev.* 1995, *13*, 19–50.

[2] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[3] Rabilloud, T., Adessi, C., Giraudel, A.,. Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[4] Herbert, B. R., Molloy, M. P., Walsh, B. J., Gooley, A. A., Williams, K. L., *Electrophoresis* 1998, *19*, 845–851.

[5] Molloy, M. P., Herbert, B. R., Walsh, R. J., Tyler, M. I., Traini, M., Sanchez, J.-C., Hochstrasser, D. F., Williams, K. L., Gooley, A. A., *Electrophoresis* 1998, *19*, 837–844.

[6] Rabilloud, T., Valette, C., Lawrence, J. J., *Electrophoresis* 1994, *15*, 1552–1558.

[7] Görg, A., Boguth, G., Obermaier, C., Weiss, W., *Electrophoresis* 1998, *19*, 1516–1519.

[8] Wilkins, M. R., Gasteiger, E., Sanchez, J.-C., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 1501–1505.

[9] Adessi, C., Miege, C., Albrieux, C., Rabilloud, T., *Electrophoresis* 1997, *18*, 127–135.

[10] Herbert, B. R., Sanchez, J.-C., Bini, L., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 13–33.

[11] Yerushalmi, H., Lebender, M., Schuldiner, S., *J. Biol. Chem.* 1995, *270*, 6856–6863.

[12] Fillingame, R., *J. Bioenerg. Biomembr.* 1992, *24*, 485–491.

[13] Neidhardt, F. C., Bloch, P. L., Smith, D. F., *J. Bacteriol.* 1974, *119*, 736–747.

[14] Pasquali, C., Frutiger, S., Wilkins, M. R., Hughes, G. J., Appel, R. D., Bairoch, A., Schaller, D., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 547–555.

[15] Traini, M., Gooley, A. A., Ou, K., Wilkins, M. R., Tonella, L., Sanchez, J.-C., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1998, *19*, 1941–1949.

[16] Chevallet, M., Santoni, V., Poinas, A., Rouquié, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J.,. Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1909.

[17] Ames, G. F.-L., Prody, C., Kustu, S., *J. Bacteriol.* 1984, *160*, 1181–1183.

[18] Sato, T., Ito, K., Yura, T., *Eur. J. Biochem.* 1977, *78*, 557–567.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 705–711

Véronique Santoni[1]
Thierry Rabilloud[2]
Patrick Doumas[1]
David Rouquié[1]
Monique Mansion[1]
Sylvie Kieffer[3]
Jérôme Garin[3]
Michel Rossignol[1]

[1]Biochimie et Physiologie Moléculaire des Plantes, INRA/ENSA-M/CNRS URA 2133, Montpellier, France
[2]Bioénergétique Cellulaire et Pathologique, DBMS, CEA-Grenoble, France
[3]Laboratoire de Chimie des Protéines, CEA-Grenoble, Grenoble, France

Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels

An extensive proteomic approach relies on the possibility to visualize and analyze various types of proteins, including hydrophobic proteins which are rarely detectable on two-dimensional electrophoresis (2-DE) gels. In this study, two methods were employed for the purification of hydrophobic proteins from *Arabidopsis thaliana* leaf plasma membrane (PM) model plants, prior to analysis on 2-DE immobilized pH gradient (IPG) gels. Solubilization efficiency of two detergents, (3-[(3-cholomidopropyl)-1-propanesulfonic acid (CHAPS) and C8Ø, were tested for the recovery of hydrophobic proteins. An immunological approach was used to determine the efficiency of the above methods. Fractionation of proteins by Triton X-114 combined with solubilization with CHAPS resulted in the inability to detect hydrophobic proteins on 2-DE gels. The use of C8Ø for protein solubilization did not improve this result. On the contrary, after treatment of membranes with alkaline buffer, the solubilization of PM proteins with detergent C8Ø permitted the recovery of such proteins on 2-DE gels. The combination of membrane washing and the use of zwitterionic detergent resulted in the resolution of several integral proteins and the disappearance of peripheral proteins. In the resolution of expressed genome proteins, both large pH gradients in the first dimension and various acrylamide concentrations in the second dimension must be used. Notwithstanding, it is important to combine various sample treatments and different detergents in order to resolve soluble and hydrophobic proteins.

Keywords: *Arabidopsis thaliana* / Hydrophobic protein / Plasma membrane / Proteome / Two-dimensional electrophoresis
EL 3392

1 Introduction

Proteome research provides a powerful tool to complement other approaches currently used in molecular biology through analysis of the actively translated portion of the genome. The use of IPG (immobilized pH gradient) has greatly improved the feasibility of such approaches and interlaboratory reproducibility [1]. Currently most of the proteomes published have been achieved across pH range 4–8 and using only one given concentration of acrylamide in the second dimension. However, analysis of the expressed genome needs to be as complete as possible and should not be limited to proteins with a particular p*I* and molecular weight [2]. Recently, large immobilized pH gradients and different acrylamide concentrations were used [3, 4] in order to recover the maximum amount of proteins possible. An extensive proteomic approach also relies on the possibility to visualize specific types of proteins (*i.e.*, hydrophobic proteins, glycosylated or membrane-anchored proteins). However, such hydrophobic proteins are rarely detectable on 2-DE gels [5, 6]. In fact, a recent study on the characterization of the plasma membrane (PM) from the plant model *Arabidopsis thaliana* demonstrated that approximately 80% of PM-specific proteins detected in 2-DE gels corresponded nearly exclusively to extrinsic proteins [7]. A combination of recently synthetized new detergents [8] and thiourea was proposed to improve the recovery of hydrophobic proteins on 2-DE gels. In the case of the plant PM, two of the most abundant hydrophobic proteins (water channels and H^+-ATPase) were effectively detected after solubilization with C8Ø [8].

In this study, we further focused on hydrophobic proteins from *Arabidopsis thaliana* PM and compared the efficiency of various treatments designed to enrich the sample in hydrophobic proteins. Using an immunological approach together with protein identification to assess the presence of hydrophobic proteins, a specific procedure was derived. This procedure is compatible with the use of C8Ø and enables the selectively improved recovery of hydrophobic proteins.

Correspondence: Dr. Véronique Santoni, Biochimie et Physiologie Moléculaire des Plantes, INRA/ENSA-M/CNRS URA 2133, place Viala, 34060 Montpellier cedex 1, France
E-mail: santoniv@ensam.inra.fr
Fax: +33-4-67-52-57-37

Abbreviation: PM, plasma membrane

2 Material and methods

2.1 Plasma membrane purification and solubilization

PM purification and protein solubilization were performed according to [8]. The lysis buffer called "UChaps" contained 9 M urea, 0.5% Triton X-100, 20 mM DTT, 1.2% Pharmalytes (3–10) and 4% CHAPS. The lysis buffer called "UTC8" contained 7 M urea, 2 M thiourea, 0.5% Triton X-100, 1.2% Pharmalytes (3–10) and 2% of the new detergent C8Ø [8].

2.2 Triton X-114 fractionation

PM was suspended at 4 mg/mL in a solution containing 2% precondensed Triton X-114 in TBS (150 mM NaCl, 10 mM Tris-HCl, pH 7.6) and was incubated for 15 min in ice [9]. The treated PM was centrifuged at 10 000 × *g* for 10 min and the supernatant was collected and incubated at 37°C. The cloudy solution obtained was centrifuged at 1000 × *g* for 10 min. The lower phase was collected, precipitated with cold acetone and the final pellet was dissolved in UTC8 or UChaps lysis buffer.

2.3 Carbonate washing of membrane

PM fractions were resuspended in 10 mM Tris-HCl, pH 7, 0.3 M sucrose. The sample was diluted 200 times with 0.1 M Na_2CO_3 and kept in ice for 30 min. After centrifugation (15 000 × *g* for 60 min), the pellet was suspended in a storage buffer containing 10 mM Mes-Tris, pH 6.5, 20% v/v glycerol, 250 mM sorbitol, 1 mM PMSF and 1 mM DTT.

2.4 Two-dimensional gel electrophoresis and data analysis

2-D gel electrophoresis was carried out according to [8] with slight modifications. Briefly, IEF was done with commercially available preformed immobilized pH gradients (nonlinear pH gradient 3–10, 18 cm length). The gels were rehydrated overnight in UChaps or UTC8 buffer. The sample was applied at the cathodic side of the gel and was submitted to focusing until 90 000 Vh were reached. After the IEF run, the IPG gel strips were incubated at room temperature in solutions containing DTT and then in iodoacetamide according to [8]. The gels were then submitted to a second dimension run and silver-stained according to [8] or transferred for Western blot experiments. The gels were scanned (ImageMaster Desk Top scanner) and analyzed using BioImage software [10].

2.5 Western blotting

After transferring the proteins on Immobilon (Millipore, Bedford, MA, USA) using a semidry electrophoretic apparatus (Pharmacia, Uppsala, Sweden), the membrane was blocked for 60 min in phosphate-buffered saline (4 mM KH_2PO_4, 16 mM Na_2HPO_4, 115 mM NaCl, pH 7.4) containing 0.1% v/v Tween-20 and 1% bovine serum albumin (BSA), called PBSTB. The blot was then incubated for 60 min in the presence of primary antibody. After washing (2 × 10 min) in PBSTB, the blot was incubated for 45 min with a peroxidase-labeled secondary antibody at 1:30 000 dilution in PBSTB. Secondary anti-rabbit antibody was used to detect H^+-ATPase and cellulase; secondary anti-chicken antibody was used to detect water channels; secondary anti-mouse antibody was used to detect actin and α-tubulin. After washing (2 × 10 min) in PBS, the blot was incubated for 1 min in a chemiluminescent substrate (Super Signal; Pierce, Rockford, IL, USA). The blot was then exposed to an autoradiographic film for 15 s to 5 min.

2.6 Protein identification

Chemical sequencing was performed according to [11]. Identification by MALDI-TOF mass spectrometry was performed according to [12] with a slight modification concerning the matrix: 0.4 µL of a mixture containing 30 mg/mL α-cyano-4-hydroxy-*trans* cinnamic acid (HCCA) and 5 mg/mL of nitrocellulose in isopropanol/acetone (1:1) was used.

3 Results

3.1 Comparative analysis of 2-DE gels from total PM proteins and Triton X-114 fractionated PM proteins

Around 700 spots were usually detected in 2-D gels from PM *Arabidopsis* leaves using a pH gradient of 4–8 and 12% acrylamide in the second dimension (Fig. 1a). Only 400 spots were detected in 2-DE gels loaded with Triton X-114 fractionated proteins (called "T114 gels"; Fig. 1b). When comparing T114 gels with PM gels we found nine spots specific for T114 gels, five spots which had a significant increase in protein levels, and eight spots which had a significant decrease in protein levels (Fig. 1b). In a previous study [7], spots of 2-DE gels from *Arabidopsis* PM were classified into four categories according to their abundance of PM and of cytosolic fraction, *i.e.* (i) specific for PM fraction when not detected in the soluble fraction, (ii) enriched in the PM fraction when compared to the soluble fraction, (iii) of similar abundance in the two fractions, and (iv) enriched in the soluble fraction when compared to the PM fraction. We used this classification to study the spots whose abundance was modified in these gels. The five spots whose amount was increased in the T114 gel corresponded to spots previously classified as being either specific for PM or enriched in PM

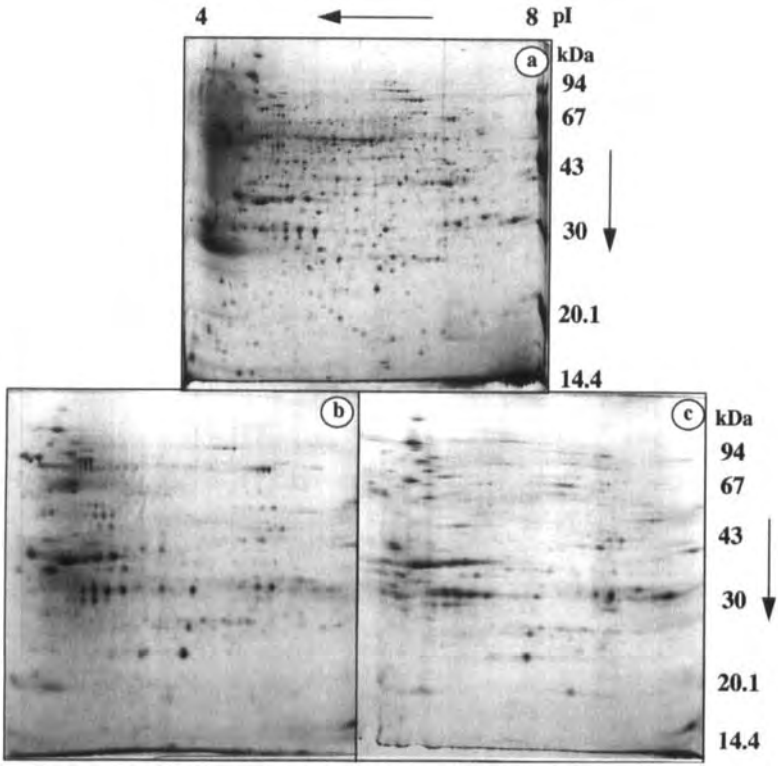


Figure 1. 2-DE gels of Triton X-114 fractionated proteins. (a) 2-DE gel of total PM solubilized in UChaps lysis buffer. Triton X-114 fractionated proteins were solubilized either in (b) UChaps or (c) UTC8 lysis buffers.

when compared to the cytosolic fraction. The eight spots whose amount decreased in the T114 gels corresponded to spots that displayed a similar abundance in PM and soluble fraction or were enriched in the soluble fraction. Therefore, the Triton X-114 fractionation of PM proteins allowed an enrichment in some of the proteins specifically associated with the PM and an impoverishment in some of the proteins characterized as being soluble contaminants of PM.

Nevertheless, only nine spots (less than 3% of the spots from T114 gel) characterized the Triton X-114 fraction in comparison to the total PM fraction. Furthermore, water channels, being the most abundant intrinsic PM proteins, could not be detected on these 2-DE gels (data not shown). On the other hand, approximately a 50% decrease in spots was observed in T114 gels *versus* PM gels. Both gels were equally loaded with PM proteins. In addition, the total integrated optical density of T114 gel did not compensate this low number of spots. These results suggested that most of the proteins from the Triton X-114 fraction were not recovered on the 2-DE gels. The failure to recover the major plant PM intrinsic protein and the general poor recovery of proteins in T114 gels suggested that hydrophobic proteins, classically enriched in this type of fraction, were in fact underrepresented on these 2-DE gels.

3.2 Compared efficiency of the detergents CHAPS and C8Ø to solubilize Triton X-114-fractionated PM proteins

The use of the new detergent C8Ø improved the recovery of some hydrophobic proteins on 2-DE gels such as water channels and H^+-ATPase of PM from *Arabidopsis* leaf [8]. A comparison of changing lysis buffer resulted in no overall change in the total number of spots detected on 2-DE gels. However, 40% of the spots were differentially localized (data not shown). By contrast, the 2-DE pattern of the soluble fraction remained unchanged upon modification of the lysis buffer (data not shown). All these results suggested a specific influence of the C8Ø detergent on the recovery of hydrophobic proteins on 2-D gels. Hence a test to solubilize Triton X-114 fractionated proteins was performed. Figure 1c shows a typical 2-DE gel of Triton X-114-fractionated proteins performed with UTC8 buffer. The comparative analysis of T114 gels performed with UChaps or UTC8 buffers showed that both gels contained around the same number of spots, with no significant quantitative or qualitative differences. Furthermore, we

were unable to detect water channels or H^+-ATPase on these gels.

3.3 2-DE gel recovery of hydrophobic PM proteins extracted with sodium carbonate at pH 11

An enrichment for hydrophobic proteins was sought in the absence of sample pretreatment with any detergent. For this purpose, PM was extracted with sodium carbonate at pH 11, which was shown to be efficient for stripping peripheral proteins from a membrane without affecting the disposition of integral components [13]. The enrichment in hydrophobic proteins was estimated by performing SDS-PAGE Western blots with antibodies directed against known hydrophobic PM proteins (water channels, H^+-ATPase, cellulase) and PM extrinsic proteins (actin, α-tubulin). Water channels and H^+-ATPase possess several transmembrane helixes; cellulase was recently shown to be an integral PM membrane protein with an endoglucanase activity [14]. The same amount of protein from total PM and carbonate-treated PM was loaded onto the gel. The abundance of water channels, H^+-ATPase and cellulase was clearly higher in carbonate-treated sample, whereas the amount of actin and α-tubulin was decreased. Therefore, these results suggested that treatment of PM with sodium carbonate at pH 11 was efficient in both the reduction in the amount of poorly associated PM proteins and enrichment of the sample in hydrophobic proteins. This fraction was then used to perform 2-DE gel.

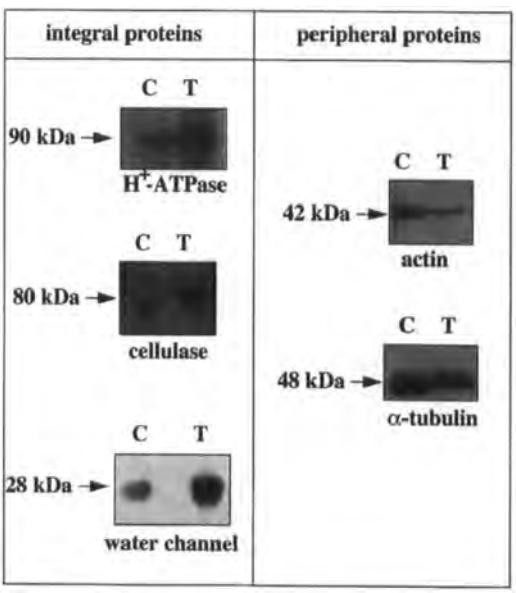


Figure 2. SDS-PAGE Western blots of total PM and PM treated with sodium carbonate at pH 11. Western blots were performed with antibodies against integral plant PM (H^+-ATPase, cellulase, water channel) and against peripheral proteins (actin, α-tubulin). The antibody against water channel was used at 1:10 000 dilution and the other antibodies were used at 1:1000 dilution. "C" and "T" mean, respectively, control PM and carbonate-treated PM.

Typical 2-DE gels of total PM, and of PM treated with carbonate and solubilized in UTC8 buffer, are shown in Fig. 3. Approximately 270 spots were recovered on the 2-DE gel from treated PM. Fifty-four spots were either specific or displayed increased abundance. Simultane-

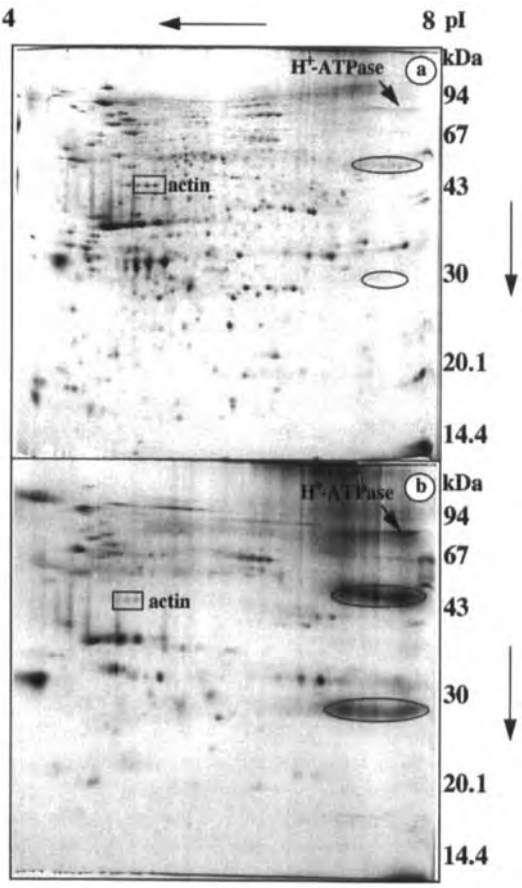


Figure 3. 2-DE gel of (a) total PM and (b) carbonate-treated PM solubilized in UTC8 buffer. Arrows indicate examples of spots whose amount was increased (H^+-ATPase) or decreased (actin) in the 2-DE gels of carbonate-treated PM. The monomer and dimer of water channels are surrounded [8].

Table 1. Identification of proteins from 2-DE gels of carbonate-treated PM

Behavior after carbonate treatment	Protein identification
New proteins and proteins with higher abundance	annexin[a)] H^+-ATPase[c)] 2 × water channels[c)] 9 × proteins of unknown function[a)b)]
Undetected proteins and proteins with decreased abundance	2 × α-tubulin[b)] 3 × actin[a)b)] 2 × BiP[b)] 3 × carbonic anhydrase[a)] HSP70[a)] ERD14-like protein[a)] 3 × GST[a)] 2 × PDI[a)] Methionine synthase[a)] 2 × PGM[a)] Rubisco large subunit[a)] Rubisco small subunit[a)] Calreticulin[a)] S-AMS[a)] Actin depolymerizing factor[a)] Major latex protein[a)] Thioredoxine[a)] Profilin[a)] V-ATPase subunit B[a)] Ascorbate peroxidase[a)] Fructose diphosphate aldolase[a)] 17 × proteins of unknown function[a)b)]

BiP, luminal binding protein; ERD, early responsive to dehydration; GST, glutathione S-transferase; PDI, protein disulfide isomerase; PGM, phosphoglycerate mutase; S-AMS, S-adenosylmethionine synthetase. Proteins were identified either by

a) chemical sequencing

b) MALDI-TOF mass spectrometry

c) Western blot

ously, more than 500 spots from the total PM were no longer detected or displayed lower abundance. The characterization of 61 spots was performed by Western blot, MALDI-TOF-MS or protein sequencing. Among the 54 spots specific for a carbonate-treated or enriched sample set, 13 spots were analyzed (Table 1); three corresponded to the proton pump and water channels, one matched with an annexin and nine spots were not identified (*i.e.*, 69% of the analyzed spots). Among spots that disappeared or decreased in the treated fraction, 31 spots corresponded to proteins with known function and 17 were not identified (*i.e.*, 35% of the analyzed spots).

4 Discussion

Peripheral proteins are mainly detected on 2-DE gels of PM preparations [7]. In addition, since many integral proteins are glycoproteins and poorly stained, a true picture of the actual abundance of integral membrane proteins can not be obtained using classical procedures. The resolving power of electrophoretic techniques would be greatly increased if membrane proteins could first be separated into peripheral and integral families. Various studies showed that membrane integral proteins can be fractionated by a temperature-induced phase separation in Triton X-114 [15]. This strategy was used to separate integral proteins from the PM of *Arabidopsis* leaf. The lower phase of this fractionation (containing hydrophobic proteins) was submitted to classical 2-DE (using CHAPS as the solubilizing agent in the lysis buffer). An analysis comparing this fraction and the PM sample, which was not submitted to prior Triton X-114 fractionation, was performed. Both H^+-ATPase and water channels (strongly hydrophobic) could not be recovered in the T114 gel. Furthermore, some previously identified peripheral proteins were still present in a high amount on T114 gels (Santoni

et al., submitted). Since very few spots characterized from this fraction were integral proteins, it seemed that the classical 2-DE technology was unable to recover hydrophobic proteins from a fraction enriched in such proteins. Therefore, we used the recently synthesized detergent C8Ø which is likely to recover hydrophobic proteins on 2-DE gels [8]. A lysis buffer containing the detergent C8Ø and thiourea (instead of CHAPS and urea) was used to solubilize Triton X-114 fractionated proteins. Surprisingly, the patterns looked similar whichever lysis buffer was used, with no improvement in the recovery of hydrophobic proteins. This suggested that Triton X-114 might interact too strongly with proteins to be displaced by C8Ø.

In a previous study, plant PM hydrophobic proteins were isolated by washing the PM with Triton X-100 in the presence of the KBr (chaotropic agent; Santoni *et al.*, submitted). This method is commonly used to enrich the PM fraction in H^+-ATPase [16]. The 2-DE patterns of samples, solubilized with C8 and pretreated with either Triton X-100/KBr or alkaline buffer, were compared (Santoni *et al.*, submitted). Consequently, the treatment with Triton X-100/KBr removed a greater number of peripheral proteins than did the alkaline treatment (Santoni *et al.*, submitted). The amount of H^+-ATPase detected by immunoblot experiment was lower on 2-DE gels of PM treated with Triton X-100 and KBr than on 2-DE gels of carbonate-treated PM. Besides, there was no apparent change in the amount of water channels. This result suggested that the pretreatment of sample with Triton X-100 prevented the recovery of some but not all hydrophobic proteins. The results obtained here with Triton X-114 show that the sample pretreatment with detergents can affect the recovery of hydrophobic proteins on 2-DE gels in a selective manner. Therefore, the alkaline treatment of membranes described here, despite its inability to remove all peripheral proteins (Fig. 2), appears to be a good compromise to allow their partial release while preserving integral proteins. In the comparison of 2-DE patterns of total PM extract with carbonate-treated PM solubilized in the presence of C8Ø we found an increase in the amount of H^+-ATPase and water channels in the carbonate-treated sample. These two types of proteins possess several transmembrane domains. The detection of H^+-ATPase and water channels occurs when the lysis buffer contained C8Ø but not CHAPS (Santoni *et al.*, submitted). Also detected was an annexin; in the animal field, proteins of this family have been shown to form ion channels *via* an oligomeric structure [17]. Protein sequence analysis of spots present on 2-DE gels of carbonate-treated PM, solubilized in UTC8, showed that one third of identified proteins enriched in this fraction correspond to hydrophobic proteins. The remaining two thirds of the proteins in this fraction were unknown. This treatment reduced the number of spots by 65%, which corresponded to peripheral proteins of the PM. This is exemplified by components of microfilaments and microtubules for which there is evidence of links with PM that can be broken with very alkaline buffers [18]. On the other hand, only one third of the proteins from this class were unknown. This twofold increase of unknown proteins (enrichment of hydrophobic proteins) is in agreement with our poor knowledge of plant PM composition. These results therefore suggest that the carbonate treament of PM at high pH, with the solubilization with C8Ø, favors the isolation of integral proteins and the release of peripheral proteins. However, the low number of spots detected on 2-DE gel under these conditions suggest that a number of hydrophobic proteins are not visualized with this approach. Determination of p*I* for ion and solute transporters in plants have shown a predicted p*I* in the alkaline range (over 8). As this study used a pH range of 4–8, one hypothesis is that most of the isolated hydrophobic proteins would not be detected; therefore, 2-DE gels with a pH range over 8 would be necessary. Preliminary results effectively showed that this pH range could indeed resolve a high number of spots. An alternative explanation could arise from the low abundance of hydrophobic proteins as could be expected from proteins involved in the transport or the perception and transduction of signals. Such limitations could be partly circumvented by increasing the sensitivity of the detection.

The IPG technology and the use of CHAPS-containing lysis buffer is well established in proteome approaches and allows easy comparison of gels. This technology can be expanded to explore extreme p*I* or molecular weight proteins. However, additional and more complex procedures remain in order to investigate special classes of proteins such as hydrophobic proteins. Preliminary enrichment of such proteins by an alkaline wash and solubilization with C8Ø detergent can be suggested as a promising procedure, being easily standardizable, for investigation of hydrophobic proteins and the construction of composite proteomes.

This work was funded by the European Community's BIOTECH program (contract BIO4-CT95-0147). The authors are grateful to Dr. A. R. Schäffner (Institut für Biochemie, Ludwig-Maximilians-Universität, Karlstr. 23, D-80333 München, Germany), Dr. R. Serrano (Instituto de Biologia Molecular y Celular de Planta, Valencia, Spain) and Dr. H. Höfte (INRA, Laboratoire de Biologie Cellulaire, route de S^t-Cyr, 78026 Versailles cedex 1, France) for the gift of antibodies against water channel, H^+-ATPase, and cellulase.

Received September 1, 1998

5 References

[1] Corbett, J. M., Dunn, M. J., Posh, A., Görg, A., *Electrophoresis* 1994, *15*, 1205–1211.

[2] Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., Payne, W. E., *Electrophoresis* 1997, *18*, 1347–1360.

[3] Wasinger, V. C., Bjellqvist, B. J., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[4] Urquhart, B. L., Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B. J., Britton, W. L., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1383–1392.

[5] Adessi, C., Miege, C., Albrieux, C., Rabilloud, T., *Electrophoresis* 1997, *18*, 127–135.

[6] Wilkins, M. R., Gasteiger, E., Sanchez, J.-C., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 1501–1505.

[7] Santoni, V., Rouquié, D., Doumas, P., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouzé, P., Dehais, P., Sahnoun, I., Barlier, I., Rossignol, M., *Plant J.* 1998, *16*, 633–641.

[8] Chevallet, M., Santoni, V., Poinas, A., Rouquié, D., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1909.

[9] Bordier, C., *J. Biol. Chem.* 1981, 256, 1604–1607.

[10] Santoni, V., Delarue, M., Caboche, M., Bellini, C., *Planta* 1997, *202*, 62–69.

[11] Rouquié, D., Peltier, J. B., Marquis-Mansion, M., Tournaire, C., Doumas, P., Rossignol, M., *Electrophoresis* 1997, *18*, 654–660.

[12] Rabilloud, T., Kieffer, S., Procaccio, V., Louwagie, M., Courchesne, P. L., Patterson, S. D., Martinez, P., Garin, J., Lunardi, J., *Electrophoresis* 1998, *19*, 1006–1014.

[13] Fujiki, Y., Hubbard, A. L., Fowler, S., Lazarow, P. B., *J. Cell Biol.* 1982, *93*, 97–102.

[14] Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H., Höfte, H., *EMBO J.* 1998, *17*, 5563–5576.

[15] Pryde, G. J., *TIBS* 1986, *11*, 160–163.

[16] Grouzis, J.-P., Gibrat, R., Rigaud, J., Ageorges, A., Grignon, C., *Plant Physiol.* 1990, *93*, 1175–1182.

[17] Moss, S. E., *Nature* 1995, *378*, 446–447.

[18] Sonesson, A., Berglund, M., Staxén, I., Widell, S., *Plant Physiol.* 1997, *115*, 1001–1007.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 712–717

Angelika Görg
Christian Obermaier
Günther Boguth
Walter Weiss

Technical University of Munich, Department of Food Technology, Freising-Weihenstephan, Germany

Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: Wide pH gradients up to pH 12, longer separation distances and simplified procedures

Wide-range immobilized pH 3–12 and 6–12 gradients were generated. Depending on the extraction method of sample preparation, proteins with p*I*s up to pH 11.7 were resolved. Highly reproducible protein patterns, focused to the steady-state with round-shaped spots up to the basic end were obtained. Moreover, because a strong water transport from cathode to anode (reverse electroendosmotic flow) inherent to narrow immobilized pH gradients (IPGs) exceeding pH 11, such as IPG 10–12, was negligible, the wide-range IPGs 3–12 and 6–12 could be run under standard conditions as originally described by Görg *et al.* (*Electrophoresis* 1988, *9*, 531–546). The wide-range immobilized pH gradient 3–12 proved to be perfectly suited for an overview separation of total cell extracts. Resolution could be increased by extending the separation distance from 18 to 24 cm. Furthermore, two-dimensional gel electrophoresis with IPGs (IPG-Dalt) was simplified by the use of an integrated system (IPGphor) where sample application by in-gel rehydration and isoelectric focusing (IEF) are performed automatically in a one-step procedure, overnight, without human assistance.

Keywords: Alkaline proteins / Automation / IPGphor / Ribosomal proteins / Two-dimensional polyacrylamide gel electrophoresis / Wide-range immobilized pH gradients EL 3348

1 Introduction

For the optimization of 2-D electrophoretic separations thereare two different approaches to attain maximum resolution: (i) using narrow pH gradients with overlapping intervals, or (ii) generating wide-range pH gradients with extended separation distances. Furthermore, there is a demand for new equipment in order to simplify the multistep procedure of 2-D electrophoresis. Blow-up experiments, using narrow immobilized pH gradients (IPG), *e.g.* IPG 5.5–6.5 over a separation distance of 11 or 18 cm [1], clearly demonstrated the high resolving power of narrow IPGs for 2-D electrophoresis. By using an IPG 4.35–4.55 over an 18 cm separation distance, a Δ p*I* of 0.001 could be obtained [2]. Alternatively, wide-range pH gradients for overview 2-D patterns of total cell extracts can be generated. More recently, a wide-range immobilized pH gradient up to pH 12, an IPG 4–12, was described for an overview 2-D separation of mouse liver proteins [3]. This gradient has two significant features: because the pH gradient is flattened at the basic end (between pH 9.5 and 12), very alkaline proteins such as ribosomal proteins are highly resolved and perfectly separated under steady-state conditions and, because a strong water transport from cathode to anode (reverse electroendosmotic flow) inherent to narrow IPGs exceeding pH 11, such as IPG 9–12 and 10–12 [4] is negligible, the IPG 4–12 can be run under standard conditions originally described in 1988 [1]. In the present paper, we describe two new immobilized pH gradients up to pH 12: IPG 3–12 and IPG 6–12. Additionally, extended separation distances up to 24 cm for the linear IPG 3–12 were investigated. Furthermore, a simplified procedure for first-dimensional IEF using an integrated system (IPGphor) was applied for IEF with IPGs 4–7, 4–9, 4–12, 3–10, and 3–12.

Correspondence: Prof. Dr. Angelika Görg, Technische Universität München, Lehrstuhl für Allgemeine Lebensmitteltechnologie, D-85350 Freising-Weihenstephan, Germany
E-mail: angelika.gorg@tum.de
Fax: +49-8161-714264

Abbreviation: IPG-Dalt, two-dimensional gel electrophoresis with immobilized pH gradients

2 Materials and methods

2.1 Apparatus and chemicals

IPGphor, Multiphor II horizontal electrophoresis apparatus, Dalt-multiple vertical electrophoresis apparatus, EPS 3500 XL power supply, Multitemp II thermostatic circulator, Immobiline II chemicals, Pharmalyte (pH range 3–10), IPG buffers, Immobiline DryStrips 4–7, 3–10 L, and 3–10 NL, acrylamide, bisacrylamide, ammonium persulfate, TEMED, CHAPS, and urea were from Amersham Pharmacia Biotech (Uppsala, Sweden). Acrylamido buffers p*K* 1.0, p*K* 10.3 and p*K* > 13 were a gift from Bengt Bjellqvist

(Amersham Pharmacia Biotech). Silicone oil was from Serva (Heidelberg, Germany). All other chemicals (analytical grade) for electrophoresis and for silver staining were from Merck (Darmstadt, Germany).

2.2 Sample preparation

Deep-frozen mouse liver was ground in a liquid nitrogen-cooled mortar. Mouse liver proteins were solubilized in "lysis" buffer containing 9 M urea, 2% CHAPS, 1% DTT and 2.0% v/v Pharmalyte 3–10. TCA/acetone extract of mouse liver was prepared as described [3].

2.3 Gel casting

IPG gels (180 or 240 mm long) on GelBond PAGfilm were cast, washed dried, and cut into individual 3 mm wide IPG gel strips by the procedure described previously [1, 5–7]. The recipes of IPG 3–12 and IPG 6–12 listed in Table 1 were calculated and optimized by using the computer program of Altland [8]. Horizontal and vertical SDS gels were cast as described previously [1, 6, 7].

2.4 2-D electrophoresis

The first dimension (IPG-IEF) of 2-D electrophoresis was either performed on the Multiphor II apparatus, or on the IPGphor.

2.4.1 IPG-IEF on the Multiphor

IPG dry strips were rehydrated in a reswelling solution containing 8 M urea, 1% CHAPS, 20 mM DTT, 0.5% v/v Pharmalyte 3–10. Samples (20–100 µL) of mouse liver proteins, either TCA/acetone extract or "lysis" buffer extract, were applied into sample cups near the anode using the Pharmacia (Uppsala, Sweden) DryStrip kit. IEF on IPGs 3–12 (24 cm) and 6–12 (18 cm) was performed under a layer of silicone oil which had been degassed and flushed with argon prior to use. Electrode paper wicks were soaked with deionized water and blotted against filter paper to remove excess liquid. For improved sample entry, voltage was limited to 150 V (30 min), 300 V (60 min), and 600 V (60 min) at the beginning. IEF, continued with a maximum of 3500 V to the steady state, was performed at 20°C [9]. Current was limited to 0.05 mA per IPG gel strip. Running conditions for IPGs 3–12 (24 cm) and 6–12 (18 cm) are given in Table 2.

Table 1. Recipes for casting IPG 3–12 and IPG 6–12 gels

Chemicals	IPG 3–12 Dense solution[a] pH 3	Light solution[a] pH 12	IPG 6–12 Dense solution[a] pH 6	Light solution[a] pH 12
Immobiline p*K* 1.0	858 µL	–	–	–
Immobiline p*K* 3.6	204 µL	–	911 µL	–
Immobiline p*K* 4.6	276 µL	–	–	–
Immobiline p*K* 6.2	372 µL	224 µL	125 µL	167 µL
Immobiline p*K* 7.0	331 µL	112 µL	217 µL	83 µL
Immobiline p*K* 8.5	75 µL	466 µL	243 µL	56 µL
Immobiline p*K* 9.3	56 µL	105 µL	331 µL	21 µL
Immobiline p*K* 10.3	17 µL	228 µL	33 µL	323 µL
Immobiline p*K* > 13	–	172 µL	–	230 µL
Acrylamide/Bis (29.1/0.9)	1.50 mL	1.50 mL	1.50 mL	1.50 mL
Deionized water	4.30 mL	7.20 mL	4.65 mL	7.60 mL
Glycerol (100%)	2.50 g	–	2.50 g	–
TEMED (100%)	6.0 µL	6.0 µL	6.0 µL	6.0 µL
Persulfate (40%)	10.0 µL	10.0 µL	10.0 µL	10.0 µL
Final volume	10.0 mL	10.0 mL	10.0 mL	10.0 mL

a) For effective polymerization, heavy and light solution are adjusted to pH 7 with 3 N sodium hydroxide and 3 N acetic acid, respectively, before polymerization

Table 2. Multiphor II: running conditions for IEF with rehydrated IPG strips

IPG strips	24 cm, IPG 3–12 18 cm, IPG 6–12		
Temperature	20°C		
Current maximum	0.05 mA per IPG strip		
Power maximum	5.0 W		
Sample volume	20–100 µL, sample cups		
IEF			
Voltage maximum	150 V	0.5–1 h	Sample entry
	300 V	0.5–2 h	Sample entry
	600 V	1 h	Sample entry
	3500 V	8 h	(28 000 Vh) IEF

2.4.2 IPG-IEF on the IPGphor

Prior to IPG strip rehydration, the extract (protein concentration ≈ 10 mg/mL) was diluted with reswelling solution (8 M urea, 1% CHAPS, 20 mM DTT, and 0.5% v/v IPG buffers). Dilution was 1 + 1 for micropreparative runs, and 1+19 for analytical runs. The required number of strip holders (up to 12) for 18 cm long IPG strips was put onto the cooling plate's electrode contact area of the IPGphor, and 350 µL sample-containing rehydration solution (for 180 mm long IPG strips) was pipetted carefully at a central point in the strip holder channel away from the sample application wells at the electrodes. The IPG strips were lowered, gel side down, onto the rehydration solution without trapping any air bubbles, and overlayered with 1 mL of silicone oil before the plastic cover was applied. Once the safety lid is closed, rehydration and IEF are carried out automatically according to the programmed settings (Table 3), preferably overnight, without further handling steps. As indicated in Table 3, low voltage (30–60 V) was applied already during the rehydration step for improved sample entry of high M_r proteins into the polyacrylamide gel matrix (Reiner Westermeier, personal communication). As an alternative to in-gel rehydration, the sample (≈ 5 mg protein/mL, dissolved in lysis buffer) was applied by pipetting 20 µL into the lateral sample application wells after the IPG strip had been rehydrated. After IEF, those IPG gel strips which were not used immediately for the second dimension, or kept for further reference, were stored between two sheets of plastic film at –78°C up to several months. Equilibration of the IPG strips, the second dimension (horizontal or vertical SDS electrophoresis) as well as silver staining, were performed as described previously [1, 7]. For more details see http://www.weihenstephan.de/blm/deg.

3 Results and discussion

In Fig. 1, a silver-stained 2-D pattern of mouse liver proteins is demonstrated. Isoelectric focusing was performed in a linear IPG 3–12 over a separation distance of 24 cm. The recipe of IPG 3–12 listed in Table 1 was calculated and optimized by using the computer program of Altland [8]. Although this gradient comprises 9 pH units, the resulting 2-D pattern shows highly resolved protein spots even in the crowded area between pH 4 and 9. This is due to the linearity of the gradient which, in contrast to the previously described IPG 4–12, is not flattened at the basic end. Additionally, the separation distance was enlarged from 18 to 24 cm. Similar resolution for alkaline proteins is obtained by using an IPG 6–12 over 18 cm (Fig. 2). Depending on the sample preparation method, (i) protein solubilization with lysis buffer (Fig. 2A), or (ii) after TCA/acetone precipitation (Fig. 2B), very alkaline proteins, such as ribosomal proteins, can be detected. However, it becomes evident that the calculated gradient 6–12 in reality does not exceed pH 11.7. Therefore, ribosomal proteins are better resolved by using IPG 4–12 [3], 9–12, or 10–12 [4]. Apart from this limitation, the IPG 6–12 proved to be an excellent gradient for the majority of alkaline proteins, and, most importantly, horizontal streaking at the basic end, which has been described for IPGs up to pH 10 [6], was not observed.

In Fig. 3, mouse liver proteins were focused in an IPG 4–7 over 18 cm separation distance. Sample was applied by in-gel rehydration using the IPGphor system for IEF [10], where specifically designed IPG strip holders with integrated electrodes provide both rehydration and IEF in one unit without further handling. Once the individual strip holders (up to 12 units) are placed on the IPGphor platform, reswelling and IEF are carried out automatically according to the protocol described in Table 3. For improved sample entry of high M_r proteins, the rehydration process was activated by applying low voltage during reswelling. After sample entry with limited voltage, isoelectric focusing was performed within 4 h with a maximum of 8000 V. The resulting 2-D pattern (Fig. 3) shows highly resolved protein spots between pH 4 and 7 (6 cm separation

Table 3. IPGphor: voltage settings for reswelling and IEF of dry strips

IPG strips	18 cm	
Temperature	20°C	
Current maximum	0.05 mA per IPG strip	
Sample volume	350 µL	
Reswelling	30 V for 6 h, followed by 60 V for 6 h	
IEF	200 V for 1 h	
	500 V for 1 h	
	1000 V for 0.5 h	
	1000 V → 8000 V for 30 min; then continue with:	
	8000 V for 4 h	IPG 4–7
or	8000 V for 3 h	IPG 3–10 L, 3–10 NL, IPG 3–12
		IPG 4–9, IPG 4–12

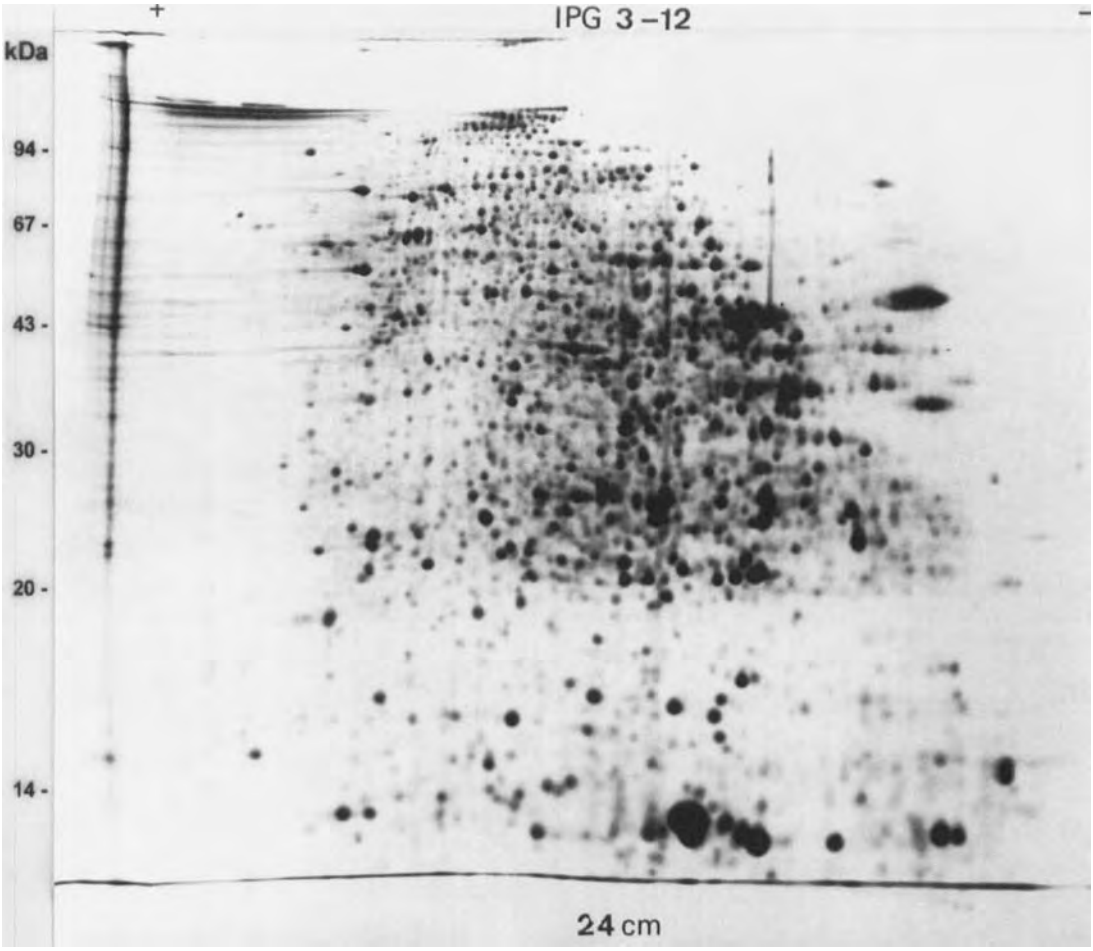


Figure 1. IPG-Dalt of mouse liver proteins (lysis buffer extract) using IPG 3–12 (24 cm separation distance) in the first dimension. Sample application: cup-loading near the anode. Second dimension: SDS-PAGE (13%T constant). Silver stain.

distance/pH unit). Additionally, excellent results were also obtained by using IPGs 4–9, 3–10 L, 3–10 NL and 3–12 (18 cm; not shown), whereas the in-gel rehydration procedure for IPGs 6–10 and 6–12 with the protocol listed in Table 3 still has to be improved.

4 Concluding remarks

Compared to narrow IPGs such as IPG 9–12 and 10–12 [4], wide gradients up to pH 12, such as IPG 4–12 [3], 6–12 and 3–12, are easy-to-use pH gradients for 2-D electrophoresis. Since no apparent reverse electroosmotic flow is observed, which would give rise to streaky 2-D patterns, no special attention has to be paid to the basic end of the gradients, and they can be run under standard conditions [1]. The linear IPG 3–12 is perfectly suited for an "overview" pattern of total cell extracts. Additionally, extended separation distances such as 24 cm can be easily verified due to the short focusing times of wide-range pH gradients and due to the size stability of IPG strips cast on plastic backings. Finally, by using the IPGphor, the multistep procedure of 2-D electrophoresis can be simplified by using IPG strip holders with integrated electrodes, and accelerated by the use of 8000 V for IEF provided by the system.

Received December 3, 1998

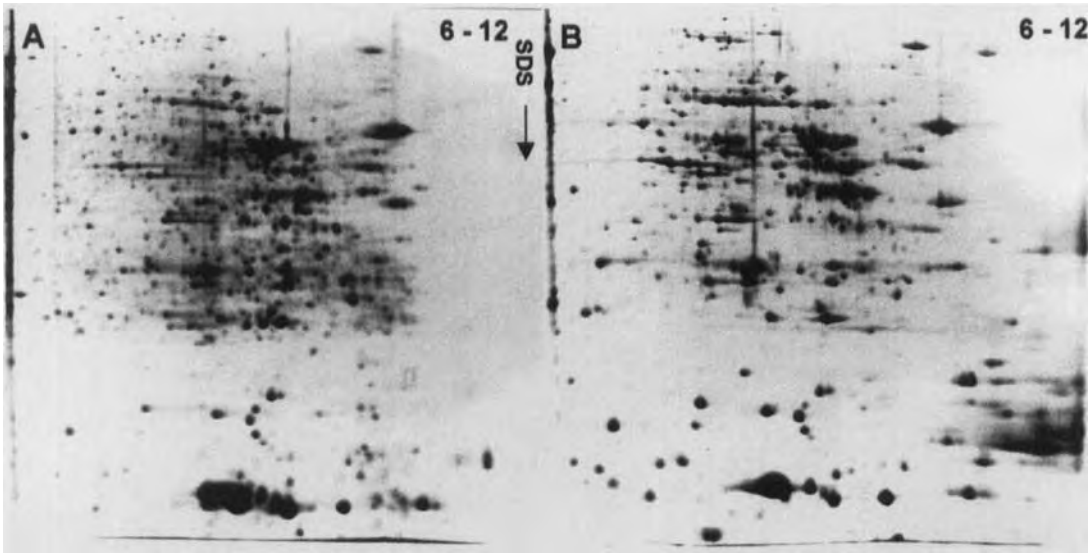


Figure 2. IPG-Dalt of mouse liver proteins using IPG 6–12 (18 cm separation distance) in the first dimension. Sample application: cup-loading near the anode. (A) Lysis buffer extract. (B) TCA/acetone extract. Second dimension: SDS-PAGE (13%T constant). Silver stain.

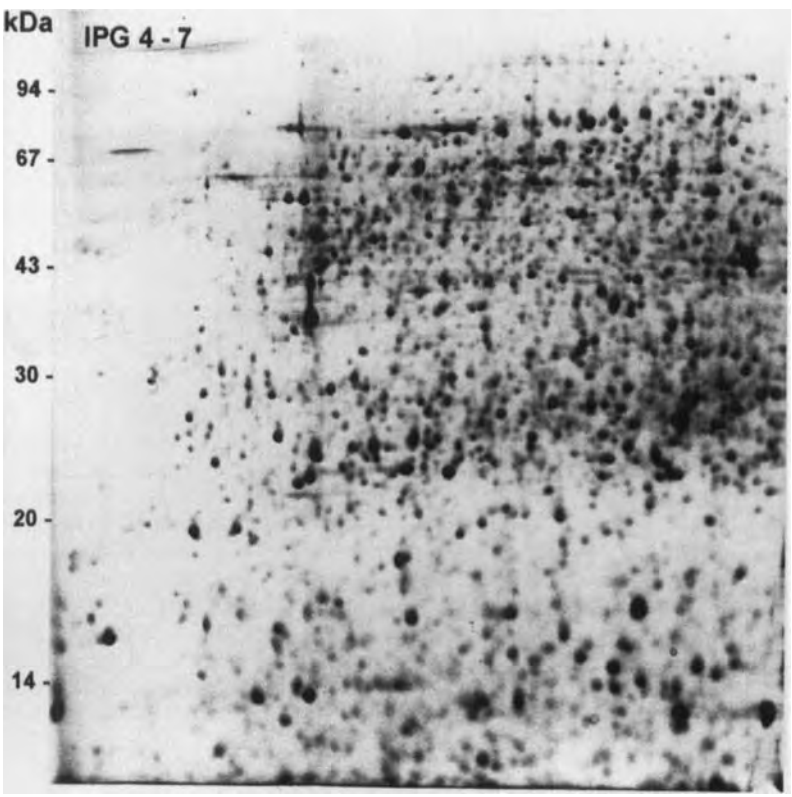


Figure 3. IPG-Dalt of mouse liver proteins (lysis buffer extract). First dimension: IPG 4–7 (18 cm separation distance), run on the IPGphor (sample application by in-gel rehydration). Second dimension: SDS-PAGE (13%T constant). Silver stain.

5 References

[1] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[2] Görg, A., Postel, W., Weser, J., Patutschnick, W., Cleve, H., *Am. J. Hum. Genet.* 1985, *37*, 922–930.

[3] Görg, A., Boguth, G., Obermaier, C., Weiss, W., *Electrophoresis* 1998, *19*, 1516–1519.

[4] Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J. J., Madjar, J. J., *Electrophoresis* 1997, *18*, 328–337.

[5] Görg, A., *Nature* 1991, *349*, 545–546.

[6] Görg, A., Boguth, G., Obermaier, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16*, 1079–1086.

[7] Görg, A., Weiss, W., in: Celis, J. (Ed.), *Cell Biology. A Laboratory Handbook*, Vol. 4, 2nd edition, Academic Press, New York 1998, pp. 386–297.

[8] Altland, K., *Electrophoresis* 1990, *11*, 140–147.

[9] Görg, A., Postel, W., Friedrich, C., Kuick, R., Strahler, J., Hanash, S. M., *Electrophoresis* 1991, *12*, 653–658.

[10] Islam, R., Ko, C., Landers, T., *Sci. Tools* 1998, *3*, 14–15.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 718–722

Alexander V. Stoyanov[1]
Pier Giorgio Righetti[2]

[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia
[2]University of Verona, Verona, Italy

Steady-state electrolysis of a solution of nonamphotheric compounds

The problem of stationary electrolysis of a solution of nonamphotheric compounds (acids or bases) is considered. The analysis is performed by taking into account the mobility dependence on pH. The properties of such a system are also compared with the ones pertaining to the solution of an amphoteric substance. It is anticipated that steady-state electrolysis of free acids and bases, in a convection-free system, might be useful for creating narrow pH gradients in rather acidic and alkaline milieus, which might be adopted for focusing without resorting to conventional carrier ampholytes or immobilized pH gradients.

Keywords: Steady state / Weak electrolytes / Electrolysis EL 3341

1 Introduction

In this article we will analyze the steady-state concentration distribution when electrolyzing a nonamphoteric substance in a convection-free environment. Such a system is not only of practical importance as an alternative way for creating pH gradients, but is also a theoretically didactic problem by itself. The experimental setup may be treated as the simplest example of "natural pH gradient" (the definition introduced by Svensson in [1]). Some other examples may be found in our recent papers dealing with the steady-state distribution of carrier ampholytes in an immobilized pH gradient [2] and in pure water [3]. There are some specific features of this electrophoretic process for nonamphoteric substances in comparison with ampholytes, the most essential difference being that the former species have no isoelectric point and thus there is no limit of the pH function as their concentration is increased. Svensson-Rilbe considered the same problem in [4]; as a result he obtained a linear concentration profile, but with some more stringent restrictions, *viz.*, he analyzed only strong electrolytes and, also, he supposed constant transference numbers (in other terms, he neglected the charge variation with pH).

2 Model description

We will use the same system of equations, as was done when analyzing the electrolysis of a solution of an amphotere [3]:

$$D\frac{dC(x)}{dx} = V(x,C)\,C(x) \tag{1}$$

$$V(x,C) = v_e\ q(x,C)\ E(x,C) \tag{2}$$

$$E(x,C) = \frac{j}{\lambda(x,C)} \tag{3}$$

Here C is the concentration, $V(x,C)$ is the velocity of electrophoretic migration, D is the diffusion coefficient and v_e is used to denote the mobility of an electrolyte ion, normalized to the electric charge value (q).

The expression for the conductivity $\lambda(x,C)$ should be written by taking into account the water ion contribution (λ_w):

$$\lambda\,(x,\,C) = \lambda_e\,(x,C) + \lambda_w\,(x,C) \tag{4}$$

while the electrolyte contribution to conductivity (λ_e) may be expressed as:

$$\lambda_e\,(x,\,C) = Fv_e\,C(x)\,q^2\,(C) \tag{5}$$

where F is the Faraday of electricity. Note that the electric charge q, the conductivity λ and, consequently, the velocity of electrophoretic transport (V) are the implicit functions of the coordinate (x); they are defined completely by the concentration C.

Let us now suppose that electrolysis of an acid solution is performed, and that the electric charge as a function of pH is:

$$q = \frac{K_a}{[H] + K_a} \tag{6}$$

In order to determine the pH we use the electroneutrality equation, the latter being a cubic equation relative to the hydrogen ion concentration:

Correspondence: Prof. P. G. Righetti, University of Verona, Department of Agricultural & Industrial Biotechnologies, Strada Le Grazie, Ca'Vignal, 37134 Verona, Italy
E-mail: righetti@imiucca.csi.unimi.it
Fax: +39-45-8098901

$$[H] = \frac{CK_a}{[H] + K_a} + \frac{K_w}{[H]} \quad (7)$$

Here, K_w is the ion water product. We may easily neglect the second term on the right side of the electroneutrality equation, provided that the acid concentration is not too small ($C >> (K_w)^{1/2}$). We thus obtain:

$$[H] = \frac{1}{2}(-K_a + \sqrt{K_a^2 - 4CK_a}) \quad (8)$$

This relationship is represented in Fig. 1 for different values of dissociation constants of hypothetical acids, spanning a range of pKs from 0 to 7 at unit increments. For the electric charge we have strong decrements, up to 0.2 M concentrations, to a small fractional charge for the weaker acids (pK 3 up to pK 7), but then this residual charge also remains different from zero at very high concentrations (see Fig. 2).

With our assumption of moderate or strong acidic conditions the expression for conductivity may be written as:

$$\lambda = F\{v_H [H(C)] + v_e C(x) q^2\} \quad (9)$$

where v_H is the relative mobility of hydrogen ions. Equation (9) predicts an unlimited conductivity increase with concentration growth (in striking contrast to the case of carrier ampholyte solution [5, 6]). As shown in Fig. 3, these conductivity increments are much steeper for stronger acids (pK 3 and lower) and shallower for weaker ones (pK 4 and higher).

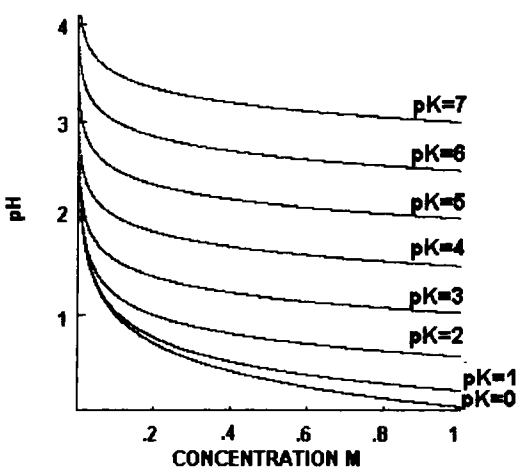


Figure 1. Prevailing pH as a function of concentration for solutions of ideal monobasic acids with different pK values spanning a range from pH 0 to 7, at unit increments (interval of high concentrations).

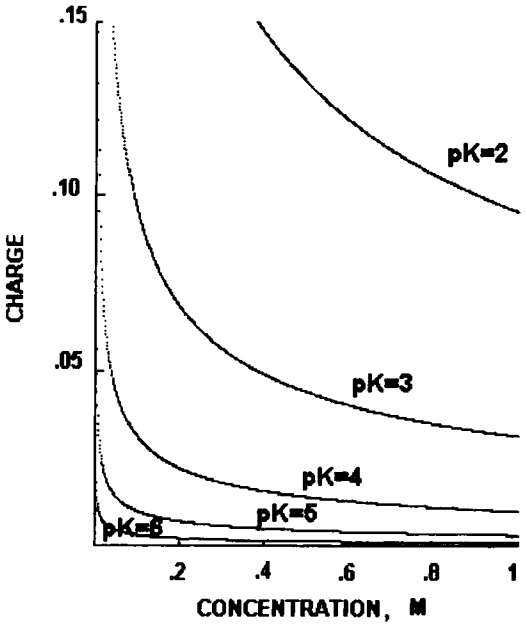


Figure 2. Electric charge *versus* concentration for solutions of hypothetical monobasic acids with different pK values spanning a range from pH 2 to 6, at unit increments (interval of high concentrations).

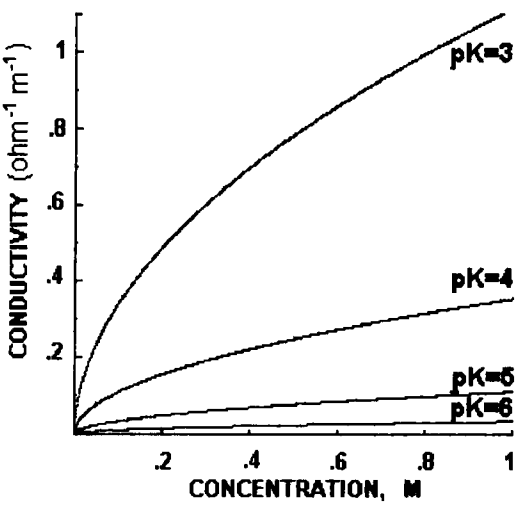


Figure 3. Conductivity as a function of concentration for solutions of ideal monobasic acids with different pK values spanning the range pH 3–6, at unit increments (interval of high concentrations). The relative mobility of anions in these simulations is 0.05.

Let us now consider the behavior of electrophoretic migration velocity; as shown in Eq. (5) it contains the ratio of q to λ, which according to our assumptions (Eq. 8 for [H]), gives the relationship of C close to a hyperbolic one ($V \sim C^{-1}$). Written in a complete form:

$$V(C) = \frac{\nu_{rel}\, j}{FC\,\{1 + \nu_{rel}\, q\,[H(C)]\}} \tag{10}$$

where ν_{rel} expressed the anion mobility in hydrogen mobility units, that is, $\nu_{rel} = \nu_e/\nu_H$. The relationship thus obtained is visualized in Fig. 4. The important result is that the velocity of electrophoretic migration is quasi independent of the dissociation constant value (by assuming rather high concentrations). In fact, the curves for all acids from pK 0 up to pK 5 are almost coincident.

It is also interesting to compare the conductivity contributions of electrolyte and water. Their ratio is expressed by:

$$\frac{\lambda_{el}}{\lambda_w} = \nu_{rel}\, C(x)\, \frac{q^2}{[H]} \tag{11}$$

By using the equation for hydrogen ion concentration [8], we obtain:

$$\frac{\lambda_{el}}{\lambda_w} = \nu_{rel}\, q \tag{12}$$

Taking into consideration that, usually, the mobility of an electrolyte ion relative to that of the hydrogen ion is lower by at least one order of magnitude, we may see that the highest degree of ionization (see Fig. 2) will correspond only to a small percent of the ratio given in Eq. 12. Note, on the contrary, that with a solution of an amphoteric substance we have a different situation: the pH approaches the pI as the concentration increases (at this point the ampholyte charge is a true zero), as also shown experimentally by Righetti and Nembri [7] and by Bossi and Righetti [8]. For nonamphoteric species, we have instead a permanent growth of hydrogen ion concentration (see Fig. 1); nevertheless, the relative contribution of electrolyte ions also becomes considerably smaller at high concentrations*. Although the concept of dividing the conductivity of an electrolyte by that of bulk water may seem somehow artificial, due to the fact that the conductivity is an additive value, it is nevertheless correct and we think it is useful for a better understanding of electrolysis phenomena.

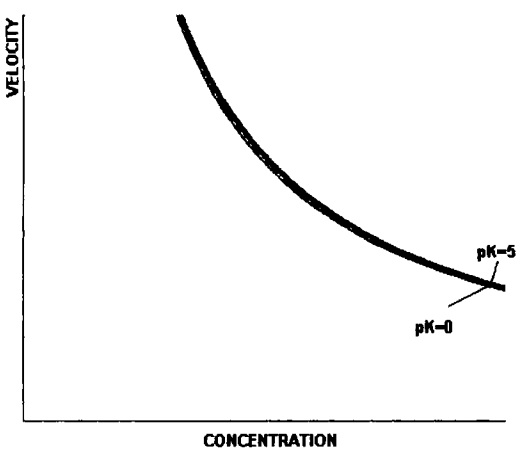


Figure 4. Velocity of electrophoretic migration *versus* concentration for hypothetical monobasic acids with different pK values spanning the pH range 0–5, at unit increments.

3 Steady-state concentration distribution

With the assumptions made above, the differential equation, expressing the steady-state condition, can be rewitten in the form:

$$\frac{dC(x)}{dx} = \frac{\nu_{rel}\, j\; q(C)C(x)}{DF\;([H] + \nu_{rel}\, q^2\,(C)C(x))} \tag{13}$$

The simplest form is obtained for the case of a strong electrolyte, in which case (H) = C and also the electric charge may be supposed equal to the unit. Then the above equation reduces to:

$$\frac{dC(x)}{dx} = \kappa \tag{14}$$

where the coefficient κ is equal to:

$$\kappa = \frac{j\nu_{rel}}{(1 + \nu_{rel})\, DF} \tag{15}$$

The solution to Eq. (14) is a linear concentration course:

$$C(x) = \kappa x + \text{Const} \tag{16}$$

where the *Const* value is defined by the total amount of electrolyte.

In a general case, however, we need to solve Eq. (13). Since the latter belongs to the type of separate variables, it may be analyzed in a rather easy way. By taking into consideration the fact that the conductivity contribution of an electrolyte anion may be omitted (see the previous section) we thus again obtain an equation of the same

* In other terms, an additional increment of concentration of electrolyte (acid or base), obviously results in an increase of conductivity, but the latter takes place mostly due to pH changes rather than to an additional amount of electrolyte (non-water ions).

form as Eq. (14). We should now take another coefficient, however κ' instead of κ:

$$\kappa' = \kappa (1 + \nu_{rel})$$

As we have seen above, the behavior of strong or weak electrolytes is similar, in the range of rather high electrolyte concentrations, but we should pay attention to the low-concentration region. Here, for the hydrogen ion concentration, we have to use another expression; by supposing that $C << K_a$, we have:

$$[H]^2 - [H(C)] - K_w = 0 \qquad (17)$$

This equation, in contrast to Eq. (8), gives an upper pH limit of pH 7 at vanishing concentrations of acid (Fig. 5). In addition, we should no longer neglect the term of conductivity, which is related to the hydroxyl ion contribution:

$$\lambda = F\nu_H \{[H(C)] + \nu_{OHrel} \frac{K_w}{[H(C)]} + \nu_{rel} C q^2 (C)\} \qquad (18)$$

where ν_{OHrel} is used to express the hydroxyl ion mobility in hydrogen ion mobility units. Thus we obtain the differential equation:

$$\left(\frac{[H(C)]}{C} + \frac{\nu_{OHrel} K_w}{[H(C)]C} + \nu_{rel}\right) dC = \left(\frac{j\nu_{rel}}{DF}\right) dx \qquad (19)$$

We will analyze the equation with the assumption of $C << 2K_w^{1/2}$. The results of calculations are given in Fig. 6, in which we can see that, by utilizing Eq. (17) for hydrogen ion concentration (in contrast to Eq. 7), there is no development of a region in solution with a zero concentration of electrolyte.

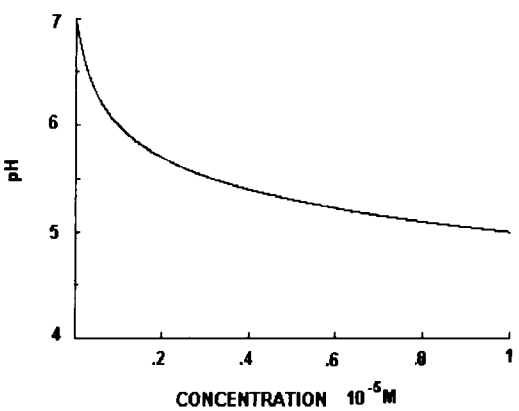


Figure 5. Prevailing pH as a function of concentration for a solution of a monobasic acid (low concentration range). Note that, for vanishing concentrations, the pH tends towards the upper limit value of pH 7.

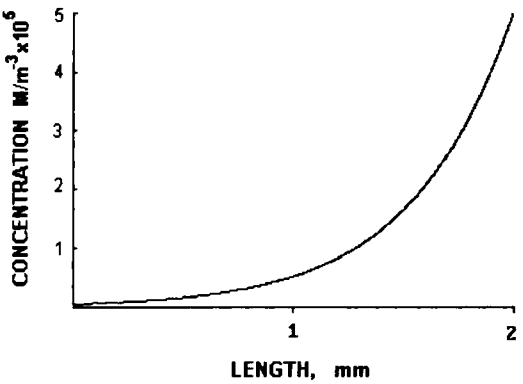


Figure 6. Concentration course of a strong acid when the steady-state electrolysis of a low concentration solution is performed. In these calculations the following parameters were used: j = 1mA/m^2 (j, current density); ν_{rel}= 0.02; D= 0.5 10^{-10} m^2/s.

4 Discussion

Our theoretical analysis has produced, as a main result, the observation that electrolysis of a weak electrolyte generates, under steady-state conditions, a concentration distribution close to linear. That is to say, although the mobility of the electrolyte does not begin to meet the condition of approaching a zero value as its concentration is increased (as in the case of amphoteric substance), nevertheless there exists a mechanism restricting the excessive concentration growth, and it is mostly connected with the increase of conductivity.

As we mentioned in the introduction, the system considered in the present paper is the simplest case of a natural pH gradient, since it represents a single-component solution subjected to electrolysis (in the sense that no other new ions, except for one, are generated when dissolving* our substance in pure water). Such systems may cause some paradox with respect to the question of terminology, *viz.*, the ion is subjected to the action of electric field in a pH gradient, but the latter is formed with the help of the same ion only (no other ions being present except for those produced by dissociation of water). On the other hand, such gradients may be used for separation purposes, provided the concentration of the gradient-creating components is considerably higher than that of the sam-

* Generally, we have a number of microstates which are in mutual equilibrium; in the framework of the so-called "highly-relaxing model" we may speak about only one "effective state" [9]. Note that a solution of any salt (one chemical substance) does not form a single component solution.

ple ions. They should probably be more effective in the case of very narrow pH gradients, especially when using only a single weak acid (or base). Although this principle has been forgotten in present-day separation science, it was tried and successfully established long ago, when attempting to create very acidic pH gradients, not covered by the carrier ampholytes commercially available in those days. Thus Pettersson [11] and Stenman and Graesbeck [12] applied this principle for generating pH gradients covering, fairly linearly, the pH 2–4 interval. In order to accomplish this task, not a single acid, but a series of acids of different strength, had to be carefully blended. These were: malic acid, formic acid, succinic acid, acetic acid, and propionic acid. Upon steady-state electrolysis, in a convection-free environment (ensured with the help of a sucrose density gradient), a pH 2–4 gradient could be generated, in which cobalophilin (or *R*-type vitamin B_{12} binding protein) of human serum could be resolved into five major components, having p*I*s of 2.5, 3.1, 3.4, 3.6 and 3.9 [12]. Such results were remarkable in those days, much as they would be unique even by today's standards. Later on, also Lundhal and Hjertén [13] attempted to create pH gradients in ordinary buffers, but with a different approach: either by means of a temperature gradient, or by means of electrophoresis in a concentration gradient of a neutral substance, such as sucrose, sucrose polymerized with epichlorohydrin (Ficoll), or polyacrylamide gels. However, neither approach was successful. To quote verbatim: thermally engendered "pH gradients are difficult to utilize for isoelectric focusing because the pH of ordinary buffers and the p*I* proteins do not exhibit sufficiently large differences in temperature coefficients". In the latter case, "semistationary pH gradients develop during electrophoresis in concentration gradients formed by neutral substances, for instance sucrose, Ficoll, polyacrylamide gels. The useful parts of these gradients in our experiments were too steep or too shallow to be utilized for isoelectric focusing".

At the end of this excursus we want to mention that the stationary electrolysis of some single component was investigated experimentally by Tiselius as early as in 1941 [10]. Tiselius studied solutions of sodium hydroxide and sulfuric acid and obtained concentration profiles close to linear (the experiments were made in a multi-compartment chamber). Our treatment was performed by taking into account free acids, as an example of electrolytes, but similar results are obtained, of course, in the case of free bases.

P.G.R. is supported by grants from Agenzia Spaziale Italiana (No. ARS-98–179) and from MURST (Coordinated Project 40%, Biologia Strutturale). A. V. Stoyanov is supported by a fellowship from Istituto Nazionale di Fisica della Materia (INFM, Genova).

Received October 31, 1998

5 References

[1] Svensson, H., *Acta Chem. Scand.* 1961, *15*, 325–361.

[2] Stoyanov, A. V., Righetti, P. G., *Electrophoresis* 1998, *19*, 1596–1600.

[3] Stoyanov, A. V., Righetti, P. G., *Electrophoresis* 1998, *19*, 1733–1737.

[4] Rilbe, H., in: Catsimpoolas, N. (Ed.), *Isoelectric Focusing*, Academic Press, New York 1976, pp. 14–52.

[5] Stoyanov, A. V., Righetti, P. G., *J. Chromatogr. A* 1997, *790*, 169–176.

[6] Stoyanov, A. V., Righetti, P. G., *J. Chromatogr. A* 1998, *799*, 275–282.

[7] Righetti, P. G., Nembri, F., *J. Chromatogr. A* 1997, *772*, 203–211.

[8] Bossi, A., Righetti, P. G., *Electrophoresis* 1997, *18*, 2012–2018.

[9] Stoyanov, A. V., Righetti, P. G., *Electrophoresis* 1998, *19*, 1944–1950.

[10] Tiselius, A., *Svensk. Kem. Tidskr.* 1941, *58*, 305–310.

[11] Pettersson, E., *Acta Chem. Scand.* 1969, *23*, 2631–2635.

[12] Stenman, U. K., Graesbeck, R., *Biochim. Biophys. Acta* 1972, *286*, 243–251.

[13] Lundahl, P., Hjertén, S., *Ann. N.Y. Acad. Sci.* 1973, *209*, 94–111.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 723–726

Jun X. Yan[1]
Jean-Charles Sanchez[2]
Veronique Rouge[2]
Keith L. Williams[1]
Denis F. Hochstrasser[2]

Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW, Australia
[2]Laboratoire Central de Chimie Clinique, Hôpital Cantonal Universitaire, Genve, Switzerland

Modified immobilized pH gradient gel strip equilibration procedure in SWISS-2DPAGE protocols

In the present paper we report a revised protocol for immobilized pH gradient (IPG) gel strip equilibration involving a procedural modification between the first- and second-dimensional separation in both analytical and preparative two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). By changing the pH of the equilibration buffer (pH 8.0), the concentration of alkylating reagent (125 mM iodoacetamide) and the time of incubation (15 min), it has been possible to achieve increased cysteine (Cys) alkylation to completion with only one adduct of carboxyamidomethyl-Cys formed. Importantly, the modification does not alter the 2-D proteome patterns and therefore maintains the integrity of the existing SWISS-2DPAGE entries. Results are presented for comparative analyses using human plasma, and for Cys analysis of human albumin to illustrate the advantages of the improved protein reduction and Cys alkylation. The modified step of IPG gel strip equilibration will assist protein digestion for matrix-assisted laser desorption/ionisation - time-of-flight - mass spectrometry analysis, and make Cys quantitation possible without further in-gel or on-blot alkylation.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Reduction / Alkylation / Cysteine / Immobilized pH gradients
EL 3370

Proteomics relies on the capability to produce high-resolution proteome maps using two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). The introduction of immobilized pH gradient (IPG) strip gels for first-dimensional isoelectric focusing [1, 2] has been a key development underpinning this capability, allowing the production of highly reproducible protein arrays for both analytical and preparative purposes. The reproducibility has led to proteome maps being available in databases accessible *via* the Internet, such as the SWISS-2DPAGE database [3] (http://www.expasy.ch/ch2d/) and HSC-2DPAGE [4, 5] (http://www.harefield.nthames.nhs.uk/nhli/protein/index.html). SWISS-2DPAGE is one example of an annotated database of reference proteome maps providing an important source of information on proteins identified from biological, physiological and pathological samples. The 2-D PAGE protocols used in compiling this database are progressively updated to incorporate best practice technique advances, including those that facilitate protein characterisation. In the last two years, peptide mass fingerprinting using MALDI-TOF (matrix-assisted laser desorption/ionisation-time-of-flight) mass spectrometry has emerged as a protein identification tool particularly suited to identification of low-abundance proteins separated on proteome maps. Peptides are produced through enzymatic digestion of in-gel or on-blot target proteins. Prior to an efficient digestion, many groups have employed reduction and alkylation steps in order to break both inter- and intra-chain disulfide bonds of the protein molecules and ensure fully alkylated cysteine (Cys) residues [6, 7]. This is believed to improve the efficiency of protein digestion, thus facilitating the production of reproducible and representative peptide fingerprints.

Limited alkylation of Cys in proteins can occur during separation in polyacrylamide gels due to the interaction of Cys with residual acrylamide to form Cys-propionamide (Cys-Pam) [8], but this is generally low in IPG strips due to extensive washing of the strips during manufacture to remove free acrylamide. During loading of the first-dimensional isoelectric focusing strips to the second-dimensional SDS-PAGE, solubilization of proteins in the IPG strips is achieved by a two-step equilibration with reducing agent o reduce disulfide bonds and alkylating reagent to block-SH groups (see http://www.expasy.ch/ch2d/technical-info.html) [2]. This is normally carried out using modified stacking buffer with a reducing agent (*e.g.*, dithioerythritol, DTE) and alkylating agent (*e.g.*, iodoacetamide) at pH 6.8 so that the strips can be loaded without further pH alteration [9]. However, the optimal conditions for reduction and alkylation are alkaline (pH 8.9) [10] so that alkylation of Cys to carboxyamidomethyl-Cys (Cys-Cam) under these conditions is incomplete. Although other groups have reported their IPG strip equilibration using pH 8.5 buffer and 15 min incubation [11–13] which did not

Correspondence: Dr. Jun X. Yan, National Heart and Lung Institute, Heart Science Centre, Harefield Hospital, Hill End Road, Harefield, UB9 6JH, UK
E-mail: jun.yan@harefield.nthames.nhs.uk
Fax: +44-1895-828-900

alter the gel patterns, detailed Cys alkylation analysis has not been carried out.

In this paper the effects of optimising alkylation and reduction during the transfer of proteins from the first to the second dimension during 2-D PAGE has been investigated. Human plasma was separated by mini 2-D PAGE [14, 15]. IPG equilibration conditions before transfer to the second dimension were carried out in original or modified equilibration buffers as shown in Table 1. The modified

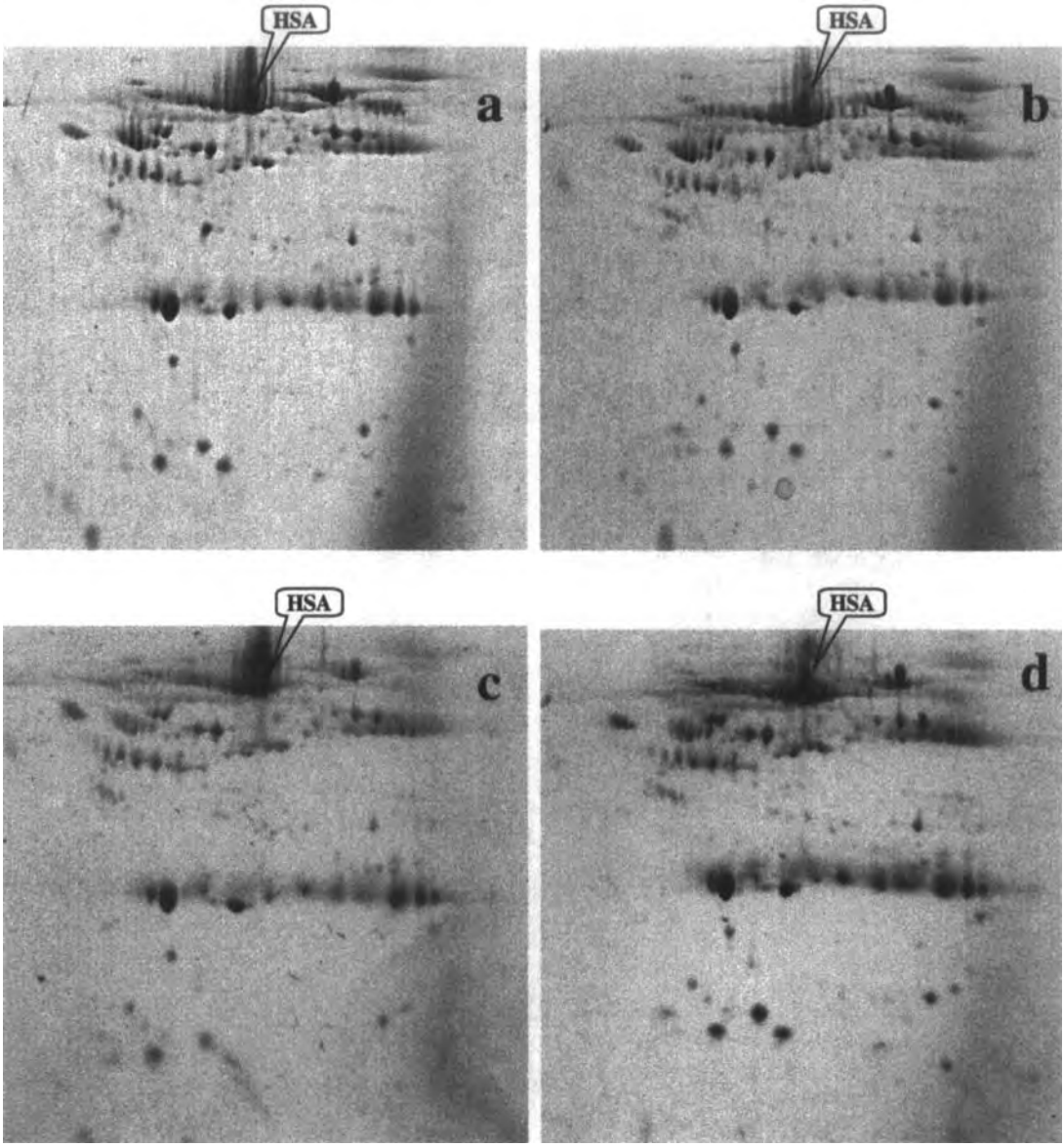


Figure 1. Comparison of images of Coomassie blue-stained gels and amido black-stained blots of human plasma 2-D PAGE. Gel (a) and blot (c) images obtained from IPG strips treated with the original equilibration method (see Table 1). Gel (b) and blot (d) images obtained from IPG strips treated with the modified equilibration method. The position of human albumin (HSA) is marked. Note the qualitative and quantitative similarities in patterns between the different equilibration methods. There is no effect of the equilibration method on protein recovery or protein transfer.

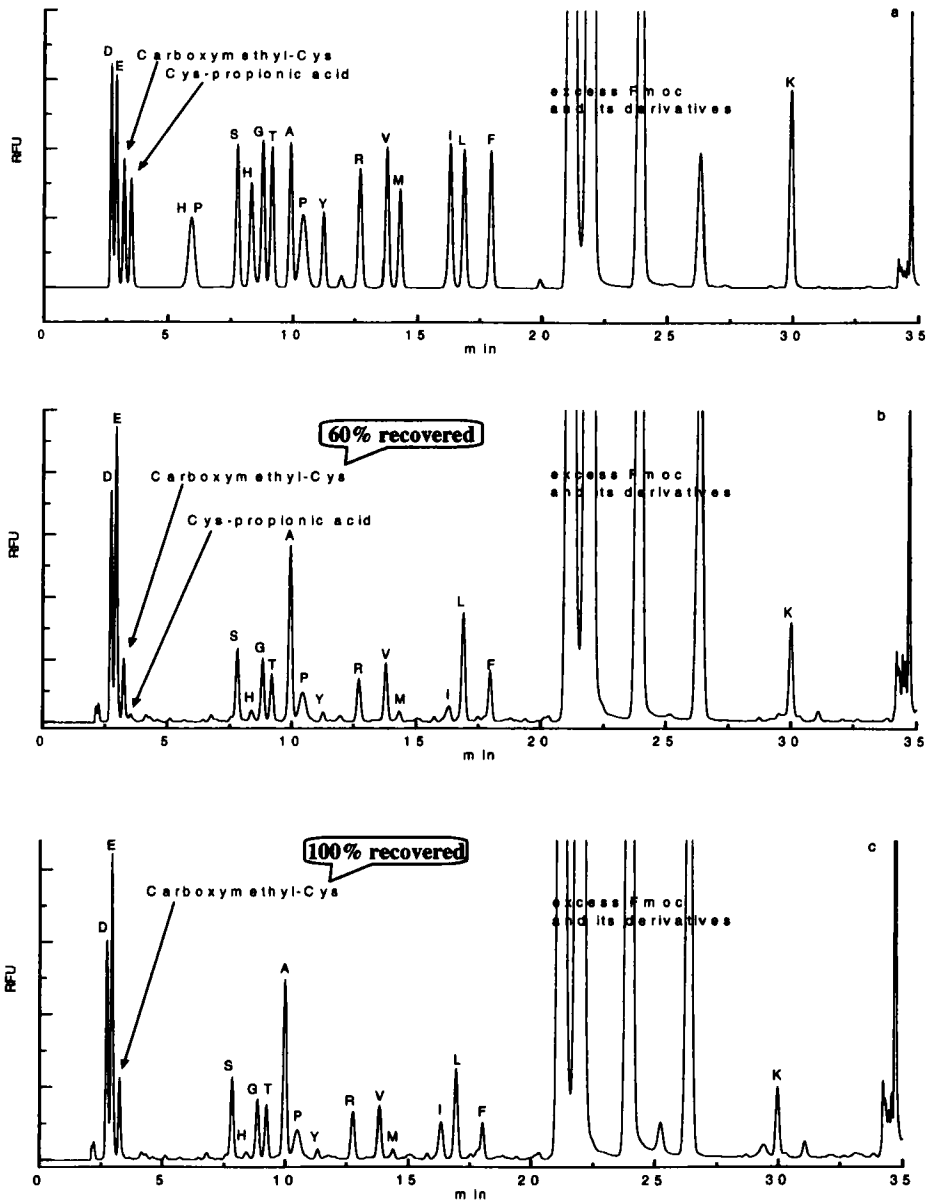


Figure 2. Chromatograms of Fmoc-amino acid separations. (a) From a standard amino acid mixture containing carboxymethyl-Cys (Cys-Cam acid) and Cys-propionic acid (Cys-Pam acid). (b), (c) From hydrolysates of human albumin, where (b) is from the 2-D PAGE using the original equilibration method and (c) is from the separation using the modified method. Note the presence of both Cys-Pam and Cys-Cam in the chromatogram of HSA prepared using the original equilibration method, but in the chromatogram prepared using the modified method only Cys-Cam is present.

conditions entailed an increase in the concentration of reducing and alkylating reagents and an increase in the pH and incubation time, but in all other respects (*e.g.*, loading, staining, *etc.*) the gels and blots were treated identically. After separation, proteins were stained with Coomassie blue [16] or blotted onto polyvinylidene difluor-

Table 1. Comparison of original and modified equilibration methods for IPG strips prior to application to the second-dimensional SDS-PAGE

Original equilibration method		Modified equilibration method	
Buffer 1	*Buffer 2*	*Buffer 1*	*Buffer 2*
65 mM DTE	65 mM iodoacetamide	125 mM DTE	125 mM iodoacetamide
50 mM Tris, pH 6.8	50 mM Tris, pH 6.8	50 mM Tris, pH 8.0	50 mM Tris, pH 8.0
6 M urea	6 M urea	6 M urea	6 M urea
30% v/v glycerol	30% v/v glycerol	30% v/v glycerol	30% v/v glycerol
2% w/v SDS	2% w/v SDS	2% w/v SDS	2% w/v SDS
10 min	5 min	15 min	15 min

ide (PVDF) membranes and stained with amido black [16]. Gels and blots were scanned and the images compared by Melanie II computing program (Bio-Rad, Uppsala, Sweden; Fig. 1). There were no significant differences in protein pattern between the two equilibration methods, and there was not significant protein loss or diffusion during the longer incubation, and the efficiency of blotting for the two treatments was the same. Modifications of Cys were determined by Fmoc (9-fluorenylmethoxycarbonyl) amino acid analysis (AAA) [10] and human albumin (HSA) was chosen as a test protein due to its high content of intrachain disulfide bond formation (see SWISS-PROT entry albu_human). HSA from plasma equilibrated in the original buffer gave two modified Cys peaks during AAA with about 60% of the total cys being recovered as Cys-Cam and a small amount of Cys-Pam also being present (Fig. 2). In the AAA of HSA treated with modified buffer there was only one Cys adduct present, with almost 100% being recovered as Cys-Cam (Fig. 2). The quantities for the remaining amino acids were essentially the same with both treatments.

In conclusion, equilibration of IPG strips in buffer with a higher pH, higher concentrations of reducing and alkylating reagents, and increased incubation time (see Table 1) results in an increase in the amount of modified Cys. There was nearly 100% conversion of Cys to Cys-Cam; thus further in-gel or on-blot reduction and alkylation is unnecessary and the number of potential products formed during peptide mass fingerprinting will be fewer, thus simplifying analysis. The treatment does not affect the 2-D PAGE protein pattern, which is in agreement with the comparison results reported previously [13], so that adoption of this methodology will be compatible with the existing SWISS-2DPAGE databases.

JXY was supported by the Australian Proteome Research and Development Grant, and would like to thank Professor Denis Hochstrasser and Dr. Mike Dunn for allowing her to work in their laboratories. JCS and DFH were supported by the Swiss Fund of Scientific Research (grant 32-49314.96)

Received January 6, 1999

References

[1] Bjellqvist, B., El, K., Righetti, P. G., Gianazza, E., Görg, A., *J. Biochem. Biophys. Methods* 1982, *6*, 317–339.

[2] Bjellqvist, B., Hughes, C., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1367.

[3] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[4] Corbett, J. M., Wheeler, C. H., Baker, C. S., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1994, *15*, 1459–1465.

[5] Evans, G., Wheeler, C. H., Corbett, J. M., Dunn, M. J., *Electrophoresis* 1997, *18*, 471–479.

[6] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[7] O'Connell, K. L., Stults, J. T., *Electrophoresis* 1997, *18*, 349–359.

[8] Chiari, M., Righetti, P. G., Negri, A., Ceciliani, F., Ronchi, S., *Electrophoresis* 1992, *13*, 882–884.

[9] Dunn, M. J., *Gel Electrophoresis: Proteins*, Bios Scientific Publishers, Oxford 1993, pp. 34–35.

[10] Yan, J. X., Kett, W. C., Herbert, B. R., Gooley, A. A., Packer, N. H., Williams, K. L., *J. Chromatogr. A* 1998, *813*, 187–200.

[11] Hughes, G. J., Frutiger, S., Paquet, N., Ravier, F., Pasquali, C., Sanchez, J.-C., James, R., Tissot, J. D., Bjellqvist, B., Hochstrasser, D. F., *Electrophoresis* 1992, *13*, 707–714.

[12] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[13] Görg, A., Celis, J. E. (Eds.), *Cell Biology*, Academic Press, New York 1994, Vol. 3, pp. 231–242.

[14] Görg, A., Boguth, G., Obermaier, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16*, 1079–1086.

[15] Sanchez, J.-C., Hochstrasser, D. F., Link, A. J., (Eds.), *Methods in Molecular Biology, 2-D Proteome Analysis Protocols*, Humana Press Inc., Totowa, NJ 1998, Vol. 112, pp. 227–233.

[16] Sanchez, J.-C., Ravier, F., Pasquali, C., Frutiger, S., Paquet, N., Bjellqvist, B., Hochstrasser, D. F., Hughes, G. J., *Electrophoresis* 1992, *13*, 715–717.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 727–731

Joël S. Rossier
Alexandra Schwarz
Frédéric Reymond
Rosaria Ferrigno
François Bianchi
Hubert H. Girault

Laboratoire d'Electrochimie,
Ecole Polytechnique
Fédérale de Lausanne,
Lausanne, Switzerland

Microchannel networks for electrophoretic separations

UV excimer laser photoablation was used to micro-machine polymer substrates not only to drill microchannel structures but also to change the surface physical properties of the substrates. We first describe how UV laser photoablation can be used for the patterning of biomolecules on a polymer and discuss parameters such as surface coverage of active antibodies and equilibration time. Secondly, we show how to design a single-use capillary electrophoresis system comprising an on-chip injector, column and electrochemical detector. The potential of this disposable plastic device is discussed and briefly compared to classical systems. Finally, preliminary results on protein separation by isoelectric focusing on a disposable microchip are presented.

Keywords: Photoablation / Disposable plastic microchip / Capillary electrophoresis / Protein separation / Immunoassay
EL 3403

1 Introduction

The increased demand for protein analysis and rapid drug discovery requires the development of novel automated techniques. The new trends to overcome the cumbersome and expensive methodologies currently used in laboratories are miniaturization and integration, providing fast screening of chemical and biochemical information on a routine basis. Intense efforts have been made to reduce whole laboratory systems onto microchip substrates, often utilizing capillary electrophoresis (CE) as the principal analytical technique [1]. These miniaturized systems have been termed microscale total analysis systems (μ-TAS) [2] and, due to high demands on contamination-free handling, they are often designed for single use only. Most μ-TAS devices to date have been produced photolithographically on substrates such as glass, quartz and silicon [3]. These devices are fabricated under clean room conditions and are normally sealed by thermal or anodic bonding to a glass cover that renders difficult the immobilization of biochemical compounds.

The UV excimer laser photoablation technique has been used in our laboratory as an alternative method for the development of microdiagnostic systems [4] in polymers such as polyethylene terephthalate (PET) or polycarbonate. Photoablated polymer channels and channel networks are further sealed by a conventional lamination process that does not require high sealing temperatures [4]. Excimer laser etching relies on the generation of high energy light pulses that are fired in nanosecond bursts and focused onto polymers possessing suitable UV absorption characteristics. The incident photons rapidly break chemical bonds within a confined space, thereby leading to a localized pressure rise resulting in mini-explosions and in the ejection of ablated material. Photoablated polymers exhibit larger surface charges and adjustable roughness or hydrophilic properties with respect to non-ablated material, and have been characterized for their ability to generate capillary as well as electroosmotic flow (EOF) [4, 5].

In this paper, we present several separation and detection methods that can be implemented on polymer substrates. First, we show that proteins can be patterned on polymer surfaces modified by photoablation in various modes maintaining part of their activity. Second, antibodies are immobilized within μ-channels, and a fast immunoassay is carried out. Third, injection, separation, and detection on photoablated μ-channel networks are described in order to demonstrate how a disposable capillary electrophoresis μ-system can be designed. Finally, electrophoretic separation of a protein mixture is achieved by isoelectric sieving, thereby showing how the conventional isoelectric focusing methodology can be transposed to a miniaturized device.

Correspondence: Prof. Hubert H. Girault, Laboratoire d'Electrochimie, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
E-mail: hubert.girault@epfl.ch
Fax: +41-21-6933667

Abbreviations: ALP, alkaline phosphatase; **DDi,** D-dimer; **PET,** polyethylene terephthalate; **μ-TAS,** microscale total analysis system

2 Materials and methods

2.1 Photoablation

The microfabrication was achieved as already described [4] by UV laser photoablation of commercially available polymers such as PET (Mylar type D from DuPont, Gene-

va, Switzerland) or polycarbonate. Briefly, 100 μm thick PET sheets were rinsed with water and ethanol and then mounted on a computer controlled *X, Y* stage (Microcontrol, Evry, France). Then, UV laser pulses (193 nm) were fired onto the polymer substrate target through a homogenizer, a copper photo mask, and a 10:1 reduction lens with a frequency of 50 Hz at 200 mJ/pulse. During the photoablation process, the polymer substrate was moved horizontally on the *X, Y* stage over 1–20 cm, thereby determining the length L_c of the μ-channels. The rate of displacement of the *X, Y* stage was set between 150 and 200 μm/s, allowing modulation of the μ-channel depth H_c between 15 and 60 μm. The copper masks were 1 cm over 0.3–2 mm rectangles, so that the μ-channel width W_c could be chosen between 30 and 200 μm. Reservoirs were opened at each extremity of the μ-channels by firing sufficient pulses to penetrate the whole polymer sheet. After washing with water and ethanol and drying under pressurized air, the μ-structures were sealed by a low temperature lamination process using a standard industrial lamination apparatus from Morane Senator (pneumatic control; Oxon, UK). The laminate used was a 35 μm thick PET film covered on one side by a 5 μm thick polyethylene (PE) adhesive layer (purchased from Morane, Oxon) that is thermally annealed to the base polymer at 125°C for 1–2 s.

2.2 Biopatterning

The patterning of proteins onto polymer substrates was achieved by the so-called lamination technique [6]. The substrate sheets (approximately 5 × 5 cm squares) were first washed with water and ethanol and directly laminated with the above PET/PE film that plays the role of a protective layer. The above photoablation procedure was then used to structure rectangular channels (L_c = 1 mm, H_c = 50 μm and W_c = 20, 40, 100 and 200 μm) through the protective laminate using the corresponding copper masks. The patterned sheets were then incubated with avidin for 2 h (100 μg/mL avidin; Sigma, St. Louis, MO, USA in PBMS solution, *i.e.*, 100 mM phosphate and 1 mM $MgCl_2$ at pH 7). After three washing steps with 200 μL PBMS, the substrates were incubated with biotin-labeled fluorescein (100 μg/mL in PBMS; Sigma), either overnight at +4°C or for 2 h at room temperature. As a last step, after washing (three times with 200 μL PBMS) and drying of the substrates, the lamination was manually peeled off and fluorescence imaging was carried out with an Axiovert LSM 410 microscope (Carl Zeiss, Jena, Germany).

2.3 D-Dimer immunoassay

Protein immobilization in the μ-channels was achieved by physisorption. The channels (L_c, 2 cm; H_c 50 μm; and W_c, 100 μm) were incubated for 1 h with 10 μg/mL of anti-D-dimer (anti-DDi) 2F7 antibody (Serbio, Gennevilliers, France). Blocking against nonspecific adsorption was carried out with 2% bovine serum albumin (BSA) (Sigma) for 1 h, and was further controlled by fluorescence imaging. Then, serial dilutions of DDi-ALP (DDi-alkaline phosphatase) were incubated for 5 min. The Vistra ECF substrate solution was finally filled and the fluorescence in the channels was measured with Storm Imager 840 (Molecular Dynamics, Sunnyvale, CA, USA). Washing steps were performed between each incubation step.

2.4 Incorporation of carbon ink electrodes

Conducting carbon paste electrodes were implemented at the end of a μ-channel (L_c, 1.7 cm, H_c, 40 μm; W_c, 40 μm) specially designed for electrophoretic separation; Fig. 1 shows the prototyping design chosen here for a single-use capillary electrophoresis μ-system. The injector is either composed of a cross (as in Fig. 1) or a double T, allowing highly reproducible injections of hundreds of picoliters. Two mM aminophenol (Fluka, Buchs, Switzerland) in 10 mM phosphate buffer at pH 8 (Fluka) was used as electroactive sample to test the final device. The electrochemical detector is composed of conductive carbon bands of 40 μm in width that are placed either on both vertical walls of the μ-channel ("face-to-face" configuration as shown in Fig. 1) or at the bottom of the μ-channel ("inlaid bands" separated by 100 μm, not shown here; Rossier *et al.*, submitted). For the integration of carbon electrodes, the μ-channels were rinsed with water and ethanol after the photoablation procedure. Then, the μ-channels were filled

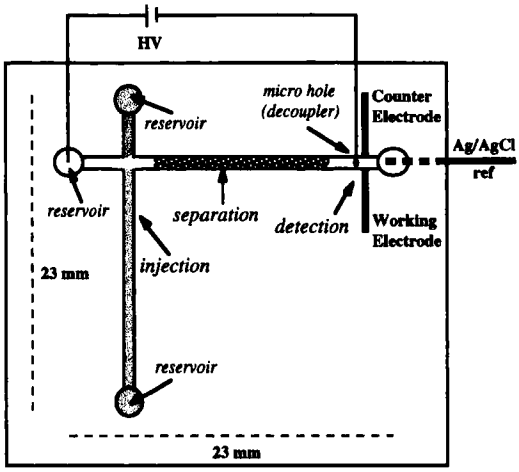


Figure 1. Schematic representation of a μ-chip with integrated injector, separation column and electrochemical detector.

with an excess of carbon ink. The whole structure was cured at 70°C for 2 h. Then, a thin layer of photoresist (1 μm) was spin-coated and cured. Finally, the main channel was laser-machined at the opposite side of the devices and the structure was sealed by a PET/PE lamination (Rossier *et al.*, submited). Pulse amperometry was performed with a computer-controlled potentiostat (AEW2-type; Sycopel Scientific, Oxford UK) in two or three electrode modes. The carbon electrodes in the channel served as working electrode (WE) and counter electrode (CE), whereas the Ag/AgCl reference electrode was placed in the reservoir drop outside the μ-channel. For the detection of aminophenol, the carbon electrodes were polarized between 0 and 400 mV against the Ag/AgCl reference.

2.5 Isoelectric sieving

Protein separation was carried out in laminated PET μ-channels (2 cm long, 100 μm wide and 50 μm deep) comprising a reservoir at each extremity. The channels were filled by capillary flow with an acrylamide/bisacrylamide solution (Sigma) already containing ammonium persulfate and *N,N,N',N'*-tetramethylethylenediamine (TEMED; Sigma) to give 6%T and 4%C. The pH of the gel was set to 5.4 by addition of 2.5% and 2.55% v/v of Immobilines (Pharmacia, Uppsala, Sweden) of pK_a 4.6 and pK_a 6.2, respectively. The polymerization was achieved in a humidification chamber at 50°C for 1 h. The anode reservoir was filled with a mixture of 1 mg/mL in water of both cytochrome *c* and β-lactoglobulin (Sigma). A constant voltage was then applied between the two reservoirs with the following power supply (CZE 1000R; Spellman, Hauppeange, NY USA): 400 V/cm were applied during the first 30 min of the experiments, and the potential was then switched to 500 V/cm for the last 15 min. At the cathode reservoir, the sample was collected at different times and analyzed by capillary electrophoresis with a Biofocus 3000 apparatus (Bio-Rad, Richmond, CA, USA).

3 Results and discussion

3.1 Patterning and adsorption of proteins on photoablated polymers

The photoablation technique was used here to create well-defined regions in which biomolecules can be adsorbed on a polymer substrate. To this aim, avidin was patterned on PET using the so-called lamination technique [6] (see Section 2.2) and further detected by fluorescence microscopy after incubation with fluorescein-labeled biotin. Microchannels separated by various distances were thus drilled by photoablation through the lamination, and the resulting dimensions of the patterned surface were calculated from fluorescence images (Fig. 2) as 34.4 (± 4.0) μm × 1 mm for the avidin line patterns and 18.8 (± 2.2) μm as smallest distance between two channels. Note that no significant fluorescence signal was observed during the control experiments in which the μ-channels contained no avidin. Therefore, the fluorescence measured in Fig. 2 results from biotin-fluoresceine bound to active immobilized avidin. This demonstrates that receptor ligand binding of avidin-biotin couples can be used as a model molecular recognition system for patterning surfaces. Furthermore, Fig. 2 shows that the resulting line-patterns in PET are very well defined, which

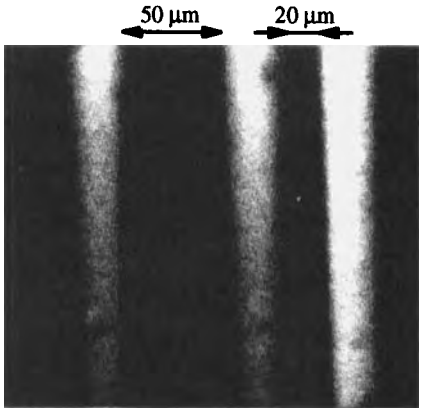


Figure 2. Fluorescence microscope picture of 20 μm wide μ-channels filled by lamination technique with avidin marked by fluorescein-labeled biotin. (The patterned μ-channels are parallel despite optical distortion).

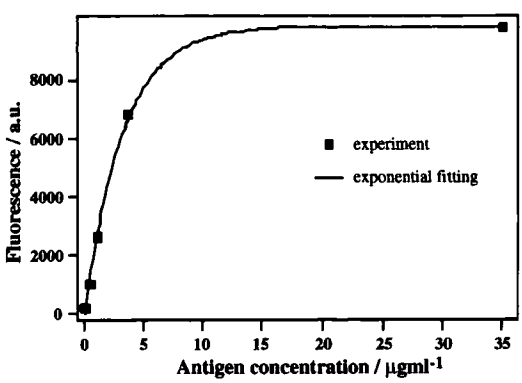


Figure 3. Detection of ALP-labeled antigen in polycarbonate μ-channels. Ten μg/mL of mouse anti-D-dimer antibody is first adsorbed for 1 h. After a washing step, the surface is blocked with a 2% BSA solution. Serial dilutions of ALP-labeled antigen are added to each channel after adsorption of anti-DDi followed by BSA blocking as detailed in Section 2.3.

signifies that no leakage underneath the lamination occurred during the incubation steps and that the lamination is completely removed during the peel-off process. It has yet been noted that the limiting distance between the patterns is ~20 μm. Apart from its excellent precision, one important advantage of the method is the unlimited variety of patterns that can be drawn.

Comparison for the adsorption of ^{14}C-labeled BSA on nonlaser-treated polymer surfaces [6] also showed that enhanced hydrophobicity and roughness of photoablated polymers increase the amount of adsorbed proteins by 2- to 3-fold. The maximum value obtained for adsorbed BSA on photoablated polyimide was ~600 ng/cm^2, which is slightly larger than the expected value for a monolayer of adsorbed BSA molecules due to increased surface area during laser treatment. Avidin-biotin binding can be easily extended to a wide variety of biomolecules such as antibodies, enzymes or DNA, and the methodology used here can also be used to adsorb protein within closed μ-channels, in which fluid handling is then possible. Figure 3, for example, shows the adsorption isotherm that was obtained by fluorescence measurements of the enzymatic product, which was generated by ALP-labeled antigens bound to mouse anti-DDimer antibodies adsorbed in polycarbonate μ-channels. Anti-DDi antibodies were not desorbed during this parallel analysis, which may then be used as a powerful method for fast immunoassays. Miniaturization of such immunoassay devices results in short equilibration times. In a first approximation, the time required for a molecule to diffuse from one wall of the μ-channel to the other is given by: $L = \sqrt{D\,t}$, where L is the μ-channel width and D is the diffusion coefficient of the compound of interest. A typical equilibration time of 5 min for the specific adsorption of antigens on immobilized anti-DDi has been measured. This is much faster than the classical microtiter plates for which the typical equilibration time is about 1 h.

3.2 Photoablated microchannel network for miniaturized capillary electrophoresis

Another advantage of miniaturized systems obtained by photoablation of PET is their ability to generate EOF when a buffer solution is placed in a high electric field [4]. A special effort has been made to integrate an electrochemical detector in the plastic μ-chips, allowing a reproducible placement and sensitive detection of electroactive species (Rossier *et al.*, submitted). A critical parameter of the electrochemical detection in an EOF-driven μ-device is the decoupling of the high voltage current path from the mass flow arriving at the detector. Indeed, in order to minimize distortion of the measurements, integrated electrochemical detection necessitates minimal interference

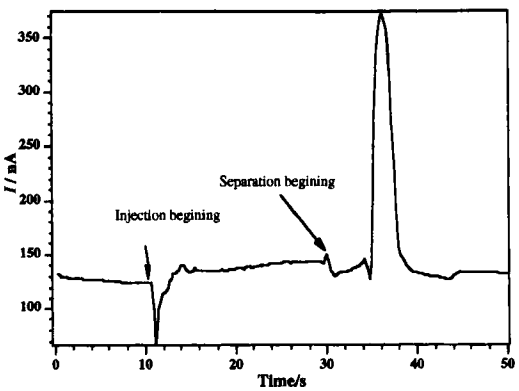


Figure 4. Detection of a plug of 2 mM aminophenol in 10 mM phosphate buffer by pulse amperometry (0–400 mV *vs.* Ag/AgCl), using the μ-electrophoresis devices shown in Fig. 1. From 0 to 11 s, no potential was applied in the μ-channels; from 11 to 30 s, 150 V/cm were applied across the injection column; after 30 s, the voltage was switched across the separation column at 900 V/cm for 5 s before being cut-off.

from the separation electric field [7]. We therefore designed a novel type of electrical decoupler which consists of an array of micro-holes of 5 μm diameter pierced through the PET/PE laminate. This decoupler is placed 1 mm in front of the electrodes, so that the effective separation length in the μ-device of Fig. 1 is 1.7 cm. The advantage of our methodology over conventional decoupling/detection systems is the fixed, stable, and reproducible placement of the electrodes as well as the possibility of working in three electrode mode to reduce the current path and hence the ohmic drop (Rossier *et al.*, submitted).

The working principle of the electrophoretic μ-system is shown (Fig. 4) by the detection of a plug of 2 mM aminophenol in 10 mM phosphate buffer. This plug was first injected by applying 150 V/cm at the extremities of a simple cross before switching the potential to the separation column in which a typical electric field of 900 V/cm was applied across the whole column. This electric field generated a sufficiently high EOF to rapidly let the plug reach the two inlaid carbon bands of 40 μm width that served as working and counter electrodes (approximately 5 s were needed by the aminophenol plug to cover the 1.8 cm separating the injector from the detector). This experiment demonstrates that the configuration shown in Fig. 1 is suitable for correct and reproducible electrochemical measurements and that it provides sensitive and fast analysis. Work is now in progress to improve each part of these μ-CE devices so as to reach theoretical plates similar to those already obtained with other μ-systems.

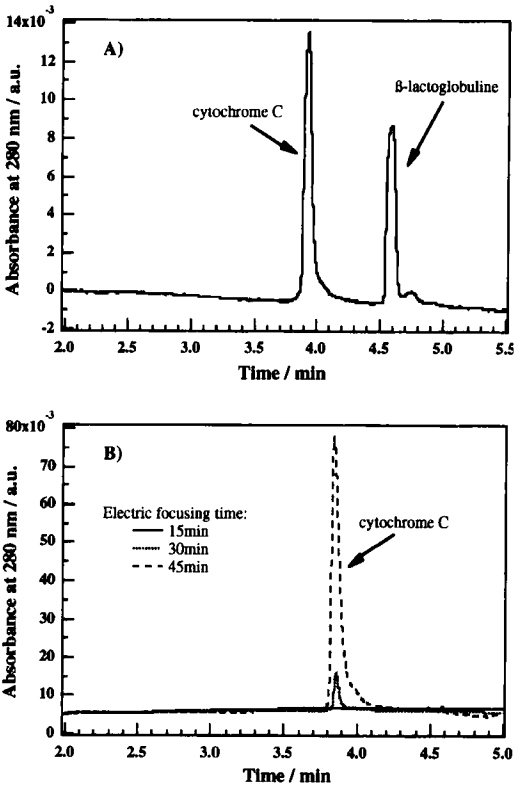


Figure 5. Electropherograms obtained from the reservoir solutions before and during isoelectric sieving of a mixture of cytochrome *c* and β-lactoglobulin following the procedure described in Section 2.5. (A) Electropherogram for a mixture of 1 mg/mL of both cytochrome *c* and β-lactoglobulin in the anodic reservoir at $t = 0$. (B) Electropherograms for cytochrome *c* collected in the cathodic reservoir of a μ-channel after separation from β-lactoglobulin with dependence on time (applied potential: 400 V/cm for 30 min and then 500 V/cm for 15 min).

3.3 Isoelectric sieving in photoablated PET microchannels

Part of our work is devoted to the transposition of 2-D gel electrophoresis to a miniaturized system specially designed to recover the separated proteins in a free-flowing solution. IEF is used here for protein separation, and the device consists of a network of μ-channels separated by reservoirs and filled with conventional polyacrylamide gels. The pH is different in each section of the global structure and can be fixed by addition of an adequate amount of Immobilines to the various polyacrylamide solutions. Electrophoresis allows then to entrap the proteins of interest in the reservoir corresponding to their respective isoelectric point (p*I*), therefore separating them from the rest of the sample. All proteins with a different p*I* from that defined by the gels either do not reach the reservoir or pass it.

Preliminary results obtained for the separation of β-lactoglobulin (p*I* 5.2) and cytochrome *c* (p*I* 9.6) are shown in Fig. 5. Before application of the electric field, CE analysis of the anodic reservoir reveals the presence of both proteins (see Fig. 5A), whereas the evolution of the cathodic solution with time (Fig. 5B) clearly shows the progressive concentration increase of cytochrome *c* and the absence of β-lactoglobulin. The main advantage of the present technique is that sample transfer is easy to perform, which means that separated proteins can be further separated, purified, or analyzed.

4 Concluding remarks

We showed the working principles of photoablated polymers to create analytical μ-devices. The patterning of active biomolecules, the development of fast immunoassays, and the use of electrophoresis and electrochemistry in μ-systems make it possible to design powerful polymer devices. The simplicity and easy handling of the described techniques facilitate their application in microdiagnostic devices. Further improvements are in progress in our laboratory. Polymer substrates are suitable for the manufacture of μ-TAS and disposable sensing devices, and laser photoablation is a rapid and convenient method for prototyping such μ-systems. Moreover, the low temperature and rapid sealing also allow the predeposition of sensitive biological reagents that have been shown to retain their desired activity after μ-channel closure.

The authors gratefully acknowledge the financial support given by the SPP Bio program of the Swiss National Science Foundation and the MedTech initiative of the Commission for Technology and Innovation.

Received September 1, 1998

5 References

[1] Arquint, P., Koudelka-Hep, M., van der Schoot, B., van der Wal, P., de Rooij, N. F., *Clin. Chem.* 1994, *40*, 1805–1809.

[2] Manz, A., Graber, N.,. Widmer, H. M., *Sens. Actuators B* 1990, 244.

[3] Manz, A., Becker, H., *Microsystem Technology in Chemistry and Life Science*, Springer, Berlin 1998.

[4] Roberts, M. A., Rossier, J. S., Bercier, P., Girault, H., *Anal. Chem.* 1997, *69*, 2035–2042.

[5] Rossier, J. S., Bercier, P., Schwarz, A., Loridant, S., Girault, H. H., *Topography, Cristallinity and Wettability of Photoablated PET Surfaces, Langmuir*, in press.

[6] Schwarz, A., Rossier, J. S., Roulet, E., Mermod, N., Roberts, M. A., Girault, H. H., *Langmuir* 1998, *14*, 5526–5531.

[7] Woolley, A. T., Lao, K., Glazer, A. N., Mathies, R. A., *Anal. Chem.* 1998, *70*, 684–688.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 732–737

Lila Castellanos-Serra[1]
Wilfredo Proenza[1]
Vivian Huerta[1]
Robert L. Moritz[2]
Richard J. Simpson[2]

[1]Center for Genetic Engineering and Biotechnology (CIGB) Havana, Cuba
[2]Joint Protein Structure Laboratory, Ludwig Institute for Cancer Research (Melbourne) and the Walter and Eliza Hall of Medical Research, Parkville, Victoria, Australia

Proteome analysis of polyacrylamide gel-separated proteins visualized by reversible negative staining using imidazole-zinc salts

Identification and characterization of proteins isolated from natural sources by polyacrylamide gel electrophoresis has become a routine technique. However, efficient sample proteolysis and subsequent peptide extraction is still problematic. Here, we present an improved protocol for the rapid detection of polyacrylamide gel-separated proteins, *in situ* protein modification, proteolytic digestion and peptide extraction for subsequent protein identification and characterization by capillary high-performance liquid chromatography/tandem mass spectrometry. This simple technique employs the rapid imidazole-zinc reverse stain, in-gel S-pyridylethylation and proteolytic digestion of microcrushed polyacrylamide gel pieces with proteases. This technique obviates the need for buffer exchange or gel lyophilisation due to all of the sample manipulation steps being carried out at near neutral pH and thus lends itself readily to automation.

Keywords: Reversible negative staining / Mass spectrometry / Capillary column chromatography
EL 3410

In 1996, Grimm and Eckerskorn [1] reported a protocol for the digestion of Coomassie blue visualized proteins in acrylamide gel bands previously crushed to yield microparticles of 32 μm [2]. By reducing the gel particle size, the diffusion processes (both of protease into the gel, and of proteins and generated peptides from the gel) was facilitated. Nevertheless, due to fixation during the gel staining process that may provoke protein precipitation, in-gel proteolysis generally requires prolonged digestion times (18–24 h) and several extraction steps for optimal peptide recovery. Previously, we have shown that proteins detected with highly sensitive imidazole-zinc reverse staining (5–10 ng/band) [3], under fully reversible fixative conditions, can be rapidly (2 × 10 min) eluted from 32 μm acrylamide gel microparticles with yields in the range of 90–98% [4, 5]. In this report, we describe the further improvement of the protocol for in-gel microdigestion of reverse-stained proteins isolated by polyacrylamide (PA) gel electrophoresis, and its application to proteomics, as exemplified by the rapid identification of PA gel separated proteins by capillary column liquid chromatography/tandem ion-trap mass spectrometry (LC-MS/MS).

A detailed procedure for the in-gel microdigestion of reverse-stained proteins, outlined schematically in Fig. 1, is as follows.

Correspondence: Dr. Richard J. Simpson, Institute for Cancer Research, P.O. Box 2008, Royal Melbourne Hospital, Parkville, Victoria, Australia 3050
E-mail: richard.simpson@ludwig.edu.au
Fax: +61-3-9341-3192

Abbreviation: PA, polyacrylamide

Step 1, electrophoresis: Proteins are loaded onto a PA gel in Laemmli sample buffer [6] either without reducing agent or fully reduced and alkylated in urea solution prior to loading. Alternately, proteins can be S-pyridylethylated with 4-vinylpyridine *in situ* in the PA gel following electrophoresis [7]. It is well recognized that incomplete S-S reduction/alkylation can lead to in-gel reaggregation of proteins, which can affect the efficiency of proteolytic digestion. Thus, special care should be taken during this step to assure full derivatization of cysteine residues.

Step 2, detection of proteins: Proteins are detected by imidazole-zinc staining. After SDS-PAGE, rinse the gel in distilled or polished water for 30 s and then incubate in 0.2 M imidazole, 0.1% SDS for 15 min. Once completed, discard the solution and incubate the gel in 0.2 M zinc sulphate until the gel background becomes a deep white consistency (~ 30 s), leaving the protein bands transparent and colourless. Stop staining by rinsing the gel with abundant distilled water. Care should be taken not to overstain the gel. This process can be monitored during development by placing the gel over a black surface as described [3–5].

Step 3, PA gel conditioning: Immediately after electrophoresis, the bands of interest are excised using a clean scalpel blade and incubated for 2 × 8 min in 1 mL 50 mM Tris buffer, 0.3 M glycine, pH 8.3, containing 30% acetonitrile. The gel pieces are then washed briefly (2 × 1 min) with 1 mL of digestion buffer (see below). Using radioiodinated total lysate proteins from *Escherichia coli* separated on a 12.5% PA gel, losses from the gel during this step were determined to be only 0.5–1%.

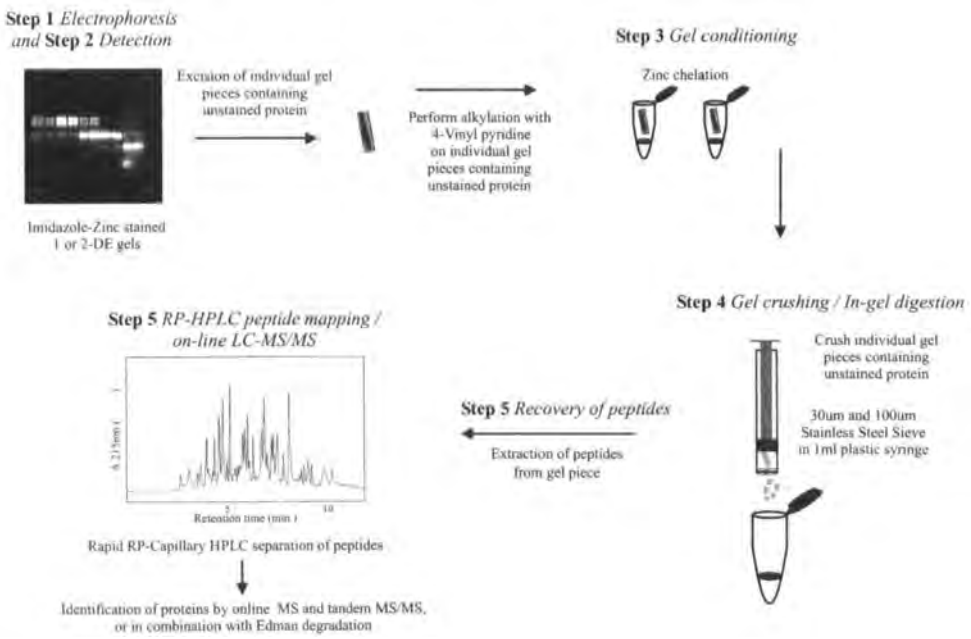


Figure 1. Summary of reverse staining/in-gel proteolysis/peptide mapping procedure.

Step 4, gel crushing/digestion: The gel band(s) are crushed by forcing them through a 1 mL syringe equipped with metal sieves of 100 and 32 μm mesh (Cat. No. FGE1000290 and FGE0320250, respectively, from F. Carl Schroeter, Hamburg, Germany; see [1, 2, 5]), to yield 32 μm gel particles, and collected into a 1.5 mL polypropylene tube. To flush out any remaining gel particles from the sieves, 30–60 μL of digestion buffer (see below) containing 0.3–0.5 μg of protease was added to the crushing syringe and collected into the same 1.5 mL tube. The volume of solution should be sufficient to cover the gel slurry. The mixture is then vortexed for 10 min and incubated at 37°C for 4–5 h. During the incubation step, care should be taken to ensure that the digestion solution fully covers the gel slurry; if necessary, additional buffer should be added. Recommended digestion buffers: trypsin, 50 mM NH_4HCO_3; Glu-C endoproteinase, 25 mM NH_4HCO_3, pH 7.8; Asp-N endoproteinase, 50 mM Na-phosphate, pH 8.0; Lys-C endoproteinase, 50 mM Tris-HCl, pH 9.0.

Step 5, recovery of peptides (elution volumes at each step are approximately twice the volume of crushed gel): The gel/digestion mixture is centrifuged (~ 5 s) and the overlaying solution collected and placed in a polypropylene tube (Eppendorf). Additional digestion buffer (~ twofold the crushed gel volume) is added and the gel mixture vortexed for 10 min, then centrifuged. The overlaying solution is collected and combined with the first extract. Then, a crushed gel volume of 50% acetonitrile, 0.1% TFA is added and vortexed for 10 min. The solution is collected and combined with the other extracts. If highly hydrophobic peptides are suspected to be present, the second extraction reagent (50 % acetonitrile) is replaced with an aqueous solution containing 20% formic acid, 15% isopropanol, 25% acetonitrile and extract for 10 min. Finally, 80% acetonitrile is added and vortexed vigorously for 2 min; then the solution is centrifuged and collected and the extracts are combined and concentrated by centrifugal evaporation.

Table 1 shows the recovery of tryptic peptides of radioiodinated bovine serum albumin (BSA) under different extraction conditions. The suspension formed by the gel microparticles and the solution behaves as a two-phase liquid-liquid extraction system. Since peptide diffusion is not a limiting factor, the relative volume of the phases as well as the number of extraction steps is more significant than the extraction time. Very high recoveries are obtained after a short incubation either in the digestion buffer or in aqueous-organic solutions. The use of 80–100% acetonitrile is not recommended during the early steps of elution, since it causes the gel particles to dehydrate and collapse, probably trapping some peptides inside. Nevertheless, after completion of the elution step, an additional

Table 1. Recovery of radioiodinated S-carboxamidomethyl-BSA digestion peptides from acrylamide gel microparticles

First eluant	Time (min)	Second eluant	Time (min)	Elution yield[a] (%)
Digestion buffer	1 × 10	Water	1 × 1	85 (1.2)
Digestion buffer	2 × 5	Water	1 × 1	96 (0.8)
ACN 15%, 0.1% TFA	2 × 5	Water	1 × 1	89 (1.3)
ACN 30%, 0.1% TFA	2 × 5	Water	1 × 1	92 (1.0)
ACN 45%, 0.1% TFA	2 × 5	Water	1 × 1	92 (0.8)
ACN 60%, 0.1% TFA	2 × 5	Water	1 × 1	96 (1.1)
ACN 50%, 0.1% TFA	1 × 10	FAPW[b]	1 × 10	98 (0.7)
ACN 100%, 0.1% TFA	2 × 5	Water	1 × 5	84 (2.4)

a) Radioiodinated S-carboxamidomethyl-BSA (3 μg per band, about 50 pmol) was loaded on a 12.5% SDS-PA gel. After detection by reverse staining, the bands were digested for 4 h according to the present procedure. Each row corresponds to an independent experiment and shows the elution yields under the extraction conditions described. Values correspond to the average from triplicate experiments (the deviation is shown in parentheses).

b) FAPW: 20% formic acid, 25% acetonitrile, 15% isopropanol, 40 % water.

incubation of the gel particles in 80–90% acetonitrile may be useful for recovering any solution remaining inside the gel particles. The efficiency of this procedure has also been confirmed by a careful study of sequence coverage obtained by MALDI-TOF-MS analysis of PA-separated model proteins. For example, using the above described reverse staining and in-gel digestion protocols, 98% sequence coverage was obtained for tryptic peptides from streptokinase (47 kDa, 0.5 μg loaded on gel), and 91% for *N*-deglycosylated erythropoietin (18.3 kDa, 3 μg loaded on gel; L. J. Gonzalez and J. P. Le Caer, personal communication). Representative peptide maps for the 47 kDa protein streptokinase, after 4 h digestion with four proteases (endoproteinase Glu-C, trypsin, endoproteinase Lys-C, endoproteinase Asp-N; Fig. 2). A remarkable similarity between in-gel and in-solution proteolysis is shown for the Glu-C peptide maps (Fig. 2A and B). In contrast to staining with Coomassie blue, the analysis of a gel blank from the reverse staining procedure shows no artifact peaks (see Fig. 2F).

The following procedure was applied for the RP-HPLC separation of peptides recovered from in-gel digestion of ZnII reverse-stained visualized proteins (Step 6, Fig. 1). The combined peptide extracts (see above) were adjusted to pH 2.0 by the addition of 0.1% aqueous TFA (monitored by spotting ~ 0.1 μL of the extract onto pH paper) and centrifuged briefly (1–5 s). For conventional narrow bore (< 2.1 mm ID) chromatography, an in-line filter (2 μm) was placed between the injector and the column. For microcolumn (< 0.3 mm ID) chromatography, a standard Hewlett-Packard (Avondale, PA, USA) HPLC was modified (Fig. 3), as described elsewhere [7], to achieve the low flow rates (~ 1.5 μL/min) required for capillary column chromatography. Note that for the microcolumn HPLC configuration, an in-line filter between the injector and the column was not used and, therefore, care should be taken when filling sample-loading syringes to avoid microgel particles.

For the MS analysis of peptides by electrospray (ESI) tandem mass spectrometry, a Finnigan-MAT LCQ ESI ion-trap mass spectrometer (San Jose, CA, USA) was employed as described elsewhere [8]. Briefly, peptides obtained from combined in-gel digest extracts were introduced into the ESI ion trap using a 0.2 mm ID column packed in-house [6] with 5 μm particle diameter, 300 Å pore size, C8 packing (Brownlee RP-300) and operated at 1.6 μL/min. The standard Finnigan-ESI source was used with a sheath flow of 2-methoxyethanol introduced at 3 μL/min, with a sheath bath gas of nitrogen set at 30 (arbitrary unit). The ESI needle was set at +4.5 kV and the heated capillary was set at 150°C. Spectra were recorded in centroid mode with three microscans summed per scan. The spectrometer was operated in automatic gain-control mode with the trapping limits set at 2×10^8 and 8×10^7 for MS and MS/MS collision-activated induced dissociation (CID), respectively, with a maximum injection time set at 200 ms for both scan modes. CID experiments were performed as a "triple-play" by selection of the most intense ion observed during MS with a 3 amu isolation window, performing a Zoom-scan to determine the charge state, and then performing dissociation using arbitrary 55% relative collision energy. For the detection of S-pyridylethylated peptides, in-source-CID was performed by setting the source voltage to 70% resonance

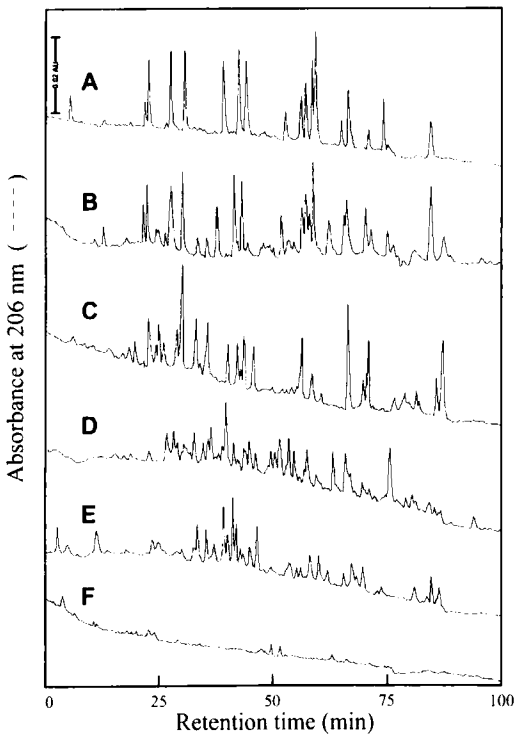


Figure 2. Microcrushed in-gel solution digestion map of streptokinase. Chromatographic conditions: column, Vydac C18, 150 × 2.1 mm ID, flow rate 200 μL/min, detection at 206 nm; buffer A, 0.1% TFA v/v/water; buffer B, 0.1% v/v/60% v/v CH_3CN; gradient, 0–100% B in 60 min. (A) In-solution-derived peptide map of 47 kDa streptokinase (50 pmol). Protease: Glu C; digestion time: 4 h. (B) In-gel digestion (same conditions as in panel A). (C) In-gel-derived peptide map of strepokinase (50 pmol). Protease: trypsin; digestion time, 4 h. (D) In-gel-derived peptide map of strepokinase (20 pmol). Protease: Lys C; digestion time: 4 h. (E) In-gel-derived peptide map of strepokinase (20 pmol). Protease: Asp N; digestion time: 4 h. (F) Control map (blank) derived by Lys C digestion of a nonprotein-containing gel band.

excitation and setting the MS scan to a limited scan range of 104.5–107.5 to detect the characteristic labile 106 Da S-pyridylethyl moiety [7].

Figure 4 shows a typical high-sensitivity ion-trap capillary LC-MS/MS analysis of a PA gel-separated protein, using recombinant interleukin-6 (IL-6) as an example. IL-6 (250 ng) was electrophoresed on 4–20% precast PA gel (Novex, San Diego, CA, USA) and, following visualization by imidazole-zinc reverse staining, subjected to in-gel digestion with trypsin. The resultant peptides were fractionated by capillary RP-HPLC that was directly coupled to an ESI-IT mass spectrometer. The total ion current for MS mode and MS/MS modes obtained from 50% (~ 5 fmol) of the total digest are shown in Figs. 4A and 3B, respectively. A representative CID spectrum of a doubly-charged m/z 1105.79 tryptic peptide (T4, calculated mass 2210.6 Da) is shown in Fig. 4C. The search routine SEQUEST [9] correctly identified peptide T4 from the OWL protein database as a tryptic fragment of IL-6, comprising residues 160–177 (Leu-Gln-Ala-Gln-Asn-Gln-Trp-Leu-Gln-Asp-Met-Thr-Thr-His-Leu-Ile-Leu-Arg; Fig. 4D). By these means an additional four peptides (T1, 2, 3, and T5) were readily ascribed to IL-6 (data not shown). By performing a full MS/MS search by SEQUEST, a total of eight peptides were identified as IL-6 tryptic peptides, where seven peptides were ranked highest and given a score of 10 and one ranked fourth and given a score of 4. A total score of 74 was ascribed by SEQUEST-SUMMARY [10], identifying IL-6 as the major component of the digest analyzed. A salient feature of the total ion current profile (MS mode) shown in Fig. 4A was the lack of PA gel-related artifact peaks normally seen at these levels with Coomassie blue visualization protocols.

Finally, the advantages of this protocol in comparison to the established methods for Coomassie blue-detected proteins are: (i) The PA gel pH is kept relatively constant during all manipulations from electrophoresis to digestion, thus avoiding drastic changes in pH that could initiate protein precipitation. (ii) There are insignificant artifact peaks from the reverse-staining procedure with the sensitivity of protein detection at 5–10 ng/band [3]. Note that we do not recommend this staining method to be used for quantitative measurements due to the limited dynamic range, and attempts to modify the reverse-staining procedure to enable better quantitation and stability of the stain for long-term storage by drying have required extra steps and the introduction of an acid "toning" step with the use of carcinogenic agents [11]. To avoid this, gels should not be dried, but can easily be stored wet. (iii) A significantly shorter working time is necessary. For instance, starting with a precast PA mini-gel, it is possible to accomplish separation, detection, digestion, extraction and analysis in one working day. The shorter digestion times have the additional advantage of reducing nonspecific cleavage and protease autoproteolysis, while increasing the chance of obtaining longer peptides that facilitate separation and analysis. (iv) The procedure is simpler compared to established positive staining methods, where the possibility of protein modification is diminished. It has been shown that during standard Coomassie blue or silver staining procedures, protein modification can still occur, which severely affects the overall sequence coverage and determination of gel-resolved proteins [12, 13].

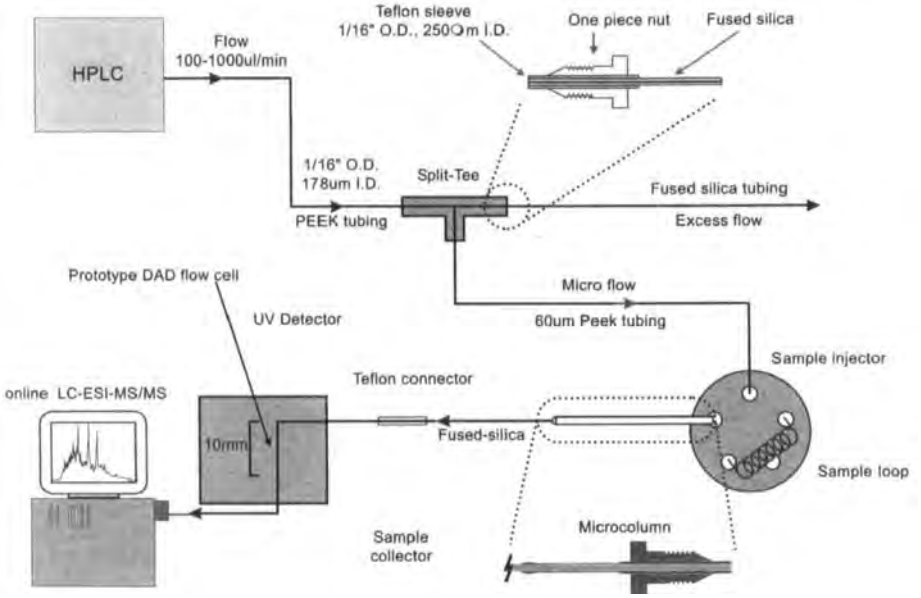


Figure 3. Construction of microcolumn HPLC for LC-MS/MS.

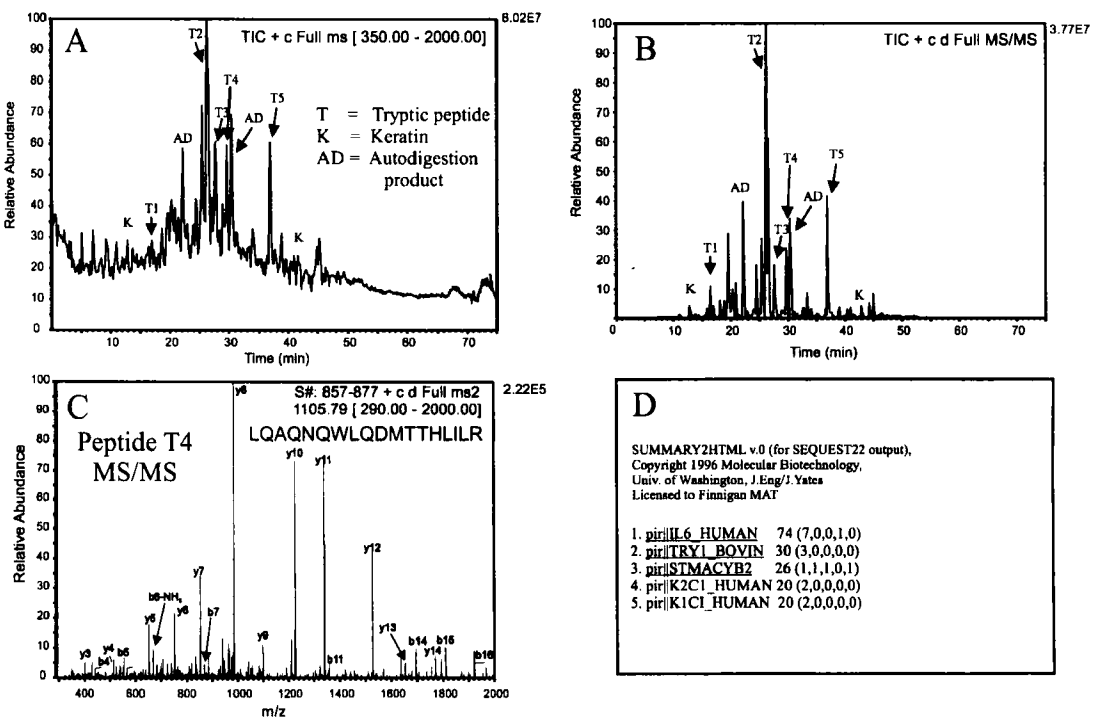


Figure 4. Microcrushed in-gel digestion map of recombinant human interleukin-6. Sample: 0.25 μg recombinant IL-6 loaded onto PA gel. Gel electrophoresis: Novex 4–20% Pre-cast; gel stain: imidazole-zinc reverse stain. Sample injection onto RP-HPLC: 20 μL/40 μL (50% of total sample digest); column : Brownlee RP-300, 50 mm × 0.2 mm ID; solvent A: 0.1% aqueous TFA; solvent B: 0.1% aqueous TFA/60% acetonitrile; flow rate : 1.4 μL/min, sheath liquid : 2-methoxyethanol, 3 μL/min. (A) Total ion current (TIC) spectrum of a tryptic digest of IL-6. (B) MS/MS TIC spectrum from (A). (C) MS/MS spectrum of tryptic peptide T4 from (A). (D) SEQUEST output of ranked identified match results.

(v) This protocol is compatible with rapid identification of proteins by high-sensitivity capillary column RP-HPLC/on-line ion trap MS [8], or direct analysis by MALDI-TOF-MS [14].

Received November 10, 1998

References

[1] Eckerskorn, C., Grimm, R., *Electrophoresis* 1996, *17*, 899–906.

[2] Heukeshoven, J., Dernick, R., *Electrophoresis Forum'91*, Technical University, Munich 1991, pp. 501–506.

[3] Fernández-Patrón, C., Calero, M., Rodríquez, P., Collazo, P., García, J., Musacchio, A., Soriano, F., Estrada, R., Frank, R., Castellanos-Serra, L., Méndez, E., *Anal. Biochem.* 1995, *224*, 203–211.

[4] Castellanos-Serra, L. R., Fernandez-Patrón, C., Hardy, E., Huerta, V., *Electrophoresis* 1996, *17*, 1564–1572.

[5] Castellanos-Serra, L. R., Fernandez-Patrón, C., Hardy, E., Santana, H, Huerta, V., *J. Protein Chem.* 1997, *16*, 415–419.

[6] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[7] Moritz, R. L., Eddes, J. S., Reid, G. E., Simpson, R. J., *Electrophoresis* 1996, *17*, 907–917.

[8] Zugaro, L. M., Reid, G. E., Hong, J., Eddes, J. S., Murphy, A. C., Burgess, A. W., Simpson, R. J., *Electrophoresis* 1998, *19*, 867–876.

[9] Eng, J. K., McCormack, A. L., Yates III, J. R., *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989

[10] Ducret, A., Van Oostveen, I., Eng, J. K., Yates III, J. R., Aebersold, R., *Protein Sci.* 1998, *7*, 706–719.

[11] Ferreras, M., Gavilanes, J. G., Garcia-Segura, J. M., *Anal. Biochem.* 1993, *213*, 206–212.

[12] Haebel, S., Albrecht, T., Sparbier, K., Walden, P., Korner, R., Steup, M., *Electrophoresis* 1998, *19*, 679–686.

[13] Scheler, C., Lamer, S., Pan, Z., Li, X. P., Salnikow, J., Jungblut, P., *Electrophoresis* 1998, *19*, 918–927.

[14] Matsui, N. M., Smith, D. M., Clauser, K. R., Fichmann, J., Andrews, L. E., Sullivan, C. M., Burlingame, A. L., Epstein, L. B., *Electrophoresis* 1997, *18*, 409–417.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 738–742

Jun X. Yan[1]
Jean-Charles Sanchez[2]
Luisa Tonella[2]
Keith L. Williams[1]
Denis F. Hochstrasser[2]

[1]Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW, Australia
[2]Laboratoire Central de Chimie Clinique, Hôpital Cantonal Universitaire, Genve, Switzerland

Studies of quantitative analysis of protein expression in *Saccharomyces cerevisiae*

In the present study amino acid analysis is applied to quantitation of *Saccharomyces cerevisiae* proteome expression. The quantitation levels obtained are compared to data using densitometric analysis of silver or amido black staining and to the theoretical expression level (codon bias) of the identified proteins determined from their amino acid analysis (AAA). The results show that relative volume ratio (%vol) using Melanie II is a better parameter for spot quantitation than relative optical density ratio (%OD), and amino black staining provides good linearity within the range 1–100 pmol protein. However, AAA shows that theoretical expression levels are not well correlated with actual protein expression level, although there is better correlation when isoforms of the expressed protein are identified and included. It is concluded that amino acid analysis provides accurate protein quantitation and has a continuing role in proteome studies in terms of the rapid and inexpensive quantitation of proteins displayed on proteome maps. We do however recognize that in the context of future clinical applications and large-scale proteome discovery projects, quantitation and post-translational modification need to be analyzed by 'proteomatic' (*i.e.*, proteome automatic bioinformatic analysis directly from the gel) techniques.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Protein expression / Quantitation / Amino acid analysis / Proteome / Yeast / Codon bias EL 3372

Identification of proteins displayed on proteome maps is a key element of proteomics and in early proteome studies amino acid analysis (AAA) proved to be a useful means to this end [1–3]. More recently peptide mass fingerprinting using matrix-assisted laser desorption/ionisation – time-of-flight (MALDI-TOF) mass spectrometry has emerged as the premier protein identification tool [4–6]. We demonstrate that AAA still has an important role to play in terms of quantitation of the expressed proteins, which is likely to assume increasing importance. For example, clinical applications of proteomics will involve the identification of protein markers for the diagnosis, treatment and management of cancer and other diseases. In such applications quantitation of the marker protein expression level is critical. It may provide insight into the developmental stage and pathogenesis of the disease. Quantitation of key protein molecules of various biological development stages is also important in gaining an understanding of the development of cellular differentiation. Formerly, analysis of mRNA levels has been used as an indirect means to quantify levels of protein expression. However, studies have shown that there is little correlation between mRNA and protein abundance in well-characterized biological systems [7]. This indicates that protein levels in the cell are influenced considerably by post-translational events. In this work we have used *Saccharomyces cerevesiae* as a model system to investigate quantitative aspects of proteome expression.

The preparation of *S. cerevesiae* (strain X2180-1A) proteins by two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and blotting onto polyvinylidene difluoride (PVDF) were performed as described previously [8, 9]. After blotting and staining with amido black, 54 protein spots (Fig. 1) were selected and quantified by densitometry. The spots were then excised and subjected to Fmoc (9-fluorenylmethoxy carbonyl) AAA to give the amino acid composition and absolute quantitation [1, 10]. Protein spots corresponding to those from the blot were also quantified by densitometry using an identical silver-stained master gel image obtained from the SWISS-2DPAGE database [8]. Proteins analyzed by densitometry were quantified as both relative volume ratios (%vol) where $vol = \Sigma_{x,y \in spot} I(x,y)$, $\%vol = (vol/\Sigma^{n}_{s=1} vol_s) \times 100$, or relative optical density ratio (%OD), where $OD = MAX_{x,y \in spot} I(x,y)$, $\%OD = (OD/\Sigma^{n}_{s=1} OD_s) \times 100$ using Melanie II computing program (Bio-Rad, Richmond, CA USA). From amino acid composition analysis 31 of the 54 proteins were identified using AACompIdent on the ExPASy Molecular Biology Server (http://www.expasy.ch/ch2d/

Correspondence: Dr. Jun X. Yan, National Heart and Lung Institute, Heart Science Centre, Harefield Hospital, Hill End Road, Harefield, UB9 6JH, UK
E-mail: jun.yan@harefield.nthames.nhs.uk
Fax: +44-1895-828-900

Abbreviations: AAA, amino acid analysis; **%OD,** relative optical density; **%vol,** relative volume ratio

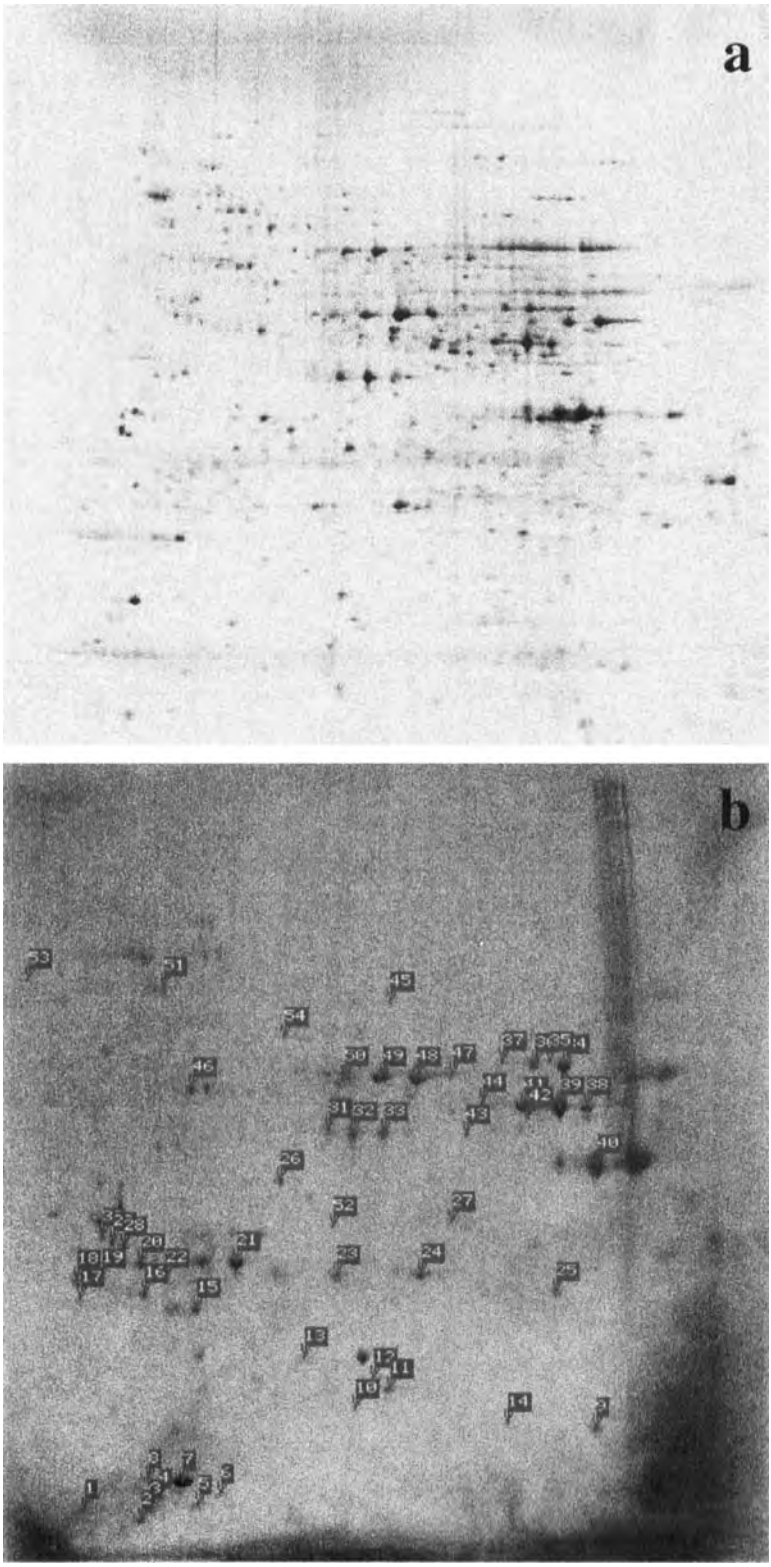


Figure 1. Images of yeast proteins separated by 2-D PAGE. (a) Image of a silver-stained SWISS-2DPAGE map (http://www.expasy.ch/ch2d/publi-yeast.html). (b) Image of the amido black-stained blot with labeling of the 54 spots which were taken for amino acid analysis in this study.

Table 1. Yeast proteins identified in this study using AACompIdent together with their codon bias values obtained from the Yeast Protein Database (Proteome Inc.). Spot numbers correspond to those labeled in Fig. 1b.

Spot Number	Estimated p*I*/M_r	Protein identity	SWISS-PROT identifier	Protein quantity (pmol)	Codon bias
3	4.76/10302	Thioredoxin I	TRX2	22.8	0.657
4	4.8/10791	Heat shock protein of 12 kD	HSP12	25.1	0.800
7	4.84/11797	Heat shock protein of 12 kD	HSP12	126.4	0.800
8	4.73/11839	Heat shock protein of 12 kD	HSP12	33.2	0.800
10	5.71/12401	Profilin	PFY1	12.6	0.618
11	5.86/13606	Copper-zinc superoxide dismutase	SOD1	39.1	0.555
12	5.82/14207	Copper-zinc superoxide dismutase	SOD1	21.2	0.555
16	4.98/22865	Protein with similarity to YCP4P	YBR052C	18.6	0.105
22	5.04/22933	Protein with similarity to YCP4P	YBR052C	12.0	0.105
21	5.33/23956	Heat shock protein 26 kD	HSP26	36.1	0.434
23	5.64/23000	Triosephosphate isomerase	TPI1	30.2	0.900
24	6.01/23117	Triosephosphate isomerase	TPI1	32.0	0.900
31	5.75/36878	Fructose-bisphosphate aldolase II	FBA1	10.8	0.935
32	5.87/36878	Fructose-bisphosphate aldolase II	FBA1	23.9	0.935
33	5.99/37187	Fructose-bisphophate aldolase II	FBA1	29.8	0.935
34	6.61/45911	Enolase 1 (2-phosphoglycerate dehydratase)	ENO1	24.7	0.930
35	6.55/45911	Enolase 1 (2-phosphoglycerate dehydratase)	ENO1	5.9	0.930
36	6.5/46174	Enolase 1 (2-phosphoglycerate dehydratase)	ENO1	13.1	0.930
37	6.37/45911	Enolase 1 (2-phosphoglycerate dehydratase)	ENO1	2.8	0.930
38	6.69/40856	Alcohol dehydrogenase I	ADH1	25.1	0.913
39	6.6/41027	Alcohol dehydrogenase I	ADH1	62.6	0.913
41	6.45/41113	Alcohol dehydrogenase I	ADH1	27.1	0.913
40	6.73/32759	Glyceraldehyde-3-phosphate dehydrogenase 3	TDH3	90.3	0.988
42	6.46/39677	Alcohol dehydrogenase III, mitochondrial	ADH3	3.2	0.397
44	6.4/41113	Alcohol dehydrogenase III, mitochondrial	ADH3	6.6	0.397
45	6.05/59064	GTPase-activating protein	GYP6	3.3	0.147
46	5.01/43786	S-adenosylmethionine synthetase 2	SAM2	9.1	0.641
47	6.27/44795	Enolase 2 (2-phosphoglycerate dehydratase)	ENO2	17.8	0.96
48	6.17/44795	Enolase 2 (2-phosphoglycerate dehydratase)	ENO2	48.3	0.960
49	6/44693	Enolase 2 (2-phosphoglycerate dehydratase)	ENO2	32.0	0.960
50	5.87/44795	Enolase 2 (2-phosphoglycerate dehydratase)	ENO2	27.9	0.960

aacompi.html) [1] and these represented products of 15 separate genes (Table 1).

Gene expression levels are correlated strongly with codon bias [11]; therefore the codon bias of *S. cerevesiae* obtained from the Yeast Proteome Database (Proteome Inc.; http://www.proteome.com/YPDhome.html) has been used as an indicator for theoretical protein expression levels. Clearly, if the codon bias for the identified proteins is compared with the actual amount of protein determined by AAA, the correlation coefficient obtained is very low because the calculation does not include the isoforms of the different proteins. However, by adding together the quantities for identified isoforms of the proteins, a correlation of 0.59 was obtained (see Fig. 2), which is higher than that previously published [7]. However, the generally low correlation between protein amount and theoretical

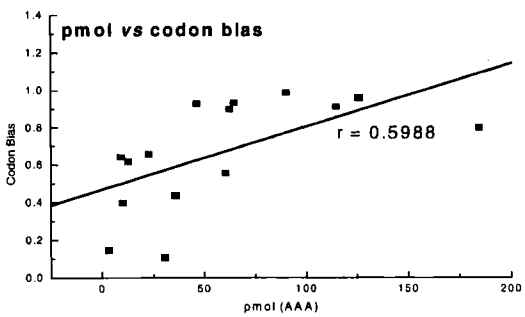


Figure 2. Regression analysis of codon bias *versus* total protein measured by amino acid analysis for the fifteen yeast gene products identified in this study. The amount of each product (pmol) is the sum of all the identified isoforms of the protein as shown in Table 1. The correlation coefficient is 0.59.

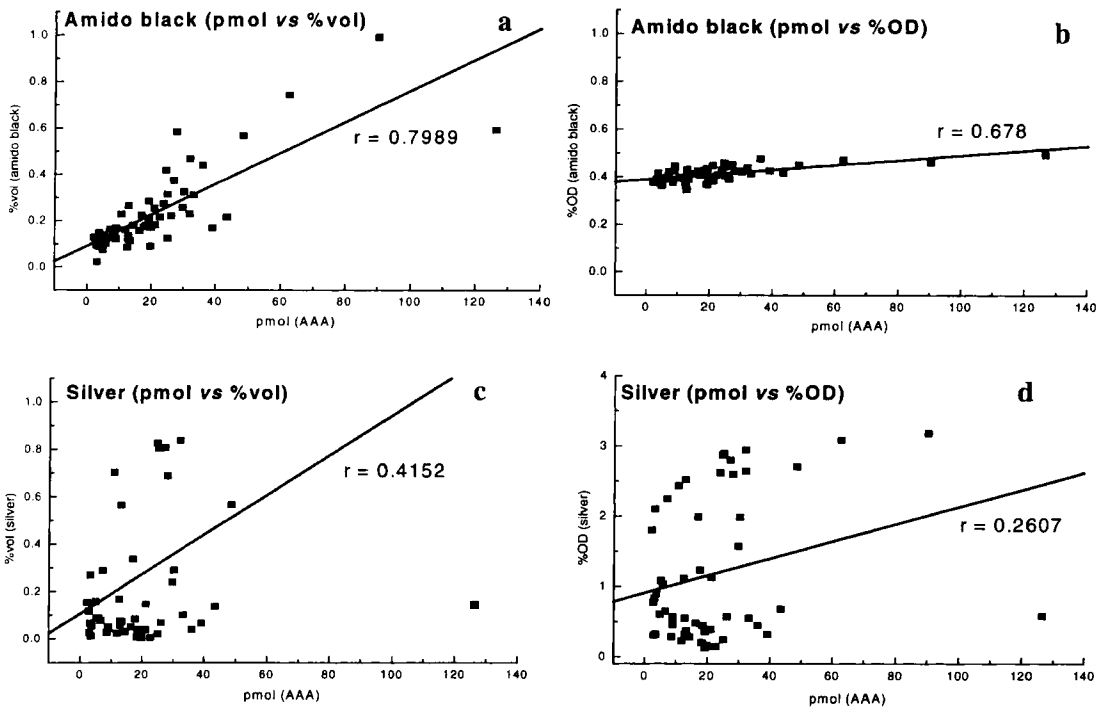


Figure 3. Regression analyses of 54 protein spots to show the correlation between Melanie II computing program quantitation values *vs.* actual protein amount determined by amino acid analysis for two staining methods. Comparing (a) with (b), and (c) with (d), shows that a better correlation is obtained using %vol measurement rather than %OD. Comparing (a) with (c), and (b) with (d), shows that the intensity of amido black staining is more linearly related to protein amount than silver staining within this protein range.

mRNA level indicates that codon bias does not accurately reflect the amount of protein in the cell. The improvement in the correlation obtained by summing the isoforms of the proteins indicates that part of the inaccuracy involved in assessing protein expression from codon bias must be due to post-translational modification of the expressed proteins. In this study only those isoforms which were positively identified were included in the calculation of protein amounts and it is certain that other isoforms were present, but not identified. If these were included these would undoubtedly improve the correlation further. In addition, it must be appreciated that there can be real and apparent protein losses during the 2-D process, due, for example, to poor solubilization, staining and transfer, and this will vary from protein to protein. If these processes were optimal the protein amount to codon bias ratio would also improve.

Quantitation by AAA was also used to determine the correlation between the absolute amount of protein on the gel with the quantities determined by densitometry using the %OD and %vol methods of Melanie II. Comparison of total protein with %OD gave a correlation of only 0.25 for silver-stained and 0.67 for amido black-stained proteins for the 54 protein spots. Comparison of total protein with %vol gave a correlation of 0.41 for silver-stained proteins and 0.79 for amido black-stained proteins. Generally, the %OD or %vol is used to determine the protein amount from densitometry [12], but these data show that %vol is more closely correlated with the actual protein amount than %OD. We presume that the total intensity of the spot represents the real protein amount better than the maximum intensity of the spot in this particular protein range. However, within different ranges, proteins can bind to the dye in a nonlinear fashion, so that the optical density of a protein spot may present a better parameter or the combination of both %OD and %vol may give more accurate quantitative data.

The data also show that amido black staining more accurately corresponds to the amount of protein present than does silver staining. This may reflect that the AAA was performed on protein spots obtained from amido black-stained blot membranes and therefore the quantitation re-

lates to the amount of protein present on the blot rather than in the gel. It is well known that Western blotting can be, and often is, nonquantitative and selective. Using the same dye (*i.e.*, amido black or Coomassie blue) for both gel and blot staining and analyzing the correlation between AAA data and gel or blot spot intensity would give us a clearer evaluation of the effect of the protein transfer on spot quantitation. The data could then be corrected for the transfer error. Amido black staining is less sensitive than silver staining, but gives good linearity in the range 1–100 pmol. These data show that a combination of densitometry using amido black staining and %vol quantitation is a suitable substitute for quantitation by AAA for high abundance proteins. Such an approach has the advantage of speed and simplicity and the proteins are analyzed directly from the gel or blot. For low abundance proteins, however, silver or high sensitivity Coomasssie blue staining remain the major method for quantitation.

It is concluded that AAA provides accurate protein quantitation and has an important continuing role in proteome studies in terms of the rapid and inexpensive quantitation of proteins displayed on proteome maps. We do however recognize that in the context of future clinical applications and large-scale proteome discovery projects, quantitation and post-translational modifications need to be analyzed by 'proteomatic' (*i.e.*, proteome automatic bioinformatic analysis directly from the gel) techniques. Therefore, we propose using AAA initially to extend SWISS-2DPAGE to include abundance data on non-post-translationally modified proteins as a first step towards developing a proteomatic approach to protein quantitation.

JXY was supported by the Australian Proteome Research and Development Grant, and would like to thank Professor Denis Hochstrasser and Dr. Mike Dunn for allowing her to work in their laboratories. JCS and DFH were supported by the Swiss Fund of Scientific Research (grant 32-49314.96).

Received January 7, 1999

References

[1] Yan, J. X., Wilkins, M. R., Ou, K., Gooley, A. A., Williams, K. L., Sanchez, J.-C., Golaz, O., Pasquali, C., Hochstrasser, D. F., *J. Chromatogr. A* 1996, *736*, 291–302.

[2] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *BioTechnol.* 1996, *14*, 61–65.

[3] Yan, J. X., Tonella, L., Sanchez, J.-C., Wilkins, M. R., Packer, N. H., Gooley, A. A., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1997, *18*, 491–497.

[4] Haag, A. M., Taylor, S. N., Johnston, K. H., Cole, R. B., *J. Mass Spectrom.* 1998, *33*, 750–756.

[5] Gharahdaghi, F., Kirchner, M., Fernandez, J., Mische, S. M., *Anal. Biochem.* 1996, *233*, 94–99.

[6] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[7] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[8] Sanchez, J.-C., Golaz, O., Frutiger, S., Schaller, D., Appel, R. D., Bairoch, A., Hughes, G. J., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 556–565.

[9] Sanchez, J.-C., Ravier, F., Pasquali, C., Frutiger, S., Paquet, N., Bjellqvist, B., Hochstrasser, D. F., Hughes, G. J., *Electrophoresis* 1992, *13*, 715–717.

[10] Ou, K., Wilkins, M. R., Yan, J. X., Gooley, A. A., Fung, Y., Sheumack, D., Williams, K. L., *J. Chromatogr. A* 1996, *723*, 219–225.

[11] Bennetzen, J. L., Hall, B. D., *J. Biol. Chem.* 1982, *257*, 3026–3031.

[12] Bini, L., Magi, B., Marzocchi, B., Cellesi, C., Berti, B., Raggiaschi, R., Rossolini, A., Pallini, V., *Electrophoresis* 1996, *17*, 612–616.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 743–748

Joseph A. Loo[1]
Jeffrey Brown[2]
Glenn Critchley[2]
Charles Mitchell[3]
Philip C. Andrews[3]
Rachel R. Ogorzalek Loo[3]

[1]Chemistry Department, Parke-Davis Pharmaceutical Research, Division of Warner-Lambert Company, Ann Arbor, MI, USA
[2]Micromass, Wythenshawe, Manchester, United Kingdom
[3]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA

High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins

The molecular weight measurement of intact *Escherichia coli* proteins separated by isoelectric focusing-immobilized pH gradient (IEF-IPG) gels and analyzed by mass spectrometry is presented. Two methods are discussed: (i) electrospray ionization (ESI) mass spectrometry (MS) of extracted proteins, and (ii) matrix-assisted laser desorption/ionization (MALDI)-MS analysis directly from IEF-IPG gels. Both ESI and MALDI methods yield sub-picomole sensitivity and good mass measurement accuracy. The use of an array detector for ESI-MS was essential to discriminate against contaminating background ions and to selectively detect high mass protein ions. MALDI-MS offers high-throughput analysis of one- and potentially two-dimensional (2-D) gels. The "virtual 2-D" gel method with first-dimensional IEF separation and the second dimension as molecular mass determination by MS, is a particularly promising method for protein analysis due to its ultra high sensitivity and correspondence to classical 2-D gels. Further sensitivity enhancements for the MALDI-MS method are provided by post acceleration detection optimized for high mass time-of-flight analysis.

Keywords: *Escherichia coli* proteins / Mass spectrometry EL 3391

1 Introduction

Applying mass spectrometry (MS) to the identification of proteins separated by one- and two-dimensional polyacrylamide gel electrophoresis (PAGE) has accelerated proteome analysis. Protein expression patterns from organisms, cells, and tissues are currently analyzed by 2-D gel electrophoresis, *in situ* proteolysis from excised protein spots, and protein identification by mass spectrometric analysis. The accurate mass measurement of the proteolysis fragments (peptide mapping), often augmented by further sequence analysis *via* tandem mass spectrometry (MS/MS) experiments, is correlated to protein sequence databases for identification. Such searching routines applying the peptide mapping and/or MS/MS sequencing results offer reliable protein identification [1–3]. However, the sequence coverage from the enzymatic digestion/MS analysis is often low, making it difficult to assess possible sequencing errors. Moreover, post-translational modifications are difficult to assess by this procedure. Mass measurement of the intact protein at the accuracy of SDS gel electrophoresis is frequently too low (~ 10%) to be useful for such analysis, and can be compromised further if proteins migrate anomalously (*e.g.*, errors to 30% for small, unmodified proteins [4]). Modern MS methods with ± 0.1% or better accuracy would clearly be useful.

Several methods for MS measurement of intact proteins separated by PAGE have been presented. Methods using solvent extraction (passive elution) [5–7], while simple, are experimentally challenging when high recovery and high sensitivity are sought from protein loadings below one microgram. Both electrospray ionization (ESI) [8] and matrix-assisted laser desorption/ionization (MALDI) [9] can be used in large protein mass analysis. In most cases, sensitivities for the extraction/MS method are in the 10–100 pmol range (loaded on the gel), although the recent reports by Cohen and Chait [10,11] demonstrated sensitivities to 1 pmol for MALDI-MS analyses. Electroelution is employed widely in solution-phase analysis of gel-isolated proteins [5, 12–15]. Recently Naylor and coworkers [16] devised an apparatus to electroelute and preconcentrate proteins for MALDI-MS to the 1 pmol level. Electroblotting to membranes is commonly used for *N*- and C-terminal sequencing, amino acid analysis, and Western blotting. The direct analysis of electroblotted membranes by MALDI-MS has also shown promise. The best sensitivity in the direct analysis of an electroblotted membrane has been demonstrated by Hillenkamp's laboratory [17–20], especially with an infrared (IR) laser rather than the ultraviolet radiation [21–24] more commonly em-

Correspondence: Dr. Joseph A. Loo, Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, MI 48105, USA
E-mail: joseph.loo@wl.com
Fax: 734-622-2716
Rachel R. Ogorzalek Loo*, Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA
E-mail: ogorzloo@umich.edu
Fax: +734-936-2638

Abbreviation: PATRIC, position and time-resolved ion counting

ployed for MALDI. Little loss in electrophoretic resolution is apparent, despite addition of MALDI matrix solution to the wet membrane [20], and the method has been applied to many 2-D isolated proteins at ± 0.5% mass accuracy [25]. In general, sensitivities for ESI-MS analysis of gel-isolated proteins lag behind those of MALDI-MS; *e.g.*, passive elution protocols yielded successful analysis for *ca.* 0.5 μg of myoglobin extracted from SDS gels [11], while a minimum of 5 μg was required to yield good ESI-MS signals from IPG gels [7].

We have investigated two methods to analyze proteins separated by 1-D gel electrophoresis. Several proteins have been identified from a 25 μg total loading of *E. coli* cell extract, separated by isoelectric focusing (IEF) PAGE in the presence of urea, detergents, and carrier ampholytes, and extracted for ESI-MS measurements. Our goal in these studies was to perform isoelectric focusing by including additives favored for separation of complex mixtures, despite potential deleterious effects on the MS. In several examples, a mass accuracy of ± 0.05% or better has been achieved using ESI-MS for proteins up to 60 kDa. A focal plane array detector on the mass spectrometer allows for sensitive detection in addition to capabilities to discriminate against background ions from extraneous buffers and detergents, allowing us to attempt analyses with minimal clean-up steps. Traditionally, ultrafiltration or reversed-phase HPLC follow extraction for ESI-MS. We wished to explore an approach in which the mass spectrometer discriminated against additives, an approach that could be useful for hydrophobic proteins poorly recovered from ultrafiltration membranes or reversed-phase packing, and when additives not easily HPLC-separated from proteins are employed. These ESI-MS results will be compared to results from the direct analysis of an IEF gel (10 μg total protein loading) by matrix-assisted laser desorption/ionization MS. Our laboratory has previously presented results from the direct desorption/ionization of proteins on 1-D gels in either the native, SDS, or the IEF format [26–28]. We extend these results with our preliminary analysis of *E. coli* cell extracts by MALDI-MS.

2 Materials and methods

2.1 Isoelectric focusing gel electrophoresis

Proteins from cell extracts from *E. coli* (K-12 type strain W3110) were separated in the IEF dimension using 18 cm length pH 4–7 linear immobilized pH gradient (IPG; Pharmacia, Uppsala, Sweden) gels (p*I* spanning pH 4–6.6; B. Bjellqvist, personal communication). Samples were loaded by rehydration in 9 M urea / 0.5% Triton X-100 / 2% Pharmalyte 3–10 / 1% w/v DTT or, alternatively, 7 M urea / 2 M thiourea / 0.6% Pharmalyte 3–10 / 0.8% CHAPS / 1% Triton X-100 / 1% w/v DTT. Gels were prefocused at 500 V for 3 h, ramped to 3500 over 5 h, and focused at 3500 V for 12.5–18 h at 20°C on a Multiphor II electrophoresis unit (Pharmacia).

2.2 ESI-MS

ESI-MS was performed with a double focusing hybrid mass spectrometer (EBqQ geometry, Finnigan MAT 900Q, Bremen, Germany) with a mass-to-charge (*m/z*) range of 10 000 at 5 kV full acceleration potential [29]. A position and time-resolved ion counting (PATRIC) scanning focal plane detector with an 8% *m/z* range of the *m/z* centered on the array detector was used [30, 31]. The ESI-MS experiments utilized a heated metal capillary inlet with a low flow micro-ESI source [32]. Control of the analyte flow rate from 50–200 nL/min was accomplished by pneumatic pressure. No countercurrent gas flow is necessary for desolvation. Gas phase collisions, controlled by adjustment of the voltage difference between the tube lens at the metallized exit of the glass capillary and the first skimmer element (ΔV_{TS}), were also used to augment the desolvation of the ESI-produced droplets and ions. A silver-stained duplicate IPG-IEF gel was run and used as a reference to locate protein bands of interest. Single bands from one unstained gel were cut with a sharp razor in 1–2 mm wide pieces and placed into a 400 μL volume polypropylene centrifuge tube. The gel slices were washed 2–3 times with 50 μL ice-cold distilled water. Afterwards, 25–50 μL of 70:30 v/v acetonitrile:water was added to the gel slice and sonicated for approximately 30 min [7]. Prior to ESI-MS measurements, acetic acid to a final volume of 5% v/v was added to the acetonitrile/water protein extract solution to aid ionization.

2.3 MALDI-MS

MALDI-MS experiments were performed with conventional UV laser irradiation (337 nm nitrogen laser) and delayed ion extraction. The TofSpec-2E time-of-flight (TOF) mass spectrometer (Micromass, Manchester, United Kingdom) was operated in the linear mode and is equipped with a microchannel plate detector with a retractable post acceleration 20 kV conversion dynode optimized for high mass (or *m/z*) ion detection. Additional data was acquired with PerSeptive-Vestec Voyager Elite and PerSeptive DE-STR MALDI-TOF mass spectrometers (Framingham, MA, USA) operating in linear mode. Gels were washed for 10 min in 15 mL of (1:1) acetonitrile:0.2% TFA followed by a 5–10 min matrix soak in saturated sinapinic acid MALDI matrix in acetonitrile:0.2% TFA 1:1. Gels were dried at room temperature. The 3 mm wide dried gel strips (plastic backing still attached) were cut to *ca.* 4 cm lengths and adhered to the milled 4 × 4 cm

sample plate with double-sided tape. Calibration spots of a mixture of bovine insulin and horse heart myoglobin were spotted at regular intervals along the edge of the gel for internal and external mass calibration [28].

3 Results and discussion

3.1 ESI-MS of *E. coli* proteins

One advantage of using ESI-MS for the measurement of large intact proteins is that mass measurement accuracy is generally ± 0.05% or better, provided that small ion adducts such as salts and buffers are not present at significant levels. In addition to reducing the sensitivity of the method by distributing the ion signal over more molecular species, the apparent mass can be shifted to higher values if they are not mass-resolved. Newer ESI sources such as nanoelectrospray [33] reduce total sample consumption, increasing the sensitivity of the analysis. The micro-ESI sources utilize analyte flow rates of 50–200 nL/min compared to the more conventional 1–50 μL/min.

ESI mass spectra from three separate IEF gel slices are shown in Figs. 1–3. These spectra show a range of multiple charge states for each of the proteins. From the measured molecular weights and the estimated p*I*, the identity of the *E. coli* proteins can be determined. Theoretically, identities can be uniquely determined for 92% of the 4400 proteins predicted from the *E. coli* genome sequence by employing data acquired at a mass accuracy of 0.1% and a p*I* accuracy of 0.05 pH units [34]. In practice, this value is lower, due to post-translational modifications, sequence errors, and our inability to predict p*I* this accurately (± 0.2–0.3 pH units may be more realistic [34]). A small change in gel position (p*I*) produces different ESI mass spectra, as illustrated in Figs. 2 and 3. The spectrum at p*I* 6.37 only shows resolved ions from the 34 kDa aspartate

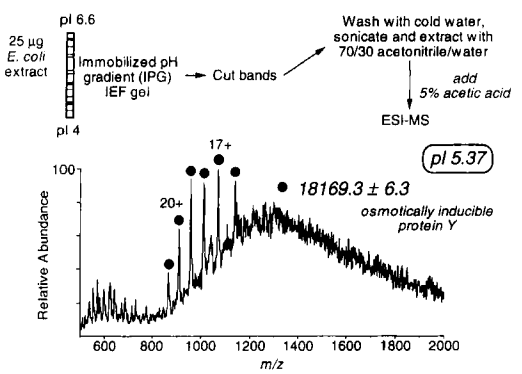


Figure 1. ESI mass spectrum of the band at p*I* 5.37 of the IPG gel of *E. coli* proteins. The protein identified by the measurement is osmotically inducible protein Y (osmY).

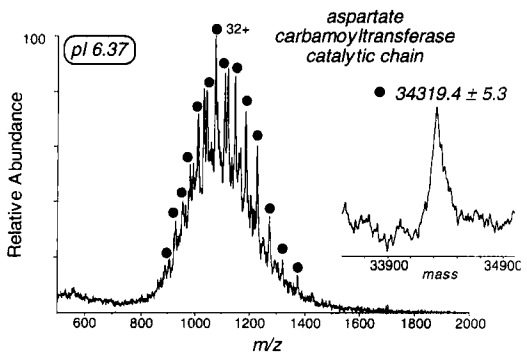


Figure 2. ESI mass spectrum of the band at p*I* 6.37 of the IEF-IPG gel of *E. coli* proteins. The measured molecular weight of the protein indicated its identity as aspartate carbamoyltransferase catalytic chain. The inset shows the deconvoluted spectrum to the mass domain.

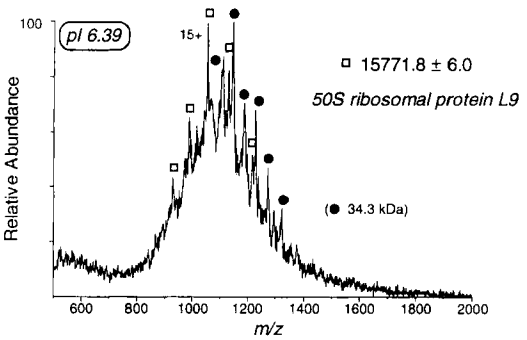


Figure 3. ESI mass spectrum of the band at p*I* 6.39 of the IEF-IPG gel of *E. coli* proteins. The identity of the major species, 50S ribosomal protein L9 (rl9), was determined from the measured molecular weight (peaks in the spectrum labeled with the □ symbol) and the estimated isoelectric point. The peaks labeled with the ● symbol are multiply charged ions from the overlapping protein, aspartate carbamoyltransferase catalytic chain (pyrB), observed at p*I* 6.37 and shown in Fig. 1.

carbamoyltransferase catalytic chain (pyrB), while both the 15.8 kDa 50S ribosomal protein L9 (rl9) and the 34 kDa protein appear at p*I* 6.39, 1.5 mm away.

Assuming 100% extraction efficiency, less than 10 fmol was consumed for each of the measurements shown in Figs. 1–3. Based on known abundances, it is estimated that proteins examined were present at the 0.3–3 pmol level on the gel. The unique low molecular weight (or high *m/z*) discrimination capabilities of the PATRIC array detector [30, 31] were essential for this analysis. Although the washing procedure eliminated many of the possible background components from interfering with the MS

analysis, it did not completely eliminate the background such as Triton-X, urea, and carrier ampholytes. This is clearly evident by the sloping baseline in the spectra. At a normal channel plate voltage of +850 V, only low molecular ions are observed (data not shown). The spectra in Figs. 1–3 were acquired with the array detector voltage less than +725 V, which is optimal for detection of more highly charged ions (with discrimination of the low or singly charged background ions). Although the lower detector voltage improves the sensitivity for large molecule measurements, it makes the analysis of complex mixtures with a large dynamic range in both abundance and molecular weight range more difficult. Each protein has an "optimum" PATRIC detector voltage, depending on the size and charge state [30, 31]. Mass spectra at various detector voltage values would need to be acquired for optimized analysis of a mixture. For example, the ESI mass spectra shown in Figs. 2 and 3 can be compared to the MALDI spectra (Fig. 4) obtained from the IPG gel for the same p*I* region (*vide infra*). The proteins of less than 27 kDa at p*I* 6.37, and masses less than 15 kDa at p*I* 6.39, observed by MALDI-MS were not detected in the ESI mass spectra because the detector voltage was "tuned" for ions from the most abundant species. In some regards, the discrimination "advantage" of the array detector alleviates some of the discrimination "disadvantages" of the ESI process and mixture analysis, as surface effects may selectively suppress ionization. It is unlikely that the array detector will be able to completely offset any serious suppression effects, and sensitivity will certainly be reduced. Obviously, information on relative abundance or quantitation may be difficult to deconvolute from such data, and the dynamic range of protein abundances remains to be evaluated. However, from the viewpoint of developing an ESI-based technique for measuring protein molecular weights from gels, the approach described provides a viable and sensitive method.

Further reduction of contaminating background by more efficient washing of the gel slices would increase the sensitivity of the ESI-MS analysis. However, the current washing protocol is sufficient in conjunction with the high mass sensitivity (and selectivity) of the PATRIC array detector. IPG gels were chosen for these experiments because higher recoveries were anticipated from their low percent acrylamide matrices and because they provide ultra-high resolution separations, especially beneficial if only one-dimensional electrophoresis is employed. Narrow range IPGs further expand capabilities for MS analyses. The size information provided by SDS PAGE is not essential if subsequent MS analysis is pursued. Ultrathin gels have demonstrated compatibility with direct desorption of proteins by MALDI, as reported by our lab [26–28]. As shown here and elsewhere [7], such gels are also applicable for ESI-MS analyses.

3.2 MALDI-MS of *E. coli* proteins on IEF-IPG gels

The direct desorption of proteins from polyacrylamide gels offers the potential for high-throughput molecular weight analysis. Commercial MALDI-TOF-MS systems have capabilities for automated data acquisition from hundreds of samples deposited on a sample stage. Automated analysis of proteins from 1-D gels should also be amenable. The methodology combining 1-D IEF-PAGE as the first dimension and MALDI-MS molecular weight measurement as the second dimension has been termed "virtual 2-D gel analysis" and can potentially be a powerful

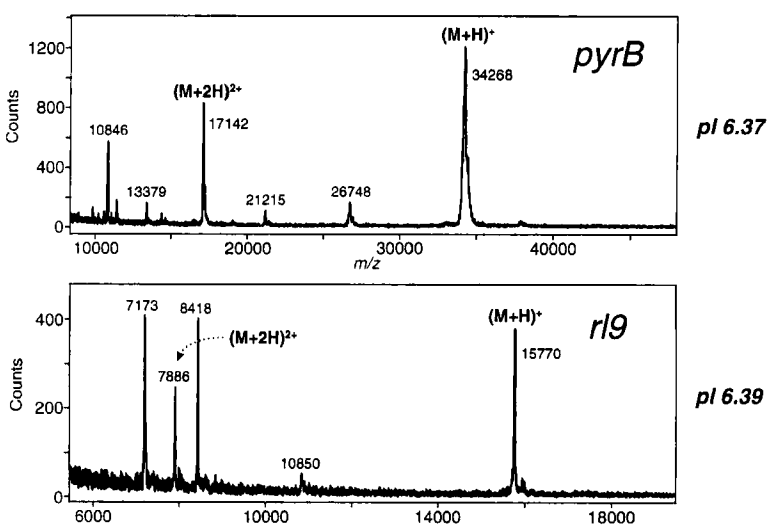


Figure 4. Direct MALDI-MS of the IEF-IPG gel of *E. coli* proteins at (a) p*I* 6.37 and (b) p*I* 6.39.

method for proteome analysis [26, 28]. In contrast to scanning from classical 2-D gels or blots, virtual 2-D gels require data acquisition along only one dimension, reducing analysis time and data storage requirements. Mass calibration is simplified because standards may be spotted at one edge of the 1-D gel, minimizing concerns about ion suppression from calibrants.

Previously, we demonstrated that sub-pmol sensitivities can be obtained by direct UV MALDI-MS of protein mixtures on ultrathin IEF gels [26–28]. The gels used for the original reports were the gels supplied for Pharmacia's Phast system. We have obtained comparable data from Serva's isoelectric focusing gels precast onto netting – a format that avoids the polyester insulator backing of the Phast and DryStrip formats. The current work analyzing proteins from *E. coli* extracts extends this project by examining IPG strips as the separation format. Sinapinic acid MALDI matrix was incorporated into the gel by soaking, as described previously [26–28]. MALDI mass spectra from approximately the same p*I* regions depicted in the ESI-MS data (Fig. 2 and 3), but spanning only a few tens of micrometers (corresponding to the laser spot size) in the p*I* dimension, are shown in Fig. 4. An abundant $(M+H)^+$ ion is observed for each protein component. Mass calibration was achieved by analyzing calibration spots composed of insulin and myoglobin adjacent to or near the band of interest. A molecular weight accuracy of better than ± 0.1% is typically obtained for proteins smaller than 20–30 kDa, and ± 0.2% for larger proteins. Accuracies are generally a factor of 2 better for abundant proteins.

The MALDI mass spectra show more protein species than the corresponding ESI mass spectra (*vide supra*). However, this does not address the question, "How many proteins should the MS-based method observe at a given spot on the 1-D gel?". This issue is a key point for the application of the virtual 2-D gel method for protein profiling. At the present time, work on comparing 2-D PAGE patterns with virtual 2-D gel results to address this issue directly is on-going. Moreover, because the MALDI desorption/ionization process depends on a number of factors, such as the interaction of analyte with matrix molecules and with the surface substrate, among many other variables, it is difficult to predict the level at which the virtual 2-D gel method can provide any information regarding the relative quantitation of proteins on the gel. We do not expect MALDI mass spectrometry to provide a viable substitute for traditional methods of quantitation, but the direct comparison of 2-D PAGE and virtual 2-D gel will provide answers to these questions.

Proteins greater than 60 kDa can be observed from direct PAGE-MALDI-MS. Large proteins such as 66 kDa albumin and 150 kDa monoclonal antibodies have been detected from gels [26–28]. Higher laser powers are needed to desorb proteins from the gels in comparison to desorption from the probe. From the perspective of automation, however, a single laser fluence setting – high enough to desorb large proteins with good sensitivity – is adequate to acquire data down an entire gel. It is not necessary to seek the optimal setting at every position; mass resolution is only sacrificed a bit for small proteins under such acquisition conditions. In general, MALDI sensitivity decreases with increasing molecular weight, but MALDI-MS sensitivities can be increased by utilizing different TOF detector technologies. Mass spectra of two proteins from IEF-IPG gels acquired with a TOF system utilizing post-acceleration and a 20 kV conversion dynode arrangement is shown in Fig. 5. An excellent signal-to-noise ratio is observed for proteins to at least 99 kDa. Although resolution is slightly sacrificed with this detector arrangement, the sensitivity enhancement is important for such large proteins.

4 Concluding remarks

Either ESI or MALDI mass spectrometry can be used to determine the molecular weights of intact proteins separated by IEF-IPG gels with accuracy far superior to

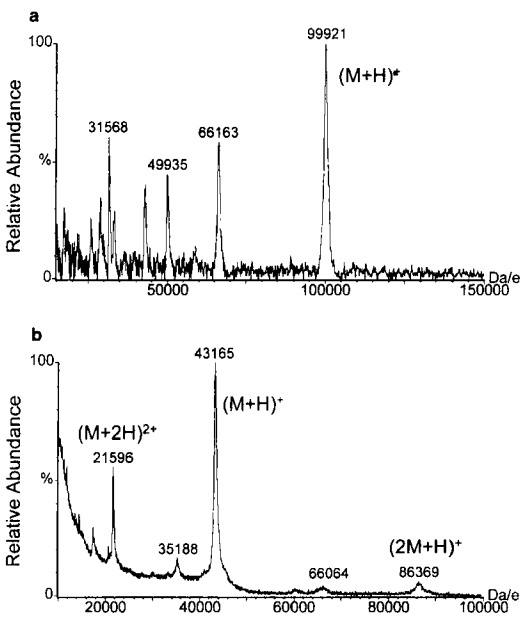


Figure 5. MALDI mass spectra of 2 spots on an IEF-IPG gel of *E. coli* proteins. Ion detection utilizing post-acceleration and a 20 kV conversion dynode increase high mass sensitivity.

SDS-PAGE. ESI-MS can analyze proteins extracted from the gels with relatively high sensitivity. Discrimination of low molecular weight contaminating background ions by array detection is extremely advantageous for such analysis, particularly for low abundance proteins separated in the presence of detergents and carrier ampholytes. The virtual 2-D gel method utilizing MALDI offers the potential for high throughput protein characterization for organisms. Both ionization techniques are suited to the identification of post-translational modifications or sequence variations, because they provide highly accurate and precise intact protein mass measurements. A major advantage of both approaches presented here is that losses associated with sample handling have been minimized, an important consideration when dealing with sample-limited situations. However, relative quantitation by either MS method and the fidelity between classical 2-D PAGE and the virtual 2-D gel method are difficult issues to assess at this time, and addressing these issues is a current major focus of our research efforts.

The authors wish to acknowledge Ruth A. VanBogelen (Parke-Davis) for the E. coli extracts and James Cavalcoli (Parke-Davis) for aid in the identification of E. coli proteins. RROL thanks Parke-Davis Pharmaceutical Research for additional financial support through a research fellowship. PCA and RROL also thank NCI and NHGRI for programmatic support.

Received September 1, 1998

5 References

[1] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[2] Roepstorff, P., *Curr. Opin. Biotechnol.* 1997, *8*, 6–13.

[3] Ducret, A., VanOostveen, I., Eng, J. K., Yates III, J. R., Aebersold, R., *Protein Sci.* 1998, *7*, 706–719.

[4] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997.

[5] Asquith, T. N., Keough, T. W., Takigiku, R., Lacey, M. P., Purdon, M. P., Gauggel, D. L., in: Angeletti, R. H. (Ed.), *Techniques in Protein Chemistry IV*, Academic Press, San Diego 1993, pp. 99–106.

[6] Sheer, D. G., in: Crabb, J. W. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego 1994, Vol. V, pp. 243–248.

[7] Breme, U., Breton, J., Visco, C., Orsini, G., Righetti, P. G., *Electrophoresis* 1995, *16*, 1381–1384.

[8] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., *Science* 1989, *246*, 64–71.

[9] Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K., Hillenkamp, F., *Anal. Chim. Acta* 1990, *241*, 175–185.

[10] Cohen, S. L., Halaas, J. L., Friedman, J. M., Chait, B. T., Bennett, L., Chang, D., Hecht, R., Collins, F., *Nature* 1996, *382*, 589.

[11] Cohen, S. L., Chait, B. T., *Anal. Biochem.* 1997, *247*, 257–267.

[12] le Maire, M., Deschamps, S., Moeller, J. V., Le Caer, J. P., Rossier, J., *Anal. Biochem.* 1993, *214*, 50–57.

[13] Klarskov, K., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 433–440.

[14] Mortz, E., Sareneva, T., Haebel, S., Julkunen, I., Roepstorff, P., *Electrophoresis* 1996, *17*, 925–931.

[15] Schuhmacher, M., Glocker, M. O., Wunderlin, M., Przybylski, M., *Electrophoresis* 1996, *17*, 848–854.

[16] Clarke, N. J., Li, F., Tomlinson, A. J., Naylor, S., *J. Am. Soc. Mass Spectrom.* 1998, *9*, 88–91.

[17] Eckerskorn, C., Strupat, K., Karas, M., Hillenkamp, F., Lottspeich, F., *Electrophoresis* 1992, *13*, 664–665.

[18] Strupat, K., Karas, M., Hillenkamp, F., Eckerskorn, C., Lottspeich, F., *Anal. Chem.* 1994, *66*, 464–470.

[19] Strupat, K., Eckerskorn, C., Karas, M., Hillenkamp, F., in: Burlingame, A. L., Carr, S. A. (Eds.), *Mass Spectrometry in the Biological Sciences*, Humana Press, Totowa, NJ 1996, pp. 203–216.

[20] Eckerskorn, C., Strupat, K., Schleuder, D., Hochstrasser, D. F., Sanchez, J.-C., Lottspeich, F., Hillenkamp, F., *Anal. Chem.* 1997, *69*, 2888–2892.

[21] Vestling, M. M., Fenselau, C., *Mass Spectrom. Rev.* 1995, *14*, 169–178.

[22] Schreiner, M., Strupat, K., Lottspeich, F., Eckerskorn, C., *Electrophoresis* 1996, *17*, 954–961.

[23] Blais, J. C., Nagnan-Le-Meillour, P., Bolbach, G., Tabet, J. C., *Rapid Commun. Mass Spectrom.* 1996, *10*, 1–4.

[24] Ogorzalek Loo, R. R., Mitchell, C., Stevenson, T. I., Loo, J. A., Andrews, P. C., *Int. J. Mass Spectrom. Ion Proc.* 1997, *169/170*, 273–290.

[25] Sutton, C. W., Wheeler, C. H., U, S., Corbett, J. M., Cottrell, J. S., Dunn, M. J., *Electrophoresis* 1997, *18*, 424–431.

[26] Ogorzalek Loo, R. R., Mitchell, C., Stevenson, T., Loo, J. A., Andrews, P. C., in: Marshak, D. R. (Ed.), *Techniques in Protein Chemistry VII*, Academic Press, San Diego, CA 1996, pp. 305–313.

[27] Ogorzalek Loo, R. R., Stevenson, T. I., Mitchell, C., Loo, J. A., Andrews, P. C., *Anal. Chem.* 1996, *68*, 1910–1917.

[28] Ogorzalek Loo, R. R., Mitchell, C., Stevenson, T. I., Martin, S. A., Hines, W., Juhasz, P., Patterson, D., Peltier, J., Loo, J. A., Andrews, P. C., *Electrophoresis* 1997, *18*, 382–390.

[29] Loo, J. A., Ogorzalek Loo, R. R., Andrews, P. C., *Org. Mass Spectrom.* 1993, *28*, 1640–1649.

[30] Loo, J. A., Pesch, R., *Anal. Chem.* 1994, *66*, 3659–3663.

[31] Loo, J. A., Ogorzalek Loo, R. R., *J. Am. Soc. Mass Spectrom.* 1995, *6*, 1098–1104.

[32] Sannes-Lowery, K. A., Hu, P., Mack, D. P., Mei, H.-Y., Loo, J. A., *Anal. Chem.* 1997, *69*, 5130–5135.

[33] Wilm, M., Mann, M., *Anal. Chem.* 1996, *68*, 1–8.

[34] Cavalcoli, J. D., VanBogelen, R. A., Andrews, P. C., Moldover, B., *Electrophoresis* 1997, *18*, 2703–2708.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 749–754

Jun X. Yan[1]
Jean-Charles Sanchez[2]
Pierre-Alain Binz[2,3]
Keith L. Williams[1]
Denis F. Hochstrasser[2]

[1]Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW, Australia
[2]Laboratoire Central de Chimie, Clinique, Hôpital Cantonal Universitaire, Genve, Switzerland
[3]Swiss Institute of Bioinformatics, Centre Medical Universitaire, Universite de Genve, Switzerland

Method for identification and quantitative analysis of protein lysine methylation using matrix-assisted laser desorption/ionization – time-of-flight mass spectrometry and amino acid analysis

Protein methylation is a post-translational modification that might have important functional roles in cell regulation. We present a new technique with sufficient sensitivity (sub-pmol level) for analysis of methylation of proteins in abundances typically found on proteome maps produced by two-dimensional (2-D) gel electrophoresis. The method involves the identification and quantitation of lysine (Lys) methylation using Fmoc (9-fluorenylmethyl chloroformate)-based amino acid analysis (AAA). Tri- and monomethyl-Lys were baseline-separated from other amino acids using a modified buffer system. Trimethyl-Lys was quantitatively recovered after acid hydrolysis and AAA of two known methylated proteins – yeast cytochome *c* and human calmodulin. The methylated peptides from tryptic digestion of those two proteins were identified by high sensitivity matrix-assisted laser desorption/ionization – time-of-flight (MALDI-TOF) mass spectrometry (MS). An automated mass-screening approach is proposed for the study of various post-translational modifications to understand the distribution of those protein isoforms separated by two-dimensional polyacrylamide gel electrophoresis. It is concluded that the combination of AAA and MALDI-TOF-MS provides a high sensitivity quantitative tool for the analysis of protein post-translational methylation in the context of proteome studies.

Keywords: Protein methylation / Lysine / Amino acid analysis / Matrix-assisted laser desorption ionization – time-of-flight mass spectrometry / Post-translational modification / Cytochrome *c* / Calmodulin
EL 3373

A key element of proteomics is the capability to identify protein post-translational modifications that can play an important role in biological and clinical functions [1, 2]. Post-translational modifications are not reliably predicted from the DNA sequence and new analytical methods are needed to identify such changes in protein arrays produced by 2-D gel electrophoresis [3]. Protein methylation is a post-translational modification that has been recognized to be important in many physiological roles. In bacteria, methylated proteins are involved in sensory adaptation to chemical stimulation [4, 5]. In eukaryotic cells, protein methylation is associated with the cellular stress response [6, 7], aging, and the repair of proteins [8, 9]; and recently, reversible protein methylation has been seen to have a potential role in signal transduction [10].

Correspondence: Dr. Jun X. Yan, National Heart and Lung Institute, Heart Science Centre, Harefield Hospital, Hill End Road, Harefield, UB9 6JH, UK
E-mail: jun.yan@harefield.nthames.nhs.uk
Fax: +44-1895-828-900

Abbreviations: AAA, amino acid analysis; **Fmoc,** 9-fluorenylmethoxycarbonyl

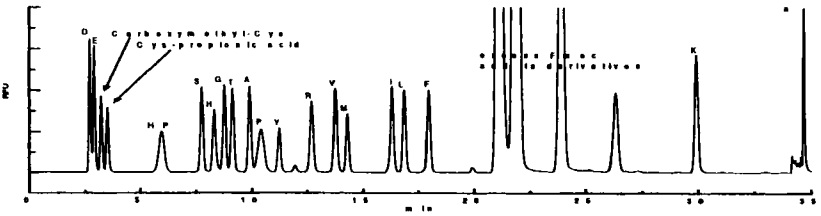
The diversity of protein methylation substrates and physiological roles of this modification suggest that it may have important roles in pathophysiological processes. The role of protein methylation is not yet determined and its function deserves further investigation.

Table 1. Modified gradient for separation of mono- and trimethylated lysine with other standard amino acids. Note the modification is the addition of 10% tetrahydrofuran in buffer A. The gradient program is essentially identical to that previously described [16].

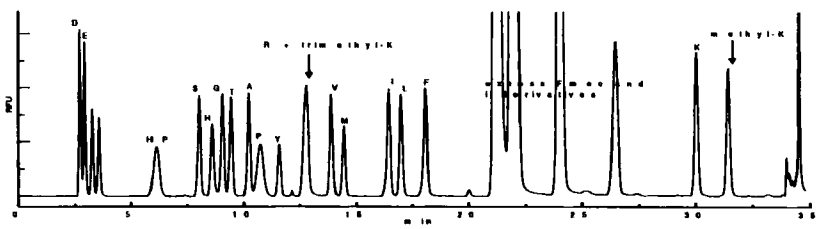
Minute	Flow rate (mL/min)	%A[a)]	%B[b)]	%C[c)]
0	1	17	68	15
1	1	17	68	15
32	1	10.8	43.2	46
32.05	1	0	0	100
34	1	0	0	100
34.05	1	17	68	15
35	1	17	68	15

a) %A, 30 mM ammonium phosphate, 10 % tetrahydrofuran, pH 6.5, in B
b) %B, 15 % v/v methanol in water
c) %C, 90 % v/v acetonitrile in water

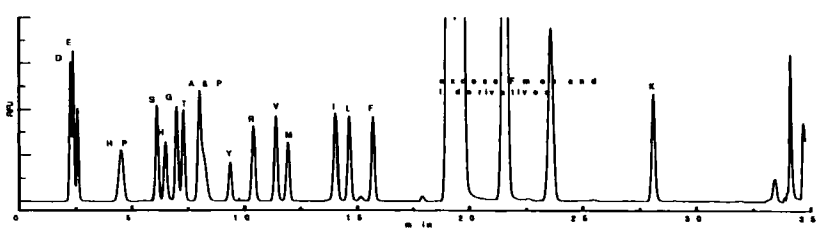
a. AA standard only



b. AA standard + trimethyl and monomethyl-Lys



c. AA standard only



d. AA standard + trimethyl and monomethyl-Lys

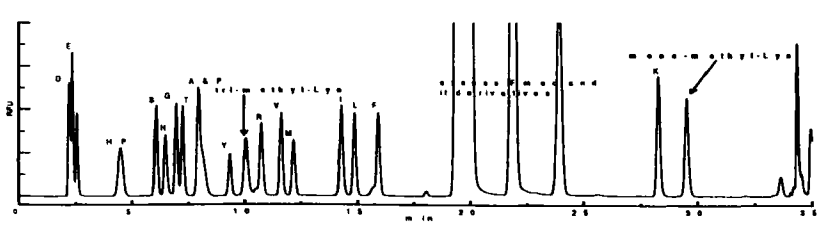


Figure 1. Chromatograms of separation from Fmoc-AA standards. Using gradient described previously [16], without 10% tetrahydrofuran, (a) AAs are well resolved (b), but arginine and triemethyl-Lys coeluted. Use of the modified gradient (addition of 10% tetrahydrofuran) allowed (d) baseline resolution of triemethyl-Lys from all other components although (c, d) proline and alanine coeluted. Monomethyl-Lys is well resolved from other AAs in both solvent systems.

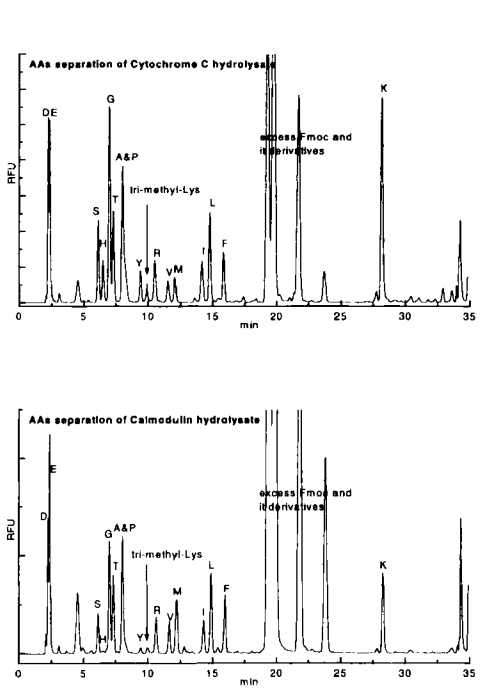


Figure 2. Chromatograms of separation of Fmoc-AA from protein hydrolysates. (a) PVDF-bound yeast cytochrome *c* stained with amido black, 12 pmol recovered after hydrolysis, 1.5% trimethyl-Lys recovered (1.4% expected) by Fmoc-AAA. (b) PVDF-bound human calmodulin stained with amido black, 6 pmol recovered after hydrolysis, 1.5% trimethyl-Lys recovered (1.2% expected) by Fmoc-AAA.

Protein methylation generally occurs on the carboxyl group or the side chain nitrogens of the amino acids lysine (Lys), arginine, or histidine, with ε-*N*-trimethyl-lysine being the predominant methyl lysine species in all cases [10]. Existing analytical methods involving radiolabelling in conjunction with methyltransferases [11], or chromatographic analysis of methylated amino acids [12–15], have a limited role and the latter does not possess sufficient sensitivity to have practical applications in proteome studies. Little has been done using mass spectrometry (MS) analysis to address this problem, although it is well known that addition of methyl groups on protein peptides is chemically stable. In this study, we developed a new identification and quantitation method for analysis of Lys mono- and trimethylation on proteins with sufficient sensitivity (sub-pmol level) to be applied to typical proteome maps produced by 2-D gel electrophoresis.

The quantitative analysis was achieved by Fmoc (9-fluorenylmethoxycarbonyl) amino acid analysis (AAA). The AAA was essentially identical to the method described previously [16]; however, with the use of 10% tetrahydrofuran in gradient buffer A (Table 1), trimethyl- and monomethyl-Lys (Sigma, St. Louis, MO, USA) were baseline-separated from other standard amino acids (Fig. 1). Two known trimethylated proteins, yeast cytochome *c* (SWISS-PROT AC: cycl_yeast, P00044; Sigma) and human calmodulin (SWISS-PROT AC: calm_human, P02593; Sigma) were taken for 1-D SDS-PAGE separation and electroblotted onto polyvinylidene difluoride membrane. Trimethyl-Lys was recovered from both amido black-stained protein bands after acid hydrolysis with quantitative recovery for low pmol amounts (Fig. 2). Coomassie blue-stained gel pieces of both proteins were also taken for tryptic digestion and high sensitivity MALDI-TOF-MS analysis [17, 18]. Tryptic peptides were well recovered with around 67% sequence coverage achieved for both proteins, and the peptide profiles are presented in Fig. 3, highlighting those peptides with the modifications predicted using ExPASy FindMod tool (http://www.expasy.ch/sprot/findmod.html). The exact masses and positions of peptides recovered from the two protein digests are shown in Fig. 4. The peptide 61–78 in the cycl_yeast and the peptide 107–126 in calm_human are known from the SWISS-PROT database to be trimethylated at Lys 77 and Lys 115, respectively. Our results confirm both trimethylations. In addition, two masses (2223.24 and 2238.66) suggest the presence of a dimethylated form of Lys 77 in the cycl_yeast (matching data not shown here), suggesting a heterogeneity of the sample as the methyl groups would not be degraded under the conditions of MALDI-MS. Note that the nonmethylated forms of the peptides are not observed, and that the peptides are not cleaved by trypsin at the modified lysines.

It is concluded that the combination of AAA combined with MALDI-TOF-MS provides a high sensitivity quantitative tool for protein methylation post-translational modification analysis in the context of proteome studies. Chemical stability of trimethyl-Lys enables AAA and MS analysis without additional treatments towards acid hydrolysis and enzymatic digestion. FindMod is a prediction tool and not an identification tool, so that potential methylation identification by peptide mass data can be followed by confirmation with AAA data. With the high usage of MALDI-TOF-MS as a primary protein identification tool [19–21], the proteomic strategy is to use peptide mass profile data of 2-D PAGE-separated proteins from MALDI-TOF-MS, automatically put in through PeptIdent tool (http://www.expasy.ch/sprot/peptident.html). Once the protein is identified, the mass profile data can be automatically put in to FindMod tool (http:///www.expasy.ch/sprot/findmod.html) for identification of any post-translational modifications. The search protocol from peptide mass,

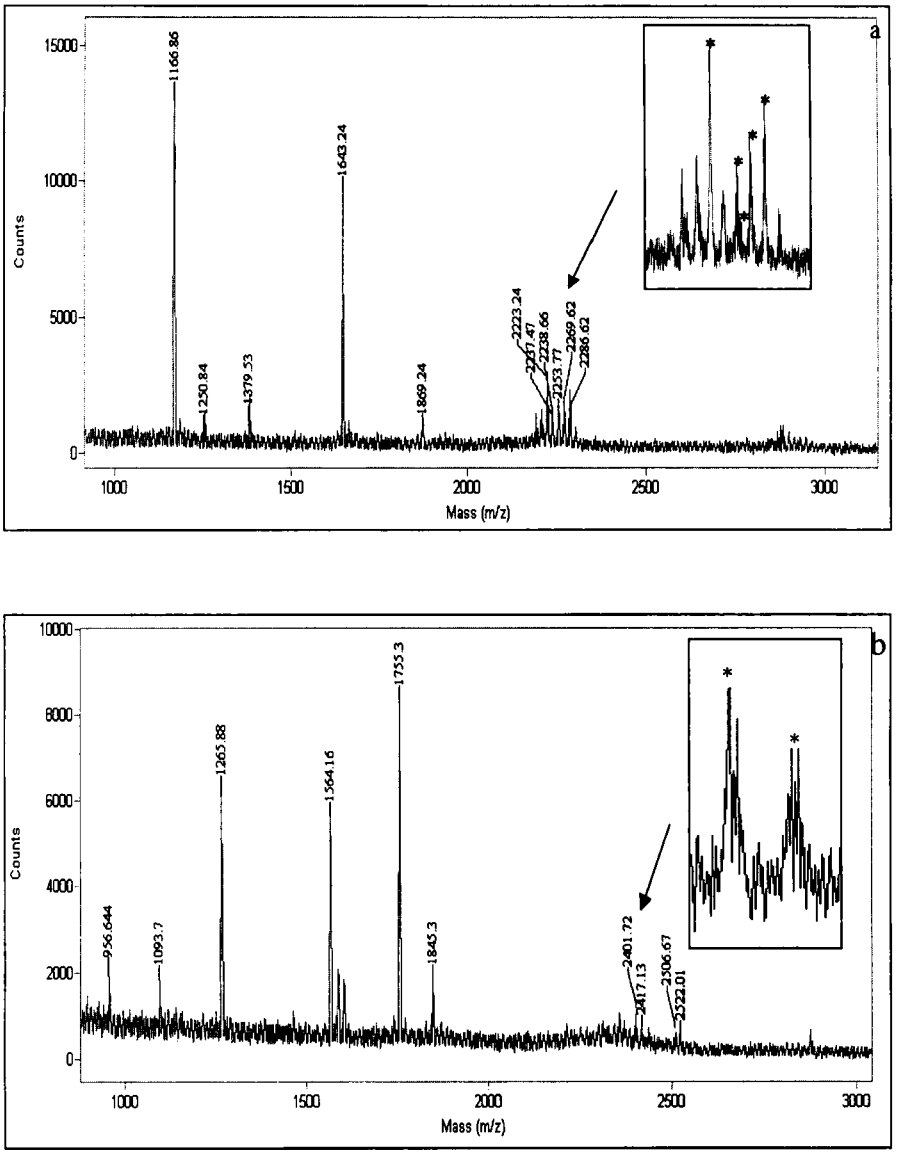


Figure 3. Mass spectra from tryptic digests of (a) cytochrome *c* and (b) calmodulin with the spectra of the modified peptides expanded (details of peptide identifications are shown in Fig. 4).

through PeptIdent to FindMod will offer great assistance in the updating of the SWISS-PROT database.

JXY was supported by the Australian Proteome Research and Development Grant, and would like to thank Professor Denis Hochstrasser and Dr. Mike Dunn for allowing her to work in their laboratories. PAB was supported by the Helmut Horten Foundation. JCS and DFH were supported by the Swiss Fund of Scientific Research (grant 32-49314.96).

Received January 12, 1999

References

[1] Yan, J. X., Packer, N. H., Gooley, A. A., Williams, K. L., *J. Chromatogr. A* 1998, *808*, 23–41.

a. Cytochrome C yeast

Matching peptides:

User mass	DB mass	Δmass (Dalton)	#MC	peptide	position	known modifications
1166.86	1166.6429	-0.217	0	VGPNLHGIFGR	33-43	
1250.92	1250.6739	-0.246	1	DRNDLITYLK	95-104	
1379.05	1379.7463	0.6963	3	MAFGGLKKEKDR	85-96	
1379.05	1378.7689	-0.281	2	DRNDLITYLKK	95-105	
1643.24	1642.8924	-0.3475	1	GGPHKVGPNLHGIFGR	28-43	
1869.41	1868.8773	-0.5326	1	HSGQAEGYSYTDANIKK	44-60	
2237.48	2237.0908	-0.3891	1	KNVLWDENNMSEYLTNPK	60-77	(1xTRIMETH)
2237.48	2237.0908	-0.3891	1	NVLWDENNMSEYLTNPKK	61-78	(1xTRIMETH)
2253.69	2253.0857	-0.6042	1	KNVLWDENNMSEYLTNPK	60-77	(1xMSO, 1xTRIMETH)
2253.69	2253.0857	-0.6042	1	NVLWDENNMSEYLTNPKK	61-78	(1xMSO, 1xTRIMETH)

b. Calmodulin

Matching peptides:

User mass	DB mass	Δmass (Dalton)	#MC	peptide	position	known modifications
956.644	956.4723	-0.1716	0	EAFSLFDK	14-21	
1093.7	1093.4644	-0.2355	0	DTDSEEEIR	78-86	
1265.88	1265.612	-0.2679	0	DGNGYISAAELR	95-106	
1564.16	1563.7537	-0.4062	0	ADQLTEEQIAEFK	1-13	(ACET: 1)
1755.3	1754.8707	-0.4292	1	VFDKDGNGYISAAELR	91-106	
1845.3	1844.8912	-0.4087	1	EAFSLFDKDGDGTITTK	14-30	
2401.73	2401.1739	-0.556	1	HVMTNLGEKLTDEEVDEMIR	107-126	(1xTRIMETH)
2417.13	2417.1688	0.0388	1	HVMTNLGEKLTDEEVDEMIR	107-126	(1xMSO, 1xTRIMETH)
2506.76	2506.0748	-0.6851	0	EADIDGDGQVNYEEFVQMMTAK	127-148	(1xMSO)
2522.01	2522.0697	0.0597	0	EADIDGDGQVNYEEFVQMMTAK	127-148	(2xMSO)

Figure 4. Peptide mass matching results from tryptic digests of (a) cytochrome and (b) calmodulin using FindMod tool on the Expasy server (http://www.expasy.ch/sprot/findmod.html). Abbreviations: ACET, acetylation; MSO, methionine sulfoxide; TRIMETH, trimethylation.

[2] Packer, N. H., Harrison, M. J., *Electrophoresis* 1998, *19*, 1872–1882.

[3] Wilkins, M. R., Sanchez, J.-C., Williams, K. L., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 830–838.

[4] Springer, M. S., Goy, M. F., Adler, J., *Nature* 1979, *280*, 279–284.

[5] Stock, A., Schaeffer, E., Koshland, Jr., D. E., Stock, J., *J. Biol. Chem.* 1987, *262*, 8011–8014.

[6] Wang, C., Lin, J. M., Lazarides, E., *Arch. Biochem. Biophys.* 1992, *297*, 169–175.

[7] Desrosiers, R., Tanguay, R. M., *J. Biol. Chem.* 1988, *263*, 4686–4692.

[8] Johnson, B. A., Langmack, E. L., Aswad, D. W., *J. Biol. Chem.* 1987, *262*, 12283–12287.

[9] Najbauer, J., Orpiszewski, J., Aswad, D. W., *Biochemistry* 1996, *35*, 5183–5190.

[10] Aletta, J. M., Cimato, T. R., Ettinger, M. J., *Trends Biochem. Sci.* 1998, *23*, 89–91.

[11] Clarke, S., Vogel, J. P., Deschenes, R. J., Stock, J., *Proc. Natl. Acad. Sci. USA* 1988, *85*, 4643–4647.

[12] Park, K. S., Lee, H. W., Hong, S. Y., Shin, S., Kim, S., Paik, W. K., *J. Chromatogr.* 1988, *440*, 225–230.

[13] Kohse, K. P., Graser, T. A., Godel, H. G., Rossle, C., Franz, H. E., Furst, P., *J. Chromatogr.* 1985, *344*, 319–324.

[14] Minkler, P. E., Erdos, E. A., Ingalls, S. T., Griffin, R. L., Hoppel, C. L., *J. Chromatogr.* 1986, *380*, 285–299.

[15] Garcia, R., Hiatt, W. R., Jasinski, S., Sypherd, P. S., *Appl. Environ. Microbiol.* 1986, *51*, 1355–1357.

[16] Ou, K., Wilkins, M. R., Yan, J. X., Gooley, A. A., Fung, Y., Sheumack, D., Williams, K. L., *J. Chromatogr. A* 1996, *723*, 219–225.

[17] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[18] Weekes, J., Wheeler, C. H., Yan, J. X., Weil, J., Eschenhagen, T., Scholtysik, G., Dunn, M. J., *Electrophoresis* 1999, *20*, 898–906.

[19] Haag, A. M., Taylor, S. N., Johnston, K. H., Cole, R. B., *J. Mass Spectrom.* 1998, *33*, 750–756.

[20] Gharahdaghi, F., Kirchner, M., Fernandez, J., Mische, S. M., *Anal. Biochem.* 1996, *233*, 94–99.

[21] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 755–765

Klaus-Peter Pleißner[1]
Frank Hoffmann[2]
Klaus Kriegel[2]
Carola Wenk[2]
Susan Wegner[1]
Anders Sahlström[1]
Helmut Oswald[2]
Helmut Alt[2]
Eckart Fleck[1]

[1]Department of Internal Medicine / Cardiology, Charité, Campus Virchow-Clinic, Humboldt University and German Heart Institute Berlin, Germany
[2]Department of Mathematics and Computer Science, Free University Berlin, Germany

New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases

Protein spot identification in two-dimensional electrophoresis gels can be supported by the comparison of gel images accessible in different World Wide Web two-dimensional electrophoresis (2-DE) gel protein databases. The comparison may be performed either by visual cross-matching between gel images or by automatic recognition of similar protein spot patterns. A prerequisite for the automatic point pattern matching approach is the detection of protein spots yielding the *x(s),y(s)* coordinates and integrated spot intensities *i(s)*. For this purpose an algorithm is developed based on a combination of hierarchical watershed transformation and feature extraction methods. This approach reduces the strong over-segmentation of spot regions normally produced by watershed transformation. Measures for the ellipticity and curvature are determined as features of spot regions. The resulting spot lists containing *x(s),y(s),i(s)*-triplets are calculated for a source as well as for a target gel image accessible in 2-DE gel protein databases. After spot detection a matching procedure is applied. Both the matching of a local pattern *vs.* a full 2-DE gel image and the global matching between full images are discussed. Preset slope and length tolerances of pattern edges serve as matching criteria. The local matching algorithm relies on a data structure derived from the incremental Delaunay triangulation of a point set and a two-step hashing technique. For the incremental construction of triangles the spot intensities are considered in decreasing order. The algorithm needs neither landmarks nor an *a priori* image alignment. A graphical user interface for spot detection and gel matching is written in the Java programming language for the Internet. The software package called CAROL (http://gelmatching.inf.fu-berlin.de) is realized in a client-server architecture.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Database comparison / Gel matching / World Wide Web / Internet / Java / Point pattern matching EL 3338

1 Introduction

Two-dimensional gel electrophoresis (2-DE) is currently one of the important methods in proteome research [1]. Together with protein identification methods such as sequencing, mass spectrometry, or amino acid composition, much information on proteins is produced and stored in World Wide Web 2-DE gel protein databases, which are available to all scientists. The software principles for the construction of such 2-DE databases are well-developed and described [2–6]. The exponential increase of stored data gives rise to many new tasks. One of these consists of comparing gel images included in 2-DE databases *via* the Internet to support or to replace the expensive identification of proteins. Another interesting and possible application of gel comparison is related to the fact that with some diseases there are associated typical deviations of certain protein spots compared to standard spot size/intensity [7, 8]. Detecting such deviations is of great importance, for example, in view of a possible drug design. By creating master gels and using the SWISS-PROT accession number of a protein, spots can be localized and identified in different 2-DE databases fulfilling the concept of federated 2-DE databases [2]. The visual inter-laboratory comparison of two gel images maintained in a 2-DE meta-database was pioneered by Lemkin [9, 10]. To overcome substantial difficulties caused by different resolution, gel size, gel distortion, and noise, several image processing techniques are applied. Image processing includes, for instance, noise reduction by smoothing, image warping transformations, contrast enhancement, and flickering. Moreover, these techniques can be currently realized on the Internet using, for instance, the programming language Java and the client-server paradigm (see, for example, flicker server at http://www-lmmb.ncifcrf.gov/flicker). This kind of cross-matching aims at searching visually for qualitative differences between two gel images at pixel level [11, 12]. Eventually, the user must decide whether spots correspond to each other or not.

Correspondence: Dr. Klaus-Peter Pleißner, Deutsches Herzzentrum Berlin, Projektgruppe Digitale Bildverareitung, Iranische Str. 2, 13347 Berlin, Germany,
E-mail: pleiss@dhzb.de
Fax: +49-30-46504-108

We address the problem of comparing/matching 2-DE gel images stored in WWW databases by recognizing similar protein spot patterns automatically. Instead of image pixels we use features derived by a new spot detection algorithm and apply a computational geometry approach for point pattern matching. Furthermore, the spot detection and matching procedure is performed directly on the Internet by means of a user interface invoked in any Java-capable Internet browser.

2 Material and methods

2.1 Protein spot detection

A prerequisite for automatic point pattern matching between a source and target gel image located in different WWW gel protein databases is the detection of protein spots. Several algorithms for spot detection [11, 13–17] exist and can mainly be divided into Gaussian fitting and Laplacian of Gaussian (LOG) spot detectors. Some of these spot detection algorithms are realized in commercial gel analysis systems as Kepler, Gellab-II, PDquest, Melanie, and Phoretix. For our task we developed a new algorithm based on a hierarchical watershed transformation (WST) combined with feature extraction methods. The spot detection for WWW-accessible gel images using the WST can be described shortly by the following steps (Fig. 1): (i) downloading of a gel image from the Internet, (ii) calculation of the gradient image, (iii) performing the WST, (iv) localization of spot regions, (v) merging of partial spot regions, and (vi) calculation of a spot list.

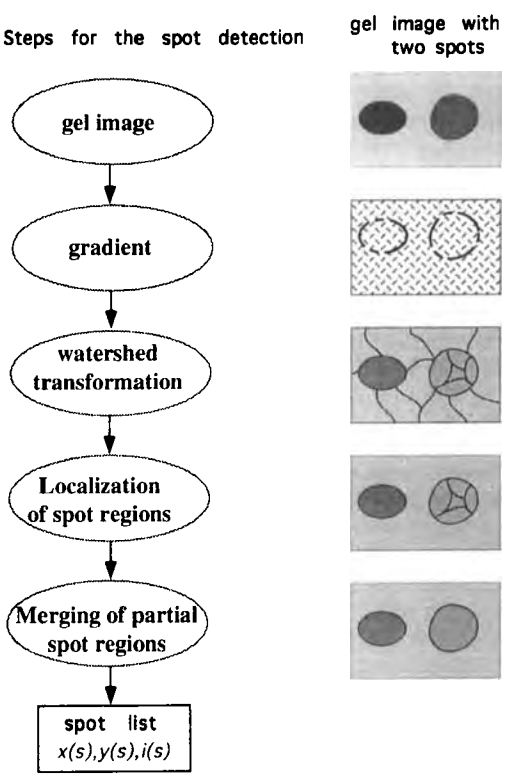


Figure 1. Steps for spot detection.

2.1.1 Watershed transformation

The WST (Fig. 2) is an image segmentation approach which can be described as an immersion process in "gray value reliefs" [18–20]. Drilling holes in each regional minimum of the relief, the surface is immersed into a lake, ensuring a constant water level in all the valleys of the gray-value mountains. The valleys correspond to the regions whereas the watersheds define the optimal contours of objects. The regions are obtained using the WST on the gradient image (first derivative). The assumption is that valleys of the gradient image correspond to the requested regions whereas the ridges -watersheds- define the optimal contour of a region. Each region is then described by a feature, achieving a so-called mosaic image. The WST results in well-located closed contours, but also in strong over-segmentation caused by the image noise amplified by the calculation of the gradient image. Additionally, Gaussian smoothing of the original image can be performed as a prefilter.

2.1.2 Localization of spot regions

Due to the over-segmentation in the mosaic image not every spot is recognized as a single region. In many cases one spot consists of several subregions ("partial spot regions"). Therefore two kinds of regions must be distinguished. First, one region already covers the whole spot, and second, the region only describes a part of the spot, *i.e.*, several regions cover one spot (Fig. 3). In the latter case all regions of a spot must be merged to obtain a

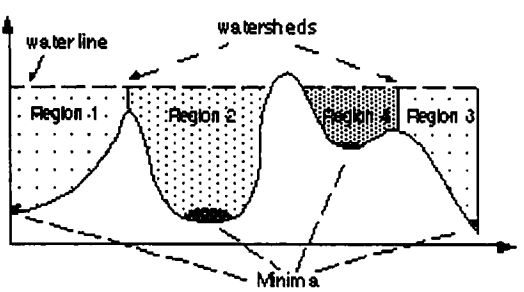


Figure 2. Principle of the watershed transformation as an immersion process.

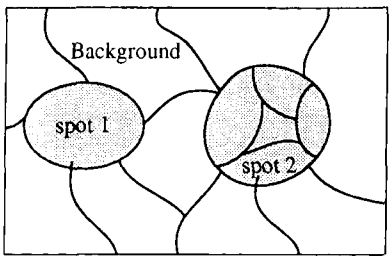


Figure 3. Mosaic image with two kinds of spots to be localized. Spot 1 is already separated by one region. Spot 2 consists of several regions (partial spot region).

complete single spot. Therefore, the spot and "partial spot" regions will be found and located (i) by a preprocessing step using gray-value thresholding and (ii) by feature analysis using the curvature criterion.

2.1.3 Preprocessing by gray-value thresholding

In this step the number of possible spot regions is reduced by a simple gray-value thresholding knowing that spots have significantly lower gray values than the background (black $\cong$ 0, white $\cong$ 255) of a gel image. For each region obtained after the WST the average gray value is determined. The applied thresholds are evaluated depending on the gray value distribution within the gel image. Each spot and partial spot region in the mosaic image borders on a background region. Consequently, only regions that have a neighboring region with a significantly higher average value could be a spot or a partial spot region. For further consideration we introduce the following variables: R, set of regions; $N = |R|$, number of regions in the mosaic image; Neighbor(r), set of neighboring regions of $r \in R$; $\mu(r)$, average gray-value of $r \in R$;

Criterion 1: The neighbor difference $D(r)$ of a region given by

$$D(r) = \inf\{|\mu(r) - \mu(r')| \; | \; r' \in \text{Neighbor}(r)\} \qquad (1)$$

A candidate for a spot or partial spot region is a region r so that $D(r) > t_1$, where t_1 is a threshold given by 40% of the gray-value variation within the considered gel image. However, this assumption is not true for all partial spot regions. Especially partial spot regions in the middle of spots do not border on a background region. In general, these types of partial spot regions have very low gray values and can be found using a second threshold value, t_2.

Criterion 2: A candidate for a spot or partial spot region is a region $r \in R$, so that $\mu(r) < t_2$, where t_2 is a threshold given by 20% of the total gray value variation within the gel image. Both threshold values (40 and 20%) were determined empirically.

2.1.4 Feature extraction of regions

Although the preprocessing step that uses two specific gray-value thresholds reduces the number of candidate regions significantly, it is not specific enough to find exactly the correct spot candidates or to remove all nonspecific regions. Regarding the gel image as a topographic surface, one of the obvious features of spots is their convex curvature. This feature is also true for partial spot regions. Since the watershed transformation is performed on the first derivative the watersheds separate regions of convex and concave curvature. Thus, we only need to find those regions among the candidates which have a convex curvature. The curvature $C(r)$ of a region $r \in R$ is defined by

$$C(r) = \Sigma f''(p) \qquad p \in r \qquad (2)$$

where the sum is taken over all pixels p in region r, and f'' is the second derivative of the image. A candidate for a spot or partial spot region is a region r that satisfies $C(r) > 0$.

Applying the two gray-value criteria as well as the curvature thresholding, only such regions remain that correspond to partial spot or spot regions. Next we describe how to merge partial spot regions.

2.1.5 Merging of partial spot regions

The merging process of the partial spot regions of a spot is based upon two additional properties of spots: (i) a spot should have an approximately elliptical shape; (ii) for neighboring partial spot regions within a spot, the local curvature C_L in convex in a small neighborhood along their common border (watershed line). Since each of these two properties can not sufficiently characterize the regions, a combination of these features is used. To examine the similarity of a region r to an ellipse, an ellipse E_r is constructed using the circumscribed rectangle of a region (Fig. 4). Then, using a χ^2 test, the difference or similarity between the constructed ellipse E_r and the region r is evaluated. A region should be merged with neighboring regions if the similarity of the merged region to an ellipse is higher than the similarity of each separate region to an ellipse. The local curvature C_L for neighboring regions r and r' is determined by the second derivative in a 3×3 pixel neighborhood along their common border. The second derivative is computed in the direction of the line connecting the centers of gravity of r and r'. Partial spot regions, together representing one spot, are character-

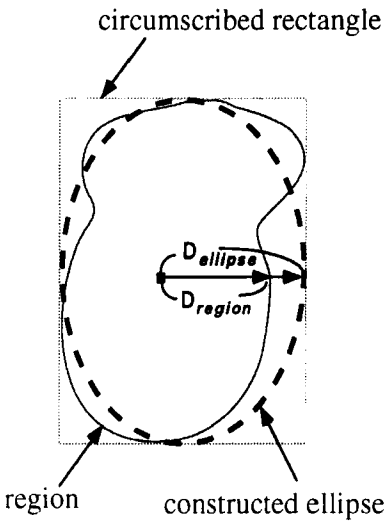


Figure 4. Examination of the similarity of a region to an ellipse. The ellipse is constructed using the circumscribed rectangle of a region. Using the χ^2 test the difference of the lengths d_{region} and $d_{ellipse}$ (the similarity between the constructed ellipse and the region) is evaluated.

ized by a positive curvature along their watershed line. In this case the partial spot regions can be merged. If, in contrast, the local curvature is negative (or concave), then the two regions should not be merged. By using both criteria the algorithm should, for instance, not merge two or more neighboring circular spots, even if they overlap; nevertheless, spots consisting of separate spot regions, resulting from the segmentation process, should be merged. Each criterion alone is too weak for the merging process. Finally, we associate with each identified spot s the coordinates $x(s)$, $y(s)$ of its center of gravity and we define its intensity $i(s)$ as the sum of all gray-values in the region. The vectors $x(s)$, $y(s)$, $i(s)$ of all spots are collected in a spot list.

2.2 Modeling the matching problem and main algorithmic ideas

2.2.1 The point pattern matching approach

One of the fundamental tasks of gel analysis is the comparison of two gel images in the sense that one wants to determine those pairs of spots which represent identical proteins. Several algorithms have been designed and implemented to solve this so-called gel matching problem [11, 13–15, 21–25]. The efficiency of most of these algorithms can be improved by preassigning several corresponding spot pairs (landmarks). Some of the algorithms even require a certain number of landmarks. Here, we present a geometrical matching approach which can be used to search for similar protein patterns in a source and a target gel image. The automatically determined corresponding spots can be taken as landmarks for global matching between entire gels. Note that the algorithm of Olson and Miller [22], which is the basis for the matching implementation in the gel analysis system Melanie II, also can be started without preassigned landmarks. The main advantage of our geometric method is that it takes into account similar geometric patterns and spot intensity relations and is independent of geometric and densitometric (gray-value) image resolution. Furthermore, this method is also robust in the presence of noise (additional / missing spots). The only assumption is a proper spot detection algorithm extracting the geometrical data $x(s)$, $y(s)$, $i(s)$ of the spots from the images. We remark that our matching algorithm uses the intensities only to get an intensity order relationship of the spots in an image (the numerical values $i(s)$ themselves are not important).

We start with some general remarks about geometric point pattern matching, which has been the subject of intensive research in the last decade [26–31]. Let P denote a point pattern (in our application P will be a set of intensive points chosen from a small window in a source image). Given another point set, T (the spots in a target image), all occurrences of P in T should now be computed. This is a similar task as looking for a given star constellation in a celestial chart. Usually, an admissible space A of transformations (*e.g.*, translations, rigid motions and/or scalings) is given which can be used to map the pattern as close as possible to the point set T. Additionally, we have a distance measure d between patterns. The Hausdorff distance H is most commonly considered [32]. In general, we want to find an $f \in A$ and a pattern Q in T for which $d(f(P),Q) \leq \varepsilon$, where ε is a prescribed error tolerance. We distinguish between exact matchings ($\varepsilon = 0$) on the one hand and approximate matching solutions on the other hand. The latter are important in most practical applications. As in our concrete application, it is sometimes only possible to find partial matchings, *i.e.*, we will be looking for as large as possible sub-patterns of P which have an approximate matching pattern in T.

Two approaches which have proven to be useful for our application are considered: the alignment method and geometric hashing. The alignment method is based on the observation that any similarity transformation is determined up to reflection by the mapping of a single line segment. Thus, we will choose two points (a, b) in the pattern P and map the edge $\overline{ab}$ to all edges $\overline{uv}$ in the target set T. For each mapping we check whether it induces a (large partial approximate) matching of P. Note that in the special situation of partial matchings it is not sufficient to con-

sider only one edge $\overline{ab}$. In a worst-case scenario one thus has to map all pattern edges to all edges in T. The situation is much easier to handle if scalings are not allowed (or strongly restricted). Then for a given edge $\overline{ab}$ it is sufficient to search for all target edges of approximately the same length. Analogously, if rotations are forbidden it suffices to search for all target edges with approximately the same slope as $\overline{ab}$. An essential speed-up of the alignment method can be obtained if the points in both the pattern and the target are labeled with positive values (intensities) such that for any valid matching the intensity order of the pattern is consistent with the intensity order in the target. Then the quadratic size search space of all target edges can, under certain circumstances, be reduced to the set of all edges in the history of the incremental Delaunay triangulation of T, which has an expected linear size. The details of this idea will be discussed below. The main drawback of the alignment method consists in the waste of time caused trying to construct a matching for each legal pair of edges $\overline{ab}$ in P and $\overline{uv}$ in T; in the end, however, only few of these attempts will be successful. This can be avoided by making use of geometric hashing. This method requires some preprocessing for a hash table. After that, the number of pattern points which are matched by a transformation (induced by a legal edge pair) can be counted efficiently. Finally, one has to compute the matches only for the best transformations. Some basic ideas and principles, such as the alignment paradigm or Delaunay triangulation graphs, have been rediscovered several times and used for the matching task. The two novelties of the algorithmic solution we present are the following: the construction process of the Delaunay triangulation graph rather than the final graph itself is used for the matching process, and the approach works in the case of noise.

2.2.2 Local matching queries

First, we address the modeling of the following algorithmic Local Matching Query Problem: Given a local spot pattern P selected from a 2-D gel source image, find all local spot patterns in a target image T that resemble at least partially both the geometric shape and the spot intensities of P. Later we describe how to use results of local matching queries to form a global match between source and target image.

The algorithm to answer local matching queries starts from two assumptions. First, the geometry of spot patterns is given by point patterns. Second, the algorithm should not use the relative position of the selected pattern within the source image or its location with respect to possibly given landmarks. Such information, if available, can be used to speed up our solution considerably, for example by restricting the search range in the target image. Moreover, source and target image are assumed to have the same bounding box; otherwise they are scaled linearly. The spot intensities induce a linear order in the spot list. The choice of admissible transformation space and a distance measure to evaluate matches is based on the following observations and assumptions.

(i) Assume we want to choose a pattern P from a small rectangular window in the source image S. Source and target image can have significantly different spot numbers but, since intensive spots tend to appear first, it only makes sense to choose and restrict oneself to such patterns P that consist of the locally most intensive spots. The default value for the pattern size in the implementation is eight. (ii) On the other hand, a matching pattern P' in the target T should also consist of locally intensive spots. Moreover, we should also accept solutions in which P' resembles only a large portion of P. In this way we can also try to correct certain errors made by the spot detection algorithm, which tends to have difficulties in correctly interpreting those spots that are very close to each other and partially overlap. (iii) To model the pure geometric resemblance between P and P' we use the following simple rule based on two real tolerance parameters λ and α. We call two line segments $\overline{st}$ and $\overline{s't'}$ (λ,α)-similar if their absolute slope difference is smaller than α and their length is: $1\text{-}\lambda \leq |\overline{st}|/|\overline{s't'}| \leq 1+\lambda$. Two point patterns P and P' are (λ,α)-similar if there is a bijection f between the point sets such that $\overline{st}$ and $\overline{f(s)\ f(t)}$ are (λ,α)-similar for all $s,t \in P$. In total we want to find (λ,α)-matchings between as large as possible subpatterns $P'' \subset P$ and target patterns P' (also compare [30]). (iv) To model the intensity resemblance between spots we do not use the absolute intensity values $i(s)$ directly. Instead, we apply the following robust heuristic ranking rule that assigns to each spot a discrete intensity integer between 1 and 10. The 500 most intensive spots in an image are distributed equally according to cardinality between 10 and 6; the remaining spots are assigned integer values ≤ 5 such that the total intensity sum in each class is the same. For the matching we use the criterion that a pattern spot s can only be matched to a spot in P' if their discrete intensities differ by at most 2. (v) Since the edge similarity constants λ and α are usually small (in the implementation the default values are $\lambda = 0.2$ and $\alpha = 0.2$), we know each (λ,α)-matching between $P'' \subset P$ and P' is close to a translation t; more exactly, the Hausdorff distance $\tilde{H}(t)(P''),P')$ between the translated P'' and P' is in the worst case bounded from above by $\max_{s,t \in P''} \varepsilon |\overline{st}|$. Besides the size of P'', another criterion for evaluating the match could be the Euclidean distance of the center of P' from the expected center position of the transformed P in the target image, provided its position in the source is known.

2.2.3 The basic algorithmic idea: using the Delaunay triangulation

Given the above-described setting, our local matching algorithm is based on the following key idea (first used in [33] where the task was to find a star constellation in a star atlas). Let us call a triplet of spots in a gel image intensive if its circumcircle does not contain a spot that is more intensive. An edge connecting spots s,t is intensive if there exists a third spot forming an intensive triplet together with s and t. This concept of intensive edges is strongly related to the Delaunay triangulation construction known in computational geometry. A triangulation of a point set S in the plane is called Delaunay triangulation if for each triangle in the triangulation its circumcircle contains only the three triangle points. One can construct such a triangulation incrementally by adding one point after the other (compare [34]). Now the straightforward but central observation of Weber *et al.* [33], in our terminology

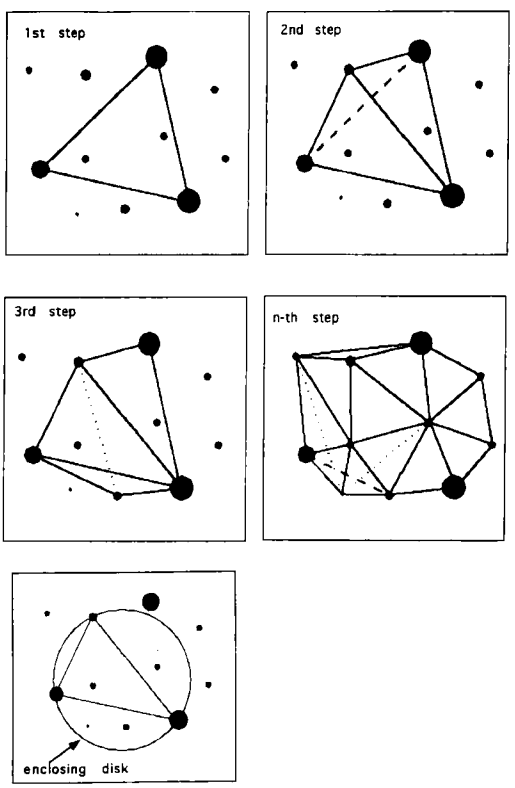


Figure 5. Construction of an intensive triangle and incremental Delaunay triangulation for spots in order of decreasing intensity. The enclosing disk does not contain a spot that is more intensive than the spots forming the triangle.

reads: Assume that the Delaunay triangulation of a gel image is computed incrementally by inserting spots in order of decreasing intensity. Then the set of all Delaunay triangles and edges occurring during the history of that process is exactly the set of intensive triangles and intensive edges.

Let Hist* (T) be a data structure representing all intensive edges together with their lengths and slopes as well as so-called flipped diagonals (see [35] for definition). Figure 5 illustrates the incremental construction. In the example of Fig. 5 the dotted edge is deleted from the current triangulation; four Delaunay edges and two flipped diagonals are added to Hist* (T). We use the edge set Hist* (T) instead of the set of all edges in the target for the following reasons: (i) While the number of all edges is quadratic in the n number of spots, it can be shown that the expected number of edges in Hist* (T) is bounded by $12n$. (ii) If a pattern P of locally intensive spots occurs in T, then one can expect that, despite the possible noise that causes changes in the intensity order, many of the edges connecting spots in P will be (λ,α)-similar to edges in Hist* (T); see [35] for empirical data from random point patterns.

This is the point where our approach and that of Weber *et al.* [33] diverge. In the latter, according to the alignment technique, one tries to extend each occurrence of a target edge that is of equal length with a source pattern edge to a matching of the complete source pattern; we have to opt for a different strategy. The main reason for this is the small but nevertheless considerable length and slope tolerance that imply an edge search range in Hist* (T) that is too large (although already of linear size). Our alternative to the alignment technique [33] is a two-step variant of geometric hashing (see [26]). First, we compute all locations within the target image where a good matching with the pattern is likely to occur. Only then we compute actual patterns that answer our local matching query.

2.2.4 Computing tentative matching pattern locations

For each pattern edge e we find all target edges e' in Hist* (T) that meet the tolerance bounds with respect to length, slope, and discrete spot intensities. However, we do not store the results of such a query as an edge list. We first compute the vector $t(e,e')$ that translates the midpoint of edge e to the midpoint of e'. For all these we maintain a scoring table that indirectly stores translation vectors and at the same time yields clusters of such vectors. Observe that such clusters correspond to possible matching locations. This is done as follows. The bounding box of T is interpreted as the possible space for translation vectors $t(e,e')$. Next we overlay a regularly spaced grid on the

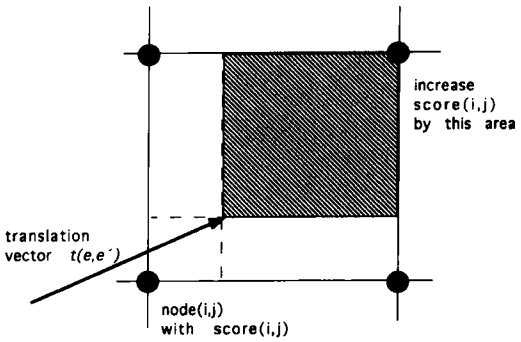


Figure 6. Updating the score for a pair e, e' of similar edges.

translation space and maintain a data structure for integer scores, initially all zero, which are defined for each grid node. Each translation vector $t(e,e')$ increases the score among the four grid nodes defining the grid cell into which the vector falls. This cell is subdivided into four rectangles by $t(e,e')$ as depicted in Fig. 6. Each of the four grid nodes adds to its current score an amount proportional to the area of the opposite rectangle given the total area by 100. Let Score (i,j) be the total score accumulated in grid node (i,j) after probing all $\binom{|P|}{2}$ pattern edges. All local maxima that are greater than a threshold value depending on the cardinality of P are considered to correspond to potential matching locations (see Fig. 7).

2.2.5 Verifying tentative local matching locations

After the scoring procedure we are given a list of putative locations of matching pattern centers (c) in the target. Recall that each such point c corresponds to a translation, say t_c, for which we compute $t_c(P)$, *i.e.*, we map the pattern into the target. For each source spot $s \in P$ we search for the neighbors of $t_c(s)$ in the Delaunay triangulation of T. These constitute a list of matching candidates for the pattern spot s. Using an exhaustive search we select from the candidate lists a maximal subset that defines a (λ,α)-matching with the corresponding subpattern of P (also compare [30]).

2.2.6 Strongly distorted images

A further problem is how to proceed when there is a more severe geometric distortion between the pattern P and its counterpart P' in the target image. Obviously, one solution would be to increase the values for slope and length tolerances, yielding increased time bounds for the geometric hashing process. Instead, we apply a standard heuristic trick (compare [27]); we distort the pattern P iteratively in a typical (application-dependent) way and then search for the distorted pattern while keeping similarity tolerances low. In our case these distortions are combinations of independent x-y scalings and shifts that transform rectangular regions into parallelograms; Distort (P) denotes the list of all patterns derived from P.

2.2.7 Global matching *via* local matching

Previous algorithms for the global matchings of gel images are based mainly on landmarks set by the user. In this context a landmark is a pair of points, one in the source image S, the other in the target T. Thus, the user fixes a partial match on a sufficiently large set of so-called support points $S_{supp} \subset S$. By triangulating S_{supp} and constructing the corresponding triangulation in the target image one gets a piecewise affine transformation f defined on the triangles. Finally, for any source point $p \in S$ one

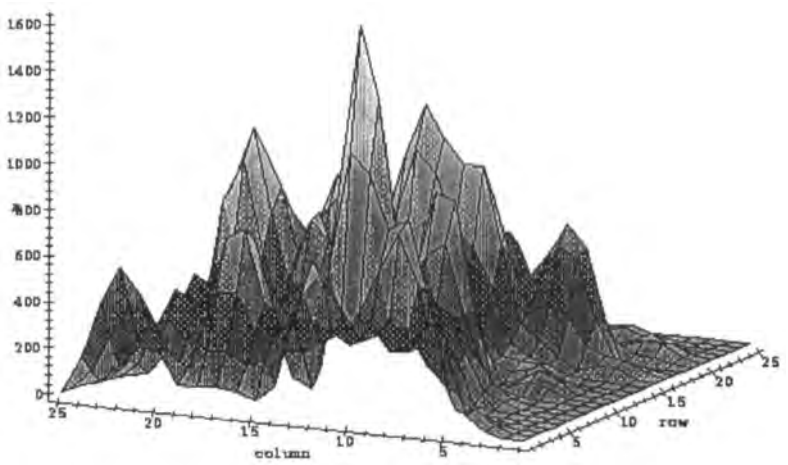


Figure 7. Local result of the scoring procedure (pseudo 3-D representation of score values).

has to search for the nearest neighbor of $f(p)$ in T where the distance is a combination of the Euclidian distance and the intensity difference. Our aim is to avoid an interactive landmark setting by making use of the local matching solution. The problem is that this algorithm computes several matches of a chosen pattern and, in general, it does not clarify which is the right one. There are two approaches to improve confidence in a found match. The first is to insert a spot $p \in S$ into different patterns $P_1,\ldots,P_m$ and to compute their best local matches independently. We considered four patterns extending a bounding box from p into the four quadrants. Then a point $q \in T$ will be accepted as an image of p if for all patterns there is at least one match (of the best) mapping p to q. Answers found this way are sufficiently satisfying. However, due to the strong restrictions within this approach there is a possibility that no answer will ensure. The second approach consists in covering the source image by patterns in a gridlike fashion. Computing the best local matches for all patterns one has to look for a consistent choice of matches. Let P_1 and P_2 be neighboring patterns in S. A matching of P_1 to P'_1 in T will be called consistent with a matching of P_2 to P'_2 if for their centers $c(P)$ it holds that the edges $\overline{c(P_1)\ c(P_2)}$ and $\overline{c(P'_1)\ c(P'_2)}$ are (λ', α')-similar. Here, we have to enlarge the tolerance bounds to reflect the fact that for each single pattern P_i the matching can stem from the list Distort (P_i).

2.2.8 Implementation and user interface

The spot detection and matching algorithms have been implemented and are part of the Carol software system [36, 37]. It has essentially two parts: The first part, the combinatorial and geometrical kernel of the matching algorithms, has been implemented in C++. It makes essential use of the Standard Template Library (STL) and of the Computational Geometry Algorithms Library [38]. The latter library provides several geometric data structures and functions and especially an implementation of the incremental Delaunay triangulation. The second part of the Carol system is the graphical user interface which has been implemented in Java. It can be run as an applet started out of an Internet browser or as an application. The communication with the matching program part is established *via* Internet sockets, whereby the C++ program works as a server which waits for matching requests from the Java client, performs the computation, and eventually sends the results back to the client. The program will be

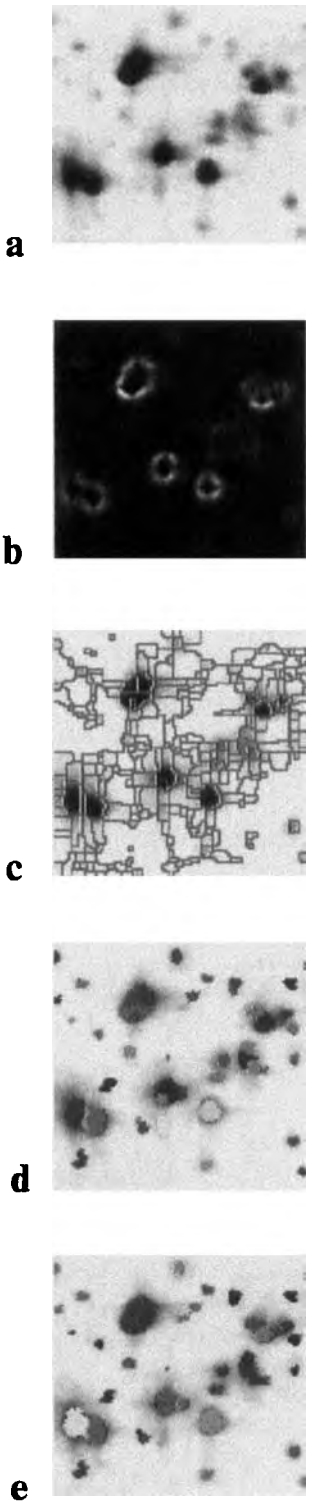


Figure 8. Steps of the spot detection using the WST for a selected part of the HEART-2DPAGE gel image. (a) Original image; (b) gradient image; (c) WST with a strong over-segmentation (mosaic image); (d) localization of spot regions; (e) merging of spot regions according to the merging criteria.

eligible to match gel images from 2-DE databases all over the Internet. This feature is strongly supported by the possibility to run the user interface as an applet and, furthermore, by the client-server architecture of the program. The user has the possibility to open GIF images from any 2-DE databases, to carry out the spot detection, to perform a local matching between the source and target image, and to set parameters like tolerance bounds, pattern size, *etc.* The software system can be launched under http://gelmatching.inf.fu-berlin.de. The local matching algorithm implemented on a Sun Sparc Ultra 1 computes the best nine matches for a pattern of eight spots in about 3 s, including the preprocessing of a 3000 spot target image. Each further pattern in the list Distort (*P*) increases this time by about 0.3 s on average.

3 Results

The spot detection using WST was tested for the original 2-DE gel images of the heart protein databases HEART-2DPAGE and HSC-2DPAGE downloaded from the Internet. For a selected part of the HEART-2DPAGE gel image the steps of the spot detection according to the principal scheme shown in Fig. 1 are illustrated in Fig. 8a–e. After the WST (Fig. 8c) the gel image (mosaic image) is strongly over-segmented. After preprocessing by gray value thresholding, spot regions are localized (Fig. 8d) and merged if merging criteria are fulfilled (Fig. 8e). The majority of spots is well detected. Difficulties arise in saturated spot regions, streaks, and from graphic overlays often seen in WWW images. In order to evaluate our algorithm we compared it with a Gaussian fitting spot detection algorithm used by the PDQuest system. The same gel image was scanned and spots were detected using this system. Figures 9a, b shows the result of that comparison for a part of the gel image. We could observe that our algorithm detects more spots in regions consisting of several overlapping spots and streak regions. Using the matching software under the graphical user interface Carol (Fig. 10), local matchings between gel images from the HEART-2DPAGE and HSC-2DPAGE databases were performed and, indeed, most of the matching results found are consistent with protein-chemical identification.

4 Discussion

For our geometric matching approach we developed a new spot detection algorithm using the watershed transformation. By employing feature extraction methods for spot regions we overcome the over-segmentation difficulties inherent in the WST. Like every spot detection algorithm, our approach also has its advantages and drawbacks, the latter pertaining especially to the detection of faint, overlapping, or saturated spots. Although the major-

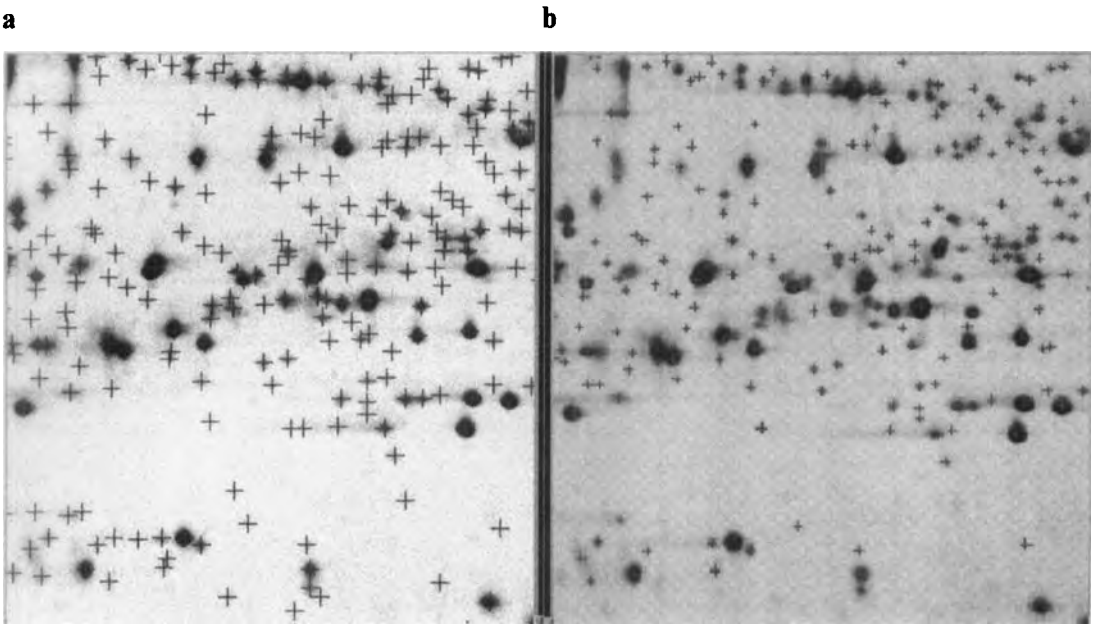


Figure 9. Comparison of (a) the WST algorithm with (b) a Gaussian fitting spot detection algorithm used by the PDQuest system.

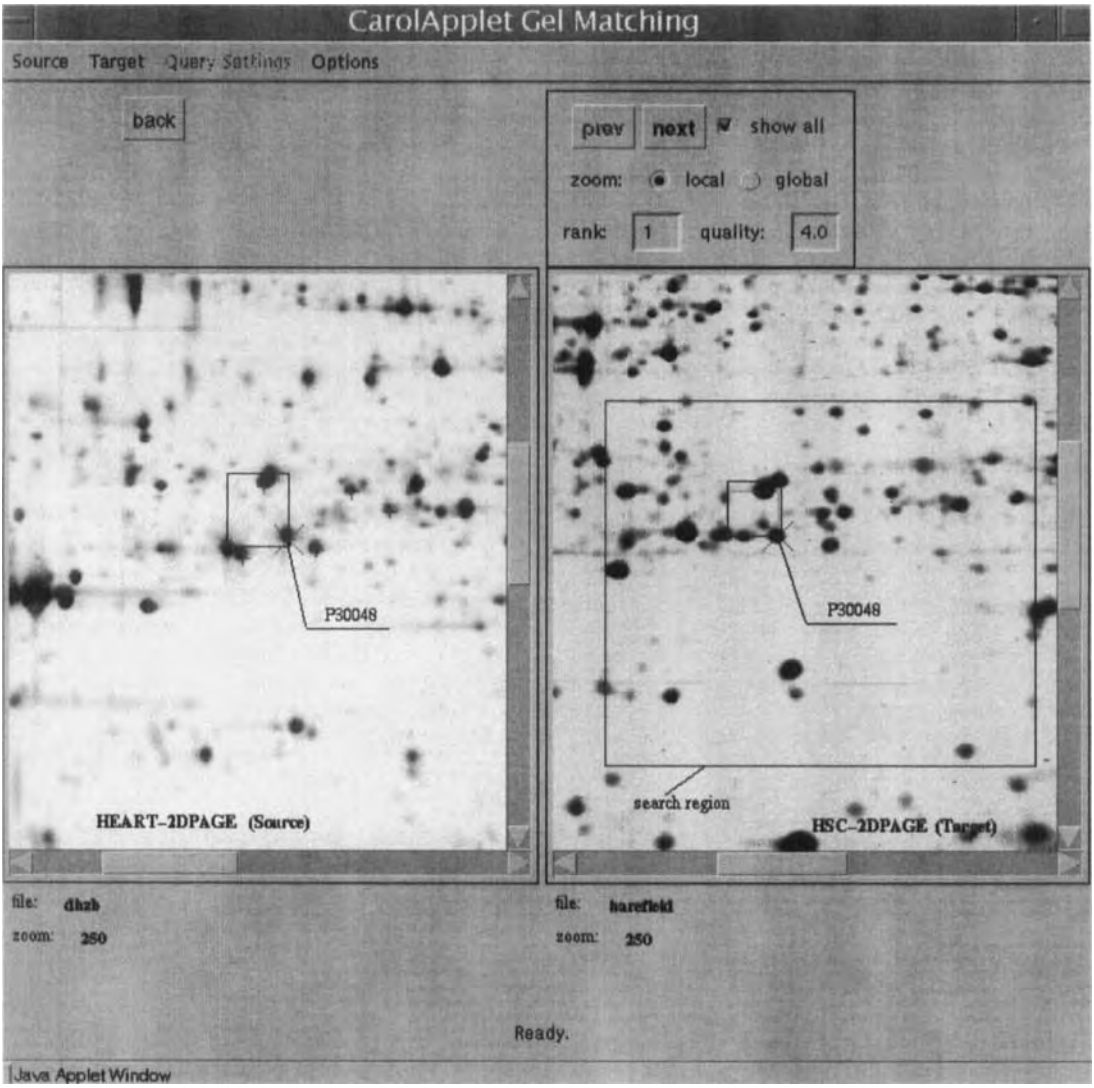


Figure 10. The Carol user interface applied to the comparison of images from the HEART-2DPAGE (source) and HSC-2DPAGE (target) databases. The search pattern (+) given in the source image and the corresponding pattern after the matching. For one match (X) the SWISS-PROT AC found in both databases is shown to illustrate the correctness of the matching result.

ity of spots are well detected, some false results still appear; we therefore plan to include a spot editing tool into the Carol system. This will offer the possibility of interactively correcting false results.

We have presented the underlying ideas for an algorithmic solution of the local matching problem of 2-D patterns of protein spots in electrophoretic images. Its main features are: (i) Local matches for a source pattern are found in the target image without knowledge of its context. (ii) The local matching algorithm works for patterns consisting of the locally most intensive spots. There are standard techniques (*e.g.* point location combined with affine approximation and nearest-neighbor search) that extend the solution to other spots. (iii) The local matching algorithm can be used as a basic step for the global matching problem for gel images. In fact, local matching is then used like a landmark setting. (iv) The central idea

for the algorithm stems from the use of the extended history of the incremental Delaunay triangulation, which proved to be a suitable structure for the local matching problem because of its expected linear size and its robustness in the presence of noise. (v) The transformation of the absolute intensity values to a discrete intensity integer, and the use of a discrete threshold, enable a matching approach that has only a weak dependence on the absolute densitometric resolution (gray-values).

The spot detection using WST and the local matching approach support the automatic comparison of gel images stored in different WWW 2-DE databases. Further work will be especially aimed at improving and refining an algorithmic tool for the global matching task between gel images as provided by other gel analysis systems like Melanie, PDQuest, *etc.* Finally, it is conceivable that our approach can be used for other applications dealing with similar matching-type problems.

We would like to thank Christian Knauer and Sven Schönherr from Freie Universität Berlin for their generous help in developing the CAROL software system. This study was supported by the Deutsche Forschungsgemeinschaft under grant FL 165/4-1.

Received September 1, 1998

5 References

[1] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer-Verlag Berlin, Heidelberg, New York 1997.

[2] Appel, R. D., Bairoch, A., Sanchez, J.-C., Vargas, J. R., Golaz, O., Pasquali, C., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 540–546.

[3] Hoogland, C., Baujard, V., Sanchez, J.-C., Hochstrasser, D. F., Appel, R. D., *Electrophoresis* 1997, *18*, 2755–2758.

[4] Latter, G. I., Boutell, T., Monardo, P. J., Kobayashi, R., Futcher, B., McLaughlin, C. S., Garrels, J. I., *Electrophoresis* 1995, *16*, 1170–1174.

[5] Pleißner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1996, *17*, 1386–1392.

[6] Evans, G., Wheeler, C. H., Corbett, J. M., Dunn, M. J., *Electrophoresis* 1997, *18*, 471–479.

[7] Giometti, C. S., Williams, K., Tollaksen, S. L., *Electrophoresis* 1997, *18*, 573–581.

[8] Pleißner, K.-P., Söding, P., Sander, S., Oswald, H., Neuß, M., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1997, *18*, 802–808.

[9] Lemkin, P. F., *Electrophoresis* 1997, *18*, 461–470.

[10] Lemkin, P. F., *Electrophoresis* 1997, *18*, 2759–2773.

[11] Appel, R. D., Vargas, J. R., Palagi, P. M., Walther, D., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2735–2748.

[12] Pleißner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1997, *18*, 480–483.

[13] Anderson, N. L., Taylor, J., Scandora, A. E., Coulter, B. P., Anderson, N. G., *Clin. Chem.* 1981, *27*, 1807–1820.

[14] Lemkin, P. F., Lipkin, L. E., *Comput. Biomed. Res.* 1981, *14*, 407–446.

[15] Garrels, J. I., *J. Biol. Chem.* 1989, *264*, 5269–5282.

[16] Prehm, J., Jungblut, P., Klose, J., *Electrophoresis* 1987, *8*, 562–572.

[17] Bettens, E., Scheunders, P., Van Dyck, D., Moens, L., Van Osta, P., *Electrophoresis* 1997, *18*, 792–798.

[18] Vincent, L., Soille, P., *IEEE Trans. Pattern Analysis and Machine Intelligence* 1991, *13*, 583–559.

[19] Meyer, F., Beucher, S., *J. Visual Communication and Image Representation 1990, 1, 21–46.*

[20] Wegner, S., Sahlström, A., Pleißner, K.-P., Oswald, H., Fleck, E., in: *Bildverarbeitung in der Medizin*, Springer Verlag, Heft 10, pp. 134–138.

[21] Miller, M. J., Olson, A. D., Thorgeirsson, S. S., *Electrophoresis* 1984, *5*, 297–303.

[22] Olson, A. D., Miller, M. J., *Anal. Biochem.* 1988, *169*, 49–70.

[23] Vincens, P., Tarroux, P., *Electrophoresis* 1987, *8*, 100–107.

[24] Lemkin, P. F., Lipkin, L., *Comput. Biomed. Res.* 1981, *14*, 355–380.

[25] Skolnick, M. M., *Clin. Chem.* 1982, *28*, 979–986.

[26] Alt, H., Guibas, L., in: Urrutia, J., Sack, J.-R. (Eds.), *Handbook for Computational Geometry*, North Holland, in print.

[27] Sonka, M., Hlavac, V., Boyle, R., *Image Processing, Analysis and Machine Vision*, Chapman and Hall 1994.

[28] Chang, S.-H., Cheng, F.-H., Hsu, W.-H., Wu, G.-Z., *Pattern Recognition* 1997, *30*, 311–320.

[29] Cheng, F.-H., *Pattern Recognition Lett.* 1997, *17*, 1429–1435.

[30] Ogawa, H., *Pattern Recognition* 1986, *19*, 35–40.

[31] Finch, A., Wilson, R., Hancock, E., *Pattern Recognition* 1997, *30*, 123–140.

[32] Huttenlocher, D. P., Klanderman, G., Rucklidge, W., *IEEE Trans. Pattern Analysis and Machine Intelligence* 1993, *15*, 850–863.

[33] Weber, G., Knipping, L., Alt, H., *J. Symbolic Computation* 1994, *17*, 321–340.

[34] Guibas, L., Knuth, D., Sharir, M., *Algorithmica* 1992, *7*, 381–413.

[35] Hoffmann, F., Kriegel, K., Wenk, C., *Proceedings of the 14th Annual ACM Symposium on Computational Geometry* 1988, *231–239.*

[36] Alt, H., Hoffmann, F., Kriegel, K., Wenk, C., Pleißner, K.-P., *Electrophorese Forum '97*, Strasbourg, France, Nov. 25–27, 1997, P21, Abstract.

[37] Pleißner, K.-P., Sahlström, A., Wegner, S., Oswald, H., Fleck, E., *Electrophorese Forum '97*, Strasbourg, France, Nov. 25–27, 1997, P20, Abstract.

[38] http://www.cs.ruu.nl/CGAL; (CGAL – The Computational Geometry Algorithms Library)

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 766–774

Jane M. C. Oh[1]
Samir M. Hanash[2]
Daniel Teichroew[1]

[1]Department of Industrial and Operations Engineering, School of Engineering
[2]Department of Pediatrics University of Michigan, Ann Arbor, MI, USA

Mining protein data from two-dimensional gels: Tools for systematic post-planned analyses

There is a considerable need to develop comprehensive, systematic mechanisms to analyze the vast number of proteins that orchestrate various cellular functions and to identify proteins associated with disease or that are affected by pharmacological agents. Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) continues to be relied upon to analyze protein constituents of cells and tissues. We have developed a Laboratory Information Processing System (LIPS) as a computer-based tool for capturing quantitative and qualitative changes in thousands of proteins detected in 2-D gels of various types. Protein databases have been developed to serve as a repository for data processing of the basic and derived data and of findings derived from different studies. There have been remarkable advances both in database technology as well as in the computer hardware that have benefited our effort at mining protein data from 2-D gels. We here review our current efforts aimed at improving the performance and features of our 2-D related protein databases, with particular emphasis on the tools we utilize for database mining *via* a systematic analysis of information known as post-planned analysis.

Keywords: Database / Data mining / Protein / Internet / Two-dimensional polyacrylamide gel electrophoresis
EL 3354

1 Introduction

High-resolution two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) currently provides the most comprehensive analysis system of the whole proteome. A systematic analysis of the human proteome by 2-D PAGE requires computer-based tools to process gel images, to construct protein databases, and to retrieve qualitative and quantitative information pertaining to experimental design, samples, gels, spots, matches, conclusions, findings, biological identifiers, *etc.* Previous tools we have utilized are described in several publications [1–3] and biological data derived from our database analyses have led to numerous findings [4–10]. The objective of our 2-D related database effort is to build database applications that will contain data on many thousands of 2-D gels already produced in our laboratory (currently numbering over 30 000). The requirements of the database are: (i) store basic data and images for our (currently > 30 000) 2-D gels; (ii) store data derived from the basic data and images; (iii) provide queries to retrieve and analyze the existing basic and derived data; (iv) provide access to data in various forms, text, image, relational tables, *etc.*; (v) provide a user-friendly, flexible user interface, including a web-based interface that allows investigators to search for, and discover, new data, facts and findings.

Traditionally, biological investigations are undertaken as a top-down, stepwise process. Such a process treats biological data as "static", *i.e.*, relatively unchangeable. However, in reality, biological research is an iterative, incremental process; particularly in the case of investigations that rely on the analysis of complex patterns as in the case of protein 2-D gels. Thus important requirements of our 2-D-related protein databases are flexibility and scaleability in order to allow multiple iterations and to provide extended capability for subsequent reinvestigations of previously prepared 2-D gels (*i.e.*, the step referred to as post-planned analysis). Using "virtual" gel analysis, an investigator may query our databases for particular findings and use the results of this query to reinvestigate gel images, possibly regrouped from different prior investigations, to generate a new set of derived data and/or findings.

Our protein databases contain all of the information gathered for a particular protein, for a particular cell type from particular projects undertaken in our laboratory. These databases currently consist of a database pertaining to secreted proteins analyzed by 2-D PAGE, a 2-D-related lymphoid protein database, and a 2-D-related cancer protein database. The user interface, which includes a World Wide Web access, provides the capability to view new and analyzed data from all the databases in a uniform, seamless fashion, including linking with other databases.

Correspondence: Samir M. Hanash M.D., Ph.D., Department of Pediatrics, Box 0684, The University of Michigan, Ann Arbor, MI 48109-0684, USA
E-mail: shanash@umich.edu
Fax: +734-647-8148

Lemkin [11, 12] described a method for the sharing of information based on the reference images.

2 Database

2.1 Database model

Our database model considers the following: (i) Data collection using spot detection and matching software (*i.e.*, spot list, matched spots) is always incomplete relative to image data in that not all spots are detected and not all spots are matched. (ii) The quality of the results depends on the quality of data. If the quality of data collected is judged not acceptable, then it is necessary to recollect data, *i.e.*, collect more accurately or more thoroughly. (iii) Judgement rules and decision-making processes are not clearly defined. Analysis is open-ended. The protein databases are the central repository for data processing and information analysis. Their value stems from the capability to track changes in protein expression and post-translational modification across many samples and to derive quantitative and qualitative information about proteins detected in 2-D gels. Our protein databases encompass four main areas. These are: (i) study design area, (ii) image analysis and matching area, (iii) conclusion area, and (iv) biology knowledge area. A top-level view of the areas and relationships is shown in Fig. 1. Each of these is decomposed into basic entities and relationships.

2.1.1 Study design area

Investigations undertaken using 2-D gels, as with other studies, are generally intended to address a particular hypothesis. In the case of 2-D gels, hypotheses relate to new biological knowledge about proteins. Data in the study design area specify what, and how, biological sources and different types of samples are selected, how 1-D and 2-D gels are processed, and what, and how, protein data are collected and analyzed. All studies undertaken use samples derived from or subject from cell lines, and usually involve study samples as well as controls. Samples are frequently processed into more than one gel. For example, an investigation may involve 20 single cell-derived clones of a cell line and each clone is processed in duplicate. In this case, 40 study images are processed and matched to the master image, and collectively are considered to belong to the investigation.

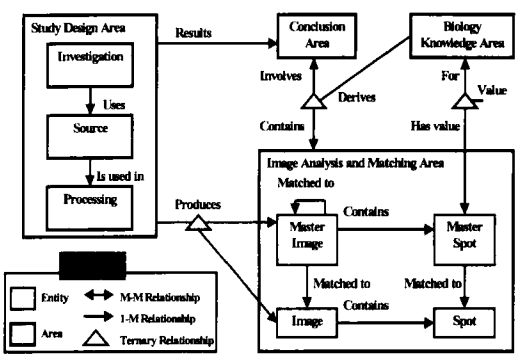


Figure 1. Entities and relationships of protein databases

2.1.2 Image analysis and matching area

This area includes image acquisition, spot quantification, and matching. Image acquisition is the process in which 2-D gels are digitized and image files produced from densitometric data or a PhosphorImager "count" type of data, usually consisting of 1024 by 1024 pixels (8 bits/pixel). Spot detection and quantification result in the spot data for an image with spot numbers, X and Y coordinates, and intensity; the spot lists are collected and stored in spot files. Matching means each spot in an image (study or master) is matched to a spot (real or virtual) on a next higher level master. Each image may be matched to one or more master images and a master image may be used as a match for two or more images. Each master image (with the exception of one) is matched to a higher-level image and each master image may be used as a match for two or more lower-level images. The number of images matched to the master image is limited to the number of images within the same predefined hierarchy after running spot detection and matching software.

2.1.3 Conclusion area

This area covers the findings and investigations are recorded once for all images within the same predefined hierarchy after running data analysis and statistical software. An investigation results in one or more conclusions and a conclusion may involve one or more protein spots catalogued in a master image as well as one or more entities from the study design and image area.

2.1.4 Biology knowledge area

The data obtained from a study may contribute knowledge about particular known proteins such as their involvement in a disease or in the cellular effect of a particular drug. This area provides for identifying and relating the protein spot on a master image to one or more known biological identifiers such as protein name, expression, *etc.*

2.2 Database schema

The schema for a particular database is derived from the top level Meta model, outlined above (Section 2.1) by

identifying the specific entities, attributes, and relationships needed for the investigations. In addition, the protein databases are designed to provide for the following requirements: (i) Accept data in a form that is recorded by image processing software, *i.e.*, attribute values. (ii) Allow simple queries (*i.e.*, number of spots detected or numer of spots matched) to be processed. (iii) Allow complex queries (*i.e.*, virtual matching or outside data linking) to provide the means for testing hypotheses and deriving new knowledge. The relational schema for the protein database is given in Fig. 2. The database needs to provide a suitable structure for deriving relationships for pairs of images. In the database, an image belongs physically to a hierarchy (a hierarchy means that matching is done and the findings are recorded); however, it can be virtually categorized or classified in many different ways.

The "front end" analysis software running on SUNs supplies the data by FTP or direct file transfer. The data consists of:

1. For each (study) image:
(1) The image is in TIFF (or other) format
(2) Image data
- Image ID
- Image type
- Master image ID
- Other values (*e.g.*, gel processing, comments, *etc.*)

(3) A spot list that consists of an entry for each spot
- Spot number
- Spot type
- *X* and *Y* coordinates
- Other values, such as integrated intensity, *etc.*
- Master spot number

(4) Sample data
- Sample number
- Sample identity, location, gel creation, loading, and storage
- Experiments (*i.e.*, differentiation/activation)
- Cell type (*i.e.*, TK-6 cell line or sheep-PB monocytes)
- Treatments (*i.e.*, not treated, day 3 control okt 3)
- Fraction (*i.e.*, whole cell)

2. For each master image
(1) The image
(2) Master image data
- Master image ID
- Other values (*e.g.*, comments)

(3) A master spot list that consists of one entity for each spot
- Master spot number
- *X* and *Y* coordinates
- Other values (*e.g.*, protein data pointer)
- Real or virtual
- (If virtual, it is necessary to know which image or images it came from)
- Remarks, findings

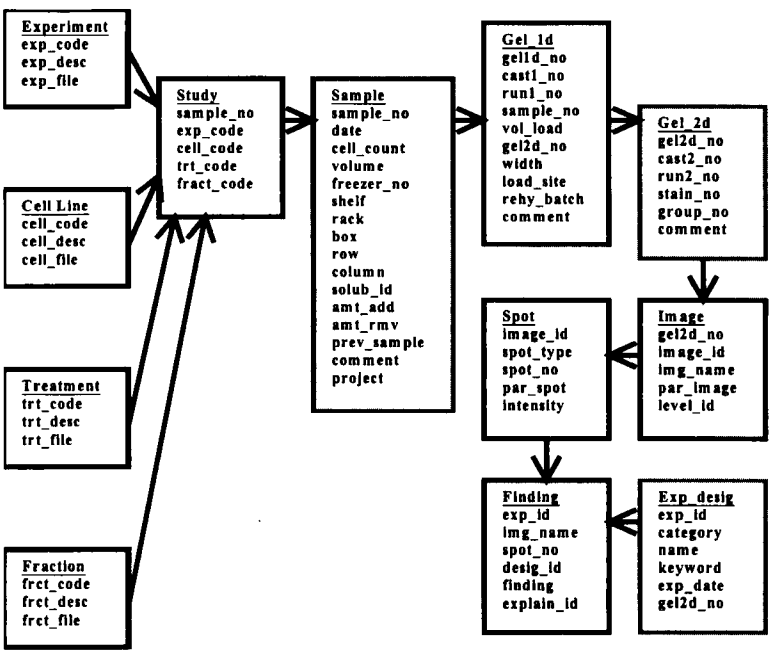


Figure 2. Schema of protein databases

(4) Protein data
- A protein name/description field
- A list of the known synonyms in our protein databases as well as other external databases
- A list of the accession numbers for the gene in GenBank, PIR-International, SWISS-PROT, YEPD (Yeast Electrophoretic Protein Database), *etc.*
- A list of the attributes of protein such as the calculated isoelectric point (p*I*) and molecular weight (M_r).
- The protein sequence data that are useful for confirmation of the identity of the protein.

3 Computing platforms

3.1 The network architecture

Computer hardware platforms have been continuously evolving and the performance and power has been dramatically improved. The reduced cost of powerful workstations, large capacity of hard disks and tapes, improved quality of scanners, and fast speed of networks, and other technologies make it possible to perform real-time window-based applications of displaying and processing text, voice, graphics, images, and other types of massive data. Distributed computing platforms, for example network protocols like TCP/IP, inter-network architecture like Internet, and distributed application architecture like Client/Server, allow investigators to use heterogeneous computer hardware, operating systems and networks and allow interaction with other geographically distributed sites. Figure 3 represents a top-level view of computer hardware platforms and network architecture that are used in our studies to develop and implement systems of interest by using increased computing power, capacity, and speed of computers and networks.

Powerful computer: Sufficient memory (*i.e.*, 128MB) and computing power are required to load multiple images at a time.

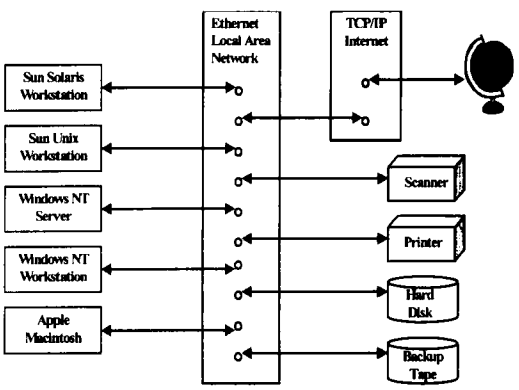


Figure 3. Network architecture

Fast connection: A fast communication connection (*i.e.*, 100 Mbps) is required to transmit images and data to remote access.
Web browser: A Java-capable web browser is required to execute Java applets.
High screen resolution: Computers which have a high screen resolution (*i.e.*, 1280x1024 SVGA for the best result) are required to display images on screen.

3.2 System architecture

Integrated development computer tools are used to design applications, to develop programs, to build user interfaces, to test both programs and interfaces, and then implement and configure software into the Run Time Environment. System development over the last few years has moved from proprietary content to open content. Creating state-of-the-art systems required understanding of computing platform and network architecture, mastery of programming language and integrated development tools. System architecture is shown below (Fig. 4).

Development software: There are several ways to develop software. The state-of-the-art technology provides an Integrated Development Environment (IDE) that supports a system development life cycle of design, program, test, and implement software.
RDBMS: There are several ways to manage and store data. Our protein databases use a relational database management system (RDBMS).
Web server: We use a Web server to host internet database applications to allow remote access to data from the database and to images from the image repository.
NT server: NT server is our operation system. It allows us to use a password protection to access data from the da-

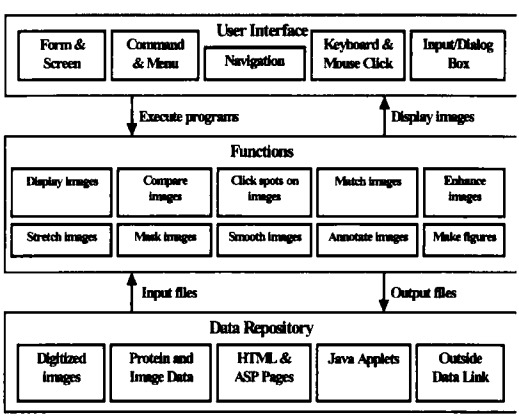


Figure 4. System architecture

tabase and to retrieve images from the image repository. Image repository: The image files are stored on our own Web server. Web servers are installed on a Microsoft Windows-NT, which is connected to the Internet *via* TCP/IP. Image formats: We generate JPEG files for 2-D gels and they are reduced in size to shorten transmission times. Image loading times: Image loading may take a long time since it needs to download the JPEG files from the Internet.

4 Application development

4.1 Application architecture

The application consists of: user interfaces running on a web browser, functions running on a web server, data stored in a database, and images stored in a repository (see Fig. 5; application architecture). Functionalities available in the application are: (i) search functions, (ii) display functions, (iii) image enhancement functions, (iv) virtual matching functions, and (v) common image processing.

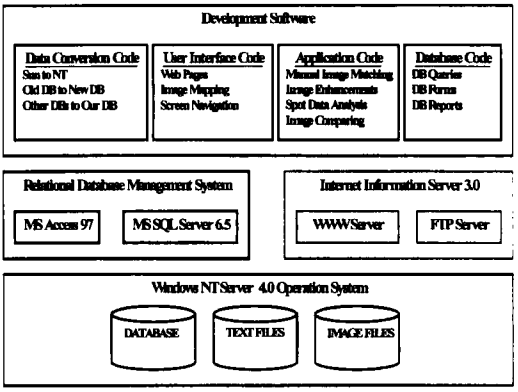


Figure 5. Application architecture

4.2 Database query

The aim of post-planned analysis is to extract nontrivial, implicit, previously unknown and potentially useful information from databases. Post-planned analysis may be referred to as data mining and knowledge discovery [13–16]. An important feature of our protein databases is that all images are matched to masters in one hierarchy so that: (i) there is only one Master Image that is not matched to any other master image; (ii) every study image is matched to one master image; (iii) every master image is matched to one (higher) master image. This allows post-planned analysis that uses the basic experimental data and derived data from a study and investigates a value beyond the context originally planned for in the study de-

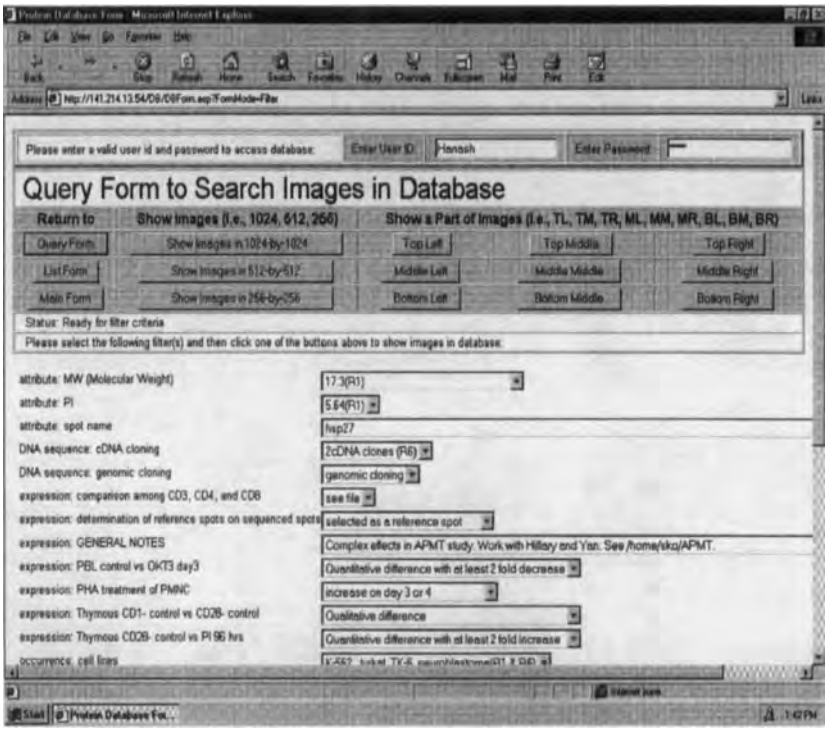


Figure 6. Search by biological identifiers

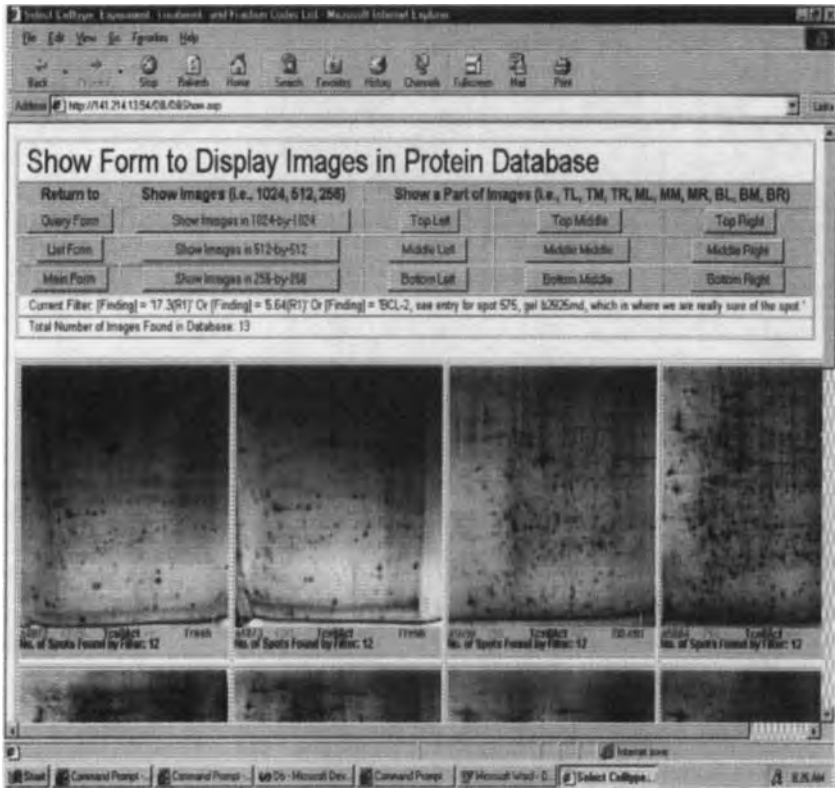


Figure 7. Display images

sign. Post-planned analysis is made possible because the database has indexing mechanisms that can relate a spot to any gel in the hierarchy.

The database application will provide the following types of queries: (i) given a spot on a master, find all spots that are matched to it, (ii) given a spot on a study or master image, find all spots that match it, (iii) given a value for a biological identifier attribute, find all spots with that value, (iv) given a spot on a study or master image, find all conclusions in which it is involved.

The criteria are as follows: (i) the term "all" in the queries can be subject to any criteria involving "source" or "investigation" values; and (ii) the term "spot" in the queries can be replaced by a set of two or more spots.
The queries will also report negative information: (i) when images that satisfy the criteria exist but no spot was found; and (ii) when there are no images that satisfy the criteria.

4.3 World Wide Web user interface

Investigators that add images and other related data to the database are presumed to be familiar with the structure of the database. Other investigators that retrieve data and formulate queries to test hypothesis and establish conclusions do not require extensive familiarity. A user-friendly interface that guides investigators through the architecture is useful for both groups. These users can enter the attributes in three ways. (i) Biological identifiers: The users are able to obtain a list of all biological identifiers for which data exists in the database and then be guided through relevant data by a series of queries that list the appropriate master images (see Fig. 6: search by biological identifiers). (ii) Source materials: The users can obtain a list of source characteristics in four major categories: experiment code, cell type, treatment code, and fraction code. The users can select a combination of some characteristics in one or more of these categories. (iii) Selecting spots on an image.

4.3.1 Displaying multiple images across the Internet

The search functions allow investigators to search the protein database and display 2-D protein electrophoretic gel images on the user's web browser. A sample web page is shown in Fig. 7. One first specifies search criteria

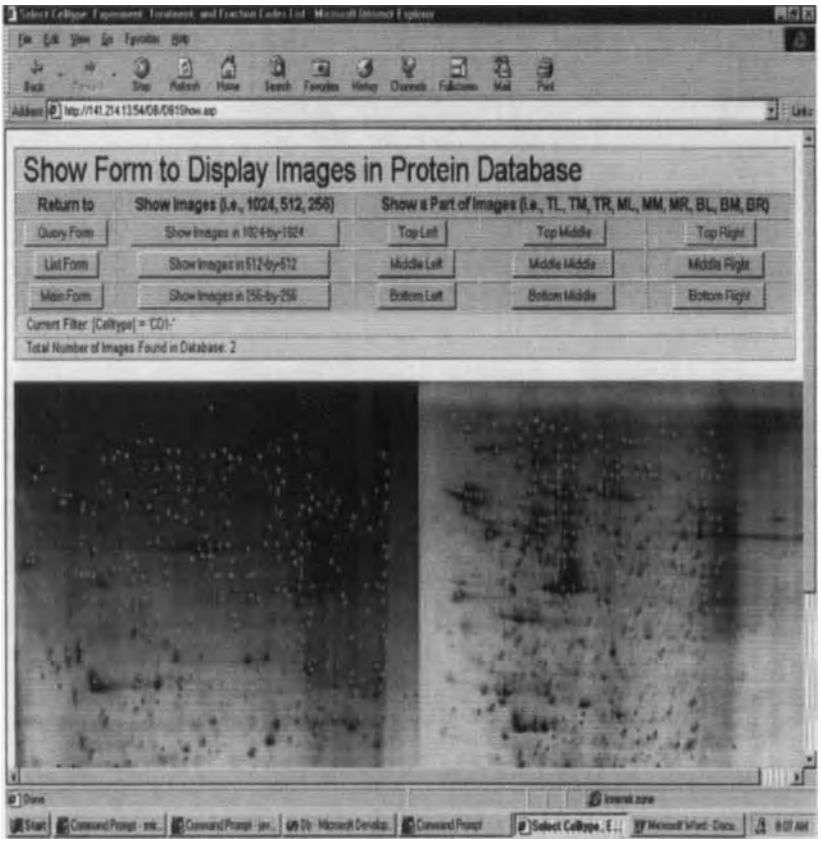


Figure 8. Compare images

for the images (*i.e.*, biological identifiers or source materials), followed by pressing the "display images" key to generate a list of images (see Fig. 7: display images).

4.3.2 Comparing multiple images across the Internet

The protein databases web application provides a method for comparing multiple 2-D gels on a user's web browser. It allows us to compare multiple images visually and to enhance them in various ways such as smoothing, stretching, enhancement, transformation, and other image processing operations. These gel images can be zoomed in and the displaying screen can be scrolled. It allows loading all gel images from our image repository based on search criteria specified. A sample web page is shown in Fig. 8.

4.3.3 Clicking spots of interest to view related information

A section containing a spot of interest can be magnified and centered in the magnified image by clicking on the desired spot. Clicking the spots allows viewing the protein information recorded. A sample web page is shown in Fig. 9.

4.3.4 Clicking spots of interest to match

The same spot can be matched in multiple images by clicking on the desired spot. This results in virtual matching of the images selected and the resulting information is recorded. A sample web page is shown in Fig. 10.

4.4 Virtual matching algorithm

An investigator designs a study, runs samples based on the study design, collects images for a biological study, selects pairs of images for matching, and specifies a sequence of matching. Then one matches all spots in the study images to the master spots in the master images within a defined hierarchy. In addition, a manual visual verification and match editing process needs to be carried out with human capability of visual inspection and knowl-

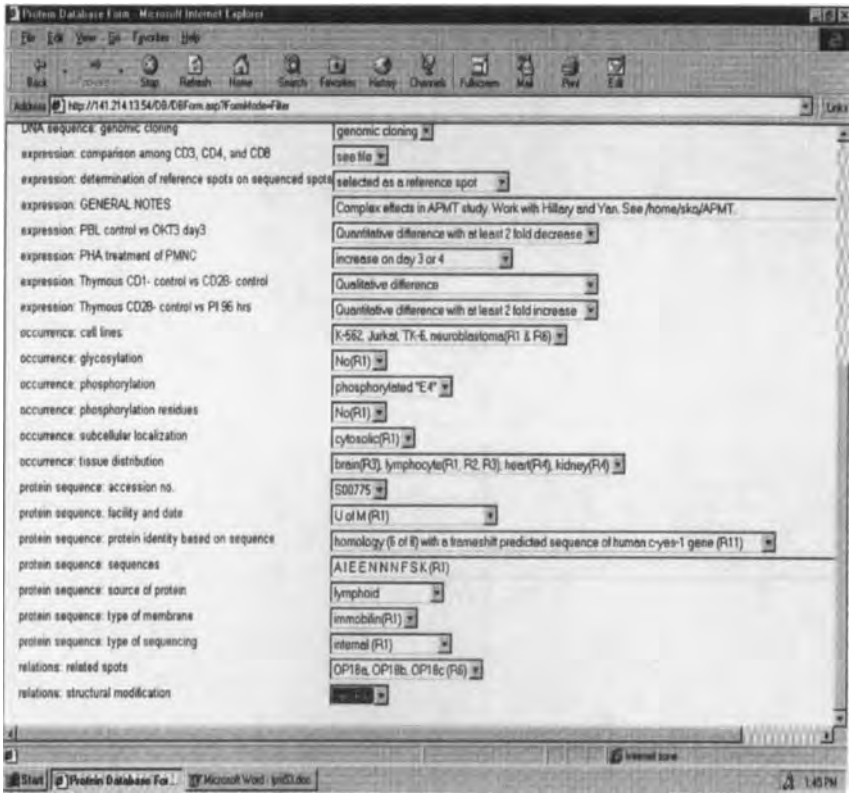


Figure 9. View protein data

edge-based reasoning. It is an exhaustive, time-consuming process and requires a great deal of manual effort for performing many iterations of this image matching process.

The major objects for grouping information are the master images. The first step will usually be for a user to determine which master images are relevant to their interests. In the database, every study image is matched to one and only one higher level master image. It is therefore possible in post-planned analyses to create virtual matching by selecting an arbitrary set of study images and creating virtual masters. This enables an investigator to create a virtual experiment and hence perform the same sequences to determine new findings and conclusions.

4.5 Outside data linking

To expand the capabilities of the databases generated by one group, it is necessary to link data to outside databases. A link to databases such as MEDLINE, GenBank, SWISS-PROT, PIR, PDB, OMIM, UniGene, GeneCards, *etc.* is simple. However linking 2-D databased generated by other groups is much more complicated and requires a great deal of effort to insure accuracy.

Received November 11, 1998

5 References

[1] Ali, I., Chan, Y., Kuick, R., Teichroew, D., Hanash, S. M., *Electrophoresis* 1991, *12*, 747–761.

[2] Hanash, S. M., Teichroew, D., *Electrophoresis* 1998, *19*, 2004–2009.

[3] Kuick, R., Skolnick, M. M., Neel, J. V., Hanash, S. M., *Electrophoresis* 1991, *12*, 736–746.

[4] Beretta, L., Singer, N. G., Hinderer, R., Gingras, A.-C., Richardson, B., Nanash, S. M., Sonenberg, N., *J. Immunol.* 1998, *160*, 3269–3273.

[5] Melhem, R., Hailat, N., Kuick, R., Hanash, S. M., *Leukemia* 1997, *11*, 1690–1695.

[6] Diaz Jr., L. A., Friedman, A. W., He, X., Kuick, R. D., Hanash, S. M., Fox, D. A., *Int. Immunol.* 1997, *9*, 1221–1231.

[7] Haftel, H. M., Chang, Y., Hinderer, R., Hanash, S. M., Holoshitz, J., *J. Clin. Invest.* 1994, *94*, 1365–1372.

[8] Ungar, D. R., Hailat, N., Strahler, J. R., Kuick, R. D., Brodeur, G. M., Seeger, R. C., Reynolds, C. P., Hanash, S. M., *J. Natl. Cancer Inst.* 1994, *86*, 780–784.

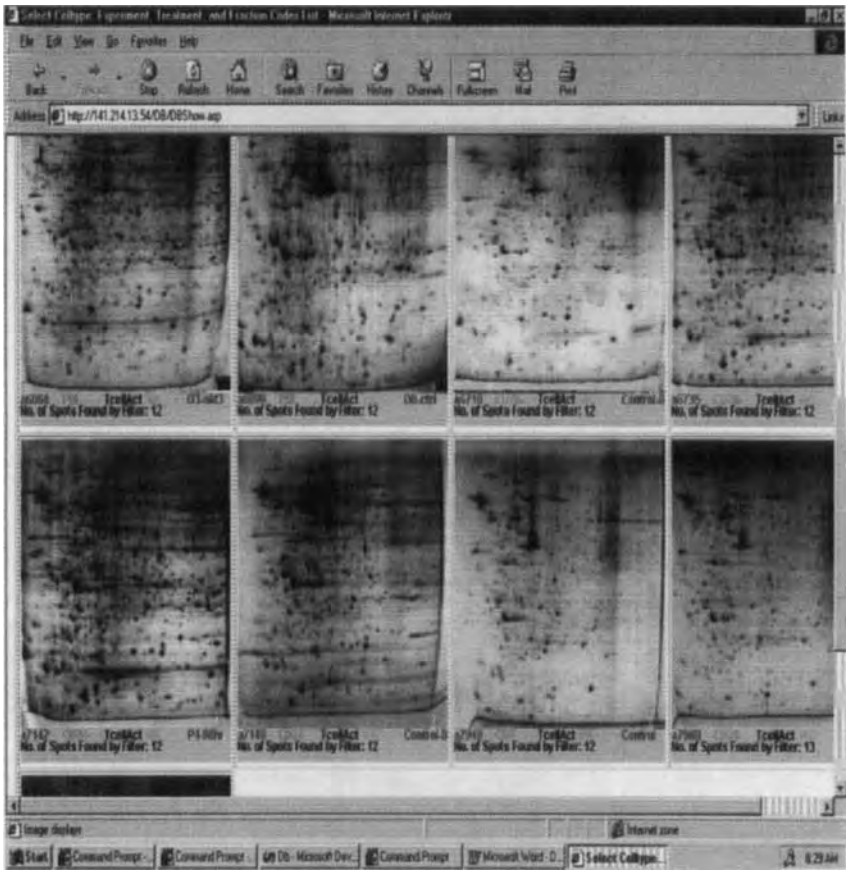


Figure 10. Match images

[9] Strahler, J. R., Zhu, X.-X., Hora, N., Wang, Y. K., Andrews, P. C., Roseman, N. A., Neel, J. V., Turka, L., Hanash, S. M., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 4991–4995.

[10] Hanash, S., Strahler, J., Chan, Y., Kuick, R., Teichroew, D., Neel, J. V., Hailat, N., Keim, D., Gratiot-Deans, J., Ungar, D., Melhem, R., Zhu, X. X., Andrews, P., Lottspeich, F., Eckerskorn, C., Chu, E., Ali, I., Fox, D., Richardson, B., Turka, L., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 3314–3318.

[11] Lemkin, P. F., *Electrophoresis* 1997, *18*, 2759–2773.

[12] Lemkin, P. F., *Electrophoresis* 1997, *18*, 461–470.

[13] Alnahi, H., Alshawi, S., *Computer Methods and Programs in Biomedicine* 1993, *39*, 343–349.

[14] Conklin, D., Fortier, S., Glasgow, J., *IEEE Trans. on Knowledge and Data Engineering* 1993, *5*, 985–987.

[15] Han, J. W., Huang, Y., Cercone, N., Fu, Y. J., *IEEE Trans. on Knowledge and Data Engineering* 1996, *8*, 373–390.

[16] Conklin, D., *Machine Learning* 1995, *21*, 125–150.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 775–780

Allan Christian Shaw[1,2]
Gunna Christiansen[2]
Svend Birkelund[1]

[1]Department of Medical Microbiology and Immunology, University of Aarhus, Denmark
[2]Department of Molecular and Structural Biology, University of Aarhus, Denmark

Effects of interferon gamma on *Chlamydia trachomatis* serovar A and L2 protein expression investigated by two-dimensional gel electrophoresis

Chlamydia trachomatis is an obligate intracellular bacterium causing human ocular and genital disease. The lymphokine interferon gamma (IFN-γ) is an important immune effector exerting antimicrobial effects towards several intracellular parasites, the chlamydia included. IFN-γ has been reported to inhibit the chlamydial replication *in vitro* in part by depleting intracellular levels of tryptophan in a dose-dependent manner. In addition, down-regulation of important immunogens has been described. These findings are extended in this paper, in which we are combining pulse labeling with [^{35}S]methionine and two-dimensional gel electrophoresis with immobilized pH gradients in order to investigate changes in the protein expression of *C. trachomatis* serovar A and L2 caused by treatment with IFN-γ. In contrast to what was observed in *C. trachomatis* L2, our results showed that, in *C. trachomatis* A, down-regulations of the chlamydia major outer membrane protein and of several other proteins were detectable upon IFN-γ treatment. In addition, we report the up-regulations of *C. trachomatis* A and L2 proteins with molecular masses of approximately 30 kDa and 40 kDa which may be part of an, as yet, uncharacterized chlamydial response to IFN-γ treatment.

Keywords: *Chlamydia trachomatis* / Interferon gamma / Two-dimensional polyacrylamide gel electrophoresis / Immobilized pH gradient / Pulse labeling EL 3406

1 Introduction

Chlamydiae is a family of Gram-negative, obligate intracellular bacteria. *Chlamydia trachomatis* is currently divided into 14 serovars; serovar A-C is the major cause of preventable blindness (trachoma), serovar D-K is responsible for sexually transmitted disease and serovar L1-L3 inflicts the systemic disease lymphogranuloma venereum. Chlamydia is characterized by a developmental cycle, which alternates between two morphologically distinct forms, the infective, but metabolically inert elementary body (EB) and the noninfectious, replicative and metabolically active reticulate body (RB). Upon attachment to the host cell membrane, EB induce their own phagocytic uptake. Shortly after entry, the structural integrity of the chlamydia membrane changes, leading to a transformation of EB to RB, which utilize host cell nutrient and divides by binary fission in a specialized phagosome called the chlamydial inclusion. After a period of growth the RB reorganize into EB, and at the end of the cycle a new generation of infectious progeny is released upon disruption of the host cell [1].

Interferon gamma (IFN-γ) is a potent immunoregulatory protein, synthesized by activated T-lymphocytes and NK-cells and plays a major role in the control of chlamydial diseases. IFN-γ added prior to infection inhibits the multiplication of *C. trachomatis* in human cell lines in a dose-dependent manner [2]. IFN-γ applied at the time of or after infection results in a persistent state at which the RB are morphologically abnormal and unable to reorganize into EB [3]. The appearance of atypical chlamydial forms was associated with a down-regulation of the important chlamydial immunogens, the major outer membrane protein (MOMP), the 60 kDa outer membrane protein (OMP2), and chlamydial LPS. In contrast, the expression of the chlamydial GroEl (homolog to the eukaryotic 60 kDa heat-shock protein) was unaltered. This suggests an important role for IFN-γ in persistent infections with *C. trachomatis*, where the Chlamydiae are maintained for a long period of time in a resting, but viable state [3]. In human cell lines, the depletion of endogenous tryptophan correlates with the inhibition of Chlamydiae [4–8], and removal of IFN-γ or exogenous addition of tryptophan is able to restore the normal chlamydial developmental cy-

Correspondence: Allan Christian Shaw, Department of Medical Microbiology and Immunology, The Bartholin Building, University of Aarhus, 8000 Aarhus C, Denmark.
E-mail: shaw@medmicro.aau.dk
Fax: +45-86-19 61 28

Abbreviations: **EB**, elementary body; **FCS**, fetal calf serum; **h.p.i.**, hours post infection; **IDO**, indoleamine 2.3-di-oxygenase; **IFN-γ**, interferon gamma; **IFU**, inclusion forming unit; **MOMP**, major outer membrane protein; **OMP2**, 60 kDa outer membrane protein; **RB**, reticulate body

cle [9, 10]. IFN-γ is responsible for the up-regulation of the enzyme indoleamine 2,3 dioxygenase (IDO), which catalyzes the breakdown of tryptophan to *N*-formylkynurenine [11], thereby restricting the availability of this amino acid for the chlamydia [4, 12–14]. However, the extent and the mechanisms by which the chlamydial replication is inhibited depends on the species, the serovar, and the host cell line used, indicating that additional factors are involved [15, 16]. In this study we investigated IFN-γ-specific changes in protein expression of *C. trachomatis* A and L2 by high-resolution two-dimensional immobilized pH gradient polyacrylamide gel electrophoresis (2-D PAGE (IPG)).

2 Materials and methods

2.1 Bacteria and host cells

HeLa 229 cells were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). Cells were cultivated at 37°C in a 5% CO_2 atmosphere in RPMI 1640 (Gibco BRL, Grand Island, NY, USA) containing 10% fetal calf serum (FCS; Gibco BRL) and 10 μg/mL gentamicin and were found free of mycoplasma by Hoechst No. 33258 staining. *C. trachomatis* serovar A/HAR-13 and *C. trachomatis* serovar L2 (434/Bu) were obtained from ATCC.

2.2 Indirect immunofluorescence microscopy

Semi-confluent monolayers of HeLa cells were infected with one inclusion forming unit (IFU) of *C. trachomatis* serovar A or serovar L2 as described [17]. Cells were washed twice in PBS and fixed in methanol 24 h post infection (h.p.i.). Chlamydiae were visualized, using a monoclonal antibody (mAb 32.2) directed against the *C. trachomatis* MOMP. A fluorescein isothiocyanate-conjugated goat anti-mouse monoclonal antibody (DAKO, Glostrup, Denmark) was used as secondary antibody.

2.3 Pulselabeling with [^{35}S]methionine

Semi-confluent monolayers of HeLa cells were infected with 1 IFU of *C. trachomatis* serovar A or serovar L2 in a medium containing RPMI 1640, 25 mM HEPES, 10% FCS, 1% w/v glutamine, 10 μg/mL gentamicin (serovar A) or RPMI 1640, 25 mM HEPES, 5% FCS, 1% w/v glutamine, 10 μg/mL gentamicin (serovar L2) with or without addition of 100 U/mL human recombinant IFN-γ (Endogen, Woburn, MA, USA). For detection of chlamydial protein synthesis a methionine/cysteine-free RPMI 1640 medium containing 10 μg/mL gentamicin, 40 μg/mL cycloheximide, 100 μCi/mL [^{35}S]methionine/cysteine (Promix, Amersham, UK) with or without addition of 100 U/mL IFN-γ was used. In an additional experiment 200 μg/mL L-tryptophan (Merck, Darmstadt, Germany) was added, together with IFN-γ, to the infection medium and the labeling medium. At the end of each labeling period, cells were washed in PBS and loosened with a rubber policeman in a lysis buffer containing 9 M urea, 4% 3-[(3-cholamidopropyl)dimethylammonium]-1-propanesulfonate (CHAPS), 40 mM Tris base, 65 mM DTE and Pharmalyte 3–10 (Pharmacia Biotech, Uppsala, Sweden). Samples were sonicated and centrifuged at 10 000 × *g* for 10 min. Supernatants containing the labeled chlamydial proteins were stored at –70°C until used.

2.4 Two-dimensional gel electrophoresis with immobilized pH gradient

For isoelectric focusing 18 cm long nonlinear immobilized pH 3–10 gradient drystrips (Pharmacia) were used. Each strip was soaked overnight with a ^{35}S-labeled sample in lysis buffer by means of a reswelling tray (Pharmacia). The amount of ^{35}S-labeled chlamydial protein was adjusted to 150 000 cpm per strip. Rehydrated strips were run in the first dimension at 300 V for 1 h, 400 V for 1 h, 500 V for 1 h, 3500 V for 3 h, and 5000 V for 24 h at 15°C. After the focusing was completed the strips were equilibrated in a buffer containing 6 M urea, 30% w/v glycerol, 2% w/v SDS, 0.05 M Tris-HCl, pH 6.8, and 2% w/v DTE for 15 min. The strips were subsequently equilibrated for an additional 15 min in a buffer in which DTE was replaced by 2.5% w/v iodoacetamide. In the second-dimensional run the proteins were separated on 9–16% linear gradient SDS-polyacrylamide gels (18 cm × 20 cm × 1 mm) until the bromophenol blue front reached the bottom of the gel. Gels were fixed in a solution containing 10% acetic acid and 25% 2-propanol for 30 min, treated with Amplify (Amersham, Uppsala, Sweden) for 30 min and vacuum-dried before exposure to Kodak Biomax MR X-ray films.

2.5 Computer analysis

X-ray films exposed for different time lengths were scanned on an HP Scanjet 3c/T. Analysis of changes in protein synthesis due to IFN-γ and estimation of p*I*/M_r coordinates for proteins was done using the Melanie II software (Bio-Rad, Richmond, CA, USA).

3 Results

3.1 Pulselabeling and 2-D PAGE (IPG) of *C. trachomatis* proteins

C. trachomatis A or L2 infected HeLa cells were either treated or untreated with 100 U/mL IFN-γ at the time of infection and pulselabeled with [^{35}S]methionine/cysteine 10–12 h.p.i. and 22–24 h.p.i. Cycloheximide, an inhibitor

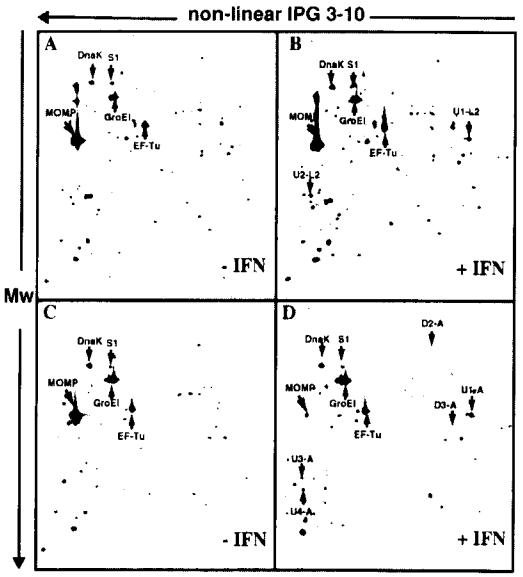


Figure 1. Two-dimensional gels (IPG 3–10 nonlinear) of [^{35}S]methionine/cysteine-labeled *C. trachomatis* serovar L2 (A and B) and serover A (C and D) proteins at 24 h.p.i., treated with (+) or without (–) IFN-γ. Arrows show the positions of the abundant chlamydial proteins MOMP, DnaK, GroEl, EF-Tu (elongation factor Tu) and the ribosomal protein S1. (B) U1-L2 and U2-L2 and (D) U1-A, U3-A, U4-A show the position of up-regulated proteins and D2-A, D3-A down-regulated proteins in the two serovars due to IFN treatment.

of eukaryotic protein synthesis, was added to prevent incorporation into HeLa cell proteins. Two sample preparations were prepared per experimental condition. Samples containing the labeled proteins from the two serovars were separated by 2-D PAGE (IPG) resulting in high-resolution 2-D gels as exemplified in Fig. 1. Incorporation of [^{35}S]methionine/cysteine into host cell proteins was estimated from 2-D gels with proteins from uninfected HeLa cells treated or untreated with IFN-γ and labeled without addition of cycloheximide. We observed eight HeLa cell proteins (three of which were the abundant proteins actin, β-tubulin, and α-tubulin) that could be detected on 2-D gels with the labeled chlamydial proteins. These were predominantly visible on 12 h.p.i. gels, presumably due to low metabolic activity of the chlamydia compared to that of the host cell at this stage (data not shown). By means of the Melanie II software we were able to detect approximately 600 labeled *C. trachomatis* L2 proteins on the gels in agreement with the number found on silver-stained gels reported by Bini *et al.* [18]. Although the two different serovars result in different diseases the overall protein pattern for the two serovars was found to be quite similar in most areas of the gels (Fig. 1A, serovar L2 and Fig. 1C, serovar A). The majority of differences were small deviations in p*I*/M_r coordinates for some of the proteins. The positions of the abundant chlamydial proteins such as GroEl, DnaK (homolog to the eukaryotic 70 kDa heat-shock protein), MOMP, elongation factor EF-Tu and the ribosomal protein S1 could be located by comparison to the 2-D PAGE (IPG) map of *C. trachomatis* L2 proteins [18] and were used as molecular markers. The cysteine-rich OMP2 cluster was only slightly detectable on the gels due to the low expression of this protein at 24 h.p.i., but could be mapped by comparison to 2-D gels with chlamydial proteins that were labeled strictly with [^{35}S]cysteine (data not shown).

3.2 Indirect immunofluorescence microscopy with mAbs against MOMP

In order to confirm that a dose of 100 U/mL IFN-γ was affecting the growth of chlamydia, we performed immunofluorescence microscopy using antibodies directed against MOMP. We did not detect any significant changes in the morphology of *C. trachomatis* L2 due to treatment with IFN-γ (Fig. 2, serovar L2). The inclusions at 24 h.p.i. displayed normal morphology of RB and EB regardless of IFN-γ treatment. In contrast, IFN-γ treatment of *C. trachomatis* A resulted in large abnormal RB, with diffuse cell walls, that stained weakly with mAb 32.2 (Fig. 2, serovar A). These results are in agreement with previously obtained results [3, 6, 15].

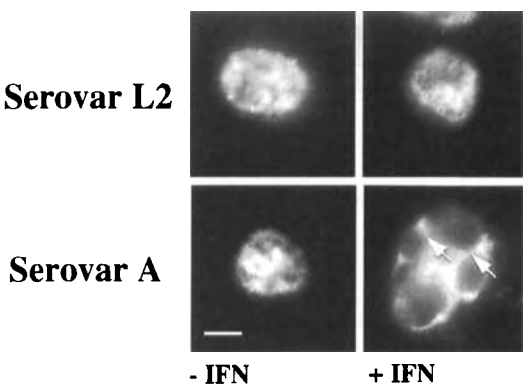


Figure 2. Indirect immunofluorescence microscopy showing the morphology of representative inclusions from *C. trachomatis* serovar L2 (upper row) and serovar A (lower row) 24 h.p.i., when treated with (–) or without (+) IFN-γ (IFN). A monoclonal antibody directed against the chlamydial MOMP was used as primary antibody. White arrows point to the borders of the atypical reticulate bodies observed in *C. trachomatis* serovar A. The white bar indicates 5 μm.

3.3 IFN-γ-dependent down-regulation of chlamydial proteins

When 2-D gels of *C. trachomatis* serovar A proteins treated or untreated with IFN-γ were compared, the expression of MOMP was markedly decreased at 24 h.p.i. (Fig. 1 D and Fig. 3, serovar A), when compared to DnaK and GroEl expression levels. However, the reduction in MOMP levels was already evident after 12 h.p.i., suggesting that IFN-γ is affecting MOMP expression early in the developmental cycle of *C. trachomatis* A (Fig. 3, serovar A). Two additional protein spots on the acidic side and in proximity to MOMP were co-regulated by IFN-γ (Fig. 3, open arrows). No significant alterations in MOMP expression due to IFN-γ treatment were found in *C. trachomatis* serovar L2, neither at 12 nor at 24 h.p.i. (Fig. 1B and Fig. 3, serovar L2).

Several other proteins displaying variability in coordinates on the gels were observed to be down-regulated upon IFN-γ treatment in addition to MOMP in *C. trachomatis* A at 24 h.p.i, most notably a 130 kDa (D2-A) and a 38 kDa (D3-A) protein (Fig. 1D and Fig. 4C, D, serovar A). In contrast to this finding the corresponding proteins in *C. trachomatis* L2 were not significantly affected by IFN-γ (Fig. 1B and Fig. 4C, D, serovar L2).

3.4 IFN-γ-dependent up-regulation of chlamydial proteins

Upon treatment with IFN-γ, a pronounced up-regulation of one protein, U1-L2, with the estimated coordinates 42 kDa/p*I* 7 was observed in *C. trachomatis* L2 at 24 h.p.i. (Fig. 1B and Fig. 4A, serovar L2). The induction of this protein was barely detectable as a faint streak on the 2-D gels of IFN-γ-untreated cells. In *C. trachomatis* A the up-regulation of U1–A was observed at the same coordinates, indicating that these proteins are identical in the two serovars (Fig. 1D and Fig. 4A, serovar A). A faint spot at the acidic side of U1-L2 and U1-A and with the same molecular mass was also up-regulated (Fig. 4A, open arrows). Whether this is another chlamydial protein or a result of a post-translational modification remains to be elucidated. In *C. trachomatis* L2 another protein U2-L2 with the estimated coordinates 29 kDa/p*I* 4.6 was strongly induced by IFN-γ at 24 h.p.i. (Fig. 1B and Fig. 4B, serovar L2). Two proteins, U3-A and U4-A, within the same p*I* region, but in a molecular mass region ~5 kDa lower than observed for U2-L2, were found up-regulated in serovar A (Fig. 1D and Fig. 4B, serovar A). We did not detect any up-regulation of either U1-L2, U2-L2, U1-A, U3-A or U4-A at 12 h.p.i. and IFN-γ treatment (data not shown).

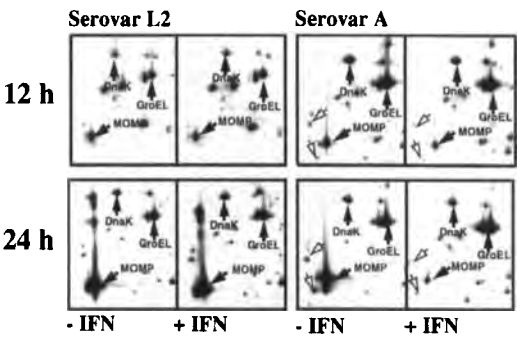


Figure 3. Enlargements from the total 2-DE images in Fig. 1 showing changes in expression of the chlamydia MOMP compared to chlamydia GroEl and DnaK in *C. trachomatis* serovar L2 (left) and serovar A (right) at 12 h.p.i. (upper row) or at 24 h.p.i. (lower row) with (+) or without (–) treatment of IFN-γ (IFN). Note that the down-regulation of MOMP due to IFN is evident already at 12 h.p.i. infection in serovar A, in contrast to serovar L2, where no significant regulation is seen, neither at 12 h.p.i. nor at 24 h.p.i.

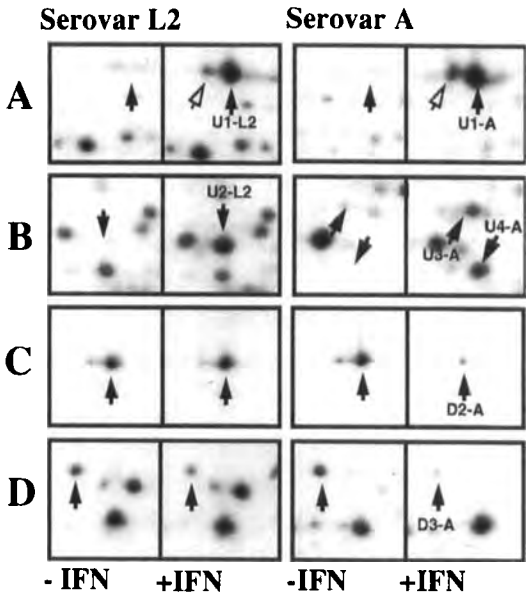


Figure 4. Enlargements from the total 2-DE images in Fig. 1 showing up-regulated (A and B) and down-regulated (C and D) chlamydial proteins in serovar L2 (left) and serovar A (right) upon IFN-γ treatment. (A) U1-L2 and U1-A and a small protein to the acidic side (open arrows) were located at the same positions on the gels from the two serovars and most likely represents the same protein. (B) Although having a similar p*I*, U2-L2 in serovar L2 was located in a molecular mass region ~ 5 kDa above U3-A and U4-A in serovar A. (C, D) The down-regulation of D2-A and D3-A is only significant in serovar A.

Figure 5. The antagonizing effect of 200 μg/mL exogenously added L-tryptophan on IFN-γ-mediated up-regulation of (A) U1-A, (B) U4-A in *C. trachomatis* A and (C) U2-L2 in *C. trachomatis* L2. Arrows indicate positions of up-regulated proteins upon treatment with IFN-γ alone (see Fig. 4).

The map of *C. trachomatis* L2 is, at the present, not detailed enough to find the identity of these proteins. However, the total genome of *C. trachomatis* on the NCBI Entrez server was recently released [19]. In order to search for potential protein candidates for the up-regulated proteins, we used the p*I*/M_w tool [20] from the ExPASy server to investigate the molecular coordinates for translated gene products from the *C. trachomatis* genome in the region 42 ± 4kDa/p*I* ± 0.3 in proximity to U1-L2 and U1-A and in the region 29 ± 4 kDa/p*I* 4.6 ± 0.3 in proximity to U2-L2. Interestingly, these investigations showed that tryptophan synthase β-chain had the theoretical coordinates 42.6 kDa/p*I* 6.79, close to those observed for U1-L2 and U1-A. The tryptophan synthase α-chain had the theoretical coordinates 28 kDa/p*I* 4.8, resembling those observed for U2-L2. If the up-regulation of tryptophan synthesizing apparatus is a chlamydial response to IFN-γ induced tryptophan-depletion, then superphysiological concentrations of tryptophan in the growth medium should prevent this. As depicted in Fig. 5, the up-regulation of U1-A/L2 and U2-L2 and U4-A is antagonized at 24 h.p.i. if 200 μg/mL L-tryptophan is added together with IFN-γ to the growth medium.

4 Discussion

Investigations by pulselabeling and 2-D PAGE (IPG) indicate that the expression of several chlamydial proteins, in addition to those previously described, is influenced upon treatment with IFN-γ. Our results are in agreement with the investigations done by Beatty *et al.* [3, 6, 10] showing the correlation between IFN-γ treatment and the development of abnormal RB and MOMP expression in *C. trachomatis* serovar A. In contrast to our investigations, the effect of IFN-γ was not monitored at 24 h.p.i. or earlier in these experiments. Takikawa *et al.* [21] found that IDO expression induced by IFN-γ was not evident before 24 h of treatment in HeLa cells even when using a dose of 1000 U/mL IFN-γ. Interestingly, we did observe an evident down-regulation of MOMP already at 12 h.p.i., indicating that tryptophan catabolism alone may not be enough to diminish the expression of MOMP. The two proteins in close proximity to MOMP were coregulated, but whether these were fragments or modifications of MOMP was not determined (Fig. 3, serovar A, open arrows). Several proteins in addition to MOMP were down-regulated in *C. trachomatis* A. Two of these, D1-A and D2-A, were not affected before 24 h.p.i., suggesting that the host cell protein responsible is not significantly induced by IFN-γ before 24 h. In contrast, we did not observe any significant down-regulation of serovar L2 proteins, including MOMP, and no changes in the morphology of RB. The last finding supports results from similar morphological studies obtained by Rasmussen *et al.* [15] showing merely a decrease in the size of serovar L2 inclusions due to IFN-γ at 24 h.p.i.

We report here the up-regulation of chlamydial proteins due to IFN-γ. One ~ 42 kDa/~ p*I* 7 protein is highly up-regulated in both serovars (U1-L2 in *C. trachomatis* L2 and U1-A in *C. trachomatis* A), suggesting that this is a general chlamydial response towards IFN-γ. In addition, U2-L2 in *C. trachomatis* L2 and U3-A and U4-A in *C. trachomatis* A were also up-regulated. U3-A and U4-A in *C. trachomatis* A migrated to an area ~5 kDa lower than U2-L2 in serovar L2. The protein pattern in this region differs between the two serovars, making them difficult to compare. Therefore it cannot be excluded that at least one of the *C. trachomatis* A proteins corresponds to U2-L2.

The *C. trachomatis* genome, used to find potential protein candidates, was established using *C. trachomatis* serovar D. However, restriction endonuclease analysis shows close similarity between the different human *C. trachomatis* biovars and serovars [22]. In view of the tryptophan depletion of the host cell induced by IFN-γ, the tryptophan synthase α- and β-chains must be considered as potential candidates for the up-regulated *C. trachomatis* A and L2 proteins. These two subunits form the functional tryptophan synthase and have distinct theoretical p*I*/M_r coordinates, which, in turn, resemble those found for the up-regulated *C. trachomatis* proteins on our 2-D gels. In addition, none of the up-regulated proteins was detected at 12 h.p.i. (data not shown), when the induction of IDO was not reported significant in HeLa cells [21]. The host cells tryptophanyl-tRNA synthetase is up-regulated in a fashion similar to IDO in many human cell lines [23]. Tryptophanylated host cell tRNA is not usable for prokaryotes, and some controversy exists as to whether this provides a means of ensuring continued tryptophan for host cell protein synthesis during the battle against the intracellular parasite [23–25].

Taken together, our findings strongly indicate that the chlamydia may up-regulate its own enzymes necessary

for tryptophan synthesis in order to circumvent the tryptophan depletion caused by increased expression of IDO in the host cell. This is supported by the observation that the IFN-γ-dependent up-regulation of these proteins is prevented by addition of superphysiological amounts of L-tryptophan. Exogenously added L-tryptophan antagonizes the up-regulation of U4-A but not U3-A, suggesting that U4-A in *C. trachomatis* A might correspond to a low molecular mass form of U2-L2 in *C. trachomatis* L2.

C. trachomatis L2 is recognized as being more invasive than *C. trachomatis* A. Whether the missing down-regulation events seen in *C. trachomatis* L2 are associated with the lower sensitivity towards the cellular immune response remains to be elucidated. However, if the lower molecular mass of U4-A is due to a truncated version of the tryptophan synthase α-chain, the generation of a functional enzyme in *C. trachomatis* A could be impaired, resulting in a higher sensitivity towards IFN-γ treatment compared to *C. trachomatis* L2.

The application of 2-D PAGE (IPG) in the analysis of changes in chlamydial protein expression due to IFN-γ has led to the novel findings of regulated chlamydial proteins. Identification of these proteins by MALDI MS or *N*-sequencing will elucidate their possible functional roles in evading the inhibiting effects of the IFN-γ-mediated host cell response.

We are grateful to Karin Skovgaard Sørensen, Inger Andersen and Lisbet Wellejus Pedersen for technical assistance. The study was supported financially by the Danish Health Research Council (grants 12-0850-1, 12-0150-1), the Danish Veterinary and Agricultural Research Council (grants 20-3503-1), Aarhus University Research Foundation, 'Nationalforeningen til Bekæmpelse of Lungesygdomme', the Velux Foundation, and The Danish Pasteur Society.

Received October 20, 1998

5 References

[1] Shacter, J., *Curr. Topics Microbiol. Immunol.* 1988, *138*, 109–139.

[2] Shemer, Y., Sarov, I., *Infect. Immun.* 1985, *48*, 592–596.

[3] Beatty, W. L., Byrne, G. I., Morrison, R. P., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 3998–4002.

[4] Byrne, G. I., Lehman, L. K., Landry, G. J., *Infect. Immun.* 1986, *53*, 347–351.

[5] Shemer, Y., Kol, R., Sarov, I., *Curr. Microbiol.* 1987, *16*, 9–13.

[6] Beatty, W. L., Belanger, T. A., Morrison, R. P., Byrne, G. I., *Infect.Immun.* 1994, *62*, 3705–3711.

[7] Summersgill, J. T., Sahney, N. N., Gaydos, C. A., Quinn, T. C., Ramirez, J. A., *Infect. Immun.* 1995, *63*, 23801–2803.

[8] Rapoza, P. A., Tahija, S. G., Carlin, J. P., Miller, S. L., Padilla, M. L., Byrne, G. I., *Invest. Ophthalmol. Vis. Sci.* 1991, *32*, 2919–2923.

[9] Methal, S. J., Miller, R. D., Ramirez, J. A., Summersgill, J. T., *J. Infect. Dis.* 1998, *177*, 1326–1331.

[10] Beatty, W. L., Morrison, R. P., Byrne, G. I., *Infect. Immun.* 1995, *63*, 199–205.

[11] Feng, G. S., Taylor, M. W., *FASEB J.* 1991, *5*, 2516–2522.

[12] Carlin, J. M., Borden, E. C., Sondel, P. M., Byrne, G. I., *J. Interferon Res.* 1989, *9*, 329–337.

[13] Thomas, S. M., Garrity, L. F., Brandt, C. R., Scorbert, C. S., Feng, G. S., Taylor, M. W., Carlin, J. M., Byrne, G. I., *J. Immunol.* 1993, *150*, 5529–5534.

[14] Gupta, S. L., Carlin, J. M., Pyati, P., Dai, W., Pfefferkorn, E. R., Martin, J., Murphy, J. R., *Infect. Immun.* 1994, *62*, 2277–2284.

[15] Rasmussen, S. J., Timms, P., Beatty, R. P., Stephens, R. S., *Infect. Immun.* 1996, *64*, 1944–1949.

[16] de la Maza, L. M., Peterson, E. M., Fennie, C. W., Czarniecki, C. W., *J. Immunol.* 1985, *135*, 4198–4200.

[17] Ripa, T. K., *Scand. J. Infect. Dis. Suppl.* 1982, *32*, 25–29.

[18] Bini, L., Sanchez-Campillo, M., Santucci, A., Magi, B., Marzocchi, B., Comanducci, M., Christiansen, G., Birkelund, S., Cevenini, R., Vretou, E., Ratti, G., Pallini, V., *Electrophoresis* 1996, *17*, 185–190.

[19] Stephens, R. S., Kalman, S., Lammel, C. J., Fan, J., Marathe, R., Aravind, L., Mitchell, W. P., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V., Davis, R. W., *Science* 1998, in press.

[20] Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1023–1031.

[21] Takikawa, O., Kuroiwa, T., Yamazaki, F., Kido, R., *J. Biol. Chem.* 1988, *263*, 2041–2048.

[22] Peterson, E. M., de la Maza, L. M., *J. Clin. Microbiol.* 1988, *26*, 625–629.

[23] Flecner, J., Mortensen, P. M., Tolstrup, A. B., Kjeldgaard, N. O., Justesen, J., *Cytokine* 1995, *7*, 70–77.

[24] Xue, H., Wong, J. T., *Gene* 1995, *165*, 335–339.

[25] Flohr, T., Bange, F. C., Kiekenbeck, M., Bottger, E. C., *Infect. Immun.* 1992, *60*, 4418–4421.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 781–789

Dirk Benndorf
Norbert Loffhagen
Wolfgang Babel

Umweltforschungszentrum
Leipzig-Halle GmbH,
Sektion Umweltmikrobiologie,
Leipzig, Germany

Induction of heat shock proteins in response to primary alcohols in *Acinetobacter calcoaceticus*

Cells of *Acinetobacter calcoaceticus* 69-V, a species able to metabolize a range of aliphatic hydrocarbons and alcohols, were confronted with ethanol, butanol, hexanol or heat shock during growth on acetate as sole source of carbon and energy. The primary alcohols and the heat shock led to the synthesis of new proteins or amplified expression of specific, common and general proteins, which were detected by silver staining after two-dimensional gel electrophoresis. Some of the alcohol-inducible proteins were identified as heat shock proteins by comparing protein patterns of alcohol-shocked cells with those of heat-shocked cells, and by *N*-terminal amino acid sequencing. DnaK was found to be amplified after all treatments, but GroEl only after heat shock and ethanol treatment. The *N*-terminal amino acid sequence of the protein, which was considerably amplified after alcohol treatment and heat shock, shows homology to HtpG (high temperature protein G). Some of the heat shock proteins induced by ethanol differ from those induced by butanol and hexanol, suggesting there are at least two different signals for the induction of some heat shock proteins by primary alcohols. This could be due to the different localization of ethanol, butanol and hexanol in the membrane, or because higher cytoplasmic concentrations of ethanol than of butanol or hexanol were applied in these tests in order to keep concentrations of the alcohols in the membrane roughly similar. Besides heat shock proteins, a group of proteins were observed which were only induced by butanol and hexanol, possibly indicating the existence of a further defense mechanism against high concentrations of hydrophobic substrates preventing protein denaturation and membrane damage.

Keywords: Heat shock response / Primary alcohols / Membrane / Heat shock proteins GroEl / DnaK / High temperature protein

EL 3367

1 Introduction

Bacteria respond to changing environmental conditions with the synthesis of specific proteins. The most studied mechanism is the heat shock response. The heat shock proteins, *e.g.* DnaK and GroEl, play an important role during "normal" growth of bacteria [1], as well as during stress conditions [2]. Their sequences are highly conserved. In analogy to a temperature increase, a great variety of chemicals (*e.g.*, ethanol, antibiotics and heavy metals) were found to induce heat shock proteins as well [3–5]. Changes in the membrane physical state caused by heat or lipophilic chemicals [6, 7] were discussed as further mechanisms for the induction of the heat shock proteins, besides the classical theories involving the accumulation of denaturated proteins or the function of DnaK as an intracellular thermometer [8]. Ethanol has been known to induce heat shock proteins for a long time. Surprisingly, the effect of higher primary alcohols (the lipophilicity of which increases with carbon number) on protein synthesis of bacteria has not been investigated. Most results showing induction of stress proteins after treatment with other lipophilic chemicals were obtained with organisms which were not able to utilize them as sole sources of carbon and energy. Therefore, we chose to study stress protein induction by higher alcohols in *Acinetobacter calcoaceticus* [9]. This species, which can easily be isolated from soil, water, or sewage [10], is known to be able to utilize a number of aliphatic hydrocarbons and corresponding alcohols, adapting to growth on these substrates by changing the composition of its membrane fatty acids [11].

Correspondence: Prof. W. Babel, Umweltforschungszentrum Leipzig-Halle GmbH, Sektion Umweltmikrobiologie, PF2, 04301 Leipzig, Germany
E-mail: babel@umb.ufz.de
Fax: +49-341-2247

2 Materials and methods

2.1 Bacterial strain, culture conditions, exposure to stress

Acinetobacter calcoaceticus 69-V was grown at 30 °C in 100 mL shake flasks in a minimal medium [12] without yeast extract containing 5 g/L sodium acetate as sole source of carbon and energy. Growth was measured spectrophotometrically by monitoring the optical density at 700 nm. When the culture reached the early exponen-

Table 1. Influence of primary alcohols and heat shock on growth of *Acinetobacter calcoaceticus* and the induction ratio of DnaK

	Added alcohol (%)	Concentration water [M]	Concentration membrane [M]	Growth rate (%) of control	Induction ratio of DnaK
	2.5	0.543	0.07	68	1.4
Ethanol	5	1.087	0.13	41	1.9
	7.5	1.630	0.20	No growth	1.4
	0.8	0.108	0.18	58	2.1
Butanol	1.4	0.189	0.31	12	0.4
	2	0.270	0.44	No growth	0.4
	0.1	0.010	0.15	39	1.5
Hexanol	0.15	0.015	0.22	No growth	1.4
	0.2	0.020	0.29	No growth	0.5
Heat shock	–	–	–	22	2.6

One μg of protein per lane was separated by SDS-PAGE. DnaK was identified on the silver-stained gel (gel image not shown) by molecular weight. The bands were quantified using the Intelligent Quantifier Software (Bio Image). Membrane concentration of primary alcohols was calculated as described by Hahn *et al.* [18].

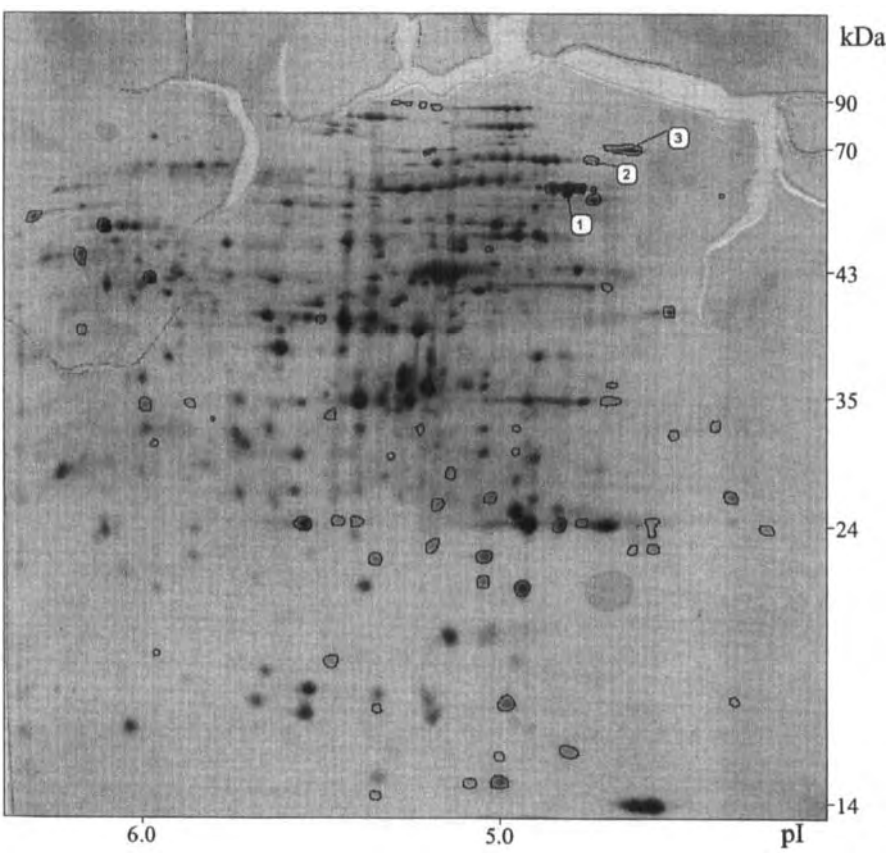


Figure 1. 2-D gel of silver-stained proteins of *Acinetobacter calcoaceticus* grown on sodium acetate. Spots bordered with a black line represent amplified proteins after various stresses. Numbered spots: 1, GroEl; 2, HtpG; 3, DnaK.

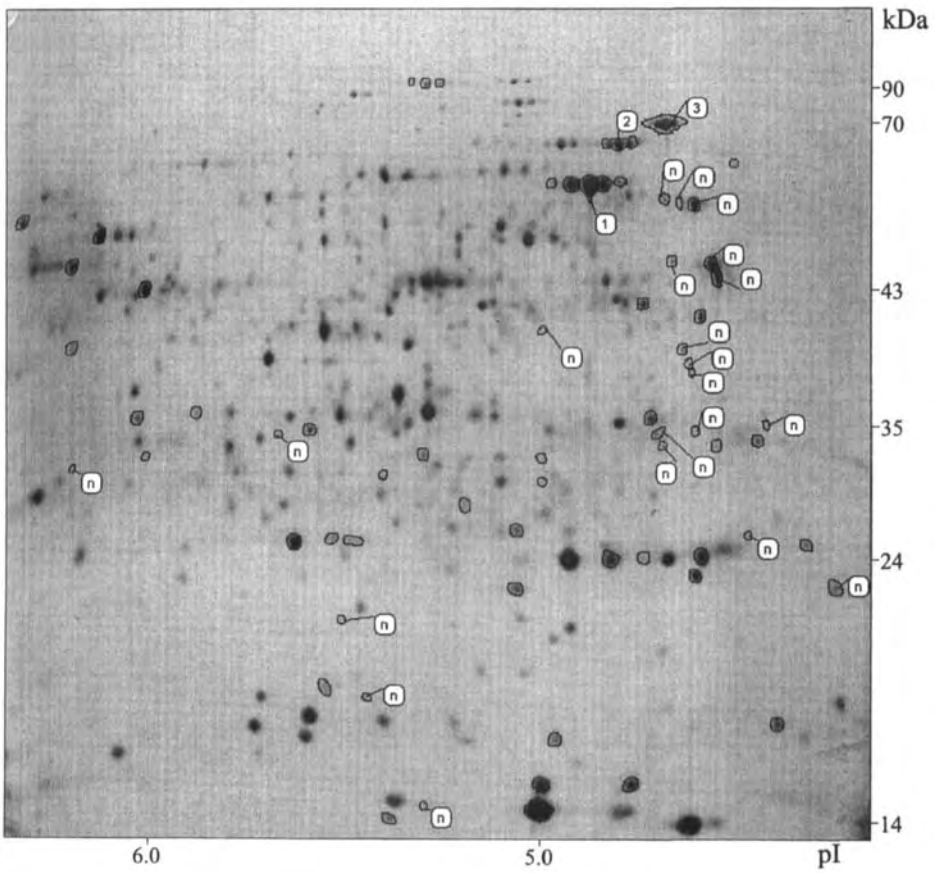


Figure 2. 2-D gel of silver-stained proteins of *Acinetobacter calcoaceticus* grown on sodium acetate 1 h after a temperature shift from 30 °C to 45 °C. Spots bordered with a black line represent amplified proteins. Spots marked with "n" represent new proteins. Numbered spots: 1, GroEl; 2, HtpG; 3, DnaK.

tial phase, the culture was divided and exposed for 1 h to 45 °C heat shock or one of the following reagents: 2.5–7.5% ethanol, 0.8–2% butanol, or 0.1–0.2% hexanol.

2.2 Sample preparation and 2-DE

The bacteria were harvested by centrifugation for 10 min at 8000 × *g* (at 4 °C). They were washed twice with buffer containing 50 mM Tris-HCl, pH 7.5, 0.1 mg/mL chloramphenicol and 1 mM PMSF, and resuspended in 0.4 mL lysis solution containing 50 mM Tris-HCl, pH 7.5, 1 mM PMSF, 0.5 mg/mL $MgCl_2$ and 1.2 μL/mL benzonase (Merck, Darmstadt, Germany). To prepare cell extracts, the cells were disrupted on ice by sonication (Branson Sonifier 250). Unbroken cells and cell debris were removed by centrifuging at 8000 × *g* for 10 min. Cell-free extracts were stored at −18 °C. The protein content was determined using Bradford's method [13]. Cell-free extracts containing 60 μg protein were precipitated by addition of 1 mL ice-cold acetone, incubated for 20 min on ice and centrifuged at 12 000 × *g* for 10 min. The precipitated proteins were resuspended in 50 μL of a solution containing 10 mM Tris-HCl, pH 6.8, 2% mercaptoethanol and 1% Nonidet P-40, and heated for 5 min at 95 °C. Then 350 μL of a resolubilizing solution was added, containing 8 M urea, 0.5% Nonidet P-40, 1.5% CHAPS, 1% IPG buffer, pH 4–7 L (Pharmacia, Uppsala, Sweden), 0.2% dithioerythritol, 1 mM PMSF and 0.002% bromophenol blue. The complete solution was loaded on an 18 cm long Immobiline DryStrip, pH 4–7 L, for equilibration overnight. The proteins were separated by isoelectric focusing using the Immobiline DryStrip Kit (Pharmacia) and a Multiphore II electrophoresis unit (Pharmacia). The gels were run for a total of 71 750 Vh, with the voltage increasing linearly

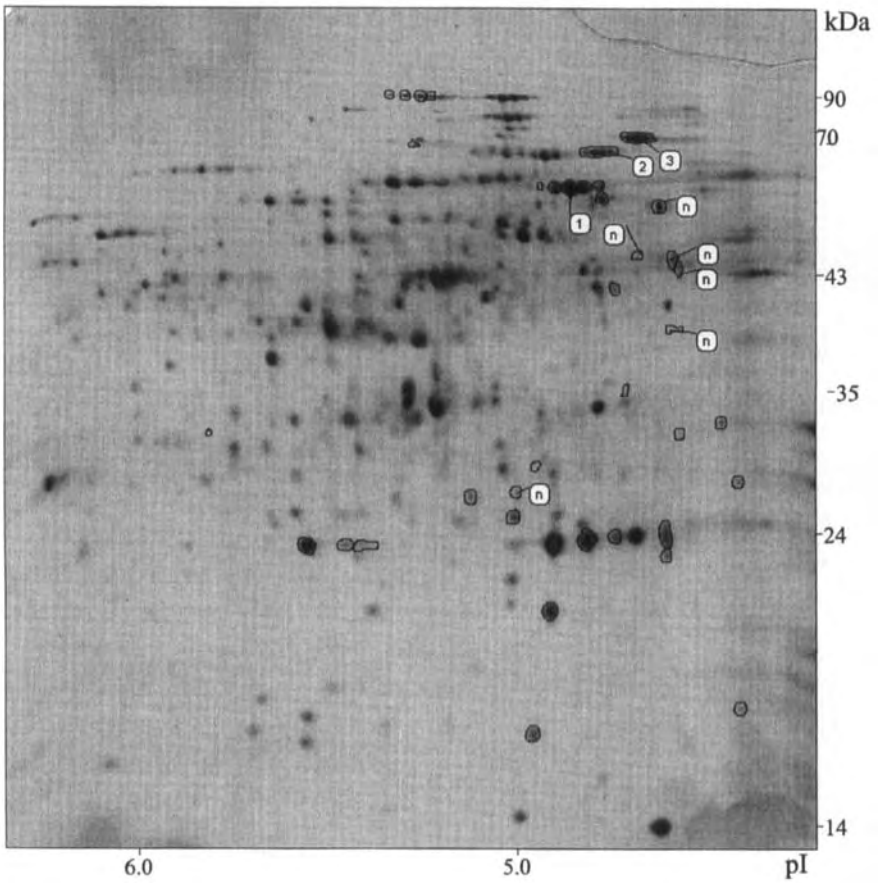


Figure 3. 2-D gel of silver-stained proteins of *Acinetobacter calcoaceticus* grown on sodium acetate 1 h after addition of 5% ethanol. Spots bordered with a black line represent amplified proteins. Spots marked with "n" represent new proteins. Numbered spots: 1, GroEl; 2, HtpG; 3, DnaK.

from 500 to 3500 V during the first 5 h and then maintained for 15.5 h at 3500 V. After the isoelectric focusing, the gels were equilibrated by gentle orbital shaking, first in a solution containing 6 M urea, 30% glycerol, 4% SDS, 2% DTT and 0.05 M Tris-HCl, pH 6.8, for 15 min and then in a solution containing 6 M urea, 30% glycerol, 4% SDS, 2.5% iodoacetamide, 0.05 M Tris-HCl, pH 6.8, and a trace of bromophenol blue for 10 min. The equilibrated gels were stored at –18 °C until used in second-dimensional electrophoresis. For the second dimension, the gels were loaded onto a layer of buffer containing 0.125 M Tris-HCl, pH 6.8, 0.1% SDS and 0.5% w/v agarose adjusted to 80 °C on 12%T, 1%C polyacrylamide slab gels (200 × 183 × 1 mm) without a stacking gel. The gels were run in a Protean II Multi-Cell (Bio-Rad, Richmond, CA, USA) instrument, first at a current of 20 mA/gel for 1 h and then at 40 mA/gel for 4 h and a temperature of 12 °C. After completion of SDS-PAGE, the gels were fixed and silver-stained as described by Blum *et al.* [14]. After staining, the gels were dried in a stream of unheated air from a Gel Air Dryer (Bio-Rad).

2.3 Comparison of 2-D protein patterns

The air-dried gels were scanned with a UMAX Power Look 2000 scanner with a resolution of 127 µm and a dynamic range of eight bits. The gels were analyzed and compared by the Phoretix 2D Full and Phoretix 2D Database Software programs (NonLinear Dynamics Ltd.). Molecular weight and p*I* calibration was done using internal standards from a 2D SDS-PAGE Standards Kit (Bio-Rad). Spots with a twofold, or greater, volume than the

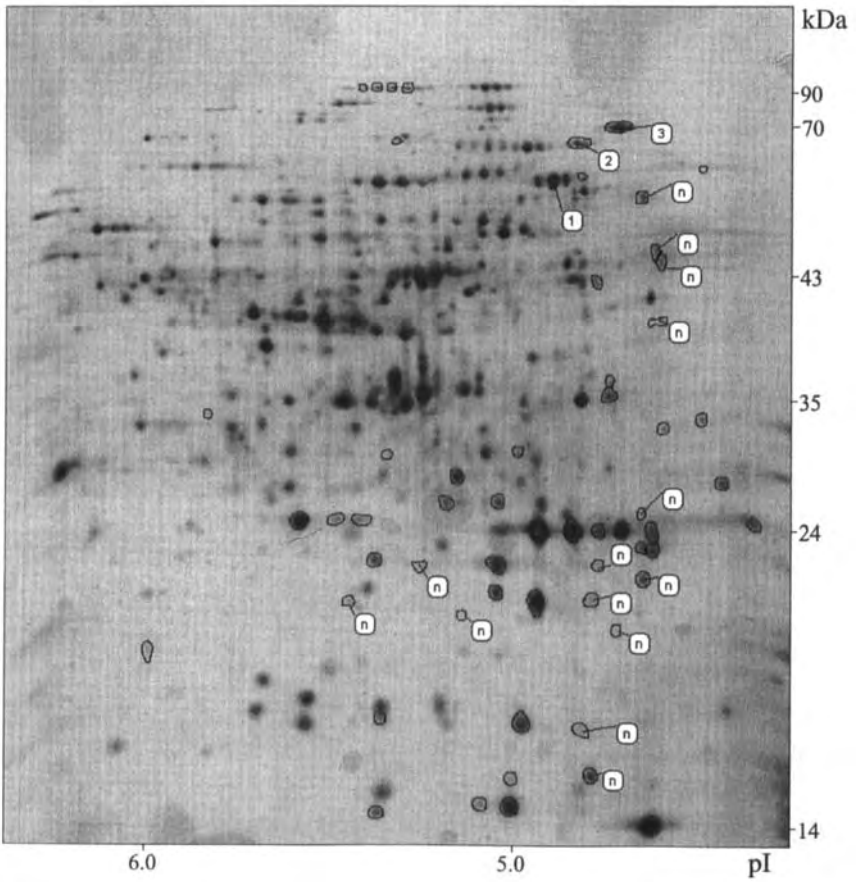


Figure 4. 2-D gel of silver-stained proteins. of *Acinetobacter calcoaceticus* grown on sodium acetate 1 h after addition of 0.8% butanol. Spots bordered with a black line represent amplified proteins. Spots marked with "n" represent new proteins. Numbered spots: 1, GroEl; 2, HtpG; 3, DnaK.

corresponding spot in control gels were considered to be amplified.

2.4 Micropreparative 2-D PAGE, electroblotting and *N*-terminal sequence analysis

Cell-free extracts containing 500 μg of protein were precipitated by addition of 1 mL ice-cold acetone, incubated for 20 min on ice and centrifuged at 12 000 × *g* for 10 min. The precipitate was resuspended in 50 μL of a solution containing 8.3 M urea, 3.6% CHAPS, 1.2 mg PMSF, 3.6% Bio-Lyte 5–7 carrier ampholytes 0.4 Bio-Lyte 3–10 carrier ampholytes, 4% Triton X-100, and 2% mercaptoethanol. The protein was loaded at the cathodic end of IEF tube gels (5 × 125 mm) with 3.5% acrylamide containing 8.1 M urea, 4.5% Bio-Lyte 5–7 carrier ampholytes, 2% Bio-Lyte 3–10 carrier ampholytes, and 2% Triton X-100. The cathodic electrode solution was 0.02 M NaOH and the anodic electrode solution was 0.06% H_3PO_4. The gels were run for a total of 9600 Vh, first with a voltage of 400 V for 16 h and then with 800 V for 4 h. After first-dimensional separation, the gels were equilibrated for 45 min in a solution containing 0.0625 M Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 5% mercaptoethanol and 0.005% bromophenol blue. The tube gels were transferred through a buffer containing 0.125 M Tris-HCl, pH 6.8, 0.1% SDS and 0.5% w/v agarose at a temperature of 80 °C onto 12%T, 1%C slab gels (200 × 160 × 3 mm) with 4%T stacking gels. Then 0.1 mM thioglycolate was added to the upper running buffer to prevent *N*-terminal blocking by reactive compounds left in the gel. The gels were run with a constant current of 80 mA/gel for 7 h. The whole

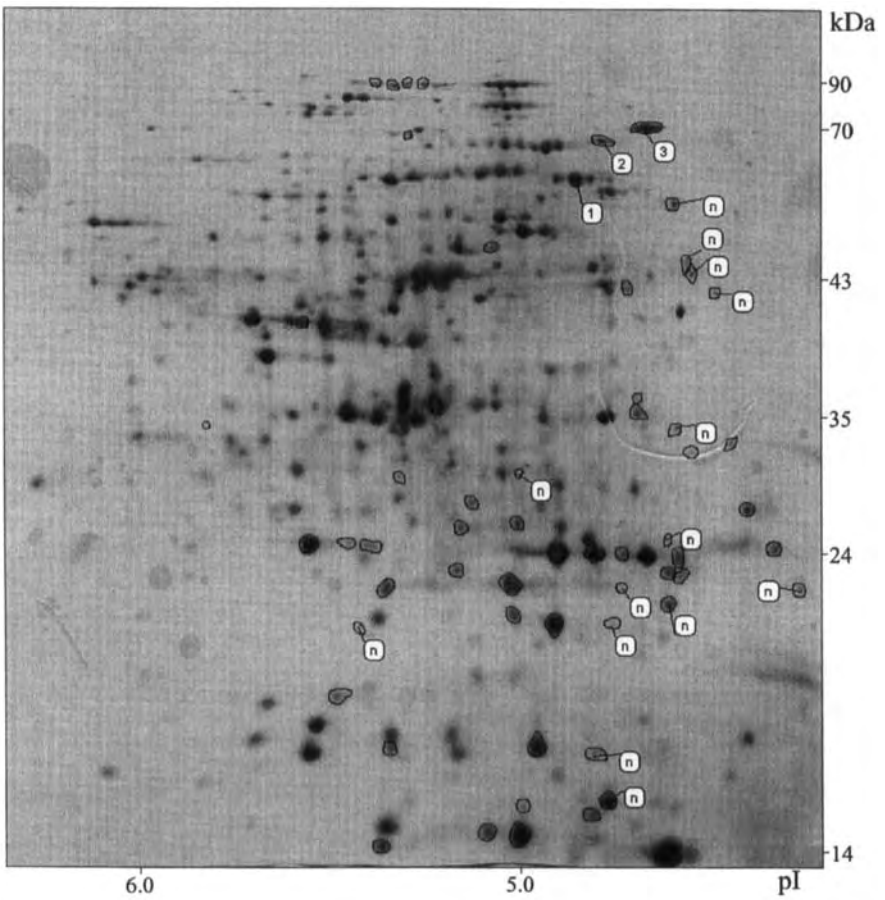


Figure 5. 2-D gel of silver-stained proteins of *Acinetobacter calcoaceticus* grown on sodium acetate 1 h after addition of 0.1% hexanol. Spots bordered with a black line represent amplified proteins. Spots marked with "n" represent new proteins. Numbered spots: 1, GroEl; 2, HtpG; 3, DnaK.

gels were electroblotted onto PVDF membranes (Trans-Blot Transfer Medium, Bio-Rad) using a Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) and 10 mM CAPS, pH 11, with 10% methanol electrode buffer as described by Jin and Cerletti [15]. The electroblotting lasted 2 h with a current of 2 mA/cm^2. The blots were stained with 0.025% w/v Coomassie Blue R-250 in a solution of 40% methanol for 10 min and destained with a solution of 50% methanol. Membranes were dried and stored at −18 °C. Spots of interest were excised and submitted to *N*-terminal sequencing using a model 473A protein sequencer (Applied Biosystems, Foster City, CA, USA). Identity searches of the amino acid sequence data obtained, in relation to proteins in the SWISS-PROT database, were performed with BLAST [16].

3 Results

3.1 Influence of primary alcohols and heat shock on growth and the induction of DnaK

After heat shock from 30 °C to 45 °C (the maximum growth temperature of *Acinetobacter calcoaceticus*) and treatment with several concentrations of primary alcohols, the growth rates of *Acinetobacter calcoaceticus* were decreased (Table 1). The concentrations of alcohols in the medium, which completely inhibited the growth, were 7.5% ethanol, 2% butanol, and 0.15% hexanol. The induction of one protein, which was identified as the heat shock protein DnaK (for identification of DnaK see Section 3.2), was already detectable by SDS-PAGE. Con-

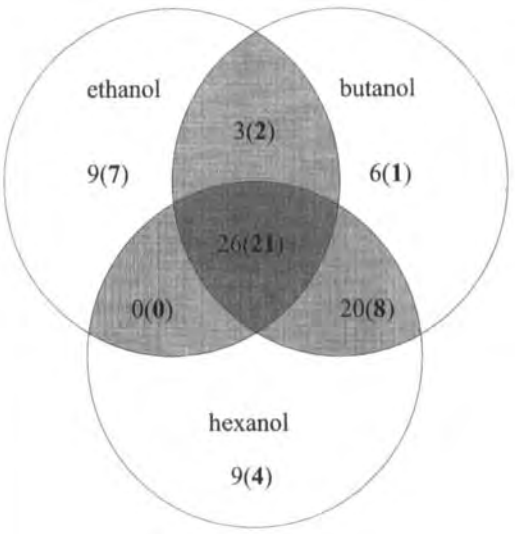


Figure 6. Induction pattern of the alcohol-inducible proteins of *Acinetobacter calcoaceticus*. Numbers in white areas represents the proteins induced only by one of the primary alcohols. Numbers in areas with grey backgrounds represent proteins that were induced by two or more alcohols. The second bold numbers written in brackets shows how many of these proteins are also induced after heat treatment.

centrations of primary alcohols, which clearly inhibit (although only incompletely) the growth of *Acinetobacter calcoaceticus* caused the highest induction ratios of DnaK (Table 1).

3.2 Comparison of 2-D gels

Acinetobacter calcoaceticus growing on acetate was treated with 5% ethanol, 0.8% butanol, 0.1% hexanol, which are the appropriate concentrations of primary alcohols for the induction of DnaK (Table 1), and heat; a comparison of the 2-D gels resulted in the detection of 91 proteins, which were newly synthesized or were amplified at least twofold (Figs. 1–5). Twenty-one proteins are generally induced by all kinds of stresses (Fig. 6). The number of proteins induced specifically by each of the primary alcohols is small; after heat shock, however, 25 proteins were induced specifically (Table 2). Nevertheless, the high number of general proteins indicate that there should be a general response mechanism. Apart from the general proteins there are two different groups of alcohol-inducible proteins which are also induced by heat shock. The first group consists of seven proteins (Fig. 6), which were only induced by heat shock or ethanol, and the second group consists of eight proteins which were only induced by heat shock or treatment with butanol or hexanol. A comparison of the induction pattern of the three primary alcohols (Fig. 6) shows that there is less similarity in the patterns induced by butanol and ethanol (three common proteins) or in the patterns induced by hexanol and ethanol (no common protein). However, there is considerably more similarity between the butanol and hexanol patterns, since 20 proteins are commonly induced by both stresses.

3.3 Microsequencing

Proteins 1 and 3 (Figs. 1–5) were identified as the heat shock proteins DnaK and GroEl by *N*-terminal amino acid sequencing and identity searches in the SWISS-PROT database (Table 3). The sequences of the two proteins have a relatively low similarity to sequences from a halophilic *Acinetobacter*, which were supplied by Tokunaga *et al.* [16]. The proteins most strongly homologous to the DnaK and GroEl of *Acinetobacter calcoaceticus* 69-V are the DnaK of *Legionella pneumophila* and the GroEl of *Pseudomonas putida*. Protein 2 has been identified as the heat shock protein HtpG by a similarity of 46% to the HtpG of *Helicobacter pylori* and by similarity of 53% to the Hsp82 of *Saccharomyces cerevisiae* (Table 3).

4 Discussion

Both heat and primary alcohols led to the induction of heat shock proteins. Although the primary alcohol con-

Table 2. Number of amplified and new protein of *Acinetobacter calcoaceticus* after heat shock and treatment with primary alcohols

	Total number of amplified/ new proteins	Number of specific amplified/ specific new proteins
Heat shock	47/21	11/14
Ethanol stress	32/6	1/1
Butanol stress	41/15	1/4
Hexanol stress	41/14	2/3

More than 70% of the new or amplified proteins after each treatment were also found in a second experiment

Table 3. Results of *N*-terminal sequence analysis of proteins of *Acinetobacter calcoaceticus* and of the identity search in the SWISS-PROT database

Spot	Protein; accession number	M_r (kDa); p*I*	*Species*		% Identity
1	DnaK		*Acinetobacter 204-I*	1 G R I I G I D L G T T N S D V A V L D G 20	60
	DnaK	71; 4.6	*Acinetobacter calcoaceticus 69-V*	1 A K I I G I D G L T T N S W V A V L E S D K V H V 25	
	DnaK; O32482		*Legionella pneumophila*	2 A K I I G I D L G T T N S C V A V M E G D K P K V 26	72
	DnaK; P04475		*Escherichia coli*	3 K I I G I D L G T T N S C V A I M D 20	66
2	HtpG; P56116		*Heliobacter pylori*	2 S N Q E Y T F Q T E I N Q 14	46
	HtpG	70; 4.75	*Acinetobacter calcoaceticus 69-V*	1 S E X A S Q N Y S F Q A E V A Q 14	
	Hsp82; P02829		*Saccharomyces cerevisiae*	3 A S E T F E F Q A E I T Q 14	53
3	GroEL		*Acinetobacter 204-I*	1 A A K Q V K F S X D A R T R M A K G V D 20	55
	GroEl	59; 4.82	*Acinetobacter calcoaceticus 69-V*	1 S A K D V K F G D S A R S M M I A G V N V I A D 24	
	GroEl; P48216		*Pseudomonas putida*	2 A A K D V K F G D S A R K K M L V G V N V L A D 25	75
	GroEl; P06139		*Escherichia coli*	2 A A K D V K F G N D A R V K M L R G V N V L A D 25	66

The one-letter code for amino acids was used. X represents an unidentified amino acid.

Table 4. Induction ratio of proteins 1–3 (Figs. 1–5) after heat shock and treatment with primary alcohols (mean of two different experiments)

Spot	Protein	Induction ratio Heat shock	Ethanol	Butanol	Hexanol
1	GroEl	3.2	2.4	1.0	0.8
2	HtpG	15.5	13.2	4.0	3.3
3	DnaK	5.3	5.1	2.2	2.0

centrations were adjusted so that the membrane concentrations of the primary alcohols would be nearly the same, the response to ethanol differs from the response to butanol or hexanol. The facts that DnaK and HtpG were amplified after all stresses, and that GroEl was only amplified after heat shock and ethanol treatment (Table 4), indicates that the mechanism of induction of GroEl may be different from the mechanisms of the induction of two other proteins. One reason could be that ethanol accumulates in a different region of the membrane lipid bilayer to butanol and hexanol [19]. In contrast to the equal concentration of primary alcohols added to the membrane in these tests, their concentrations in the cytoplasm, considered as an aqueous phase, should correspond to the (very different) concentrations added to the culture medium (1 M ethanol, 0.1 M butanol, and 0.01 M hexanol). The much larger concentration of alcohol in the aqueous phases, in the ethanol treatment, may have caused a stronger induction of the heat shock response, *via* accumulation of denaturated proteins in the cytoplasm. Nevertheless, we could show that heat shock proteins potentially help the bacteria adapt to butanol and hexanol. The great number of proteins induced only by these stressors, which are not identical with proteins strongly induced after growth on these alcohols (unpublished data), indicates that the resistance against the effect of higher primary alcohols is not caused by enzymes responsible for the degradation of these alcohols.

We would like to thank Dr. J. Bär (Institut für Biochemie, Universitätsklinikum, Universität Leipzig) for the determination of N-terminal amino acid sequences.

Received September 1, 1998

5 References

[1] Bukau, B., Walker, G. C., *J. Bacteriol.* 1989, *171*, 2337–2346.

[2] Paek, K.-H., Walker, G. C., *J. Bacteriol.* 1987, *169*, 283–290.

[3] Van Bogelen, R. A., Kelley, P. M., Neidhardt, F. C., *J. Bacteriol.* 1987, *169*, 26–32.

[4] Blom, H., Harder, W., Matin, A., *Appl. Environ. Microbiol.* 1992, *58*, 331–334.

[5] Van Dyk, T. K., Reed, T. R., Volmer, A. C., LaRossa, R., *J. Bacteriol.* 1995, *177*, 6001–6004.

[6] Tanji, K., Mizushima, T., Natori, S., Sekimizu, K., *Biochim. Biophys. Acta* 1992, *1129*, 172–176.

[7] Horvath, I., Glatz, A., Varvasovski, V., Török, Z., Pali, T., Balogh, G., Kovacs, E., Nadasdi, L., Benkö, S., Joo, F., Vigh, L., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 3513–3518.

[8] Craig, E. A., Gross, C. A., *TIBS* 1991, *16*, 135–140.

[9] Benndorf, D., Loffhagen, N., Babel, W., *J. Basic Microbiol.* 1997, *37*, 167–174.

[10] Towner, K. J., Bergogne-Bérézin, E., Fewson, C. A., in: Towner, K. J., Bergogne-Bérézin, E., Fewson, C. A. (Eds.), *The Biology of Acinetobacter*, Plenum Press, New York 1991, pp. 1–24.

[11] Loffhagen, N., Härtig, C., Babel, W., *Appl. Microbiol. Biotechnol.* 1995, *44*, 526–531.

[12] Müller, R. H., Babel, W., *Arch. Microbiol.* 1986, *144*, 62–66.

[13] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[14] Blum, H., Beier, H., Gross, H. J., *Electrophoresis* 1987, *8*, 93–99.

[15] Jin, Y., Cerletti, N., *Appl. Theor. Electrophor.* 1992, *3*, 85–90.

[16] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* 1990, *215*, 403–410.

[17] Tokunaga, M., Matsuoka, K., Tokunaga, H., *Biosci. Biotechnol. Biochem.* 1997, *61*, 1388–1390.

[18] Hahn, M. G., Shiu, E. C., West, B., Goldstein, L., Li, G. C., *Cancer Res.* 1985, *45*, 4138–4143.

[19] Ingram L. O., *J. Bacteriol.* 1976, *125*, 670–678.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 790–797

Bo Franzén[1]
Susanne Becker[2]
Riitta Mikkola[2]
Kenneth Tidblad[2]
Agneta Tjernberg[3]
Staffan Birnbaum[2]

[1]Astra Arcus AB, Preclinical R&D, Södertälje, Sweden
[2]Process R&D, Biopharma Bulk, Global Supply Europe, Pharmacia & Upjohn, Stockholm, Sweden
[3]Rockefeller University, New York, NY, USA

Characterization of periplasmic *Escherichia coli* protein expression at high cell densities

We have used two-dimensional electrophoresis (2-DE) to analyze changes in protein expression profiles during a microbial cultivation process on an industrial scale. An *Escherichia coli* strain W3110 containing the gene for recombinant human growth hormone production was used. Samples were taken at time intervals ranging from fast to slow growth rate (late growth phase at high cell density/starvation) and 2-DE analysis combined with image analysis using the PDQuest software showed significant alterations in expression levels of a number of proteins. Twenty-four protein spots were identified using a combination of matching with SWISS-2DPAGE *E. coli* map, *N*-terminal sequence analysis and mass spectrometry matrix-assisted laser desorption/ionization (MALDI). Two of the most abundant proteins expressed at late growth phase (p*I* 5.4/28 kDa and p*I* 5.5/28 kDa) were subjected to *N*-terminal sequence analysis after electrotransfer of the proteins from a preparative 2-DE gel to polyvinylidene difluoride (PVDF) membrane. Sequence tags of five amino acids in combination with approximate p*I* and M_r identified both proteins as deoxyribose phosphate aldolase (gene name deoC). In addition, both spots were subjected to tryptic in-gel digestion and analyzed using MALDI. Peptide mass fingerprints from both spots showed similar MALDI spectra and 10 of 10 tryptic fragments confirmed the identity as deoC. The identification of the acidic variant of deoC on 2-DE gels and the observation of this variant as induced during late growth phase is novel.

Keywords: *Escherichia coli* / Protein expression EL 3412

1 Introduction

Industrial production of recombinant proteins is a complex process. Bacterial cells are frequently exposed to different kinds of limiting conditions during fermentation on a large scale [1]. For example, because of inefficient mixing in large fermentors, different kinds of gradients like pH, oxygen and glucose gradients may be formed. These conditions may in turn induce stress responses in bacterial cells and consequently change the cellular protein composition. Many stress proteins possess protease activities and this may have an impact on product quality or quantity [2, 3]. Therefore, powerful bioanalytical tools are needed for characterization of the cultivation process; this in turn is a prerequisite for increased understanding and future development of a bioprocess. Periplasmic *E. coli* proteins (PECP) represent approximately 10% of all cellular *E. coli* proteins. It is an advantage for the selection of purification strategies if a recombinant protein is exported to this compartment. However, if proteins with protease activities are expressed at elevated levels during limiting growth conditions, this may cause problems with yield and quality.

In this study, we have used two-dimensional polyacrylamide gel electrophoresis (2-DE) to analyze changes in protein expression profiles during a controlled industrial scale cultivation process. A strain of *E. coli* W3110 containing the gene for recombinant human growth hormone (rhGH) was used. Samples were taken at time intervals ranging from low to high cell density and 2-DE analysis combined with image analysis using the PDQuest software showed significant alterations in expression levels of a number of cellular proteins. We have focused on the identification of (i) a number of proteins from a periplasmic fraction resolved by 2-DE, (ii) proteins that may serve as markers for intracellular or periplasmic proteins, and (iii) proteins showing increased expression levels during late growth phase of the fermentation process when cell densities are high.

Correspondence: Dr. Susanne Becker, Process R&D, Biopharma Bulk, Global Supply Europe, Pharmacia & Upjohn, S-11287 Stockholm, Sweden
E-mail: susanne.becker@eu.pnu.com
Fax: +46-8695-4146

Abbreviations: ECP, *E. coli* protein; **HCP**, host cell protein; **PECP**, periplasmic E-coli protein; **rhGH**, recombinant human growth hormone

2 Materials and methods

2.1 Bacterial cultures and cell preparations

A recombinant *E. coli* W3110 strain was used for identification of proteins by matching with SWISS-2DPAGE. Cells were fermented at laboratory-scale conditions using

MOPS medium [4] and harvested at OD_{550nm} 1.0. An industrial-scale fermentation for production of recombinant human growth hormone was used. Cell samples were obtained from the culture in sequence at different time points: T1 and T1 + 3.5 h (T2) the first day of culture, T1 + 24 h (T3) the second day, and T1 + 48 h (T4) at harvest of the culture the third day. All cell samples were washed with cold PBS three times and cell pellets were stored at −70°C, awaiting further processing. Whole *E. coli* proteins (ECP) were prepared as described ([5], Luisa Tonella, Geneva University Hospital, personal communication). Samples for PECP preparation were taken at T4. Briefly, periplasmic extracts were prepared by extraction of frozen cell pellets by shaking with buffer and glass beads. The PECP fractions were collected after centrifugation.

2.2 Two-dimensional polyacrylamide gel electrophoresis

Immobilized pH gradients (IPG), 4–7 L (linear) or 3–10 NL (nonlinear), were used for isoelectric focusing (the first dimension). The procedures used were as recommended by the manufacturer (Amersham Pharmacia Biotech, Uppsala, Sweden [6]). We applied 20–40 μg protein per analytical separation; preparative scale isoelectric focusing was performed after rehydration of IPG strips together with samples representing approximately one mg total protein [7]. The Investigator system (ESA) was used for the second dimension of electrophoresis. For comparisons with the SWISS-2DPAGE *E. coli* database we ran ECP samples that were prepared according to Pasquali *et al.* [5], and using 3–10 NL IPG strips combined with 9–16%T linear gradient acrylamide-PDA gels. We also used 4–7 L IPG strips (PECP analysis) or 3–10 NL IPG strips (ECP analysis) for the first dimension, and 12%T homogeneous acrylamide-PDA gels in the second dimension. These conditions have been established as in-house standard conditions for 2-DE analyses of bioprocess samples. All gels were stained with silver according to [8]. Gels were then scanned using the 420oe densitometer and analyzed using the PDQuest software [9], both purchased from PDI, New York, NY, USA).

2.3 Identification of ECP

A number of proteins were identified by matching with the SWISS-2DPAGE *E. coli* map (http://expasy.hcuge.ch/ch2d/ch2d-top-html). Samples were prepared similar to the procedure for the SWISS-2DPAGE *E. coli* map.

2.4 *N*-terminal amino acid sequencing

Preparative 2-DE gels were stained with CBB (Phast Gel Blue R; Pharmacia Biotech), destained and washed with (i) water and (ii) Tris/glycine transfer buffer including 20% methanol and 0.2% SDS. Proteins were electrotransferred at 200 V/1.2 A (1.5 h, +10°C) to Pro Blott membranes (Applied Biosystems, Foster City, CA, USA) using a Hoefer TE transfer tank (Scientific Instruments, San Francisco, CA, USA). A number of protein spots were cut out and washed with 200 μL methanol for 2 min and vortexed. Then 800 μL Milli-Q-grade water were added and the samples were vortexed again. This was done to remove the surplus CBB on the pieces of membrane analyzed. The membranes were then picked up and positioned in the sample cartridge for analysis on the instrument. The instrument used was a Procise 494 Protein Sequencer, equipped with a model 140C Microgradient Delivery System, Model 785A Programmable Absorbance Detector, and Model 610A Data Analysis Program. All parts were from Perkin-Elmer (Foster City, CA, USA). Five to six amino acids were analyzed and the obtained *N*-terminal sequence tags together with M_r/*pI* were used for searching SWISS-PROT using TagIdent (http://expasy.hcuge.ch/www/tools.html).

2.5 Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)

PECP proteins were separated using preparative scale 2-DE and stained with CBB. After excision of spots, we washed out CBB using 0.2 M ammonium carbonate / 50% acetonitrile and dried the gel pieces. After reduction (10 mM dithiothreitol, 100 mM ammonium carbonate, 60 min at 56°C), alkylation (55 mM iodoacetamide, 100 mM ammonium carbonate, 45 min at room temperature), washing with 100% acetonitrile, and drying the gel pieces, we performed overnight in-gel digestion at 30°C using 1 μg trypsin (Boehringer Mannheim, Indianapolis, IN, USA) per gel piece. Finally, after extraction of peptides (60% acetonitrile, 0.1% trifluoroacetic acid, 2 × 60 min at 30°C), evaporation of acetonitrile, and addition of 5% acetonitrile in 0.1% trifluoroacetic acid, samples were desalted (using a C18 HPLC column) and mixed with the matrix, α-cyano-4-OH cinnamic acid. A MALDI TofSpec SE reflectron was used and data from MS spectra were matched using MS-Fit (http://falcon.ludwig.ucl.ac.uk/cgi-bin/msfit). A delta mass of < 0.3 Da was accepted as positive search hit for matched peptide sequences. Theoretical M_r/*pI* values for matching candidates were compared with calculated values obtained from our 2-DE gels.

3 Results and discussion

3.1 Alterations in protein expression levels

Samples were taken from one industrial-scale rhGH producing culture at different time points: T1 and T1 + 3.5 h

(T2) the first day of the culture process, T1 + 24 h (T3) the second day and T1 + 48 h (T4) at harvest of the culture the third day. Samples for PECP preparation were taken at T4. The optical densities (OD_{550nm}) at these four time points were 0.7, 3.0, 57 and > 100 units, respectively. Analysis by 2-DE showed more than 1000 protein spots per gel, representing the whole *E. coli* cell lysates (ECP). Figure 1 illustrates original silver-stained 2-DE gels (3–10

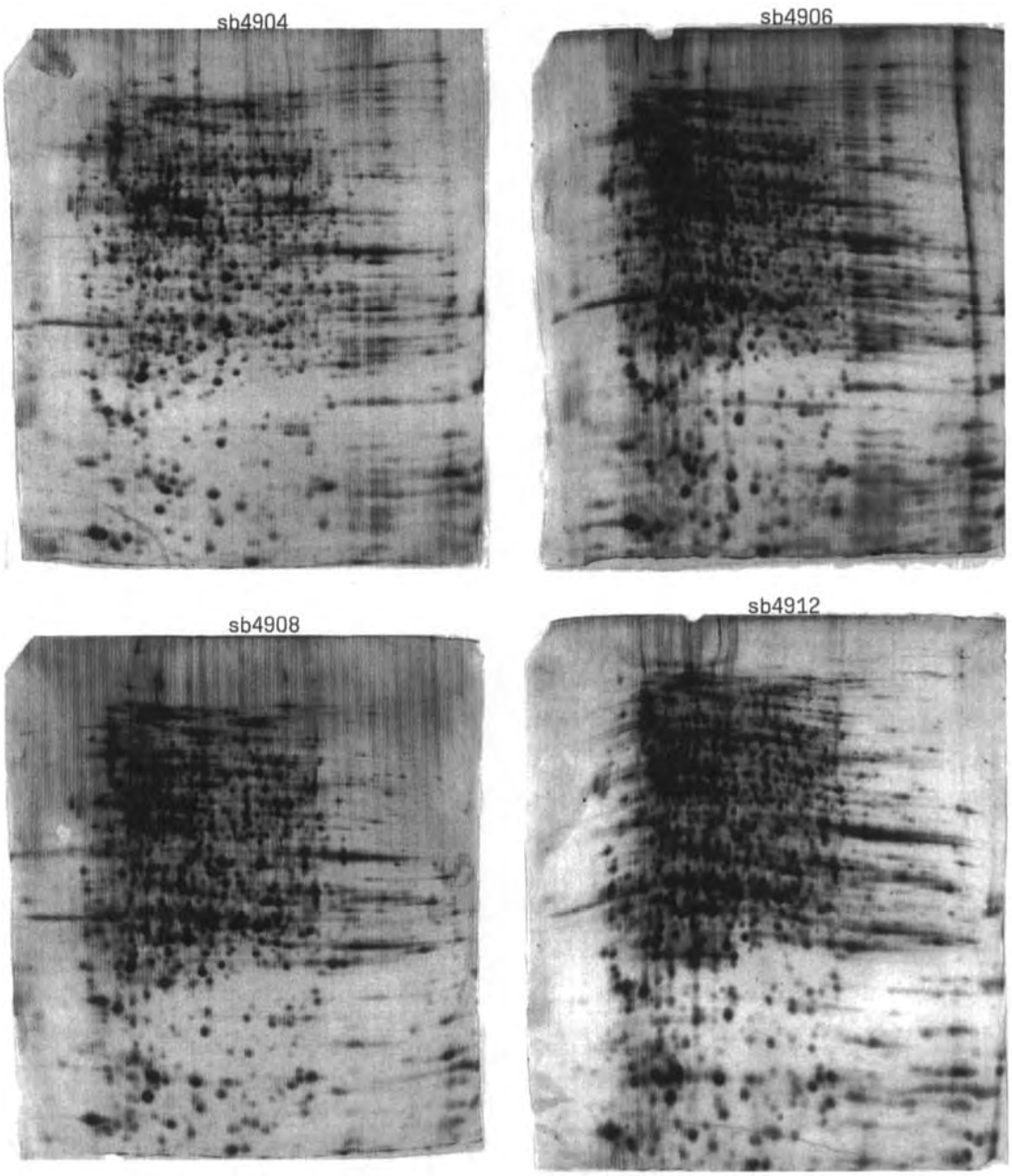


Figure 1. 2-DE profiles of *E. coli* proteins. Acidic polypeptides are to the left. 2-DE gels representing various fermentation times: T1 (gel No. sb4904), T2 (gel No. sb4906), T3 (gel No. sb4908) and T4 (gel No. sb4912).

NL) representing time points T1 (gel No. sb4904), T2 (gel No. sb4906), T3 (gel No. sb4908) and T4 (gel No. sb4912). Many alterations in expression levels were observed already during the first culture period (from T1 to T2). Analysis of PECP fractions resolved an average of 450 periplasmic proteins per 2-DE gel. One 2-DE gel representing T1 was chosen as the ECP reference map and a matchset was constructed, including two PECP 2-DE maps. Approximately 200 PECP proteins were matched to the ECP reference map. The expression level of each protein spot was normalized and expressed as parts per million (ppm) of the total optical integrated density of the corresponding 2-DE gel. The analysis focused on the search for clear trends in expression levels, going from T1, *via* T2 and T3, to T4. Only > 3-fold differences between T1 and T4 were considered. A large number of proteins showed significant differences in expression levels when 2-DE gels representing slow growth rate / high density cultures (T3 and T4) were compared to rapid growth rate / low density cultures (T1 and T2). If low abundance proteins (< 1000 ppm) were excluded, approximately 50 proteins were detected as 3- to 4-fold differentially expressed between these two groups. A majority of these proteins showed increased expression levels at high cell densities.

3.2 Matching with the SWISS-2DPAGE *E. coli* map

Almost 300 spots were automatically matched by the software between the SWISS-2DPAGE *E. coli* map and our 9–16%T *E. coli* map (Fig.2). A number of differences observed may be explained by genetic differences between the different W3110 strain variants used in-house and for the SWISS-2DPAGE map, differences in growth conditions, and electrophoretical conditions. Fifty of the matched proteins were annotated in SWISS-PROT; however, reliable high quality matching was valid for 30 protein spots. Nineteen of these spot identities were transferred *via* a 12%T map representing *E. coli* cells grown at laboratory scale using MOPS medium, to the 12%T reference map representing cells grown under industrial scale conditions (Fig. 2 and Table 1).

3.3 Identification of PECP expressed at stable or decreasing levels by gel matching

A series of 2-DE gels (IPG 4–7 and 12%T) were analyzed, representing various proportions between PECP and ECP. Using PDQuest-assisted matching, we transferred 14 protein identities to our PECP reference map (Fig. 3). A number of these proteins were intracellular proteins showing high/intermediate and stable expression in all samples (T1–T4), *e.g.*, elongation factor EF-TU (gene name *tufA*), 30S ribosomal protein S1 (*rpsA*), and ATP syntase beta-chain (*atpD*) (Table 1, Figs. 1 and 2). As expected, these proteins showed low expression level in PECP samples, on average less than 10–20% of the levels observed in ECP samples. These proteins may well serve as indicators of the purity of the PECP preparations. Four proteins, pyruvate dehydrogenase E1 component (*aceE*), asparaginyl tRNA synthetase (*asnS*), inosine 5′-monophosphate dehydrogenase (*guaB*) and pyruvate kinase (*pyrF*), showed significant decreasing expression level profiles, indicating down-regulation of these proteins during starvation. The levels of these proteins were low or not detected in PECP. A number of proteins showed stable levels through all samples, PECP included, *e.g.*, phosphoglycerate kinase (*pgk*), transaldolase B (*talB*), and a previously unidentified protein spot, No. 4213, which was identified using sequencing (see below). Two spots, periplasmic oligopeptide binding protein (*oppA*) and No. 3431, showed low or intermediate levels in all cultures, but high levels in PECP. These may be useful as indicators of the PECP preparation process.

3.4 Identification of proteins expressed at high or increasing levels using *N*-terminal amino acid sequencing and MALDI

A number of periplasmic proteins showing high expression levels, or other interesting characteristics, were excised from PVDF membranes and subjected to *N*-terminal sequencing. Five spots were identified on the basis of a 5–6 amino acid sequence tag combined with approximate M_r/p*I* data. Spot No. 2119 showed a markedly increased expression level (> 5-fold) at slow growth rate / high cell density (T3, T4) compared to rapid growth rate / low cell density (T1, T2). This protein was identified as LAO-binding periplasmic protein (gene name argT; Fig. 2, Table 1). Spot No. 3431 showed stable low/intermediate expression levels in all T1–T4 samples, but high levels (> 6-fold) in the PECP fraction. This spot was identified as a periplasmic protein (maltose binding periplasmic protein, gene name malE). Spot No. 4213 showed high expression levels in all samples. *N*-terminal sequence analysis identified this protein as deoxyribose-phosphate aldolase (deoC), a protein which previously has not been annotated on published *E. coli* 2-DE maps. Spot No. 4213/deoC has a p*I* of 5.5, which corresponds well with published data (SWISS-PROT accession No. P00882).

Spot No. 3209 shows the same M_r as No. 4213 but was more acidic (p*I* 5.4) and showed 4-fold higher expression levels in high density cultures compared to low density cultures. The expression level of this protein in the PECP fraction was essentially the same as for No. 4213. This spot was also identified as deoxyribose-phosphate aldolase by sequence analysis. This result was surprising and to be sure of the identification, both spots were excised

from gels, in-gel digested by trypsin and analyzed using MALDI. Both MS spectra were similar and ten peptide sequence masses were entered into MS-fit for search in the database and both matched perfectly to deoxyribose-phosphate aldolase. The MS-Fit search results showed two first-ranking proteins, NCBI accession No. 729314

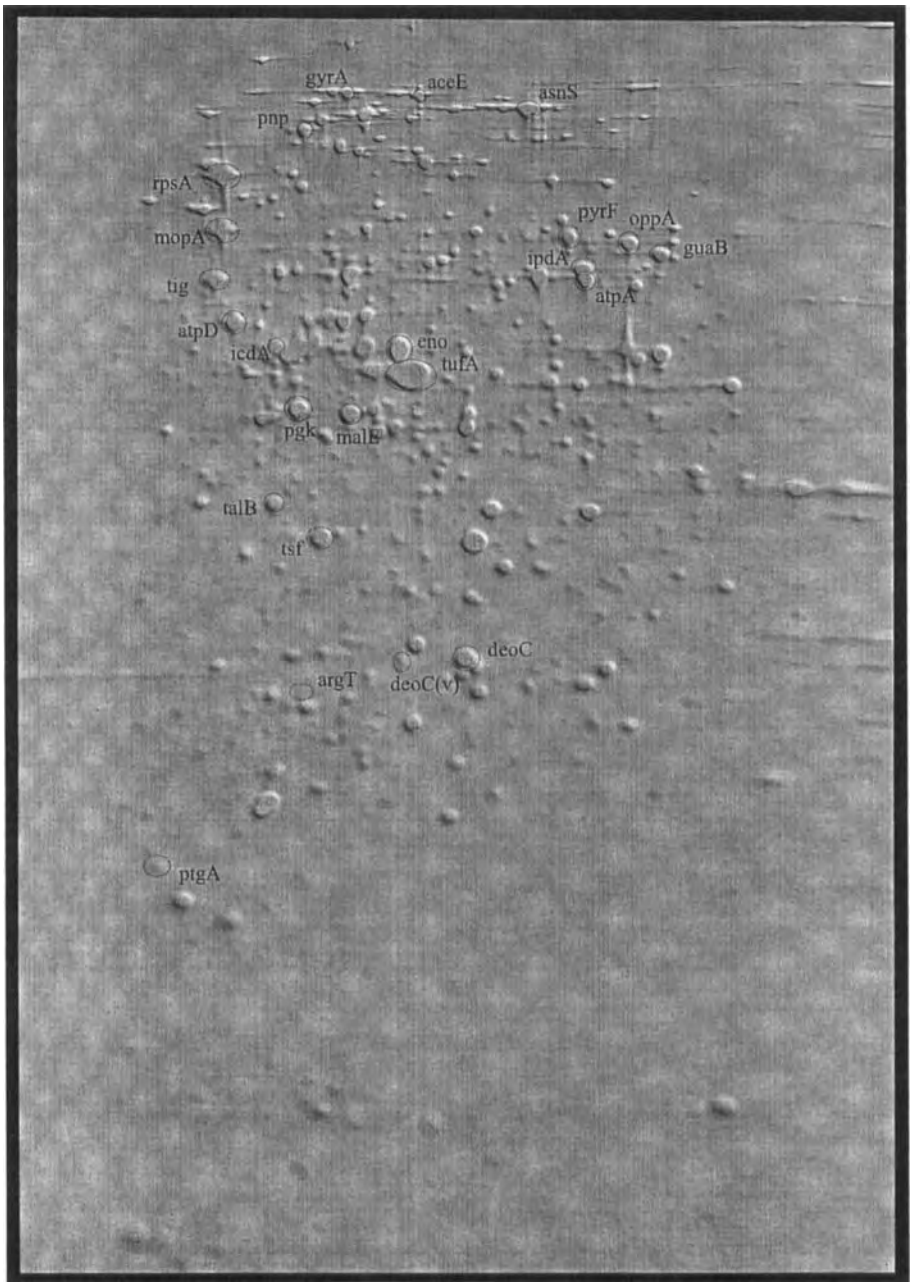


Figure 2. 2-DE profile of *E. coli* proteins representing the reference maps for our "ECP identified spots matchset". Acidic polypeptides are to the left. Identified proteins are circled and given by their gene name.

Table 1. Identified spots and change in expression levels during the fermentation process from rapid growth rate / low cell density to slow growth rate / high cell density (starvation)

Gene name	SSP	Method for identity	AC SWISS PROT No. 34	p*I* and molecular mass (kDa)	Protein description	Expression profile and level in PECP fraction
aceE	4933	Gel match[a)]	P06958	5.43 110.6	Pyruvate dehydrogenase E1 component	Decreasing profile Low or not detected in PECP
argT	2119	AA sequence	P09551	5.09 25.6	LAO-binding periplasmic protein	Increasing profile High in PECP
asnS	5921	Gel match	P17242	5.75 98.6	Asparaginyl TRNA synthetase	Decreasing profile Low or not detected in PECP
atpA	6518	Gel match	P00822	5.86 52.7	ATP syntase alpha-chain	Intermediate expression profile Low in PECP
atpD	1505	Gel match	P00824	5.00 47.7	ATP syntase beta-chain	High expression profile Low in PECP
deoC	4213	AA sequences + MS	P00882	5.50 27.7	Deoxyribose phosphate aldolase	High expression profile High in PECP
deoC (v)	3209	AA sequence + MS[b)]		5.40 27.5	Deoxyribose phosphate aldolase, variant	Increasing profile High in PECP
eno	3513	Gel match	P08324	5.40 46.3	Enolase	High expression profile Low or not detected in PECP
guaB	7602	Gel match	P06981	6.00 55.8	Inosine 5'-monophosphate dehydrogenase	Decreasing profile Not detected in PECP
gyrA	2914	Gel match	P09097	5.31 112.5	DNA gyrase subunit A	High expression profile Low in PECP
icdA	1522	Gel match	P08200	5.05 46.5	Isocitrate dehydrogenase	High expression profile Low or not detected in PECP
lpdA	6627	Gel match	P00391	5.85 54.0	Dihydrolipoamine dehydrogenase	Intermediate expression profile Low in PECP
malE	3431	AA sequence	P02928	5.32 43.1	Maltose-binding periplasmic protein	Low expression profile High in PECP
mopA	1601	Gel match	P06139	4.99 58.5	60 kD a chaperonin	High expression profile High in PECP
oppA	6628	Gel match	P23843	5.94 56.9	Periplasmic oligopeptide binding protein	Intermediate expression profile High in PECP
pgk	2417	Gel match	P11665	5.08 43.4	Phosphoglycerate kinase	High expression profile High in PECP
pnp	2705	Gel match	P05055	6.11 86.8	Polyribonucleotide nucleotidyl transferase	High expression profile Low or not detectable in PECP
ptgA	0014	AA sequence	P08837	4.73 18.1	Glucose-specific phosphotransferase	Low expression profile Low in PECP
pyrF	6602	Gel match	P08244	5.83 57.6	Pyruvate kinase	Decreasing profile Low or not detected in PECP
rpsA	1725	Gel match	P02349	4.99 66.2	30S ribosomal protein S1	High expression profile Low or not detected in PECP
talB	1308	Gel match	P30148	5.05 36.6	Transaldolase B	High expression profile High in PECP
tig	0513	Gel match	P22257	4.96 52.6	Trigger factor	High expression profile Low or not detectable in PECP
tsf	2323	Gel match	P02997	5.18 34.5	Elongation factor EF-TS	Decreasing profile Low in PECP
tufA	3430	Gel match	P02990	5.41 45.2	Elongation factor EF-TU	High expression profile Low in PECP

a) Identification using gel matching with SWISS-2DPAGE *E. coli* map
b) Identification using *N*-terminal amino acid sequence analysis and/or MALDI
AA, amino acid
SSP, matchset spot number

and No. 41252 (Table 2). A likely explanation for this is that threonine at position 18 of this aldolase has been replaced by asparagine in the acidic variant of deoC. Theoretically, the delta mass between these two proteins is 13 Da. However, we did not detect any peptide sequence covering this region of the protein by MALDI analysis.

Our results confirm that deoC exists as two variants and indicate that the acidic variant of deoC is expressed at elevated levels during late growth phase of the cultivation process. Deoxyribose-phosphate aldolase is involved in the synthesis of thymidine and its concentration is not expected to decrease with decreasing growth rate because the DNA/mass ratio of bacterial cells increases with decreasing growth rate [10]. The physiological importance of the two different forms of the enzyme is unclear. The fact that the more acidic form (spot No. 3209) increases during starvation / late growth phase of the cultivation

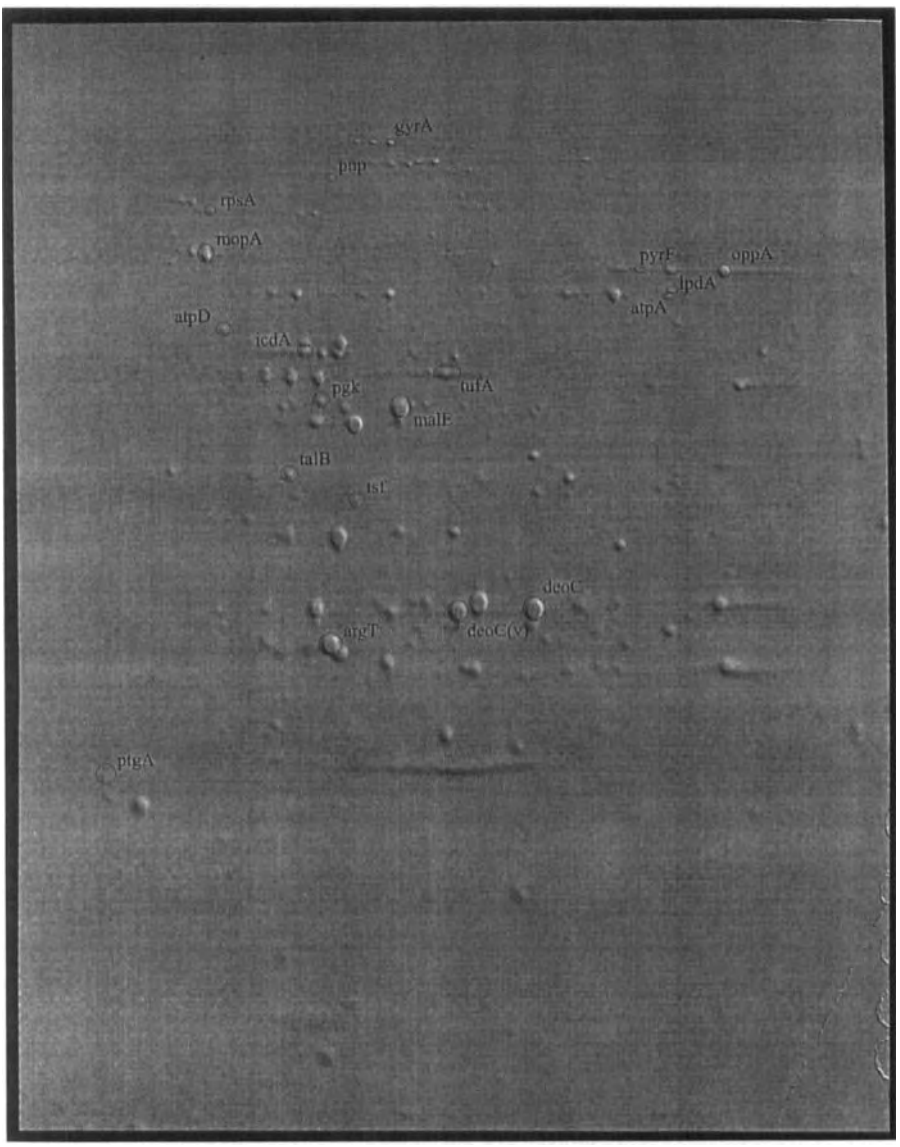


Figure 3. 2-DE profile showing periplasmic *E. coli* proteins representing the reference map for our "PECP identified spots matchset". Acidic polypeptides are to the left. Identified proteins are circled and given by their gene name.

Table 2. MS-Fit search results

Data submitted	MH+ matched	Delta Da	Start	End	Peptide sequence
716.5	716.4459	0.0541	52	57	(R)FIPIAR(K)
716.5	716.3943	0.1057	147	152	(K)DEALIR(K)
929.6	929.4692	0.1308	62	69	(K)EQGTPEIR(I)
1088.7	1088.6217	0.0783	197	207	(K)TVGFKPAGGVR(T)
1206.7	1206.7098	–0.0098	235	246	(R)FGASSLLASLLK(A)
1532.8	1532.7895	0.0105	38	51	(K)TPVGNTAAICIYPR(F)
1698.9	1698.9642	–0.0642	138	152	(K)VIIETGELKDEALIR(K)
1855.9	1855.9230	–0.0230	92	108	(R)AAIAYGADEVDVVFPYR(A)
1896.9	1896.9132	–0.0132	215	231	(K)YLAIADELFGADWADAR(H)
2383.1	2383.1894	–0.0894	70	91	(R)IATVTNFPHGNDDIDIALAETR(A)

E. coli deoxyribose-phosphate aldolase (deoC), 10 of 10 peptides match (Delta < 0.3 Da). Matched peptides cover 47% of the protein (122 of 259 amino acids).

process suggests that it may have a role in stress response. Alternatively, it may be a specific response for the cultivation conditions used or the limiting substrate environment experienced.

4 Concluding remarks

Many proteins may serve as markers for various culture conditions and purity of subcellular fractions. Only a few of these have been described in this study. The remaining number of potential markers represents a powerful resource for future analysis and monitoring of industrial-scale production of recombinant proteins. We have given one example of an application of proteome analysis in bioprocess characterization. In combination with other process-related information, cell modeling, and multivariate analysis, proteome analysis will most likely become a powerful tool for increased understanding of recombinant protein production processes in *E. coli*. This, in turn, will facilitate metabolic engineering endeavors for process optimization.

Received January 5, 1999

5 References

[1] Bylund, F., Collet, E., Enfors, S.-O., Larsson, G., *Bioprocess Engineer.* 1998, *18*, 171–180.

[2] Matin, A., *Mol. Microbiol.* 1991, *5*, 3–10.

[3] Groat, R. G., Schultz, J. E., Zychlinsky, E., Bockman, A., Matin, A., *J. Bacteriol.* 1986, *168*, 486–493.

[4] Cayley, S., Record, Jr., M. T., Lewis, B. A., *J. Bacteriol.* 1989, *7*, 3597–3602.

[5] Pasquali, C., Frutiger, S., Wilkins, M. R., Hughes, G. J., Appel, R. D., Bairoch, A., Schaller, D., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 547–555.

[6] Instruction manual: *Immobiline DryStip Kit for 2-D Electrophoresis with Immobiline DryStrip and ExcelGel SDS*, Pharmacia Biotech No. 18-1038-63.

[7] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[8] Johansson, S., Skoog, B., *J. Biochem. Biophys. Methods* 1987, *14*, 33.

[9] Garrels, J. I., *J. Biol. Chem.* 1979, *16*, 7961–7977.

[10] Bremer, H., Dennis, P. P., in: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella*, ASM Press, Washington DC 1997, Vol. 2, pp. 1553–1569.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 798–805

Leila H. Choe[1]
Wilfred Chen[2]
Kelvin H. Lee[1]

[1]School of Chemical Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Chemical and Environmental Engineering, University of California, Riverside, CA, USA

Proteome analysis of factor for inversion stimulation (Fis) overproduction in *Escherichia coli*

The factor-for-inversion stimulation protein (Fis) is a global regulatory protein in *Escherichia coli* that activates ribosomal RNA (rRNA) transcription by binding to three upstream activation sites of the rRNA promoter and enhances transcription 5- to 10-fold *in vivo*. Fis overexpression results in different effects on cell growth depending on nutrient conditions. Differential proteome analysis of Fis-expressing cells shows ten protein spots corresponding to Fis overexpression in both rich (YT) and minimal (M9+glucose) media. Three of these spots have been identified as elongation factor TS, histidine-binding periplasmic protein precursor, and ketol-acid reductoisomerase.

Keywords: Proteome analysis / Fis / *Escherichia coli* EL 3339

1 Introduction

The desire to control recombinant protein production in bacterial cell culture necessitates an understanding of the control of ribosome synthesis. The coordination of ribosome synthesis is a complex physiological task for cells and it is known that the regulation of ribosomal RNA (rRNA) synthesis is a critical determinant within this process [1]. Indeed, ribosomes and associated factors can account for up to 50% of cell mass at high growth rates [2] while rRNA constitutes over half of the total cellular RNA in cells under these same conditions [3]. *Escherichia coli* has seven rRNA operons (*rrn* operons) in its genome and the synthesis of most (not all tRNAs are cotranscribed with rRNA) transfer RNAs (tRNA) is coregulated with that of rRNA by cotranscription. Variation in rRNA expression appears as a consequence of nutrient conditions which affects growth rate. In particular, the number of ribosomes and the synthesis rates of rRNA are roughly proportional to the growth rate squared (μ^2) and rRNA synthesis is the rate-limiting step in ribosome synthesis in *E. coli* [4]. Furthermore, it is known that rRNA expression is feedback-regulated by growth rate control [5]. Experiments show that this feedback regulation depends on the number of actively translating ribosomes [6, 7]; however, the mechanism by which translating ribosomes are coupled to growth control has not been completely elucidated. rRNA expression is further, but separately, regulated by a number of repressing and activating mechanisms such as ppGpp levels as part of the stringent response [8]. The *rrn* operon contains upstream promoter sites which are responsible for strong stimulation of promoter activity but are not required for growth rate regulation. These sites include binding sites for the factor-for-inversion stimulation (Fis) protein.

Fis is an 11.2 kDa, p*I* 9.34, relatively abundant DNA-binding protein. Large fluctuations in Fis expression occur during the growth cycle, with Fis becoming less abundant in late-log-phase cells [9, 10]. Fis is also less abundant in cells grown on poor medium and this scarcity appears to cause a decrease in activity of most growth rate-regulated promoters of stable rRNA operons [11, 12]. Fis has the ability to stimulate site-specific DNA inversion reactions by binding to an enhancer sequence and bending the DNA [13, 14]. Although Fis is not required for cell growth [15], it has been shown to stimulate stable rRNA synthesis both in rich medium and under conditions of nutrient upshift [16]. Furthermore, Fis-dependent activation is crucial for providing the high rate of rRNA synthesis required forrapid cell growth. We have previously studied the effect of Fis overproduction on growth and ribosome synthesis for different media (submitted). Fis-expressing cells demonstrate significant different effects on growth characteristics on rich media and minimal media. Here, we employ the tools of proteome analysis to help elucidate the biochemical basis for this phenomena.

2 Materials and methods

2.1 Strains and vectors

Wild-type Fis was expressed in *E. coli* W1485 △H LAM-rpoS396(Am) rph-1 [17]. A positive mutant of Fis used in this study is designated GS. The GS gene contains a G-to-S substitution at amino acid 72 of the wild-type Fis protein which renders the GS protein much less able to activate expression of the *rrnB* promoter as compared to

Correspondence: Dr. Kelvin H. Lee, Chemical Engineering, 120 Olin Hall Cornell University, Ithaca, NY 14853-5201, USA
E-mail: khlee@cheme.cornell.edu
Fax: +1-607-255-9166

Abbreviations: Fis, factor for inversion stimulation; **IPTG,** isopropyl-β-D-thioglactopyranoside; **PTS,** phosphotransferase system; **rRNA,** ribosomal RNA; **tRNA,** transfer RNA; **YT,** yeast-tryptone

wild-type Fis [18]. Both the Fis and GS genes were subcloned into the control of the *lac-tac* promoter for isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible expression. Cells were cultivated at 37°C on yeast-tryptone (YT) media or minimal M9 media supplemented with 0.2% glucose or 0.2% glycerol [19]. OD measurements were made on a Beckman DU 640 spectrophotometer (Palo Alto, CA).

2.2 Proteome analysis

2.2.1 Sample preparation and electrophoresis

Samples for proteome analysis were taken at approximately 3 h post-induction with 0.25 mM IPTG. The culture was pelleted and washed four times in a solution containing 3.0 mM KCl, 1.5 mM KH_2PO_4, 68 mM NaCl, and 9.0 mM NaH_2PO_4. The washed pellet was resuspended in a solution containing 10 mM Tris-HCl, pH 8.0, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM Pefabloc SC (Boehringer, Indianapolis, IN) and 0.1% SDS. This solution was sonicated at full power on ice for 30 s in a Fisher Model F550 sonifier (Pittsburgh, PA) and stored at –75°C until use. One hundred μg of protein (corresponding to 40 μL of sample) was mixed with 20 μL of 8 M urea, 4% w/v CHAPS, 65 mM DTT, 67 mM Tris, pH 8.0, and a trace of bromophenol blue. This mixture was loaded onto pH 3–10 nonlinear Immobiline gels (Amersham-Pharmacia-Hoefer, Piscataway, NJ) by in-gel rehydration. The reswelling solution contained 8 M urea, 2% CHAPS, 0.3% DTT, 1.33% BioLyte 3–10 and 0.67% BioLyte 5–7 (Bio-Rad Laboratories, Richmond, CA), and a trace of bromophenol blue. Isoelectric focusing was performed for 71 750 Vh. Gels were subsequently equilibrated for 15 min in a solution containing 6 M urea, 2% DTT, 30% glycerol, 2% SDS and 0.05 M Tris, pH 6.8, and for 5 min in a solution containing 6 M urea, 2.5% iodoacetamide, 30% glycerol, 2% SDS and 0.05 M Tris, pH 6.8. Strips were transferred to a vertical SDS-PAGE tank (Bio-Rad) and covered with a solution of 0.5% agarose, 25 mM Tris (pH 8.3), 198 mM glycine, and 0.1% SDS. The 1.5 mm thick 12%T gels [20] were run at 40 mA per gel until the dye front migrated to the end of the gel. Gels were stained with ammoniacal silver as described previously [21].

2.2.2 Computer-assisted analysis

Computer-assisted gel analysis (Melanie II, Bio-Rad Laboratories) was performed on images captured with a Molecular Dynamics Personal Densitometer. Default parameters were used for feature detection and matching and corrected by visual inspection. An estimate of the relative quantitative changes was made based on the change in percent volume among silver-stained gels. These quantitative data were obtained from multiple gel runs and from a Melanie II analysis of the gel images. Data from spots that were very faint or very strongly stained were not included in quantitative comparisons. Spot changes of interest were tested on multiple gels for reproducibility. The genetic basis for spot changes was made based on a comparison to the *E. coli* 2-DE database available at SWISS 2-D PAGE (www.expasy.ch). It is noted that existing *E. coli* databases use strain W3110 whereas this study was performed in strain W1485. Some strain-specific differences in these proteomes exist and *N*-terminal sequence tagging of landmark proteins was performed to help establish reference points for comparison between W1485 and W3110 proteomes.

3 Results and discussion

3.1 Growth studies

Fis-expressing cells were grown in YT media containing 0.25 mM IPTG and compared to GS-expressing cells (GS-expressing cells are positive mutants of Fis) grown in the same media. GS-expressing cells have a post-induction growth rate of 0.119 h^{-1} while Fis-expressing cells demonstrate a post-induction growth rate of 0.100 h^{-1} (see Fig. 1 and Table 1). This 16% decrease in growth

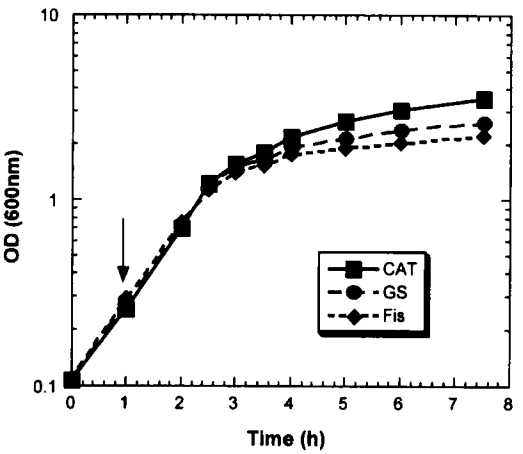


Figure 1. Growth curves for Fis-, GS-, and CAT-expressing *E. coli* grown on YT medium. The arrow indicates the time of IPTG-induced protein induction.

Table 1 Post-induction growth rates (h^{-1}) of Fis- and GS-expressing cells on different culture medium

	YT 0.25 mM IPTG	M9+glycerol 0.25 mM IPTG	M9+glucose 0.25 mM IPTG
GS	0.119	0.245	0.319
Fis	0.100	0.165	0.389
Change	16% decrease	33% decrease	22% increase

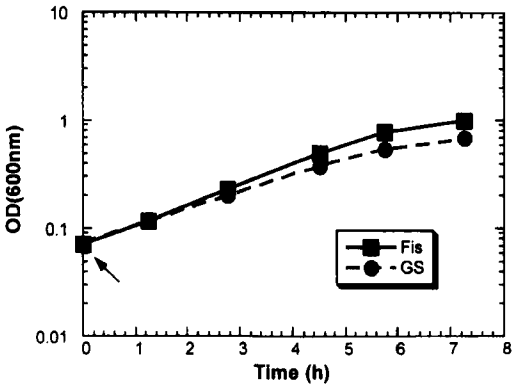


Figure 2. Growth curves for Fis- and GS-expressing *E. coli* grown on minimal M9 + glucose medium. The arrow indicates the time of IPTG-induced protein induction.

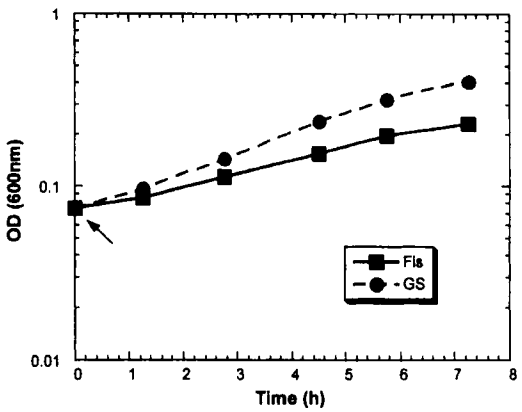


Figure 3. Growth curves for Fis- and GS-expressing *E. coli* grown on minimal M9 + glycerol medium. The arrow indicates the time of IPTG-induced protein induction.

rate parallels an expected increase in rRNA synthesis as measured directly by primer extension (data not shown). Elevated rRNA synthesis is an expected observation based on the known function of Fis. In contrast, CAT-expressing control cells also grew 40% faster than Fis-expressing cells on YT media (0.167 h^{-1}) and cell growth was proportionally inhibited at higher levels of gene induction (data not shown).

Fis-expressing cells grown on minimal M9 media supplemented with glucose demonstrate enhanced growth characteristics compared to GS-expressing cells (see Fig. 2 and Table 1). The Fis-expressing cells, 3 h post-induction, grow 22% faster than GS-expressing cells (0.319 h^{-1} *versus* 0.389 h^{-1}). Here again Fis-expressing cells have elevated rRNA synthesis as expected. In contrast, Fis-expressing cells grown on M9 media supplemented with glycerol (see Fig. 3) demonstrated a 33% decrease in growth rate as compared to GS-expressing controls (0.245 h^{-1} *versus* 0.165 h^{-1}). A decrease in growth rate is also observed for these cells growing on YT medium. This observation suggests a link between the phosphotransferase system (PTS) and Fis activity in cells. When glucose is present in the medium, intracellular cyclic adenosine monophosphate (cAMP) levels are relatively low and the PTS system is active; however, when glucose is not present, cAMP levels are relatively high and the PTS system is inactive. The observed growth rate differences are consistent with low intracellular cAMP levels in cells grown on glucose-supplemented medium. It is known that several Fis-regulated promoters contain putative cAMP-cAMP receptor protein (CRP) binding sites (CAP binding sites) overlapping the Fis-binding sites [22]. It is possible that a depletion in intracellular cAMP enables Fis to bind more readily to these promoters, thus activating these pathways and enabling faster cell growth (compared to GS-expressing cells).

3.2 Proteome analysis

The presence of various nutrients in YT-rich medium (as compared to minimal M9 medium) suggests that *E. coli* grown on YT will demonstrate a different physiology from their M9 counterparts. In particular, YT-grown cells will need to retrieve various amino acids and other nutrients from the media for incorporation into biomass; yet these cells will not need to synthesize these molecules from precursors. In contrast, cells grown on M9 media will actively synthesize many molecules from precursors while devoting significant effort to retrieve a particular carbon source from the medium. To elucidate how these physiological differences may contribute to growth-rate differences seen among the Fis-expressing strains, we performed two-dimensional protein electrophoresis on samples obtained from Fis-expressing and GS-expressing cells. Cells grown in YT media (Figs. 3 and 4) were compared, as were cells grown in minimal M9+glucose medium (Figs. 6 and 7).

A direct comparison between Fis-expressing and GS-expressing strains grown on YT media and at 3 h post 0.25 mM IPTG induction reveals the presence of approximately 1600 protein spots. Of these, we observe 23 spots which are qualitatively different or quantitatively upregulated at least 2-fold in Fis-expressing cells (see Fig. 4) as compared to GS-expressing cells (see Fig. 5) and six spots which are likewise different in GS-expressing cells as compared to Fis-expressing cells. These differences

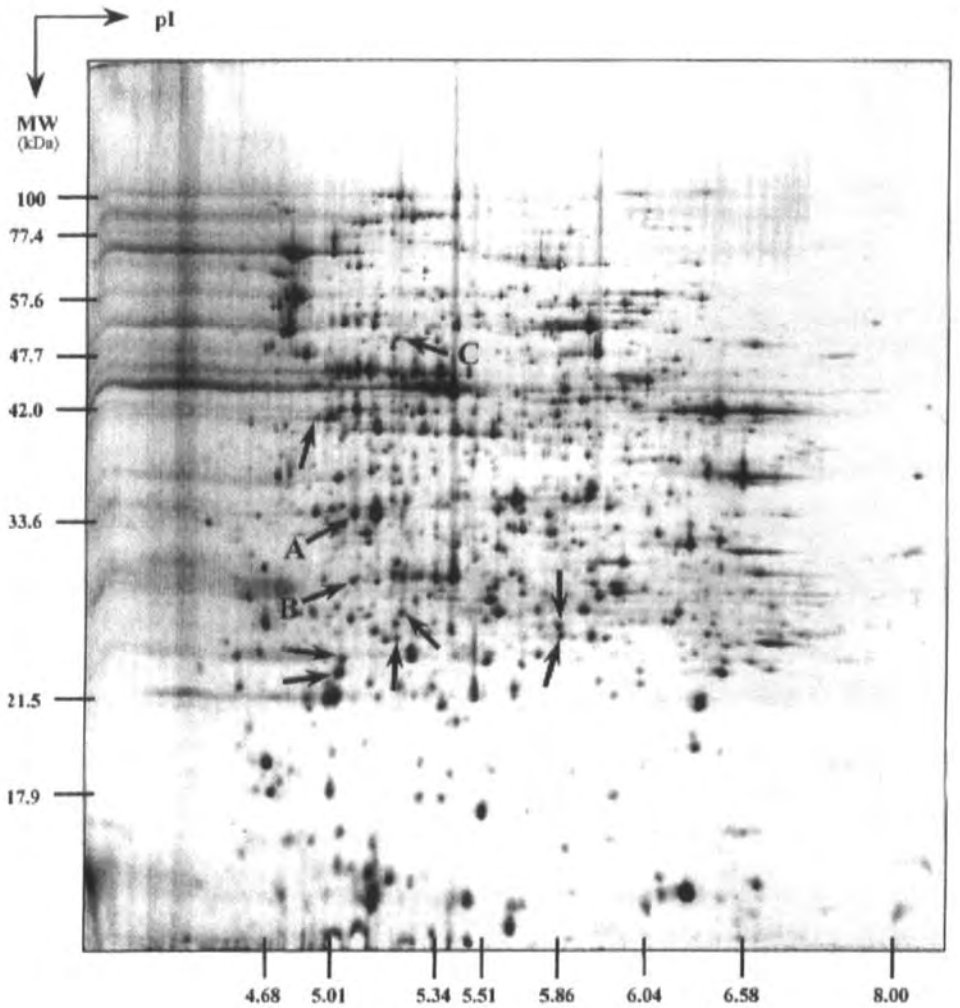


Figure 4. Proteome of Fis-expressing *E. coli* grown on YT medium. The location of 10 spots that appear to be involved in the observed decrease in growth rate on YT medium and the observed increase in growth rate on M9 medium are noted. Three of these 10 spots have been characterized as A, elongation factor TS; B, histidine-binding periplasmic protein precursor; and C, ketol-acid reductoisomerase.

can be attributed to the presence or absence of Fis-binding and subsequent gene activation, which results in the observed decrease in growth rates for Fis-expressing cells on YT medium.

Fis-expressing cells (see Fig. 6) were compared to GS-expressing cells (see Fig. 7) grown on M9 media supplemented with glucose. The Fis-expressing cells demonstrate enhanced growth rate as compared to the GS-expressing cells. In this comparison 19 spots appear significantly upregulated in Fis-expressing cells as compared to GS-expressing cells while only one spot has that property in GS-expressing cells. Those spots which appear in Fis-expressing cells and not in GS-expressing cells grown on both YT media and M9 media likely correspond to genes which demonstrate enhanced expression due to increased Fis-expression. There are ten spots with this characteristic (as labeled on Figs. 4 and 5). These genes, and their corresponding pathways, are thus related to the observed differences in growth characteristics on M9 and YT media. To date, we have characterized three of these ten spots. They are: elongation factor TS

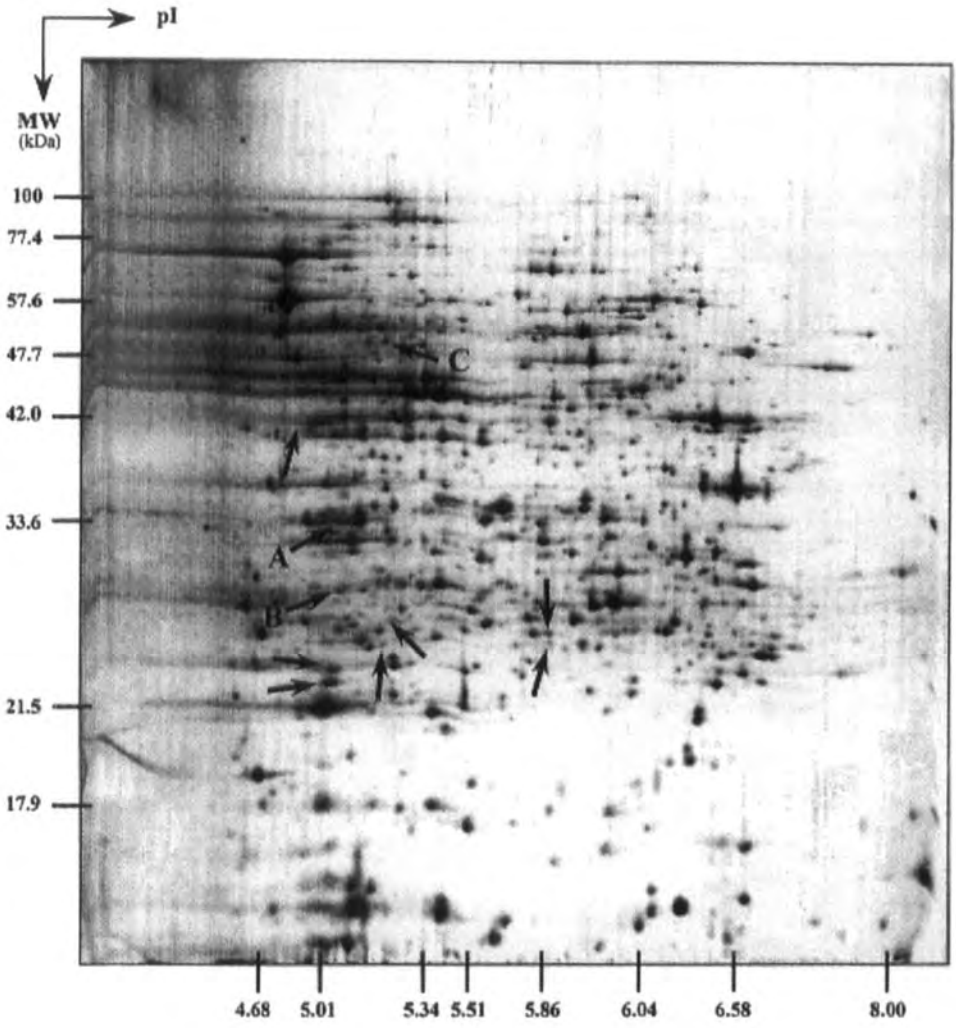


Figure 5. Proteome of GS-expressing *E. coli* grown on YT medium.

(EFTS; SWISS-PROT Accession No. P02997), histidine-binding periplasmic protein precursor (HisJ; SWISS-PROT Accession No. P39182), and ketol-acid reductoisomerase (IlvC; SWISS-PROT Accession No. P05793). We are in the process of characterizing the other changes of interest which are not present in current databases.

3.3 Characterized spot changes

Elongation factor TS associates with elongation factor-Tu and forms part of the ribosomal complex. As such, its elevated expression in Fis-expressing cells is expected because Fis activates the rRNA promoter. The histidine-binding periplasmic protein precursor is a component of the high-affinity histidine permease, which is a binding-protein-dependent transport system. The family of proteins which make up this transport system (including HisJ) has been implicated in the active transport of amino acids and sugars. Elevated expression of this protein precursor in Fis-expressing strains leading to different growth characteristics on media with and without glucose suggests a nonlinear relationship between amino acid import and the PTS uptake system. Ketol-acid reductoisomerase performs the second step in valine and isoleucine biosynthesis. Elevated expression of this protein is also consistent with the observed phenomena because cells grown on

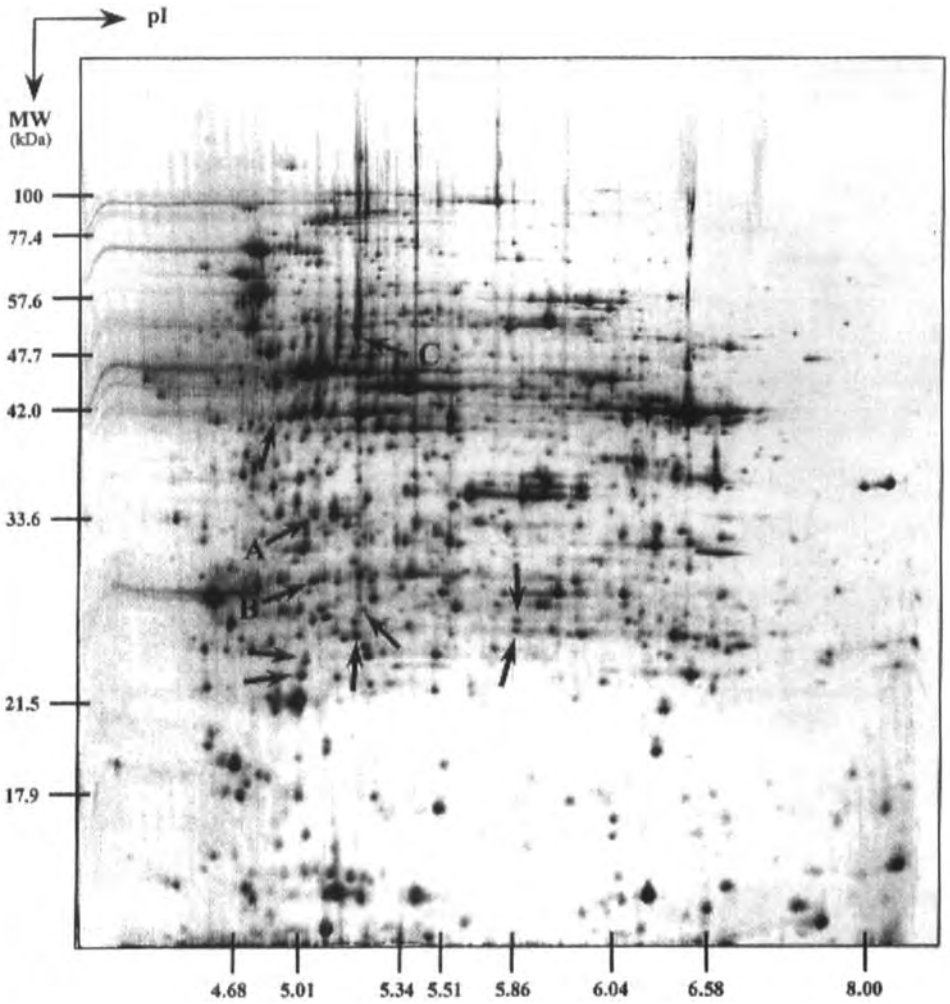


Figure 6. Proteome of Fis-expressing *E. coli* grown on M9+glucose medium. Nineteen spots are significantly increased as compared to GS-expressing cells. The location of 10 of these is depicted on Fig. 4.

M9 media must synthesize valine and isoleucine for incorporation into biomass. Fis-expressing cells grown on YT medium can recruit these amino acids from the medium.

4 Concluding remarks

In an attempt to elucidate the biochemical coordination of rRNA synthesis and growth rate control we have studied the effects of Fis-overexpression on *E. coli* physiology under different growth conditions. Differential proteome analysis suggests several key changes which involve various aspects of cellular physiology including carbon metabolism, nutrient uptake, and translation among others. We are currently investigating the genetic basis for uncharacterized spot changes as well as developing a predictive model for use in the engineering of cells with enhanced recombinant protein production.

This work was supported by an NSF Career Award (BES9700981) to WC, by a grant from Intel Corporation (98–238) to KHL and by a generous gift from Merck Corporation to KHL. We thank Dr. R. L. Gourse for providing the GS gene.

Received September 1, 1998

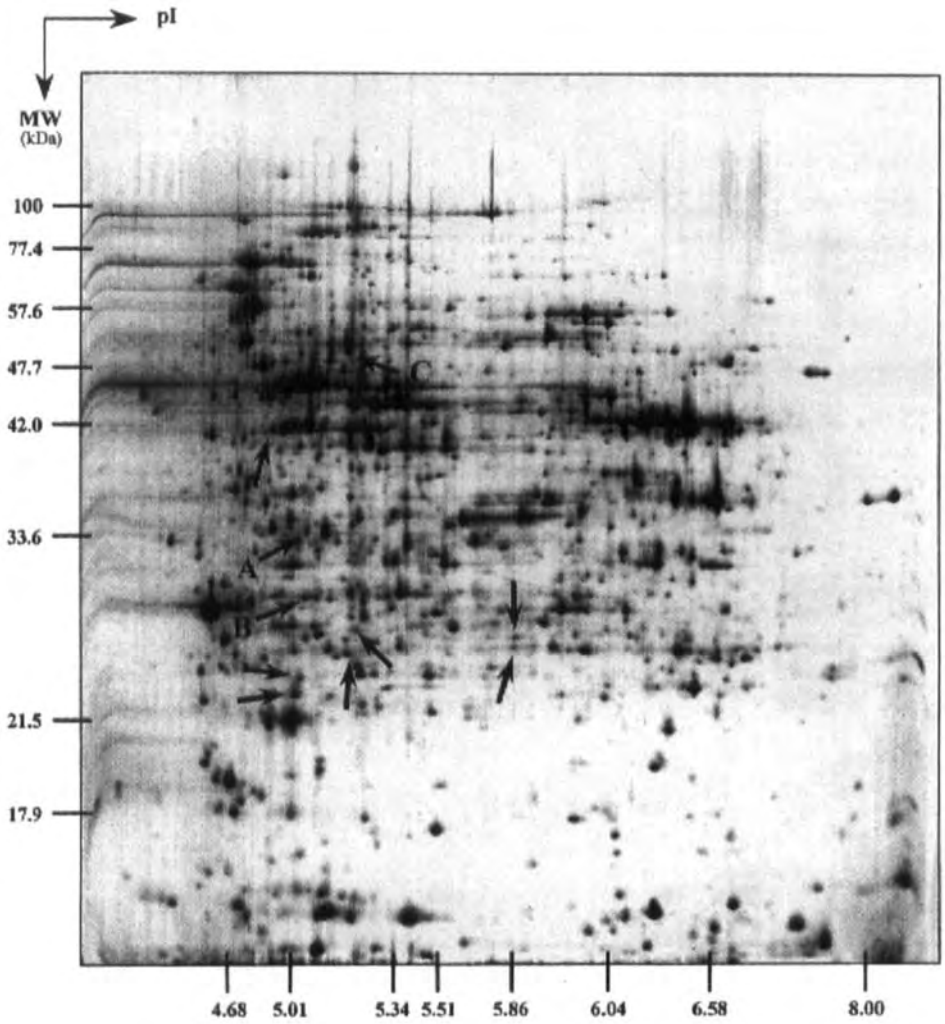


Figure 7. Proteome of GS-expressing *E. coli* grown on M9+glucose medium.

5 References

[1] Cohn, W. E., Moldave, K. (Eds.), *Progress Nucleic Acid Res. Mol. Biol.* 1994, *47*, 331–370.

[2] Neidhardt, F. C. (Ed.), *Escherichia coli and Salmonella Cellular and Molecular Biology*, American Society for Microbiology Press, Washington DC 1996, *Vol. 1*, pp. 1458–1496.

[3] Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasnaik, B., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella Cellular and Molecular Biology*, American Society for Microbiology Press, Washington, DC, 1987, *Vol. 1*, pp. 1358–1385.

[4] Ross, W., Thompson, J. F., Newlands, J. T., Gourse, R. L., *EMBO J.* 1990, *9*, 3733–3742.

[5] Gaal, T., Gourse, R. L., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 5533–5537.

[6] Cole, J. R., Olsson, C. L., Hershey, J. W. B., Grunberg-Manago, M., Nomura, M., *J. Mol. Biol.* 1987, *198*, 371–381.

[7] Yamagishi, M., de Boer, H. A., Nomura, M., *J. Mol. Biol.* 1987, *198*, 547–550.

[8] Neidhardt, F. C. (Ed.), *Escherichia coli and Salmonella Cellular and Molecular Biology*, American Society for Microbiology Press, Washington DC, 1996, *Vol. 1*, pp. 1417–1431.

[9] Ball, C. A., Osuna, R., Ferguson, K. C., Johnson, R. C., *J. Bacteriol.* 1992, *174*, 8034–8056.

[10] Thompson, J. F., de Vargas, L. M., Koch, C., Kahmann, R., Landy, A., *Cell* 1987, *50*, 901–908.

[11] Nilsson, L., Emilsson, V., *J. Biol. Chem.* 1994, *269*, 9460–9465.

[12] Nilsson, L., Verbeek, H., Vijgenboom, E., van Drunen, C., Vanet, A., Bosch, L., *J. Bacteriol.* 1992, *174*, 921–929.

[13] Johnson, R. C., Bruist, M. F., Simon, M. I., *Cell* 1986, *46*, 531–539.

[14] Kahmann, R., Rudt, F., Koch, C., Merlens, G., *Cell* 1985, *41*, 771–780.

[15] Johnson, R. C., Ball, C. A. Pferrer, D., Simon, M. I., *Proc. Natl. Acad. Sci. USA* 1988, *85*, 3484–3488.

[16] Nilsson, L., Vanet, A., Vijgenboom, E., Bosch, L., *EMBO J.* 1990, *9*, 727–734.

[17] Condon, C., Philips, J., Fu, Z. Y., Squires, C., Squires, C. L., *EMBO J.* 1992, *11*, 4175–4185.

[18] Gosink, K. K., Gall, T., Bokal, A. J., Gourse, R. L., *J. Bacteriol.* 1996, *178*, 5182–5187.

[19] Sambrook, J., Fritsch, E. F., Maniatis, T., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Press, New York 1989.

[20] Lee, K. H., Harrington, M. G., Bailey, J. E., *Biotech. Bioengineer.* 1996, *50*, 336–340.

[21] Harrington, M. G., Gudeman, D., Zewert, T., Yun, M., Hood, L. E., *Methods: Comp. Methods Enzymol.* 1991, *3*, 98–109.

[22] Gonzalez-Gil, G., Bringmann, P., Kahmann, R., *Mol. Microbiol.* 1996, *22*, 21–29.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 806–812

Petra Theresia Schindler
Franz Macherhammer
Sabine Arnold
Matthias Reuss
Martin Siemann

Institut für
Bioverfahrenstechnik,
Universität Stuttgart,
Stuttgart, Germany

Investigation of translation dynamics under cell-free protein biosynthesis conditions using high-resolution two-dimensional gel electrophoresis

A cell-free extract from *Escherichia coli*, generated through a routine procedure according to Chen and Zubay (*Methods Enzymol.* 1983, *101*, 674–690), was used for an *in vitro* protein synthesis. High-resolution two-dimensional gel electrophoresis (2-DE) was exploited to investigate the protein composition of the cell-extract and its dynamic development during a 24 h-period of cell-free protein synthesis performed in a membrane reactor device. Green fluorescent protein (GFP) was chosen as a target protein to be produced in a cell-free reactor because of its functional activity, which can easily be monitored by measurement of fluorescence, and because of its high sensitivity. GFP synthesis was observed by a standard fluorescence assay and was correlated to a quantitative assessment of the silver-stained GFP spot appearing on 2-DE gel maps. A constant protein synthesis rate was obtained for at least 8 h of process operation. While declining continuously, protein synthesis stopped entirely after 24 h. Both, the total protein content and total number of detectable spots were found to decrease over the reaction time, due to proteolytic digestion and protein precipitation. Certain proteins taking part in the translation process, such as the elongation factors (EF-Tu, EF-Ts) and the ribosomal protein RP-L9, were identified by Edman *N*-terminal sequencing and have thus been considered for reaction evaluation. The dynamics obtained during the entire process suggest that these translational factors were likewise affected by proteolytic decay.

Keywords: Green fluorescent protein / *in vitro* transcription/translation / *Escherichia coli* / S30-extract / Cell-free protein biosynthesis / Elongation factors EL 3420

1 Introduction

Cell-free transcription/translation – the *in vitro* synthesis of proteins using cellular extracts – has been a routine synthesis technique in molecular biology laboratories for several decades [1–3]. Most advantageous applications of this method include the production of toxic proteins, incorporation of amino acid derivatives into newly synthesized proteins, or directed evolution, based on ribosomal or polysomal display [4, 5]. Above that, cell-free synthesis systems comprise accessible model systems for investigations on protein maturation, genome/proteome research and cell therapy (*e.g.*, using antisense technology). However, despite its usefulness, the *in vitro* technique is inherently afflicted with a number of counterproductive reactions, *i.e.*, RNAse, DNAse and protease activity or the problem of sufficient energy charge maintenance; all of these – if not dealt with – can lead to decreased productivity and eventually to a collapse of the synthesis reaction. The cause of product limitations can be pinpointed and quantified using the tools of stoichiometric network analysis together with metabolic flux investigations, and dynamic modeling combined with sensitivity analysis [6]. The identification of key reactants and their interactions during cell-free transcription/translation makes it possible to derive model-based construction rules for the assembly of a tailored protein-synthesizing machinery. High-resolution two-dimensional electrophoresis (2-DE) has been shown to be a powerful tool for proteome analysis (proteomics). We have thus applied this technology for reaction analysis of the complex cell-free protein synthesis system in order to identify system-immanent bottlenecks limiting the efficiency of *in vitro* protein synthesis.

Correspondence: Dr. Martin Siemann, Institut für Bioverfahrenstechnik, Universität Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany
E-mail: siemann@ibvt.uni-stuttgart.de
Fax: +49-711-685-5164

Abbreviations: **EF**, elongation factor; **GFP**, green fluorescent protein; **RP**, ribosomal protein

2 Materials and methods

2.1 Cell-free biosynthesis reaction

Cell from *Escherichia coli* (strain A19) [10] were disrupted using a French press. S30-lysate was prepared according to Zubay [7]. Protein concentration was determined as described by Bradford [11]. The biosynthesis reaction was

carried out following the protocol described by Spirin [8] and Metzler [Roche Diagnostics, Germany, personal communications] in a DispoDialyzer membrane reactor (Spectra-Pores, USA) with a reaction volume of 1 mL (the dialysis volume was 5 mL) and a molecular weight cut-off of 10 kDa. Lysate concentration in the reaction volume was 20% v/v, which corresponded to 5.4 mg total protein. Plasmid pMGFP, which was a kind gift from A. Spirin (Institute of Protein Research, Pushchino, Russia), containing the structural gene for green fluorescence protein (GFP), was added to obtain a final concentration of 15 μg/mL. After 0, 2, 4, 8, and 24 h of incubation, two identical samples of 20 μL were removed from the system for further analysis. After storage at 4°C for at least 12 h, the first sample volume was used for activity determination of GFP in a fluorometer (Kontron Instruments, Fluorometer SMF 25) at the excitation of 485 nm and an emission wavelength of 530 nm. For quantification, the measured activities were correlated against a GFP standard of known concentration (Standard recombinant GFP; Cat. No. 1814524 Roche Diagnostics). The second sample volume was immediately frozen at –80°C for later use in high-resolution 2-DE. All samples for 2-DE analysis were collected, thawed on ice, and afterwards desalted in Microcon Microconcentrators (molecular weight cut-off 10 kDa; Millipore, Bedford, MA, USA). Samples were resuspended in an appropriate buffer according to Bjellqvist *et al.* [12]. Equal volumes of 124 μL were loaded onto first-dimensional gels. The protein amount loaded on the initial gel was 70 μg for the sample taken at 0 h of reaction time.

2.2 2-DE

High-resolution 2-DE was performed as described by Görg *et al.* [13]. The equipment for electrophoresis (Multiphor II electrophoresis unit, IsoDalt system, EPS 2A200 and 3500XL power supplies, automated gel stainer, MultiTemp thermostatic circulator, PlusOne silver staining kit, Immobiline dry strip kit) was supplied from Pharmacia Biotech (Uppsala, Sweden). Precast Immobiline dry strip gels used for the first dimension were 18 cm long, exhibiting a nonlinear pH gradient reaching from 3.5 to 10. The equilibration procedure was performed according to Westermeier [14]. The dimensions used for homogenous 13% SDS-PAGE gels were 25 × 20 cm, with 1.5 mm spacers. An automated gel stainer was used according to the instructions supplied by the manufacturer, in order to obtain the most reproducible silver-staining procedure.

2.3 Image analysis

Silver-stained gels were scanned with a Sharp JX-330 color image scanner at 400 dpi using the software LabScan 2.01 (Pharmacia Biotech). The scanner was calibrated with a DeskTop scanning package (Pharmacia Biotech) including the film scanning unit JX-3F6 (Sharp) and the photographic step tablet No. 2 (21 steps, density range approximately 0.05–3.05, Eastman Kodak Company, Rochester, NY, USA). Analysis of the digitalized gel images, including spot identification, editing, and volumetric quantification, was performed using ImageMaster 2D Elite Software, version 2.00 (Pharmacia Biotech).

2.4 Identification of spots

Proteins were identified on the 2-DE gels by blotting and subsequent *N*-terminal sequencing (Edman degradation) or MALDI-MS fingerprinting. For additional calibration of gel maps according to their molecular weight and p*I* values, 15 proteins, which were evenly distributed on the gel, were isolated from the gel and identified by MALDI-MS fingerprinting (data not shown). Amino acid sequences were aligned using the TagIdent data base (http://expasy.hcuge.ch/sprot/tagident.html). The analyzed *N*-terminal sequence obtained for EF-Tu was KEKFERTKPF (aligned to ID-No. P02990, SwissProt). For EF-Ts it was ITASLVKELR (ID-No. P02997, SwissProt), while RP-L9 gave an *N*-terminal sequence of MQVILLDKVAN (ID-No. P02418, SwissProt).

2.5 Detection of proteolytic activity

For demonstration of the proteolytic potency of the lysate, a commercially available protease activity detection system based on a protease-sensitive fluorescent label was applied, following the manufacturers manual (EnzChek-Kit, Molecular Probes, Leiden, Netherlands). At room temperature, cell extracts with final protein concentrations of 5–1500 μg/mL were incubated for 0, 2 and 6 h. The increase of fluorescence due to proteolysis was measured at 485 nm excitation and 530 nm emission. The reduction of proteolytic decay was investigated using two different protease inhibitor mixes (BM1 and BM2), both a kind gift from Roche Diagnostics (Penzberg, Germany). To the reaction mixture described in Section 2.1, which had a total volume of 200 μL, 10 μL of the respective protease inhibitor was added. A reference experiment was carried out without any addition of protease inhibitors. After 6 h of incubation at 30°C, samples were taken for 2-DE image analysis.

3 Results and discussion

3.1 Cell-free synthesis of GFP

The cell-free system studied in this work was based on an S30-extract from *Escherichia coli*, performed according to a routine procedure described by Zubay [7]. The cell-ex-

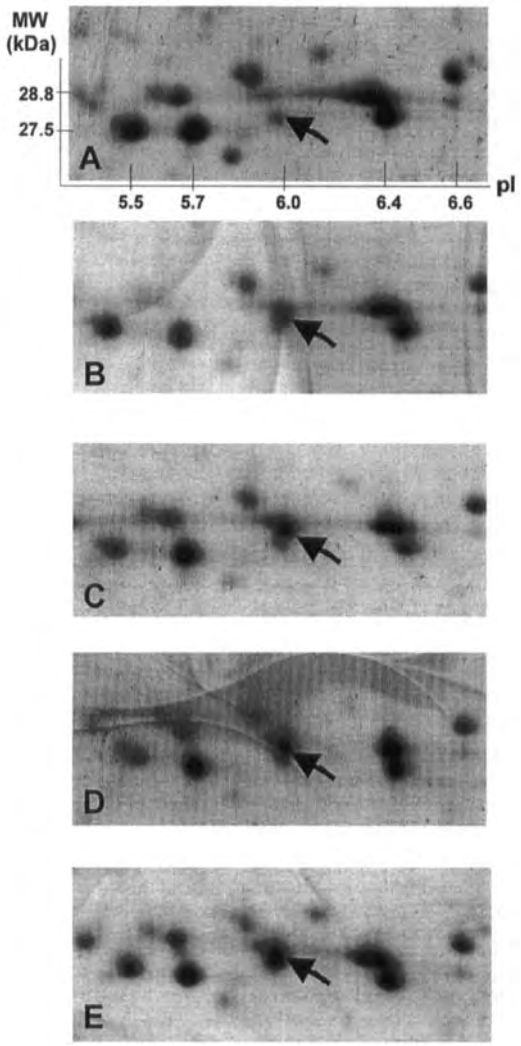


Figure 1. Cell-free biosynthesis of GFP in a protein bioreactor and observation of its production dynamics through identification in a silver-stained 2-DE gel map. The respective GFP spot is indicated by an arrow. (A) reaction beginning at 0 h, (B) after 2 h, (C) 4 h, (D) 8 h, and (E) after 24 h of incubation at 30°C, respectively. In contrast to the majority of proteins, a permanent increase of this product spot could be observed during the entire incubation period (see also Table 1). Data of the quantitative evaluation of GFP spot intensities (spot volumes) are given in Fig. 2 and Table 1.

tract was supplied with a feeding solution containing all necessary nucleotides (ATP, CTP, GTP, and UTP), all natural amino acids, all species of tRNAs and an endogenous energy-regenerating system (acetylphosphate/acetate kinase) that recovers nucleoside triphosphates from the corresponding diphosphates [1, 8]. Plasmid pMGFP, containing the structural gene for GFP [9], was added to this reaction mixture as a template for transcription. Since the structural gene is under the control of a T7 promoter, *in vitro* transcription was initiated by the addition of T7-RNA polymerase. Samples from the coupled transcription/translation reaction were taken over a period of 24 h and separated on 2-DE gels. These gels were examined according to the dynamics of key factors involved in cell-free protein biosynthesis. GFP was used as a model protein, because the product activity can be sensitively monitored in a standard fluorometic assay. Together with the proteins present in the complete reaction system, the concentration of GFP could also be calculated from the size, shape and optical density of silver-stained protein spots on the respective 2-DE gel map. Quantification of protein concentration was subsequently performed with commercially available image analysis software. The results were compared with a standard fluorometric measurement of GFP, aiming at the quantitative correlation between these two analytical tools.

The coordinates characteristic for GFP on the corresponding silver-stained 2-DE gel maps are given by the SwissProt database to a molecular mass of 26.9 kDa and an isoelectric point of 5.7. This information was used for identification of the GFP spot in Fig. 1 and 3. A purified external standard of GFP was additionally used to support the identification of the *in vitro* synthesized protein. Experimentally determined values of the molecular mass of GFP from three independent gel maps were in a range of 24.6–27.8 kDa and showed a p*I* of 5.7–6.0. The time-dependent optical density change of the GFP spot reflected the time-dependent GFP formation. The optical density of spots is referred to as 'spot volume'. The data from image software analysis are given in Table 1 and Fig. 2. GFP concentration increased continuously for at least 24 h of the cell-free synthesis process (Fig. 2A). A significant but rather low initial signal of 2% (initial GFP spot volume) was nonetheless obtained, which was presumably caused by a time delay during sampling and preservation. However, measurements from replica plates (reaction performed with omission of a GFP-encoding plasmid) indicated a negligible background signal of 0.02% at this very spot position (data not shown).

3.2 Comparison of detection methods for GFP

Fully matured GFP shows sensitive fluorescence activity, and thus its functional activity can easily be monitored over the course of *in vitro* translation. Fluorescence activity can be correlated with protein concentration by use of a purified external GFP standard of known protein concen-

Table 1. Evaluation of silver-stained spots on 2-DE gel maps during cell-free synthesis of GFP by use of a DispoDialyzer membrane reactor[a)]

Reaction time (h)	Total number of spots	Total spot volume (%)	Spot volume of GFP (%)	Spot volume of EF-Tu (%)	V_{spec}	Spot volume of EF-Ts (%)	V_{spec}	Spot volume of RP-L9 (%)	V_{spec}
0	660	100	2	100	1.0	100	1.0	100	1.0
2	521	76	8	97	1.3	i.r.		96	1.3
4	541	67	40	i.r.		77	1.2	87	1.3
8	432	64	49	79	1.2	71	1.1	92	1.4
24	516	68	100	68	1.0	72	1.1	i.r.	

a) Identification and quantification of spots were performed using ImageMaster software (Pharmacia Biotech) and were calculated according to their respective optical density (spot volume, units)

V_{spec}, specific spot volume (spot volume of protein species per respective total spot volume)

i.r., insufficient resolution due to interference with other protein spots; values are omitted

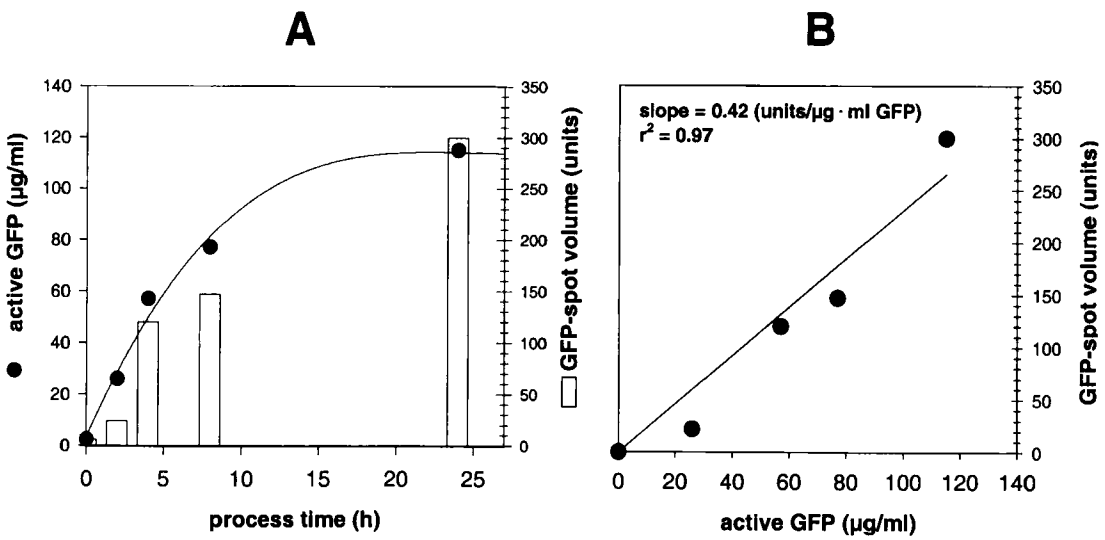


Figure 2. Comparison of two analytical methods for the detection of GFP. (A) Vertical white bars represent the total spot volume of GFP in the 2-DE gel, measured as raw spot volume in pixels with background subtracted. The black symbols reveal the time-dependent occurrence of fluorometric activity of GFP measured at 530 nm, which was correlated to an external GFP standard in order to give concentrations of active protein (µg/mL). (B) Correlation between GFP concentration and GFP spot volume within a reaction time of 24 h. The correlation was estimated from linear regression and was 0.42 µg/mL GFP per unit spot volume with a regression coefficient of $r^2 = 0.97$.

tration. Measured protein concentration was correlated to the estimated corresponding spot volume obtained from the 2-DE gel maps. The respective values from fluorescence measurements and spot-volume determination are found to correlate sufficiently well during the 24 h of synthesis (Fig. 2A and B). A correlation factor of 0.42 µg/mL GFP per units of spot volume with a regression coefficient, $r^2 = 0.97$, was obtained in Fig. 2B. Systematic errors of spot volumes, which might be due to slight variations of performance during the 2-DE process [15], were minimized, since all gels of one experiment were simultaneously made from the same stock solution and subsequently run in parallel using the IsoDalt system and automated gel-staining procedure (see also Section 3.5 and Table 2).

The maturation time of GFP is known to be long, independent of *in vivo* [16] or *in vitro* conditions. According to Kolb (Institute of Protein Research, Pushchino, Russia), the GFP encoded on the plasmid pMGFP is fully functionally active after at least 4 h (unpublished data). Thus, all samples had to be preincubated for more than 4 h at moderately (4°C) low temperature, to ensure full maturation of the protein before measurement. This inevitable, pro-

Table 2. Estimation of standard errors from independent measurements[a)]

Protein	Spot volume	Mean spot volume	Standard deviation Absolute	(%)
RP-L9	Gel No. 1: 3706 Gel No. 2: 3563 Gel No. 3: 3434	3568	136	3.8
EF-Ts	Gel No. 1: 1649 Gel No. 2: 1648 Gel No. 3: 1749	1682	58	3.4
EF-Tu	Gel No. 1: 2665 Gel No. 2: 2726 Gel No. 3: 2699	2697	31	1.1

a) Three independent samples of the same crude S30 lysate were simultaneously separated on three different 2-DE gels, matched, and evaluated by use of the ImageMaster software. Systematic errors were minimized since the entire procedure for gel preparation, separation, and staining was performed simultaneously using the IsoDalt device. The average standard errors were all below 5%, and thus demonstrate sufficient reproducibility of the procedure.

longed incubation of GFP, seemed nonetheless surprisingly unproblematic, since preliminary data revealed a sufficient resistance of GFP against, for instance, proteolytic decay (unpublished results). Nonetheless, for quantification of the total synthesis of GFP under *in vitro* conditions, the maturation rate, proteolytic decay, and amount of fully active protein have to be balanced. The quantification of spot volumes might bear additional sources of variation due to possible occurrence of nonlinearity in the silver staining process [15, 17]. However, Giometti *et al.* [18] previously described a linear relationship between the absolute amount of proteins loaded on a 2-DE gel and the respective final spot volume for at least 80% of human leucocyte proteins. Similar findings are – to the best of our knowledge – not described for prokaryotic systems, and thus it is still unclear whether these findings are equally valid for *Escherichia coli* proteins. Since most of the above discussed parameters are unknown for the described system, further investigations, aiming at their quantitative determination, are needed.

3.3 Dynamic proteome development during cell-free biosynthesis

After initiating cell-free biosynthesis, a considerable sample volume was taken from the reactor device and was later subjected to high-resolution 2-DE. The total amount of proteins (660 spots) served as a reference for the evaluation of all samples taken over the course of the GFP-synthesis reaction (see Table 1 and Fig. 3A). After 2 h of incubation, 290 spots out of the 521 remaining spots were already reduced to 75% of their initial intensity. After 4 h, the intensity of 205 spots had even dropped to 50% of their initial value (Table 1). Obviously, a vast protein decay occurred during the first hours of incubation, which might have been caused by protein precipitation (data not shown) and proteolytic decay. After 8 h of incubation, the total number of initial spots decreased again to nearly half of the initial value, while some novel spots appeared, mainly in the alkaline region of the gel map (see also marked zone in Fig. 3B). This phenomenon is further reflected by a rise of total number of spots from 432 at 8 h to 516 after at least 24 h of incubation. This increase is significant because it is greater than can be explained simply by measurement uncertainty.

Only proteins within a range of 10–100 kDa could be resolved on the 2-DE gel. Since an increasing occurrence of high molecular weight spots over the entire process time was obtained, this might suggest that in the functionally active crude extract, a significant amount of proteins greater than 100 kDa was originally present, but was beyond the resolution range of the prepared gel. These proteins might have been digested to lower size during prolonged incubation, and were thus finally appearing on the gel map. Polypeptides of low molecular mass (especially those smaller than 10 kDa), which might occur during proteolytic decay, will not appear on the gel, since these will be removed either during the process (cut-off of the dialyzer membrane was 10 kDa) or sample preparation (ultrafiltration by use of a 10 kDa membrane). The proteolytic potency of the lysate was qualitatively determined using a commercially available protease activity detection system based on a protease-sensitive fluorescent label. According to the test procedure described in Section 2.5, a high proteolytic activity of the tested S30 lysate used for *in vitro* protein biosynthesis was obtained (data not shown). The addition of two different kinds of protease inhibitor mixtures reduced, but did not entirely prevent, the general proteolytic potency of the cell extract.

3.4 Consideration of proteins directly involved in protein biosynthesis

The dynamic behavior of three proteins particularly involved in translation were observed during the entire process. These were the elongation factors Tu (EF-Tu, 44 kDa, p*I* 5.4), Ts (EF-Ts, 30 kDa, p*I* 5.1), and the ribosomal protein L9 of the 50S subunit (RP-L9, 16 kDa, p*I* 6.2). These proteins have been further identified on the gel map by blotting, subsequent *N*-terminal sequencing and final sequence alignment (see also Fig. 3A and B). Relative units of spot volume, which are defined as the ratio of the absolute value of spot volume to the initial spot volume, are summarized in Table 1. Due to the proteolytic

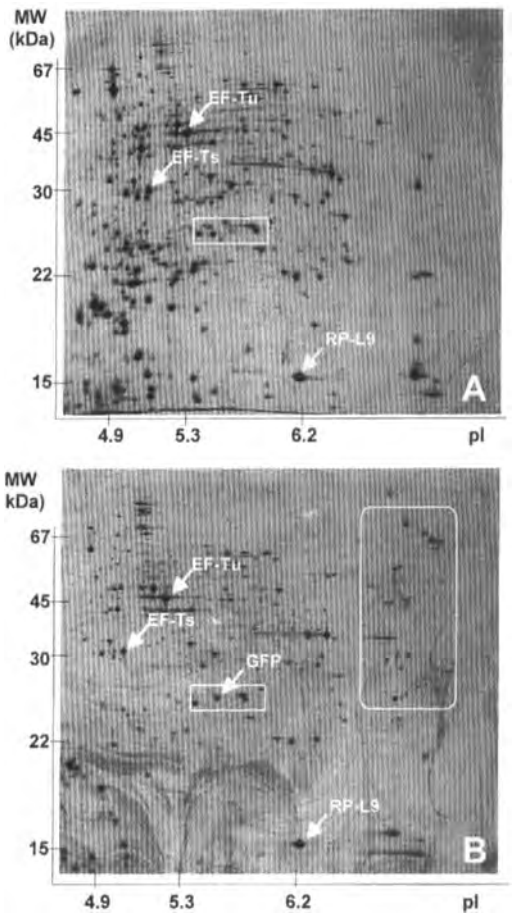


Figure 3. Silver-stained 2-DE protein patterns of the lysate from *Escherichia coli* strain A19 used for cell-free protein biosynthesis of GFP in a protein bioreactor. (A) Protein pattern immediately after initiation of protein synthesis (0 h of incubation). The marked section indicates the region around the GFP spot, especially highlighted in Fig. 1. (B) Protein pattern after 24 h of prolonged incubation under process conditions. Presentation of gels maps after 2, 4, and 8 h is omitted. The marked section shows a region covering the appearance of new spots. Identified spots are denoted with an arrow and were: GFP (26.9 kDa/p*I* 5.7), elongation factors EF-Tu (44 kDa, p*I* 5.4) and EF-Ts (30 kDa, p*I* 5.1), and ribosomal protein RP-L9 (16 kDa, p*I* 6.22).

decay, all values of the above-mentioned proteins decreased over the time course of the process. EF-Tu, which is responsible for the association of amino-acylated tRNAs to the ribosomal complex, was equally reduced by proteolytic decay (68% of the initial EF-Tu concentration remained in the system), and thus behaved similarly to the majority of the investigated lysate proteins (specific spot volume of EF-Tu, Vspec = **1**). The same was true for EF-Ts, which catalyzes the regeneration of EF-Tu during protein synthesis. Vspec is defined as the relation of the spot volume of an individual protein related to the total spot volume. Thus, a value equal to 1 indicates that none of the proteins is selectively removed, while values below 1 would indicate a specific disappearance of the individual protein. In contrast, a value greater than 1 denotes a further concentration of a certain protein. This, however, was only obtained for the ribosomal protein L9, which might be stabilized in the complex ribosomal structure itself, and might thus be protected from proteolysis (remaining specific spot volume of RP-L9, Vspec = 1.4).

3.5 Reproducibility

As stated in Sections 2.2 and 3.2, systematic errors (mainly caused by incorrect pipetting, or general handling of the procedure) can be reduced to a minimum when the measurements are performed from simultaneously prepared gels. In order to estimate these standard errors, three samples of the same crude S30 lysate were independently prepared and afterwards simultaneously separated on three different 2-DE gels using the IsoDalt device. These gels were matched and evaluated by use of the ImageMaster software. The standard errors given in Table 2 were below 5%, and thus demonstrated sufficient reproducibility of the procedure.

4 Concluding remarks

(i) GFP, which was chosen as target protein due to its high sensitivity and ease of measurement, was successfully synthesized under *in vitro* conditions over a period of 24 h by use of a dialyzer membrane reactor device. The synthesized functional protein was quantified according to its fluorescence abilities. These data correlated well with values received from quantification by use of silver-stained high-resolution 2-DE gel maps. Despite this analytical advantage, the application of GFP as a reporter protein seemed to be unfavorable, particularly for the purpose of reaction analysis and reaction engineering as long as GFP is afflicted with a prolonged maturation time.

(ii) Most critical for the performance of *in vitro* protein synthesis is the occurrence of proteolytic decay and precipitation during the entire reaction process. This was both confirmed from analysis using high-resolution 2-DE and from proteolytic assays. Both phenomena will lead to reduced productivity of the *in vitro* reaction system and should thus obviously be avoided for further system optimization.

(iii) The obtained dynamic behavior of the identified ribosomal protein L9 and both elongation factors Tu and Ts showed no indication of being a limiting step during *in vitro* synthesis, since these seemed to be maintained over the course of the synthesis process. Obviously these initially identified proteins cannot represent all of the majority of important factors of the *in vitro* protein biosynthesis. Nonetheless, since high-resolution 2-DE is a powerful analytical tool for observing the dynamics of translation (proteomics), a more extended view, mainly concentrated on further identification of the majority of ribosomal proteins, translation factors (*i.e.*, all initiation, elongation, and release factors) and amino acryl-tRNA synthetases, will be taken into account in future investigations. Especially the analysis and establishment of a well-balanced, stoichiometric value of all these factors might the helpful for further improvement of the *in vitro* protein synthesis machinery.

The authors would like to honor all members of the participating groups in the joint project on in vitro protein biosynthesis by use of bioreactors, project number FKZ 0 311 302, represented by Dr. Gerd Kleinhammer (Roche Diagnostics, Penzberg, Germany), Prof. Dr. Alexander Spirin from the Institute of Protein Research (Pushchino, Russia) and Prof. Dr. Volker Erdmann from the Institute of Biochemistry (Freie Universität Berlin, Germany). Special thanks are addressed to Dr. Thomas Metzler and Rolf Reichhuber from Roche Diagnostics for their assistance in cell-free GFP synthesis. The project was supported by the German Ministry of Research.

Received November 25, 1998

5 References

[1] Chen, H. Z., Zubay, G., *Methods. Enzymol.* 1983, *101*, 674–690.

[2] Kameyama, T., Novelli, G. D., *Biochem. Biophys. Res. Comm.* 1960, *2*, 292–396.

[3] Patnaik, R., Swartz, J. R., *BioTechniques* 1998, *24*, 862–868.

[4] Mattheakis, L. C., Bhatt, R. R., Dower, W. J., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 9022–9026.

[5] Tawfik, D. S., Griffiths, A., *Nature Biotechnol* 1998, *16*, 652–656.

[6] Mauch, K., Arnold, S., Reuss, M., *Chem. Eng. Sci.* 1997, *52*, 2589–2598.

[7] Zubay, G., *Annu. Rev. Genet.* 1973, *7*, 267–287.

[8] Spirin, A. S., in: Todd, P., Sidkar, S. K., Bier, M. (Eds.), *Frontiers in Bioprocessing II*, American Chemical Society, Washington, DC 1992, pp. 31–43.

[9] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., Prasher, D. C., *Science* 1994, *263*, 802–805.

[10] Gesteland, R. F., *J. Mol. Biol.* 1966, *16*, 67–84.

[11] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[12] Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.

[13] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[14] Westermeier, R., *Electrophoresis in Practice*, VCH, Weinheim 1997.

[15] Lottspeich, F., in: Lottspeich, F., Zorbas, H. (Eds.), *Bioanalytik*, Spektrum Verlag, Heidelberg 1998, pp. 815–827.

[16] Cormack, B. P., Valdivia, R. H., Falkow, S., *Gene* 1996, *173*, 33–38.

[17] Herbert, B. R., Sanchez, J.-C., Bini, L., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 13–30.

[18] Giometti, C. S., Gemmell, M. A., Tollaksen, S. L., Taylor, J., *Electrophoresis* 1991, *12*, 536–543.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 813–817

Jörg Deiwick[1,2]
Michael Hensel[1]

[1]Lehrstuhl für Bakteriologie, Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie der Ludwig-Maximilians-Universität München, Munich, Germany
[2]FB Biologie, Chemie, Pharmazie, Freie Universität Berlin, Berlin, Germany

Regulation of virulence genes by environmental signals in *Salmonella typhimurium*

Sensing and responding to environmental signals is a crucial element of bacterial pathogenicity. For a successful progression of infection, virulence gene expression is coordinated in response to habitat-specific environmental signals from the host organism. We are interested in identifying environmental cues affecting the expression of genes within *Salmonella* Pathogenicity Island 2 (SPI2), a virulence locus important for systemic infections by *S. typhimurium*. We describe our approach starting with the identification of new virulence genes, and analysis of the regulation of these genes by environmental signals leading to the proteome analysis in order to define the SPI2 regulon.

Keywords: *Salmonella* virulence / Environmental signals / *Salmonella* Pathogenicity Island 2
EL 3411

The expression of bacterial virulence genes during infection of a host organism is regulated in a spatial and temporal fashion. This observation has become the basis of several methods for the identification of virulence genes. *Salmonella typhimurium*, a facultative intracellular pathogen of animals and humans, has been used frequently as a model organism to study host-pathogen interactions and for the evaluation of new *in vivo* selection technologies for the identification of virulence genes. In the model system of murine salmonellosis, the function of a large number of metabolic, regulatory, and specific virulence genes is required for pathogenesis (for review see [1]). Hallmarks of *Salmonella* pathogenicity are the invasion of nonphagocytic cells and the ability to survive and replicate within host cells. A large number of invasion genes encoding a type-III secretion system, secreted proteins, and regulatory proteins are located within '*Salmonella* pathogenicity island 1' (SPI1). The function of various other gene loci is required for intracellular pathogenesis of *Salmonella*.

Despite the detailed analysis of some stages of the interaction between *Salmonella* and the infected host, many aspects of the pathogenesis of salmonellosis are still not well understood. Three genetic approaches were devised to identify virulence factors important for the infection of the host organism. Both '*in vivo* expression technology' (IVET) [2] and 'signature- tagged mutagenesis' (STM) [3] use the infected host organism as a selective environment. The IVET approach selects for *in vivo* induced promoters, whereas STM selects for mutants unable to survive in the host due to transposon insertions in virulence genes. 'Differential fluorescent induction' (DFI) [4], a further approach similar to IVET, selects promoters that are activated by intracellular bacteria using green fluorescent protein (GFP) as a reporter and fluorescence-assisted cells sorting for selection of induced fusions. All these approaches result in the identification of a large number of potential virulence genes, whose molecular function now has to be elucidated. For example, mutants highly attenuated in virulence in the model of murine salmonellosis were isolated by STM [3]. Further characterization of these mutants led to the identification of '*Salmonella* Pathogenicity Island 2' (SPI2) [5]. SPI2 contains 31 genes encoding a second type-III secretion system (Ssa) and the two-component regulatory system (SsrAB). Type-III secretion systems of Gram-negative bacteria have a complex composition with more than 20 subunits located in the cytoplasm, the inner membrane, and the outer membrane (see [6] for review). *Salmonella* strains harboring mutations in *ssa* or *ssr* genes have severe defects in systemic infection of the host, and also show a reduced accumulation in macrophages. This observation indicates that SsrAB has a key role in the expression of virulence genes.

We set out to analyze the function of SPI2 for pathogenesis of salmonellosis, the regulation of SPI2 genes, and to identify proteins under control of the SPI2 regulon. While the infected host was obviously the proper environment for selection of new virulence genes, additional *in vitro* assays have to be applied to reveal the molecular basis of their function. An important clue for the understanding of the regulation of SPI2 genes came from the application of the DFI approach. This resulted in the identification of the

Correspondence: Jörg Deiwick, Lehrstuhl für Bakteriologie, Max von Pettenkofer-Institut, Pettenkoferstr. 9a, D-80336 München, Germany
E-mail: joergd@m3401.mpk.med.uni-muenchen.de
Fax: +49-89-51605223

Abbreviations: DFI, differential fluorescent induction; **IVET**, *in vivo* expression technology; **SPI**, *Salmonella* pathogenicity island; **STM**, signature-tagged mutagenesis

GFP fusion to a SPI2 gene which was highly induced inside the phagosome of infected cells, but not induced in common culture media [4]. The phagosomal compartment of infected cells represents an environment highly restrictive for bacterial growth. However, intracellular pathogens such as *S. typhimurium* respond to signals of the phagosomal compartment by expressing specific subsets of genes and they are able to survive and replicate in this environment. In order to characterize these specific signals and the chemical composition of the vacuolar lumen, intracellular bacteria harboring reporter gene fusions have been used recently. For an indirect estimation of the amounts of trace elements, Garcia del Portillo *et al.* [7] determined vacuolar amounts for Fe^{2+} and Mg^{2+} as 1 μM and 10–50 μM, respectively. Furthermore, *in vitro* analysis showed that the two component regulatory system PhoPQ responds to limitations in the amounts of divalent cations and mild acidic conditions [8]. PhoPQ has been characterized as a global regulator of *Salmonella* virulence. It was demonstrated that Mg^{2+} deprivation (*i.e.*, media containing 8 μM Mg^{2+}) is the signal sensed by PhoQ *in vitro* resulting in activation of the subset of genes of the PhoPQ regulon (*pho*-activated genes). In contrast, high concentrations of Mg^{2+} result in the repression of the subset of *pho*-repressed genes. Garcia-Vescovi *et al.* [9] showed the binding of Mg^{2+} to the extracellular part of PhoQ.

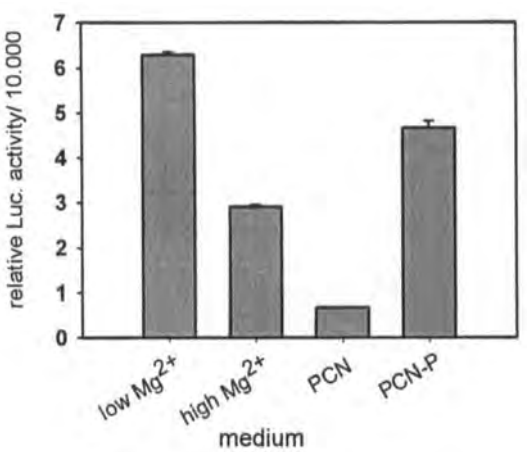


Figure 1. Effect of phosphate starvation and Mg^{2+} deprivation on SPI2 expression. The expression of a luciferase fusion to the *ssaB* gene of SPI2 was analyzed using the Promega luciferase assay kit according to the manufacturer (Promega, Heidelberg, Germany). Bacterial cultures were grown overnight in different media and luciferase activity of lysates was determined. Growth conditions: low Mg^{2+}, *N*-minimal media containing 8 μM Mg^{2+} [9]; high Mg^{2+}, *N*-minimal media containing 10 mM Mg^{2+}; PCN, MOPS-buffered minimal media without limitation in phosphate, carbon or nitrogen source [15]; PCN-P, MOPS-buffered minimal media with phosphate limitation.

In order to identify environmental signals inducing SPI2 expression we analyzed reporter gene fusions to SPI2 genes and used antibodies to monitor the amount of SPI2 gene products under various growth conditions. Low expression of SPI2 genes was detected after growth in rich media and minimal media containing high amounts of Mg^{2+}; however, expression of SPI2 genes was observed after growth in minimal media with limiting amounts of Mg^{2+} (Fig. 1). This observation indicates that SPI2 expression is under control of either the PhoPQ system, or a second regulatory system responding to low concentrations of Mg^{2+}. In addition, starvation for phosphate was identified as a second signal independently activating SPI2 expression (Fig. 1). Phosphate starvation is not commonly considered as an environmental signal for the induction of virulence genes. However, a gene encoding a phosphate transport protein has been identified as induced inside macrophages [4], indicating that the intracellular environment is also restricted in phosphate. Further analysis is required to determine the role of phosphate in gene regulation of intracellular pathogens. We observed that SPI2 mutant strain P8G12 is unable to respond to either of these inducing signals. This mutant strain harbors a transposon insertion in *ssrB*, encoding the transcriptional regulator of the two-component system of SPI2. *ssrB* is the terminal gene of a cluster of virulence genes in SPI2.

Based on the analysis of the expression of SPI2 within infected host cells and the identification of specific signals inducing SPI2 expression, we applied the proteome approach to the characterization of the SPI2 regulon. Previous applications of the proteome approach contributed to the understanding of the complex regulatory network of intracellular bacteria. Based on the pioneering work of Buchmeier and Heffron [10], Abshire and Neidhardt [11] analyzed the changes in expression patterns of intramacrophage *S. typhimurium* by selective radioactive labeling and compared the intracellular protein patterns to those obtained with several stress conditions *in vitro*. Several similarities were observed between protein expression inside macrophages and protein expression in culture media under starvation, *e.g.*, for phosphate. However, no distinct stress signal resulted in patterns of protein synthesis entirely matching the set of macrophage-induced proteins. The analysis of the regulatory mutant strain *pho-24*, constitutively expressing PhoPQ-activated genes, and a *phoP*$^-$ strain by two-dimensional gel electrophoresis (2-DE) revealed the presence of at least 40 proteins whose synthesis is PhoPQ-activated and repressed [12].

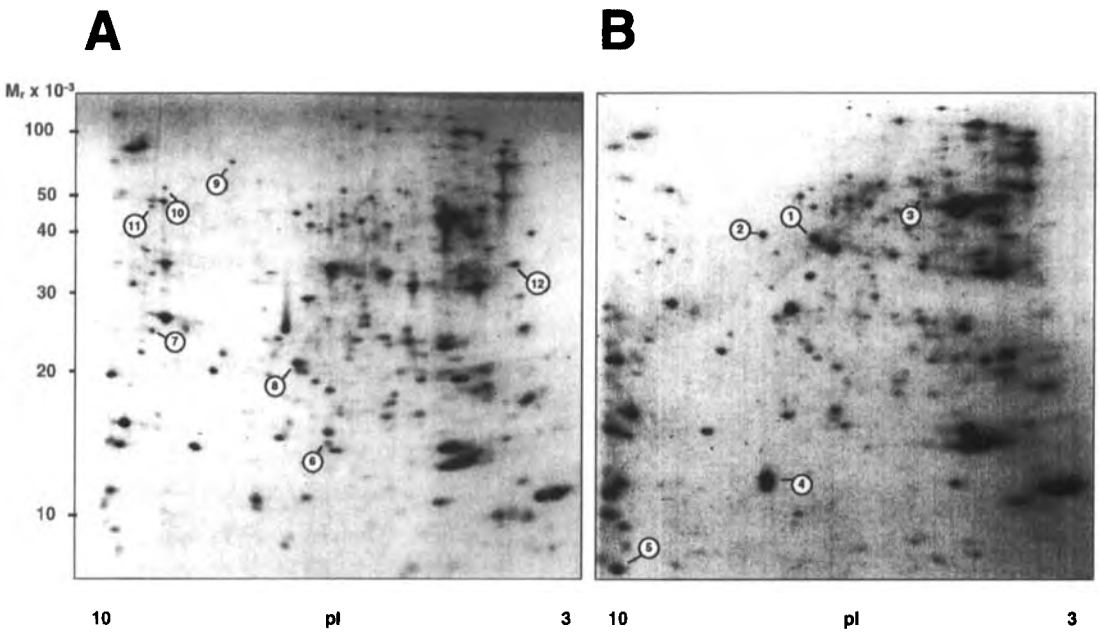


Figure 2. Intracellular labeling of (A) *S. typhimurium* wild-type and (B) the SPI2 regulatory mutant P8G12. Following infection of J774.1 (ATCC) macrophages (infection for 1 h, multiplicity of infection = 10) extracellular bacteria were killed by addition of gentamicin (20 μg/mL). Cytosolic protein biosynthesis was suppressed by addition of 200 μg/mL cycloheximide 0.5 h before labeling with 100 μCi/mL [^{35}S]methionine/cysteine (Amersham, Braunschweig, Germany). Bacteria were labeled from 9 to 12 h after infection. By addition of 0.1% Triton X-100, 10^7 cells were lysed. The Triton X-100 insoluble material was pelleted, washed, and separated using a pH 3–10 dry strip (Pharmacia, Freiburg, Germany) for the first and a 12% tricine gel for the second dimension according to published protocols [16–18]. Dried gels were exposed to X-Omat Ar films (Kodak) for 3 days. Protein spots with significantly different intensities between the strains are indicated by numbers.

The cell envelope proteins [13] and intra-macrophage-induced proteins [11] of *phoPQ* mutant strains of *S. typhimurium* were also characterized by 2-DE. To adopt 2-DE analysis to the investigation of the SPI2 regulon, we first evaluated the time course of SPI2 gene expression in the macrophage-like cell line J774A.1. Analysis of a transcriptional reporter fusion to an SPI2 secretion apparatus gene indicated highest expression of SPI2 about 6 h after infection of macrophages with *S. typhimurium* [14]. To evaluate the expression of SsrAB-regulated genes inside macrophages, we subsequently applied 2-DE on immobilized pH gradients (IPG) using radiolabeling of intracellular bacteria for periods of 6–9 h and 9–12 h after infection. Several differences in the patterns of protein synthesis were detected by comparison between *S. typhimurium* wild-type and the regulatory mutant P8G12 (Fig. 2). Several protein spots were absent in the 2-DE pattern obtained for P8G12 (Fig. 2A, spots 6–12). However, we also identified proteins preferentially labeled in P8G12 but absent in the wild-type (Fig. 2B, spots 1–5). These differences indicate that the SsrAB system has inducing as well as repressing effects on protein synthesis inside the host macrophage. Such differences in protein synthesis can either result from direct regulatory effects of the SsrAB system or from indirect effects of global regulatory systems such as PhoPQ responding to the intracellular environment.

For a more detailed characterization of the role of the SsrAB and PhoPQ systems for the regulation of SPI2-encoded proteins, we compared protein synthesis of *S. typhimurium* wild-type, SPI2-mutant strain P8G12 and mutants in the PhoPQ system under *in vitro* growth conditions to intracellular growth conditions. A *phoP*$^-$ mutant is unable to respond to PhoPQ-inducing signals and does not grow in media with low concentrations of Mg^{2+}. To allow a comparative analysis of the above strains under *in vitro* conditions, bacterial cultures were shifted from high to low concentrations of Mg^{2+} at mid-logarithmic growth phase. SPI2 gene expression was induced under these experimental conditions as demonstrated by West-

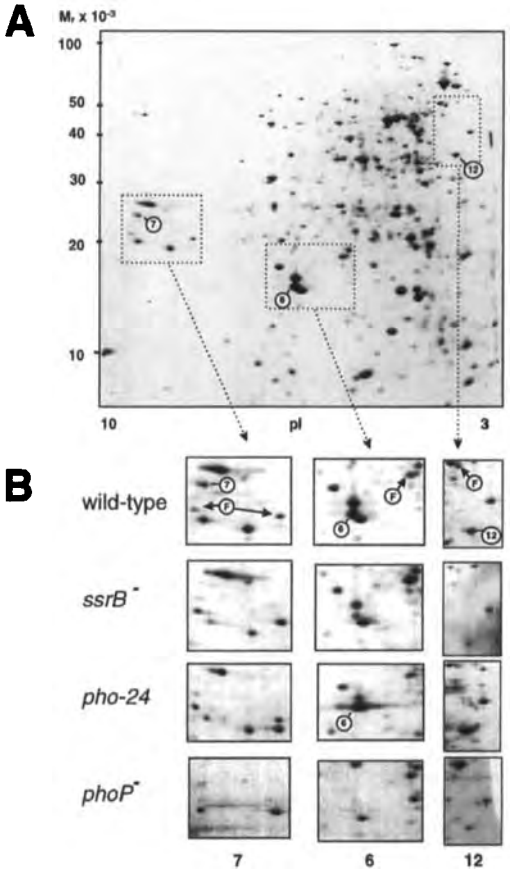


Figure 3. Role of SsrAB and PhoPQ. *S. typhimurium* wild-type (strain ATCC 14028), an SPI2 mutant strain harboring an mTn*5* insertion in *ssrB* (strain P8G12 [5]), a *pho-24* strain (CS022 [12]), and a *phoP*⁻ strain (CS015 [19]) were analyzed. Bacterial cultures were grown to mid-log phase (OD_{600} = 0.5) in minimal medium containing 10 mM $MgCl_2$ including 18 amino acids (0.005% each) without methionine and cysteine, pelleted, washed three times in minimal medium with 8 μM $MgCl_2$, and resuspended in fresh minimal media containing 8 μM $MgCl_2$. Cultures were incubated at 37°C, and aliquots were withdrawn in intervals of 1 h and labeled with 40 μCi [^{35}S]methionine/cysteine/mL for 10 min. At 2 and 3 h after the shift to low Mg^{2+} media, bacteria were harvested, washed and resuspended in denaturing solubilization buffer as previously described [16] and cells were lysed by two freeze/thaw cycles to obtain whole cell lysates. About 10 μg solubilized protein was separated by 2-DE as described in the legend to Fig. 2. Poor labeling of proteins was observed for the *phoP*⁻ strain 3 h after the shift. In (B) the respective areas of separation obtained with wild-type strain in (A) are shown. (Numbering of spots according to Fig. 2; 'F' denotes protein spots present in all samples serving as fixed points for orientation).

ern blotting using antibodies against SPI2-encoded proteins [14]. To evaluate changes in the overall protein synthesis, radiolabeling with [^{35}S]methionine/cysteine and subsequent 2-DE separation was performed. This analysis revealed that approximately 90% of the apparent proteins that are synthesized under intracellular conditions are also synthesized under the defined *in vitro* conditions, confirming previous observations of Abshire *et al.* [11] (see Figs. 2A, 3A). We identified three distinct proteins whose synthesis was significantly lower or entirely abrogated in the SPI2 regulatory mutant P8G12 (Figs. 2, 3; spots 6, 7 and 12). This effect was observed for intracellular bacteria as well as for *in vitro* cultures shifted to low Mg^{2+} concentrations. More dramatic differences in the 2-DE protein patterns were observed by comparison of *S. typhimurium* wild-type to the mutants strain *phoP*⁻ and *pho-24*. The synthesis of the three macrophage-induced proteins (spots 6, 7 and 12 in Fig. 3B) appeared to be affected by the regulatory systems SsrAB and PhoPQ to a different extent. These observations suggest a modulating effect of the PhoPQ system on the SPI2 regulatory system SsrAB. Alternatively, the expression of the corresponding genes may be directly affected by both systems or mutations in the SsrAB and PhoPQ result in alterations of the global regulatory network.

The characterization of proteins identified here may contribute to the understanding of the high attenuation of SPI2 mutants. In addition, analysis of the cell envelope fraction of *S. typhimurium* may help to provide further information on changes in protein levels induced by stress conditions with respect to the recently published reference map [13]. Finally, the separation and identification of the components of the translocation machinery of the type-III secretion system encoded by SPI2 may become feasible. Further comparative analyses of *in vitro* and *in vivo* conditions inducing SPI2 expression have to reveal changes in the 2-DE pattern of the regulatory mutant in SPI2. This will improve our understanding of regulation as a prerequisite for biological function of virulence factors of intracellular pathogens.

We described an experimental approach for the analysis of the SP2 regulon that may be applicable to the characterization of other virulence factors identified by techniques such as IVET, STM or DFI (Fig. 4). Applications of these techniques yield large numbers of candidate virulence genes whose function has to elucidated. As a next step towards the analysis of these new genes, the characterization of environmental signals or signals of the infected host affecting the expression of the new genes is required. This information is essential to define *in vitro* conditions for the subsequent application of the proteome

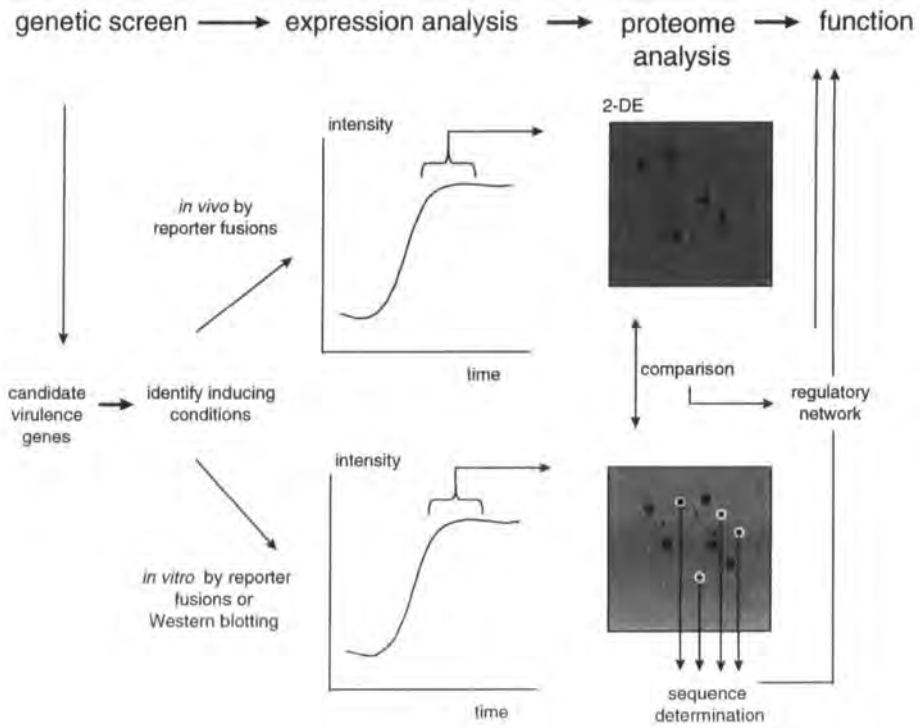


Figure 4. Schematic presentation of an approach to the characterization of candidate virulence genes.

approach. Based on such conditions, the proteome analysis will allow the identification and characterization of the corresponding proteins and contribute to the understanding of regulation and function of new bacterial virulence determinants.

This work was supported by DFG grant He1964/2-2 and a travel grant of the DGHM.

Received December 9, 1998

References

[1] Foster, J. W., Spector, M. P., *Annu. Rev. Microbiol.* 1995, *49*, 145–174.

[2] Mahan, M. J., Slauch, J. M., Mekalanos, J. J., *Science* 1993, *259*, 686–688.

[3] Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., Holden, D. W., *Science* 1995, *269*, 400–403.

[4] Valdivia, R. H., Falkow, S., *Science* 1997, *277*, 2007–2011.

[5] Shea, J. E., Hensel, M., Gleeson, C., Holden, D. W., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 2593–2597.

[6] Hueck, C. J., *Microbiol. Mol. Biol. Rev.* 1998, *62*, 379–433.

[7] Garcia del Portillo, F., Foster, J. W., Maguire, M. E., Finlay, B. B., *Mol. Microbiol.* 1992, *6*, 3289–3297.

[8] Miller, S. I., *Mol. Microbiol.* 1991, *5*, 2073–2078.

[9] Garcia Vescovi, E., Soncini, F. C., Groisman, E. A., *Cell* 1996, *84*, 165–174.

[10] Buchmeier, N. A., Heffron, F., *Science* 1990, *248*, 730–732.

[11] Abshire, K. Z., Neidhardt, F. C., *J. Bacteriol.* 1993, *175*, 3734–3743.

[12] Miller, S. I., Mekalanos, J. J., *J. Bacteriol.* 1990, *172*, 2485–2490.

[13] Qi, S. Y., Moir, A., O'Connor, C. D., *J. Bacteriol.* 1999, *178*, 5032–5038.

[14] Deiwick, J., Nikolaus, T., Erdogan, S., Hensel, M., *Mol. Microbiol.* 1999, in press.

[15] Neidhardt, F. C., Bloch, P. L., Smith, D. F., *J. Bacteriol.* 1974, *119*, 736–747.

[16] Görg, A., Postel, A., Gunther, W. S., *Electrophoresis* 1988, *9*, 531–546.

[17] Schägger, H., von Jagow, G., *Anal. Biochem.* 1987, *266*, 368–379.

[18] Rabilloud, T., Valette, C., Lawrence, J. J., *Electrophoresis* 1994, *15*, 1552–1558.

[19] Miller, S. I., Kukral, A. M., Mekalanos, J. J., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 5054–5058.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 818–825

Nelson Guerreiro
Michael A. Djordjevic
Barry G. Rolfe

Plant-Microbe Interaction Group, Research School of Biological Sciences, Australian National University, Canberra City, Australia

Proteome analysis of the model microsymbiont *Sinorhizobium meliloti:* Isolation and characterisation of novel proteins

Sinorhizobium meliloti is an agriculturally and ecologically important microbe due to its capacity to establish nitrogen-fixing symbiosis with plant legumes. Two-dimensional gel electrophoresis of total cellular protein was used to establish a proteome reference map for the model microsymbiont *Sinorhizobium meliloti* strain 1021. The extent of changes in the gene expression of cells grown in a defined medium at different growth phases was established. After examination of over 2000 resolved protein spots, a minimum of 52 reproducible changes in protein expression levels were detected when early exponential phase cells were compared to late exponential phase cells. In contrast, induction of nodulation gene expression by the addition of the flavonoid luteolin to cells did not result in detectable changes in protein expression at either early or late exponential phase. *N*-terminal microsequencing of eighteen unknown constitutive proteins plus four proteins, induced or up-regulated in late exponential phase cells, allowed the identification of proteins not previously described in rhizobia. These included an amide-binding protein, a putative hydrolase of the glyoxalase II protein family, a nucleoside diphosphate kinase, and a 5′-nucleotidase. *N*-terminal microsequencing was also valuable in revealing *N*-terminal post-translational processing and assigning a subcellular location to the analysed protein. Proteome analysis will provide a powerful analytical tool to complement the sequencing of the genome of strain 1021.

Keywords: Rhizobia / *Sinorhizobium meliloti* / *Rhizobium meliloti* / Protein identification / Amide-urea binding protein / Two-dimensional polyacrylamide gel electrophoresis EL 3359

1 Introduction

Microbes of the genera *Sinorhizobium, Rhizobium, Bradyrhizobium* and *Azorhizobium* (collectively termed rhizobia), can occupy several major environmental niches including the soil. In the soil, these rhizobia exist as saprophytes, whereas they can adopt a commensal or parasitic lifestyle when growing on the surface or inside plant root cells. *Sinorhizobium meliloti* (previously known as *Rhizobium meliloti*) can also form a symbiotic relationship with leguminous plants belonging to the genera *Medicago, Melilotus* and *Trigonella* where, in a differentiated state called a bacteroid or symbiosome, it is capable of fixing atmospheric nitrogen inside legume nodules (reviewed in [1, 2]). *S. meliloti* becomes capable of initiating infection on legume hosts once the genes required for infection are induced by plant-excreted flavonoids, such as luteolin [3].

S. meliloti is a model microsymbiont and efforts are now underway to completely sequence its genome. Although a complete genome sequence will reveal essentially all the putative open reading frames (ORFs), as well as operon structures, G–C content, origin(s) of replication, and possible regulatory circuits *via* the identification of common *cis*-acting regulatory elements, several questions concerning the function of higher order processes would remain. These include: which ORFs are authentic genes, what are the functions of the gene products, when and to what level are gene products expressed, where are the gene products localised and are the gene products post-translationally processed or modified? Functional genome analysis using high density oligonucleotide probe arrays (chip technology) has offered a promising method for the quantitative analysis of the transcriptional activity of bacterial genomes [4]. However, analysis of gene expression at the transcript mRNA level alone does not take into account the occurrence of gene regulation at the post-transcriptional level, *e.g.*, Anderson *et al.* [5, 6] and Haynes *et al.* [7] showed a general lack of correlation between the levels of mRNA and the resulting protein levels that occur in liver and yeast cells, respectively. This, along with the occurrence of post-translational processing and modification events in prokaryotes (as one mRNA can give rise to several protein products), would reinforce the necessity for direct determination of expressed proteins. Therefore, a comprehensive understanding of cellular activity will

Correspondence: Dr. Michael Djordjevic, Plant-Microbe Interaction Group, Research School of Biological Sciences, Australian National University, P.O. Box 475, Canberra City, A.C.T 2601, Australia
E-mail: michael@rsbs.anu.edu.au
Fax: +61-2-6249-0754

only arise if DNA sequence information and mRNA transcriptional patterns are complemented with gene expression analysis at the protein level, that is, by linking a protein to the protein encoding gene.

Proteome analysis has rapidly become a powerful tool for investigating global changes in the gene expression program of prokaryotic organisms and for understanding the complex regulatory networks which coordinate gene expression. This approach, utilising high-resolution two-dimensional gel electrophoresis (2-DE), provides a snapshot of the protein gene products expressed in a cell population at a given time under a defined physiological conditions. A proteome has been defined as the protein complement expressed by a genome [8]. Proteomes are dynamic and reflect the activity state of a biological system, and the analysis of the proteome of an organism under different conditions can contribute to an holistic understanding of how a microbe reorganises and adapts to new environments. More recently, the combination of 2-DE with sensitive post-translational technologies, including *N*-terminal Edman microsequencing, amino acid composition analysis [9], peptide-mass fingerprinting and peptide sequencing *via* mass spectrometry-based approaches [10–12], and post-translational modification analysis [13, 14] has provided a rapid means of characterising the function or functional aspects of unknown gene products. Microbial proteomes are being determined for bacterial groups including *Haemophilus influenzae* [15, 16], *Spiroplasma melliferum* [11], *Bacillus subtillus* [17] and *Salmonella typhimurium* [18]. The most extensive of these is currently that developed for the K12 strain of *Escherichia coli* where proteins have been catalogued according to their expression and regulation pattern under defined stimuli [19–21].

This proteomic approach for analysing genomic expression and function has already been used to describe a *Rhizobium leguminosarum* bv *trifolii* strain ANU843 2-D reference map, the response of this strain to flavonoid exposure [22], and the identification of putative plasmid-encoded functions [23]. Results obtained have suggested the existence of global regulatory interactions occurring between plasmids and chromosomal replicons which would have been difficult to show with a nucleic-based technology. In this study we have initiated a proteome approach on *S. meliloti* strain 1021 to investigate the patterns of gene expression and regulation. We present the first 2-D protein map of *S. meliloti* and characterise the *N*-terminal sequences of newly identified proteins not previously described in rhizobia. The changes in the pattern of proteins synthesised during early exponential phase growth and late exponential phase growth, and in response to the flavonoid luteolin were also investigated.

2 Materials and methods

2.1 Growth conditions and sample preparation

The reagents and apparatus used have been described in detail [22]. Cultures of *S. meliloti* strain 1021 were grown in defined BIII medium [24] at 28°C. At an OD of 0.08 (late lag phase) at 600 nm, the *nod* gene inducer, luteolin was added to the bacterial cultures at a concentration of 5 μM. Incubation was continued for 6 h and 17 h to an OD of 0.3 (early exponential phase) and 1.2 (late exponential, entering stationary phase), respectively, at 600 nm, before harvesting. Control cultures were treated in an identical way, except that luteolin was omitted. Cells were harvested, washed and lysed as previously described [22]. Protein concentrations of the cell lysates were determined using a Bradford protein assay (Bio-Rad, Hercules, CA, USA) with BSA as the standard.

2.2 2-DE and *N*-terminal sequencing

2-DE and electrotransfer to PVDF were done according to previously described methods [22]. Isoelectric focusing in the first dimension was carried out on linear pH 4–7, 18 cm immobilised pH gradient (IPG) strips (Pharmacia-Biotechnology, Uppsala, Sweden) loaded with 100 μg of total cellular protein at the anodic application site and run for 200 kVh. For narrow-range focusing, linear pH 4.5–5.4, 11 cm IPG strips were focused for 46 kVh. Horizontal second-dimensional SDS-PAGE was carried out on ready-made gels (ExcelGel SDS, 12–14%T acrylamide, from Pharmacia-Biotechnology). For *N*-terminal sequence analysis of protein spots, 0.5–1 mg of total protein was loaded in a volume of 100 μL and the gel electroblotted onto PVDF membrane. *N*-terminal sequencing was done on a PROCISE-HT sequencer system or on a PROCISE-CLC for increased sensitivity (both machines from Perkin-Elmer Applied Biosystems, Foster City, CA, USA). Typically, 16 cycles of Edman degradation were performed on the selected spots and the sequences were used to search a nonredundant protein database (SWISS-PROT, PIR, TREMBL and GenPept) using the FASTA program in order to establish the identity of the polypeptides and assign a putative function. The theoretical isoelectric point and molecular mass of the matched sequence in the database and sequence alignment at the *N*-terminus of the corresponding homologous protein were also used to determine the significance of the matches.

2.3 Staining and image analysis

Proteins on analytical 2-D gels were visualised by silver staining [25] and digitised at 600 dpi with a UMAX PS-

2400X lamp scanner. Spot detection, gel alignment and gel-to-gel protein spot matching were performed with the use of the Melanie II 2-D image analysis software program (Bio-Rad). The apparent molecular masses of the proteins were extrapolated from standard protein markers coelectrophoresed with the sample using Melanie II.

3 Results

3.1 Analysis of the 2-D profile of *S. meliloti* strain 1021 proteins

The 2-D protein profile in Fig. 1 was selected as the reference proteome map for *S. meliloti* strain 1021. Approximately 2000 protein spots were separated on the silver-stained gel within the pH range of 4–7 and size range of 10–122 kDa (Fig. 1). The 2-D pattern was highly reproducible with intra- and inter-samples of strain 1021 protein preparations, since the protein patterns obtained were indistinguishable from one another on several 2-D gels. Narrow-range IPG strips were used for the first dimension of 2-DE to further resolve the proteins occurring between the p*I* range of pH 4.5–5.4 and a proteomic contig [26] representing a protein expression window was established (Fig. 2). A further 65 proteins were resolved in the pH range of 4.5–5.4 when using the narrow-range IPG strips (572 proteins detected) compared to the broader-range IPG strip (507 proteins detected).

3.2 Identification of constitutive proteins by *N*-terminal microsequencing and assignment to the proteome map of *S. meliloti* strain 1021

To initiate the development of a 2-D protein database for *S. meliloti* strain 1021, 18 abundant protein spots were selected for *N*-terminal amino acid sequencing. These protein spots are numbered in Fig. 1 and the homology of these protein sequences to a nonredundant protein database is shown in Table 1. Of the 18 amino acid sequen-

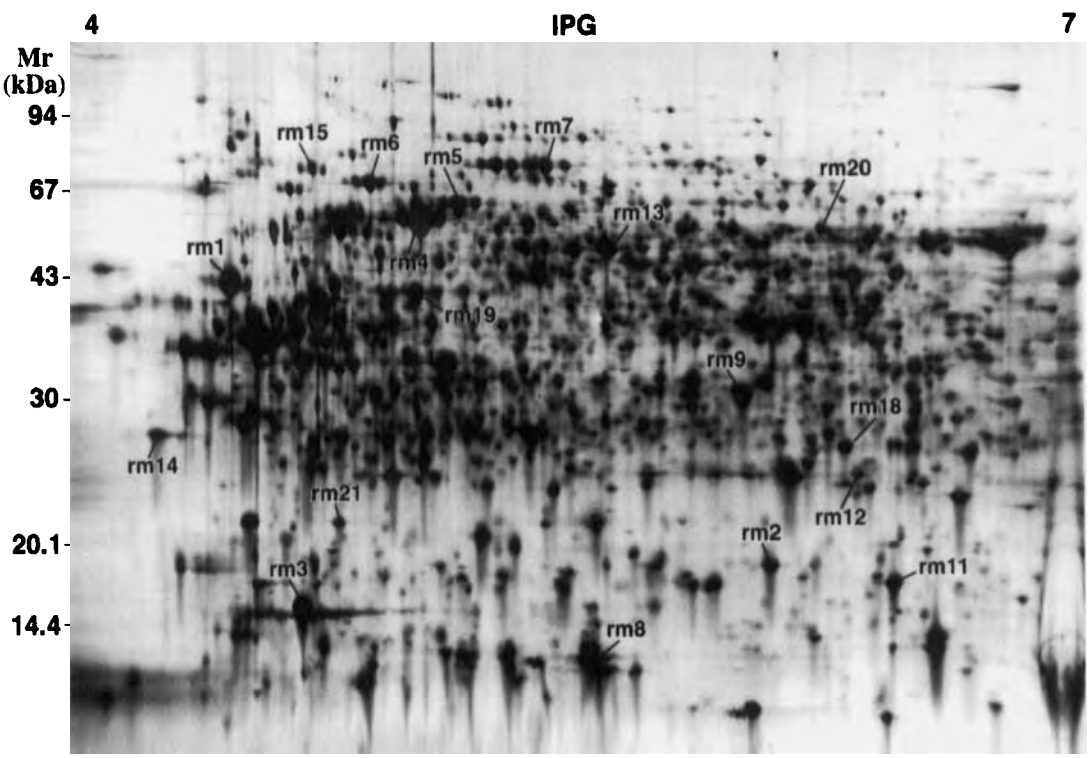


Figure 1. 2-D reference map of proteins synthesised in *S. meliloti* strain 1021 at early exponential phase. Proteins selected for *N*-terminal microsequencing were assigned arbitrary numbers. Isoelectric focusing in the first dimension was on IPG strips with a linear gradient ranging from pH 4–7 (18 cm) and loaded with 100 µg of total cellular protein. For the second dimension 12–14%T SDS-PAGE gels were used. Proteins were visualised by silver staining.

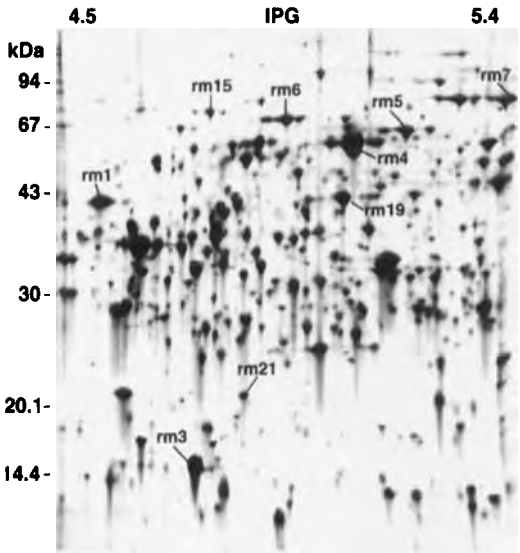


Figure 2. 2-DE gel image representing a protein expression window showing total cellular protein from *S. meliloti* between the p*I* range of 4.5–5.4. Narrow-range IPG 4.5–5.4 (11 cm) strips, loaded with 100 μg of protein, were used for the first-dimension separation of 2-DE.

ces obtained, six matched to the sequences of known *Rhizobium* and *Sinorhizobium* proteins. Three of these proteins were identified as the molecular chaperones GroES A (spot 8), GroEL A (spot 4), and DnaK (spot 6). These highly abundant chaperones assist in protein folding, the assembly of oligomeric protein complexes, and the export of proteins. Also identified were the 30S ribosomal protein (spot 5), the succinyl-CoA synthetase beta subunit (spot 19), which is involved in the tricarboxylic acid cycle, and a putative transcription elongation factor (spot 21).

Nine proteins were assigned putative functions based on their homology to known sequences from species other than rhizobia. These included spots 3 and 7 which matched to the 50S ribosomal protein L7/L12 from *Brucella abortus* and the elongation factor EF-G from *Agrobacterium tumefaciens*, respectively. Both proteins had also been previously identified on the proteome map of *R.l.* bv *trifolii* [22]. Also spots 3, 4, 6, 7 and 8 exhibited identical electrophoretic mobilities to their *R.l.* bv *trifolii* equivalents previously analysed by proteome analysis [22]. The *N*-terminal sequence of spot 2 displayed significant homology to the cytoplasmic nucleoside diphosphate kinase of *Rhodobacter sulfidophilus* which is involved in the synthesis of nucleoside triphosphates other than ATP. Spot 18 shared strong similarity to a variety of phosphoglycerate mutases which are involved in glycolysis. The *N*-termini of several protein spots (spots 1, 13 and 15) were most likely *N*-terminally processed since alignment began 24–51 residues downstream of the predicted methionine initiation site translated from the DNA sequence of the matched protein. Of these, spot 15 showed homology to a variety of 5′-nucleotidase proteins belonging to the 5′-nucleotidase family that catalyse the hydrolysis of phosphate esterified at the carbon 5′ of nucleotide molecules [27]. Consistent with this, spot 15 is most likely *N*-terminally processed since the first residue was 27 residues downstream from the predicted initiator methionine, as has been shown for other 5′-nucleotidases, and also indicative of proteins being targeted to the periplasmic space and outer membrane of bacterial cells [28]. Similarly, the *N*-terminal sequence of spot 13 matched to that of a surface-exposed outer membrane 38 kDa lipoprotein from *Pasteurella haemolytica* 24 residues downstream of the predicted initiator methionine. This putative 38 kDa lipoprotein is similar to the *E. coli* polyamine transport proteins PotD and PotF [29]. Also, the *N*-terminal sequence of spot 1 showed strong homology to some periplasmic amide-urea binding protein precursors which form part of an active transport system for short-chain amides and urea. Spots 11, 14 and 20 did not display any significant similarity to proteins currently available in the databases, indicating that they may be novel.

3.3 Application of proteome analysis for the identification of growth-phase-regulated proteins

We analysed the changes in the 2-D protein pattern in strain 1021 in response to two physiological conditions. First, we used proteome analysis to investigate the regulation of gene expression during different phases of cell growth. The expression pattern of proteins during early exponential cell growth was compared with that of late exponential cell growth. Figure 3 shows the expression pattern of proteins from late exponential phase cells. A minimum of 52 protein differences, defined as up- or down-regulated or induced or lost, were observed when comparing the protein expression patterns of *S. meliloti* in early exponential phase growth (Fig. 1) with that of late exponential phase growth. The *N*-terminal sequences of four up-regulated or induced proteins were determined and these are numbered in Fig. 3. The *N*-terminal sequences of these proteins are shown in Table 2. Spot i1 and i2 were highly abundant proteins that were not observed in early exponential phase cells and sequence data from these spots did not reveal any homology to database sequences. Protein spot u3 shared high sequence homology to a heat shock protein from *E. coli* which belongs to the small heat shock protein (HSP20) family. Members of the HSP20 family are induced by a variety of environmen-

Table 1. *S. meliloti* proteins analysed by *N*-terminal microsequencing and FASTA search of a nonredundant protein database

Spot	M_r/p*I*[e]	*N*-terminal sequence[a]	Homology (% identity)	Organism	M_r/p*I*[b]	SWISS-PROT[c] Accession No.
1[f]	41881/4.68	ADETIKVGILHSLSG	Hypothetical amide-urea binding protein (84.6%)	*Synechocystis sp.*	48359/4.82	P74390
			Amide-urea binding protein precursor (71.4%)	*Methylophilus methyotrophus*	41896/7.85	O50371
2	16921/6.15	AIERTFSMIKPDATK	Nucleoside diphosphate kinase (86.6%)	*Rhodobacter sulfidophilus*	15254/5.25	P95653
3	14479/4.88	ADLAKIVEDL	50S ribosomal L7/L12 protein (100%)	*Brucella abortus*	12546/4.79	P41106
4	54965/5.21	AAKEVKFGRS	GroEL A (100%)	*Sinorhizobium meliloti*	57673/5.11	P35469
5	60765/5.32	SATNPTRDDFAALLE	30S ribosomal protein S1 (100%)	*Sinorhizobium meliloti*	*62639/5.24*	*P14129*
6	66823/5.07	AKVIGIDLGT	DnaK (100%)	*Sinorhizobium meliloti*	69072/4.94	P42374
7	74114/5.56	AREKYIEDYRNFGIM	Elongation factor EF-G (100%)	*Agrobacterium tumefaciens*	78044/5.13	P70782
8	11691/5.68	ASTNFRPLHD	GroES A (100%)	*Sinorhizobium meliloti*	10587/5.43	P35473
9	30234/6.08	ALPDFTMRQLLEAGV	30 S ribosomal protein S2 (57.1%)	*Spirulina platensis*	27776/6.24	P34831
11	16025/6.48	ATRKTEDAFSLSFDP	Unknown			
12	22984/6.39	SEGVDLKELKRMMDG	Enoyl-CoA hydratase (53.3%)	*Trypanosoma cruzi*	NA[d]	Q95045
13[f]	48424/5.72	KAATPKETAAATGP	38 kDa lipoprotein (64.2%)	*Pasteurella haemolytica*	40147/5.05	Q51848
14	26329/4.47	SDLPDLGGKTVVVVT	Unknown			
15[f]	72323/4.91	DYELNILHINDLHSR	5′-Nucleotidase precursor (71.4%)	*Homo sapiens*	60823/6.34	P21589
18	25246/6.35	SGTLVLVRHGQSDWN	Phosphoglycerate mutase (73.3%)	*Agrobacterium tumefaciens*	*27215/5.38*	*Q11140*
19	40444/5.20	MNIHEYQAKALLKSY	Succinyl-CoA synthetase beta subunit (93.3%)	*Rhizobium leguminosarum*	NA[d]	O33526
20	52971/6.29	AVASKKIPAPDEVRI	Unknown			
21	19596/4.98	VDKVPMTQGGVNLQEE	Elongation factor homolog (70.5%)	*Rhizobium leguminosarum*	17444/4.88	O68546

a) *N*-terminal sequences are shown by using the single-letter code for amino acid residues.
b) The theoretical molecular mass (M_r) and isoelectric point (p*I*) calculated for the predicted amino acid sequence of the protein. The units for molecular mass are Da.
c) For proteins which have no SWISS-PROT accession number, the EMBL/GenBank accession number is given.
d) Unable to calculate the theoretical p*I* and M_r values for these proteins because only fragments of their sequence were available in the database.
e) Observed M_r and p*I* obtained from analysis of the 2-DE spot mobility.
f) Sequence alignment inferred the proteolytic cleavage of a signal peptide from these proteins.

tal stresses and act as chaperones that can protect proteins against denaturation and degradation. Protein spot u4 shared homology to a hypothetical protein from *Haemophilus influenzae* which shows strong similarity to hydrolases belonging to the glyoxalase II family. Members of the glyoxalase II family are involved in the hydrolysis of S-lactoylglutathione to lactic acid and the restoration of reduced glutathione in the glyoxal pathway [30].

The expression pattern of cellular proteins in response to the *nod*-gene inducer, luteolin, was also investigated. Early exponential phase *S. meliloti* cells were exposed to luteolin for a period of 6 h and 17 h before harvesting and the proteins subsequently extracted for 2-DE analysis. No detectable change was observed in the level of protein synthesis between treated and nontreated cells upon 6 h or 17 h luteolin exposure (data not shown).

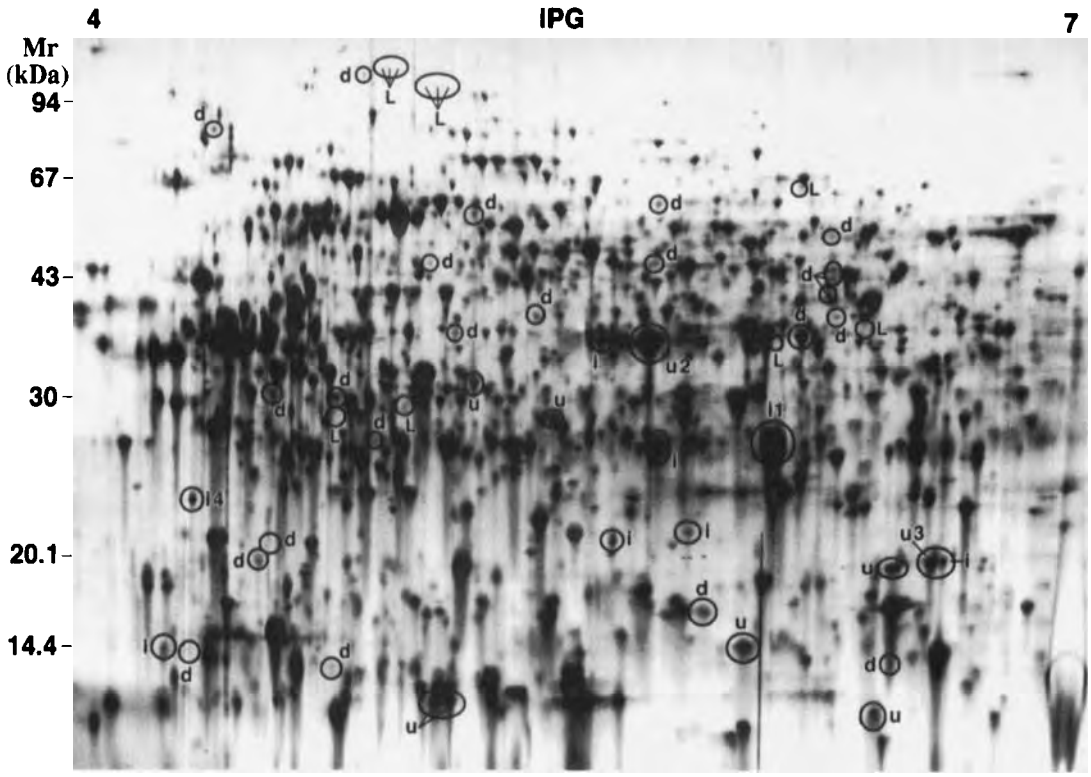


Figure 3. 2-D map of proteins expressed in *S. meliloti* strain 1021 at late exponential phase. The protein differences between early and late exponential phase growth are circled and those proteins selected for *N*-terminal microsequencing were assigned arbitrary numbers. The letter code indicates the following changes in protein synthesis: d, down-regulation; u, up-regulation; i, induced; L, loss of protein.

Table 2. List of growth -phase-regulated proteins for which *N*-terminal sequence was obtained

Spot[a)]	M_r/p*I*	*N*-terminal sequence	Homology	Organism	M_r/p*I*	SWISS-PROT Accession No.
i1	26095/6.21	AEPATKLPIKSEEKGVE RRA	Unknown			
u2	36762/5.89	HATFETGSAPAETTVN ATL	Unknown			
u3	17802/6.57	MRHFDFSPLYRSTV	16 kD heat shock protein A (71.4%)	*Escherichia coli*	15774/5.57	P29209
u4	22157/4.66	MLQAGIIPVTHFEQN	Hypothetical protein HI1663 (60%)	*Haemophilus influenza*	23882/5.12	Q57544

a) Arbitrary numbers assigned to proteins which were up-regulated or induced in late exponential growth phase. The letter code indicates the following changes in protein synthesis: u, up-regulation; **i**, induced.

4 Discussion

We present the beginning of a proteomic approach to the global analysis of gene expression in *S. meliloti.* We established a 2-D protein map of *S. meliloti* containing approximately 2000 gene products. This 2-D protein map could be further resolved into proteomic contigs using narrow-range IPG strips which provided enhanced resolution ability. The *S. meliloti* genome is 6.5 Mbp in size [31], which could code for approximately 6500 average-sized

proteins (each of 330 amino acids and equivalent to 1 kbp). Assuming that one gene codes for one gene product, we calculate that about 31% of the *S. meliloti* coding capacity was visualised on one 2-D gel within the pH range of 4–7 and size range of 10–122 kDa. This level of detection and resolution was similar to that obtained with *R.l.* bv *trifolii* [22].

Eighteen abundant constitutive proteins were *N*-terminal microsequenced and assigned to the *S. meliloti* reference map. Fifteen of these were identified on the basis of their sequence homology, with six matching to the sequences of known rhizobia proteins and nine matching to proteins found in other prokaryotes but not previously reported in rhizobia. *N*-terminal sequencing of mature proteins was valuable in revealing *N*-terminal post-translational processing and in turn assigning a subcellular location to the analysed protein. We identified several novel proteins with putatively cleaved signal peptides, indicative that they are exported across the inner membrane. Of the identified proteins, spot 1 is of particular interest since it showed strong homology to a periplasmic amide-urea binding protein, thereby possibly representing the binding protein component of a novel transport system in rhizobia. This solute-binding protein forms part of an active transport system for the uptake of short-chain amides and urea in *M. methylotrophus* [32]. Short-chain amides and urea are used as a source of nitrogen (and carbon in the case of amides) for growth and amino acid synthesis once hydrolysed by their respective hydrolases [33, 34]. Urea hydrolysis may also lead to the neutralisation of the microenvironment surrounding the bacterium under acidic conditions which has been shown with the pathogen *Helicobacter pylori* [35], and such a protective mechanism could be utilised by *S. meliloti* under acidic soil conditions. Interestingly, the relative abundance of spot 1 suggests that this protein may play an important role in cellular function. Furthermore, the amide-urea binding protein was also shown to be maximally expressed in *M. methylotrophos* during growth under amide or urea limitation [32], which was similar to the *S. meliloti* growth medium used where glutamate was the sole nitrogen source. Five highly abundant proteins involved in protein chaperoning and translation exhibited identical electrophoretic mobilities to homologues in *R.l.* bv *trifolii* [22], indicating that these proteins may be highly conserved within rhizobia. Several of these proteins, GroEL, DnaK and EF-G, also appeared to be present in several isoforms as they were in *R.l.* bv *trifolii*.

Many proteins were observed for which the synthesis rates changed during the transition of growth phases. These results show the dynamic nature of the *S. meliloti* proteome which changes in response to varying environmental conditions. A number of proteins which were synthesised during early exponential phase growth were completely repressed or synthesised at reduced levels in late exponential phase growth, while a number of other proteins were either up-regulated or induced. Some of these growth-phase-related differences could be associated with the quorum sensing system which activates the expression of genes, operons and regulons at high cell densities [36]. We identified an up-regulated protein which showed similarity to hydrolases belonging to the glyoxalase II family, and also a small heat shock protein which functions to protect proteins against degradation. Glyoxalase II hydrolases are involved in the restoration of reduced glutathione into the glyoxal pathway which is involved in detoxification of toxic metabolites, such as alpha-oxoaldehydes. These toxic metabolites induce DNA mutagenesis and lead to protein degradation [37]. Toxic metabolites are likely to reach high levels as cells approach stationary phase; therefore a variety of protective mechanisms need to be employed to maintain cell viability. The induction of both a heat shock protein and a putative hydrolase that generates reduced glutathione suggests that cells in late exponential phase may be subjected to higher metabolic stress levels than early exponential phase cells.

The addition of the flavonoid luteolin to *S. meliloti* 1021 did not result in any detectable changes in gene expression on the 2-D protein maps. This is in contrast to *R.l.* bv *trifolii* strain ANU843 where the flavonoid-induced nodulation proteins NodA, NodB and NodE are clearly visible on 2-D protein maps upon the addition of the flavonoid 7,4′-dihydroxyflavone [22]. Translational *lacZ* fusions have been used to measure the level of expression of the common *nod* genes in free-living rhizobia in response to flavonoid inducers. The level of expression of the inducible *nod* genes has been shown to significantly vary amongst the different species of rhizobia. *S. meliloti* 1021 has been shown to exhibit very low levels of flavonoid-induced *nod* gene induction (1.6-fold increase) unless extra copies of the *nodD1* gene that encodes a positive *nod* gene regulator are artificially elevated [3, 38, 39]. Such low levels of *nod* gene products would be difficult to visualise on the 2-D gels by the detection method used. This is in contrast to *R.l.* bv *trifolii*, which exhibits high levels of *nod* gene induction (up to 100-fold) [40]. It appears that low levels of gene expression in *S. meliloti* also lead to low levels of the corresponding gene products.

As the complete nucleotide sequence of nitrogen-fixing bacteria become available, proteome analysis will be an essential tool in understanding the functional and molecular basis of microbial gene expression under nutrient limiting, plant-associated and symbiotic growth conditions. Genomics complemented with proteomics provides a

powerful tool for the comprehensive study of complex regulatory processes in microbes. A completed genomic sequence will allow for the use of alternative protein identification approaches such as mass spectrophotometry, which offers high throughput and sensitivity, therefore enabling a more vigorous analysis of the *S. meliloti* proteome.

We wish to thank Jill McGovern from the Biomolecular Resource Facility, Australian National University, for performing the N-terminal sequence analysis of proteins.

Received November 12, 1998

5 References

[1] Spaink, H. P., *Crit. Rev. Plant Sci.* 1996, *15*, 559–582.

[2] Pueppke, S., *Crit. Rev. Biotechn.* 1996, *16*, 1–51.

[3] Peters, N. K., Frost, J. W., Long, S. R., *Science* 1986, *233*, 977–980.

[4] Desaizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W., Mous, J., *Nature Biotech.* 1998, *16*, 45–48.

[5] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[6] Anderson, N. G., Anderson, N., *Electrophoresis* 1998, *19*, 1853–1861.

[7] Haynes, P., Gygi, S., Figeys, D., Aebersold, R., *Electrophoresis* 1998, *19*, 1862–1871.

[8] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. K., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *BioTechnology* 1996, *14*, 61–65.

[9] Yan, J., Wilkins, M. R., Ou, K., Gooley, A. A., Williams, K. L., Sanchez, J.-C., Golaz, O., *J. Chromatogr. A* 1996, *736*, 291–302.

[10] Yates III, J. R., *J. Mass Spectrom.* 1998, *33*, 1–19.

[11] Cordwell, S., Basseal, D., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1335–1346.

[12] Patterson, S., Thomas, D., Bradshaw, R., *Electrophoresis* 1996, *17*, 877–891.

[13] Packer, N. H., Wilkins, M. R., Golaz, O., Lawson, M. A., Gooley, A. A., Hochstrasser, D. F., Redmond, J. W., *BioTechnology* 1996, *14*, 66–70.

[14] Udiavar, S., Apffel, A., Chakel, J., Swedberg, S., Hancock, W. S., Pungor, E., *Anal. Chem.* 1998, *70*, 3572–3578.

[15] Cash, P., Argo, E., Langford, P., Kroll, J., *Electrophoresis* 1997, *18*, 1472–1482.

[16] Link, A., Hays, L., Carmack, E., Yates III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[17] Antelmann, H., Bernhardt, J., Schmid, R., Mach, H., Volker, U., Hecker, M., *Electrophoresis* 1997, *18*, 1451–1463.

[18] Qi, S.-Y., Moir, A., O'Conner, D., *J. Bacteriol.* 1996, *178*, 5032–5038.

[19] VanBogelen, R., Abshire, K., Moldover, B., Olson, E., Neidhardt, F., *Electrophoresis* 1997, *18*, 1243–1251.

[20] VanBogelen, R., Olsen, E., Wanner, B., Neidhardt, F., *J. Bacteriol.* 1996, *178*, 4344–4366.

[21] Link, A., Robison, K., Church, G., *Electrophoresis* 1997, *18*, 1259–1313.

[22] Guerreiro, N., Redmond, J. W., Rolfe, B. G., Djordjevic, M. A., *MPMI* 1997, *10*, 506–516.

[23] Guerreiro, N., Stepkowski, T., Rolfe, B. G., Djordjevic, M. A., *Electrophoresis* 1998, *19*, 1972–1979.

[24] Dazzo, F. B., in: Burns, R. G., Slater, J. L. (Eds.), *Experimental Microbiol. Ecology*, Blackwell Scientific, Oxford, UK 1982, pp. 431–446.

[25] Rabilloud, T., Vuillard, L., Gilly, C., Lawrence, J. C., *Cell. Mol. Biol.* 1994, *40*, 57–75.

[26] Humphery-Smith, I., Blackstock, W., *J. Prot. Chem.* 1997, *16*, 537–544.

[27] Zimmermann, H., *Biochem. J.* 1992, *285*, 345–365.

[28] Murphy, C. K., Beckwith, J., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM press, Washington DC 1996, pp. 967–978.

[29] Pandher, K., Murphy, G., *Vet. Microbiol.* 1996, *51*, 331–341.

[30] Kim, N.-S., Umezawa, Y., Ohmura, S., Kato, S., *J. Biol. Chem.* 1993, *268*, 11217–11221.

[31] Sobral, B. W., Honeycutt, R. J., Atherly, A. G., McClelland, M., *J. Bacteriol.* 1991, *173*, 5173–5180.

[32] Mills, J., Wyborn, N., Greenwood, J., Williams, G., Jones, C., *Eur. J. Biochem.* 1998, *251*, 45–53.

[33] Mobley, H. L. T., Island, M. D., Hausinger, R. P., *Microbiol. Rev.* 1995, *59*, 451–480.

[34] Wyborn, N. R., Mills, J., Williams, S. G., Jones, C. W., *Eur. J. Biochem.* 1996, *240*, 314–322.

[35] Skouloubris, S., Labigne, A., De Reuse, H., *Mol. Microbiol.* 1997, *25*, 989–998.

[36] Fuqua, C., Greenberg, E. P., *Curr. Opinion Microbiol.* 1998, *1*, 183–189.

[37] Thornalley, P. J., *Chem. Biol. Interactions* 1998, *112*, 137–151.

[38] Mulligan, J., Long, S., *Proc. Natl. Acad. Sci. USA* 1985, *82*, 6609–6613.

[39] Kondorosi, E., Gyuris, J., Schmidt, J., John, M., Duda, E., Hoffnung, B., Schell, J., Kondorosi, A., *Eur. Mol. Biol. Org. J.* 1989, *8*, 1331–1340.

[40] Djordjevic, M. A., Redmond, J. W., Batley, M., Rolfe, B. G., *Eur. Mol. Biol. Org. J.* 1987, *6*, 1173–1179.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 826–829

Alois Harder[1]
Robert Wildgruber[1]
Arek Nawrocki[2]
Stephen J. Fey[2]
Peter Mose Larsen[2]
Angelika Görg[1]

[1]Technical University of Munich, Department of Food Technology, Freising-Weihenstephan, Germany
[2]Center for Proteome Analysis in Life Sciences, Odense, Denmark

Comparison of yeast cell protein solubilization procedures for two-dimensional electrophoresis

Three different procedures for the solubilization of yeast (*S. cerevisiae*) cell proteins were compared on the basis of the obtained two-dimensional (2-D) polypeptide patterns. Major emphasis was laid on minimizing handling steps, protein modification or degradation, and quantitative loss of high molecular mass proteins. The procedures employed were sonication, followed by (i) protein solubilization with "standard" lysis buffer (9 M urea, 2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 1% dithiothreitol (DTT), 2% v/v carrier ampholytes, (ii) presolubilization of proteins with sodium dodecyl sulfate (SDS) buffer, consisting of 1% SDS and 100 mM tris(hydroxymethyl)aminomethane (Tris)-HCl, pH 7.0, followed by dilution with "standard" lysis buffer, and (iii) boiling the sample with SDS during cell lysis, followed by dilution with thiourea/urea lysis buffer (2 M thiourea / 7 M urea, 4% w/v CHAPS, 1% w/v DTT, 2% v/v carrier ampholytes). All procedures tested were rapid and simple. However, with the first procedure (i), considerable degradation of high M_r proteins occurred. In contrast, protein degradation was minimized by boiling the sample in SDS buffer immediately after sonication (method ii). Protein disaggregation and solubilization of high M_r proteins were further improved by pre-boiling with SDS and using thiourea/urea lysis buffer instead of "standard" lysis buffer (procedure iii).

Keywords: Immobilized pH gradient / *Saccharomyces cerevisiae* / Two-dimensional polyacrylamide gel electrophoresis / Yeast protein solubilization EL 3347

1 Introduction

High-resolution two-dimensional (2-D) gel electrophoresis is the core technology in proteome analysis. Since thousands of gene products can be separated and, in combination with computer-aided image analysis, analyzed simultaneously, this approach is now being widely used to establish protein databases from all kinds of samples. Theoretically, the information stored in these 2-D databases can be exchanged between different laboratories, but in practice the exchange of data is hampered by the fact that in many cases the 2-D polypeptide patterns of the same sample analyzed in different laboratories are not identical. Although one reason responsible for this lack of reproducibility, namely the spatial irreproducibility of carrier ampholyte-generated pH gradients, has been overcome by the introduction of 2-D PAGE with immobilized pH gradients (IPG-Dalt) [1, 2], the problem of efficient and reproducible sample preparation prior to 2-D PAGE is still unsolved for a variety of samples. Typically, sample preparation includes the following steps: (i) lysis (breakage) of the cell wall, (ii) inactivation or removal of interfering substances (*e.g.*, proteases, nucleic acids or plant phenols), and (iii) protein solubilization (*i.e.*, disruption of protein aggregates or complexes into a solution of individual polypeptides) [3]. Unfortunately, there is no single sample preparation method which can be applied for all kinds of samples. Nevertheless, independently from the sample preparation method used, it is of great importance to minimize protein modification or degradation which otherwise would result in artifactual spots on the 2-D patterns, and to employ a sample solubilization procedure as efficient as possible to avoid quantitative loss of high molecular mass proteins, membrane proteins, and/or nuclear proteins.

Depending on the nature of the sample, more or less widely used methods for the disruption of cell walls are autolysis (*e.g.*, by incubation with toluene), enzymatic lysis (*e.g.*, with lysozyme or β-glucanase), grinding of the deep-frozen cell pellet in a liquid nitrogen-cooled mortar, glass-bead beating, or sonication. Tissues are often disrupted in the presence of sample solubilization buffer. All these procedures have their advantages and disadvantages. Proteolytic enzymes present within the sample have to be inactivated by adding (during cell lysis) proteinase inhibitors, Tris base, or high amounts of urea and DTT, by TCA/acetone precipitation, or by boiling the sam-

Correspondence: Prof. Dr. Angelika Görg, Technische Universität München, Lehrstuhl für Allgemeine Lebensmitteltechnologie, D-85350 Freising-Weihenstephan, Germany
E-mail: angelika.gorg@tum.de
Fax: +49-8161-714264

Abbreviation: IPG-Dalt, two-dimensional gel electrophoresis with immobilized pH gradient

ple in SDS buffer. Yet, it should be kept in mind that certain proteases retain their activity even under rather harsh conditions.

Following cell disruption and inactivation of interfering substances, proteins have to be solubilized. The most common protein solubilization procedures are based on O'Farrell's method [4], using mixtures of 2–4% nonionic (*e.g.*, NP-40 or Triton X-100) or zwitterionic (*e.g.*, CHAPS) detergent, 9–10 M urea, 1% DTT, and 2% carrier ampholytes. Although these procedures work well with the majority of samples, there are protein complexes which are not disrupted by these agents, so that certain proteins are either missing or insufficiently resolved on the 2-D pattern. In contrast to urea, SDS disrupts almost all noncovalent protein interactions. However, because of its deleterious effects on IEF, it has to be displaced from the proteins with a nonionic or zwitterionic detergent (such as NP-40 or CHAPS) prior to IEF. In practice, this is achieved by first solubilizing the proteins with SDS buffer, followed by dilution with excess lysis buffer. Although SDS is an efficient solubilizing agent, one problem of this approach is that certain proteins tend to reaggregate after SDS has been replaced by a less efficient detergent, causing streaks or precipitates on the 2-D gel. In order to increase the solubilizing power of the "standard" lysis buffer, Rabilloud and co-workers [5] have tested a range of detergents and chaotrops. As a result, they proposed to replace the standard lysis buffer by a mixture of 2 M thiourea and 7 M urea, plus a high concentration of an efficient detergent (*e.g.*, 4% CHAPS), 1% DTT, and 2% carrier ampholytes.

The aim of the present study was to compare three different procedures for solubilizing yeast cell proteins for IPG-Dalt analysis. Besides simplicity and reproducibility, special emphasis was placed on minimizing protein degradation and quantitative loss of high molecular mass proteins. The procedures employed were cell lysis by sonication and (i) protein solubilization with "standard" lysis buffer, (ii) presolubilization of proteins with SDS, followed by dilution with "standard" lysis buffer, and (iii) boiling the sample with SDS during cell lysis, followed by dilution with thiourea/urea lysis buffer.

2 Materials and methods

2.1 Apparatus and chemicals

All apparatus for IEF and horizontal electrophoresis (Multiphor II electrophoresis chamber, Macrodrive V power supply, DryStrip kit, gradient mixer), Immobilines, Pharmalyte pH 3–10 and GelBond PAGfilm were from Amersham Pharmacia Biotech (Uppsala, Sweden). The Dalt apparatus for multiple runs was from Hoefer Amersham Pharmacia Biotech (San Francisco, CA, USA). Acrylamide (2 × crystallized), *N,N'*-methylenebisacrylamide and SDS were from Serva (Heidelberg, Germany). Tris (hydroxymethyl)aminomethane (Tris), iodoacetamide, dithiothreitol (DTT) and CHAPS were from Sigma (St. Louis, MO, USA). Thiourea was from Fluka (Buchs, Switzerland). Synthetic medium (YNB) for yeast cell growth was from Difco (San Francisco, CA, USA). All other chemicals (analytical or biochemical grade) for yeast cell growth, protein solubilization, electrophoresis and silver staining were from Merck (Darmstadt, Germany).

2.2 Yeast cell growth

The haploid wild-type strain S288c (suc2 mal gal cup1) (Deutsche Sammlung für Mikroben und Zellinien, Braunschweig, Germany) of *Saccharomyces cerevisiae* was grown at 30°C in defined synthetic YNB medium (6.7 g/L) supplemented with 20 g/L w/v glucose and buffered with sodium hydroxide (6 g/L w/v) / succinic acid (10 g/L w/v) to pH 5.8. Cells were harvested by centrifugation in the late mid-exponential phase when the culture had reached an OD of 1.0, washed in ice-cold deionized water, and spun (16 000 × *g*, 15 min). Yeast cell pellets were transferred into individual 1.5 mL Eppendorf tubes and stored at –78°C if not processed immediately.

2.3 Cell lysis and protein solubilization

Three different protein solubilization procedures were compared: (i) To the yeast cell pellet (5 mg dry weight) were added 200 µL of "standard" lysis buffer (9 M urea, 1% w/v DTT, 2% w/v CHAPS, 2% v/v carrier ampholytes, pH 3–10, and 10 mM Pefabloc® proteinase inhibitor) and sonicated three times for 1 s each (60 W, 20 kHz) with intermediate cooling on ice. After cell lysis, the sample was spun (16 000 × *g*, 20 min, 15°C). (ii) To the yeast cell pellet (5 mg dry weight) were added 200 µL of SDS sample buffer (1 % w/v SDS, 100 mM Tris-HCl, pH 7.0); sonication was performed as above. After cell lysis, the sample was boiled for 5 min, cooled in an ice-bath, diluted with 500 µL of "standard" lysis buffer, and spun as above. (iii) To the yeast cell pellet were added 200 µL of hot (95°C) SDS sample buffer (1% w/v SDS, 100 mM Tris-HCl, pH 7.0) and sonified (3 × 1 s). After cell lysis, the sample was additionally boiled for 5 min, cooled in an ice-bath, diluted with 500 µL of thiourea/urea lysis buffer (2 M thiourea, 7 M urea, 4% w/v CHAPS, 1% w/v DTT and 2% v/v carrier ampholytes, pH 3–10), shaken for 1 h, and spun (16 000 × *g*, 20 min). The clear supernatants were removed and stored at –78°C until analyzed.

2.4 IPG-Dalt of yeast cell proteins

IPG-Dalt was carried out according to Görg *et al.* [1, 2, 6] (for further details see also http://www.weihenstephan.de/blm/deg/manual/manfrm.htm). IPG gels with linear gradients pH 4–9 (4%T, 3%C; size, 250 × 180 × 0.5 mm) were cast as described [6]. The washed and dried IPG gels were then cut into 3 mm wide IPG strips which were rehydrated overnight to their original thickness of 0.5 mm with a solution consisting of 8 M urea, 0.5% CHAPS, 0.2% DTT and 0.2% Pharmalytes, pH 3–10. The rehydrated IPG gel strips were placed onto the cooling plate of a flatbed IEF chamber. Sample volumes of 20 µL were applied into silicone frames directly placed onto the surface of the rehydrated IPG strips (anodic sample application), whereas larger sample volumes (up to 100 µL) were applied with the help of sample application cups (Pharmacia DryStrip kit). Sample load was about 60 µg of protein per IPG strip. IEF was performed at 20°C [7] for 28 000 Vh [8]. Following IEF, the IPG gel strips were stored in a plastic bag at –78°C. Prior to the second dimension of IPG-Dalt, the IPG gel strips were equilibrated for 2 × 10 min in 2 × 10 mL equilibration solution (50 mM Tris-HCl buffer, pH 8.8, containing 6 M urea, 30 % w/v glycerol, 2% w/v SDS). DTT (1% w/v) was added to the first equilibration solution, and iodoacetamide (4% w/v) to the second [1]. Second-dimensional SDS-PAGE (13%T, 2.7%T; Laemmli [9] buffer system) was carried out simultaneously on ten slab gels (210 × 185 × 1.0 mm) in a Dalt tank [6, 10]. Silver staining was done according to a modified procedure of Blum *et al.* [11].

3 Results and discussion

Three different procedures for the solubilization of yeast cell proteins were compared on the basis of their 2-D polypeptide patterns. Major emphasis was laid on minimizing protein modification and degradation, and high yield of high M_r proteins. All methods tested were rapid and simple. However, there were considerable differen-

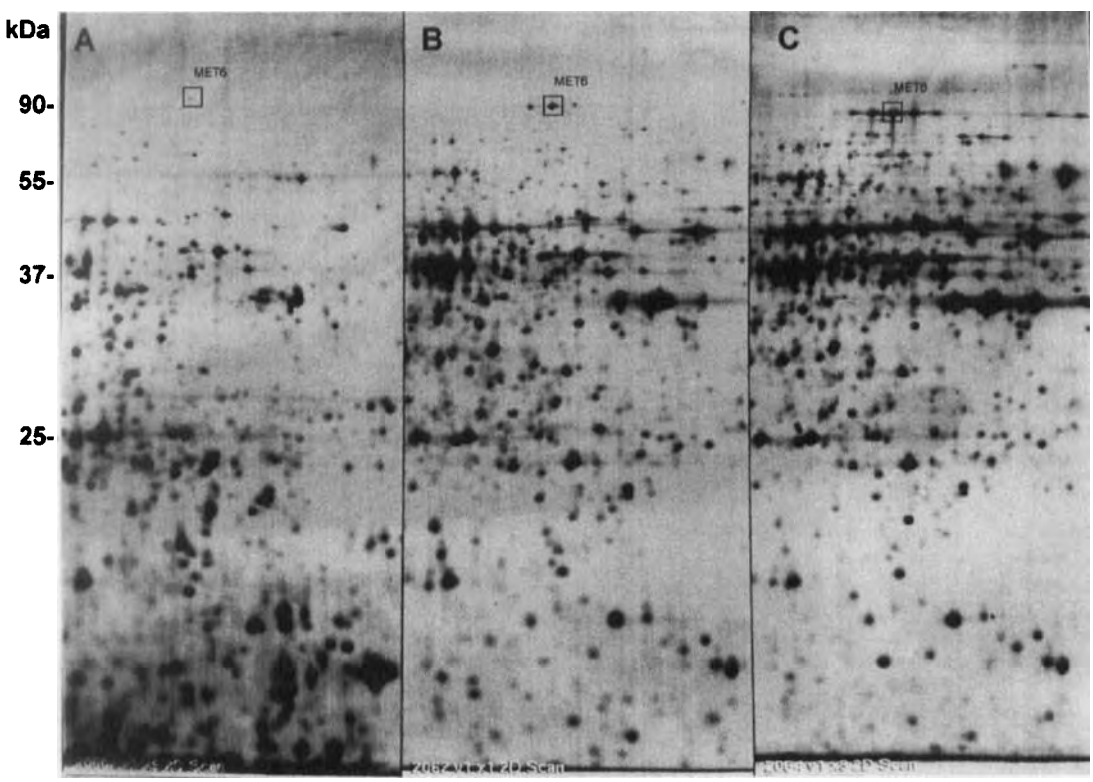


Figure 1. Comparison of three different solubilization procedures for yeast cell proteins. Yeast cell were sonicated in the presence of (A) "standard" lysis buffer, (B) SDS, followed by dilution with "standard" lysis buffer, and (C) boiled with SDS during cell lysis, followed by dilution with thiourea/urea lysis buffer. IPG-Dalt according to Görg *et al.* [1, 2, 6]. First dimension: IPG-IEF 4–9 (section). Second dimension: SDS-PAGE (13%T, 2.6%C). Silver stain [11]. MET6, methionine synthase.

ces in protein degradation and quantitative loss of high M_r proteins between the individual procedures. With procedure (i), *i.e.*, sonication of yeast cells in the presence of "standard" lysis buffer, low M_r polypeptides (<30 kDa) were the predominant species on the 2-D gel, whereas only a small number of higher M_r proteins > 50 kDa was detectable (Fig. 1A). Obviously, significant protein degradation had occurred during (or after) cell disruption due to insufficient inactivation of endogenous proteases.

In order to minimize proteolytic breakdown, a modified procedure (ii) was employed in which yeast cell proteins were disrupted in the presence of SDS instead of urea. Immediately after cell disruption, the extract was boiled for 5 min for inactivation of endogenous proteolytic enzymes as well as for improved protein solubilization (Fig. 1B). From a comparison between Figs. 1A and 1B it is obvious that protein degradation was considerably reduced due to the SDS treatment, and – in contrast to procedure (i) – proteins in the M_r range between 25 and 50 kDa were the predominant species on the 2-D pattern. Despite these improvements, a number of proteins in the 2-D gel region above 60 kDa seemed to be missing, as could be calculated from the sequenced yeast genome. Possibly, "standard" lysis buffer is not efficient in keeping certain high M_r proteins in a soluble state after the removal of SDS from the proteins.

In order to increase the solubilizing power of the lysis solution, procedure (ii) was modified in such a way that the "standard" lysis buffer was replaced by a lysis buffer which contained 2 M thiourea, 7 M urea; additionally, the CHAPS concentration was increased from 2% to 4% (Fig. 1C). To be sure that endogenous proteases were inactivated during cell disruption as far as possible, procedure (ii) was also modified in such a way that cell lysis was performed by sonication in presence of hot SDS buffer instead of working at low temperatures. In comparison to procedure (ii), a larger number of high M_r proteins in the range above 60 kDa was detectable (Fig. 1C). Computer densitometric evaluation revealed that the number of spots detectable on the 2-D pattern shown on Fig. 1C exceeded the number of spots present on the 2-D pattern of Fig. 1B by more than 50%.

4 Concluding remarks

In yeast cells, endogenous proteolytic enzymes comprise the major source of problems with respect to protein degradation and modification during sample preparation for 2-D electrophoresis. Given that these proteases have retained at least part of their activity during cell disruption, raising the temperature (even in presence of SDS) prior to electrophoresis can activate them and generate artifacts [12]. This formation of artifacts can be avoided by heating the sample in SDS buffer prior to, and during, cell disruption. To keep the extracted proteins in solution after removal of SDS, superior results were obtained with thiourea/urea buffer in comparison to "standard" lysis buffer not containing any thiourea. In order to further simplify the procedure, investigations are underway in which yeast cells are rapidly disrupted in the presence of thiuorea/urea buffer with a more powerful sonifier than the one used in the present study.

This work was supported by the EU Biotechnology program (1994–1998; Contract No. BIO4-97-2361).

Received November 22, 1998

5 References

[1] Görg, A, Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[2] Görg, A., *Nature* 1991, *349*, 545–546.

[3] Dunn, M. J., *Gel Electrophoresis of Proteins*, BIOS Scientific Publishers Ltd. Alden Press, Oxford 1993.

[4] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[5] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[6] Görg, A., Weiss, W., in: Celis, J. (Ed.), *Cell Biology. A Laboratory Handbook*, Academic Press, New York 1998, pp. 386–397.

[7] Görg, A., Postel, W., Friedrich, C., Kuick, R., Strahler, J. R., Hanash, S. M., *Electrophoresis* 1991, *12*, 653–658.

[8] Görg, A., *Biochem. Soc. Transactions* 1993, *21*, 130–132.

[9] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[10] Anderson, N. G., Anderson, N. L., *Anal. Biochem.* 1978, *85*, 331–354.

[11] Blum, H., Beier, H., Gross, H. J., *Electrophoresis* 1987, *8*, 93–99.

[12] Jones, W. R., *Methods Enzymol.* 1991, *194*, 428–453.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 830–835

Takashi Manabe
Hitomi Mizuma
Kenji Watanabe

Department of Chemistry, Faculty of Science, Ehime University, Matsuyama, Japan

A nondenaturing protein map of human plasma proteins correlated with a denaturing polypeptide map combining techniques of micro two-dimensional gel electrophoresis

Human plasma proteins were separated by combining four types of two-dimensional electrophoresis (2-DE) techniques to obtain systematic information on proteins and their constituent polypeptides. A micro gel system was employed to facilitate the analysis. A plasma sample was first analyzed under nondenaturing conditions of electrophoresis (Type I 2-DE) to characterize the properties of proteins under physiological conditions. The sample was then analyzed, employing nondenaturing isoelectric focusing in the first dimension and sodium dodecyl sulfate (SDS) electrophoresis in the second dimension (Type II 2-DE), to study the dissociation of noncovalently bound protein subunits. In the third type of 2-DE (Type III 2-DE), proteins were separated by nondenaturing isoelectric focusing and treated with urea/mercaptoethanol/SDS and then subjected to second-dimension SDS electrophoresis, to study the dissociation of disulfide-bonded polypeptides. In the fourth type of 2-DE (Type IV 2-DE), the conditions of denaturing 2-DE were employed; the sample was treated with SDS-mercaptoethanol-urea-Nonidet P-40, separated by denaturing isoelectric focusing, and then subjected to SDS electrophoresis. The combined 2-DE technique will be useful to construct a comprehensive database of plasma proteins combining a "nondenaturing protein map" (a protein map) and a "denaturing protein map" (a polypeptide map).

Keywords: Nondenaturing conditions / Denaturing conditions / Protein map / Polypeptide map / Human plasma proteins / Two-dimensional polyacrylamide gel electrophoresis EL 3321

1 Introduction

A technique of two-dimensional polyacrylamide gel electrophoresis (2-DE), which employs denaturing agents in both dimensions [1], has been successfully used to analyze the structure of polypeptides expressed in cells or in other complex protein systems, making it possible to correlate gene structures with their expression products. The analyzed polypeptide structure includes information on various post-translational modifications. However, further information is required if we aim to reconstruct the structure of proteins with physiological activities. Human plasma proteins have been analyzed by denaturing 2-DE [2–5] and information on denatured proteins (polypeptides) has been accumulated as databases [6, 7]. In order to acquire information on proteins in physiological conditions, we have developed a nondenaturing 2-DE technique and applied it to analyze proteins in various body fluids including human plasma [8, 9]. Major plasma proteins have been identified by electrophoretic blotting followed by immunochemical staining [10] and the results have been summarized as a "protein map". Enzyme activities can be detected on the 2-D gels [11] and the dissociation process of lipoproteins into apolipoproteins can be analyzed by modifying the process of IEF gel equilibration [12, 13]. In this paper, we describe attempts to correlate a protein map obtained by nondenaturing 2-DE of plasma proteins with a map obtained by denaturing 2-DE [6, 7]. Four types of 2-DE technique have been combined to correlate the protein map with the polypeptide map and the dissociation of proteins into their constituent polypeptides have been traced.

Correspondence: Dr. Takashi Manabe, Department of Material Science (Chemistry Group), Faculty of Science, Ehime University, Matsuyama-City, 790-77 Japan
E-mail: manabet@dpc.ehime-u.ac.jp
Fax: +81-89-927-9590

Abbreviations: α_2**M**, α_2-macroglobulin; **Apo B**, apolipoprotein B-100; **CBB**, Coomassie Brilliant Blue; **HDL**, high density lipoprotein; **IgA**, immunoglobulin A; **IgG**, immunoglobulin G; **IgM**, immunoglobulin M; **LDL**, low density lipoprotein

2 Materials and methods

2.1 Materials

Ammonium persulfate, *N,N'*-methylenebisacrylamide (Bis), and *N,N,N',N'*-tetramethylethylenediamine (TEMED), all special grade for electrophoresis, were from Wako Pure Chemical Industries (Osaka, Japan). Acrylamide, special

grade for electrophoresis, was from Daiichi Pure Chemicals (Tokyo, Japan). Ampholine pH 3.5–10 was from Pharmacia LKB Biotechnology (Uppsala, Sweden). Coomassie Brilliant Blue R-250 (CBB) was from Fluka Chemie AG (Buchs, Switzerland). Human blood from an apparently healthy individual (23 years, female) collected in heparinized tubes was centrifuged at 3000 × *g* for 10 min. Blood collected in dry tubes was kept for 4 h in order to achieve complete coagulation. Plasma and sera supplemented with sucrose to a final concentration of 40% w/v were ivided into small aliquots and kept at –20 °C.

2.2 Micro 2-DE

The apparatus for micro 2-DE [14, 15] and the automated apparatus for gradient gel preparation [16] were used. For micro 2-DE, four different combinations of first- and second-dimensional electrophoresis were employed, namely, (i) without denaturants in both dimensions, (ii) without denaturants in the first dimension and with SDS in the second dimension, (iii) without denaturants in the first dimension, the IEF gel equilibrated with urea and mercaptoethanol, and with SDS in the second dimension, and (iv) in presence of urea and NP-40 in the first dimension and with SDS in the second dimension.

2.2.1 Type I

The first type of micro 2-DE (Type I) was performed as described previously [10, 15]: The plasma sample (2.0 μL) was subjected to isoelectric focusing in the absence of denaturants employing column gels (1.3 ϕ × 35 mm) which contained Ampholine pH 3.5–10 and Ampholine pH 3.5–5 in final concentrations of 2% and 0.5%, respectively; isoelectric focusing was run at 0.1 mA/tube constant current until a voltage of 300 V was reached (about 30 min) and continued at 300 V/cm constant voltage for 40 min. The focusing gels were transferred onto the second-dimensional slab gels, equilibrated for 10 min with 0.01 M glycine-0.076 M Tris (pH 8.3), and electrophoresis was run in 4–17 % acrylamide linear gradient microslab gels without denaturants (38 × 38 × 1 mm) at 10 mA/slab constant current. Electrophoresis was continued until the spot of albumin, visualized by binding with bromophenol blue (BPB) added in the electrophoresis buffer, moved to about 15 mm from the bottom end of the slab gels (about 50 min).

2.2.2 Type II

In the second type of micro 2-DE [12] (Type II), isoelectric focusing was performed as in technique (i), except the sample volume was 1.0 μL. The focusing gels were transferred, equilibrated for 20 min with 2% SDS – 0.01 M Tris – 0.076 M glycine buffer (pH 8.3), and electrophoresis was run in the gradient micro slab gels containing 0.1% SDS at 10 mA/slab constant current. Electrophoresis was continued until the band of BPB moved to about 3 mm from the bottom end of the slab gels (about 40 min).

2.2.3 Type III

In the third type of micro 2-DE (Type III), isoelectric focusing was performed as in technique (ii). The focusing gels were transferred, equilibrated for 20 min with 0.026 M Tris-HCl buffer (pH 6.8) containing 2% SDS-5% mercaptoethanol-8 M urea, and electrophoresis was run as in technique (ii).

2.2.4 Type IV

The fourth type of micro 2-DE (Type IV) was performed by modifying the method of Anderson and Anderson [3]. One volume of plasma sample, containing 40% sucrose, was mixed with one volume of solubilizing solution of 4% SDS-10% mercaptoethanol. The mixture was heated at 95 °C for 5 min and centrifuged at 10 000 × *g* for 3 min; then an 0.8 μL aliquot was subjected to IEF. Isoelectric focusing was run in micro column gels, containing 8 M urea, 1% NP-40, and 2% Ampholine pH 3.5–10, at 0.1 mA/tube constant current until 300 V was reached (about 40 min), then at 300 V constant voltage for an additional 80 min. The focusing gels were transferred and equilibrated with 0.026 M Tris-HCl buffer (pH 6.8) containing 2% SDS-5% mercaptoethanol-10% glycerol for 20 min. The second-dimensional electrophoresis was run as technique (ii). The slab gels were stained in 0.1% CBB R-250 in 50% v/v methanol/7% v/v acetic for 15 min, and destained in 20% v/v methanol/7 % acetic acid for 2 h. SDS-containing slab gels were washed in 50% v/v methanol/7% v/v acetic acid for 15 min before staining to remove carrier ampholytes.

2.3 Estimation of isoelectric points and molecular weights of plasma proteins

Isoelectric points of proteins were estimated according to the "normalized map" [9] of human plasma proteins. Molecular weights of plasma proteins in SDS slab gels were estimated by coelectrophoresis of high molecular weight and low molecular weight calibration kits (Pharmacia LKB Biotechnology) and preparation of a standard curve of molecular weight *versus* mobility.

2.4 Identification of plasma proteins and polypeptides

Proteins on "Type I" (nondenaturing) and "Type II" (nondenaturing IEF, then SDS electrophoresis) 2-DE patterns

were identified referring to the "identification map" of human plasma proteins [10]. Also, proteins on "Type II"- and "Type III" (nondenaturing IEF, equilibration with urea and reducing agent, then SDS electrophoresis) 2-DE patterns were identified by blotting-immunochemical identification [10]. Proteins on "Type IV" (denaturing) 2-DE patterns were identified referring to the papers which dealt with the plasma protein identification on denaturing 2-DE [3, 6, 7].

3 Results and discussion

3.1 Comparisons between a nondenaturing (Type I) 2-DE pattern and a nondenaturing/SDS (Type II) 2-DE pattern

A nondenaturing (Type I) 2-DE pattern is shown in Fig. 1. Major plasma proteins have been identified on the pattern referring to the results of blotting-immunochemical staining [10] and some of the major plasma proteins are located in the figure. When a nondenaturing IEF gel was equilibrated with the SDS solution and subjected to slab gel electrophoresis in the presence of SDS (Type II 2-DE), the pattern shown in Fig. 2 was obtained. Comparing the two separation patterns, the following information is obtained: (i) Most of the major plasma proteins do not dissociate into subunits by the addition of SDS. These results are in accordance with their structural information – that they have interchain disulfide bonds. Then major plasma proteins can be identified on the Type II 2-DE pattern, since the relative location of proteins in the IEF axis is not changed. (ii) The estimation of molecular mass for proteins which have p*I* values larger than six became possible by the addition of SDS in the second dimension. In the pattern of Type I 2-DE (Fig. 1), the bands of immunoglobulin G (IgG) and immunoglobulin M (IgM) are curved since the more basic species have the smaller number of negative charges. These bands showed uniform molecular mass distribution in the pattern of Type II 2-DE (Fig. 2) because of the large negative charge acquired by the binding of many SDS molecules. (iii) High density lipoproteins (HDL) dissociated into constituent apolipoproteins, Apo A-I Apo A-II, and Apo Cs [12]. The distribution patterns of these apolipoproteins in the IEF axis are not uniform, suggesting that they are weakly associated and may be present as heterogeneous complexes under

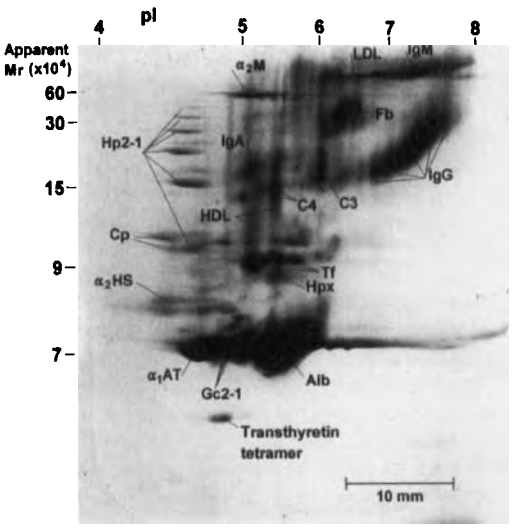


Figure 1. Nondenaturing (Type I) 2-DE of human plasma proteins. A sample of normal human plasma was subjected to isoelectric focusing (IEF) in the absence of denaturants, the IEF gel (35 mm long × ID 1.3 mm) was set on a micro slab gel (38 × 38 × 1 mm) of polyacrylamide pore gradient (4.2%T–17.85%T linear gradient, 5%C), and the second-dimensional electrophoresis was run in the absence of denaturants. Abbreviations: α_2M, α_2-macroglobulin; LDL, low density lipoprotein; IgM, immunoglobulin M; HP 2-1, haptoglobin phenotype 2-1 polymer series; IgA, immunoglobulin A; Fb, fibrinogen; IgG, immunoglobin G; HDL, high density lipoprotein; C4, complement factorC4; C3, complement factor C3; Cp, ceruloplasmin; Tf, transferrin; Hpx, hemopexin; α_2HS, α_2-HS glycoprotein; α_1AT, α_1-antitrypsin; Gc2-1, Gc-globulin phenotype 2-1; Alb, albumin.

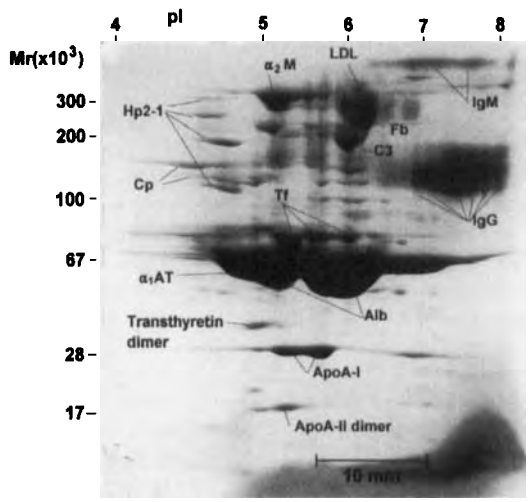


Figure 2. Type II (nondenaturing/SDS) 2-DE of human plasma proteins. The IEF conditions were the same as in Fig. 1, the IEF gel was set on a micro slab gel, equilibrated with a buffer which contained 2% SDS, then the second-dimensional electrophoresis was run in the presence of 0.1% SDS. The size and the acrylamide gradient of the slab gel were the same as in Fig. 1.

physiological conditions. (iv) SDS equilibration leads to appearance of about twenty spots of molecular mass smaller than albumin, all only weakly stained by CBB staining. These spots must be the constituents of protein complexes which are formed by noncovalent bonds. The identification of these proteins awaits further investigation. (v) The size distribution of haptoglobin polymer series (phenotype 2-1) did not change with the SDS treatment, since the subunits are disulfide-bonded. In contrast, α_2 macroglobulin (α_2M), apparent molecular mass in nondenaturing 2-DE (Fig. 1) being about 600 000 Da, seems to have dissociated into its half-molecule (apparent molecular mass about 320 000 Da). These results are inconsistent with the fact that it is a homotetramer consisting of two pairs of disulfide-linked monomers [17].

3.2 Comparisons between a nondenaturing/SDS (Type II) 2-DE pattern and a nondenaturing/reduced-SDS (Type III) 2-DE pattern

When a nondenaturing IEF gel was equilibrated with 8 M urea-5% mercaptoethanol-2% SDS and subjected to the second-dimensional SDS electrophoresis, the pattern shown in Fig. 3 was obtained. Major plasma proteins were identified on the pattern referring to the identification results shown in Fig. 2. Some protein spots disappeared, but new spots appeared at the corresponding p*I* position with different molecular mass positions. These spots were assumed to be dissociated by the equilibration procedure and were listed as the candidate spots of the constituent polypeptides of the original (vanished) protein identified in Fig. 2. Since the information on subunit structure and on amino acid sequence is available for major plasma proteins, the molecular mass values of the constituent polypeptides were calculated and were used to identify the polypeptides among the candidate spots. Further, the molecular mass values of the polypeptides in the literature on denaturing 2-DE of human plasma proteins [6, 7] were also used for the identification. Comparing the Type II (Fig. 2) with the Type III pattern (Fig. 3), the following major changes were observed: (i) Haptoglobin polymer series (phenotype 2-1) completely dissociate into constituent polypeptide chains, α^1, α^2, and β-chains, which appear at the same p*I* position (about p*I* 4.5) in Fig. 3. (ii) α_2M further dissociates into its monomers, showing apparent molecular mass of about 170 000 Da [17]. (iii) The molecular mass of transthyretin (prealbumin) varied depending on the time length of equilibration. Longer equilibration (30 min) caused two spots, at p*I* 4.8/molecular mass 34 000 Da and p*I* 4.8/molecular mass 17 000 Da; shorter equilibration (20 min) caused one spot (Fig. 3) at the larger molecular mass (34 000 Da). Since transthyretin is a homotetramer of subunits of molecular mass of about

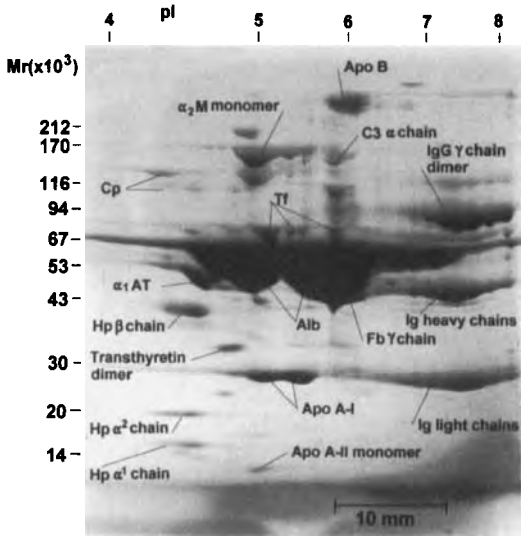


Figure 3. Type III (nondenaturing/reduced-SDS) 2-DE of human plasma proteins. The IEF conditions were the same as in Fig. 1, the IEF gel was set on a micro slab gel, equilibrated with a buffer which contained 8 M urea-5 % mercaptoethanol-2% SDS, and then the second-dimensional electrophoresis was run in the presence of 0.1% SDS. The size and the acrylamide gradient of the slab gel were as in Fig. 1. Abbreviations: Apo B, apolipoprotein B-100.

16 000 Da and antiparallel β-sheets of the monomers link monomers into dimers [18], further treatment of the gel with the urea-containing equilibration solution will fully dissociate transthyretin into monomers. (iv) Apo A-II dimer dissociates into monomers as shown in Fig. 3. (v) Fibrinogen (molecular mass about 300 000 Da in Type II pattern; Fig. 2) dissociates into constituent polypeptides; the α-chains are tentatively located at the spot at p*I* 5.8/molecular mass 94 000 Da and γ-chains at p*I* 5.8/molecular mass 45 000 Da. However, β-chains could not be identified, possibly because the spot overlapped with the dense spot of albumin. (vi) Apo B-100, which is the protein moiety of low density lipoprotein (LDL; Fig. 1), was clearly observed at the corresponding p*I* position (p*I* 6.0) and molecular mass about 500 000 Da. (vii) IgG and IgM dissociate into constituent polypeptides, heavy chains (molecular mass 50 000 Da) and light chains (25 000 Da). However, a series of dense bands appeared around p*I* 6–8 and molecular mass about 100 000 Da, suggesting that the reducing process (5% mercaptoethanol, room temperature, on top of the second-dimensional slab gel, for 20 min) is not sufficient to dissociate all the IgG polypeptides.

3.3 Comparisons between a nondenaturing/reduced-SDS (Type III) 2-DE pattern and a denaturing (Type IV) 2-DE pattern

When the plasma sample was subjected to Type IV 2-DE, modifying the method reported by Anderson and Anderson [3] for our micro gel system [12], the pattern shown in Fig. 4 was obtained. Since the relative locations of spots on the micro slab gels (40 × 40 mm) were similar to those on larger gels (160 × 160 mm) [3], we could identify major plasma polypeptides referring to the identification results on denaturing 2-DE patterns [3, 6, 7] (some of which are shown in Fig. 4). Comparing the pattern in Fig. 4 with that in Fig. 3, it is obvious that the locations of proteins in the IEF dimension are quite different from each other. The two patterns can be correlated if the Type III 2-DE pattern (Fig. 3) has been analyzed by correlating it with Type I (Fig. 1) and Type II (Fig. 2) patterns. Since the patterns in Figs. 1–3 are obtained by using the same IEF conditions (nondenaturing), correlation of the three patterns is relatively easy. We have identified and correlated the locations of major plasma proteins on Type I, Type II, and Type III patterns by using specific antibodies. Moreover, on the Type I (nondenaturing) 2-DE pattern, proteins can be identified by their biological functions. Therefore, Type I, Type II, and Type III patterns are to be used as a set to analyze the proteins present under physiological conditions. Type IV (denaturing) 2-DE patterns provide the information on p*I* molecular mass, and partial amino acid sequence of each polypeptide which is important to analyze the process of gene expression. Amino acid sequence analyses of polypeptides separated on Type III 2-DE patterns will provide more detailed information on the construction of proteins from their constituent polypeptides.

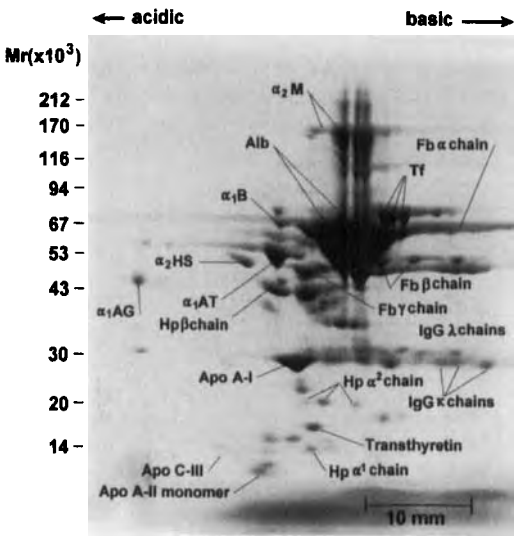


Figure 4. Type IV (denaturing) 2-DE of human plasma proteins. The plasma sample was treated with urea-mercaptoethanol-NP-40 as described in Section 2.2, subjected to IEF in the presence of urea and NP-40; the IEF gel was set on a micro slab gel, equilibrated with a buffer which contained 8 M urea – 5% mercaptoethanol – 2% SDS, and then the second-dimensional electrophoresis was run in the presence of 0.1% SDS. The size and the acrylamide gradient of the slab gel were as in Fig. 1. Abbreviations: α_1B, α_1-glycoprotein B; α_1AG, α_1-acid glycoprotein; Apo C-III, apolipoprotein C-III.

The patterns in Fig. 3 (Type III 2-DE) and Fig. 4 (Type IV 2-DE) are complementary in providing information on polypeptides in human plasma. Note, however, that the pattern in Fig. 3 shows many dense spots with a molecular mass larger than 100 000 Da, whereas the pattern in Fig. 4 has fewer spots in this region. In nondenaturing IEF, plasma samples were subjected to IEF without treatment, most of the proteins entered the IEF gel, and all the proteins in the IEF gel moved the second-dimensional slab gel. However, in denaturing IEF, a part of the plasma proteins became insoluble by treatment with urea-reducing agent NP-40. The use of SDS in sample treatment improved the solubility of plasma proteins; nevertheless, a fraction of insoluble proteins remained at the top end of the IEF gel. This might be the reason that the spots of high molecular weight polypeptides, apo-B, α_2-macroglobulin, complement C3 α-chain, ceruloplasmin, *etc.* were not detected or only faintly stained in the denaturing 2-DE pattern (Fig. 4).

4 Concluding remarks

Human plasma proteins were analyzed by combining four types of two-dimensional electrophoresis. Type I, Type II, and Type III 2-DE patterns are useful to know how polypeptides associate to form functional proteins, and the three patterns could be correlated with Type IV 2-DE patterns, which had been proved to be useful in obtaining structural information of plasma polypeptides. Comparative studies of the 2-DE patterns were facilitated by using the micro gel system. The combined technique of 2-DE presented above will provide information to construct a comprehensive database on human plasma proteins. Also, the technique should be applicable to the comprehensive analysis of other complex systems of soluble proteins, such as cytosol proteins of various cells.

Received September 1, 1998

5 References

[1] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[2] Dale, G., Latner, A. L., *Clin. Chim. Acta* 1969, *24*, 61–68.

[3] Anderson, N. L., Anderson, N. G., *Proc. Natl. Acad. Sci. USA* 1977, *74*, 5421–5425.

[4] Taylor, J., Anderson, N. L., Scandara, A. E. J., Willard, K. E., Anderson, N. G., *Clin. Chem.* 1982, *28*, 861–866.

[5] Hochstrasser, D. F., Harrington, M. G., Hochstrasser, A. C., Miller, M. J., Merril, C. R., *Anal. Biochem.* 1988, *173*, 424–435.

[6] Anderson, N. L., Anderson, N. G., *Electrophoresis* 1991, *12*, 883–906.

[7] Hughes, G. H., Frutiger, S., Paquet, N., Ravier, F., Pasquali, C., Sanchez, J.-C., James, R., Tissot, J.-D., Bjellqvist, B., Hochstrasser, D. F., *Electrophoresis* 1992, *13*, 606–714.

[8] Manabe, T., Tachi, K., Kojima, K., Okuyama, T., *J. Biochem.* 1979, *85*, 649–659.

[9] Manabe, T., Kojima, K., Jitzukawa, S., Hoshino, T., Okuyama, T., *J. Biochem.* 1981, *89*, 1317–1323.

[10] Manabe, T., Takahashi, Y., Higuchi, N., Okuyama, T., *Electrophoresis* 1985, *6*, 462–467.

[11] Kadofuku, T., Sato, T., Manabe, T., Okuyama, T., *Electrophoresis* 1983, *4*, 427–431.

[12] Manabe, T., Visvikis, S., Steinmetz, J., Galteau, M. M., Okuyama, T., Siest, G., *Electrophoresis* 1987, *8*, 325–330.

[13] Manabe, T., Visvikis, S., Dumon, M. F., Clerc, M., Siest, G., *Clin. Chem.* 1987, *33*, 468–472.

[14] Manabe, T., Hayama, E., Okuyama, T., *Clin. Chem.* 1982, *28*, 824–827.

[15] Manabe, T., Okuyama, T., in: Dunn, M. J. (Ed.), *Two-Dimensional Polyacrylamide Gel Electrophoresis '91*, Zebra Printing, Middlesex 1991, pp. 7–11.

[16] Manabe, T., Okuyama, T., in: Dunn, M. J. (Ed.), *Electrophoresis '86*, VCH Verlagsgesellschaft, Weinheim 1986, pp. 613–616.

[17] Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Lonblad, P. B., Magnusson, S., Petersen, T. E., *J. Biol. Chem.* 1984, *259*, 8318–8327.

[18] Blake, C. C., Swan, I. D., Rerat, C., Berthou,J., Laurent, A., Rerat, B., *J. Mol. Biol.* 1971, *61*, 217–224.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 836–845

Ingrid Miller[1]
Paul Haynes[2]
Ivano Eberini[3]
Manfred Gemeiner[1]
Ruedi Aebersold[2]
Elisabetta Gianazza[3]

[1]Veterinärmedizinische Universität, Institut für Medizinische Chemie, Wien, Austria
[2]University of Washington, School of Medicine, Department of Molecular Biotechnology, Seattle, WA, USA
[3]Universitá degli Studi, Istituto di Scienze Farmacologiche, Milano, Italy

Proteins of rat serum: III. Gender-related differences in protein concentration under baseline conditions and upon experimental inflammation as evaluated by two-dimensional electrophoresis

We have previously described the major components of rat serum (*Electrophoresis* 1998, *19*, 1484–1492 and 1493–1500). In this report we examine sex-related differences in protein concentrations, both in control animals and upon experimentally induced inflammation. Under baseline conditions approximately one third of the spots resolved in serum by two-dimensional electrophoresis (2-DE) are expressed at levels ≥ 25% higher in female rats than in male rats and a further 10% at levels ≥ 25% lower. Inflammation increases the expression of the positive acute-phase reactants: hemopexin, ceruloplasmin, α_1-antitrypsin (all approximately 2-fold), C-reactive protein (3- to 5-fold), serine protease inhibitor-3 (4- to 5-fold), thiostatin (> 5-fold in females, > 20-fold in males), clusterin, orosomucoid, haptoglobin chains and α_2-macroglobulin. The baseline level of the last four markers is below the detection limit, hence no percent increase can be computed. Conversely, negative acute-phase reactants are reduced on inflammation: α_1-inhibitor III, α_2-HS-glycoprotein, kallikrein-binding protein and transthyretin (all reduced to between 1/2 to 1/3 of the baseline levels), retinol-binding protein (to about 1/2 to 1/4) and albumin (to 2/3). Except for thiostatin, the changes in acute-phase protein levels are similar in male and female rats.*

Keywords: Rat serum / Two-dimensional polyacrylamide gel electrophoresis / Inflammation / Sex / Acute-phase reactants
EL 3364

1 Introduction

Our groups have recently detailed the 2-DE map of rat serum by identifying the major protein components through a combination of immunoblotting and HPLC-MS/MS-based protein identification [1]. In a further step, the proteins typical of the acute-phase response to experimental inflammation have been identified [2]. Altogether, the above include 28 proteins, corresponding to about 80 individual spots. A compilation of all our rat data can be accessed [3, 4]. We wish to extend this database through the addition of further spot identifications and, furthermore, of annotations on physiological and pathological properties for individual components. One condition of special interest to our current research is experimental inflammation. We examine whether evaluation of serum levels of the acute-phase reactants may become a useful tool for the screening of anti-inflammatory drugs [5]. Detailed quantitative parameters characterizing the acute-phase response are a preliminary step towards this goal. Additional aspects of this investigation are sex-related differences in serum protein expression after inflammation as well as under control conditions; a typical animal model for chronic inflammation is adjuvant arthritis in female Lewis rats. Preliminary data [2] imply that the baseline concentrations of some of the acute-phase markers, *e.g.*, kallikrein-binding protein and thiostatin, differ significantly between male and female rats. For humans, in the compilation by Putnam [6], different normal concentration ranges are reported for men and women for α_2-macroglobulin, IgM, and β-lipoprotein.

Correspondence: Dr. Elisabetta Gianazza, Istituto di Scienze Farmacologiche, Facoltá di Farmacia, Universitá degli Studi di Milano, via Balzaretti 9, I-20133 Milano, Italy
E-mail: elisabetta.gianazza@unimi.it
Fax: +39-02-29404961

Abbreviations: apo, apolipoprotein; **CRP**, C-reactive protein; **Gc**, Gc-globulin; **Hp**, haptoglobin; **Hpx**, hemopexin; **α_1-I$_3$**, α_1-inhibitor III; **α_1-M**, α_1-macroglobulin; **α_1-MAP**, thiostatin (1 and 2); **α_2-M**, α_2-macroglobulin; **NC**, nitrocellulose; **OD**, optical density; **PAA**, polyacrylamide; **RBP**, retinol binding protein; **Tf**, transferrin

* Procedures involving animals and their care were conducted at Istituto di Scienze Farmacologiche, Universitá di Milano, in conformity with the institutional guidelines that comply with national (D.L. No. 116, G.U. Suppl. 40, February 18, 1992 and Circolare No. 8, G.U. July 1994) and international laws and policies (EEC Council Directive 86/609, OJ L 358.1, December 12, 1987, and Guide for the Care and Use of Laboratory Animals, US National Research Council, 1996).

2 Materials and methods

2.1 Samples

Three female and three male Sprague Dawley rats, 10 weeks old, 230–250 g, on a standard chow diet, were probed both before (control) and 48 h after intramuscular injection of 5 mL turpentine/kg body weight (as a stimulus inducing inflammation) [7]. Blood was drawn from the caudal vein under diethylether anesthesia at the beginning of the test period; the animals were killed by decapitation at the end of the treatment. Aliquots of the sera were stored at –20°.

2.2 Electrophoresis

Most analytical procedures were identical to those already detailed in the previous papers of this series [1, 2]. 2-DE maps were obtained by IPG-DALT [8] on 5 μL aliquots of sera. Under standard conditions, whole serum proteins were resolved on a nonlinear pH 4–10 IPG [9] and SDS-PAGE was then run on 7.5–17.5%T polyacrylamide (PAA) gradients; the slab size was 16 × 14 cm. Protein patterns were stained with 0.3% w/v Coomassie; α-mannose-containing glycoproteins were revealed by concanavalin A/peroxidase interaction [10, 11] after blotting onto a nitrocellulose (NC) membrane. In addition to the standard 2-DE protocol given above, enhanced resolution of individual areas or for a given class of proteins was achieved by variation of pH gradient or acrylamide concentration of the gels. Orosomucoid from 7.5 μL serum was focused on a pH 2.5–5 IPG [12] prior to SDS-PAGE as above. For glycoprotein detection (required for theresolution of even the basic isoforms of hemopexin), 2 μL of serum were run under standard conditions, then blotted on an NC membrane and affinity-stained with Concanavalin A-peroxidase [10]. For proteins in the crowded area around α_2-HS glycoprotein, 2-DE was also performed on a pH 4–6 IPG (the acidic 60% of the wide range above) followed by SDS-PAGE on 4.5–9%T PAA gradient gels. With high M_r proteins, showing minimal overlap in size, 1-DE gave satisfactory results even on 6 × 8 cm slabs. The best resolution was achieved by SDS-PAGE on a 4–6%T PAA gradient, with sample loads of 0.1 μL per lane for Coomassie stain and of 0.002 μL per lane for glycoprotein detection. Ceruloplasmin was immunostained with an antiserum raised in goats against the human protein (Sigma, St. Louis, MO, USA).

2.3 Protein quantitation

The 2-DE gels and the blots were scanned with a video camera (Sony, Japan) under the control of NIH Image, release 1.52, and analyzed with the software PDQUEST Version 5.1 (PDI, Huntington Station, NY, USA) run on a Sun SPARCstation 4 (Sun Microsystems, Mountain View, CA, USA). A matchset was created from the patterns of the four pools of rat sera (control males, inflamed males, control females and inflamed females) in a given run. A standard gel (master) was generated out of the image with the highest spot number by including additional spots from the other three gels [13]. As all gels within one series were processed "in one batch", results were evaluated in terms of spot volumes without further correction/normalization. The reports for individual proteins (spots, or spot chains, already identified by immunological or physicochemical means [1, 2]) (or single, unidentified spots) are referred to by percentage values of the highest figure among the four samples. Albumin was determined by the immediate bromocresol green assay [14, 15].

2.4 Protein identification by peptide mapping through HPLC-MS/MS

Some proteins influenced in their expression by inflammation and not already included in our database [3, 4] were identified by HPLC-MS/MS. After blotting and staining with Amido Black, material from individual spots was digested with trypsin according to [16]. The resulting peptides were resolved by chromatography on microbore HPLC system (Ultrafast Microprotein Analyzer; Michrom Bioresources, Auburn, CA, USA) equipped with a 100 Å, 5 μM Reliasil C18 BDX column (0.5 × 150 mm; Michrom). Peptides were eluted using buffer A (6% v/v acetonitrile, 0.1% acetic acid, 0.005% v/v heptafluorobutyric acid) and buffer B (80% v/v acetonitrile, 0.1% acetic acid, 0.005% v/v heptafluorobutyric acid), at a flow rate of 15 μL/min. The gradient used was 0% B for 2 min, followed by a linear increase to 60% B over 48 min. Tandem mass spectrometry was performed on-line in a triple quadrupole mass spectrometer (TSQ 7000; Finnigan, San Jose, CA, USA) connected to the HPLC *via* a microspray interface [17]. Spectra were scanned over a range of 380–1500 mass units at 1.5 s intervals. Automatic peak recognition and daughter ion scanning were performed using the built-in instrument control language (ICL) [18]. Tandem mass spectra were analyzed using SEQUEST, which allows the correlation of experimental data with theoretical spectra generated from known protein sequences [19, 20]. Briefly, SEQUEST matches a precursor ion of a given mass with any peptide present within the protein sequences of a given database, within a specified mass accuracy. An initial ranking is obtained by considering the presence and intensity of the peptide fragment ions in the experimental spectrum. A theoretical collision-induced dissociation (CID) spectrum of the five hundred best-ranked peptides is then generated and correlated with the experimental CID spectrum to find the best match. All

spectra were searched against the OWL protein database, release V30.1 [21, 22]. The criteria used for positive peptide identification were a correlation factor greater than 2.0, a delta cross-correlation factor greater than 0.1 (indicating a significant difference between the best match reported and the next best match), and a high preliminary score [17]. The search data were summarized using the sequest_summary program [19, 20], which assigns a dimensionless numerical value to search results based on a sliding scale where 10 units are given for a peptide ranked as the best match, 8 units for a second-best match, 6 for a third-best match, 4 for a fourth-best match and 2 for a fifth-best match. The scores assigned to peptides from a common protein are summed, to give a sequest_summary score for each potential protein match. The criterium we used for a positive protein identification was a sequest_summary score of greater than 20, indicating at least two peptides from a potential matching protein had been positively identified [17].

2.5 Lipoprotein purification and analysis

Total lipoproteins were collected by flotation on a KBr solution ($d = 1.21$ g/mL) and delipidated by treatment with diethylether:ethanol 3:1 v:v [23]. Equal amounts of the purified apoproteins (10 μg) were run in SDS-PAGE on a 4–20%T PAA gradient in order to resolve components of such different size as apoB and apoC at the same time.

3 Results

3.1 HPLC-MS/MS protein identifications

Three proteins not included in our previous database were identified by HPLC-MS/MS as being among the spots whose expression was modulated upon inflammation. Their positions in the 2-DE map are indicated (by ovals) in Fig. 1. The identifications of these three (Table 1) were α_1-inhibitor III [24–26], C-reactive protein (CRP) [27, 28] and clusterin [29, 30]. α_1-Inhibitor III was analyzed on a blot from 2-DE as well as on the SDS-PAGE of a control serum; clusterin was identified next to orosomucoid on a map from a 2.5–5 IPG. The component identified as CRP is located in the immunoglobulin light chain region (results from immunoblotting [1]) which extends to acidic p*I*s. One spot close to CRP had been previously identified as belonging to the immunoglobulin κ isotype (spot 50 of Fig. 2 in [1]) by HPLC-MS/MS in the serum of normal male rats. No peptides of this protein were found in the present analysis when investigating samples from normal and inflamed animals. Since immunoglobulin populations are known to be heterogeneous and variable, depending on the individuals (outbred mice share only 30%

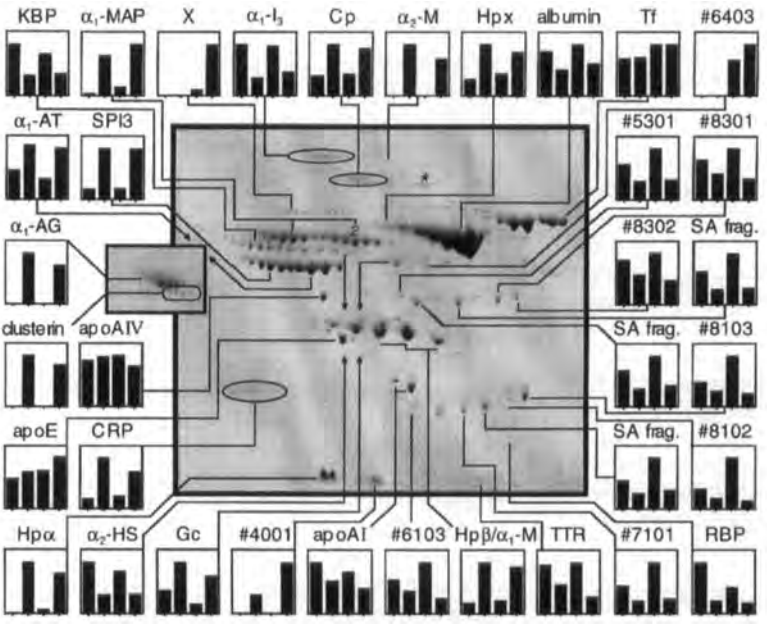


Figure 1. 2-DE map of serum proteins from female rats 48 h after the intramuscular injection of turpentine; Coomassie stain on a 5 μL sample. The inset is from a 2-DE gel whose first dimension was run on a 2.5–5 IPG, with a serum load of 7.5 μL. Histograms of the abundance of the various proteins, as marked on the top of each panel, in the sera of (from left to right) control males, inflamed males, control females, inflamed females. All data are referred to as % value to the highest figure among the four samples. Arrows point to the prolongation of broken lines (for α_1-AT, α_2-HS and Gc). α_1-AG, α_1-acid glycoprotein; α_1-AT, α_1-antitrypsin; α_1-I$_3$, α_1-inhibitor III; α_1-M, α_1-macroglobulin; α_1-MAP, thiostatin (1 and 2); α_2-HS, α_2-HS-glycoprotein; α_2-M, α_2-macroglobulin; apo, apolipoprotein; Cp, ceruloplasmin; CRP, C-reactive protein; Gc, Gc-globulin; Hp, haptoglobin; Hpx, hemopexin; KBP, kallikrein-binding protein; RBP, retinol-binding protein; SPI3, serine protease inhibitor-3; Tf, transferrin; TTR, transthyretin. The spots outlined by ovals correspond to proteins newly identified by HPLC-MS/MS; the asterisk marks a further spot identified as ceruloplasmin.

Table 1. Proteins identified by HPLC-MS/MS

Protein (abbreviation)	ID OWL [21]	Sequest_scores[a)]	Notes
α_1-Inhibitor III (α_1-I$_3$)	A1I3_RAT	Total of 368	32 peptide best matches + 3 peptides of s 15904 (a variant of A1I3_RAT)
α_2-Macroglobulin (α_2-M)	A2MG_RAT	144	Confirmation of data from immunoblotting [1]
Ceruloplasmin (Cp)	CERU_RAT	130	Overlap with RNITIH4
Cp*	CERU_RAT	128	
C-reactive protein (CRP)	CRP_RAT	118, 144, 146, 78	4 different analyses
Clusterin	CLUS_RAT	48, 68	2 different analyses
Inter-α-inhibitor H4	RNITIH4	52	

a) See Section 2.4 for further details

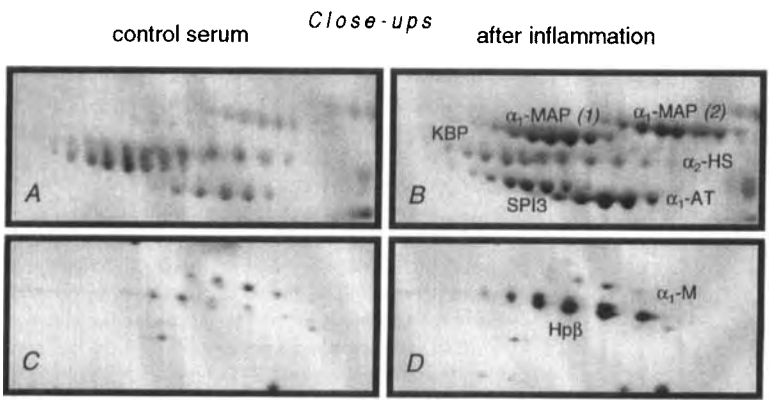


Figure 2. (A, B) Close-up of the 2-DE region of pH 4.5–5.5 and M_r 45–65 kDa for control female (on the left) and inflamed female animals (on the right). The second dimension was run on a 4.5–9%T slab. (C, D) Details from 2-DE using standard IPG, but 10–15% gels in second dimension, run for an extended period (migration distance for apoAI 12 cm instead of 6 cm) and silver-stained; patterns correspond to 0.2 µL serum. α_1-M light chain (upper row) and Hp β-chain (lower double row) are resolved. Control female animal on the left, inflamed female on the right.

of their Ig light chain spots, and inbred mice 70–80% [31]), it appears that this isotype was not present at significant levels in the animals used in these studies. HPLC-MS/MS was also used to identify some of the high M_r spots (Fig. 1). Ceruloplasmin [32, 33] was found to overlap with an inter-α-inhibitor H4 [34, 35]. Ceruloplasmin was also detected at the position marked with an asterisk. One further protein, previousy detected only by immunoblotting, was confirmed by HPLC-MS/MS to be α_2-macroglobulin.

3.2 Protein evaluation

2-DE evaluation of the rat sera for spot quantitation was only performed on gel series which had been processed in the same batch (in all steps, *i.e.*, from IPG to destaining after 2-DE and scanning) to keep variation levels as low as possible. A constant volume of serum was loaded onto the gel, although samples from inflamed animals are known to vary slightly in protein concentration and distribution [7]. All quantitations were based on (absolute) spot volumes; it was *a priori* not clear which protein was unaffected in its level and could serve as internal standard. As serum protein content varies upon inflammation [7], % based on total optical density (OD) in a given slab appear less reliable than constant serum volume. The reproducibility of the system was tested by running several gels of the same sample (male rat serum), in different batches. Gels were evaluated in the way described, and variation was determined for all proteins. Variation coefficients whithin one batch were 20% on average, smaller for compact, round and darker spots, *e.g.*, transferrin (Tf) 7–9%, apolipoprotein (apo) AIV 7–11%, haptoglobin (Hp) β/α_1-M 9% and higher for the more diffuse, fainter, and oval-shaped ones, *e.g.*, α_1-inhibitor III (α_1-I$_3$) within 36%. No correlation between M_r of the protein and variation coefficient was found.

3.3 Protein concentration in control and inflamed rats

The master gel pattern produced by PDQUEST software contains 151 different single spots resolved under standard conditions (as in Fig. 1). For males the number of

quantitated spots is lower (113 for control animals, 124 for inflamed) than for females (120 for control, 148 for inflamed animals), with more saturated spots for females than for males (mainly albumin and Tf). Under baseline conditions, 29% of the single spots are expressed at levels at least 25% higher in females than in males, 10% at levels at least 25% higher in males than in females. Upon experimental treatment, an increase above control levels exceeding 25% is recorded for 25% of the protein spots in males and for 28% in females, a decrease exceeding 25% for 32 and 34% of the spots, respectively.

The largest number of protein spots is found in the serum of inflamed female rats; the 2-DE map of such a sample is shown in Fig. 1, with an inset from a 2-DE gel from a 2.5–5 IPG to fully resolve orosomucoid and clusterin. A close-up details the region with the most dramatic changes in protein patterns due to the treatment, *i.e.*, the region of p*I* 4.5–5.5 and molecular mass 45–65 kDa (Fig. 2, panels A–B). For identified proteins, single spot volumes of the respective chain were summed and plotted, while unidentified spots were treated individually. Figure 1 shows the identified proteins and single spots of high intensity (cut-off chosen arbitrarily). Histograms are linked to the protein, or protein spot, to which each refers, and are marked with the protein abbreviation, or a spot number (according to PDQUEST listing). For one as yet unidentified protein, designated X, the sum of volumes for all spots in a row was computed. In each histogram, the series, from left to right, includes: control males, inflamed males, control females, inflamed females.

Thiostatin (α_1-MAP) is five times higher in females than in males, hemopexin (Hpx) almost 50% higher, CRP and apoE *ca.* 25% higher in females. Gc-globulin (Gc) and retinol-binding protein (RBP) are about twice as high in males, and α_2-HS glycoprotein and α_1-antitrypsin about 25 and 30% higher in males than in females, respectively. Once expressed after inflammation, the concentration of α_2-macroglobulin is about 40% higher, and that of orosomucoid about 30% higher in male rats. Among the nonidentified proteins, protein X and spot 6403 are found only in females, being either strongly or slightly up-regulated in inflammation. Spot 4001 appears only in the course of treatment, and several others are reduced in concentration (#5301, 7101, 8103, 8301, 8302).

Variations in serum levels are brought about by the experimental treatment for virtually all proteins, although the order of magnitude of the changes differs widely. Among the major identified proteins only transferrin does not seem to change its absolute concentration (also see [5]). Inflammation increases the expression of the positive acute-phase reactants hemopexin, ceruloplasmin and α_1-antitrypsin (all approximately 2-fold), C-reactive protein (3- to 5-fold), serine protease inhibitor-3 (4- to 5-fold), thiostatin (> 5-fold in females, > 20-fold in males), clusterin, orosomucoid, haptoglobin chains, and α_2-macroglobulin. The baseline level of the last four markers is below the detection limit, hence no percent increase may be computed. Upon inflammation orosomucoid represents *ca.* 2–3%, haptoglobin α-chain 3–4%, haptoglobin β (together with α_1-macroglobulin (α_1-M) 13–14% and α_2-macroglobulin 5% of total serum proteins. Among the nonidentified components, spot #4001 is only detected after inflammation. Conversely, negative acute-phase reactants are reduced on inflammation: for example, α_1-inhibitor III, α_2-HS glycoprotein, kallikrein-binding protein and transthyretin (all of them being reduced to 1/2 or 1/3 of their baseline levels), retinol binding protein (to 1/2 or 1/4), transthyretin (to 1/2 to 1/3) as well as a number of relatively faint, unidentified protein spots (not included in the histograms of Fig. 1). Albumin is reduced to about 2/3 of its concentration by inflammation; the same is observed for some spots previously identified as albumin fragments [1].

As for apolipoproteins, the amount of apoE is increased, while that of apoAI is decreased by inflammation (Fig. 1). If the percent concentrations of the various apoproteins in a purified lipoprotein preparation are computed (Fig. 3), the reduction for apoAI equals 50%, while the level of apoB doubles. As discussed (see [5]), the distribution among p*I* isoforms is not influenced by the experimental treatment (not shown).

Concordance is usually observed between the sexes as for the extent of the changes following inflammation. A strong difference is only found for thiostatin, for which the increase in males is four times higher than in females.

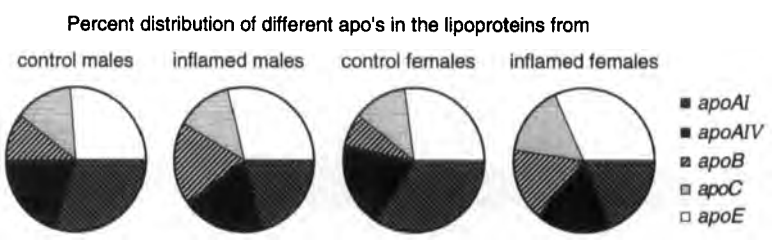


Figure 3. Percent distribution of the various apoproteins in the lipoproteins of (from left to right) control males, inflamed males, control females, inflamed females.

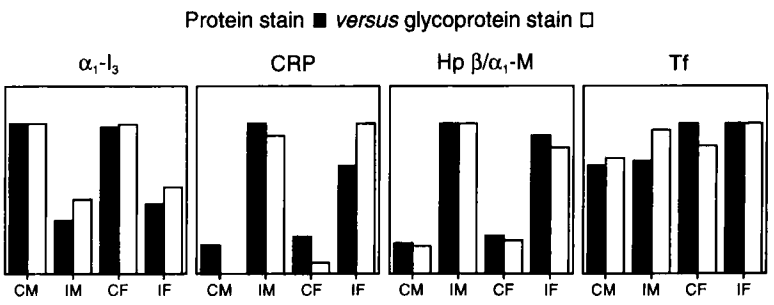


Figure 4. Comparison between the quantitations from Coomassie-stained pattern (solid bars) and concanavalin-positive affinity blotting for the glycyl moiety (open bars) of some glycoproteins. In each series the sequence of the samples is, from left to right: control males (CM), inflamed males (IM), control females (CF), inflamed females (IF); for each type of staining, data are referred to as percentage values relative to the highest figure among the four samples.

4 Discussion

4.1 Assessment of the quantitation procedures

As detailed under Section 2.3, as a second step after 2-DE under standard conditions, the quantitations for many proteins were repeated under electrophoretic conditions maximizing resolution in specific areas. A countercheck of values revealed consistency of the trends observed throughout all the methods, although absolute spot volumes differed. The spots of the two proteins with the highest concentration in serum, albumin and Tf, were usually saturated in 2-DE with standard sample loads. Transferrin could be evaluated after 1-D SDS-PAGE on a 4–6%T gel. Albumin was determined by a photometric assay based on the reaction with bromocresol green [36]. Two proteins could not be fully resolved from neighbors with overlapping molecular properties. In the Hpx/albumin pair, lectin affinity staining was able to clearly differentiate the glycoprotein from the nonglycoprotein component and was used for quantitation (histogram in Fig. 1). The spots of α_1-macroglobulin light chain and haptoglobin β-chain overlap almost completely in the standard 2-DE system; both proteins are glycosylated and no cross-reactive antibody is available [2]. Improved resolution in 2-DE was obtained by prolonging the second-dimension runs, lowering sample loading amounts, and using silver staining as a detection method [37, 38]. In this way, α_1-M (a single row) was found to remain constant, whereas the levels of Hp β (a double chain) rose 9–10 times (Fig. 2 C–D).

4.2 Newly identified proteins

In the course of our new experiments several spots were investigated in more detail, especially those showing variations in serum levels upon treatment. All of the proteins newly identified by HPLC-MS/MS showed considerable differences from the human homologs. CRP, a pentameric glycoprotein with identical subunits, displays a much more acidic p*I* in rats than in humans. It is one of the early markers of inflammation in humans [39] and is reported to be little affected in rats (by a factor of 2 *versus* 1000 [40]). In contrast to the human counterpart, rat CRP is present in serum at higher baseline levels [41], it is a glycoprotein (see Fig. 4) [28], and its binding properties for phosphorylcholine differ (it contains three instead of five binding sites [42]). Rat clusterin is also much more acidic than the human homolog. In mice and humans, this protein, an extensively glycosylated disulfide-linked heterodimer with subunits of similar M_r, is known as apo J and is associated with apoptosis and inflammation [29]. α_1–I$_3$ is a rat-specific protein, belonging to the macroglobulin family and acting as a protease inhibitor [25]. Whereas the first two proteins described are increased upon inflammation, the level of α_1–I$_3$ drops.

During the search for new inflammation markers, two high M_r spots were identified as ceruloplasmin (Cp) by HPLC-MS/MS. The previously described immunostaining of ceruloplasmin [1] had been performed with a weakly cross-reacting antibody raised against the human protein. It produced multiple spots or spot chains of different intensity, among them the position indicated in Fig. 2 of [1] and the spot named Cp in Fig. 1 of this paper. When performing the immunoblot after 1-D SDS-PAGE, there was also a fainter double band with lower M_r. As it could not be verified whether the latter was identical to spot Cp* (Fig. 1), quantitation was only based on the main cross-reactive component. Human Cp is thought to be a single-chain polypeptide, but was found as a major type I with four sugar chains and a minor type II with three sugar chains [43], which could also correspond to the general Cp pattern given in SWISS-2DPAGE [42]. Rat Cp is reported to be more acidic than the human homolog [32] and contains an even larger number of potential glycosylation sites (6 *versus* 4). However, these different properties cannot completely explain the spot pattern in our 2-DE runs.

Table 2. Changes reported for rat serum proteins upon inflammation

Protein	Type of change	Reference
α_1-AG	↑ 20 ×	Urban *et al.*, 1979 [60]
	↑ 20 × in 24 h,	
	↑ 14 × in day 3	Schreiber *et al.*, 1982 [7]
α_1-AT	↑ 2 × in day 2	Gauthier *et al.*, 1978 [24]
	↑ 2 ×	Urban *et al.*, 1979 [60]
	↑ <1.50 ×	Koj *et al.*, 1982 [61]
α_1-I_3	↓ 0.3 × in day 2	Gauthier *et al.*, 1978 [24]
	↓ 0.5 × in day 2	Lonberg-Holm *et al.*, 1987 [25]
	↓ 0.3–0.25 × in 24 h	Braciak, 1988 [26]
	↓ 0.25–0.2 × in 24 h	Aiello *et al.*, 1988 [62]
α_1-M	↑ 1.4 × in day 3	Gordon *et al.*, 1976 [63]
	↔	Gauthier *et al.*, 1978 [24]
	↔	Lonberg-Holm *et al.*, 1987 [25]
	↑ <1.5 ×	Koj *et al.*, 1982 [61]
α_1-MAP	↑ 20 × in 24 h	Urban *et al.*, 1979 [60]
	↑ 20 × in 24 h,	
	↑ 18 × in day 3	Schreiber *et al.*, 1982 [7]
	↑ 20 ×	Cole *et al.*, 1985 [64]
α_2-HS	↓	Schreiber *et al.*, 1988 [45]
α_2-M	↑ 90 × in day 2	Gordon *et al.*, 1976 [63]
	↑ (rapid)	Gauthier *et al.*, 1978 [24]
	↑ 110–140 × in 24 h	Okubo *et al.*, 1981 [65]
	↑ 70 × in day 2	Lonberg-Holm *et al.*, 1987 [25]
Albumin	↓ 0.6 ×	Urban *et al.*, 1979 [60]
	↓ 0.4 × in 24 h,	
	↓ 0.6 × in day 3	Schreiber *et al.*, 1982 [7]
ApoAI	↔	Schreiber *et al.*, 1989 [45]
ApoAIV	↓	Schreiber *et al.*, 1989 [45]
ApoE	↓	Schreiber *et al.*, 1989 [45]
Clusterin	Associated with	
	apoptosis	[42]
Cp	↑	Schreiber *et al.*, 1989 [45]
CRP	↑ 2 ×	Kushner *et al.*, 1988 [41]
	Undetectable	Schreiber *et al.*, 1989, [45]
Gc	↔	Schreiber *et al.*, 1989 [45]
Hp	↑ 10 ×	Lombart *et al.*, 1965 [66]
	↑ 3 ×	Koj *et al.*, 1982 [61]
Hpx	↑	Kushner, 1984 [40]
KBP	↓	Chao *et al.*, 1990 [67]
	↓	Pages *et al.*, 1990 [68]
	↓	Chai *et al.*, 1991 [47]
RBP	↓	Schreiber *et al.*, 1989 [45]
SPI-3	↑	Pages *et al.*, 1990 [68]
Tf	↑ 1.3 × in 24 h	Schreiber *et al.*, 1982 [7]
Transthyretin	↓	Schreiber *et al.*, 1989 [45]

↑, increased
↔, no change
↓, decrease in concentration

4.3 Changes in protein levels

With ELISA, Ikawa and Shozen [44] have determined normal values of several acute phase proteins in an (unspecified) rat strain. The detection system was validated and found to have a variation of less than 10% for most proteins, but the concentration ranges reported for some of the analytes vary between 10 and 25% (*e.g.*, Hp, Hpx, orosomucoid). Sprague-Dawley as an outbred strain may also be expected to display a certain degree of variation in serum protein levels. Most changes noticed in inflammation are much more marked than the error in our quantitation (determined in the reproducibility testing) or the normally expected variation in the respective rat population.

Several procedures have been reported in the literature to induce inflammation processes, most often injection of turpentine. However, the strength of the stimulus used varied among authors, as well as the time intervals after which the shifts in protein levels were recorded (usually one to three days), which results in differences in the magnitude of the changes described. For an overview on previous results, see Table 2. As for the major proteins, our present findings confirm current knowledge about positive and negative acute-phase reactants, although the extent of variation we report sometimes differs from literature data. As for CRP, we could only confirm a much lower rise than in humans [40] but still a clear-cut increase in its levels. Contrary to Schreiber *et al.* [45] who report Gc and apoAI not to be affected by the acute-phase response, we observe an increase for the former and a decrease for the latter. Rat Tf has been reported to change little in inflammation [7] and has therefore been suggested for reference as an internal standard. Tf data at the time point chosen for our present investigation would support this hypothesis; however, detailed studies in time course experiments provide evidence for a more complex behavior [5]. Some of the changes in protein levels brought about by acute experimental inflammation are reproduced by such different noxious stimuli as exposure to solvents in toxicological tests (dimethylformamide [46]).

Slight to moderate differences were found in baseline levels for single proteins when comparing male and female rats. Sevenfold higher kallikrein-binding protein (KBP) and α_1-antitrypsin (α_1-AT) values have been reported for male rats [47]; a similar trend was noticed in our animals, but not exceeding a factor of 2 [2]. Gc and RBP level of our male and female controls also varied. The most marked difference was the almost complete absence of thiostatin (α_1-MAP) in the serum of healthy males, whereas it was clearly detectable in control females [2]. Except for the latter, the changes in acute-phase protein levels were similar in male and female rats. α_2-Macroglobulin was higher in inflamed males, in agreement with earlier reports [48].

The major acute-phase protein (α_1-MAP) of rat serum has been given different synonymous designations, to stress one or another of its functions. It is also named T-kininogen 1 precursor, after the sequence of T-kinin it contains, and thiostatin, after its inhibitory effect on cysteine proteases. Two genes exist for this protein, resulting from a duplication event: MAP1 and MAP2 [49, 50]. Two types of proteins are indeed found in plasma, differing in their p*I* by about 0.3 pH units, in agreement with computation based on amino acid composition [42]. Their baseline levels are different: α_1-MAP (2) is more abundant than α_1-MAP (1) in female rats, and is also the only form present (in very small amounts) in control males. The concentration of each peptide and, in detail, of each glycoform, appears to be raised to a similar extent by acute inflammation (not shown). Strict coregulation of α_1-MAP (1) and α_1-MAP (2) is not a rule, however. Pharmacological treatments [5] as well as different phlogistic stimuli may bring about a differential modulation of the two peptides.

The glycosylation pattern of serum proteins has been reported to change under various physiological and pathological conditions [51], including inflammation [52]. In humans, detailed studies have been devoted, for instance, to Tf [53] and to orosomucoid [54]. In the present investigation no systematic comparison of the differences in glycosylation between control and inflamed pattern was performed since the concentrations were adjusted to the detection of few proteins of specific interest while the stainability of the various glycoproteins differs widely (as shown in [2]). Quantitations after protein blotting are further complicated by the limited binding capacity of the NC membrane as well as by the narrow dynamic range of the zymograming techniques. However, among the assayed serum components, for α_1–I_3 and Tf the protein secreted during inflammation appears more glycosylated than under basal conditions (Fig. 4).

4.4 Biological aspects

It has been shown that the concentration of about 30% of the liver proteins differs between male and female Wistar rats, the reason for this gender-related distinction being the level of estrogen [55]. It is not surprising that a similar trend is likewise observed for the plasma proteins, also synthesized by the liver and then secreted to the circulatory stream instead of being released within the parenchymal cells. With a cut-off at 25% difference in baseline levels, and with reference to individual spots, approximately 40% of the serum components were found to be present at higher concentrations in females, while only 10% of the serum components were found to be higher in males. For almost all proteins the responses to the inflammatory stimulus were found to be similar in male and female rats. The effects of the mediators of inflammation and of the steroid hormones thus appear additive and not synergistic. The only major exception to this trend is thiostatin, for which, however, the baseline level for most of the isoforms in males is almost below the detection limit so that computations of percent increase are highly inaccurate.

The overall impression from the observations above is that serum proteins behave as a coordinate system. In fact, the concentration of virtually all serum components is affected by inflammation, which means – owing to our selection on data reduction – that their synthesis rate is modified. The total protein concentration in inflamed sera changes slightly [7]. A significant drop is observed for the most abundant among the serum proteins, albumin. It might be speculated that its levels could decrease during inflammation by a compensatory effect, so that down-regulation would not occur at the genetic level, but rather as a result of competition for available resources, such as aminoacyl-tRNA or ATP. The opposite situation has been described in detail, namely the compensatory adjustments observed in the serum of analbuminemic rats [56]; the concentration of globulins (α_1-AT, Tf, Cp, IgG, IgA and IgM but not orosomucoid) rises [57], with especially large effects on high M_r components such as α_2-M (whose concentration increases 12-fold) and α_1-I_3 (whose concentration doubles) [58]. Also, in humans, lack of albumin is compensated for so as to obtain unchanging total protein concentration and oncotic pressure [59].

The authors thank Profs. Philippe Arnaud and Adriana Maggi for fruitful discussion. When Papers I and II of this series were prepared, E.G. was awfully inept with all image treatment programs; she is very grateful to Mr. Antonio Del Rio for all his teaching and all his help. P.H. and R.A. wish to acknowledge funding support from the National Science Foundation as part of the Science and Technology Center for Molecular Biotechnology at the University of Washington.

Received October 31, 1998

5 References

[1] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, M. R., Manzoni, C., Vignati, M., Gianazza, E., *Electrophoresis* 1998, *19*, 1484–1492.

[2] Miller, I., Haynes, P., Gemeiner, M., Aebersold, R., Manzoni, C., Lovati, M. R., Vignati, M., Eberini, I., Gianazza, E., *Electrophoresis* 1998, *19*, 1493–1500.

[3] http://weber.u.washington.edu/~ruedilab/aebersold.html

[4] http://imiucca.csi.unimi.it/~ratserum/homeframed.html

[5] Eberini, I., Miller, I., Zancan, V., Bolego, C., Puglisi, L., Gemeiner, M., Gianazza, E., *Electrophoresis* 1999, *20*, 846–853.

[6] Putnam, F. W. (Ed.), *The Plasma Proteins* Academic Press, Orlando 1975, 1977, 1984, Vol. I–IV.

[7] Schreiber, G., Howlett, G., Nagashima, M., Millership, A., Martin, H., Urban, J., Kotler, L., *J. Biol. Chem.* 1982, *257*, 10271–10277.

[8] Gianazza, E., in: Link, A. J. (Ed.), *2-D Proteome Analysis Protocols*, Humana Press, Totowa 1998.

[9] Gianazza, E., Giacon, P., Sahlin, B., Righetti, P. G., *Electrophoresis* 1985, *6*, 53–56.

[10] Hawkes, R., *Anal. Biochem.* 1982, *123*, 143–146.

[11] Taketa, K., *Electrophoresis* 1987, *8*, 409–414.

[12] Gianazza, E., Celentano, F., Ettori, C., Righetti, P. G., *Electrophoresis* 1989, *10*, 806–808.

[13] Garrels, J. I., *J. Biol. Chem.* 1989, *25*, 5269–5282.

[14] Rodkey, F. L., *Clin. Chem.* 1965, *11*, 478–487.

[15] Doumas, B. T., Watson, W. A., Biggs, H. G., *Clin. Chim. Acta* 1971, *31*, 87–96.

[16] van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Biochem.* 1997, *247*, 310–318.

[17] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, *19*, 939–945.

[18] Ducret, A., van Oostveen, I., Eng, J. K., Yates, J. R. I., Aebersold, R., *Prot. Sci.* 1997, *7*, 706–719.

[19] Eng, J., McCormack, A. L., Yates, J. R. I., *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989.

[20] Yates, J. R. I., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1426–1436.

[21] http://bmbsgi11.leeds.ac.uk/bmb5dp/owl.html

[22] Bleasby, A. J., Akrigg, D., Attwood, T. K., *Nucleic Acids Res.* 1994, *22*, 3574–3577.

[23] Schumaker, V. N., Puppione, D. L., *Methods Enzymol.* 1986, *128*, 155–170.

[24] Gauthier, F., Ohlsson, K., *Hoppe-Seylers Z. Physiol. Chem.* 1978, *359*, 987–992.

[25] Lonberg-Holm, K., Reed, D. L., Roberts, R. C., Hebert, R. R., Hillman, M. C., Kutney, R. M., *J. Biol. Chem.* 1987, *262*, 438–445.

[26] Braciak, T. A., Northemann, W., Hudson, G. O., Shiels, B. R., Gehring, M. R., Fey, G. H., *J. Biol. Chem.* 1988, *263*, 3999–4012.

[27] Rassouli, M., Sambasivam, H., Azadi, P., Dell, A., Morris, H. R., Nagpurkar, A., Mookerjea, S., Murray, R. K., *J. Biol. Chem.* 1992, *267*, 2947–2954.

[28] Sambasivam, H., Rassouli, M., Murray, R. K., Nagpurkar, A., Mookerjea, S., Azadi, P., Dell, A., Morris, H. R., *J. Biol. Chem.* 1993, *268*, 10007–10016.

[29] Collard, M. W., Griswold, M. D., *Biochemistry* 1987, *26*, 3297–3303.

[30] Bettuzzi, S., Hilipakka, R. A., Gilna, P., Liao, S. T., *Biochem. J.* 1989, *257*, 293–296.

[31] Anderson, N. L., *Immunol. Lett.* 1981, *2*, 195–199.

[32] Manolis, A., Cox, D. W., *Prep. Biochem.* 1980, *10*, 121–132.

[33] Fleming, R. E., Gitlin, J. D., *J. Biol. Chem.* 1990, *265*, 7701–7707.

[34] Daveau, M., Jean, L., Soury, E., Olivier, E., Masson, S., Lyoumi, S., Chan, P., Hiron, M., Lebreton, J. P., Husson, A., Jegou, S., Vaudry, H., Salier, J. P., *Arch. Biochem. Biophys.* 1998, *350*, 315–323.

[35] Soury, E., Olivier, E., Daveau, M., Hiron, M., Claeyssens, S., Risler, J. L., Salier, J. P., *Biochem. Biophys. Res. Commun.* 1998, *243*, 522–530.

[36] Keay, G., Doxeny, D. L., *Res. Vet. Sci.* 1983, *35*, 58–60.

[37] Heukeshoven, J., Dernick, R., in: Radola, B. J. (Ed.), *Elektrophorese Forum '86*, Technical University, Munich 1986, pp. 22–27.

[38] Miller, I., Gemeiner, M., *Electrophoresis* 1992, *13*, 450–453.

[39] Schwick,H. G., Haupt, H., *Behring Inst. Mitt.* 1986, *80*, 1–10.

[40] Kushner, I., in: Arnaud, P., Benvenu, J., Laurent, P. (Eds.), *Marker Proteins in Inflammation*, W. deGruyter, Berlin 1984, Vol. 2, pp. 3–14.

[41] Kushner, I., *Methods Enzymol.* 1988, *163*, 373–383.

[42] http://expasy.hcuge.ch/sprot/

[43] Arnaud, P., Gianazza, E., Miribel, L., *Methods Enzymol.* 1988, *163*, 441–452.

[44] Ikawa, M., Shozen, Y., *J. Immunol. Methods* 1990, *143*, 101–106.

[45] Schreiber, G., Tsykin, A., Aldred, A. R., Thomas, T., Fung, W.-P., Dickson, P. W., Cole, T., Birch, H., De Jong, F. A., Milland, J., *Ann. N. Y. Acad. Sci.* 1989, *557*, 61–85.

[46] Marshall, T., Williams, K. M., Vesterberg, O., *Electrophoresis* 1985, *6*, 392–398.

[47] Chai, K. X., Ma, J.-X., Murray, S. H., Chao, J., Chao, L., *J. Biol. Chem.* 1991, *266*, 16029–16036.

[48] Baldo, B. A., Chow, S. C., Euers, C., *Agents Actions* 1981, *11*, 482–489.

[49] Kageyama, R., Kitamura, N., Ohkubo, H., Nakanishi, S., *J. Biol. Chem.* 1985, *260*, 12060–12064.

[50] Kitagawa, H., Kitamura, N., Hayashida, H., Miyata, T., Nakanishi, S., *J. Biol. Chem.* 1987, *262*, 2190–2198.

[51] Baenziger, J. U., in: Putnam, F. W. (Ed.), *The Plasma Proteins*, Academic Press, Orlando 1984, Vol. 4, pp. 271–315.

[52] Van Dijk, W., Mackiewicz, A., *Ann. N. Y. Acad. Sci.* 1995, *762*, 319–330.

[53] de Jong, G., van Noort, W. L., Feelders, R. A., de Jeu-Jaspars, C. M., van Eijk, H. G., *Clin. Chim. Acta* 1992, *212*, 27–45.

[54] Hansen, J. E., Bøg-Hansen, T. C., Pedersen, B., Neland, K., *Electrophoresis* 1989, *10*, 574–578.

[55] Anderson, N. L., Giere, F. A., Nance, S. L., Gemmell, M. A., Tollaksen, S. L., Anderson, N. G., in: Galteau, M.-M., Siest, G. (Eds.), *Recent Progresses in 2D Electrophoresis*, Presses Universitaires de Nancy, Nancy 1986, pp. 253–260.

[56] Nagase, S., Shimamune, K., Shumiya, S., *Science* 1979, *205*, 590–591.

[57] Emori, T., Takahashi, M., Sugiyama, K., Shumiya, S., Nagase, S., *Jikken-Dobutsu* 1983, *32*, 123–132.

[58] Stevenson, F. T., Greene, S., Kaysen, G. A., *Kidney Int.* 1998, *53*, 67–75.

[59] Peters, T., *Adv. Protein Chem.* 1985, *37*, 161–245.

[60] Urban, J., Chan, D., Schreiber, G., *J. Biol. Chem.* 1979, *254*, 10565–10568.

[61] Koj, A., Dubin, A., Kasperczyk, H., Bereta, J., Gordon, H., *Biochem. J.* 1982, *206*, 545–553.

[62] Aiello, L. P., Shia, M. A., Robinson, G. S., Pilch, P. F., Farmer, S. R., *J. Biol. Chem.* 1988, *263*, 4013–4022.

[63] Gordon, A. H., *Biochem. J.* 1976, *159*, 643–650.

[64] Cole, T., Inglis, A. S., Roxburgh, C. M., Howlett, G. J., *FEBS Lett.* 1985, *182*, 57–61.

[65] Okubo, H., Miyanaga, O., Nagano, M., Ishibashi, H., Kudo, J., Ikuta, T., Shibata, K., *Biochim. Biophys. Acta* 1981, *668*, 257–267.

[66] Lombart, C., Dautrevaux, M., Moretti, J., *Biochim. Biophys. Acta* 1965, *97*, 270–274.

[67] Chao, J., Chai, K. X., Xiong, W., Chao, S., Woodley-Miller, C., Wang, L., Lu, H. S., Chao, L., *J. Biol. Chem.* 1990, *265*, 16394–16401.

[68] Pages, G., Rouayrenc, J. F., Le Cam, G., Mariller, M., Le Cam, A., *Eur. J. Biochem.* 1990, *190*, 385–391.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 846–853

Ivano Eberini[1]
Ingrid Miller[2]
Valeria Zancan[1]
Chiara Bolego[1]
Lina Puglisi[1]
Manfred Gemeiner[2]
Elisabetta Gianazza[1]

[1]Università degli Studi, Istituto di Scienze Farmacologiche, Milano, Italy
[2]Veterinärmedizinische Universität, Institut für Medizinische Chemie, Wien, Austria

Proteins of rat serum IV. Time-course of acute-phase protein expression and its modulation by indomethacine

Changes in the concentration of major serum proteins were monitored from day 0 to day 4 in three experimental groups: rats injected with turpentine, rats receiving the turpentine shot and daily doses of indomethacine, and rats given indomethacine alone. In inflamed animals, peak changes for acute-phase reactants, evaluated by two-dimensional electrophoresis (2-DE), were usually observed between 48 and 72 h after the phlogistic stimulus. By itself, indomethacine was found to affect the synthesis of most proteins (except one of the thiostatin variants and ceruloplasmin); the changes in serum levels, whether positive or negative, were the same as upon inflammation (except for kallikrein-binding protein), but their extent and/or timing usually differed. When inflamed animals were given indomethacine, a clear-cut difference in the concentration of some proteins was observed *versus* inflamed rats not given medication, at 24 h after the start of the treatments. Proteins mainly affected were α_2-macroglobulin, α_2-HS-glycoprotein, C-reactive protein and kallikrein-binding protein.*

Keywords: Rat serum / Two-dimensional polyacrylamide gel electrophoresis / Inflammation / Anti-inflammatory drug / Indomethacine
EL 3365

1 Introduction

Human pathology recognizes both acute and chronic inflammatory diseases [1]. The former, often of bacterial or viral origin (carditis, appendicitis, meningitis, mucositis, hepatitis), may be life-threatening, with massive and severe symptoms (including high fever) and a short-term evolution. They entail accumulation of fluid and plasma components in the affected tissue, intravascular stimulation of platelets, and the migration of polymorphonuclear leukocytes. The etiology of the chronic inflammations is usually unknown, so no causal but only symptomatic therapy may be envisaged. Their onset is gradual; the main symptom, delayed in time, is pain, with occasional fever; the involved cell populations are lymphocytes, plasma cells, and macrophages. Examples of chronic inflammations are arthritis, psoriasis, and lupus erithematosus.

Both *in vivo* and *in vitro* models of inflammation are in current use. The former rely on the development of acute edema after intramuscular injection of carrageenin or subcutaneous insertion of cotton pellets (both causing an aseptic granuloma) [2–5]; swelling of the tissues as assessed by pletismometry is the reference parameter in these experimental systems. In the latter, both mediators of inflammation (eicosanoids, kinins, cytokines) and the enzymes involved in their synthesis are quantitated after suitable stimulation of cells in culture [6–9]. To reliably evaluate a systemic response such as inflammation and the interference by drugs on its progression, only the *in vivo* approach seems completely appropriate. Experimental protocols that could reduce cost and labor while not reducing the sensitivity of the assay would represent a major advancement. The present investigation is indeed meant as a first step towards the assessment of whether the analysis of the 2-DE pattern of serum proteins may be a novel approach for the screening of anti-inflammatory drugs, possibly recognizing both their therapeutic value and/or their toxicity.

Quantitative evaluation of 2-DE maps of biological samples was introduced as a powerful tool for the assessment of physiological and pathological states [10], and has

Correspondence: Dr. Eliseabetta Gianazza, Istituto di Scienze Farmacologiche, Facoltà di Farmacia, Università degli Studi di Milano, via Balzaretti 9, I-20133 Milano, Italy
E-mail: elisabetta.gianazza@unimi.it
Fax: +39-02-29404961

Abbreviations: APR, acute-phase reactant; **CRP,** C-reactive protein; α_1-**I$_3$,** α_1-inhibitor III; **KBP,** kallikrein-binding protein; **LDL,** low density lipoprotein; α_2-**M,** α_2-macroglobulin; **NSAID,** nonsteroidal anti-inflammatory drug; **PAA,** polyacrylamide; **SPI3,** serine protease inhibitor-3; **Tf,** transferrin; α_1-**MAP,** thiostatin

* Procedures involving animals and their care were conducted at Istituto di Scienze Farmacologiche, Universitá di Milano, in conformity with the institutional guidelines that comply with national (D.L. No. 116, G.U. Suppl. 40, February 18, 1992 and Circolare No. 8, G. U. July 1994) and international laws and policies (EEC Council Directive 86/609, OJ L 358.1, December 12, 1987, and Guide for the Care and Use of Laboratory Animals, US National Research Council, 1996).

been applied extensively to pharmacological and toxicological investigations [11–21]. One of the main bonuses of such a technique is that tens to hundreds of individual components may be evaluated at once, and that no previous selection of the analytical target is required. In all the cases referred to above, the analysis was performed on liver homogenates. Serum would in principle be a most expedient substitute for autoptic specimens. In fact, repeated sampling from the same animals could allow, without major trauma, a direct comparison of baseline and post-treatment levels of the test proteins, thus reducing the deviance caused by intraspecific individual variability.

Some toxicological investigations using rat serum have been performed using 2-DE [22, 23]. In a few cases, markers of inflammation in rat serum have been addressed with (simpler) methodological approaches. Electrophoresis on cellulose acetate was used for the differentiation of various experimentally induced pathological processes, including adjuvant arthritis as a model of chronic inflammation [24]. Qualitative and quantitative changes of Con A-reactive proteins and α_2-macroglobulin were examined by lectin- and immunoblots and by radioimmunoassay, after chronic inflammation by injection of Freund's complete adjuvant; a test drug, bindarit (2-[(1-benzyl-indazol-3-yl)methoxy]-2-methyl propionic acid), was given to the animals in comparison with indomethacine [25].

Our detailed knowledge of the major proteins of rat serum both in control conditions [26] and upon experimental inflammation [27, 28] should allow adequate monitoring of changes in their levels, reflecting both therapeutic and toxic effects of test drugs. In this first phase of the investigation the same stimulus was used that had already been assayed, an intramuscular injection of turpentine. The progression of the acute-phase response was assessed with time, and the effects of a typical anti-inflammatory drug, indomethacine, when given either together with the inflammatory shot or by itself, was monitored.

2 Materials and methods

2.1 Samples

Male Sprague-Dawley rats, 250 g in weight, were used in all tests. Three experimental groups, $N = 4$, were designed as follows: 1, inflammation; 2, inflammation *plus* treatment with an anti-inflammatory drug; and 3, administration of the anti-inflammatory drug alone. Acute inflammation was induced by intramuscular injection of 0.5 mL turpentine / kg body weight. Indomethacine, selected as model anti-inflammatory drug, was given by gavage, twice a day (10 a.m. and 5 p.m.), at dosages of 3.5 mg/kg; in group 2, the rats received their first medication at the same time as the turpentine shot. Blood was drawn from the caudal vein under diethyl ether anesthesia, before the treatments and then every 24 h, for 4 days.

2.2 Protein electrophoresis and pattern quantitation

The main analytical procedures for sera were identical to those reported in the previous papers of this series [26–28]. 2-DE maps were obtained by IPG-DALT [29] on 5 µL aliquots of the samples. Isoelectric focusing was performed on a nonlinear pH 4–10 IPG [30], and SDS-PAGE was then run on 7.5–17.5%T polyacrylamide (PAA) gradients. Protein patterns were stained with 0.3% w/v Coomassie. Orosomucoid pattern was analyzed by running 7.5 µL serum per sample on a 2.5–5 IPG [31], and staining with Coomassie according to [32]. The gels were scanned with a video camera (Sony, Japan) under the control of NIH Image, release 1.52, and analyzed with the software PDQUEST Version 5.1 (PDI, Huntington Station, NY, USA) run on a Sun SPARCstation 4 (Sun Microsystems, Mountain View, CA). All results were evaluated and reported in terms of (absolute) spot volumes. Albumin concentration was evaluated by the immediate bromocresol green assay [33, 34].

2.3 Lipoprotein analysis

Ten microliter aliquots of the sera were run in zonal electrophoresis on 0.8% w/v agarose gels, with a pH 8.6 buffer [35]. The lipoprotein pattern was stained with Sudan Black [36]. Total lipoproteins were purified from sera of control and inflamed animals (2 days after turpentine shot) by flotation on a KBr solution, $d = 1.21$ g/mL, and delipidated by treatment with diethylether:ethanol 3:1 v/v [37]. Equal amounts of the purified apoproteins (50 µg) were run in 2-DE as above.

3 Results

Preliminary experiments helped select working conditions. They included the following parameters: (i) decreasing intensity of the stimulus, from 5 to 0.5 mL turpentine per kg body weight, and (ii) varying doses of the reference drug, from 1.4 to 7 mg indomethacine per kg body weight, given in the drinking water starting 24 h after the turpentine shot. A correlation was observed between the serum concentration of some protein bands (as resolved by SDS-PAGE) with either turpentine stimulus (α_1-inhibitor III (α_1-I$_3$) and, to a lower extent, α_2-macroglobulin (α_2-M)) or indomethacine intake (transferrrin (Tf) and albumin), demonstrating that over the specified range the response of the test system could be modulated by the

experimental parameters. Comparisons at days 1 and 3 of drug treatment showed an overall similar peaking of time-dependent changes with all protocols. The lowest intensity of the stimulus, corresponding to the smallest volume of turpentine that could be reproducibly injected with a 1 mL syringe, and the average dose of indomethacine, 3.5 mg/kg, typical for therapeutic use in humans, were then selected. Since our aim was to evaluate a possible application of the present protocol into routine screening, the duration of the test was limited to five working days. Ether anesthesia, given to the animals to reduce distress from the treatments, was checked so as not to alter the pattern of serum proteins: band densities as evaluated after SDS-PAGE varied < 10% among the samples from following days.

Figure 1 shows the changes in serum along the progression of inflammation, from day 0 (top) to day 4 (bottom). The panels are cropped from 2-DE with SDS-PAGE run on 4.5–9%T PAA gradients, to encompass M_r ranging between *ca.* 30 000 and 70 000. Figure 2 compares the whole protein patterns from treated animals on day 1 (groups 1–3, from top to bottom). Figure 3 displays concentration *versus* time of (some of) the most abundant serum proteins. The top row plots data for the time-course of inflammation, *i.e.*, the reaction to the intramuscular injection of turpentine; the middle row, the response to the concurrent administration of indomethacine; the bottom row, the effects of indomethacine alone. The sequence of the panels is according to alphabetical order of the abbreviations for the protein names. For thiostatin (α_1-MAP) two curves are plotted, one for each of the spot rows in the 2-DE pattern (α_1-MAP (1) and α_1-MAP (2)). In the inflamed animals, the timing of the peak effect on serum concentration varies from one protein to another. It is observed after 24 h for kallikrein-binding protein (KBP) and transferrin; between 36 and 48 h for α_1-inhibitor III, serine protease inhibitor 3 (SPI3), haptoglobin chains, orosomucoid, α_2-macroglobulin, α_2-HS-glycoprotein and α_1-antitrypsin; and between 60 and 72 h for C-reactive protein (CRP), albumin, hemopexin and thiostatin.

Indomethacine by itself is able to influence the synthesis rate of most of the proteins, with minor effects only on thiostatin (α_1-MAP (2)) and ceruloplasmin. In some instances the effects of indomethacine on protein concentration have the same direction (increase or decrease) as the inflammatory stimulus, but the extent is lower and the peak is delayed. This, for instance, is the case with CRP, haptoglobulin and transferrin. For α_2-HS-glycoprotein, α_2-macroglobulin, and orosomucoid, the effect neither plateaus, nor reverts to baseline at the end of the experimental period, hence an even larger peak effect of indomethacine *versus* turpentine may be expected over an extended observation period. With KBP, the overall changes with time are symmetrical with either turpentine alone or indomethacine alone, with a negative and a positive peak, respectively, 24 h from the onset of treatment.

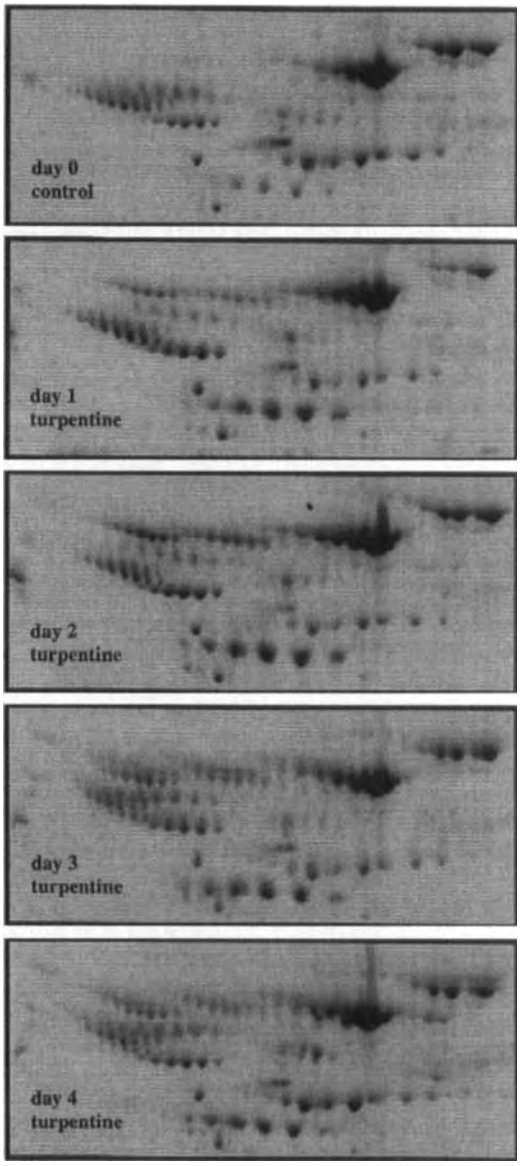


Figure 1. 2-DE maps of serum proteins from animals of experimental group 1, corresponding to acute inflammation after intramuscular injection of 0.5 mL turpentine / kg body weight, from day 0 (top panel) to day 4 (bottom panel). Second dimension run on 4.5–9%T PAA gradients. Coomassie stain on 5 µL samples. For all panels only the region between *ca.* 30 and 70 kDa is shown.

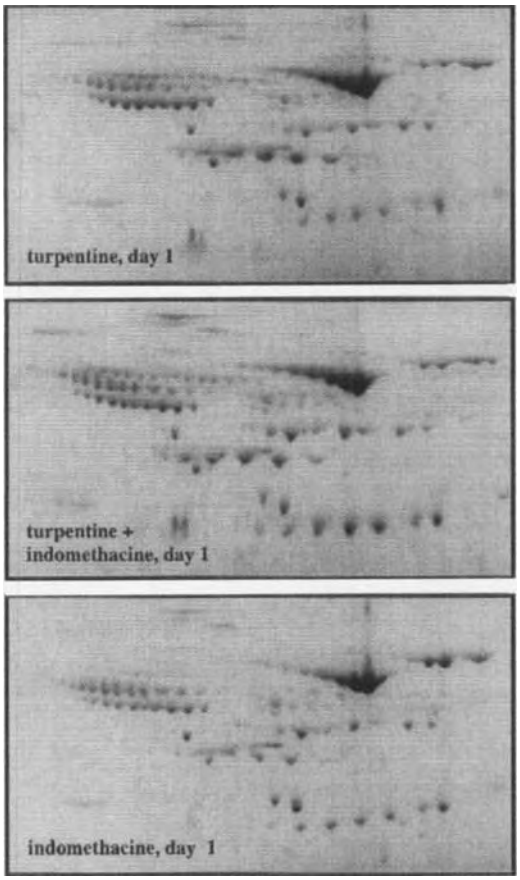


Figure 2. 2-DE map of serum proteins from animals of the three experimental groups, on day 1 of treatment: group 1, inflammation after intramuscular injection of 0.5 mL turpentine / kg body weight, in top panel; group 2, inflammation treated with indomethacine, in middle panel; treatment with indomethacine alone, in bottom panel. Second dimension run on 7.5–17.5%T PAA gradients. Coomassie stain on 5 μL samples.

In the inflamed animals receiving indomethacine, the observed changes in protein concentration are not always equivalent to the algebraic sum of the individual treatments. This is indeed the case with α_2-macroglobulin, orosomucoid, and thiostatin: the early peak, associated with inflammation overlaid to the delayed rise caused by indomethacine, results in a slow but sustained increase of the protein concentrations. This is also found for haptoglobin chains and SPI3: peaks with different onset and, for the former, of different height, give rise to broader plateauing when both treatments are given together. As an extreme case, opposite trends for the shifts in KBP concentration brought about by turpentine or indomethacine alone result in a sigmoidal pattern, the timing of whose bends is intermediate between the peaks (positive and negative) for the individual treatments. At 24 h a difference in concentration of $\geq$ 50% between inflammation and inflammation plus drug is observed for α_2-macroglobulin, α_2-HS-glycoprotein, CRP and KBP (also refer to Fig. 2).

A peculiar pattern is displayed by the lipoproteins of the test animals: as shown by Fig. 4, the mobility of the various classes changes with time and with treatment. In the inflamed animals, decreased migration is observed on day 1, reverting to control mobility on day 3. With inflammation plus drug treatment, the difference in migration is reduced but not abolished, even on day 4. With indomethacine alone, the effects on protein migration become apparent from day 3 on. Analysis of the apolipoprotein (apo) moieties from total lipoporoteins purified from sera of inflamed animals in comparison with control rats does not show any change in the relative abundance of the various charge isoforms for the different apoproteins (not shown).

4 Discussion

4.1 Comparison with literature data

The time-course of the expression of acute-phase reactants (APR) in the rat was reviewed by Schreiber *et al.* [38], detailing both the transcription of mRNAs and the translation into serum proteins. A relevant difference in our experimental setup is the choice of a stimulus (volume of injected turpentine per kg body weight) reduced to one tenth of the standard value. A further difference is the parallel evaluation of the changes brought about by a reference anti-inflammatory drug such as indomethacine. This permits discussing the serum pattern of the inflamed rats given the drug with relation to both the phlogistic stimulus and the antiphlogistic treatment. Moreover, with 2-DE it is possible to monitor, at the same time, a large number of spots, among which we selected data for 16 gene products *versus* the seven proteins discussed by Schreiber *et al.* [38]. With either standard (5 mL turpentine/kg [38]) or reduced stimulus (0.5 mL/kg in this report) the peak effects are obtained almost at the same time for most proteins. For example, the overall time-course for the concentration changes is similar for albumin and orosomucoid. Contrary to our findings, and to the results of Lonberg-Holm *et al.* [39] on α_2-M, Schreiber *et al.* [38] report an earlier rise for α_1-antitrypsin and a faster return to baseline for α_2-M, while transferrin is found in increased concentration from day 3 on.

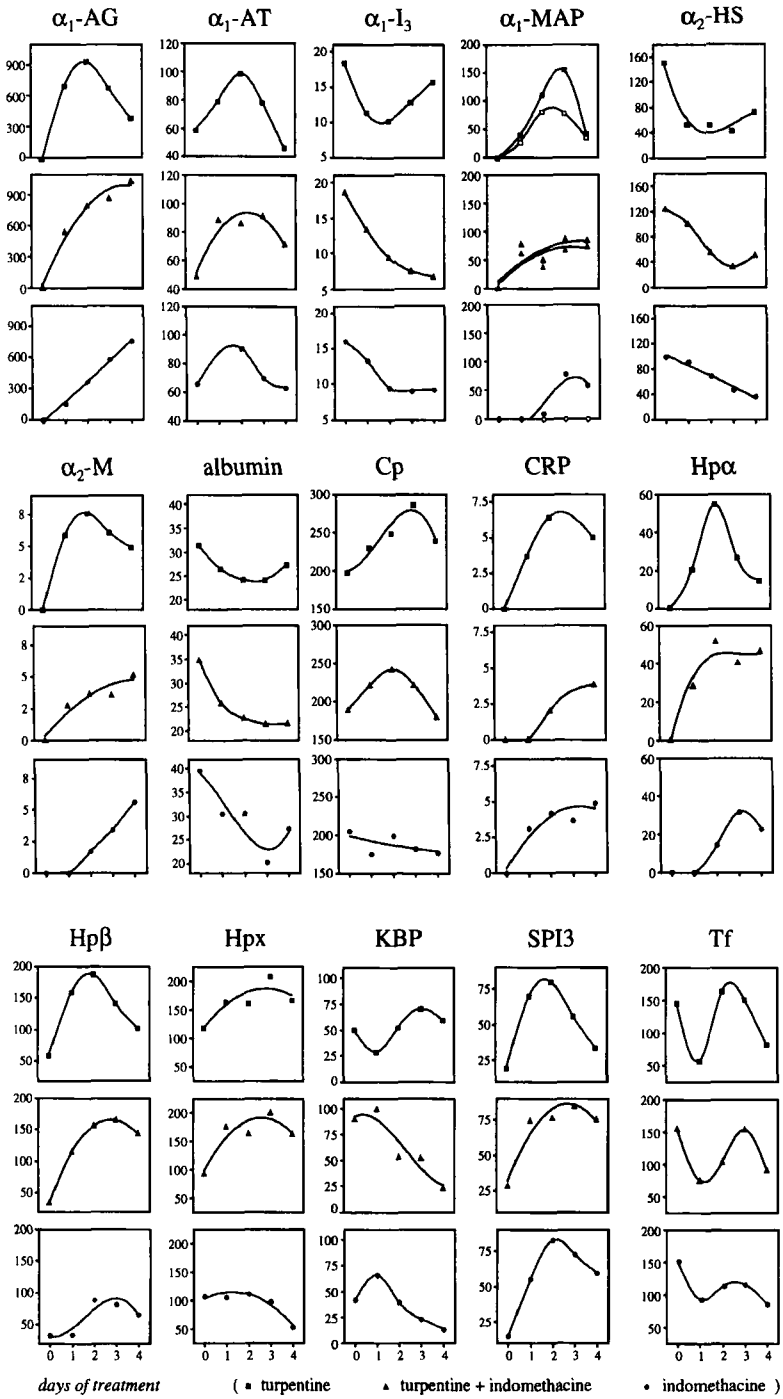


Figure 3. Time-course of the changes in concentration for the major proteins in rat serum, after intramuscular injection of 0.5 mL turpentine / kg body weight (top panel in each series ■), after injection of turpentine plus the administration of daily doses of indomethacine (middle panel in each series ▲) or after treatment with indomethacine alone (bottom panel in each series ●). α_1-AG, α_1-acid glycoprotein; α_1-AT, α_1-antitrypsin; α_1-I_3, α_1-inhibitor III; α_1-MAP, thiostatin (1, solid symbols; 2, open symbols); α_2-HS, α_2-HS-glycoprotein; α_2-M, α_2-macroglobulin; Cp, ceruloplasmin; CRP, C-reactive protein; Hp, haptoglobin; Hpx, hemopexin; KBP, kallikrein-binding protein; SPI3, serine protease inhibitor-3; Tf, transferrin. Results are plotted as spot volumes from the densitometric evaluation of Coomassie-stained 2-DE gels, except for albumin, whose concentration, mg/mL, was assayed through a dye-binding test in solution.

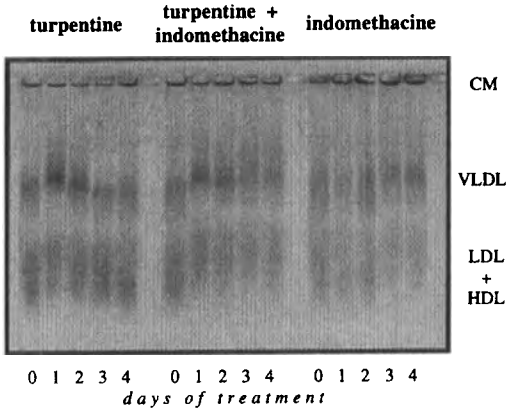


Figure 4. Agarose electrophoresis and lipoprotein stain with Sudan Black of the sera from the time-course experiment above. From left to right: animals only receiving a turpentine shot, inflamed animals treated with indomethacine, animals given indomethacine alone. CM, chylomicrons; VLDL, very low density lipoprotein; LDL, low density lipoprotein; HDL, high density lipoprotein.

When comparing the figures of day 2 in the present report with our own data after standard phlogistic stimulus at the single time point of 48 h [28], agreement is observed for almost all findings. Day 2 either corresponds to the peak effect brought about by inflammation on the APR concentration, or is at least within the span over which the typical shift, whether an up- or a down-regulation, is observed. This is not the case for transferrin, which shows no difference from baseline on day 2 whereas its concentration decreases on day 1 and towards the end of the test period. Hence, our observation of a nonchanging concentration upon inflammation [28] is only confirmed for the specified time point of 48 h. A decrease in concentration on day 1 had also been reported by Schreiber *et al.* [38]. The only major difference between our observations after either 5 or 0.5 mL/kg of turpentine involves KBP. The higher dose results in a significant decrease of this protein's concentration (still on day 2) whereas with the low dose the drop in concentration is limited to day 1.

4.2 Effects of nonsteroidal anti-inflammatory drugs (NSAIDs)

It has been shown that some NSAIDs, while treating the inflammatory pathology, paradoxically raise the levels of circulating pro-inflammatory cytokines such as interleukin-6 [40]. Most NSAIDs inhibit, to a different extent, both isozymes of cyclooxigenase (COX-1 and -2) [41, 42], the inhibition of COX-1 likely being responsible for some of the side effects [43]. Indomethacine reproduces some but certainly not all the effects of the inflammatory stimulus.

As examples of different behavior, one can contrast α_1-antitrypsin and transferrin (whose concentrations change with almost identical pattern after either turpentine or indomethacine), with hemopexin and thiostatin (two APRs whose synthesis is unaffected by indomethacine) and with KBP (on whose expression inflammation and NSAID treatment have opposite effects). Also the extent of the up- or down-regulation associated with phlogosis and antiphlogosis and the timing of the maximal effect usually differ. Delay is observed in our experimental group 3 in comparison with group 1 for the decrease in concentration of α_2-HS-glycoprotein and for the increase of α_2-macroglobulin; delay and lower peak concentration is found for haptoglobin chains. The rise in cytokine levels obviously does not have a comparable effect on the promoters of all APRs [44]. An extreme case is thiostatin, for which two genes exist as a result of a duplication event [45, 46]. While the expression of the more acidic peptide, α_1-MAP (1) (labeled as 1 in Fig. 1 of [28]), is increased by indomethacine, the level of α_1-MAP (2) (labeled as 2) is unaffected by the drug treatment. Furthermore, on days 2 and 3 splitting of the curve for α_1-MAP (1) is observed.

In spite of the complexity of the response to NSAID treatment, some clear-cut differences between the serum pattern of group 2 *versus* group 1 rats may be observed. For α_2-macroglobulin and CRP the increase, and for α_2-HS-glycoprotein the decrease, in concentration is at least 50% lower in the former than in the latter; for KBP again a difference in concentration of $\geq$ 50% between the two experimental groups results from changes in opposite directions in comparison with baseline levels. All of the above are observed 24 h after the start of treatment; as a result of the specific time-courses for the expression of the various proteins, day 1 is indeed the most appropriate time to evaluate the effects of indomethacine treatment of APRs (Fig. 2). The finding of maximum discrimination at such an early time would make the detailed protocol very convenient if it could be applied to a screening program.

4.3 Diagnostic relevance of APR levels

For the diagnosis of inflammatory diseases in which pain is the dominant symptom, as well as for the follow-up of their therapy, easily quantified clinical parameters are needed. In the clinic, the markers in general use are CRP and orosomucoid serum levels and erythrocyte sedimentation rate (ESR), but no consensus exists as for their absolute and relative importance [47]. In agreement with Tishler *et al.* [48], who suggest that in patients treated with NSAIDs no relationship holds between CRP and ESR, the increase of CRP and orosomucoid levels in our control rats after indomethacine suggests caution when

evaluating clinical markers of rheumatic disease during the administration of NSAIDs.

4.4 Lipoproteins in inflammation

In the three experimental groups in this investigation, all classes of lipoproteins display a most unusual change in mobility, with a decrease in anodic migration, that is modulated by time and by treatment. The microheterogeneity of the purified apolipoproteins was found to be similar for samples from control and from inflamed animal sera (on day 2 from turpentine shot; not shown). Even with high sample loads, only apolipoproteins could be detected, in contrast to the hypothesis of the possible binding of further serum components, as had been reported for CRP in rabbits [49]. Furthermore, the *in vitro* binding of CRP to rat low density lipoprotein (LDL) results in an increased, instead of decreased, electrophoretic mobility [50]. It thus seems likely that the changes in lipoprotein mobility result from alterations in the composition of the lipid portion, as phosphocholine-containing phospholipids, which bear a positive charge, significantly increase after inflammation [49].

4.5 Variance

Comparison of data from time-courses are best achieved when the same animals can be used throughout the entire study, possibly by drawing blood on different days, starting with baseline conditions. Changes as a result of experimental treatment can thus be easily distinguished from variations in the normal ranges of single proteins [28]. Our own Sprague-Dawley rats showed a dispersion of data for baseline concentrations in different pools (compare the starting values at day 0 for individual proteins in Fig. 3), resulting in an average coefficient of variation for all proteins of 20%. Ikawa and Shozen [51] have determined differences of 10–25% in concentration levels of some proteins in healthy individuals of an (unspecified) rat strain.

4.6 Conclusions

While the alterations in the serum protein makeup after acute or chronic inflammation have also been investigated in mice [52] and in humans [53, 54], we feel our rat model may be a useful tool for the primary screening of molecules with potential antiphlogistic activity. After the effects of a model drug such as indomethacine, the target serum proteins for such a screening appear to be α_2-macroglobulin, α_2-HS-glycoprotein, C-reactive protein and kallikrein-binding protein. Owing to the specific effects of the drug itself on most serum components, sampling for differential evaluation, NSAID therapy *versus* inflammation, should be performed 24 h from the onset of treatment. Although technically demanding, the detailed protocol appears cost-effective for the large-scale screening of new drugs, since information is provided at once on a large number of proteins, and unusual/unexpected modifications will not escape detection. One of the concerns about the use of animal models is the cost of their care. The possibility to use the same animals as a control as well as a test group allows a limited number of rats to be treated even in the presence of the relatively large baseline variability typical of an outbred population. The short treatment time required is also instrumental in cutting the costs involved in the experimental protocol. Ongoing research already involves the effects of NSAIDs on adjuvant arthritis (to be published with Laboratorio di Neuroimmunologia, Istituto di Ricerche Farmacologiche "Mario Negri",Milano).

The authors wish to thank Prof. G. Folco, Drs. Pia Villa, Maddalena Fratelli, Davide Agnello, Giulia Chiesa, Laura Calabresi and Livia Tonti for fruitful discussion, and Dr. Paola Bezzi for providing the first lot of test animals. E.G. is grateful to Prof. Rodolofo Paoletti for his encouragement.

Received October 31, 1998

5 References

[1] Rubin, E., Farber, J. L., *Essential Pathology*, Lippincott-Raven, Philadelphia 1995.

[2] Engelhardt, G., Homma, D., Schlegel, K., Utzmann, R., Schnitzler, C., *Inflamm. Res.* 1995, *44*, 423–433.

[3] Carvalho, J. C., Silva, M. F., Marciel, M. A., Pinto, A. C., Nunes, D. S., Lima, R. M., Bastos, J. K., Sarti, S. J., *Planta Medica* 1996, *62*, 402–404.

[4] Ionac, M., Parnham, M. J., Plauchithiu, M., Brune, K.,. *Pharmacol. Res.* 1996, *33*, 367–373.

[5] Ozaki, Y., Xing, L., Satake, M., *Biol. Pharm. Bull.* 1996, *19*, 1046–1048.

[6] Elliott, H. G., Elliott, M. A., Watson, J.,. Steele, L., Smith, K. D., *Biomed. Chromatogr.* 1995, *9*, 199–204.

[7] Zhang, S., Howarth, P. H., Roche, W. R., *J. Pathol.* 1996, *180*, 95–101.

[8] Menzel, J. E., Kolarz, G., *Inflammation* 1997, *21*, 451–461.

[9] Viganò, T., Habib, A., Hernandez, A., Bonazzi, A., Boraschi, D., Lebret, M., Cassina, E., Maclouf, J., Sala, A., Folco, G., *Am. J. Resp. Crit. Care Med.* 1997, *155*, 864–868.

[10] Anderson, N. G., Anderson, N. L., *Clin. Chem.* 1982, *28*, 739–748.

[11] Anderson, N. L., Giere, F. A., Nance, S. L., Gemmell, M. A., Tollaksen, S. L., Anderson, N. G., *Fund. Appl. Toxicol.* 1987, *8*, 39–50.

[12] Anderson, N. L., Copple, D. C., Bendele, R. A., Probst, G. S., Richardson, F. C., *Fund. Appl. Toxicol.* 1992, *18*, 570–580.

[13] Richardson, F. C., Strom, S. C., Copple, D. M., Bendele, R. A., Probst, G. S., Anderson, N. L., *Electrophoresis* 1993, *14*, 157–161.

[14] Richardson, F. C., Hom, D. M., Anderson, N. L., *Carcinogenesis* 1994, *15*, 325–329.

[15] Cunningham, M. L., Pippin, L. L., Anderson, N. L., Wenk, M. L., *Toxicol. Appl. Pharmacol.* 1995, *131*, 216–223.

[16] Myers, T. G., Dietz, E. C., Anderson, N. L., Khairallah, E. A., Cohen, S. D., Nelson, S. D., *Chem. Res. Toxicol.* 1995, *8*, 403–413.

[17] Steiner, S., Wahl, D., Varela, M. C., Aicher, L., Prieto, P., *Electrophoresis* 1995, *16*, 1969–1976.

[18] Anderson, N. L., Taylor, J., Hofman, J. P., Esquer-Blasco, R., Swift, S., Anderson, N. G., *Toxicol. Pathol.* 1996, *24*, 72–76.

[19] Anderson, N. L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharmacol.* 1996, *137*, 75–89.

[20] Steiner, S., Wahl, D., Margold, B. L., Robison, R., Raymackers, J., Meheus, L., Anderson, N. L., Cordier, A., *Biochem. Biophys. Res. Commun.* 1996, *218*, 777–782.

[21] Myers, T. G., Anderson, N. L., Waltham, M.,. Buolamwini, J. K., Scudiero, D. A., Paull, K. D., Sausville, E. A., Weinstein, J. N., *Electrophoresis* 1997, *18*, 647–653.

[22] Marshall, T., Vesterberg, O., *Electrophoresis* 1983, *4*, 363–366.

[23] Zastrow, G., Günther, S., Postel, W., Görg, A., Diehl, H. A., Jansen, E. H. J. M., *Electrophoresis* 1990, *11*, 655–657.

[24] Rothkopf-Ischebeck, M., *Laboratory Animals* 1980, *14*, 153–165.

[25] Guglielmotti, A., Silvestrini, B., Saso, L., Zwain, I., Chen, C. Y., *Biochem. Mol. Biol. Int.* 1993, *29*, 747–756.

[26] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, M. R., Manzoni, C., Vignati, M., Gianazza, E., *Electrophoresis* 1998, *19*, 1484–1492.

[27] Miller, I., Haynes, P., Gemeiner, M., Aebersold, R., Manzoni, C., Lovati, M. R., Vignati, M., Eberini, I., Gianazza, E., *Electrophoresis* 1998, *19*, 1493–1500.

[28] Miller, I., Haynes, P., Eberini, I., Gemeiner, M., Aebersold, R., Gianazza, E., *Electrophoresis* 1999, *20*, 836–845.

[29] Gianazza, E., in: Link A. J. (Ed.) *2-D Proteome Analysis Protocols*, Humana Press, Totowa 1998.

[30] Gianazza, E., Giacon, P., Sahlin, B., Righetti, P. G., *Electrophoresis* 1985, *6*, 53–56.

[31] Gianazza, E., Celentano, F., Ettori, C., Righetti, P. G., *Electrophoresis* 1989, *10*, 806–808.

[32] Righetti, P. G., Drysdale, J. W., *J. Chromatogr.* 1974, *98*, 271–321.

[33] Rodkey, F. L., *Clin. Chem.* 1965, *11*, 478–487.

[34] Doumas, B. T., Watson, W. A., Biggs, H. G., *Clin. Chim. Acta* 1971, *31*, 87–96.

[35] Monthony, J. F., Wallach, E. G., Allen, D. M., *Clin. Chem.* 1978, *24*, 1825–1827.

[36] Sargent, J. R., George, S. G., *Methods in Zone Electrophoresis*, BDH Chemicals Ltd, Poole 1975.

[37] Schumaker, V. N., Puppione, D. L., *Methods Enzymol.* 1986, *128*, 155–170.

[38] Schreiber, G., Tsykin, A., Aldred, A. R., Thomas, T., Fung, W.-P., Dickson, P. W., Cole, T., Birch, H., De Jong, F. A., Milland, J., *Ann. N. Y. Acad. Sci.* 1989, *557*, 61–85.

[39] Lonberg-Holm, K., Reed, D. L., Roberts, R. C., Hebert, R. R., Hillman, M. C., Kutney, R. M., *J. Biol. Chem.* 1987, *262*, 438–445.

[40] Sipe, J. D., Bartle, L. M., Loose, L. D., *J. Immunol.* 1992, *148*, 480–484.

[41] Smith, W. L., Meade, E. A., DeWitt, D. L., *Ann. N. Y. Acad. Sci.* 1994, *714*, 136–142.

[42] Smith, W. L., Garavito, R. M., DeWitt, D. L., *J. Biol. Chem.* 1996, *271*, 33157–33160.

[43] Vane, J. R., *Lancet* 1995, *346*, 1105–1106.

[44] Arai, K.-I., Lee, F., Miyajima, A., Miyatake, S., Arai, N., Yokota, T., *Annu. Rev. Biochem.* 1990, *59*, 783–836.

[45] Kageyama, R., Kitamura, N., Ohkubo, H., Nakanishi, S., *J. Biol. Chem.* 1985, *260*, 12060–12064.

[46] Kitagawa, H., Kitamura, N., Hayashida, H., Miyata, T., Nakanishi, S., *J. Biol. Chem.* 1987, *262*, 2190–2198.

[47] Wollheim, F. A., Eberhardt, K. B., *Baillieres Clin. Rheumatol.* 1992, *6*, 69–93.

[48] Tishler, M., Caspi, D., Yaron, M., *Clin. Rheumatol.* 1985, *4*, 321–324.

[49] Cabana, V. G., Gewurz, H., Siegel, J. N., *J. Immunol.* 1983, *130*, 1736–1742.

[50] Mookerjea, S., Francis, J., Hunt, D., Yang, C. Y., Nagpurkar, A., *Arterioscler. Thromb.* 1994, *14*, 282–287.

[51] Ikawa, M., Shozen, Y., *J. Immunol. Methods* 1990, *143*, 101–106.

[52] Pluschke, G., Jenni, L., van Alphen, L., Lefkovits, I., *Clin. Exp. Immunol.* 1986, *66*, 331–339.

[53] Fritz, P., Arold, N., Mischlinski, A., Wisser, H., Oeffinger, B., Neuhoff, V., Laschner, W., Koening, G., *Rheumatol. Int.* 1990, *10*, 177–183.

[54] Doherty, N. S., Littman, B. H., Reilly, K., Swindell, A. C., Buss, J. M., Anderson, N. L., *Electrophoresis* 1998, *19*, 355–363.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 854–859

Bridget A. Lollo
Sheryl Harvey
Jane Liao
Anthony C. Stevens
Raymond Wagenknecht
Rick Sayen
Justine Whaley
Fereydoun G. Sajjadi

Proteomix, Inc.,
San Diego, CA, USA

Improved two-dimensional gel electrophoresis representation of serum proteins by using ProtoClear™

ProtoClear™ is a proprietary technique for clearing albumin and immunoglobulin G (IgG) from human serum samples. Albumin constitutes 57–71% of total serum protein and IgG ranges from 8–26%. Removal of these two proteins alone clears ~75% of the total protein present in serum and allows the detection of the remaining proteins that are present in far lower concentrations. ProtoClear effectively removed > 95% of human serum albumin (HSA) and > 97% of human IgG as measured by an anti-HSA competitive immunoassay and a radial immunodiffusion assay, respectively. ProtoClear was far more specific at removing albumin and IgG than Cibracon Blue Dye chromatography (Cibracon Blue), the typically utilized alternative. Comparing two-dimensional (2-D) gels of serum cleared by either Cibracon Blue or by ProtoClear, it was apparent that Cibracon Blue removed a number of proteins in addition to albumin. Following removal of albumin and IgG from serum, we found a significant improvement in the resolution of polypeptide spots detected on two-dimensional gels.

Keywords: Serum / Two-dimensional gel electrophoresis / Albumin / Immunoglobulin G / Proteomics
EL 3393

1 Introduction

Serum is known to be difficult to resolve by 2-D gel electrophoresis (2-DE), largely due to the abundance of serum albumin and immunoglobulin G (IgG). Albumin constitutes 51–71% of the total protein present in human serum and immunoglobulin G constitutes 8–26% [1]. The presence of these two proteins obscures other proteins that migrate to the surrounding area and limits the amount of serum that can be resolved on a 2-D gel. Theoretically, by removing albumin and immunoglobulin G, that together make up 59–97% of the protein present in serum, 3–7 times more serum can be analyzed.

The current method available for removal of albumin from serum is highly nonspecific. Cibracon Blue chromatography is commonly used to bind albumin; however, this material is known to bind numerous other proteins [2]. Cibracon Blue chromatography is based on the Cibracon Blue 3G dye immobilized onto a matrix. Cibracon Blue mimics nicotinamide adenine dinucleotide and purine dinucleotides and, therefore, it binds most proteins that contain a dinucleotide fold through a specific affinity interaction [3]. Other proteins bind to its planar ring structure and negatively charged sulfate groups through a complex combination of electrostatic, hydrophobic, and hydrogen bonding interactions. It is known that albumin binds to this matrix with a very high affinity and the binding of various other plasma proteins such as interferon, lipoproteins, haemopexin, antithrombin III, and blood coagulation factors have been reported [2, 4]. Therefore, although Cibracon Blue binds albumin with a relatively high affinity, the fact that it also binds many other proteins makes it undesirable for studying serum proteins by 2-D gel electrophoresis. Furthermore, bilirubin and fatty acids complexed to albumin interfere with the binding of albumin to Cibracon Blue [4]. The amount of bilirubin and fatty acids complexed to albumin can vary from person to person and therefore the extent of removal of albumin by Cibracon Blue would be expected to be variable.

ProtoClear™ is a proprietary polypeptide affinity matrix designed to remove albumin and immunoglobulin G from human serum with minimal nonspecific protein removal. Data presented here show that ProtoClear is highly effective at removing albumin and immunoglobulin G from human serum. We compare ProtoClear to the Cibracon Blue-based method commercially available in the Albumin Removal Kit from ESA, and demonstrate that ProtoClear resulted in much greater recovery of serum proteins other than albumin. Finally we demonstrate the improvement that ProtoClear makes in 2-D gel images of serum.

Correspondence: Bridget A. Lollo, 2132 Pleasant Grove Road, Encinitas, CA 92024, USA
E-mail: blollo@tns.net

Abbreviation: IgG, immunoglobulin G

2 Materials and methods

2.1 ProtoClear

ProtoClear™ technology is a proprietary technique [5] for clearing interfering proteins from a liquid sample. Proto-

Clear is a "collapsible affinity matrix" that is specific for a particular protein or proteins of interest. It consists of a polypeptide that has an affinity for the protein(s) of interest attached to a collapsible matrix. The collapsible matrix is formed by combining one member of a high affinity binding pair with another member of a high affinity binding pair system. The resulting complex is rigid enough to be centrifuged and washed, resulting in a matrix with minimal interstitial space and with specificity towards the protein of interest. The version of ProtoClear™ described here (and referred to in the text simply as ProtoClear) is designed to remove human serum albumin (HSA) and immunoglobulin G (IgG) from serum. When the serum sample is incubated with the matrix, centrifugation results in a pelleted matrix containing HSA and IgG and a supernatant greatly depleted in HSA and IgG. The pellet was discarded and the supernatant material was the ProtoClear-treated serum. The volume of serum treated ranged between 10–100 μL per tube. Complete, Mini, EDTA-free Protease Inhibitor Cocktail (Boehringer Mannheim, Mannheim, Germany) was added to prevent proteolysis.

2.2 HSA and IgG quantification

HSA removal was quantified by using an immunoassay specific for human serum albumin, developed in-house. Human IgG was quantified by using a Human IgG ML Radial Immunodiffusion Kit (The Binding Site).

2.3 Cibracon Blue removal of HSA

For albumin removal from serum using the Albumin Removal Kit (ESA, St. Ives, UK), directions from the kit were followed except that in-house formulations were used for sample dilution into immobilized pH gradient (IPG) rehydration buffer.

2.4 Sample preparation and isoelectric focusing

Serum treated with either ProtoClear, the ESA Albumin Removal Kit, or not treated, was diluted into IPG rehydration buffer containing 8 M urea, 4% CHAPS, 10 mM DTT, 0.4% Pharmalytes 3–10 (Pharmacia, Uppsala, Sweden), 0.1% sodium taurodeoxycholate, and bromophenol blue. Three hundred and sixty μL of sample in rehydration buffer was absorbed into an Immobiline Dry Strip (Pharmacia Biotech; Piscataway, NJ; pH range indicated in figure legends) overnight and the strips focused on a MultiPhor II (Pharmacia Biotech) for 100 kVh at 15°C. The focused strips were reduced, alkylated, and equilibrated in SDS by the method of [6]. Briefly, reduction was with 0.8% DTT in equilibration buffer (30% glycerol, 2.5% SDS, 6 M urea, 0.15 M Bis-Tris and 0.1 M HCl) and alkylation was with 4% iodoacetamide in equilibration buffer. Strips were stored at –70°C until later use.

2.5 Second dimension

The equilibrated IPG strips were placed onto a resolving acrylamide gel (percentages are noted in figure legends) and overlaid with 1% low melt agarose (FMC Bioproducts, Rockland, ME) in 0.2% SDS, 0.15 M Bis-Tris, 0.1 M HCl, bromophenol blue at 60°C [6]. Electrophoresis was in 0.1 M Tris base, 0.1 M Tricine, 0.1% SDS buffer overnight at 22°C. Gels were silver stained by the method of Rabilloud [7] and dried in a BioRad Gel Dryer (Richmond, CA).

2.6 Image analysis

Gels were digitized using a Sharp JX-330 Scanner. Scanned images were analyzed with the ImageMaster Software (Pharmacia Biotech). After ImageMaster identification of polypeptide spots, the operator made any necessary modifications. Gels were compared to each other by the operator first seeding all obvious matches and then allowing ImageMaster to illuminate spots unique to each gel. Molecular weights and isoelectric points were estimated using known serum proteins as internal standards.

3 Results

The degree of removal of albumin and IgG from serum by ProtoClear was analyzed by quantitative methods. An immunoassay, developed to measure HSA, was used to compare the amount of HSA before and after treatment by ProtoClear. The evaluation of five different lots of ProtoClear demonstrated that 98 ± 1% of the HSA was removed, with the range of removal being 96–99%. A radial immunodiffusion assay was used to quantitate the amount of IgG in serum before and after ProtoClear treatment. The percent removal of IgG from four different human sera samples was 97.6 ± 0.2%. Greater than 97% of the IgG was removed irrespective of the initial IgG concentration, which varied more than twofold (initial IgG range = 8.1–21.1 mg/mL).

We compared the effectiveness of ProtoClear with Cibracon Blue chromatography, the currently available alternative for albumin removal. We evaluated the Albumin Removal Kit from ESA, which is based on a Cibracon Blue resin spin column to remove albumin from small quantities of serum. As shown in Table 1, the two techniques removed nearly equivalent amounts of albumin; however, the total protein recovery after accounting for albumin removal was nearly twofold greater in the serum treated

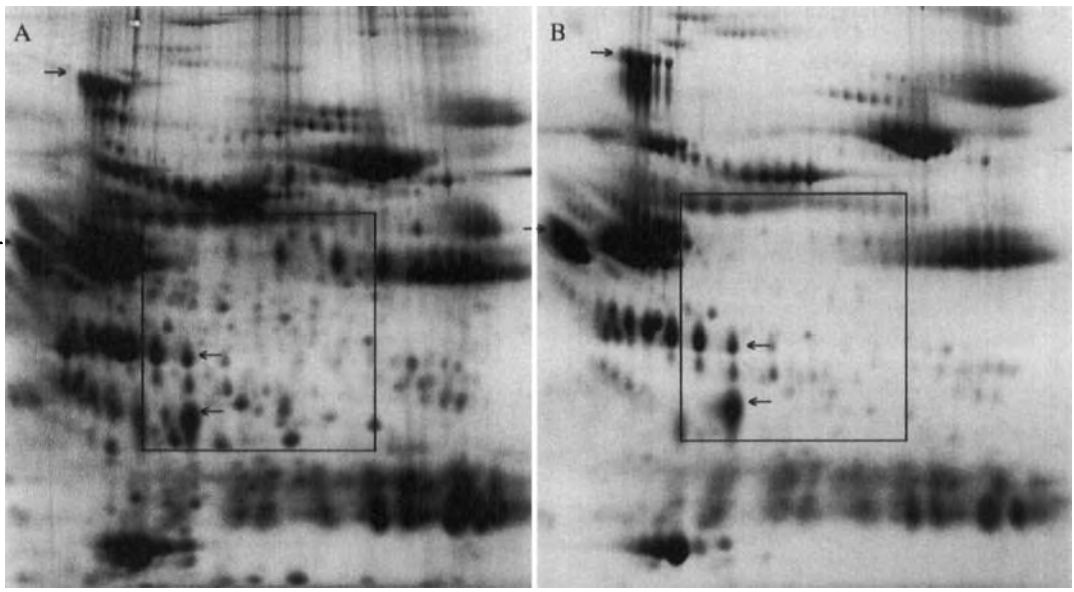


Figure 1. 2-D gel comparison of two albumin removal techniques. Equivalent volumes of serum (25 μL) cleared of albumin by either (A) ProtoClear or (B) the ESA Albumin Removal Kit were loaded onto pH 3–10 nonlinear IPG strips, focused and run on 10% SDS-PAGE gels. Boxed regions are analyzed and presented in Fig. 2. Arrows point to proteins that exhibit equivalent protein load on both gels.

with ProtoClear than in the serum treated with the ESA Albumin Removal Kit. In Fig. 1, equivalent volumes of serum cleared of albumin by either ProtoClear or the ESA Albumin Removal Kit are resolved by 2-D gel electrophoresis and compared. It is apparent that there are more polypeptides in serum treated with ProtoClear (Fig. 1A) than in serum treated with the Albumin Removal Kit (Fig. 1B). Arrows point to proteins that exhibit equivalent protein load on both gels to substantiate that an equivalent volume of sera was loaded onto each gel. The two uppermost proteins are α_1-antitrypsin dimer and α_1-acid glycoprotein, both of which are reported to flow through a Cibracon Blue matrix under low salt conditions [2]. Boxed regions in Fig. 1A and 1B are enlarged in Fig. 2A and 2B, respectively. When the spots in these equivalent regions were quantified, 97 spots were counted in the gel of serum treated with ProtoClear and 41 spots were counted in the gel of serum treated with the ESA kit. The total volume of the spots in Fig. 2A was 9047 and in Fig. 2B, 4699. These results correlate well with the twofold greater amount of protein recovered in the ProtoClear sample when compared to the sample treated with the ESA kit, as measured by bicinchoninic acid (BCA) protein assay (Table 1). This suggested that ProtoClear treatment resulted in greater nonalbumin protein recovery than the Cibracon Blue-based technique.

Figures 2C and D represent overlays of the two boxed regions in Figs. 2A and B. The open circles in Figs. 2C and D are spots that were found in both Figs. 2A and B. These spots were used as reference spots to match the two gels. The solid circles in Fig. 2C are polypeptide spots unique to the gel of serum treated with ProtoClear and likely represent polypeptides that are removed by the ESA kit (56 spots). Similarly, the solid circles in Fig. 2D are polypeptides that are unique to serum cleared by the ESA kit and presumably represent polypeptides that are nonspecifically removed by the ProtoClear treatment (9 spots). Upon closer examination of the actual gels, rather than the scanned images, it is apparent that most of the spots missing from the ProtoClear gel are actually present but are hidden in the background of more prominent spots. The gel with ProtoClear-treated serum has many spots of much larger volume that obscure the relatively tiny spots. In the gel of Cibracon Blue-treated serum, which has less background as a result of fewer polypeptides, the very tiny spots stand out better. In fact, upon close observation only one spot is clearly absent in the serum treated with ProtoClear: the centermost spot in Fig. 2B. This is a distinct spot and is likely to be a polypeptide that nonspecifically adheres to the ProtoClear matrix. Since ProtoClear removed only one polypeptide nonspecifically, out of 106 total polypeptides, we feel that ProtoClear has

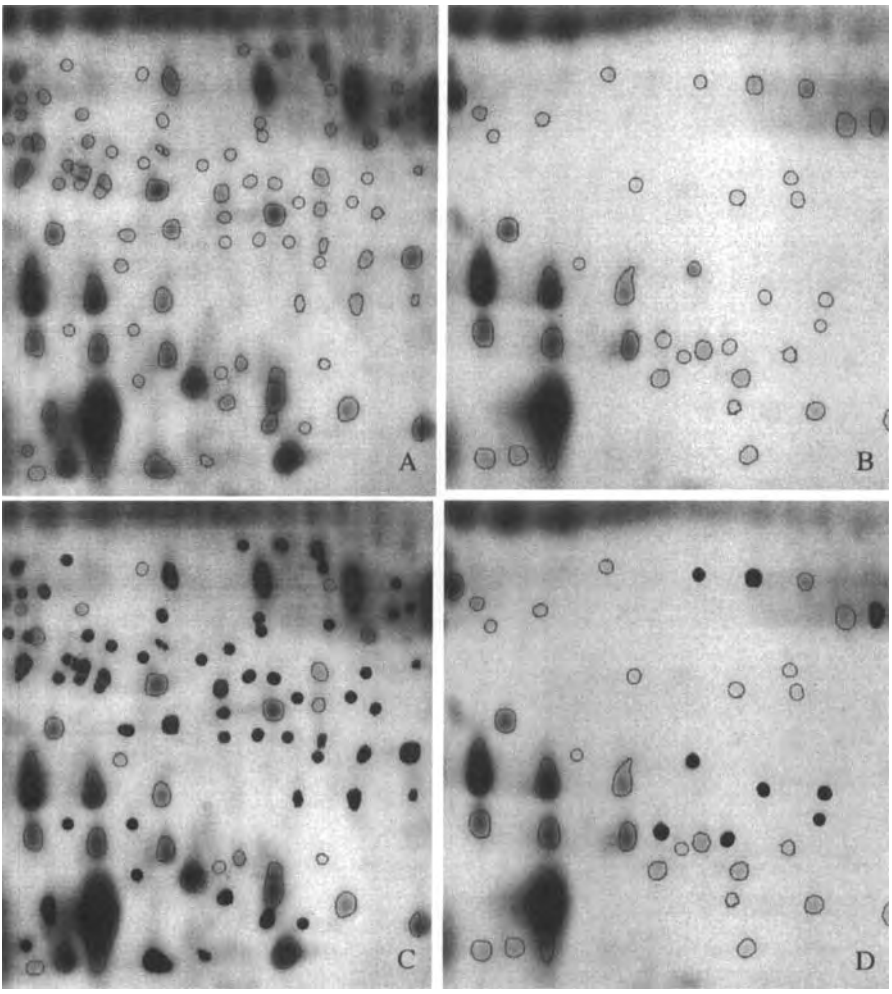


Figure 2. Image analysis of two albumin removal techniques. (A) and (C) are the boxed region shown in Fig. 1A of serum treated with ProtoClear; (B) and (D) are the boxed region shown in Fig. 1B of serum treated with the ESA Albumin Removal Kit. In (A) and (B), open circles surround spots identified by ImageMaster software as described in Section 2.6. In (C) and (D), open circles are spots found on both gels and used as reference spots to match the two regions. Closed circles are spots unique to that gel.

minimal nonspecific protein binding and offers a significant advantage over the currently available method.

Additionally, numerous polypeptides were of much lower intensity in the gel of serum treated with the ESA Albumin Removal Kit than in the gel of serum treated with ProtoClear. Six of the reference spots had a 10-fold or greater spot volume in the gel of ProtoClear-treated serum. This could be due to an unequal loading of the two gels or due to the fact that the Cibracon Blue technique removed partial quantities of these proteins. Care was taken to load the gels with equivalent volumes of treated sera by correcting for any dilution that occurred throughout the procedures. The proteins highlighted with arrows in Fig. 1 appear to be of equivalent intensity in both gels, suggesting that an equal amount of protein was indeed applied. Additionally, 15 of the reference spots used to match these regions had equivalent spot volumes and three of the spots in the gel of the Cibracon Blue-treated serum had larger volumes (~2-fold larger), as measured by ImageMaster.

Table 1. Comparing albumin removal and total protein recovery for different albumin removal techniques

Clearing method	Albumin removed[a] (%)	Serum protein recovered after albumin removal[b] (%)
ProtoClear	98 ± 2%	23 ± 2%
ESA Albumin Removal Kit	93 ± 4%	11 ± 2%

a) Albumin removal was quantified by an albumin immunoassay
b) Total protein was quantified by BCA protein assay
Percent serum protein recovered = (mg protein in cleared serum/mg protein in untreated serum)

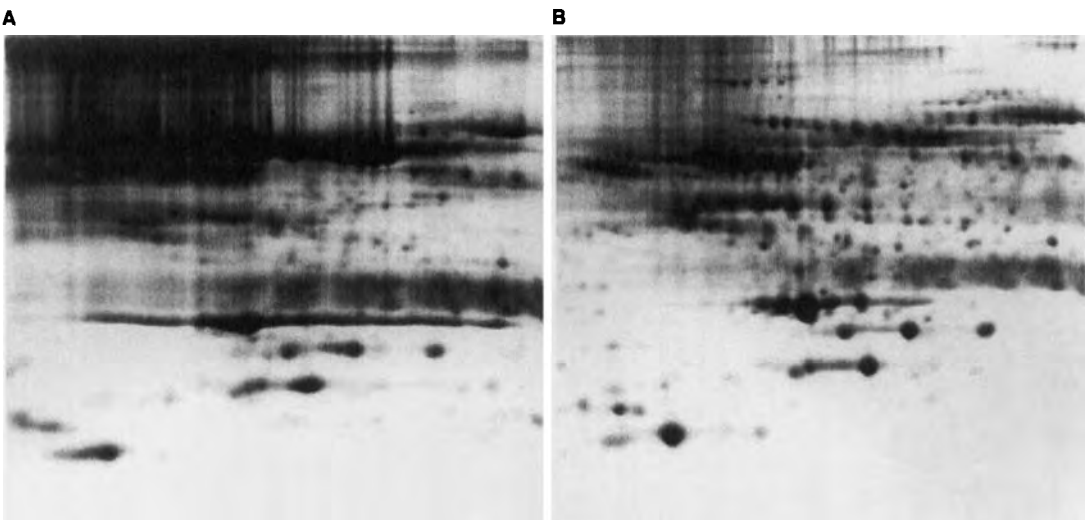


Figure 3. 2-D gel comparison of untreated serum and serum treated with ProtoClear. (A) Serum, dialyzed against 5 mM Tris, 10 mM NaCl, pH 7.5, is compared to (B) serum treated with ProtoClear. Shown is 50 µL serum equivalent, after correcting for dilution factor, loaded onto a commercial pH 4–7 IPG strip, focused and run on 14% SDS-PAGE gels.

This provides objective evidence that equal volumes were applied to the gels. Therefore, it is more likely that those polypeptides having a lower affinity for Cibracon Blue were partially removed. In total, the large degree of partial or complete removal of serum proteins makes Cibracon Blue unsatisfactory for serum-based proteomics where consistency is critical for direct quantitation of different samples. The high degree of specificity for albumin and immunoglobulin G by the ProtoClear technique is necessary to achieve our goal of identifying disease-related proteins in serum.

ProtoClear was developed to enhance the resolution of 2-DE gels of serum. A comparison of 2-D gels of serum with and without ProtoClear treatment is shown in Fig. 3. From these gels it is apparent that the large amount of albumin and IgG present in the untreated serum adversely affected the focusing of proteins over the whole pH range.

4 Discussion

We find that approximately 25–30 µL of serum (~ 2 mg of protein) can be well resolved on a commercial 3 pH unit strip (pH 4–7) and that with ProtoClear treatment the volume can be increased to 100 µL (~ 2 mg of protein). Therefore ProtoClear treatment allows 3–4 times more serum to be applied to a 2-D gel. We are developing custom Proteomix, Inc. IPG strips with an even narrower pH range to allow further increases in serum load. By decreasing the pH range to 1 pH unit and by increasing the

width of the strips we intend to increase the amount of serum that can be applied by an additional 3- to 4-fold. These techniques should enable us to load 400 μL of serum (after ProtoClear treatment), or 8 mg of protein, onto one 2-D gel and will undoubtedly allow us to visualize more of the less abundant proteins in serum. Most previous publications report analytical loads of plasma onto 2-D gels. For example, the SWISS-2DPAGE map of plasma was prepared using only 0.75 μL of human plasma [9]. These authors subsequently developed narrow pH range IPG strips with a wide cathodic end that reportedly allows up to 15 mg of plasma proteins to be resolved [10]. This technique appears to work well; however, it requires a large number of Vh (up to 500 kVh) for focusing. With currently available equipment this then requires four days for focusing, increasing the chance of protein degradation [6]. The combination of ProtoClear and narrow range strips should greatly assist in the discovery of novel serum proteins.

It is envisioned that ProtoClear can have additional uses besides removal of albumin and IgG from serum and plasma. Albumin and IgG are also found in significant concentrations in other bodily fluids such as cerebral spinal fluid, amniotic fluid, seminal fluid, and milk [1]. Removal of albumin and IgG from these fluids should enable greater visualization of the remaining proteins by 2-DE. Further experiments will validate the use of ProtoClear in these additional matrices. In its current version, ProtoClear is specific for human albumin and IgG. However, a variation of ProtoClear is under development that will bind and remove serum albumin from other species. Bovine albumin is being targeted for use in removing albumin from cell culture supernatants. In particular, this would be useful for purifying protein from genetically engineered cells that secrete only low amounts of protein. Other species that are commonly used as animal models in pharmaceutical research are also being targeted. In addition, we are working towards another variation of ProtoClear that removes the remaining isotypes of serum immunoglobulin present, *i.e.*, immunoglobulin M, immunoglobulin A, and immunoglobulin D. This will improve the visibility of proteins that resolve near the μ, α, and δ immunoglobulin heavy chains and those that resolve underneath the haze from the immunoglobulin light chains. Furthermore, it will increase the amount of serum that can be loaded onto a gel. ProtoClear was developed to assist in the search for unique disease-related proteins in human serum by 2-D gel electrophoresis. By comparing serum from normal individuals to serum from affected individuals we hope to identify disease-specific proteins. We are currently using ProtoClear to search for disease-related proteins in osteoporosis, prostate cancer, breast cancer, colon cancer, ovarian cancer and dilated cardiac hypertrophy.

Received November 3, 1998

5 References

[1] Putnam, R. W., *The Plasma Proteins*, Academic Press, New York 1975.

[2] Gianazza, E., Arnaud, P., *Biochem. J.* 1982, *201*, 129–136.

[3] Thompson, S. T., Cass, K. H., Stellwagen, E., *Proc. Natl. Acad. Sci. USA* 1975, *72*, 669–672.

[4] Gianazza, E., Arnaud, P., *Biochem. J.* 1982, *203*, 637–641.

[5] U.S. Patent Application Serial No. 09/017,284. File date 2/2/98.

[6] Rabilloud, T., Valette, C., Lawrence, J., *Electrophoresis* 1994, *15*, 1552–1558.

[7] Rabilloud, T., *Electrophoresis* 1992, *13*, 429–439.

[8] Scopes, R. K., *Protein Purification. Principles and Practice*, Springer-Verlag, New York 1987, pp. 147–148.

[9] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[10] Bjellqvist, B., Sanchez, J.-C., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1375–1378.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to. *Electrophoresis* 1999, *20*, 860–865

Carol Lynn Nilsson
Maja Puchades
Ann Westman
Kaj Blennow
Pia Davidsson

Department of Clinical Neuroscience, Unit of Neurochemistry, Göteborg University, Sahlgrenska University Hospital/Mölndal, Mölndal, Sweden

Identification of proteins in a human pleural exudate using two-dimensional preparative liquid-phase electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry

Pleural effusion may occur in patients suffering from physical trauma or systemic disorders such as infection, inflammation, or cancer. In order to investigate proteins in a pleural exudate from a patient with severe pneumonia, we used a strategy that combined preparative two-dimensional liquid-phase electrophoresis (2-D LPE), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and Western blotting. Preparative 2-D LPE is based on the same principles as analytical 2-D gel electrophoresis, except that the proteins remain in liquid phase during the entire procedure. In the first dimension, liquid-phase isoelectric focusing allows for the enrichment of proteins in liquid fractions. In the Rotofor cell, large volumes (up to 55 mL) and protein amounts (up to 1–2 g) can be loaded. Several low abundance proteins, cystatin C, haptoglobin, transthyretin, β_2-microglobulin, and transferrin, were detected after liquid-phase isoelectric focusing, through Western blotting analysis, in a pleural exudate (by definition, > 25 g/L total protein). Direct MALDI-TOF-MS analysis of proteins in a Rotofor fraction is demonstrated as well. MALDI-TOF-MS analysis of a tryptic digest of a continuous elution sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) fraction confirmed the presence of cystatin C. By applying 2-D LPE, MALDI-TOF-MS, and Western blotting to the analysis of this pleural exudate, we were able to confirm the identity of proteins of potential diagnostic value. Our findings serve to illustrate the usefulness of this combination of methods in the analysis of pathological fluids.

Keywords: Two-dimensional liquid-phase electrophoresis / Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry / Peptide mass mapping / Protein identification / Biological fluid

EL 3344

1 Introduction

Pleural effusion may occur in patients suffering from physical trauma or systemic disorders such as infection, inflammation, or cancer. The differential diagnosis of these various etiologies is a difficult clinical problem [1, 2]. Analysis of pleural effusion in clinical chemistry and pathology has traditionally relied upon immunological and cytological assays in order to determine the cause of the accumulation of fluid around the lung [3–7]. The origin of proteins in this fluid is primarily plasma, although malignant cells or microorganisms may contribute as well. Several thousand proteins can be resolved and visualized using two-dimensional gel electrophoresis [8–10]. The combination of 2-D gel electrophoresis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) has been demonstrated to be useful for the rapid identification of proteins in cell cultures [11] and tissue extracts [8–10]. But because low amounts of total protein can be loaded onto analytical 2-D gels, insufficient amounts of low abundance proteins in biological fluids such as pleural exudates may be enriched for further characterization by mass spectrometry. The protein composition in this pathological fluid is typically 70–80% albumin and immunoglobulins, whereas other proteins of potential diagnostic importance are found in trace amounts (ng/L-mg/L).

The use of a new strategy employing preparative two-dimensional liquid-phase electrophoresis (2-D LPE) and MALDI-TOF-MS was recently shown to be useful in the characterization of low abundance proteins in cerebrospinal fluid [12]. 2-D LPE is based on the same isoelectric fo-

Correspondence: Dr. Carol L. Nilsson, Department of Clinical Neuroscience, Unit of Neurochemistry, Sahlgrenska University Hospital/Mölndal, SE-43180 Mölndal, Sweden
E-mail: carol.nilsson@ms.se
Fax: +46-31-3432421

Abbreviations: CHCA, α-cyano-4-hydroxy-cinnamic acid; **2-D LPE,** two-dimensional liquid-phase electrophoresis

cusing and gel electrophoresis principles as the widely used analytical 2-D gel electrophoresis. However, in the first dimension it is possible to load large sample volumes (up to 55 mL) and protein amounts (up to 1 g), which yields larger amounts of low abundance proteins for further purification and characterization. Previously, the usefulness of the combination of liquid-phase IEF and immunoblotting was demonstrated in the enrichment of trace amounts of synaptic proteins, present in ng/L, from human cerebrospinal fluid [13]. During 2-D LPE, proteins remain in liquid phase during the entire procedure. Thus, proteins can be monitored easily, both in the first and second dimension. Extra steps necessary for the hyphenation of gel electrophoresis to mass spectrometry (electroblotting, extraction, *etc.*) are obviated when the proteins remain in liquid phase. However, 2-D LPE does not have the high resolution that characterizes 2-D gel electrophoresis. We used 2-D LPE, Western blotting, and MALDI-TOF-MS to characterize proteins in a pleural exudate (by definition, > 25 g/L total protein), which contained a large number of inflammatory cells. The aim was to study whether proteins could be identified and characterized in this complex pathological fluid, which contained high amounts of albumin and immunoglobulins.

2 Materials and methods

2.1 Materials

A 10 mL sample of pleural exudate was obtained by thoracocentesis of a patient with severe pneumonia and X-ray verified pleural effusion. Routine clinical chemical analysis showed 42 g/L of total protein, and direct microscopy showed a high number of polymorphonuclear granulocytes (1.1×10^{10}/mL) and erythrocytes (1.6×10^{6}/mL). The sample was centrifuged at 2000 × *g* at +4°C in order to remove cells and cellular debris, and kept at –70°C until analysis. Porcine trypsin, equine myoglobin, cytochrome *c*, angiotensin II, and ACTH clip (18–39) were purchased from the Sigma (St. Louis, MO, USA). The MALDI matrices used were 4-hydroxy-3-methoxy-cinnamic acid (ferulic acid; Fluka, Buchs, Switzerland) and α-cyano-4-hydroxy-cinnamic acid (CHCA; Aldrich, Steinheim, Germany).

2.2 Two-dimensional liquid phase electrophoresis

2-D LPE was performed according to experimental procedures as described previously [12]. In short, proteins in 2 mL of pleural exudate were precipitated by 10% trichloroacetic acid (TCA) and brought to a volume of 10 mL with 6 M urea, 20 mM dithioerythritol and 2.5% BioLyte carrier ampholytes, pH range 3–10 (Bio-Rad Laboratories, Hercules, CA, USA). The sample (10 mL) was loaded into the Rotofor cell (Bio-Rad Laboratories) and constant power (10 W) was applied for 4 h. Twenty separate 0.5 mL fractions were rapidly harvested and pH values determined immediately. Each fraction was analyzed using MALDI-MS, and by 12 or 15% SDS-PAGE followed by Western blotting. The remainder was dried (Speedvac). The dried Rotofor fractions containing cystatin C, as determined by Western blotting, were dissolved in 1 mL of SDS sample buffer (0.06 M Tris-HCl, pH 6.8, containing 2% SDS, 3% dithioerythritol, 10% glycerol, and 0.025% bromophenol blue), boiled for 5 min and applied to model 491 PreCell (Bio-Rad Laboratories) according to the instruction manual. The gel composition was 17%T / 2.67%C (height, 10 cm; tube size, 28 mm ID). The %T is the percentage by weight of total monomer, including crosslinker *N,N'*-methylenebisacrylamide, and %C is the proportion of crosslinker as a percentage of total monomer. The stacking gel composition was 4%T / 2.67%C (height, 2.0 cm). One hundred twenty fractions containing 2.5 mL buffer eluate were collected. To identify fractions containing cystatin C, 40 μL aliquots from every fifth fraction were concentrated and analyzed by Western blotting.

2.3 SDS-PAGE and Western blotting

Ten μL samples were electrophoresed through a 12 or 15% gel using the buffer systems of Laemmli [14]. The proteins were transferred from the gel to a PVDF membrane (Millipore, Bedford, MA, USA) using the semidry technique that utilized the Nova Blot System (Pharmacia, Uppsala, Sweden) at 0.8 mA/cm^2 for 30 min, blocked with 5% milk powder in phosphate buffered saline (58 mM $Na_2HPO_4 \cdot 2H_2O$, 17 mM $Na_2HPO_4 \cdot H_2O$, 68 mM NaCl, pH 7.4) containing 0.05% Tween-20. After blocking, the membranes were incubated overnight with a rabbit antiserum (Dakopatts, Glostrup, Denmark) against haptoglobin, cystatin C, transthyretin, β_2-microglobulin, or transferrin, diluted to 1 μg/mL, then incubated with a secondary alkaline phosphatase-conjugated antibody (Bio-Rad). The color reaction was developed with 0.015% 5-bromo-4-chloro-3-indolyl phosphatase, and 0.030% nitroblue tetrazolium in 0.1 M carbonate buffer containing 1.0 mM $MgCl_2$ [15].

2.4 SDS removal

SDS was removed by a modification [16] of the extraction method described by Wessel and Flügge [17]. The samples were centrifuged at 14 000 × *g*.

2.5 Tryptic digest

PreCell fraction 68/100, which showed the presence of cystatin C by Western blotting, was dried and dissolved in

25 μL of digestion buffer containing 0.1 mM $CaCl_2$ and 0.1 M NH_4HCO_3 in water. Trypsin (1 g/L) dissolved in 1 mM HCl and 0.1 mM $CaCl_2$ was added. A protein-enzyme ratio on the order of 500:1 was used. The sample was incubated at 37°C for 4 h.

2.6 MALDI-TOF-MS

Samples were prepared using the seed layer method [18]. In short, a matrix seed layer was created by depositing 0.5 μL of a 1 g/L solution of matrix dissolved in acetonitrile on the probe. Then, equal volumes of matrix and analyte solutions were mixed and 0.5 μL of the mixture was deposited on the matrix seed layer and dried under ambient conditions. For analysis of proteins in Rotofor fractions, ferulic acid (30 g/L) in 0.1% TFA in acetonitrile/water (1:1) was used. For tryptic peptides, CHCA (15 g/L) in 0.1% TFA in acetonitrile/water (1:1) was used. All mass spectrometric analyses were performed using a Reflex MALDI-TOF mass spectrometer (Bruker-Franzen Analytik, Bremen, Germany). Samples were irradiated with a nitrogen (337 nm) laser. The instrument is equipped with a gridless two-stage electrostatic reflectron and a pulsed ion extraction ion source. Mass spectra were analyzed using Bruker software on a Sun Sparc station, and calibrated using external calibration.

3 Results

3.1 Western blotting of proteins in Rotofor fractions

Cystatin C was detected in Rotofor fractions 7–20 as a single band at 15 kDa, covering a pH range of 7.6–8.5 (Fig. 1). Haptoglobin was detected in Rotofor fractions 3–11, pH 5.8–6.5 (Fig. 2). The heterogeneity of haptoglobin found in this pleural exudate is evident in this figure, because bands can be observed at 15, 25, and 55 kDa.

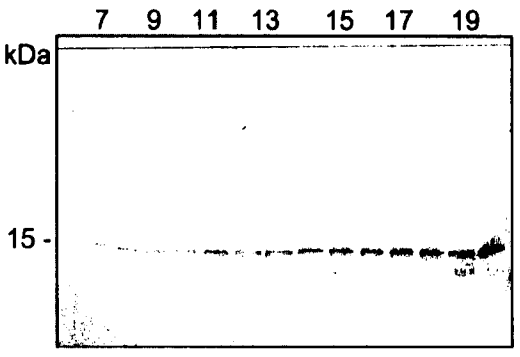


Figure 1. Western blotting of cystatin C in Rotofor (liquid-phase preparative IEF) fractions 1–20.

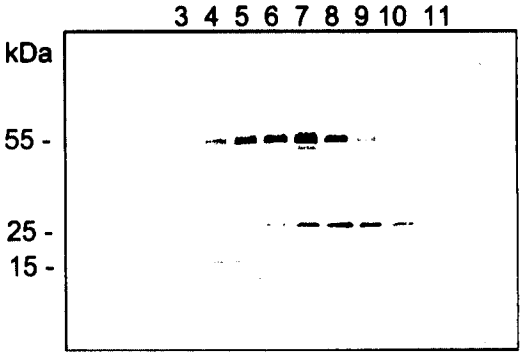


Figure 2. Western blotting of haptoglobin in Rotofor (liquid-phase preparative IEF) fractions 1–20.

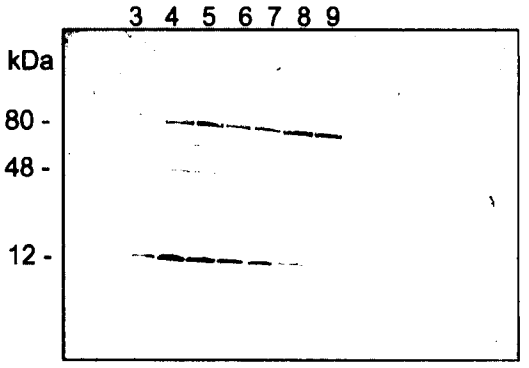


Figure 3. Western blotting of transthyretin in Rotofor (liquid-phase preparative IEF) fractions 1–20.

Haptoglobin is comprised of two alpha chains and two beta chains. In the presence of free hemoglobin, haptoglobin forms a high affinity complex with this protein. Transthyretin was localized in Rotofor fractions 3–9, pH 5.8–6.2 (Fig. 3). As in the case of haptoglobin, this protein has a heterogeneous appearance; bands may be observed at 12, 48, and 80 kDa. Transthyretin is composed of four identical polypeptide chains, each weighing 13.8 kDa. Therefore, these three bands probably correspond to the transthyretin monomers, dimers, and teramers. In addition, β_2-microglobulin and transferrin were detected at pH 6.2–7.8 and pH 6.0–7.0, respectively (data not shown).

3.2 Direct MALDI-TOF-MS analysis of Rotofor fraction 7

Several peaks were apparent in the MALDI-TOF-MS analysis of Rotofor fraction 7 (Fig. 4). Proteins at *m/z* 4606, 9141, 9197, 9374, 13766, and 15950 were observed. No further attempt was made to identify these proteins in this step. Proteins could be observed by MAL-

DI analysis of several other Rotofor fractions as well (data not shown).

3.3 Western blotting of continuous elution SDS-PAGE fractions

Rotofor fractions 12–15, which were positive for cystatin C as determined by Western blotting, were pooled and loaded on the PrepCell. Fraction numbers 65–75 were positive for cystatin C.

3.4 MALDI-TOF-MS analysis of a tryptic digest

MALDI-TOF-MS analysis of a tryptic digest of PrepCell fraction 68/120, which showed the presence of cystatin C by Western blotting, confirmed the presence of this protein (Fig. 5). The sequence coverage provided by this analysis was 44.2%. A comparison of the observed and theoretical masses of the tryptic peptides is presented in Table 1.

4 Discussion

We used preparative 2-D LPE combined with Western blotting and MALDI-TOF-MS in order to identify and characterize proteins in a pleural exudate. Compared to analytical 2-D gel electrophoresis, it is possible to load larger sample volumes and protein amounts for the first-dimensional separation. However, 2-D gel electrophoresis offers superior resolution to 2-D LPE. Liquid-phase IEF has been demonstrated to be useful in the enrichment of synaptic proteins from cerebrospinal fluid (CSF) for identification by immunoblotting [13]. 2-D LPE in combination with MALDI-TOF-MS has been used successfully to identify and characterize low abundance proteins in CSF [12]. In the present study, we have shown that proteins in other complex biological fluids such as pleural exudates can be characterized using the same combination of methods. After liquid phase IEF, cystatin C, haptoglobin, transthyretin, β_2-microglobulin, and transferrin could be detected using Western blotting.

Although it is possible to obtain mass spectra of proteins in complex biological mixtures without previous purification steps [18, 19], simple fractionation often results in a dramatic improvement of the signal-to-noise ratio. We have demonstrated that it is possible to analyze Rotofor fractions of a pleural exudate using MALDI-MS directly. To our knowledge, this is the first time that liquid-phase IEF using the Rotofor has been used as a fractionation step prior to MALDI-MS analysis. Tentative identification of proteins can be made on the basis of the experimental-

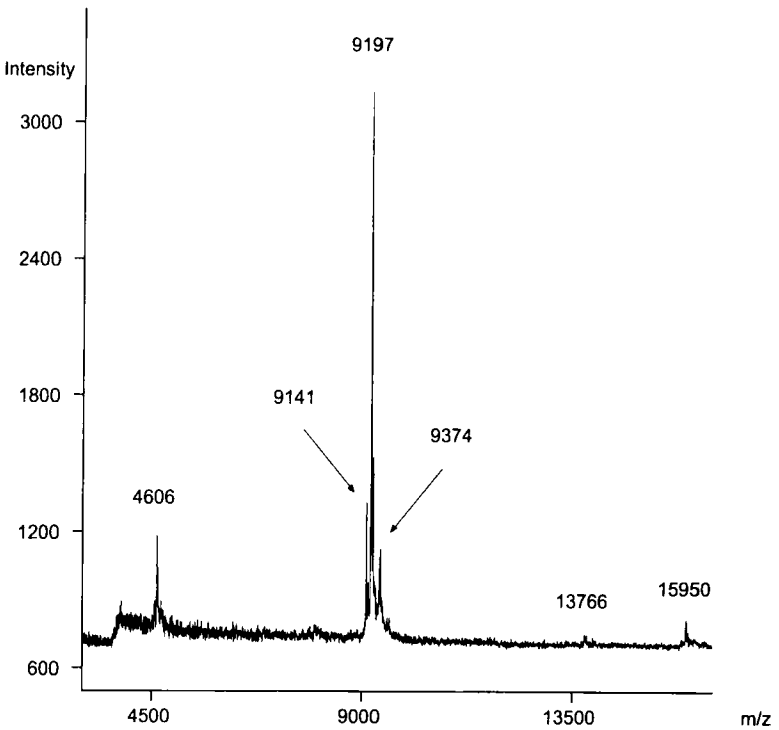


Figure 4. Direct MALDI-MS analysis of proteins in Rotofor fraction 7. The spectrum was acquired in linear mode using ferulic acid as a matrix, at an accelerating voltage of 25 kV, and is the sum of 300 laser shots.

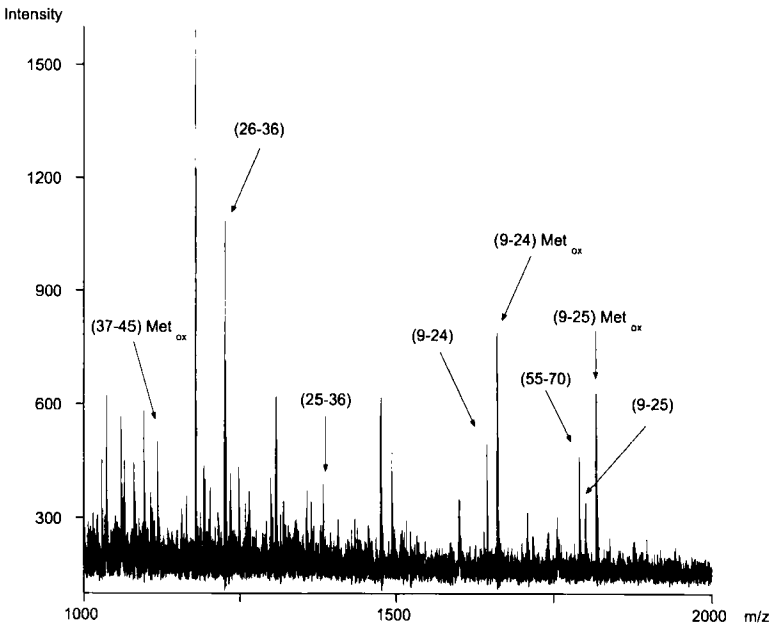


Figure 5. Tryptic digest of PreCell fraction 68/120. The spectrum was acquired in reflectron mode using CHCA as a matrix, at an accelerating voltage of 20 kV, and is the sum of 100 laser shots. Peptide sequences from cystatin C are indicated with arrows.

Table 1. Peptides observed by MALDI-TOF-MS analysis of PrepCell fraction 68/120. The monoisotopic masses of peptide sequences derived from cystatin C are indicated.

Sequence	*m/z* observed	*m/z* calculated
(9–24)	1644.4	1644.7
(9–24) Met-ox	1660.4	1660.7
(9–25)	1800.5	1800.5
(9–25) Met-ox	1816.5	1816.9
(25–36)	1382.1	1382.7
(26–36)	1226.2	1226.6
(37–45) Met-ox	1096.2	1096.4
(55–70)	1792.3	1792.9

ly determined masses, but further procedures are required for certain identification. For instance, the protein with observed monoprotonated mass 13 766 u may be transthyretin (Fig. 4; theoretical monoprotonated mass 13 762 u). Transthyretin was also detected using immunoblotting in this fraction (Fig. 3). In the case of cystatin C, the amount of protein from a single PrepCell fraction was sufficient to obtain a peptide map by enzymatic digest followed by MALDI-MS analysis of the tryptic peptides (Fig. 5, Table 1). It is possible to combine the first dimension (Rotofor) with an alternative second-dimensional separation. MALDI-TOF-MS is easily interfaced with the electrophoretic methods described here, and can be used to detect proteins in both the first (Rotofor fractions) and second (PrepCell) dimensions. Because the proteins are eluted in buffer from the PrepCell, extra steps such as electroblotting or extraction from gels are avoided. However, 2-D LPE lacks the high-quality resolution that 2-D gel electrophoresis offers.

The usefulness of this strategy has been demonstrated by the identification of proteins in a pleural exudate. In the sample studied, the total protein content (42 g/L) was more than half the protein content of plasma (about 75 g/L), and many of the proteins found in pleural exudates reflect the distribution of plasma proteins. The average values for albumin and immunoglobulins in human plasma are 45 g/L and 11 g/L, respectively. Other important diagnostic proteins are found in much lower amounts. The average plasma concentration of cystatin C and β_2-microglobulin is 1 mg/L, whereas the concentrations of transferrin, haptoglobin, and transthyretin are 2.3 g/L, 1.5 g/L, and 0.3 g/L. In inflammatory fluids such as pleural exudates, haptoglobin complexes with free hemoglobin and is rapidly cleared. Therefore, the concentration of haptoglobin in this sample was probably lower than would be expected in a plasma sample from a healthy individual. In the clinical setting, ascertaining the cause of pleural effusions continues to present diagnostic difficulties. Several markers have been evaluated for use in the clinical chemistry laboratory for distinguishing between different types of pleural effusions. For instance, haptoglobin and transferrin are used as diagnostic markers in pleural exudates [1], but more specific markers are needed. Even when invasive diagnostic procedures are used, up to 20% of pa-

tients remain without diagnosis. We have shown that proteins in pleural exudates can be characterized using the methods described. Similar procedures may also prove useful in proteome studies of cell cultures, cell media, tissue homogenates, and other complex biological fluids.

This study was funded by the Medical Society of Göteborg, the Swedish Society for Medical Research, the Swedish Medical Research Council (grant numbers 07517, 12575, 11560, 12103, 12769), under the LUA agreement (98-280, R. Ekman), Bohuslandstingets FoU fond, Sweden, the Wilhelm and Martina Lundgren Fund, Sweden, and Magnus Bergvalls Stiftelse, Stockholm, Sweden.

Received October 26, 1998

5 References

[1] Alexandrakis, M., Coulocheri, S., Kyriakou, D., Bouros, D., Xirouhaki, N., Siafakas, N., Castanas, E., Elipoulos, G. D., *Resp. Med.* 1997, *91*, 517–523.

[2] Salama, G., Miédougé, M., Rouzaud, P., Mauduyt, M.-A., Pujazon, M.-C., Vincent, C., Carles, P., Serre, G., *Br. J. Cancer* 1998, *77*, 472–476.

[3] Assi, Z., Caruso, J. L., Herndon, J., Patz, E. F., *Chest* 1998, *113*, 1302–1304.

[4] Brown, G. A., Ginsberg, P. C., Harkaway, R. C., *Urol. Int.* 1998, *60*, 197–198.

[5] Marie, C., Losser, M.-R., Fitting, C., Kevmarrec, N., Payen, D., Cavaillon, J.-M., *Am. J. Respir. Crit. Care Med.* 1997, *156*, 1515–1522.

[6] Romero, S., Fernández, C., Arriero, J. M., Espasa, A., Candela, A., Martín, C., Sánchez-Payá, J., *Eur. Respir. J.* 1996, *9*, 17–23.

[7] Villena, V., López-Encuentra, A., Echave-Sustaeta, J., Martin-Escribano, P., Ortuño-de-Solo, B., Estenoz-Alfaro, J., *Cancer* 1996, *78*, 736–740.

[8] Celis, J. E., Gromov, P., Ostergaard, M., Madsen, P., Honoré, B., Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D. B., Gromova, I., Haunsø, A., van Damme, J. V., Puype, M., Vandekerckhove, J., Rasmussen, H. H., *FEBS Lett.* 1996, *398*, 129–134.

[9] Klose, J., Kobaltz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[10] Yan, J. X., Tonella, L., Sanchez, J.-C., Wilkins, M. R., Packer, N. H., Gooley, A. A., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1997, *18*, 491–497.

[11] Matsui, N. M., Smith, D. M., Clauser, K. R., Fichmann, J., Andrews, L. E., Sullivan, C. M., Burlingame, A. L., Epstein, L. B., *Electrophoresis* 1997, *18*, 409–417.

[12] Davidsson, P., Westman, A., Puchades, M., Nilsson, C. L., Blennow, K., *Anal. Chem.* 1999, *71*, 642–647.

[13] Davidsson, P., Puchades, M., Blennow, K., *Electrophoresis* 1999, *20*, in press.

[14] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[15] Leary, J. J., Brigati, D. J., Ward, D. C., *Proc. Natl. Acad. Sci. USA* 1983, *80*, 4045–4049.

[16] Puchades, M., Westman, A., Blennow, K., Davidsson, P., *Rapid Commun. Mass Spectrom.*, in press.

[17] Wessel, D., Flügge, U. I., *Anal. Biochem.* 1984, *138*, 141–143.

[18] Westman, A., Nilsson, C. L., Ekman, R., *Rapid Commun. Mass Spectrom.* 1998, *12*, 1092–1098.

[19] Beavis, R. C., Chait, B. T., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 6873–6877.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 866–869

Michele Mortarino
Daniele Vigo
Giovanni Maffeo
Severino Ronchi

Istituto di Fisiologia Veterinaria e Biochimica, Università degli Studi di Milano, Milano, Italy

Two-dimensional polyacrylamide gel electrophoresis map of bovine ovarian fluid proteins

Molecular mechanisms underlying the cystic degeneration of ovarian follicles in the dairy cow have not been clarified yet. A useful approach for complementing endocrinological and clinical studies could be represented by the systematic analysis of the protein patterns in follicular and cystic fluid. With this aim, a two-dimensional polyacrylamide gel electrophoresis map of proteins contained in fluid from bovine ovarian follicles at different stages of development and from bovine ovarian cyst has been obtained. About 200 spots were detected after silver staining. Polypeptides from nine spots or series of spots have been identified by *N*-terminal sequencing, and further analysis of the map has been performed by gel comparison. Alpha-1-antitrypsin, albumin, serotransferrin and apolipoprotein A-I and A-IV were located on the map. Comparison between protein patterns revealed the differential expression of some spots among follicles of smaller diameter, follicles of larger diameter, and cysts. This could represent the first step toward the identification of proteins differentially expressed and associated with ovarian cyst development.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Bovine ovarian fluid proteins
EL 3342

1 Introduction

Ovarian cystic disease, characterized by a lack of normal ovarian activity and a low conception rate, is the main cause of reduced reproductive efficiency in the dairy cow. Follicle anovulation and degeneration occur mostly during the first ovarian cycle after parturition. This leads to the formation of a cystic structure, persisting up to 140 days. Clinical, cytological and endocrine aspects of ovarian cystic disease have been widely investigated, but the molecular mechanisms that result in the anovulation of Graaf follicle and in the development of a degenerative cystic form are not well understood. In particular, a major role in determining the ovulation of the follicle has been assigned to plasminogen activator synthesized by granulosa cells, leading to plasmin-mediated disruption of cell-cell interactions in the follicle wall [1]. A link between reduced concentration of plasminogen in follicular fluid and the onset of ovarian cystic degeneration has been postulated [2], but detailed studies of the protein composition of fluid contained in the Graaf follicles and in the ovarian cysts have not been performed yet.

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) represents a useful tool for the separation, comparison, and characterization of proteins from complex biological samples. In this work, we have established the 2-D PAGE map of proteins expressed in fluid contained in the dairy cow follicle and ovarian cyst. Our aim was to determine approximately how many proteins are expressed, to identify the most abundant polypeptides, and to compare the protein profile in connection with the different stages of ovarian follicle development, up to cystic degeneration. This approach could provide a useful key for the study of the nature and function of the proteins expressed in bovine ovarian fluids.

Correspondence: Prof. Severino Ronchi, Istituto di Fisiologia Veterinaria e Biochimica, Via Celoria 10, 20133 Milano, Italy
E-mail: sronchi@imiucca.csi.unimi.it
Fax: +39-2-2666301

2 Material and methods

2.1. Apparatus

The IPG separation was performed using a Multiphor™ II Electrophoresis Cell (Pharmacia, Uppsala, Sweden). Vertical slab gels (160 × 200 × 1.5 mm) were cast in a Protean™ Multi-Gel Casting Chamber and SDS-PAGE was performed using a Protean™ Multi-Cell apparatus (Bio-Rad, Hercules, CA). Electroblotting was performed with a Trans-Blot™ Cell (Bio-Rad). *N*-terminal sequencing was performed on an Applied Biosystems model 477/A protein sequenator (Foster City, CA, USA). All reagents were of electrophoresis or sequencing grade.

* Present address: Intervet Italia S.r.l., Milano, Italy

2.2 Sample preparation

Ovarian fluid samples from four Holstein Fresian cows were obtained by two-dimensional ultrasonographic-monitored transvaginal collection: from each animal, samples were collected from low-dimension (diameter less than 10 mm), medium-dimension (diameter from 10 mm to 1 cm) and high-dimension (diameter greater than 1 cm) follicles, and from ovarian cysts (diameter 2.5 cm or more), and immediately frozen. Protein concentration

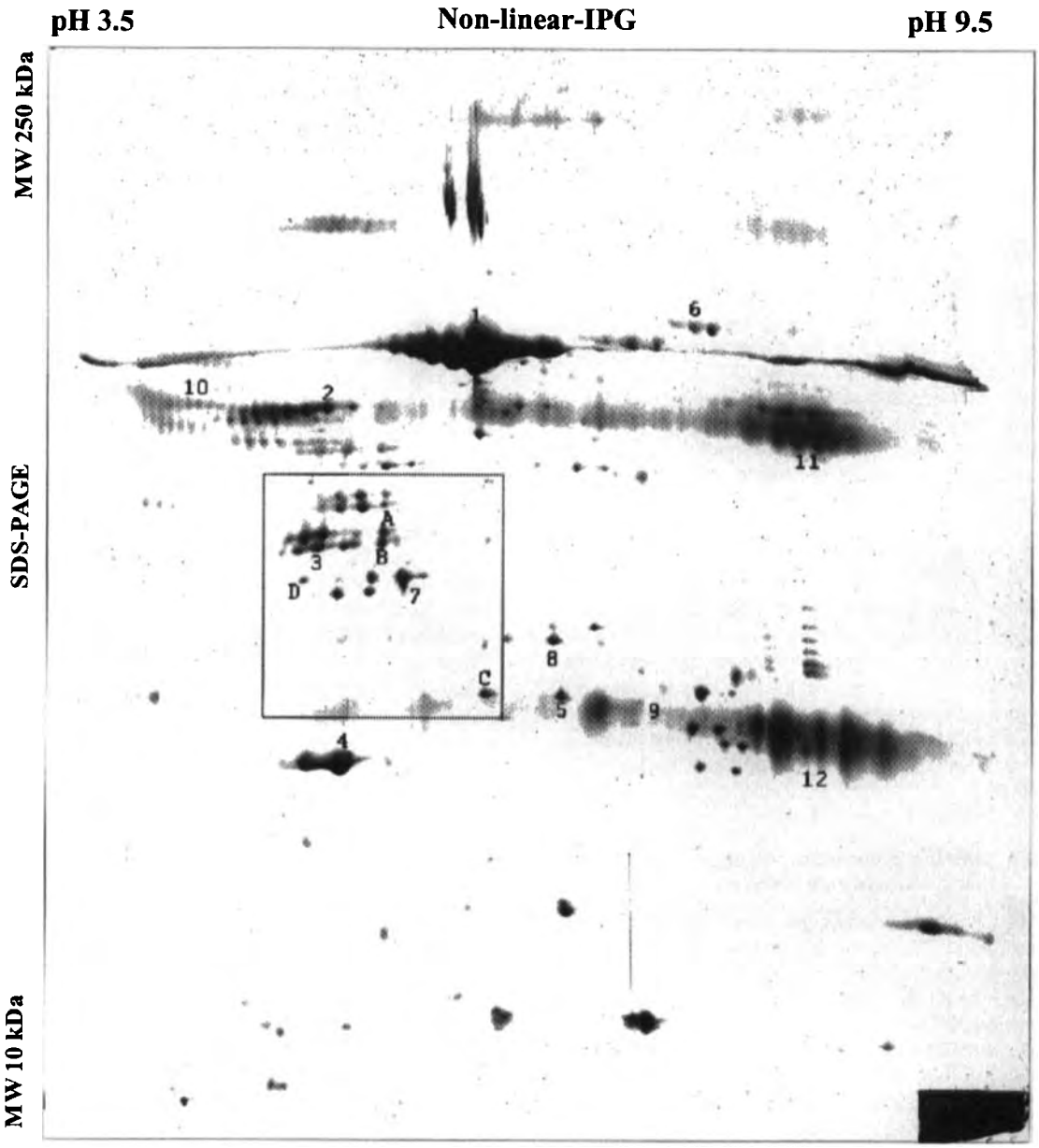


Figure 1. Silver-stained ovarian cystic fluid proteins following 2-D PAGE. Gels were scanned at 84 μm resolution using a model GS-700 imaging densitometer (12 bits/pixel; Bio-Rad) and images were processed with Melanie II [12]. Numbers refer to polypeptides listed in Table 1. See text for comments about spots marked A, B, C and D.

Table 1. List of polypeptides identified on 2-D protein map of ovarian cystic fluid

Spot	Polypeptide	Sequence
1	Albumin	DTHKS
2	Alpha-1-antitrypsin	GLVQGHAVQE
3	Apolipoprotein A-IV	EVNADQVATVXXDYK
4	Apolipoprotein A-I	DDPQSSWDRV
5	Albumin putative fragment 358–683	LAKEYEAXXEE
6	Serotransferrin	DPERTVRXXT
7	Albumin putative fragment 279–683	DKPLLEKXXXIAEVE
8	Albumin *N*-terminal fragment	DTHKXEIAHRFKDLG
9	Albumin putative fragment 356–683	LRLAKEYEAXXEE
10	Alpha-1-antichymotrypsin	Gel comparison
11	IgG heavy chains	Gel comparison
12	IgG light chains	Gel comparison

Letter X represents an unidentified residue.

was determined by spectrophotometric evaluation of absorbance at 280 nm, considering the value Abs 1 = 1 mg/mL total protein. One µL (for analytical gels) or 80 µL (for micropreparative gels) of ovarian (follicular or cystic) fluid were diluted to 350 µL with sample buffer containing 7M urea, 2 M thiourea, 1% dithioerythritol (DTE), 4% CHAPS, 0.8% Ampholine, pH 4–8, and a trace of bromophenol blue, and loaded onto commercial 3–10 NL IPG strips by in-gel rehydration as described in [3].

2.3 2-D PAGE

Protein were separated by 2-D PAGE essentially as described in [4] for the first dimension, and in [5] for the second dimension. In particular, total voltage applied was 100 kVh both for analytical IEF and for preparative IEF. The second dimension was performed on a 9–16% polyacrylamide linear gradient gel. The temperature was maintained at 15 °C during IEF and at 9 °C during SDS-PAGE.

2.4 Staining, electroblotting and microsequencing procedures

The 2-D PAGE analytical gels were silver-stained as described [6]. The micropreparative gels were alternatively stained with 0.2% Coomassie blue in 40% MeOH / 10% acetic acid / 50% water or electroblotted onto PVDF membranes as described [7]. Membranes were stained with 0.2% Coomassie blue in 40% MeOH / 60% water, destained with 40% MeOH / 60% water and dried under a stream of air. Spots of interest were excised and *N*-terminally sequenced [8]. Proteins were identified by *N*-terminal sequencing of excised spots from the PVDF membrane and comparison of the corresponding amino acid sequences (at least five residues) with those reported in SWISS-PROT entries.

3 Results and discussion

The protein pattern of each type of fluid obtained from the same cow or from different cows was highly reproducible, except in the overall intensity of staining among the maps, due to the increasing protein concentration ranging from about 30 mg/mL in the smallest follicle fluid to about 55 mg/mL in cystic fluid. Silver staining of the 2-D maps revealed about 200 spots (Fig. 1). Among these, nine spots (single spots or a series of spots) have been successfully analyzed by *N*-terminal sequencing after electroblotting onto PVDF membrane. A series of spots numbered 10, 11 and 12 was tentatively identified by gel comparison with the human plasma 2-D map [9].

Protease inhibitor alpha-1-antitrypsin and highly expressed serum-related polypeptides as albumin, serotransferrin and apolipoprotein A-I were identified (Table 1). In particular, the polypeptide corresponding to spot 3 was related to apolipoprotein A-IV, not reported yet in bovine, sharing a similar *N*-terminal sequence with human apolipoprotein A-IV. Albumin fragments were also constantly detected, despite the use of protease inhibitors. It seemed useful to compare the obtained map with published high-resolution 2-D PAGE protein maps. Among these, the human plasma 2-D PAGE map was chosen as reference map for computer matching, since bovine ovarian fluid protein samples were prepared using substantially the same method, and the overall protein pattern results were similar. In particular, bovine alpha-1-antichymotrypsin has been located in a well-defined area on the map, according to the one-dimensional SDS-PAGE migration pattern and to the high degree of amino acid sequence homology shared with the human protein [10, 11]. Moreover, putative identification of spots 10, 11 and 12 must be regarded in a critical way, and final confirmation by sequence analysis and/or other methods is needed.

Figure 2. Silver-stained 2-D protein map from the fluid contained in ovarian follicles with a diameter <1 cm. Close-up corresponding to the boxed area in Fig. 1. Spots A, B, C and D are practically absent.

Gel comparison among the 2-D protein maps from follicular and cystic ovarian fluid, both from the same cow and from different cows, revealed that the protein patterns were substantially similar. Moreover, at least two low-abundance spots of apparent p*I* 5.5, M_r 27 000 and apparent p*I* 4.8, M_r 40 000 (letters C and D in Fig. 1, respectively) are expressed only in ovarian cyst fluid and not in follicular fluid, irrespective of the follicle diameter. Furthermore, two other low-abundance spots (letters A and B in Fig. 1) are expressed only in the largest follicular (diameter > 1 cm) fluid and in cystic fluid, and are absent in smaller follicular fluid (Fig. 2). Polypeptides related to these four differently expressed spots have not been identified yet. At this early stage, there are no findings about changing of plasminogen/plasmin levels, or appearance of fragments from proteins other than albumin, in ovarian fluids analyzed. Moreover, further identification of polypeptides will be needed in order to clarify the possible role of proteases and antiproteases in altering normal follicle development.

4 Concluding remarks

These results indicate that 2-D PAGE constitutes a useful key to analyze the protein expression pattern in fluid from a Graaf follicle and from an ovarian cyst. This could represent a starting point in the search for uncommon and/or unknown proteins, and for the identification of proteins differently expressed and associated with ovarian cyst development.

We would like to thank Dr. Maurizio Chiesa for sample collection. This work was supported by a grant from the Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

Received October 23, 1998

5 References

[1] Strikland, S., Beers, W. H., *J. Biol. Chem.* 1976, *251*, 5694–5702.

[2] Hirschel, M. D., Junter, A. G., *J. Anim. Sci.* 1980, *51*, Suppl. 1, 285.

[3] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M., Hochstrasser, D. *Electrophoresis* 1997, *18*, 324–327.

[4] Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.

[5] Hochstrasser, D. F., Harrington, M. G., Hochstrasser, A.-C., Miller, M. J., Merril, C. R., *Anal. Biochem.* 1988, *173*, 424–435.

[6] Hochstrasser, D. F., Patchornik, A., Merril, C. R., *Anal. Biochem.* 1988, *173*, 412–423.

[7] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[8] Edman, P., Begg, G., *Eur. J. Biochem.* 1967, *1*, 80–91.

[9] Hughes, G. J., Frutiger, S., Paquet, N., Ravier, F., Pasquali, C., Sanchez, J.-C., James, R., Tissot, J. D., Bjellqvist, B., Hochstrasser, D. F., *Electrophoresis* 1992, *13*, 707–714.

[10] Sinha, D., Yang, X., Emig, F., Kirby, E. P., *J. Biochem.* 1994, *115*, 387–391.

[11] Hwang, S.-R., Kohn, A. B., Hook, V. Y., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 9579–9583.

[12] Appel, R. D., Hochstrasser, D. F., Funk, M., Vargas, J. R., Pellegrini, C., Müller, A. F., Scherrer, J.-R., *Electrophoresis* 1991, *12*, 722–735.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing the article, please refer to: *Electrophoresis* 1999, *20*, 870–874

Marcia Goldfarb

Anatek-EP, Portland, ME, USA

Two-dimensional electrophoresis and computer imaging: Quantitation of human milk casein

Because human casein does not precipitate from milk at its isoelectric point as does bovine casein, there is no easy method of quantitation. Casein represents only approximately 30% of the protein fraction in human milk, and the complex methods necessary for isolation cannot be used easily with small samples in a survey of a large number of mothers. Two-dimensional electrophoresis coupled with computer imaging has the potential to compare and quantitate proteins expeditiously using a small sample size. IsoDalt, a denaturing methodology, separates the casein micelle into its component parts, beta-casein, kappa-casein, *para*kappa-casein and casomorphins. Identification of these spots was made by immunoassay of a Western blot with monoclonal anti-human casein. Two spots at 24 kDa and 26 kDa, thought to be phosphorylated isomers of beta casein, were selected for quantitation. Milk samples from 20 mothers, 8 weeks *post partum*, were run on two-dimensional (2-D) gels; a slide was taken of each silver-stained gel with a Kodak control strip; the slide was scanned into powerMac Photoshop 3 with a Polaroid-Sprintscan; spots were isolated using "threshold", "mask" with IPTK (Imaging Processing Tool Kit, Reindeer Games) a Photoshop plug-in, and transferred to the NIH-Image program. Using an NIH-Image gel macro (Thomas Seebacher), the area and integrated density of the spots were measured. The Kodak control scale provided calibration and conversion to OD units. Visual scanning of the gels and computer units indicated a wide range of concentrations. To understand the range in units of weight, a standard was generated using bovine alpha casein (Sigma). Measurements will be used in a statistical program, Statview (Abacus), in an attempt to correlate information from a questionaire with casein concentration.

Keywords: Two-dimensional electrophoresis / Milk / Computer imaging EL 3366

1 Introduction

Casein is the major protein nutrient in human milk. A visual scan of 2-D gels suggests a wide variation in casein concentrations. Computer imaging and measurement, recently made possible with the development of sensitive scanning techniques and sophisticated software accompanied by manuals such as *The Image Processing Handbook* [1], give a numerical basis for determining differences. Imaging program algorithms enable standardization of gels, mitigating the problem of gel-to-gel comparison, which arises, for the most part, from silver staining. All samples used were from mothers 8 weeks *post partum*; this is essential because milk composition changes dramatically over time from colostrum through a transition phase to mature milk. Two samples, am and pm, from 20 mothers (40 samples total) were electrophoresed and quantitated. Casein is secreted as a micelle. It is denatured by 2-D methodology and is seen on the gel as approximately 14 areas related to the basic fragments κ-casein, *para*-κ-casein, β-casein, casein phosphopeptides, and casomorphins [2]. Two spots at approximately 26 kDa and 24 kDa were selected for quantitation. These were identified as phosphoserine isoforms of β-casein by immunoassay of a 2-D blot with anti-phosphoserine. The spots were chosen because they were isolated, simplifying the imaging procedure for this beginning attempt. The spots represent an estimated 75% of the β-casein isoform, that portion of the micelle which is internalized (*i.e.*, the nutrient). Using the program NIH-Image [3], area and density measurements were made of these spots. In addition, serial dilutions of weighed bovine α-casein were electrophoresed and imaged in an attempt to correlate computer units with micrograms.

Correspondence: Marcia Goldfarb, Anatek-EP, 17 Bishop Street, Portland, ME 04103, USA
E-mail: anatekep@maine.rr.com
Fax: +207-772-5644

2 Materials and methods

2.1 Materials and analyses

Human milk samples, from 2 to 8 mL, were collected in 8 mL polypropylene tubes by mothers at home and held refrigerated. The samples were picked up by this lab and centrifuged in the collection tube at 3000 × *g*, at 10°C. The lipid formed a firm band at the top; it was removed

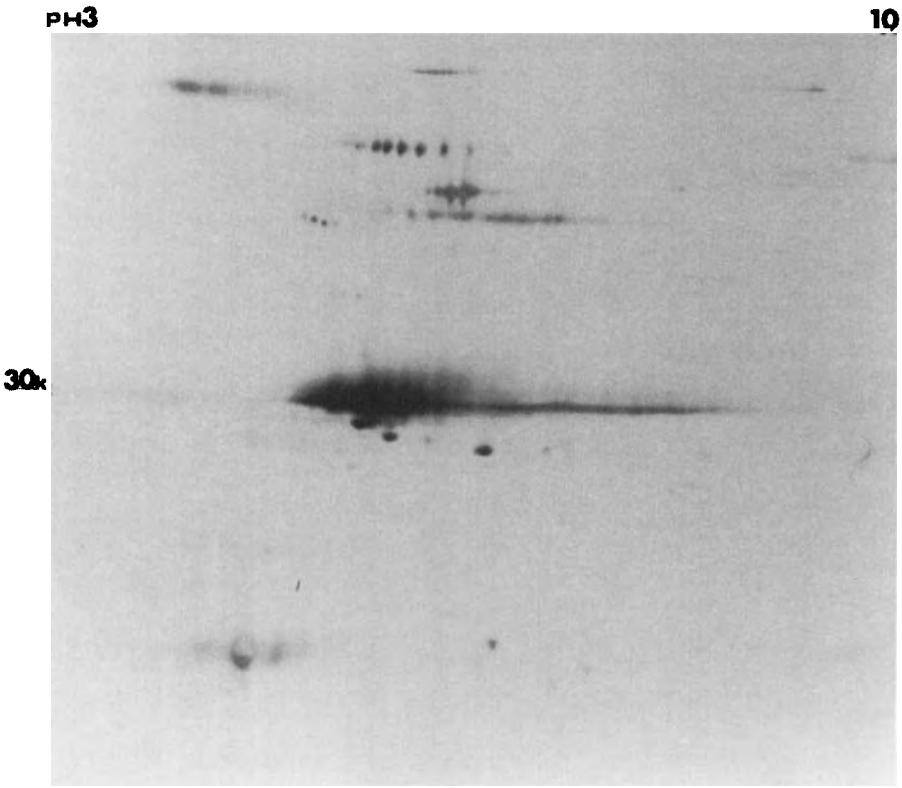


Figure 1. 2-D electrophoresis of a 10 μL human milk sample from a mother 8 weeks *post partum*. The blot in Fig. 2 identifies areas that are casein.

with a spatula and fractions were frozen at –80°C. Bovine α_s-casein was purchased from Sigma (C6780; St. Louis, MO). Dilutions were made in 0.05 M Tris-HCl, pH 8.0, and solubilized as the human milk samples. Two-dimensional electrophoresis was performed using the Iso-Dalt method [4] with a pH gradient 3.5–10 (nonlinear) in the first dimension (isoelectric focusing), and a 14 × 16 cm, 10–20% acrylamide gradient gel in the second dimension (SDS-PAGE). Samples were solubilized in 9 M urea, 4% NP-40, 2% carrier ampholytes (pH 3.5–10), 1% dithioerythritol (DTE). Silver staining was according to Morrissey [5]. Western blot was according to Towbin *et al.* [6]. Western blots were assayed with monoclonal antibody to casein (clone F20.14, Biogenesis, Sandown, NH) diluted 1:750 in PBS–0.3% Tween for 1 h, washed 4 × 5 min with PBS–0.3% Tween, 37°C, incubated with appropriate second antibody horseradish peroxidase conjugate 1:500 in PBS, 1 h, washed 4 × 5 min with PBS, and reacted with substrate TMblue (Moss, Pasadena, MD). Milk blots were also assayed with mouse anti-phosphothreonine and rabbit anti-phosphoserine, Zymed (San Francisco, CA). Analysis was performed on a Macintosh computer using the public domain NIH Image program (developed at the US National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/nihimage/).

2.2 Computer imaging

A slide was taken of each gel with a control grey scale (Kodak T-14) included, and then scanned into a Power Mac with a Polaroid Sprintscan and "saved" in Photoshop 3. Using IPTK 2.1 (Image Processing Tool Kit, Reindeer Games) a plug-in to Photoshop, the following steps were used to isolate two β-casein spots for measurement. Several duplicates of the image were made, and a histogram was looked at. The "histogram" produces a graph with the *x*-axis representing a grey scale from 0, darkest, to 256, lightest. The *y*-axis shows the number of pixels at each value. The mean and standard deviation are noted. A decision was made to consider values greater than 2 SD as background. Using that number "threshold" was applied to one duplicate and then used as "mask" for the other duplicate. This step is a method of standardizing gels, *i.e.*,

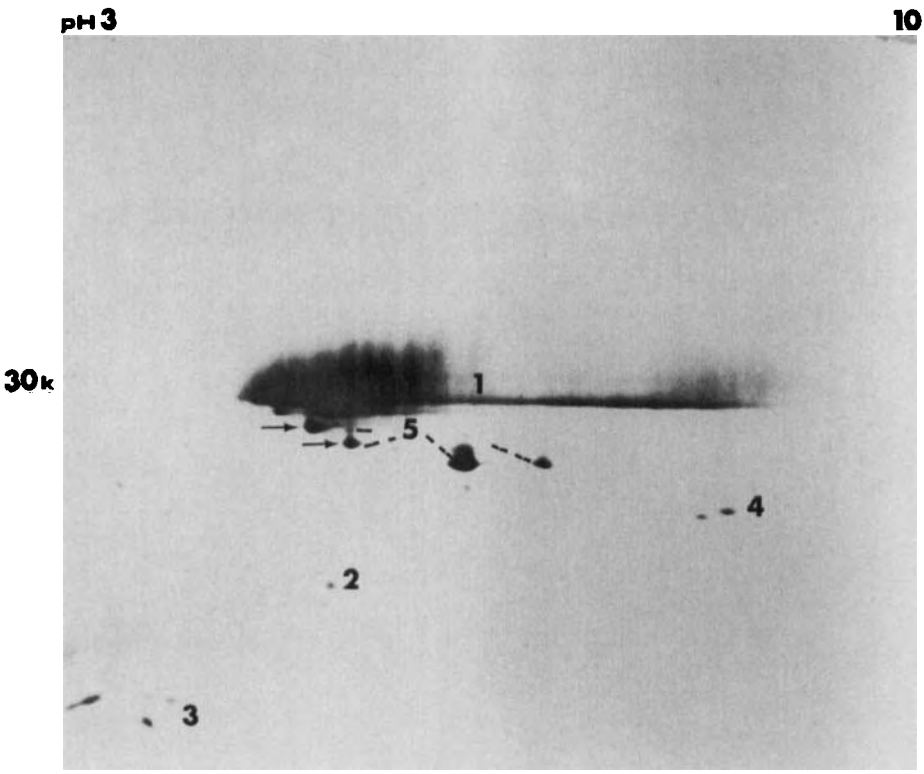


Figure 2. Western blot of a 10 μL human milk sample assayed with monoclonal antibody to casein. Casein is secreted as a micelle. Descriptions of subunits in the literature suggest the following identification of areas: 1, κ-casein; 2, para-κ-casein; 3, casein phosphopeptides/casomorphins; 4, γ-casein (beta fragments); 5, β-casein isoforms. Arrows point to the 2 spots selected for quantitation.

eliminating the differences in gel handling. The resulting image was saved (tiff) in the program NIH-Image. A macro [7] in NIH-Image was used to measure area and integrated density (mean density × number of pixels) of spots selected with "density slice". Mean density was calculated from a calibration of the Kodak scale using measured "step" values *vs.* the given "step" values and converting this from greyscale to optical density units using Rodbard, an equation which takes into account that the conversion is not linear. This equation is part of the NIH-image program. Data was exported to a StatView (Abacus) spreadsheet.

3 Results and discussion

A standard curve from bovine α-casein was generated with samples weighed and dissolved at 0.5–0.01 mg/mL, which, after solubilization and 10 μL sample load,yielded actual micrograms on gel from 1.67 to 0.03 μg. Imaging of α-casein at 5 mg/mL and 1 mg/mL indicated the α-casein spot was only 85% of the pattern, whereas, at the lower concentrations of the working range, these contaminants did not show. The concentration of the lower range was corrected to 1.42–0.025 μg, reflecting the correct percent of α-casein. Bovine α-casein was chosen as standard because it gave a single spot, whereas bovine β-casein, probably because it is fully phosphorylated and commercial samples are from many cows, had a complex pattern. While human and bovine casein are almost identical in protein sequence, human casein does not have an α-isomer. Therefore, the standard may not be optimum. Commercial human casein is available and will be looked at in the future. The linear regression had a 0.998 correlation coefficient with 100 integrated density units equal to 0.1 μg casein.

Sample spot 26 kD had an integrated density range of 17–1492 units. This was converted to concentration dividing by 100 units = 0.1 μg, giving a microgram range of 0.02–1.49 μg with a mean 0.63 μg. Spot 24 kDa had a range of 39–940 or 0.04–0.94 μg with a mean of 0.36 μg. The mean ratio of spot 24 kDa/26 kDa was 60%. For 18

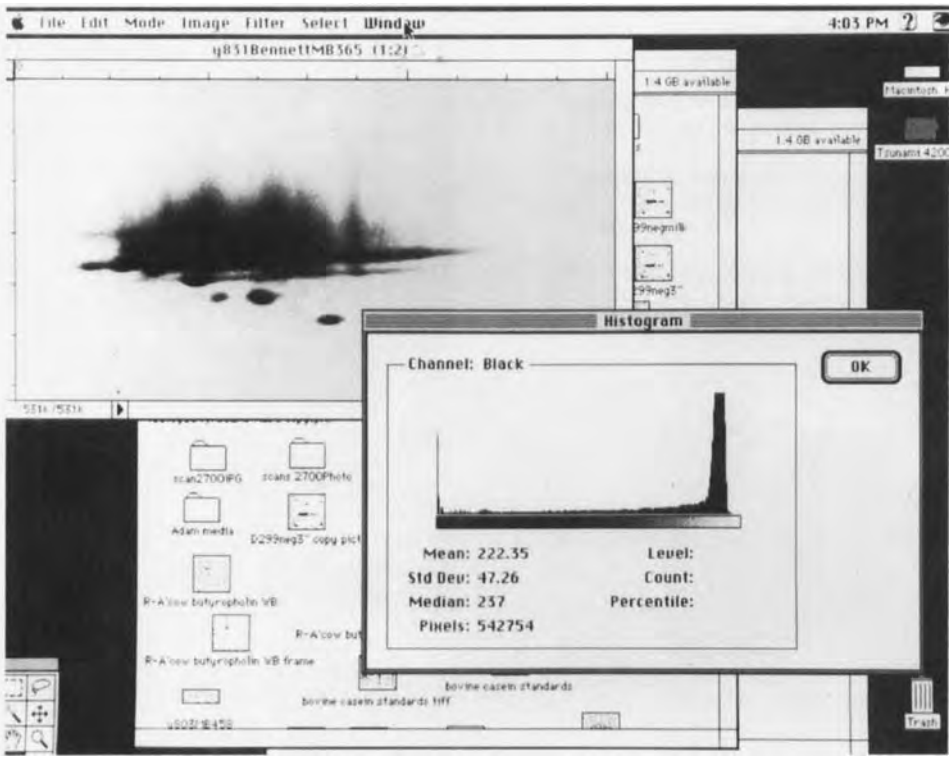


Figure 3. Computer window showing "histogram" function; *x*-axis shows computer grey scale 0–256 (black to white) and *y*-axis indicates number of pixels at each level. Background of image was determined as pixels 2 SD from the mean; "threshold" was applied with that number.

mothers, the 26 kDa spot was larger. For two mothers with the samples of lowest concentration, 24 kDa was larger than 26 kDa. Excluding these two mothers (4 samples) the ratio is 50%. Each mother gave two samples (morning and evening). The morning sample was larger, but not significant using a paired T-test. A creamatocrit measurement had been taken for each sample. This measures the lipid content as a percentage of total volume. A linear regression indicates no relationship between protein concentration and lipid concentration. This agrees with all previous research, which indicates that the major milk components, lactose, lipid, and protein, are synthesized by separate mammary functions. The 4 samples with the lowest concentration visually appear to represent a casein synthesis problem, not only because of the low concentration, but because 24 kDa area is larger than 26 kDa, the opposite of other samples. Statistical analysis of frequency distribution does not indicate that the low concentration of these samples is statistically significant. A Z-score histogram shows only the largest concentration being greater than 2 SD from the mean. Excluding the largest sample, all concentrations fall within 2 SD. The relationship of measured spots to total casein concentration needs to be calculated. Measuring the entire casein pattern, particularly the kappa casein area, will be difficult. Kappa-casein is highly glycosylated and presents as a diffuse area. A major β-casein spot, other than the two which were analyzed, was absent in two samples, suggesting genetic influence, and the difficulty of comparison.

4 Concluding remarks

The program NIH-Image has many functions for isolating the "area of interest", which, in this instance, were done using IPTK software prior to importing the image. It has the important function of converting greyscale to optical density, the units that have a relationship to concentration. The "density slice" function easily identifies saturation, a problem which has to be corrected by adjustment of 2-D parameters. It is a useful tool for quantifying 2-D gels. This data will be used to analyze information from a questionaire filled out by the mothers.

Received September 1, 1998

5 References

[1] Russ, J. C., *The Image Processing Handbook*, CRC Press, Boca Raton 1995.

[2] Lonnerdal, B., Atkinson, S., in: Jensen, R. G. (Ed.), *Handbook of Milk Composition*, Academic Press, New York 1995, pp. 353–358.

[3] NIH-Image program developed at the US National Institutes of Health is available at http://rsb.info.nih.gov/nih-image/.

[4] Anderson, L., *Two-Dimensional Elelctrophoresis; Operation of the Iso-Dalt System*, Large Scale Biology Press, Washington DC 1991.

[5] Morrissey, J. H., *Anal. Biochem.* 1981, *17*, 307–310.

[6] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.

[7] Seebacher, T., Bade, E. G., *Electrophoresis* 1996, *17*, 1573–1574.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 875–880

Franz-H. Grus
Albert J. Augustin

Department of
Ophthalmology, University
of Mainz, Germany

Analysis of tear protein patterns by a neural network as a diagnostical tool for the detection of dry eyes

The electrophoretic patterns of tears from patients with dry-eye disease (n = 43) and from healthy subjects (n = 17) were analyzed by means of multivariate statistical methods and an artificial neural network (ANN), following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). From each electrophoretic pattern a data set was created, randomly divided into test (unknown samples) and training patterns (known samples), with ANN training by one of these sets. After training, the performance of the ANN was checked by presenting the test data set to the ANN. Furthermore, the data was classified using multivariate analysis of discriminance. The groups were significantly different from each other ($P < 0.05$). The statistical procedure yielded 97% (known samples) and 71% (unknown samples) correct classifications. The ANN revealed 89% of correct classifications using the test set (unknown samples). The use of pruning algorithms (optimization procedure which automatically eliminates small weighted neurons) or genetic algorithms (optimization procedure which performs genetically induced changes of the neural net) resulted in a slight decrease of correct classifications compared to those of the nonoptimized neural network. The results reveal significant differences between the two groups. Using the ANN we were able to classify the electrophoretic tear protein pattern for diagnostic purposes.

Keywords: Tear proteins / Neural network / Dry-eye syndrome EL 3337

1 Introduction

The study of dry-eye syndrome is gaining in importance because of the high frequency of occurrence and the difficulties in treating the disease. Patients can have considerable discomfort with this syndrome. The National Eye Institute/Industry Workshop adopted the following definition of dry eye [3]. Dry eye is a disorder of the tear film due to tear deficiency or excessive tear evaporation which causes damage to the interpalpebral ocular surface and is associated with symptoms of ocular discomfort. Patients suffering from dry-eye syndromes show several clinical signs resulting from aqueous, mucin or lipid deficiency. Those deficiencies can be caused by a large variety of diseases such as Sjögren's syndrome. However, most patients do not suffer from other diseases [1, 2]. Knowledge about tear deficiencies has increased during the last years [3–8]. On the other hand, the variety of symptoms and diseases associated with dry eyes complicates the diagnosis and treatment of this disease. The clinical diagnosis of "dry-eye" is commonly based on the patient's history, slit-lamp examination, determination of tear film break-up time (BUT), Schirmer's test and Schirmer's test with anesthesia, and the so-called basic secretory test (BST); however, there is only poor correlation between the different tests [9, 10]. Recently, it was shown that the analysis of tear protein patterns based on digital image analysis of the electrophoretic patterns of tear proteins and subsequent multivariate statistical calculations can be a useful diagnostic tool for the detection of dry eyes [11]. This technique has also been successfully used in myastenia gravis, Graves' disease, and experimental uveitis [12–15].

In the present study a completely different approach for classification of the electrophoretic pattern was used: the artificial neural network (ANN). Furthermore, the efficacy of the ANN was compared to that of the multivariate analysis. The design of ANNs began in the 1940s [16, 17]. Neural networks are used in everyday life to make predictions and understand data in almost every field of human endeavor – investments, medicine, science, engineering, *etc.* We have demonstrated that neural networks are able to classify patterns of Western blots in autoimmune diseases [18, 19]. Neural networks learn from experience, not from programming. They are fast, tolerant of imperfect data, and do not need formulas or rules. ANNs are able to

Correspondence: Dr. Dr. F.-H. Grus, Universitäts-Augenklinik der Universität Mainz, D-55131 Mainz, Germany
E-mail: fgrus@csi.com
Fax: +49-6131-175566

Abbreviations: ANN, artificial neural network; **BST,** basic secretory test; **BUT,** tear film break-up time; **CTRL,** control; **DRY,** dry-eye disease

generalize and extract consistent features of patterns used to train them. The present study demonstrates that an ANN can be "trained" to recognize the electrophoretic patterns of tear proteins from tears of patients suffering from dry-eye or from healthy subjects. Because the network can classify unknown patterns to known groups learned before, this technique could be a helpful diagnostic tool.

2 Materials and methods

2.1 Patients

The tears were obtained from 60 patients, 43 patients with dry-eye symptoms (DRY) and 17 healthy subjects (CTRL). The tears were sampled using a 5 μL glass capillary. Then the tears (volume ≈ 5 μL) were stored at –20°C until use. At the time of presentation the tear film BUT and the BST were performed. The initial clinical diagnosis of "dry-eye" was based on the BST value and the presence of subjective symptoms such as burning, foreign body sensations, tearing, and "dryness" of the eyes: patients were categorized as "dry-eye" with a BST value ≤ 10 mm/5′ plus subjective symptoms.

2.2 Biochemical procedures

Tear samples were centrifuged at 12 000 *g* for 3–5 min. Of each tear sample, 0.5 μL were diluted with 2.5 μL sample buffer (62.5 mM Tris, pH 6.8, 5% v/v 2-mercaptoethanol, 10% w/v saccharose, 2% SDS, 0.005% bromophenol blue). The tear proteins were separated by SDS-PAGE on discontinuous slab gels ([20]; stacking gel: 125 mM Tris, pH 6.8, 0.1% w/v SDS, separating gel: 375 mM Tris, pH 8.8, 0.1% SDS; electrode buffer: 192 mM glycine, 25 mM Tris, pH 8.3, 0.1% SDS, MultiGel-Long; Biometra Göttingen, Germany). Molecular weights were estimated using marker proteins (molecular weight standards "broad range"; Bio-Rad, Munich, Germany). The electrophoretic separations were stained using the standard Coomassie blue procedure [21].

2.3 Digital image analysis of electrophoretic separations

Digital image analysis and evaluation of the densitometric data of the electrophoretic separations were performed by ScanPacK™ (Biometra, Göttingen, Germany) and have been described elsewhere [22–24]. ScanPacK created densitometric data files for each electrophoretic lane (separation of one tear sample), which show the grey-intensity values (8-bit grey values) *versus* the R_f values (relative mobility, *x*-axis). ScanPacK evaluated the height, area, molecular weight, R_f value, *etc.*, of all peaks in this densitometric data file and also included a photographic-quality half-tone bitmap. From each densitometric data file, two vectors containing 50 variables were built. Each variable of each vector corresponds to an R_f region (the R_f axis was broken into 50 classes; each variable of the vector represents 1/50 of the R_f region between 0 and 1). First vector (ARVEC): For the R_f value of each electrophoretic lane, the corresponding class of the 50-variables vector was determined. This class of the data vector was increased by the percent AUC (area under the curve) of this peak, which was calculated by ScanPacK. Thus, each variable of the data vector represents the percent area of the peaks of the electrophoretic lane at the corresponding R_f region. Second vector (DENSVEC): For each class of this vector, the corresponding R_f region was determined. For this particular R_f region, the average grey intensity of the electrophoretic lane was calculated. Thus, each variable of this vector contains the average grey intensity of this particular R_f region. Both data vectors were compiled into a database for subsequent calculations and each one was randomly divided in two subsets: the test (unknown data, not used in the calculation procedure) and the training set (known data, used in the calculation procedure).

2.4 Calculation procedures

2.4.1 Multivariate analysis of discriminance

As described above, ScanPacK generated densitometric data files for each electrophoretic lane. Each data file was assigned to solely one of both predefined group: CTRL or DRY. Based on these data vectors created by ScanPacK, a multivariate analysis of discriminance was performed using the ARVEC. The analysis of discriminance can test the null-hypothesis that data vectors of the groups arise from a multivariate normally distributed population. The calculation procedures have been described in detail [11, 12, 25–27]. Figure 1 shows a synopsis of the calculation procedure.

2.4.2 ANN

The human brain is a complex biological network of hundreds of billions of specialized neurons. These neurons send information back and forth to each other through connections; the result is an intelligent being capable of learning, analysis, prediction, and recognition. Artificial neural networks are built from simulated neurons which mimick "real-life" neurons in the human brain. They are connected in the same way as the brain's neurons and are able to learn in a similar manner [28]. The most widely used ANN model is the multiple layer feedforward network (MLFN) with the back-propagation training algo-

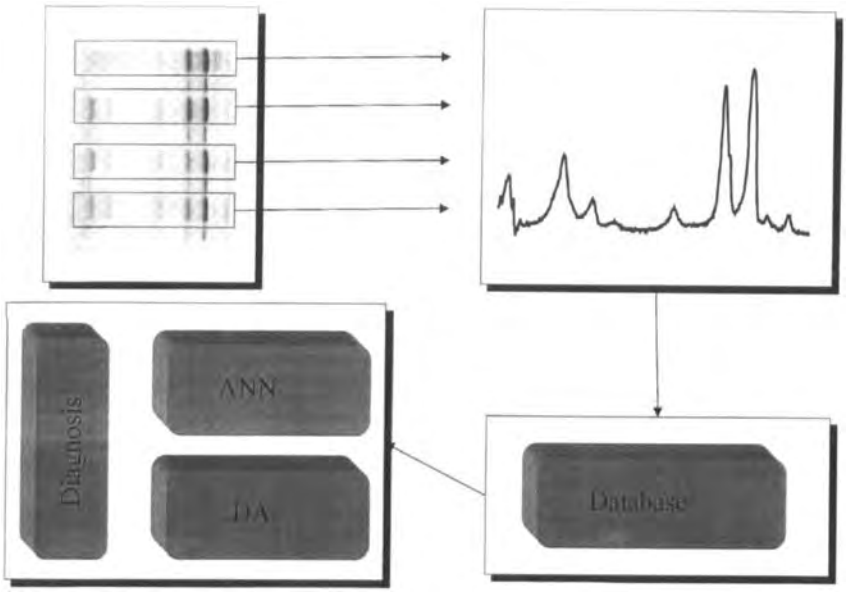


Figure 1. Synopsis of calculation procedure: scanning of SDS-PAGE, from each lane a densitograph was created. Each peak in the densitographs was quantified: (height, R_f (relative mobility), area under curve. From each densitograph, two data vectors were built, presenting (i) the mean grey values for each R_f, (ii) the areas under the curve of each peak at this particular R_f position. These vectors were compiled into a database. Based on these, a multivariate analysis of discriminance (DA) was performed and the data were presented to the ANN. Both try to classify unknown samples for diagnostic purposes.

rithm. This kind of network is typically formed by three layers: the input layer receives information from the "external world"; the output layer presents the results to the connected devices; a layer of hidden neurons is sandwiched between them. Networks are trained by presenting known samples to them. The network tries to change the function (weight) of each neuron until all training samples are classified correctly. The procedure has been described in detail [11, 18]. In this study, the Brainmaker Software (MLFN; California Scientific Software, Nevada City, CA) was used. Both ARVEC and DENSVEC are presented to the network to be learned. The network used had 50 input neurons, one hidden layer, and two output neurons.

To optimize the network, we tested two additonal approaches: (i) pruning algorithms; pruning a network means that the ANN tries to eliminate some small weighted neurons (preferable in the hidden layer), whose influence on the diagnostic decision is assumed to be negligible. The ANN attempts to find those neurons that contribute least to the decision of the ANN. The decrease in the number of hidden neurons can increase the generalization capability of the ANN; (ii) genetic algorithms; we used the Genetic Training Option of the Brainmaker Software (GTO, California Scientific Software). This option tries to change the trained network by mimicking some "genetic" techniques such as changing or eliminating neurons *via* mutation or cross-overs. (ANN mutation: Instead of changing nucleic acids as in biology, the connections and weights of the hidden neurons were changed, *e.g.*, some connections of the hidden neurons to other layers were lost, and others were introduced. ANN crossover requires two "parents": it is implemented by taking some neurons from one "parent" network and some from another – previously built by mutation – to produce the "child" network). After mutation, the new network is trained and tested, and if it performs better than its "parent", the "child" is taken and becomes "parent" for the next generation.

3 Results

Electrophoretic separation of all tear samples was performed. Figure 2 shows a typical densitograph of an electrophoretic lane from tears of the DRY group. The main peaks lactoferrin (LACT), albumin (ALB), tear-specific prealbumin (TSPA), and lysozyme (LYS) were detected and evaluated. Figure 3 shows the mean percent areas of

LYS, LACT, TSPA and ALB for the CTRL and DRY group. As in earlier studies, the analysis of discriminance revealed no statistical significant difference using only these main peaks. Using the ARVECs to perform the calculation (including all detected peaks), the analysis of discriminance revealed a significant difference between both groups ($p < 0.05$). Figure 4 shows the canonical roots of the discriminance analysis demonstrating a good separation between the samples of both groups in the discriminant space. Trying to classify unknown samples for diagnostic purposes, the statistical procedure yielded 97% and 71% correct results for known samples (ARVEC) and unknown samples, respectively. The ANN revealed 89% correct classifications using the test set (unknown samples, ARVEC). The use of pruning algorithms (automatic elimination of small weighted neurons) or genetic algorithms (genetically induced changes of the neural net) resulted in a slight decrease of correct classifications compared to those of the nonoptimized neural network (ARVEC, pruning: 79%; genetic algorithms: 72% correct results). The ARVEC group was derived as described above from the areas under the curve of detected peaks in the densitographs. The DENSVEC group represents the average mean grey values of the raw densitographic data. Training the ANN using the DENSVEC, 88% of unknown samples were correctly classified. Using BST and BUT as additional variables, 93% of the results were good. As with ARVEC, there was no increase in classification results using the optimizing algorithms (genetic algorithms: 86%; Table 1).

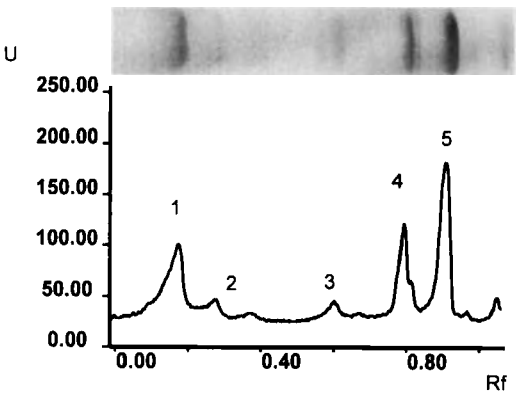


Figure 2. Densitograph and photograph of SDS-PAGE of tear proteins from the DRY group: 1, lactoferrin; 2, albumin and IgA heavy chain; 3, IgA alpha-chain; 4, tear-specific prealbumin (TSPA); 5, lysozyme. Grey values (scanner units) were plotted *vs.* R_f (relative mobility).

4 Discussion

The electrophoretic patterns of tear proteins of dry-eye patients and healthy controls were analyzed and classified by means of conventional multivariate statistical techniques and by means of artificial neural networks. The main protein peaks, lactoferrin, lysozyme, tear-specific prealbumin, and albumin were detected and quantified by digital image analysis. As in an earlier study [11], the differences between the peak areas in both groups were not significant. Using all peaks (ARVEC) detected simultaneously in the multivariate procedure, a significant difference of protein patterns between dry-eye patients and healthy controls was found. This underlines the importance of using all peak information simultaneously, because the separate analysis of each main peak resulted in no detectable significant difference. The difference found in the multivariate procedure can be used to classify the electrophoretic patterns for diagnostic purposes.

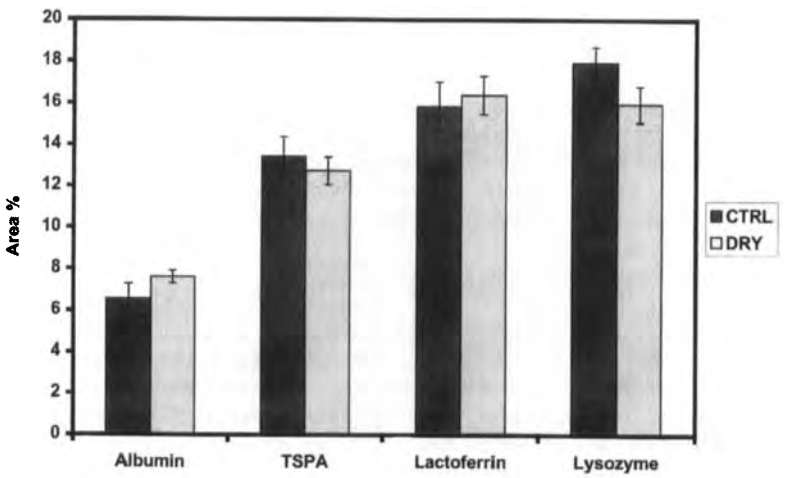


Figure 3. Mean percentage areas of the main peaks lysozyme (LYS), lactoferrin (LACT), tear-specific prealbumin (TSPA), and albumin (ALB) for the CTRL (n = 17) and DRY groups (n = 43).

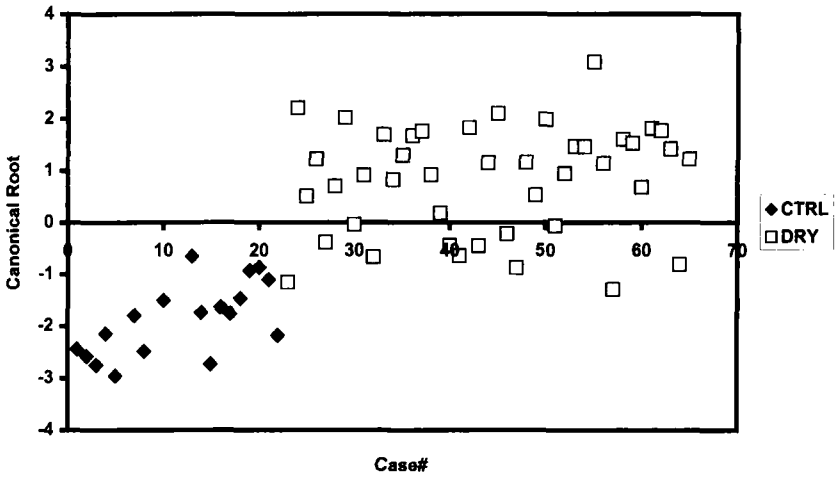


Figure 4. The canonical root of the analysis of discriminance was plotted *vs.* the case number. (The analysis was based on ARVEC data).

Table 1. Percentages of correct classifications of *unknown samples using different methods*

	%
DA	71.4
ANN-1	86
ANNGEN	72
ANNPRUN	79
ANN-2	88
ANN-2-BST	93
ANN-2-GEN	86

DA, analysis of discriminance
ANN-1, artificial neural network based on data vectors containing the peak areas
ANNGEN, ANN-1 optimized with genetics algorithms
ANNPRUN, ANN-1 optimized by pruning algorithms
ANN-2, artificial neural network based on densitographic raw data
ANN-2-BST, ANN-2 including two additional variables (BST and BUT)
ANN-2-GEN, ANN-2 changed by genetics algorithms

To classify unknown patients, the neural network and the *discriminance analysis have to learn "known" patterns of* patients with an *a priori* known group assignment. The basic secretory test and the rose bengal staining test have several limitations in the grouping of dry-eye patients [29]. We used the BST in combination with subjective symptoms as the best test for initial grouping of patients. Other clinical ELISA-based tests can only measure single parameters such as lysozyme or lactoferrin. The advantage of the methods used in this study is to include simultaneously all peaks found or even the raw data of the PAGE densitographs, *i.e.*, the complete electrophoretic pattern. *This gives the test more reliable information to decide* whether a single patient suffers from dry-eye. In comparison, in other tests the decision is based on solely one parameter.

The diagnostic advantages of our methods are demonstrated: Up to 93% of unknown patients were classified correctly. The use of statistical techniques resulted in an only 71% correct assignment. Thus, the neural network surpasses the traditional statistical technique. Furthermore, the statistical techniques have the disadvantage of *not being distribution-free. If the number of variables is* high and some variables used are similar, the results of some complex matrix calculations required in discriminance analysis can become unstable and therefore unreliable. The neural networks do not assume normal distributions and they are able to deal with even the most complex distributions, including multimodal distributions. Recently, several studies proved that neural networks can handle the classification of these data excellently [19, 27].

*Furthermore, this study attempted to train the neural net*work using two kinds of data vectors: one constaining the corresponding peak areas of the densitograph (ARVEC), the other containing the mean raw data of each densitograph. The neural network performed slightly better using the raw data (DENSVEC). Thus, no additional steps, which may lead to errors such as peak detection and quantification, are required; the neural network can deal excellently with densitographic raw data to classify electrophoretic patterns of unknown patients. This makes the

method easy to handle and allows the procedure to be automated for routine diagnostic purposes. Additionally, presented with the raw data, the neural network "decides" itself which data is important for the analysis; no subjective data reduction step is required before running the network.

In this study, all network optimization techniques, *e.g.*, pruning and genetics algorithms, failed; none could improve the performance or reliability of the net. Our method is able to automatically detect dry-eyes from the electrophoretic patterns of patients. Thus, it can become an important tool in the diagnosis of dry-eyes. Further studies have to prove the reliability of this test using a larger number of patients. In addition, those studies should provide a confidence level for the decision of the neural network.

Received September 1, 1998

5 References

[1] de Roeth, A. F. M., *Am. J. Ophthalmol.* 1952, *35*, 782–787.

[2] Van Haeringen, N. J., *Curr. Eye Res.* 1981, *26*, 84–96.

[3] Lemp, M. A., *C.L.A.O.* 1995, *2*, 221–232.

[4] Mirchoeff, A. K., Gierow, J. P., Wood, R. L., *Int. Ophthal. Clin.* 1994, *34*, 1–18.

[5] Mircheff, A. K., *Res. Prev. Blindness Sci. Writers Sem.* 1993, 51–54.

[6] Pflugfelder, S. C., Crous, C. A., Atherton, S. S., (Eds.), *Lacrimal Gland, Tear Film and Dry Eye Syndromes, Adv. Exp. Med. New York, Plenum Press, New York* 1994, pp. 641–646.

[7] Warren, D. W., *Int. Ophthalmol. Clin.* 1994, *34*, 19–266.

[8] Sullivan, D. A., *Handbook of Mucosal Immunology* Academic Press, 1994, *47*, 569–597.

[9] Holly, F. J., *Diagnosis and treatment of dry eyes in the United States*, 1st International Tear Film Symposium, Lubbock, 1984, Dry Eye Institute Inc. Lubbock.

[10] Göbbels, M., *Fortschr. Ophthalmol.* 1990, *87*, 190–197.

[11] Grus, F. H., Augustin, A. J., Evangelou, N. G., Toth-Sagi, K., *Eur. J. Ophthalmol.* 1998, *8*, 90–97.

[12] Zimmermann, C. W., Grus, F. H., Dux, R., *Electrophoresis* 1995, *16*, 941–947.

[13] Grus, F. H., Augustin, A. J., Toth-Sagi, K., Koch, F., *Adv. Ther.* 1997, *14*, 8–13.

[14] Grus, F. H., Augustin, A. J., Koch, F., Zimmermann, C. W., Lutz, J., *Adv. Ther.* 1996, *13*, 203.

[15] Grus, F. H., Augustin, A. J., Zimmermann, C. W., Spitznas, M., Sekundo, W., Lutz, J., *Graefe's Arch. Clin. Exp. Ophthalmol.* 1997, *235*, 118–123.

[16] Hebb, D. O., The Organisation of Behaviour, Wiley, New York 1949.

[17] Lashley, K. S., in: *Society of Experimental Biology Symposium, No. 4: Psychological Mechanism in Animal Behaviour*, Cambridge University Press, 1950, pp. 454–480.

[18] Grus, F. H., Zimmermann, C. W., *Electrophoresis* 1997, *18*, 1120–1125.

[19] Grus, F. H., Augustin, A. J., Toth-Sagi, K., *Ocul. Immunol. Inflamm.* 1998, *6*, 43–50.

[20] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[21] Harlow, E., Lane, D., *Antibodies: A laboratory Manual*, Cold Spring Harbor Laboratory 1988, p. 649.

[22] Grus, F. H., Nuske, J. H., in: Schemel-Trumpfheller, C. K. (Ed.), *GIT-Laborjahrbuch 1991*, GIT Verlag, Darmstadt 1991, pp. 234–235.

[23] Nuske, J. H., Grus, F. H., *BioTec* 1993, *2*, 33–35.

[24] Nuske, J. H., Grus, F. H., *BioForum* 1993, *11*, 436–437.

[25] Grus, F. H., *Dissertation*, Universität-Gesamthochschule Essen 1994.

[26] Bortz, J., *Statistik für Sozialwissenschaftler*, Springer Verlag, Berlin 1993.

[27] Søndergaard, I., Jensen, K., Krath, B. N., *Electrophoresis* 1994, *15*, 584–588.

[28] Rumelhart, D. E., Hinton, G. E., Williams, J., *Nature* 1986, *323*, 533–536.

[29] Augustin, A. J., Spitznas, M., Kaviani, N., Meller, D., Koch, F. H. J., Grus, F., Göbbels, M. J., *Graefe's Arch. Clin. Exp. Ophthalmol.* 1995, *233*, 694–698.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 881–890

Mats Lindahl
Jesper Svartz
Christer Tagesson

Department of Health and Environment, Division of Occupational and Environmental Medicine, Faculty of Health Sciences, University of Linköping, Sweden

Demonstration of different forms of the anti-inflammatory proteins lipocortin-1 and Clara cell protein-16 in human nasal and bronchoalveolar lavage fluids

The anti-inflammatory proteins lipocortin-1 and Clara cell protein-16 (CC-16) were studied in two-dimensional gel electrophoresis (2-DE) protein patterns of human nasal lavage fluids (NLFs) and bronchoalveolar lavage fluids (BALFs). Seven forms of lipocortin-1 were detected with Western immunoblots: three isoforms with an apparently normal M_r of 38 kDa and p*I* of 5.9, 6.0 and 6.1, and four truncated variants with p*I*/kDa 6.0/36, 6.4/36, 7.0/33, and 7.4/34. Four 6 kDa isoforms of CC-16 were found with p*I* 4.6, 4.8, 4.9, and 5.2. Lipocortin-1 and CC-16 were expressed in all individuals tested although not all variants were found in each individual. The overall levels of lipocortin-1 were higher in BALF than NLF and there were significant differences in the distribution of the different lipocortin-1 forms between BALFs and NLFs. One patient with occupational asthma and four children with rhinitis had increased levels of one of the truncated lipocortin-1 forms in NLF (p*I*/kDa: 7.4/34) and decreased levels of the major CC-16 form (p*I*/kDa: 4.8/6). The levels of CC-16 but not of lipocortin-1 were higher in BALF from smokers than from nonsmokers. These results indicate that the levels of lipocortin-1 and CC-16 in NLF and BALF may be altered in inflammatory airway disorders. Furthermore, the identification of different forms of the two proteins makes possible more detailed studies on the role of these proteins in inflammatory disease processes and anti-inflammatory therapies.

Keywords: Lipocortin-1 / Clara cell protein-16 / Nasal lavage fluid / Bronchoalveolar lavage fluid / Two-dimensional gel electrophoresis EL 3335

1 Introduction

Lipocortin-1 and Clara cell protein-16 (CC-16) are two anti-inflammatory proteins found in human bronchoalveolar lavage fluid (BALF) [1, 2]. The proteins, although different, appear to have structural similarities as both show amino acid sequence homology to the steroid-inducible protein, uteroglobin [3, 4]. Recombinant lipocortin-1 has been demonstrated to have anti-inflammatory properties *in vivo*, and both lipocortin-1 and CC-16 have been proposed to function as natural regulators of inflammatory reactions [2, 5]. The precise nature of the anti-inflammatory effects are unclear but direct or indirect inhibition of the pro-inflammatory enzyme phospholipase A_2 are possible mechanisms that may be common to the two proteins [2, 5–8]. Lipocortin-1 was first described as a protein released from inflammatory cells in response to steroid treatment [9]. Later it was shown to belong to a family of intracellular calcium-phospholipid binding proteins (lipocortins/annexins/calpactins) [10]. Increased levels of lipocortin-1 have been demonstrated in BALF of patients after glucocorticoid treatment [1, 11] and increased proteolytic degradation of the protein has been suggested to be of importance in different lung diseases [12, 13]. CC-16 (also described as CC-10) was first discovered as the human counterpart of rabbit uteroglobin [4]. The CC-16 gene has been linked to the atopy region of chromosome 11 [14] and the CC-16 protein suggested as a potential marker of environmental toxic exposure [15]. The possibility to analyze different forms of lipocortin-1 and CC-16 in BALF and nasal lavage fluid (NLF) may therefore have wide clinical applications.

2-DE is today becoming more and more widely used, for instance to analyze protein patterns in various biological materials, to connect the genome to the "proteome", and to characterize post-translational modifications [16]. Using 2-DE, we have previously identified a number of proteins in human NLF and BALF and been able to detect changes in the protein patterns after occupational expo-

Correspondence: Mats Lindahl, Division of Occupational and Environmental Medicine, Department of Health and Environment, Faculty of Health Sciences, University of Linköping, S-58185 Linköping, Sweden
E-mail: mats.lindahl@ymk.liu.se
Fax: +46-13145831

Abbreviations: **BALF**, bronchoalveolar lavage fluid; **CC-16**, Clara cell protein 16; **IOD**, integrated optical density; **NLF**, nasal lavage fluid

sure to noxious agents [17, 18]. Recently, lipocortin-1 and CC-16 were included in the updated list of proteins identified in 2-DE patterns of NLF and BALF [19]. In the present investigation, we used 2-DE and Western immunoblots to further investigate the expression of the two proteins in human NLF and BALF. A number of isoforms of both proteins were demonstrated and alterations in the isoform distribution were found in BALF from smokers and in NLF from subjects with asthma and allergic rhinitis.

2 Materials and methods

2.1 Samples

Nasal lavage and bronchoalveolar lavage was carried out with the informed consent of the subjects and approved by the ethical committee of the Faculty of Health Sciences, University of Linköping. NLFs were obtained from nine healthy individuals (nonsmokers) by 15 mL saline washings of the nasal mucosa performed by means of a modified "nasal pool" device [17, 20]. NLFs were also collected from one individual with occupational asthma after exposure to methyltetrahydrophtalic anhydride and four children with allergic rhinitis. The asthma patient was first referred to the Department of Occupational Medicine and then to the Pulmonary Department at the University Hospital, Linköping and diagnosed as having chronic airway obstruction with 60% of predicted FEV_1, (forced expiratory volume in 1 s). The children were referred to the Department of Pediatrics at the University Hospital, Linköping and tested for specific IgE against common allergens. All subjects tested positive against pet allergens (cat or dog) and grass pollen. BALFs were collected from nine healthy individuals (four nonsmokers and five smokers) by repeated 20 mL saline washings (100 mL total volume) with a flexible fiberoptic bronchoscope positioned in a subsegmental bronchus of the right middle or lower lobe [21]. The cell content was removed by centrifugation and the protein concentrations in the recovered NLFs and BALFs were determined with the Bio-Rad (Richmond, CA, USA) protein assay according to Bradford [22]. The protein concentrations ranged from 0.09 to 0.60 mg/mL in NLFs and from 0.22 to 1.22 mg/mL in BALFs. NLFs and BALFs were then prepared for 2-DE as described previously [17]. In order to concentrate the samples, yet avoid high salt concentrations, 2.5 mL of the samples were first desalted and then lyophilized and finally dissolved in 0.25 mL of a urea solution suitable for the first dimension, according to Görg [23].

2.2 2-DE analysis

2-DE was performed in a horizontal 2-DE setup (Multiphor from Pharmacia, Uppsala, Sweden) as described previously [17], essentially according to Görg [23]. In the first dimension (isoelectric focusing) 15 μg protein/sample was applied on IPG strips (0.5 × 3 × 180 mm), containing Immobilines giving a nonlinear pH gradient from 3–10 (Pharmacia, Uppsala, Sweden). To assure a steady state, the focusing was performed overnight (45 000 Vh). The second dimension (SDS-PAGE) was carried out by transferring the proteins to ExcelGel XL SDS 12–14 from Pharmacia (0.5 × 180 × 245 mm) or to gradient gels [24] cast on GelBond PAG film (0.5 × 180 × 245 mm, 11–18%T, 1.5%C, 33–0% glycerol) running at 20–40 mA for about 4 h. In both cases ExcelGel SDS buffer strips from Pharmacia were used. Separated proteins were detected by silver staining with a detection limit of 1–5 ng/spot [25]. Isoelectric points of the separated proteins were estimated by running a mixture of p*I* protein standards ('2D standards' excluding albumin; 4.5–8.5, Bio-Rad). Molecular mass was determined by using M_r standards in each run ('Broad range'; 200–6.5, Bio-Rad).

2.3 Evaluation of 2-DE patterns

The protein patterns in the gels were analyzed as digitized images using a CCD camera (1024 × 1024 pixels) in combination with a computerized imaging 8-bit system designed for evaluations of 2-DE patterns (Visage 4.6 from BioImage, Ann Arbor, MI, USA) [17, 18]. The amount of protein in a spot was assessed as background-corrected optical density, integrated over all pixels in the spot and expressed as integrated optical density (IOD). In order to correct for differences in total silver stain intensity between different 2-DE images, the amounts of the lipocortin-1 and CC-16 forms were expressed as percentage of the lipocortin-1 spot IOD per total spot IOD of the sample (%IOD). The distribution of different forms was calculated by dividing the IOD of one form with the ΣIOD of all the forms of the protein (%).

2.4 Western immunoblots

Lipocortin-1 and CC-16 were detected in the 2-DE patterns by Western immunoblotting. After equilibration in transfer buffer the gel was removed from its plastic bonding by a film remover (Pharmacia). The proteins were transferred in 25 mM Tris, 192 mM glycine and 20% v/v methanol [26] to a nitrocellulose membrane (0.45 μm; Bio-Rad) using semidry electrotransfer (Filtron, Bjärred, Sweden or Pharmacia), 0.8 mA/cm^2 for 1 h and the membrane was then allowed to dry overnight at room temperature. The membrane was blocked with 3% w/v gelatin and then incubated with rabbit antibodies against lipocortin-1, 1:1000 (provided by Dr. Jefferey L. Browning at Biogen Inc., Cambridge, Massachusetts, USA) [27] or CC-16,

1:4000 (urine protein 1; Dako A/S, Glostrup, Denmark) [2]. Specific spots were detected by horseradish peroxidase (HRP)-labeled goat anti-rabbit antibodies (Bio-Rad) diluted 1:20 000 (for CC-16) and 1:40 000 (for lipocortin-1), and visualized by enhanced chemiluminescence (ECL-kit from Amersham, Bucks, UK) on X-ray film. In order to translate the results from the immunodetection to corresponding silver-stained spots on the gel, a limited area of the membrane containing protein of interest was cut out and incubated with antibodies. The remainder of the membrane was stained in total by incubating with biotin + HRP-avidin (Bio-Rad) and visualized by the ECL reaction or by staining with colloidal gold (Bio-Rad). Alternatively, the area of interest was cut out directly from the gel, removed from its plastic bonding and blotted to a membrane in a mini-cell wet blotter (Mini-Protean II, Bio-Rad). The remaining gel was silver stained. Thus with both strategies, the protein spots could be localized in the 2-DE pattern, matched with a corresponding silver-stained gel and translated into the protein pattern.

2.5 Statistical analysis

The significance of the differences was calculated using a nonparametric test, Wilcoxon's rank sum test.

3 Results

3.1 Lipocortin-1 and CC-16 patterns in healthy individuals

Seven different forms of lipocortin-1 and four forms of CC-16 were demonstrated in NLF and BALF of healthy individuals (Fig. 1 and 2). Lipocortin-1 consisted of three 38

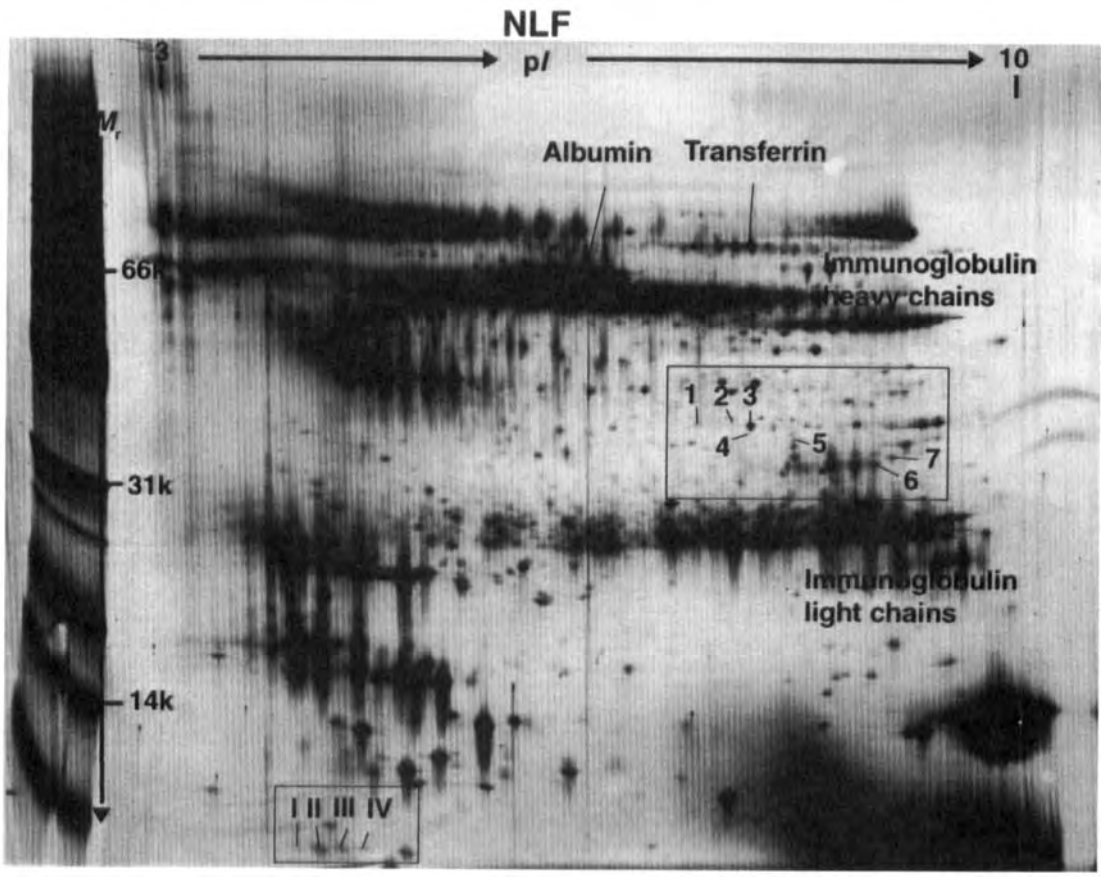


Figure 1. Localization of the seven forms of lipocortin-1 (1–7) and four forms of CC-16 (I – IV) in NLF protein patterns. Proteins were separated by 2-DE and detected by silver staining. The figure is a scanned gel presented as an unmanipulated image, except for overall contrast and light, using programs for graphic processing (Adobe Systems, San Jose, CA, USA, Photoshop 4.0 and PageMaker 6.5).

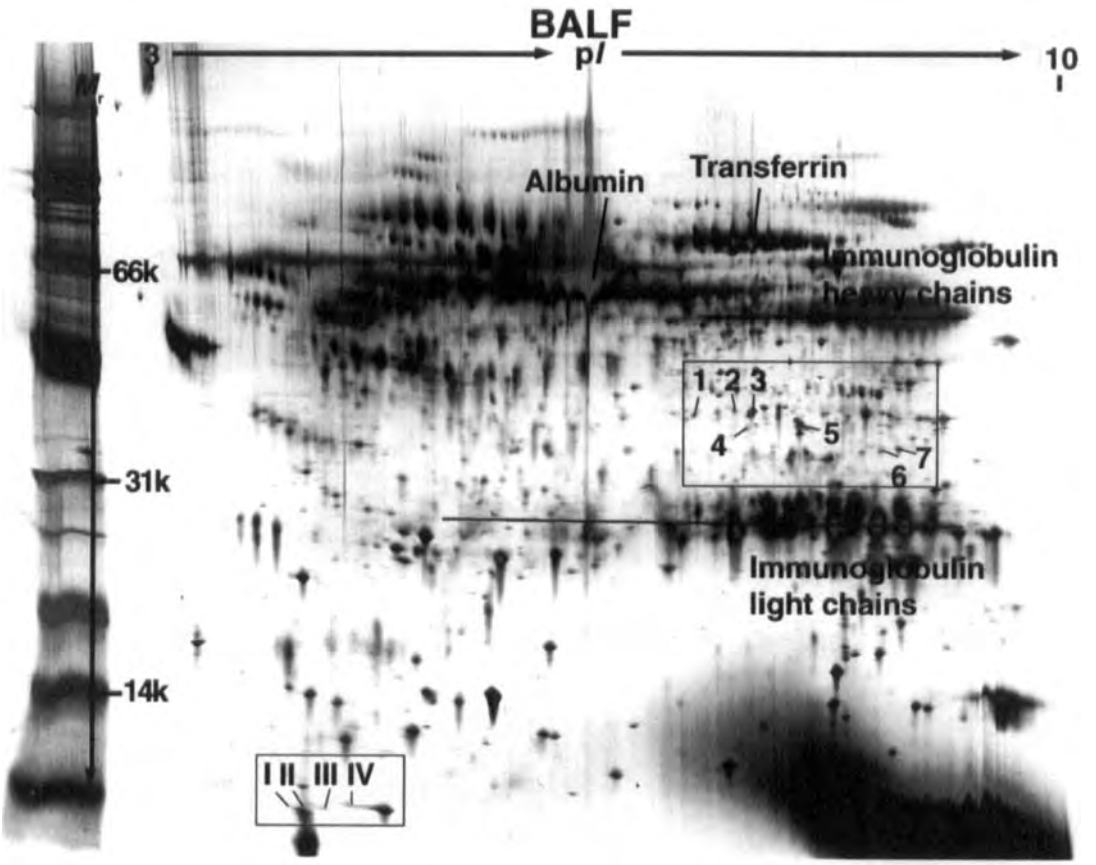


Figure 2. Localization of the seven forms of lipocortin-1 (1–7) and four forms of CC-16 (I – IV) in BALF protein patterns. Proteins were separated with 2-DE and detected by silver staining. Other conditions as in Fig. 1.

kDa isoforms and four truncated variants with molecular masses from 36 to 33 kDa, while CC-16 was expressed as four acidic 6 kDa isoforms (Table 1). Both lipocortin-1 and CC-16 were expressed in all individuals tested, although not all variants were found in all individuals (Table 1). The overall levels of lipocortin-1 were higher ($p < 0.001$) in BALF (0.31 ± 0.13 %IOD) than in NLF (0.08 ± 0.05 %IOD). In contrast to the other six forms, one lipocortin-1 form (p*I*/kDa: 7.0/33) was found in lower levels in BALF than in NLF (Fig. 3). The levels of CC-16 were the same in NLF and BALF. When comparing the BALF of smokers and nonsmokers, differences were found in CC-16, but not in lipocortin-1. The overall levels (%IOD) of CC-16 were about twofold higher in smokers (0.167 ± 0.081) than nonsmokers (0.074 ± 0.048), and one form (p*I* 4.9) was about fourfold higher in smokers (0.036 ± 0.022) than nonsmokers (0.009 ± 0.005; Fig. 4). In healthy individuals, 42 ± 17% of lipocortin-1 in NLF was found as the truncated form. The corresponding value for BALF was 57 ± 8%. Significant differences in the distribution of different lipocortin-1 forms were found between BALFs and NLFs (Table 1). As for CC-16, one form (p*I* 4.8) dominated over the others and constituted about 80% of all CC-16 in NLF and 60–70% of all CC-16 in BALF (Table 1). The distribution of the different CC-16 forms was essentially the same in NLF and BALF.

3.2 Lipocortin-1 and CC-16 patterns in patients with allergic rhinitis and asthma

In the NLF from the subjects with rhinitis, the distribution of different lipocortin-1 forms was altered (Table 2). Thus, one truncated form (p*I* 7.4; M_r 34 kDa) was found at a larger proportion than in healthy individuals ($p = 0.006$), whereas the major form (p*I* 6.1; M_r 38 kDa) was found at a lesser proportion ($p = 0.105$). The same pattern was found in a patient with occupational asthma (Table 2). The levels (%IOD, mean ± SD) of the truncated form of

lipocortin-1 was approximately twofold higher in the subjects with rhinitis (0.023 ± 0.015) and in the subject with asthma (0.022) than in healthy individuals (0.010 ± 0.035). As for the distribution of CC-16 forms in NLF, there were significant differences between healthy individuals and patients. Thus, the levels of the major CC-16

Table 1. Distribution of different lipocortin-1 and CC-16 forms in human NLF and BALF

Spot No.[a]	pI	M_r	NLF %[b]	NLF n[c]	BALF %[b]	BALF n[c]	p value[d]
Lipocortin-1							
1	5.9	38 000	4 ± 2	9/9	8 ± 4	9/9	0.019
2	6.0	38 000	4 ± 3	8/9	4 ± 4	7/9	ns[e]
3	6.1	38 000	50 ± 15	9/9	32 ± 8	9/9	0.007
4	6.0	36 000	1 ± 1	9/9	9 ± 7	4/9	< 0.001
5	6.4	36 000	8 ± 8	9/9	38 ± 10	8/9	< 0.001
6	7.0	33 000	18 ± 11	4/9	2 ± 5	9/9	0.002
7	7.4	34 000	15 ± 3	9/9	7 ± 3	9/9	0.002
CC-16							
I	4.6	6 500	4 ± 6	6/9	5 ± 3	8/9	ns
II	4.8	6 500	81 ± 15	9/9	64 ± 17	9/9	ns
III	4.9	6 500	10 ± 10	7/9	17 ± 7	9/9	ns
IV	5.2	6 500	5 ± 9	4/9	15 ± 13	8/9	ns

a) Numbers refer to Fig. 1, 2
b) Percentage of total IOD of the protein (mean ± SD)
c) Number of individuals in which lipocortin-1 form or CC-16 form was found per number examined
d) Wilcoxon's rank sum test
e) ns, nonsignificant differences

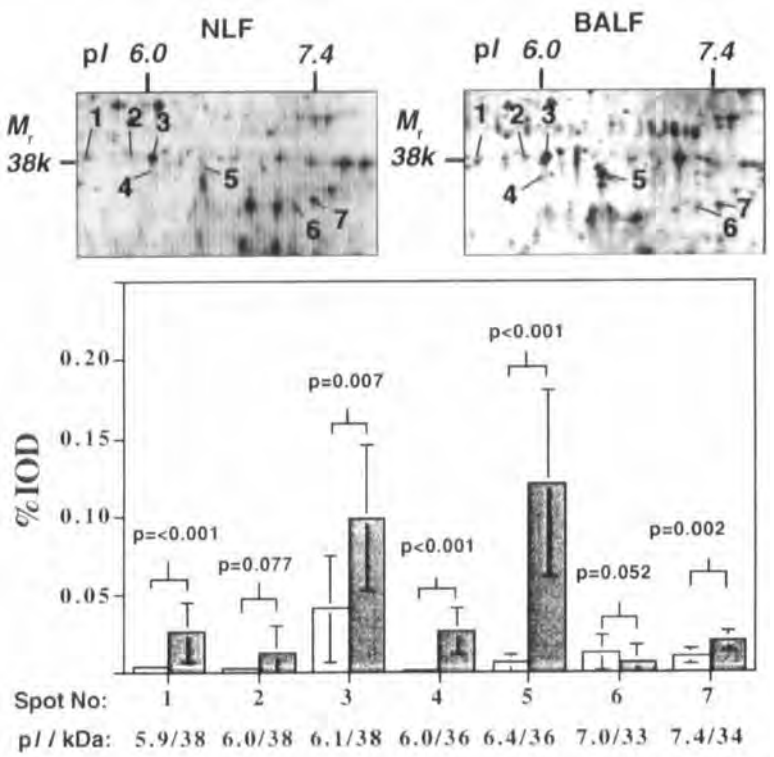


Figure 3. Levels of seven lipocortin-1 forms in NLF and BALF. NLF (open bars) and BALF (filled bars) from healthy individuals were analyzed with 2-DE. The values are mean ± SD of %IOD of nine NLFs and nine BALFs; *p* values were calculated with Wilcoxon's rank sum test. The figure also contains enlarged parts of NLF and BALF 2-DE gels of two individuals illustrating the lipocortin-1 patterns. Other conditions as in Fig. 1.

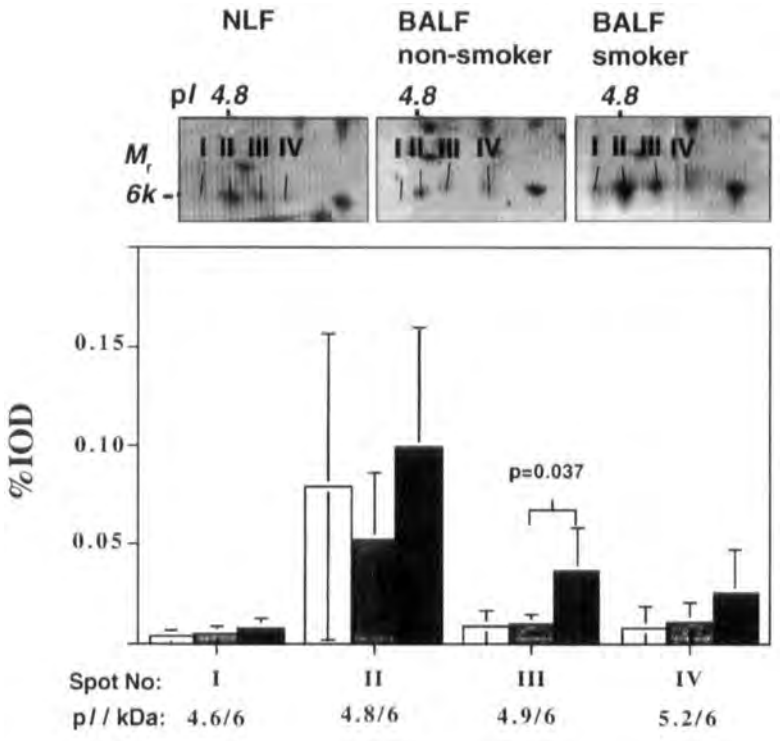


Figure 4. Levels of four CC-16 forms in NLF and BALF. NLF from nonsmokers (open bars), BALF from nonsmokers (grey bars) and BALF from smokers (black bars) were analyzed with 2-DE. The values are mean ± SD of %IOD of nine NLFs, four BALFs from nonsmokers and five BALFs from smokers; *p* values were calculated with Wilcoxon's rank sum test. The figure also contains enlarged parts of NLF and BALF 2-DE gels of three individuals illustrating the CC-16 patterns. Other conditions as in Fig. 1.

form (p*I* 4.8; M_r 6 kDa) were about 6- to 10-fold lower in subjects with rhinitis and asthma, respectively, than in healthy individuals (Fig. 5).

4 Discussion

Lipocortin-1 is an anti-inflammatory protein that is widely distributed, but particularly abundant in the lung [7]. Possible cellular sources of BALF lipocortin-1 are inflammatory cells, epithelial cells and tracheal gland cells [7, 28]. Lipocortin-1 has not been analyzed in NLF before but it is reasonable to assume that epithelial cells in the nasal cavity are the most likely source. In this study three 38 kDa isoforms (p*I* 5.9, 6.0, and 6.1) and four truncated forms (p*I*/kDa: 6.0/36, 6.4/36, 7.0/33, and 7.4/34) of lipocortin-1 were found in both NLF and BALF 2-DE patterns. An M_r of 38 kDa corresponds very well with the theoretical M_r calculated on the basis of amino acid sequence data while the apparent p*I* determined for the three 38 kDa forms is more acidic than would be expected (SWISS-PROT: //expasy.hcuge.ch/sprot/). There are numerous co- and post-translation modifications that could induce charge and mass shifts, explaining the presence of different lipocortin-1 forms. Glycosylations, acetylations, phosphorylations, and proteolytic truncations, preferentially at the *N*- or *C*-terminus, are common modifications of ex-

Table 2. Distribution of two lipocortin-1 forms in NLF from healthy individuals, and subjects with rhinitis and asthma

Spot No.[a]	p*I*	M_r (kDa)	% Controls[b]	% Rhinitis[c]	% Asthma[d]	*p* value[e]
3	6.1	38	50 ± 15	36 ± 8	42	0.105
7	7.4	34	15 ± 3	24 ± 4	25	0.006

a) Numbers refer to Fig. 1, 2
b) Percentage of total IOD of lipocortin-1, mean ± SD of 9 individuals
c) Percentage of total IOD of lipocortin-1, mean ± SD of 4 subjects
d) Percentage of total IOD of lipocortin-1, one subject
e) Controls *vs.* rhinitis, Wilcoxon's rank sum test

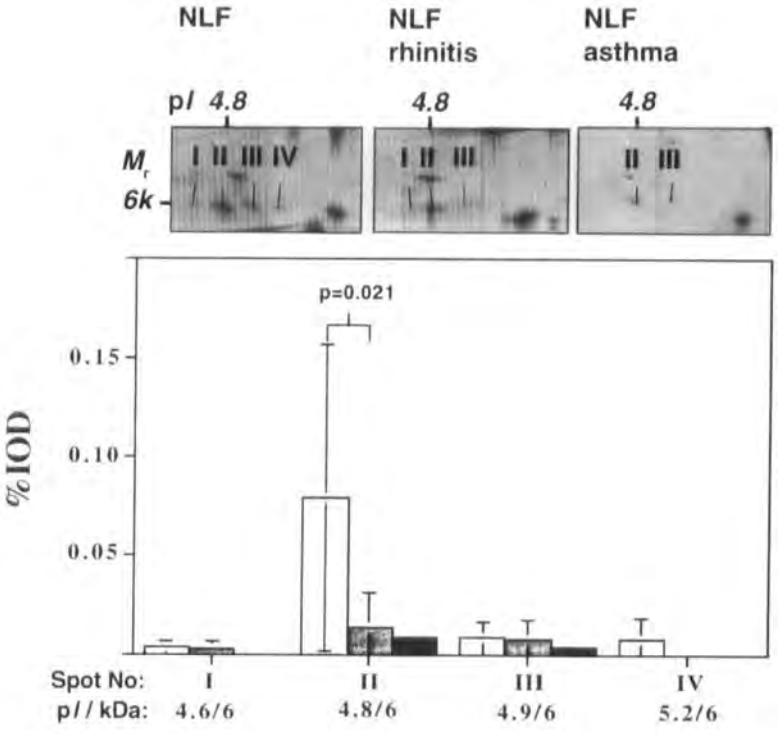


Figure 5. Levels of four CC-16 forms in NLF from healthy individuals and subjects with allergic rhinitis and asthma. NLF from healthy individuals (open bars), rhinitis (grey bars) and asthma (black bars) were analyzed with 2-DE. The values are mean ± SD of %IOD of nine healthy individuals, four subjects with allergic rhinitis and one subject with asthma, *p* values were calculated with Wilcoxon's rank sum test. The figure also contains enlarged parts of NLF 2-DE gels of three individuals illustrating the CC-16 patterns. Other conditions as in Fig. 1.

tracellular proteins [29]. While glycosylation and acetylation has not been described for lipocortin-1, both phosphorylation [5] and truncation [12, 13, 28] appear to be important modifications of lung lipocortin-1. It can therefore be hypothesized that the different forms may be derived from such modifications. To date, not more than two forms (one 37–38 kDa form and one truncated variant) have been separated and identified in extracellular fluids. Concerning cellular proteins, the only federated 2-DE database which contains lipocortin-1 is, to our knowledge, the human keratinocyte database (//biobase.dk/cgi-bin/celis). In this protein database, three 36 kDa isoforms (p*I* 6.2, 6.3, and 6.6) were identified using immunoblots with the same antibody source as in our study [30]. Truncated lipocortin-1 forms, at least in the airways, appear to exist both as intra- [28] and extracellular [13] proteins and it is unclear if some of the lipocortin-1 forms in the airways correspond to the three isoforms in keratinocytes. In human melanoma lysates, a 34 kDa form, p*I* 6.4, was separated by 2-DE and identified with MS, sequencing, and immunoblotting [31].

As judged by the IOD values of both the silver-stained and the immunoblotted 2-DE images, about 50% of lipocortin-1 appears in truncated form. This is more than we (unpublished data) and others [13] have found in BALF when analyzing lipocortin-1 with SDS-PAGE. In these analyses the truncated lipocortin-1 appears in healthy subjects as a faint band containing ≤ 10% of the total amount. One advantage of 2-DE compared with running IEF and SDS-PAGE in separate electrophoretic settings is the increased resolution in the second dimension because the proteins were previously separated by p*I*. It is possible, therefore, that the discrepancy might be caused by difficulties in discriminating between the 38 and 36 kDa forms when using only SDS-PAGE.

The human CC-16 gene is expressed in airway epithelium [14] and large amounts of CC-16 have previously been found in BALF and sputum [2, 32, 33]. Particularly Clara cells appear to be important producers of the CC-16 found in BALF [34]. As CC-16 is also expressed by nonciliated cells along the tracheobronchial epithelium [15], it is likely that similar cell types in the nasal cavity can produce the CC-16 found in NLF. The p*I* (4.6–5.2) and M_r (6.5 kDa) found in our 2-DE patterns correspond with theoretical calculations (p*I*/kDa: 4.8/8) and previously published results [15, 35]. CC-16 does not appear to be glycosylated [35]; as for lipocortin-1, however, there are several possible protein modifications that may explain

the four isoforms detected [29]. The protein is known both as CC-16 and CC-10 because it consists of an 8 kDa homodimer which migrates with an apparent M_r of about 10 kDa in SDS-PAGE [4, 15]. In some samples, especially older ones, we noted that a 16 kDa band/spot could also be detected. In separate experiments with SDS-PAGE this band disappeared when treating the samples with additional DTT. We therefore interpret this band as a dimer, noting the importance of maintaining sufficient reducing conditions.

A common problem when using Western immonoblots to identify protein spots in 2-DE is that the antibodies used may be less specific than would be desirable. The lipocortin-1 antibodies used here were polyclonal antibodies produced against human recombinant lipocortin-1, highly specific for lipocortin-1 [1, 26], and we could not detect any spots other than the ones reported herein. The CC-16 antibodies were commercially available antibodies (Dakopatts, Glostrup, Denmark), originally produced against human urinary protein-1 that later was shown to be identical to CC-16 [33]. These antibodies have been used to study CC-16 in BALF in several investigations [2, 8, 32]. In our Western blots additional bands with higher M_r were weakly stained. These bands, however, were largely removed by extensive washing after the primary and secondary antibody incubation steps, indicating that they were due to nonspecific binding of the antibodies to proteins other than CC-16. Interestingly, in NLF, but not in BALF, one additional rather strongly stained band with an apparent M_r of 20 kDa remained after washing. At present it is unclear if this protein is related to CC-16 or not. Alternative methods such as amino acid sequence analysis or peptide "fingerprinting" using MS should be used to confirm the identification made with the Western blotting technique.

When assessing the levels of lipocortin-1 and CC-16 we measured spot IOD of silver-stained gels according to a procedure used previously [17–19]. The results obtained were also confirmed by assessing IOD of Western blots of a selective number of 2-DE gels and of samples analyzed by SDS-PAGE. One advantage of analyzing the gels directly is that this is less time-consuming and more sensitive than measuring the IOD of Western blots. A problem is that silver staining is less linear than the blots. It is therefore possible that the differences in levels found in this investigation would have been even more pronounced using measurements of Western blots. For instance, the difference in CC-16 levels between the BALF of smokers and nonsmokers was about twofold higher when looking at Western blots. Remember, however, that the comparison of IOD between different forms of the proteins are based on the assumption that the different forms stained approximately equally well.

In this study we found higher levels of CC-16 in the BALF of smokers than nonsmokers. This is in contrast with other reports that have found lower levels of CC-16 in smokers [32, 36]. As the same antibodies were used in these other studies as in our investigation, we have presently no clearcut explanation for this discrepancy. Two possibilities are differences in BALF procedures or too few individuals studied in our investigation. Alternatively, the two other studies measured the CC-16 levels with immunoagglutination assay [36] and ELISA (enzyme-linked immunosorbent assay] [32] and it is possible that binding of the antibodies to other proteins than CC-16 might have occurred in these cases.

The present study demonstrated changed distribution of lipocortin-1 forms and decreased levels of CC-16 in NLF from subjects with allergic rhinitis and asthma. Although these findings are based on only few individuals and so do not allow definite conclusions, the results are of considerable interest in view of the proposed function of lipocortin-1 and CC-16 as anti-inflammatory proteins in the airways [2, 5]. It is thus possible that increased degradation of lipocortin-1 and decreased levels of CC-16 may be of significance for the inflammatory process in asthma and rhinitis. Recently, increased degradation of lipocortin-1 from a 36 kDa to a 33 kDa band was found with SDS-PAGE in BALF from patients with cystic fibrosis [13]. Experimental evidence suggested that this degradation was caused by neutrophil elastase and that the proteolytic cleavage greatly diminished the functional activity of lipocortin-1. Future studies should therefore also include the possible correlation between the number of neutrophils and the levels of the truncated form of lipocortin-1. The CC-16 gene has been identified on chromosome 11 and subchromosomally localized to a region linked to atopy and the high-affinity immunoglobulin E receptor [14]. Our finding on subjects with allergic rhinitis is the first indication on a protein level that decreased CC-16 expression in the airways may be connected to allergy.

Both lipocortin-1 and CC-16 have been proposed to function by inhibiting the proinflammatory enzyme phospholipase A_2, thereby lowering the release of arachidonic acid and lysophospholipids, with a subsequently lesser formation of inflammatory mediators such as leukotrienes, prostaglandins and PAF (platelet activating factor) [5–8]. As increased phospholipase A_2 activity has been found both in BALF and NLF from patients with allergic airway diseases [37, 38], this possibility is clearly worth considering. Both lipocortin-1 and CC-16 are also suggested to mediate anti-inflammatory effects of corticosteroids. In the

case of lipocortin-1, two investigations have shown that steroid treatment of human volunteers and patients increases the levels of the protein in BALF [1, 11], while evidence for such induction of CC-16 is still lacking in humans [2].

In summary, the identification of different forms of lipocortin-1 and CC-16 in BALF and NLF 2-DE protein patterns opens the possibility to investigate, on a molecular level, the precise role of these proteins in airway inflammation and the effect of anti-inflammatory treatments. The changed lipocortin-1 and CC-16 patterns found in NLF from patients with rhinitis and asthma also draw attention to the possibility of using 2-DE to study novel inflammatory mechanisms before and after intra-nasal anti-inflammatory treatments. Such therapies are today considered not only for treating rhinitis but also as a way to prevent the development of asthmatic conditions [39–41]. The 2-DE characterization of anti-inflammatory proteins in NLF may therefore have wide clinical applications.

Susanne Bornemo and Hadi Shahbazi are gratefully acknowledged for skillful technical assistance. Many thanks also to Bengt Ståhlbom for collecting NLFs and Professor Max Kjellman and Dr. A. H. Ferdousi at the Division of Pediatrics for their help and advice concerning the children with rhinitis. We also thank Dr. Jefferey L. Browning, Biogen. Inc., Cambridge, Massachusetts, USA for kindly providing the lipocortin antibodies. This work was supported by the Vårdal Foundation for Care and Allergy Research (A95/018), by grant B97-17x-05983-17C from the Swedish Medical Research Council and by the Swedish Foundation for Strategic Research.

Received September 1, 1998

5 References

[1] Ambrose, M. P., Hunninghake, G. W., *J. Appl. Physiol.* 1990, *68*, 1668–1671.

[2] Jorens, P. G., Sibille, Y., Goulding, N. J., van Overveld, F. J., Herman, A. G., Bossaert, L., De Backer, W. A., Lauwerys, R., Flower, R. J., Bernard, A., *Eur. Respir. J.* 1995, *8*, 1647–1653.

[3] Miele, L., Cordella-Miele, E., Facchiano, A., Mukherjee, A. B., in: Mukherjee, A. B. (Ed.), *Biochemistry, Molecular Biology, and Physiology of Phospholipase A_2 and Its Regulatory Factors*, Plenum Press, New York 1990, pp. 137–160.

[4] Sing, G., Katyal, S. L., Brown, W. E., Phillips, S., Kennedy, A. L., Anthony, J., Squegli, N., *Biochim. Biophys. Acta* 1988, *950*, 329–337.

[5] Flower, R. J., Rothwell, N. J., *TiPS* 1994, *15*, 71–76.

[6] Kim, K. M., Kim, D. K,. Park, Y. M., Kim, C.-K., Na, D. S., *FEBS Lett.* 1994, *343*, 251–255.

[7] Smith, S. F., *Thorax* 1996, *51*, 1057–1059.

[8] Lesur, O., Bernard, A., Arsalane, K., Lauwerys, R., Bégin, R., Cantin, A., Lane, D., *Am. J. Respir. Crit. Care Med.* 1995, *152*, 290–297.

[9] Blackwell, G. J., Carnuccio, R., Di Rosa, M., Flower, R. J., Parente, L., Persico, P., *Nature* 1980, *287*, 147–149.

[10] Burgoyne, R. D., Geisow, M. J., *Cell Calcium* 1989, *10*, 1–10.

[11] Smith, S. F., Tetley, T. D., Datta, A. K., Smith, T., Guz, A., Flower, R. J., *J. Appl. Physiol.* 1995, *79*, 121–128.

[12] Smith, S. F., Tetley, T. D., Guz, A., Flower, R. J., *Environm. Health Perspect.* 1990, *85*, 135–144.

[13] Tsao, F. H. C., Mayer, K. C., Chen, X., Rosenthal, N. C., Hu, J., *Am. J. Respir. Cell Mol. Biol.* 1998, *18*, 120–128.

[14] Hay, J. G., Danel, C., Chu, C. S., Crystal, R. G., *Am. J. Physiol.* 1995, *268*, L565–575.

[15] Bernard, A., Lauwerys, R., *Toxicol. Lett.* 1995, *77*, 145–151.

[16] Dunn, M. J., (Ed.), *Electrophoresis* 1997, *18*, 305–661.

[17] Lindahl, M., Ståhlbom, B., Tagesson, C., *Electrophoresis* 1995, *16*, 1199–1204.

[18] Lindahl, M., Ekström, T., Sörensen, J., Tagesson, C., *Thorax* 1996, *51*, 1028–1035.

[19] Lindahl, M., Ståhlbom, B., Svartz, J., Tagesson, C., *Electrophoresis* 1998, *19*, 3222–3229.

[20] Greiff, L., Pipcorn, U., Alkner, U., Persson, C. G. A., *Clin. Exp. Allergy* 1990, *20*, 253–259.

[21] Sörensen, J., Kald, B., Tagesson, C., Lindahl, M., *Intensive Care Med.* 1994, *20*, 555–561.

[22] Bradford, M., *Anal. Biochem.* 1976, *72*, 248–252.

[23] Görg, A., Postel, W., Günter, S., *Electrophoresis* 1988, *9*, 531–546.

[24] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[25] Swain, M., Ross, N. W., *Electrophoresis* 1995, *16*, 948–951.

[26] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.

[27] Pepinsky, R. B., Sinclair, L. K., Douglas, I., Liang, C.-M., Lawton, P., Browning, J. E., *FEBS Lett.* 1990, *261*, 247–252.

[28] Liu, L., Fisher, A. B., Zimmerman, U.-J. P., *Biochem. Mol. Biol. Internat.* 1995, *36*, 373–381.

[29] Gooley, A. A., Packer, N. H., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics, Springer-Verlag, Berlin 1997, pp. 65–91.*

[30] Celis, J. E., Rasmussen, H. H., Gromov, P., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Vorum, H., Kristensen, D. B., Östergaard, M., Haunsö, A., Jensen, N. A., Celis, A. Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kjaergaard, I., Andersen, I., Puype, M., Van Damme, J., Vanderkerckhove, J., *Electrophoresis* 1995, *16*, 2177–2240.

[31] Hall, S. C., Smith, D. M., Masiarz, F. R., Soo, V. W., Tran, H. M., Epstein, L. B., Burlingame, A. L., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 1927–1931.

[32] Shijubo, N., Itoh, Y., Yamaguchi, T., Shibuya, Y., Morita, Y., Hirasawa, M., Okutani, R., Kawai, T., Abe, S., *Eur. Respir. J.* 1997, *10*, 1108–1114.

[33] Bernard, A., Roesl, H., Lauwerys, R., Witters, R., Gielens, C., Soumillion, A., Van Damme, J., De Ley, M., *Clin. Chim. Acta* 1992, *207*, 239–249.

[34] Singh, G., Katyal, S. L., *Am. J. Respir. Cell Mol. Biol.* 1997, *17*, 141–143.

[35] Singh, G., Singal, S., Katyal, S. L., Brown, W. E., Gottron, S. A., *Exp. Lung Res.* 1987, *13*, 299–309.

[36] Lesur, O., Bernard, A. M., Bégin, R. O., *Chest* 1996, *10*, 467–474.

[37] Bowton, D. L., Seeds, M. C., Fasano, M. B., Goldsmith, B., Bass, D. A., *Am. J. Respir. Crit. Care Med.* 1997, *155*, 421–425.

[38] Stadel, J. M., Hoyle, K., Naclerio, R. M., Roshak, A., Chilton, F. H., *Am. J. Respir. Cell Mol. Biol.* 1994, *11*, 108–113.

[39] Persson, C. G. A., Svensson, C., Greiff, L., Andersson, M., Wollmer, P., Alkner, U., Erjefält, I., *Thorax* 1992, *47*, 993–1000.

[40] Corren, J., *J. Allergy Clin. Immunol.* 1997, *99*, S781–S786.

[41] Grossman, J., *Chest* 1997, *111*, 11S–16S.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 891–897

Xin Ping Li[1]
Klaus-Peter Pleißner[2]
Christian Scheler[1,4]
Vera Regitz-Zagrosek[2]
Johann Sainikow[1]
Peter R. Jungblut[3,4]

[1]Max Volmer Institute of Biophysical Chemistry and Biochemistry, Technical University Berlin, Germany
[2]Department of Internal Medicine/Cardiology, Charité, Campus Virchow-Clinic, Humboldt University and German Heart Institute, Berlin, Germany
[3]Max-Planck-Institute for Infection Biology, Berlin, Germany
[4]Wittmann Institute of Technology and Analysis of Biomolecules (WITA GmbH), Teltow, Germany

A two-dimensional electrophoresis database of rat heart proteins

More than 3000 myocardial protein species of Wistar Kyoto rat, an important animal model, were separated by high-resolution two-dimensional gel electrophoresis (2-DE) and characterized in terms of isoelectric point (p*I*) and molecular mass (M_r). Currently, the 2-DE database contains 64 identified proteins; forty-three were identified by peptide mass fingerprinting (PMF) using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), nine by exclusive comparison with other 2-DE heart protein databases, and in only 12 cases of 60 attempts *N*-terminal sequencing was successful. We used the Make2ddb software package downloaded from the ExPASy server for the construction of a rat myocardial 2-DE database. The Make2ddb package simplifies the creation of a new 2-DE database if the Melanie II software and a Sun workstation under Solaris are available. Our 2-DE database of rat heart proteins can be accessed at URL http://gelmatching.inf.fu-berlin.de/~pleiss/2d.

Keywords: / Rat / Myocardial proteins / Two-dimensional electrophoresis database / Melanie II / Gel databases / World Wide Web

EL 3331

The rat heart is an important model in studying alterations of protein expression in myocardial diseases. High-resolution two-dimensional electrophoresis (2-DE) offers a powerful tool to investigate complex protein mixtures separating up to 10 000 protein spots [1]. To store and to share this information, 2-DE databases were constructed following the concept of federated databases [2–9]. In the past years protein species were commonly identified after 2-DE separation by amino acid analysis [10, 11], Edman degradation [12, 13], internal sequencing [14], mass spectrometry [15, 16], immunoblotting [17] and comigration analysis of a known and unknown protein in 2-DE [18] shown here in a small selection of references. Here we compared both Edman degradation, which yields an exact *N*-terminal amino acid sequence, and MALDI-MS peptide mass fingerprinting (PMF), which results in better sequence coverage and higher sensitivity without the need to separate peptide mixtures. Using the Make2ddb software package we constructed our rat heart 2-DE database, allowing quick and full access to identified proteins including protein name, SWISS-PROT accession number, M_r/p*I* values and links to other protein sequence and 2-DE databases [6–8].

Correspondence: Dr. P. R. Jungblut, Max-Planck-Institute for Infection Biology, Monbijoustr. 2, D-10117 Berlin, Germany
E-mail: jungblut@mpiib-berlin.mpg.de
Fax: +49-30-28026611

Abbreviation: PMF, peptide mass fingerprinting

Rat myocardial proteins were separated by large-gel 2-DE (30 × 23 × 0.10 cm) as described by Klose and Kobalz [1] using ready-made gel solutions (WITA, Teltow, Germany). After pulverization in liquid nitrogen and dissolving in six volumes of IEF sample buffer as described previously [19], rat heart samples were applied to the anodic side of the gel and focused within a vertical isoelectric focusing chamber at 8866 Vh. Molecular mass separation was performed in the second dimension using SDS-PAGE according to Laemmli [20] with 15% acrylamide gels. The stacking gel was replaced by the equilibrated focusing gel. The master gel for computer analysis was stained with silver nitrate [21]. Proteins were either blotted onto PVDF membrane [22] and then analyzed by *N*-terminal Edman degradation (PSQ1 sequencer; Shimadzu, Kyoto, Japan) [10], or blotted onto nitrocellulose membrane and applied to PMF using MALDI-MS (Voyager™-Elite; Perseptive Biosystems, Framingham, MA, USA) after tryptic digestion (Lamer *et al.*, in preparation; [23]). Proteins were identified by searches in different databases (SWISS-PROT.r35, NCBInr.08.11.98, Genpept.r107, dbEST.07.10.98) using MS-Fit (http://prospector.ucsf.edu).

Evaluation and processing of the 2-DE rat heart gels were performed by Melanie II gel analysis software. The gel image scanned on a desktop scanner and saved in tag image file format (TIFF) was imported to Melanie II. After spot detection and labeling of known protein spots with their SWISS-PROT accession number the gel map was

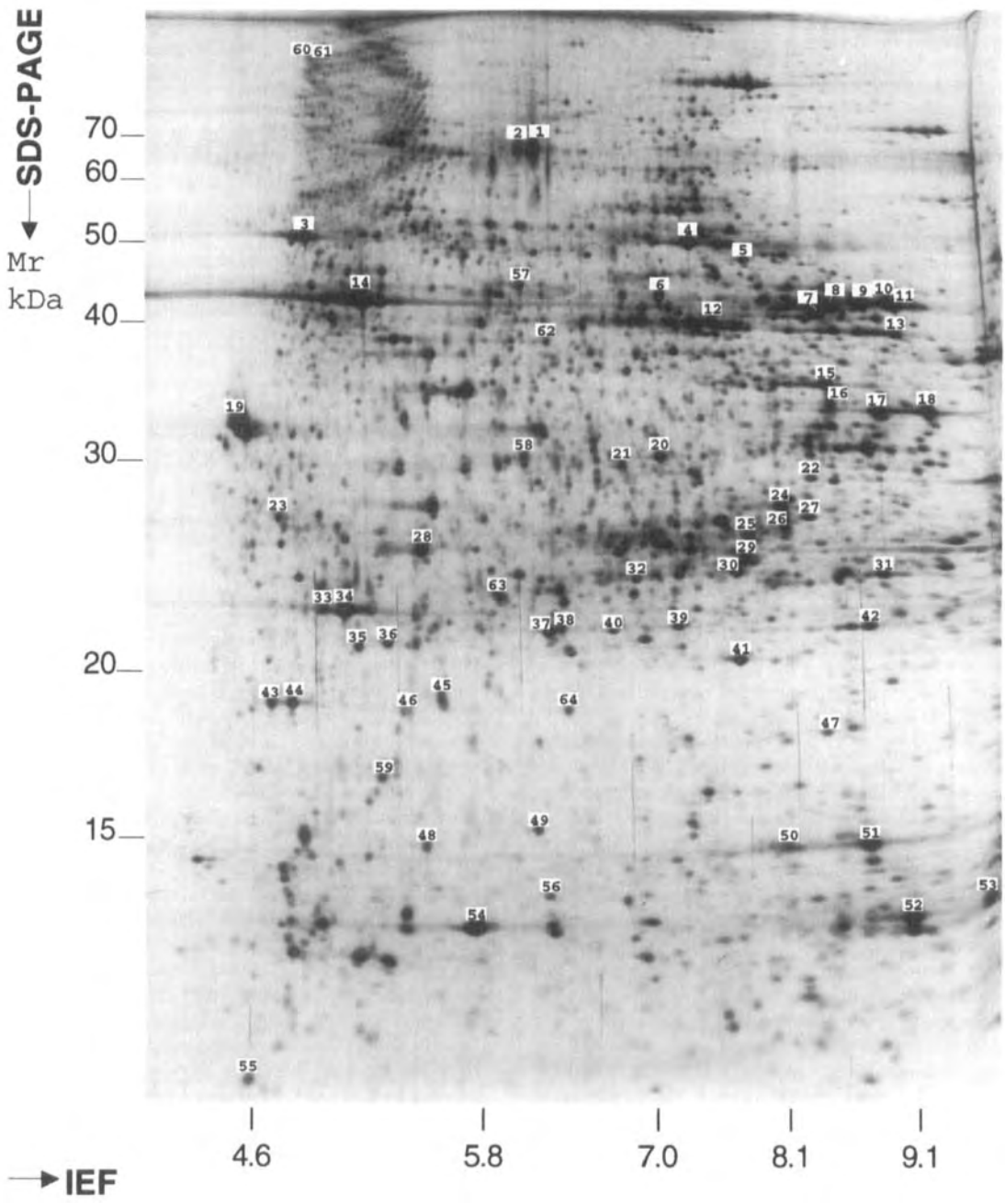


Figure 1. Standard 2-DE pattern of rat myocardial proteins. The gel size was 23 × 30 × 0.1 cm. Proteins were detected by silver nitrate staining. Calibration of isoelectric point and molecular mass was performed with theoretical values of identified proteins using Melanie II gel analysis software. Spots marked by numbers were identified. Spot numbers correspond to the spot numbers in Table 1.

Table 1. List of rat heart proteins identified in our 2-DE database

Spot No.	Database protein identification	Label number of Melanie	SWISS-PROT accession number	2-DE gel M_r	p*I*	Identification methods
1	Albumin	P02770	P02770	68700	6.1	Comparison[a)]
2	Albumin	P02770	P02770	69300	6.0	Comparison[a)]
3	ATP synthase beta subunit	P19511	P19511	52000	4.9	PMF
4	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	51021	7.3	PMF
5	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	48151	7.8	PMF
6	Thiolase	P13437	P13437	42700	7.1	PMF
7	Creatine kinase, M chain	P00564	P00564	41900	8.3	*N*-terminal sequencing
8	Unknown protein from 2-D PAGE of heart tissue (spot P11)	P11	P56576	41700	8.5	*N*-terminal sequencing Q(S/A)AREQG
9	Creatine kinase, sarcomeric mitochondrial	P09605	P09605	41800	8.7	*N*-terminal sequencing
10	Unknown protein from 2-D PAGE of heart tissue (spot P9)	P9	P56575	41800	8.9	*N*-terminal sequencing: QERRQSPE
11	Isocitrate dehdrogenase (NADP)	P56574	P56574	41600	9.0	*N*-terminal sequencing: AEKRIKVEKKVVV
12	Creatine kinase, M chain	P00564	P00564	40300	7.5	PMF
13	Fructose-bisphosphate aldolase	P05065	P05065	39300	9.0	*N*-terminal sequencing
14	Actin	P03996	P03996	42200	5.2	Comparison[a)]
15	Glyceraldehyde 3-phosphate dehydrogenase	P04797	P04797	35700	8.4	PMF
16	Glyceraldehyde phosphate dehydrogenase	P04797	P04797	34200	8.5	PMF
17	Malate dehydrogenase	P04636	P04636	33700	8.9	PMF
18	Malate dehydrogenase	P04636	P04636	33800	9.2	*N*-terminal sequencing
19	Tropomyosin	P04692	P04692	33000	4.6	PMF
20	Electron transfer flavorprotein alpha subunit precursor	P13803	P13803	31200	7.1	PMF
21	Probable peroxisomal enoyl hydratase	Q62651	Q62651	30600	6.8	PMF
22	Unknown protein from 2-D PAGE of heart tissue (spot p5)	P5	P56573	29800	8.3	*N*-terminal sequencing: FQYDSQYD(G/P)F
23	Troponin T	P50453	P50453	27700	4.9	PMF
24	Enoyl-CoA hydratase, mitochondrial precursor	P14604	P14604	28200	8.1	*N*-terminal sequencing
25	ATP synthase alpha subunit precursor	P15999	P15999	26700	7.8	PMF
26	Glyceraldehyde phosphate dehydrogenase	P04797	P04797	27100	8.1	PMF
27	Unknown protein from 2-D PAGE of heart tissue (P3)	P3	P56572	27700	8.3	*N*-terminal sequencing
28	Actin	P04270	P04270	26100	5.5	PMF, comparison[a)]
29	Ubiquinol-cytochrome *c* reductase iron-sulfur subunit precursor	P20788	P20788	25700	7.8	PMF
30	Triosephosphate isomerase	P48500	P48500	24900	7.7	PMF
31	ES1 protein (fragment)	P56571	P56571	24800	8.9	*N*-terminal sequencing
32	Creatine kinase, M chain	P00564	P00564	24600	6.9	PMF
33	Myosin light chain 1	P16409	P16409	23200	5.0	Comparison[a)]
34	Myosin light chain 1	P16409	P16409	23200	5.1	Comparison[a)]

Table 1. continued

Spot No.	Database protein identification	Label number of Melanie	SWISS-PROT accession number	2-DE gel M_r	p*I*	Identification methods
35	Phosphatidylethanolamine-binding protein	P31044	P31044	21600	5.2	PMF
36	Actin (fragment)	P04270	P04270	21700	5.4	PMF
37	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	22200	6.2	PMF
38	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	22300	6.3	PMF
39	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	22400	7.2	PMF
40	ATP synthase alpha chain mitochondrial precursor	P15999	P15999	22200	6.7	PMF
41	Alpha crystallin B chain	P23928	P23928	2100	7.8	PMF
42	Superoxide dismutase	P07895	P07895	22500	8.8	*N*-terminal sequencing
43	Myosin light chain 2	P08733	P08733	19400	4.8	PMF
44	Myosin light chain 2	P08733	P08733	19400	4.9	PMF
45	Myosin light chain 1	P16409	P16409	19700	5.6	PMF
46	Myosin light chain 1	P16409	P16409	19100	5.4	PMF
47	Creatine kinase, M chain	P00564	P00564	18322	8.4	PMF
48	Myosin light chain 1	P16409	P16409	14700	5.6	PMF
49	Cu/Zn superoxide dismutase	P07632	P07632	15200	6.1	PMF
50	Myoglobin	P04248	P04248	14700	8.1	Comparison[a)]
51	Myoglobin	P04248	P04248	14800	8.8	Comparison[a)]
52	Hemoglobin beta chain	P02091	P02091	12900	9.1	Comparison[a)]
53	Hemoglobin alpha chain	P01946	P01946	13300	9.8	Comparison[a)]
54	Fatty acid binding protein	P70623	P70623	12600	5.8	PMF
55	Myosin light chain 2	P08733	P08733	9425	4.6	PMF
56	Transthyretin precursor (Prealbumin)	P02767	P02767	13400	6.2	PMF
57	Myosin heavy chain, fragment	P80		44300	6.0	PMF
58	Myosin heavy chain, fragment	P31		31000	6.0	PMF
59	Myosin heavy chain, fragment	P61		16800	5.3	PMF
60	Myosin heavy chain, fragment	P123		92600	5.0	PMF
61	Myosin heavy chain, fragment	P124		88400	4.8	PMF
62	Myosin heavy chain, fragment	P120		38800	6.2	PMF
63	*Acidic calcium-independent phospholipase A2	P122		23800	6.0	PMF
64	*Alpha B crystallin-related protein	P14		19100	6.4	PMF

* The proteins with * were identified in NCBI instead of SWISS-PROT

a) Comparison with other 2-DE heart databases

saved in Melanie format. In the next step the 2-DE rat heart database was established using the software package Make2ddb [24] downloaded from the ExPASy server http://www.expasy.ch/ch2d/make2ddb.html. The hardware requirement for the Make2ddb software package is a Sun Microsystems workstation under the Solaris operating system. Hoogland *et al.* [24] describe its use in detail.

More than 3000 protein spots were detected on the silver-stained 2-DE master gel. The standard 2-DE pattern was calibrated by the theoretical M_r/p*I* values of identified proteins. In Fig. 1 the 2-DE master gel of rat myocardium is shown, including identified spots marked with spot numbers according to Table 1. Sixty four of the protein spots subjected were identified. Only twelve of 60 protein spots which were subjected to Edman degradation were suc-

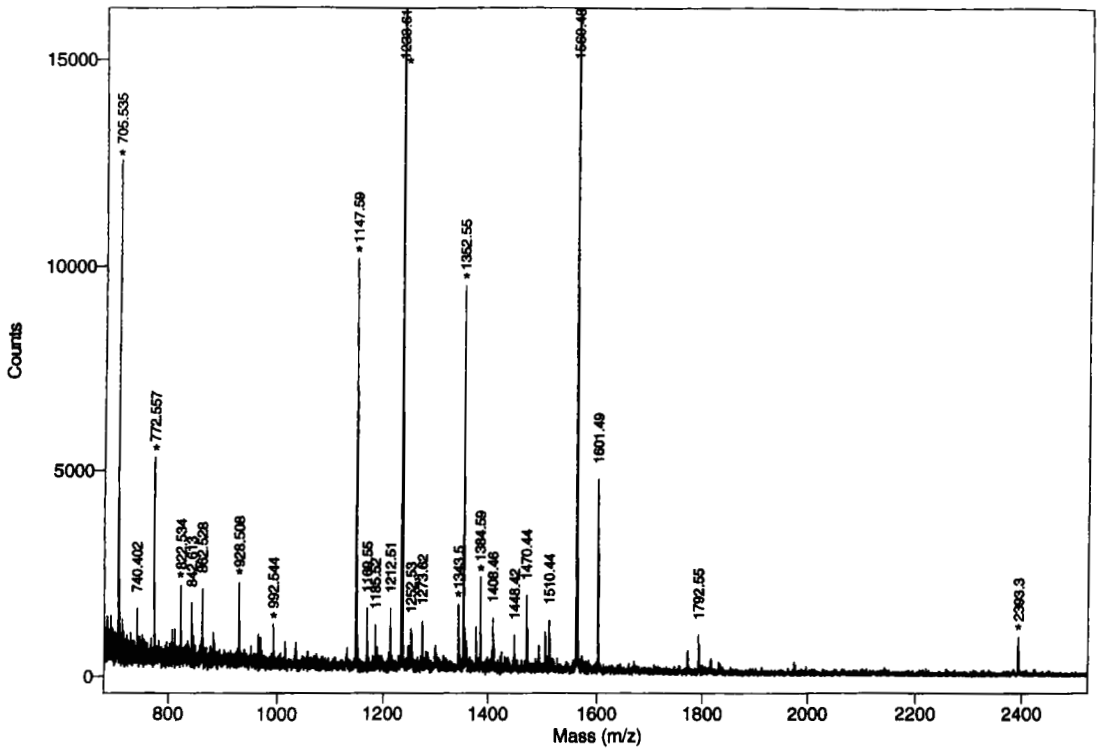
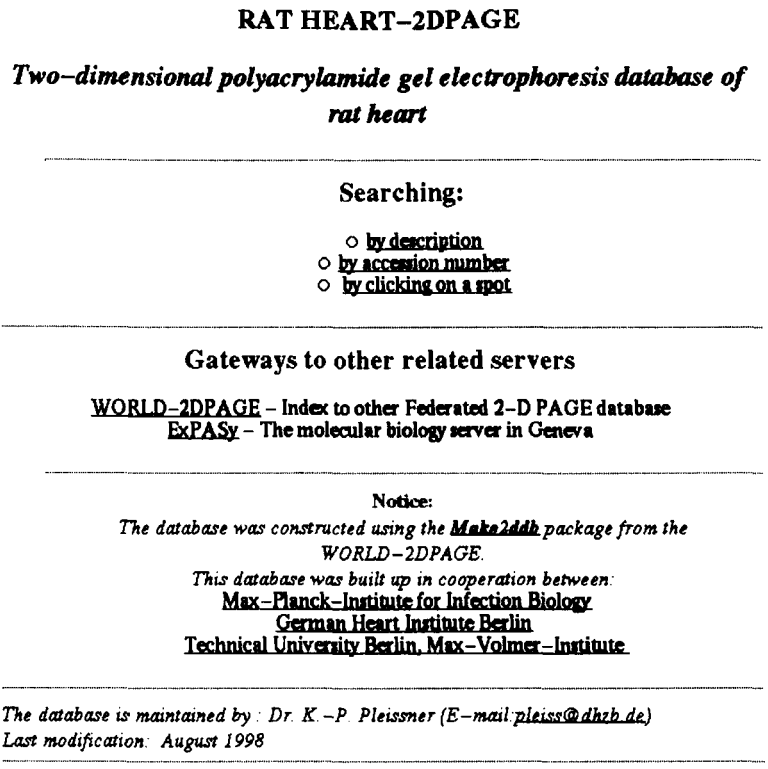


Figure 2. MALDI-MS spectrum of malate dehydrogenase after tryptic digestion. Fifteen peptide masses labeled with * matched the theoretical digestion masses yielding 35% sequence coverage. The spectrum was obtained using the following settings during measurement at Voyager™-Elite: 20 000 V accelerating voltage, 70% grid voltage, 0.05% guide wire voltage; 100 ns delay; low mass gate, 500; laser power, 2000.

cessfully *N*-terminally sequenced and have been identified by database searches. Forty-three of 60 protein spots subjected to MALDI-MS were identified by PMF after tryptic digestion. In Fig. 2 a mass spectrum of malate dehydrogenase is shown, yielding a sequence coverage of 35%. Eleven mass peaks were matched, two of them oxidized at methionine residues of peptides *m/z* 705.5 (KPGMTR) and *m/z* 1343.49 (FVFSLVDAMNGK). Besides methionine oxidation, artificial formation of pyroGlu caused by lyophilization was commonly observed in other PMF measurements. In most cases the theoretical protein mass was similar to the apparent M_r. Additionally, nine protein spots were exclusively assigned by comparison with other 2-DE databases (HP-2DPAGE Database [7] at URL http://www.mdc-berlin.de/~emu/heart, HEART-2DPAGE [2, 6] at URL http://userpage.chemie.fu-berlin.de/~pleiss, HSC-2DPAGE [4, 8] at URL http://www.harefield.nthames.nhs.uk/nhli/protein). The protein patterns of rat heart and human heart 2-DE images are highly similar. Thus, significant spots like myoglobin, hemoglobin, albumin, actin, and myosin light chain 1 were easily matched by their M_r/p*I* values. Figure 1 shows a region of many small protein spots between 60 kDa and 100 kDa at about p*I* 5.0, which also appears in the above-mentioned heart 2-DE databases. We identified two of these protein spots by PMF as myosin heavy chain, numbered No. 60 and 61 in Table 1, suggesting that many of these spots also belong to the myosin heavy chain.

For protein identification MALDI-MS PMF proved to be sensitive and rapid since 60 protein spots were investigated within two weeks. When using PMF to identify proteins, it is recommended to search in all available protein sequence databases. Seventy-two percent of the samples produced spectra from which proteins could be unequivocally identified by at least 23% sequence coverage. The database searches considered a mass accuracy of ± 0.25 Da, an M_r window of 10–100 kDa and modifications like oxidation of methionine, acetylation of the protein *N*-terminus, and formation of pyroGlu by NH_3 elimination of *N*-terminal glutamine. In contrast, *N*-terminal sequencing is laborious and slow. Also, most of the pro-



RAT HEART-2DPAGE

Two-dimensional polyacrylamide gel electrophoresis database of rat heart

Searching:

- by description
- by accession number
- by clicking on a spot

Gateways to other related servers

WORLD-2DPAGE – Index to other Federated 2-D PAGE database
ExPASy – The molecular biology server in Geneva

Notice:
The database was constructed using the Make2ddb package from the WORLD-2DPAGE.
This database was built up in cooperation between:
Max-Planck-Institute for Infection Biology
German Heart Institute Berlin
Technical University Berlin, Max-Volmer-Institute

The database is maintained by: Dr. K.-P. Pleissner (E-mail:pleiss@dhzb.de)
Last modification: August 1998

Figure 3. Homepage of the RAT HEART-2DPAGE at URL http:///gelmatching.inf.fu-berlin. de/~pleiss/2d generated by the Make2ddb software.

teins from 2-DE gels are *N*-terminally blocked or are below the detection sensitivity of *N*-terminal sequencers. Thus, *N*-terminal sequencing was only successful in 12 cases of 60 investigated proteins.

The 2-DE database was constructed on the basis of the 64 identified protein spots. Figure 3 shows the homepage of RAT HEART-2DPAGE. The data are accessible by protein description, SWISS-PROT accession number, and by clickable images. The description contains detailed information about all identified spots and links to the related SWISS-PROT database and the rat heart 2-DE database at Harefield Science Center [8]. In some cases, however, proteins were only found in the NCBI sequence database and could not be linked to the SWISS-PROT database.

Two prerequisites are evident for the creation of a 2-DE database using Make2ddb-software. (i) The package requires a Sun Microsystems workstation under the Solaris operating system; and (ii) the 2-DE map of the database has to be in the Melanie II format. Thus, the user has to consider the hardware requirements of the Make2ddb software and the usage of the Melanie II gel analysis software. It is possible to obtain the Melanie II software for Unix computers by downloading it from the Bio-Rad server URL http://www.bio.rad.com for an evaluation license limited for 30 days. A comprehensive user manual of Melanie II is directly available *via* Internet. Only four tools of Melanie II are relevant to the creation of a gel map for the database construction: (i) open a gel image, (ii) spot detection, (iii) labeling, and (iv) saving the image. In addition, the Landmark tool of Melanie II makes it possible to compute an approximated M_r/*pI* value for any point in a gel. The database creation by Make2ddb software is performed in a fast and reliable way if one considers carefully the correct format of the database text file and the installation instructions of the software package. The 2-DE database directory contains the gel map with labeled protein spots and the database text file 'rat_heart.dat' created by a text editor with the descriptive information of the labeled protein spots. The file name of the gel image has to be in coincidence with the map name in the 'existing.maps' file. If the directory structure of the Web server, 2-DE database, and Perl interpreter is assigned correctly, the creation of the database works without any problems.

Currently our 2-DE database of rat heart proteins contains more than 3000 protein spots (Fig. 1), of which 64 proteins were identified by MALDI-MS PMF, *N*-terminal se-

quencing, and by comparison with other 2-DE databases. We used the Melanie II gel analysis software for processing and evaluation of our standard rat heart 2-DE gel and the software package Make2ddb system for the database creation. According to our experience, if the Melanie II system and a Sun workstation under Solaris are available, the establishment of a 2-DE database with Make2ddb is rapid and simple. Our 2-DE database will be continually updated when new proteins are identified. It is expected to be a useful tool for elucidating proteins associated with heart failure and heart disease.

The assistance of Mrs. Carola Wenk (Free University Berlin) and Britta Seidemann (WITA GmbH) is greatly appreciated. This study was partially supported by a grant from the Deutsche Forschungsgemeinschaft (F1 165/4-1).

Received September 1, 1998

References

[1] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[2] Jungblut, P., Otto, A., Zeindl-Eberhart, E., Pleissner, K.-P., Knecht, M., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1994, *15*, 685–707.

[3] Latter, G. I., Boutell, T., Monardo, P. J., Kobayashi, R., Futcher, B., McLaughlin, C. S., Garrels, J. I., *Electrophoresis* 1995, *16*, 1170–1174.

[4] Corbett, J. M., Wheeler, C. H., Dunn, M. J., *Electrophoresis* 1995, *16*, 1524–1529.

[5] Appel, R. D., Bairoch, A., Sanchez, J.-C., Vargas, J. R., Golaz, O., Pasquali, C., Ravier, F., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 540–546.

[6] Pleißner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V., Fleck, E., *Electrophoresis* 1996, *17*, 1386–1392.

[7] Müller, E-C., Thiede, B., Zimny-Arndt, U., Scheler, C., Prehm, J., Müller-Werdan, U., Wittmann-Liebold, B., Otto, A., Jungblut, P., *Electrophoresis* 1996, *17*, 1700–1712.

[8] Evans, G., Wheeler, C. H., Corbett, J. M., Dunn, M. J., *Electrophoresis* 1997, *18*, 471–479.

[9] Lemkin, P. F., *Electrophoresis* 1997, *18*, 2759–2773.

[10] Eckerskorn, C., Jungblut, P., Mewes, W., Klose, J., Lottspeich, F., *Electrophoresis* 1988, *9*, 830–838.

[11] Jungblut, P., Dzionara, M., Klose, J., Wittmann-Liebold, B., *J. Prot. Chem.* 1992, *11*, 603–612.

[12] Vandekerckhove, J., Bauw, G., Puype, M., Van Damme, J., Van Montagu, M., *Eur. J. Biochem.* 1985, *152*, 9–19.

[13] Aebersold, R. H., Teplow, D. B., Hood, L. E., Kent, S. B. H., *J. Biol. Chem.* 1986, *261*, 4229–4238.

[14] Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., Kent, S. B. H., *Proc. Natl. Acad. Sci. USA* 1987, *84*, 6970–6974.

[15] Karas, M., Hillenkamp, F., *Anal. Chem.* 1988, *60*, 2299–2301.

[16] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., *Science* 1989, *246*, 64–71.

[17] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *79*, 4350–4354.

[18] Scheler, C., Müller, E. C., Stahl, J., Müller-Werdan, U., Salnikow, J., Jungblut, P., *Electrophoresis* 1997, *18*, 2823–2831.

[19] Jungblut, P., Thiede, B., Zimny Arndt, U., Müller, E. C., Scheler, C., Wittmann-Liebold, B., Otto, A., *Electrophoresis* 1996, *17*, 839–847.

[20] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[21] Jungblut, P., Seifert, R., *J. Biochem. Biophys. Methods* 1990, *21*, 47–58.

[22] Jungblut, P., Eckerskorn, C., Lottspeich, F., Klose, J., *Electrophoresis* 1990, *11*, 581–588.

[23] Otto, A., Thiede, B., Müller, E. C., Scheler, C., Wittmann-Liebold, B., Jungblut, P., *Electrophoresis* 1996, *17*, 1643–1650.

[24] Hoogland, C., Baujard, V., Sanchez, J.-C., Hochstrasser, D. F., Appel, R. D., *Electrophoresis* 1997, *18*, 2755–2758.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 898–906

John Weekes[1]
Colin H. Wheeler[1]
Jun X. Yan[1]
Joachim Weil[2]
Thomas Eschenhagen[2]
Günter Scholtysik[3]
Michael J. Dunn[1]

[1]Department of Cardiothoracic Surgery, National Heart and Lung Institute, Imperial College School of Medicine, Heart Science Centre, Harefield Hospital, Harefield, United Kingdom
[2]Department of Pharmacology, Pathology Institute, Eppendorf University Hospital, Hamburg, Germany
[3]Veterinary Pharmacology Institute, University of Bern, Switzerland

Bovine dilated cardiomyopathy: Proteomic analysis of an animal model of human dilated cardiomyopathy

Bovine hereditary dilated cardiomyopathy (bCMP) is endemic in Switzerland and hearts from diseased animals display important clinical and biochemical similarities to human DCM. Recent research has identified at least one protein (myoglobin) to be significantly reduced in bovine DCM. Using a proteomic approach, we have separated over 1125 protein species from bovine ventricular tissue. Gel analysis and protein characterisation have identified a number of proteins whose abundance is significantly altered in bovine DCM. Twenty-four proteins are of decreased abundance in diseased tissue, whilst 11 proteins are of increased abundance in the diseased state. A combination of amino acid compositional analysis, peptide mass profiling, *N*-terminal microsequencing and MultiIdent (http://www.expasy.ch/sprot/multiident.html) has been employed in order to elucidate the identities of the differentially expressed proteins. Using these techniques we have currently determined the identity of 12 of the 35 altered proteins. We have also detected three proteins that are differentially expressed in genotypically diseased but phenotypically normal animals, identifying a possible mechanism for the onset of the disease. The possibility that inappropriate ubiquination of proteins plays an important role in the disease is discussed. A database of bovine proteins is currently being established. The identity of the proteins affected, together with a comparison of the human and bovine expression patterns, is displayed.

Keywords: Dilated cardiomyopathy / Bovine heart / Protein expression / Heart proteins / Proteomics
EL 3371

1 Introduction

Dilated cardiomyopathy (DCM) in humans is a disease of unknown aetiology and is characterised by impaired systolic function with reduced ejection fraction and increased end-systolic blood pressure as the heart adapts to maintain a normal stroke volume [1]. To date, no specific diagnostic marker for DCM has been found and the disease can be regarded as a syndrome in which a variety of factors give rise to a common cardiac dysfunction. While a direct causal link has yet to be established, viral infections, toxic agents, chronic alcohol abuse and genetic factors have been implicated in the onset of the disease. A large number of qualitative and quantitative changes in cardiac proteins have been reported from both animal models [2, 3] and human disease samples [4], including alterations in the levels of G proteins [5], adrenergic receptors [6], calcium regulatory proteins [7–9] and contractile proteins [10, 11]. Whilst these changes are clearly of importance in explaining the contractile dysfunction in failing hearts, their relevance to the pathogenesis of DCM remains to be determined. We have recently attempted to address this problem by searching for global quantitative and qualitative changes in protein abundance in ventricular tissue from patients with DCM compared with ischaemic heart disease and undiseased tissue using two-dimensional electrophoresis (2-DE) [4]. Using this proteomic approach we have demonstrated a number of quantitative alterations in protein abundance specific to the DCM heart of humans and a fully accessible database of human heart proteins (HSC-2DPAGE) has been established (http://www.harefield.nthames.nhs.uk/nhli/protein). Due to the limitations of studies using human tissues, further models of DCM have been developed, including idiopathic DCM in turkey [12] and bovine DCM [13–15]. Whilst these studies are of value, they rely in the main upon a single defined trigger for disease onset, whereas DCM in humans is clearly of multifactorial origin. It is therefore apparent that the discovery of further animal models that better mirror human DCM will forward our understanding of the disease.

Correspondence: Dr. Michael J. Dunn, Protein Biochemistry, Heart Science Centre, Harefield Hospital, Hillend Road, Harefield, Middlesex, UB9 6JH, United Kingdom
E-mail: mike.dunn@harefield.nthames.nhs.uk
Fax:+44-1895-828900

Abbreviations: bCMP, bovine hereditary dilated cardiomyopathy; **DCM,** dilated cardiomyopathy; **GD-PN,** genotypic diseased phenotypic normal

In this study we use the techniques of proteomics to investigate a bovine model of DCM, namely bovine hereditary cardiomyopathy (bCMP). Bovine hereditary dilated cardiomyopathy was first described by Tontis *et al.* (1990) [13] and is endemic in Switzerland, occurring mainly in the Simmentaler × Red Holstein crossbreed of cattle. The disease occurs in 1- to 4-year-old cattle and progresses rapidly to global heart failure. Hearts from diseased animals display important clinical and biochemical similarities to human DCM [15, 16]. Recent research has identified significant reductions in myoglobin from left ventricles of animals with bovine DCM [14] and it appears that bovine DCM could serve as a useful model of human DCM. As the disease is thought to be inherited in an autosomally recessive fashion [17], it is possible to examine tissue from younger animals that are "genotypically diseased" but show no signs of disease development (GD-PN animals). Using 2-D gel electrophoresis, gel analysis, and protein characterisation, we have identified a number of proteins whose abundance is altered significantly in bovine DCM. We have also identified proteins that are differentially expressed in genotypically diseased but phenotypically normal animals, identifying a possible mechanism for the onset of the disease. A database of bovine proteins is currently in development. The identity of the proteins affected, together with a comparison of the human and bovine expression patterns, is presented.

2 Materials and methods

2.1 Animals and procurement of tissue

Hearts from six untreated cattle (Simmentaler × Red Holstein) with end-stage bCMP (age 1.5–4.5 years, female) were used as the diseased group. All animals showed the characteristic clinical signs of heart failure with marked peripheral and pulmonary edema, ascites and cardiomegaly. Nonfailing control hearts were obtained from the local slaughterhouse in Hamburg (n = 6, age 1.5–4 years, female). Hearts were rapidly removed within 2–4 min after killing of the animals and deblooded in ice-cold 0.9% NaCl solution. Tissue samples (2–4 g each) from the left ventricles were freed of fat and connective tissue, immediately frozen in liquid nitrogen, and stored at –80°C.

2.2 Preparation of tissue samples

All tissue samples were stored in liquid N_2 prior to processing. The frozen tissue specimens (typically 0.2–0.4 g) were ground to a fine powder under liquid N_2 using a mortar and pestle. The resulting powder was collected into 1.5 mL microcentrifuge tubes and homogenized using a handheld homogenizer for 1 min in 1 mL lysis buffer, containing 9.5 M urea, 1% DTT, 2% CHAPS, and 0.8% Pharmalyte pH 3–10 (Pharmacia, Uppsala, Sweden). After vortexing for 30 s samples were centrifuged at 15 000 rpm for 1 h and the resulting supernatants collected. Total homogenate protein concentration was measured in duplicate using a modification of the method described by Bradford [18]. Briefly, the BSA used for preparation of standard curves and the protein samples to be measured were made up to 10 μL with lysis buffer prior to the addition of 10 μL of 0.1 M HCl and 80 μL H_2O. Bradford reagent diluted 1 in 4 with H_2O was added to a final volume of 3.5 mL and the protein concentrations were determined spectrophotometrically.

2.3 Two-dimensional electrophoresis

Isoelectric focusing (IEF) was performed using immobilized pH gradient (IPG) strips (Pharmacia), which had pH ranges of 4–7 (linear) and 3–10 (nonlinear). The solubilized protein sample was applied to the strips using an in-gel rehydration method, as described by Rabilloud *et al.* [19] and Sanchez *et al.* [20]. The samples were diluted with rehydration solution containing 8 M urea, 0.5% CHAPS, 0.2% DTT and 0.2% Pharmalyte pH 3–10 prior to rehydration overnight in a reswelling tray (Pharmacia). For analytical gels, total protein loaded was 100 μg in 450 μL and preparative gels, 3 mg in 450 μL. The strips were focused at 0.05 mA/IPG strip for 60 kVh at 20°C [21]. After IEF the strips were equilibrated in 1.5 M Tris, pH 8.8, buffer containing 6 M urea, 30% glycerol, 2% SDS and 0.01% bromophenol blue, with the addition of 1% DTT for 15 min, followed by the same buffer with the addition of 4.8% iodoacetamide for 15 min [22]. SDS-PAGE was performed using 12%T, 2.6%C separating polyacrylamide gels without a stacking gel using a Hoefer DALT system [23]. The second-dimension separation was carried out overnight at 20 mA/gel at 8°C and was stopped as the bromophenol blue dye front just left the bottom of the gels.

2.4 Protein visualisation

Analytical gels were fixed after electrophoresis for between 40 min and overnight in 50% methanol, 10% acetic acid in glass dishes. Gels were silver-stained using the Daiichi 2-D Silver Staining Kit (Insight Biotechnology, Wembley, UK). Micropreparative gels were stained using 0.1% colloidal Coomassie Brilliant Blue G-250 [24] in 2% phosphoric acid, 10% ammonium sulphate and 20% methanol. Gels were placed into covered glass dishes and adequate staining was achieved after 48 h. Gels were washed with distilled water to remove any surface-bound dye particles and either used immediately or placed individually in sealed plastic bags.

2.5 Densitometry and computer analysis

All silver- and Coomassie-stained gels were scanned at 100 μm resolution using a Molecular Dynamics Personal SI Laser Densitometer (Sunnyvale, CA, USA). Western blots were scanned using a GS-710 optically enhanced imaging densitometer (Bio-Rad, Richmond, CA, USA). Gels were analysed using Melanie II image analysis software [25, 26] running on a Sun Sparc Ultra I workstation. After detection of spots the gels were aligned, landmarked and matched. Gels were then placed into the appropriate experimental classes and differential analysis was performed as described [25]. All spots that differed between the classes by 50% or more and passed a Student T-test ($p < 0.05$) were accepted as significant differences. All gel spots detected as significantly different between the groups were then highlighted and checked manually to eliminate any artefactual differences due to gel pattern distortions, abnormal silver staining and inappropriately matched or badly detected spots. For simplicity the optical density values for each spot were averaged across each group compared in Table 1.

2.6 Protein blotting

Immediately after electrophoresis, micropreparative gels were equilibrated for 15 min in 50 mM Tris, 50 mM boric acid buffer, pH 8.5. Semidry blotting of the gels onto PVDF membranes (FluoroTrans) was achieved using a Multiphor II Nova-Blot (Pharmacia) with a current of 0.8 mA/cm^2 for 90 min. The blots were then stained using a 0.05% solution of Coomassie Brilliant Blue R-250 in 50% methanol, 10% acetic acid. Blots were destained by washing for 2 × 5 min with 50% methanol, 10% acetic acid, followed by rinsing twice in double distilled H_2O. After drying between sheets of filter paper, blots were either used immediately or sealed individually in plastic bags for storage.

2.7 *N*-terminal protein microsequencing

N-terminal protein microsequencing was carried out as previously described [27]. Briefly, protein spots were excised from the Coomassie-stained PVDF membranes after Western blotting and destained in 50% v/v methanol. After drying, individual spots were sequenced using an ABI model 477A pulsed liquid protein sequencer equipped with an ABI model 120A PTH analyser. The sequence data produced was used to search the OWL nonredundant protein sequence database [28] using the BLAST algorithm (http://www.seqnet.dl.ac.uk/dbsearch.html#BLAST).

Table 1. Left ventricular proteins significantly altered in bovine-dilated cardiomyopathy[a)]

Spot number	$M_r(10^3)$	p*I*	Diseased ($n = 6$)	Control ($n = 6$)	GD-PN ($n = 4$)
Proteins decreased in diseased left ventricle					
80	66	5.9	19.6	75	68
333	49	6.1	89	184	159
372	50	5.7	120	369	458
379	46	6.2	11.4	24	20.5
401	45	6.7	71	159	175
428	44	6.5	31	54	32
443	51	6.38	22	122	117
538	37	5.7	14	28	22
551	41.7	6.6	53	143	123
556	36.5	5.9	35	61	83
557	36	5.8	123	190	226
574	35	6.3	100	158	169
649	34.7	5.8	69	192	182
658	32	5.4	27	80	96
815	26.3	6.4	113	384	278
831	26	5.2	164	292	340
842	25	6	21	44	35
854	24.7	5.1	86	176	241
908	17	4.3	667	1275	1230
912	16	6.7	363	764	447
919	15.4	6.8	295	718	700
920	15	6.6	408	827	833
924	15	6.6	1065	1664	1789
943	13	6.1	102	235	44
Proteins increased in diseased left ventricle					
274	60	4.6	207	99	102
279	54	5	357	132	160
281	54	5.3	143	64,5	98
327	56	5.7	365	157	125
408	42	6	902	390	516
712	31.8	6	195	41	46
713	28	6.7	117	53	56
749	28.7	5.8	102	31	27
762	27.9	6.7	268	139	138
794	26	6.1	814	284	333
795	26.9	5.1	200	28	16

a) Protein levels are mean values from silver-stained gels. Spot numbers refer to numbers in Fig. 1.

2.8 MALDI-MS peptide mass fingerprinting

Protein spots were excised from 2-DE gels stained with colloidal Coomassie Brilliant Blue G-250, cut into 1 mm cubes, and destained by washing in 20 μL aliquots of 50 mM ammonium bicarbonate in 30% v/v acetonitrile; this was repeated for 1 h (each time) until colourless. The samples were then dried in a centrifugal evaporator. Modified (methylated) porcine trypsin (Promega, Madison, WI, USA) was prepared as a stock solution in water

(0.1 μg/μL). For digestion, 4 μL of trypsin solution were added to 21 μL Tris buffer (5 mM, pH 8.8, prepared fresh for each use) and added to the gel pieces before incubation overnight at room temperature. Digestion was stopped by addition of 15 μL of 50% acetonitrile, 0.1% TFA. Tubes were then sonicated in a water bath for 10 min to extract peptides immediately before spotting onto MALDI targets. MALDI mass spectrometry was performed using a Voyager Elite mass spectrometer (PE Biosystems, Framingham, MA, USA). Two microlitre aliquots of sample were spotted and dried onto the MALDI targets followed by addition and drying of 2 μL of matrix α-cyano-hydroxycinnamic acid, 4 mg/mL). Data were averaged for 50 laser shots by delayed extraction in positive ion mode using the reflectron with an accelerator voltage of 21 kV. Peptide masses were used to search the SWISS-PROT protein database using the PeptidIdent tool on ExPASy (http://www.expasy.ch/sprot/peptidident.html). Searching was performed using a mass uncertainty of ±1 Da and a molecular weight range of ±20% of the M_r and ±1 p*I* unit determined from 2-DE, and the output consisted of a list of proteins ranked by a statistical score.

2.9 Amino acid compositional analysis

Amino acid analysis of PVDF-bound proteins was performed as described previously [29]. Single protein spots were excised from the PVDF membrane and hydrolysed in 6 M HCl at 112°C for 18 h using a Pico-Tag workstation (Waters, Watford, UK). Amino acid composition was determined using 9-fluorenylmethoxycarbonyl (Fmoc) precolumn derivatisation on a GBC Automated Aminomate System (GBC Scientific Equipment, Dandenong, Victoria, Australia) [30] and chromatography performed using the method of Yan *et al.* [16]. Percentage amino acid composition was determined for 16 recovered amino acids by comparing the pmol yield of each amino acid to the total pmol yield of all amino acids. Amino acid composition and estimated M_r and p*I* from 2-DE for each protein were used to search the SWISS-PROT protein database (Release 35) using the programme AACompID [31, 32] accessed *via* the WWW (http://WWW.expasy.ch/ch2d/aacompi.html). Searches were made using constellation 2 (Cys, Gly, Trp not included) or constellation 4 (Cys, Lys, Trp not included) with windows of p*I* ±1 unit and M_r ±20%. Analyses were calibrated using a human serum albumin sample that had been hydrolysed, extracted and analysed at the same time as the other spots.

2.10 Database searching using MultiIdent

If searches using PeptidIdent or AACompID alone failed to identify the protein spots, peptide fingerprinting and amino acid analysis data were used, in combination, to search the SWISS-PROT database using MultiIdent (http://www.expasy.ch/ch2d/multiident.html) using the same search parameters as for the individual searches above.

3 Results

3.1 Protein changes in left ventricle tissue

Figure 1 shows the pattern of proteins from normal bovine left ventricular tissue separated by 2-D gel electrophoresis using a linear IPG IEF gradient of pH 4–7. On average a total of 1125 protein spots was observed in normal ventricle patterns and 1191 protein spots were seen in diseased ventricle patterns. No significant difference in the total number of spots was observed in the GD-PN ventricles when compared to either the normal or diseased tissue. The patterns of protein separation were highly reproducible and did not differ significantly in at least three different separation runs. When the 2-DE patterns of proteins from the normal and diseased tissues are compared, it is clear that the expression of the vast majority (98%) of proteins remains unaltered. However, we have identified a group of 35 proteins, the abundance of which is significantly altered between normal and diseased ventricles. Of this group, the majority of spots (24) are of decreased abundance in the ventricle of diseased animals (indicated in red in Fig. 1). Eleven protein spots were of increased abundance in diseased tissue (indicated in blue in Fig. 1). There were no spots that were unique to either the normal, diseased or GD-PN patterns. Two of the most striking alterations in protein abundance are highlighted in greater detail in Fig. 2 a, b. A summary of the spot number, M_r, p*I* and the abundance of each altered protein in both normal, diseased and GD-PM tissues is shown in Table 1. There does not appear to be any pattern in the distribution of the altered spots over the gels, and proteins of both high and low abundance are equally likely to be changed. The pattern of differential protein expression observed between normal and diseased tissue is conserved for the majority of spots (91%) when the diseased and GD-PN patterns are compared. In fact, the 2-D PAGE patterns of normal and GD-PN tissue are almost identical. There are, however, three notable exceptions to this trend, namely spots 428, 912 and 943. These spots are of decreased abundance in both the diseased and the GD-PN patterns relative to the normal group. No other significant differences between the GD-PN and the normal protein patterns were observed.

3.2 Protein identification

A combination of amino acid compositional analysis, peptide mass profiling, *N*-terminal microsequencing and MultiIdent (http://www.expasy.ch/sprot/multiident.html) was

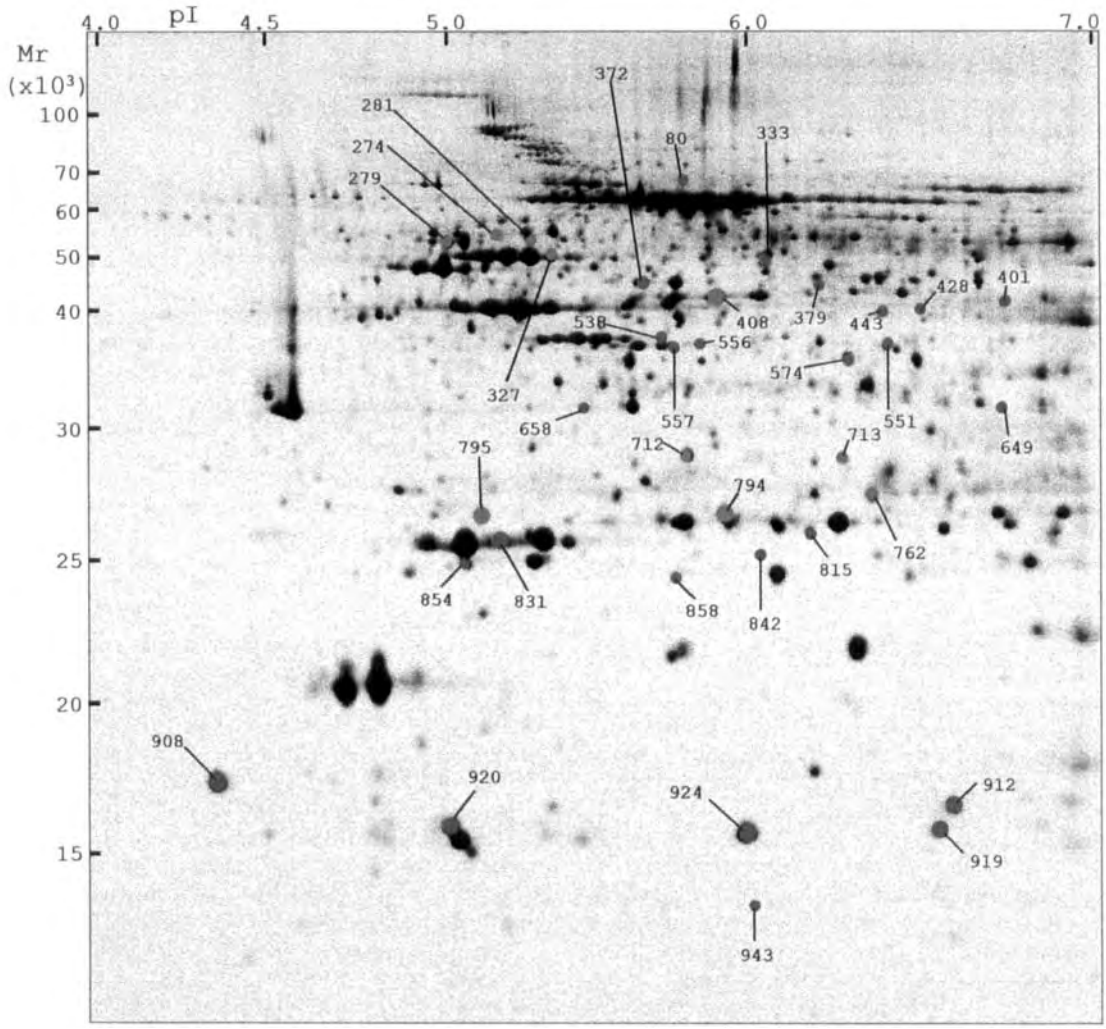


Figure 1. Standard pattern of expression of bovine left ventricle proteins separated by 2-DE using a linear pH 4–7 IPG strip in the first dimension. Proteins were detected by silver staining. Calibration of isoelectric point and molecular mass was performed using Melanie II and theoretical values from identified proteins. Spot numbers refer to numbers in Tables 1 and 2. Spots showing significant ($p < 0.05$) differences in disease compared to normals are highlighted. Spots coloured red are decreased in disease, spots coloured blue are increased in disease.

used to elucidate the identities of the differentially expressed proteins. Using these techniques we have currently determined the identity of 12 of the 35 altered proteins. *N*-terminal microsequencing successfully identified five protein spots. Three proteins were *N*-terminally blocked and one sequence did not match any protein sequence in publicly available databases. Six spots were identified using either amino acid compositional analysis, MultiIdent or peptide mass profiling alone, while one spot required amino acid compositional analysis and MultiIdent for identification. For the one protein identified using peptide mass profiling alone, 15 peptide matches were produced using PeptideIdent, with sequence coverage of 45.8%. A summary of this data is shown in Table 2. For each spot number, the M_r, p*I* and protein identity is shown, together with the SWISS-PROT identifier of either the bovine protein or the closest homologue. The method of protein identification used and the alteration in disease is also shown in Table 2. A number of proteins that showed no difference in abundance between normal and

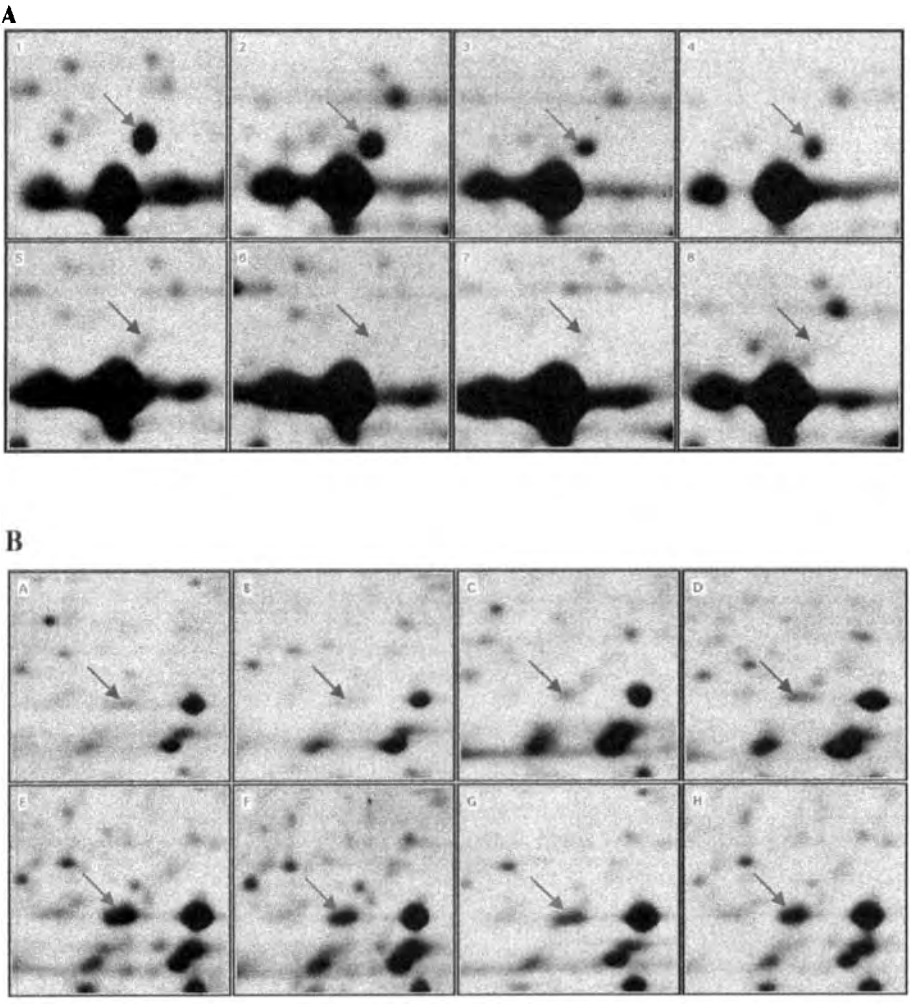


Figure 2. Close-ups of areas of the gel shown in Fig. 1 showing a spot (a) increasing and (b) decreasing in disease compared to normal patterns. The upper set of (a) and (b) represents the spots in the diseased state. The identity of the protein in (a) is ubiquitin C-terminal hydrolase (spot 795). The protein spot shown in (b) is α-I antiproteinase precursor (spot 372).

diseased tissue were also identified. This data is not shown in this publication. Work is continuing to characterise the remaining 23 differentially expressed proteins. Two of the most striking differences in protein abundance between diseased and normal tissue are seen with spots 372 and 795. These spots were identified as α-1-antiproteinase precursor protein and ubiquitin C-terminal transferase, respectively. Both spots are highlighted in greater detail in Fig. 2 a, b. Unfortunately, identification of the three spots observed to decrease in both the diseased and GD-PN groups relative to normal groups was not successful using any of the above techniques. This was due mainly to the low abundance of these proteins.

4 Discussion

This study represents the first in-depth analysis of global protein alterations in any bovine tissue. Using a proteomic approach we have demonstrated that the abundance of a number of left ventricular proteins is significantly altered in bCMP. A combination of protein characterisation techniques has revealed the identities of a number of these

Table 2. Identification of left ventricular proteins significantly altered in bovine-dilated cardiomyopathy

Spot number	M_r (10^3)	p*I*	Consensus protein identity	SWISS-PROT identifier	Identification method[a)]	Change in disease[b)]
80	66	5.9	Mitochondrial stress-70 protein precursor	GR75_HUMAN	4	▼
274	60	4.6	Alpha-1-antiproteinase precursor	A1AT_BOVIN	4	▲
327	56	5.7	Desmin	DESM_BOVIN	2	▲
372	50	5.7	Alpha-1-antiproteinase precursor	A1AT_RABIT	1	▼
379	49	6.2	Ubiquinol cytochrome *c* reductase complex core protein 1	UCR1_BOVIN	4	▼
443	51	6.38	Isovaleryl CoA dehydrogenase	IVD_RAT	1	▼
551	41.7	6.6	Isocitrate dehydrogenase	IDHA_BOVIN	3	▼
649	34.7	5.8	Blocked		4	▼
712	31.8	6.0	Not identified		4	▼
749	28.7	5.8	Blocked		4	▲
795	26.8	5.1	Ubiquitin carboxyl terminal hydrolase	UBL1_HUMAN	4	▲
815	26.3	6.4	Mitochondrial thioredoxin-dependent peroxide reductase	TDXM_BOVIN	4	▼
854	24.7	5.1	Blocked		4	▼
919	15.4	6.8	Fatty acid binding protein, heart	FABH_BOVIN	1.3	▼
920	15	6.6	Cytochrome *c* oxidase polypeptide VA	COXA_BOVIN	1	▼
924	15	6.0	Fatty acid binding protein, heart	FABH_BOVIN	1	▼

a) Key to protein identification methods: 1, amino acid compositional analysis; 2, peptide mass profiling; 3, MultiIdent (http://www.expasy.ch/sprot/multiident.html); 4, *N*-terminal microsequencing,

b) ▼ Decrease in disease; ▲ increase in disease

proteins and work to characterise the remaining proteins is currently underway. At least three proteins have been detected that are reduced in both the diseased and GD-PN hearts relative to the control group.

4.1 Comparison of bCMP and normal protein expression

Left ventricle tissue samples from a total of six animals showing the typical pathophysiological symptoms of bCMP at slaughter were compared with left ventricle tissue from six normal animals using 2-DE and computer-aided analysis of protein expression patterns. This approach has identified significant changes in the levels of 35 proteins in bCMP. Of these proteins a total of 24 were of decreased abundance and 11 were of increased abundance. When analysing the expression patterns using the Melanie II software, spots that differed between the groups by more than 50% were chosen. The actual degree of change between the spots, however, was often markedly greater than this. Examination of the results shown in Table 1 shows that only 9 of the 35 proteins showed less than 2-fold differences between the groups. Indeed, as is the case for 12 of the spots, these differences approach or exceed 3-fold. The greatest differences are seen for spots 443, 712 and 795, which show changes of 5.5-, 4.8- and 7.1-fold, respectively.

4.2 Significance of the GD-PN group

Careful analysis of the GD-PN patterns relative to the normal and diseased groups has revealed that the overwhelming majority of proteins remain unchanged. However, three spots have been identified that show identical changes to those seen in the disease group. These spots are of interest as they occur in animals that, whilst showing none of the symptoms of bCMP at the time of slaughter, would presumably have gone on to develop the disease in time. It is therefore tempting to suggest that the changes in abundance of these proteins may represent the initial biochemical abnormality which, through as yet undescribed mechanisms, leads to the gross changes in cardiac physiology and function seen in bCMP. As previously mentioned, the current work has not yet yielded the identities of these proteins, due mainly to the low abundance of these proteins in 2-DE separations. Work to characterise these proteins by pooling multiple gels is ongoing. As bCMP is a hereditary disease, thought to be inherited in an autosomally recessive manner [17], it may be possible to isolate a group of GD-PN cattle and follow the expression of ventricular proteins as the disease develops and progresses, using sequential biopsies. A control group of normal cattle of the same age group would be required to eliminate any artefactual developmental changes in protein abundance. It would also be of interest to examine ventricular tissue by 2-DE techniques from the relatively small number of cases of human familial DCM that occur each year.

4.3 Significance of protein alterations

The majority of altered proteins identified to date in this study are of decreased abundance in diseased tissue. This is in agreement with previous animal [2, 3] and

human [4] studies. It has previously been suggested that increased proteolysis may at least in part be responsible for these findings. The role of calcium-activated neutral proteases (caspases) has been widely established in myopathies of skeletal muscle [33–37], although their function in cardiac myopathies has yet to be determined. The greatest alteration in protein abundance found in this study is that of the enzyme ubiquitin *C*-terminal hydrolase, which is increased by over 7-fold in diseased ventricles. This enzyme is thought to release free ubiquitin from poly-ubiquinated proteins, thereby increasing the pool of free ubiquitin in the cytoplasm. It is possible that such a large increase in ubiquitin *C*-terminal hydrolase could increase the cellular concentration of ubiquitin to such an extent that inappropriate ubiquination of proteins occurs, leading to the subsequent proteolysis of such proteins by the 26S proteasome [38, 39]. Indeed, it has recently been suggested that inappropriate ubiquination of proteins may have a contributory role in the development of heart failure [40]. The dramatic increase in ubiquitin *C*-terminal hydrolase demonstrated in this study represents the first experimental evidence that increased or inappropriate ubiquitin-directed proteolysis may contribute towards heart disease.

Many of the proteins found to be altered in this study are found exclusively in the mitochondria. This is consistent with the suggestion that in congestive heart failure the myocardium is unable to provide enough energy to cope with the increased mechanical stresses [41]. Many mitochondrial proteins were also found to be of lower abundance in the ventricles of dogs after pacing-induced heart failure [2]. Indeed, the levels of mitochondrial isocitrate dehydrogenase were significantly lower in both this study and in the paced dog ventricle.

4.4 Conclusion

Bovine-dilated cardiomyopathy results in the altered abundance of a number of proteins in the left ventricle. The onset and progression of this disease may possibly be caused by increased proteolysis involving the ubiquitin-proteasome pathway. The identification of the remaining altered proteins will contribute to the understanding of the biochemical mechanisms in this disease and also DCM of the human heart.

We would like to thank the British Heart Foundation for financial support, Professor Denis Hochstrasser (University Hospital, Geneva) for allowing us access to the mass spectrometry facility and Professor Claude Gaillard (Institute of Animal Breeding, University of Berne) for providing us with the diseased hearts.

Received January 9, 1999

5 References

[1] Dec, G. W., Fuster, V., *New Engl. J. Med.* 1994, *331*, 1564–1575.

[2] Heinke, M. Y., Wheeler, C. H., Chang, D., Einstein, R., Drake-Holland, A., Dunn, M. J., dos Remedios, C. G., *Electrophoresis* 1998, *19*, 2021–2030.

[3] Whipple, G. H., Sheffield, L. T., Woodman, E. G., Theophilis, E. G., Freidman, S., *New Engl. Cardiovasc. Soc.* 1962, *20*, 39–43.

[4] Corbett, J. M., Why, H. J., Wheeler, C. H., Richardson, P. J., Archard, L. C., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1998, *19*, 2031–2042.

[5] Bohm, M., Gierschik, P., Jakobs, K. H., Pieske, B., Schnabel, P., Ungerer, M., Erdmann, E., *Circulation* 1990, *82*, 1249–1265.

[6] Steinfath, M.,, Lavicky, J., Schmitz, W., Scholz, H., Doring, V., Kalmar, P., *Eur. J. Clin. Pharmacol.* 1992, *42*, 601–611.

[7] Rasmussen, R. P., Minobe, W., Bristow, M. R., *Biochem. Pharmacol.* 1990, *39*, 691–696.

[8] Mercadier, J. J., Lompre, A. M., Duc, P., Boheler, K. R., Fraysse, J. B., Wisnewsky, C., Alled, P. D., Komajda, M., Schwartz, K., *J. Clin. Invest.* 1990, *85*, 305–309.

[9] Arai, M., Alpert, N. R., MacLennan, D. H., Barton, P., Periasamy, M., *Circ. Res.* 1993, *72*, 463–469.

[10] Schaub, M. C., Hirzel, H. O., *Basic Res. Cardiol.* 1987, *82 Suppl. 2*, 357–367.

[11] Mercadier, J. J., Bouveret, P., Gorza, L., Schiafino, S., Clark, W. A., Zak, R., Swynghedauw, B., Schwartz, K., *Circ. Res.* 1983, *53*, 52–62.

[12] Gwathmey, J. K., Davidoff, A. J., *Curr. Opin. Cardiol.* 1993, *8*, 480–495.

[13] Tontis, A., Zwahlen, R., Lobsiger, C., Luginbuhl, H., *Schweiz. Arch. Tierheilk.* 1990, *132*, 105–116.

[14] Weil, J., Eschenhagen, T., Magnussen, O., Mittmann, C., Orthey, E., Scholz, H., Schafer, H., Scholtysik, G., *J. Mol. Cell. Cardiol.* 1997, *29*, 743–751.

[15] Eschenhagen, T., Deiderich, M., Kluge, S. H., Magnussen, O., Mene, U., Müller, F., Schmitz, W., Scholz, H., Weil, J., Sent, U., Schaad, A., Scholatysik, G., Wuthrich, A., Gaillard, C., *J. Mol. Cell. Cardiol.* 1995, *27*, 357–370.

[16] Tschudi, P., Martig, J., *J. Vet. Med.* 1989, *36*, 612–620.

[17] Dolf, G., Stricker, C., Tontis, A., Martig, J., Gaillard, C., *J. Anim. Sci.* 1998, *76*, 1824–1829.

[18] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[19] Rabilloud, T., Valette, C., Lawrence, J. J., *Electrophoresis* 1994, *15*, 1552–1558.

[20] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[21] Görg, A., Boguth, G., Obermaier, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16*, 1079–1086.

[22] Görg, A., Postel, W., Weser, J., Günther, S., Strahler, J. R., Hanash, S. M., Somerlot, L., *Electrophoresis* 1987, *8*, 122–124.

[23] Anderson, N. L., *Two-Dimensional Electrophoresis: Operation of the ISO-DALT System*, Large Scale Biology Press, Washington DC 1988.

[24] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., *Electrophoresis* 1988, *9*, 255–267.

[25] Appel, R. D., Palagi, P. M., Walther, D., Vargas, R. J., Sanchez, J.-C., Ravier, F., Pasquali, C., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2724–2734.

[26] Appel, R. D., Vargas, J. R., Palagi, P. M., Walther, D., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2735–2748.

[27] Corbett, J. M., Wheeler, C. H., Baker, C. S., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1994, *15*, 1459–1465.

[28] Bleasby, A. J., Akrigg, D., Atwood, T. K., *Nucleic Acids Res.* 1994, *22*, 3574–3577.

[29] Wheeler, C. H., Berry, S. L., Wilkins, M. R., Corbett, J. M., Ou, K., Gooley, A. A., Humphery-Smith, I., Williams, K. L., Dunn, M. J., *Electrophoresis* 1996, *17*, 580–587.

[30] Yan, J. X., Wilkins, M. R., Ou, K., Gooley, A. A., Williams, K. L., Sanchez, J.-C., Golaz, O., Pasquali, C., Hochstrasser, D. F., *J. Chromatogr. A* 1996, *736*, 291–302.

[31] Wilkins, M. R., Ou, K., Appel, R. D., Sanchez, J.-C., Yan, J. X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D. F., Williams, K. L., Gooley, A. A., *Biochem. Biophys. Res. Commun.* 1996, *221*, 609–613.

[32] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *BioTechnology* 1996, *14*, 61–65.

[33] Toth, L., Karcsu, S., Poberai, M., Savay, G., *Acta Histochem.* 1983, *72*, 71–75.

[34] Sohar, I., Nagy, I., Heiner, L., Kovacs, Z., Guba, F., *Acta Phisiol. Acad. Sci. Hung.* 1982, *60*, 43–51.

[35] Clark, A. F., Vignos, Jr., P. J., *Muscle Nerve* 1981, *4*, 219–222.

[36] Leonard, J. P., Salpeter, M. M., *J. Cell. Biol.* 1979, *82*, 811–819.

[37] Fagan, J. M., Wajnberg, E. F., Culbert, L., Waxman, L., *Am. J. Physiol.* 1992, *262*, 637–643.

[38] Ciechhanover, A., Schwartz, A., *FASEB J.* 1994, *8*, 182–191.

[39] Goldberg, A. L., *Science* 1995, *268*, 522–523.

[40] Field, M. L., Clark, J. F., *Cardiovasc. Res.* 1997, *33*, 8–12.

[41] Katz, A. M., *Am. J. Cardiol.* 1989, *63*, 12A–16A.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 907–916

Hanno Langen[1]
Peter Berndt[1]
Daniel Röder[1]
Nigel Cairns[2]
Gert Lubec[3]
Michael Fountoulakis[1]

[1]F. Hoffmann-La Roche, Pharmaceutical Research-Gene Technology, Basel, Switzerland
[2]Institute of Psychiatry, Brain Bank, London, UK
[3]University of Vienna, Department of Pediatrics, Vienna, Austria

Two-dimensional map of human brain proteins

Samples of human brain from the parietal cortex lobe were analyzed by two-dimensional gel electrophoresis, using immobilized pH gradient strips covering the various pH regions. The protein spots were visualized with colloidal Coomassie blue stain and identified by matrix-assisted laser desorption/ionization mass spectrometry. Approximately 400 spots were identified, corresponding to 180 different brain proteins. The list of identified proteins includes a large number of structural proteins and of enzymes or enzyme subunits with various catalytic activities. The majority of proteins are localized in the cytoplasma and in mitochondria. The two-dimensional map may be useful as a reference database to study changes in the protein level caused by various disorders, such as Alzheimer's disease, major depression and schizophrenia.

Keywords: Brain / Two-dimensional electrophoresis / Protein map / Differential protein expression / Matrix-assisted laser desorption/ionization mass spectrometry EL 3345

1 Introduction

Two-dimensional (2-D) gel electrophoresis is a powerful tool for proteome analysis, providing us with valuable information on differential protein expression. It finds wide application in clinical diagnosis to study protein expression levels in healthy and diseased states [1, 2]. Such studies can be facilitated by comparison with master gels of existing 2-D databases. Several partial 2-D databases are available today, including protein maps of human plasma [3], cerebrospinal fluid [4], platelets [5], liver [6], heart [7, 8], as well of various microorganisms, such as *Escherichia coli* [9, 10], *Hemophilus influenzae* [11–17], *Saccharomyces cerevisiae* [18, 19], *Bacillus subtilis* [20], and several cell lines. Most of the 2-D databases are available in an electronic form as well, for example from the ExPASy server, accessible *via* the WorldWideWeb (http://expasy.hcuge.ch/ch2d/) [21, 22]. The investigation of the protein expression levels in the brain of patients suffering from various diseases of the central nervous system, such as Down syndrome, Alzheimer's disease and schizophrenia, in comparison with the normal brain, can yield important information about the nature of the disorder. However, samples of human brain tissue are difficult to obtain and to analyze. Therefore, except for a partial 2-D map for proteins from neuroblastoma cells [23], no database for brain proteins is available. In the present study, we tried to fill this gap by constructing a 2-D reference map of human brain proteins, which may be useful in the investigation of neurological disorders.

Correspondence: Dr. Michael Fountoulakis, F. Hoffmann-La Roche, Pharmaceutical Research-Gene Technology, Building 93-444, CH-4070 Basel, Switzerland
E-mail: michael.fountoulakis@roche.com
Fax: 41-61-691-9391

2 Materials and methods

2.1 Materials

Immobilized pH gradient (IPG) strips were purchased from Amersham Pharmacia Biotechnology (Uppsala, Sweden). Acrylamide was obtained from Serva (Heidelberg, Germany) and the other reagents for the polyacrylamide gel preparation were from Bio-Rad (Hercules, CA, USA). Carrier ampholytes (Resolyte 3.5–10) were purchased from BDH Laboratory Supplies (Poole, UK). CHAPS and thiourea were from Sigma (St. Louis, MO, USA). Urea, dithioerythritol and EDTA were obtained from Merck (Darmstadt, Germany). The brain samples were previously characterized [24]. In short, postmortem brain samples were obtained from the MRC London Brain Bank for Neurodegenerative Diseases, Institute of Psychiatry. They were taken from parietal cortex of an aged control (63 years old) with no history of neurological or psychiatric illness. The cause of death was heart disease. Postmortem interval of brain dissection was 26 h. Tissue samples were stored at –70°C and the freezing chain was never interrupted.

2.2 Two-dimensional gel electrophoresis

Brain tissue was suspended in 0.5 mL of sample buffer consisting of 40 mM Tris, 7 M urea, 2 M thiourea, 4% CHAPS, 10 mM 1,4-dithioerythritol, 1 mM EDTA and a mixture of protease inhibitors, 1 mM PMSF and 1 μg/mL of each pepstatin A, chymostatin, leupeptin and antipain. The suspension was sonicated for approximately 30 s and centrifuged at 10 000 × *g* for 10 min to sediment cell debris. The supernatant was centrifuged further at 150 000 × *g* for 45 min to sediment undissolved material. The protein content in the supernatant was determined by the Coomassie blue method [25]. The protein concentra-

tion was approximately 8 mg/mL. 2-D gel electrophoresis was performed essentially as reported [26]. Samples of approximately 1.5 mg were applied on immobilized pH 3–10 nonlinear or pH 4–7 and pH 6–11 linear gradient strips at both the basic and acidic ends of the strips. The proteins were focused at 300 V for 1 h, after which the voltage was gradually increased to 3500 V within 6 h. Focusing was continued at 3500 V for 12 h and at 5000 V for 24 h. The second-dimensional separation was performed on 9–16% linear gradient polyacrylamide gels. After protein fixing with 40% methanol containing 5% phosphoric acid for 12 h, the gels were stained with colloidal Coomassie blue (Novex, San Diego, CA, USA) for 48 h. The molecular mass was determined by running standard protein markers at the right side of selected gels. The size markers (Gibco, Basel, Switzerland) covered the range 10–200 kDa. p*I* values were used as given by the supplier of the IPG strips. The gels were destained with H_2O and scanned in a Molecular Dynamics Personal densitometer. The images were processed using Photoshop (Adobe) and PowerPoint (Microsoft) software. Protein spots were quantified using the ImageMaster 2D Elite software (Amersham Pharmacia Biotechnology).

2.3 MALDI-MS

MALDI-MS analysis was performed as described [27] with minor modifications. Briefly, spots were excised, destained with 50% acetonitrile in 0.1 M ammonium bicarbonate and dried in a Speedvac evaporator. The dried gel pieces were reswollen with 3 μL of 3 mM Tris-HCl, pH 8.8, containing 50 ng trypsin (Promega, Madison, WI, USA) or 100 ng endoproteinase Lys-C (Wako, Neuss, Germany). After 15 min, 3 μL of H_2O were added and left at room temperature for about 12 h. Two μL of 30% acetonitrile, containing 0.1% trifluoroacetic acid, was added; the content was vortexed, centrifuged for 3 min and sonicated for 5 min. One μL was applied onto the dried matrix spot. The matrix consisted of 15 mg nitrocellulose (Bio-Rad) and 20 mg α-cyano-4-hydroxycinnamic acid (Sigma) dissolved in 1 mL acetone:isopropanol 1:1 v/v. A 0.5 μL portion of the matrix solution was applied on the sample target. The digest mixtures were analyzed in a time-of-flight PerSeptive Biosystems mass spectrometer (Voyager Elite, Cambridge, MA, USA) equipped with a reflectron. An accelerating voltage of 20 kV was used. Calibration was internal to the samples. The peptide masses were matched with the theoretical peptide masses of all proteins from all species of the SWISS-PROT database. For protein search, monoisotopic masses were used and a mass tolerance of 0.0075% was allowed. Missed cleavage sites or unmatched peptides were not considered. The protein search was performed with in-house developed software (U. Hobohm, unpublished results), which is similar to the PeptIdent software on the ExPASy server (http://expasy.hcuge.ch/sprot/peptident.html).

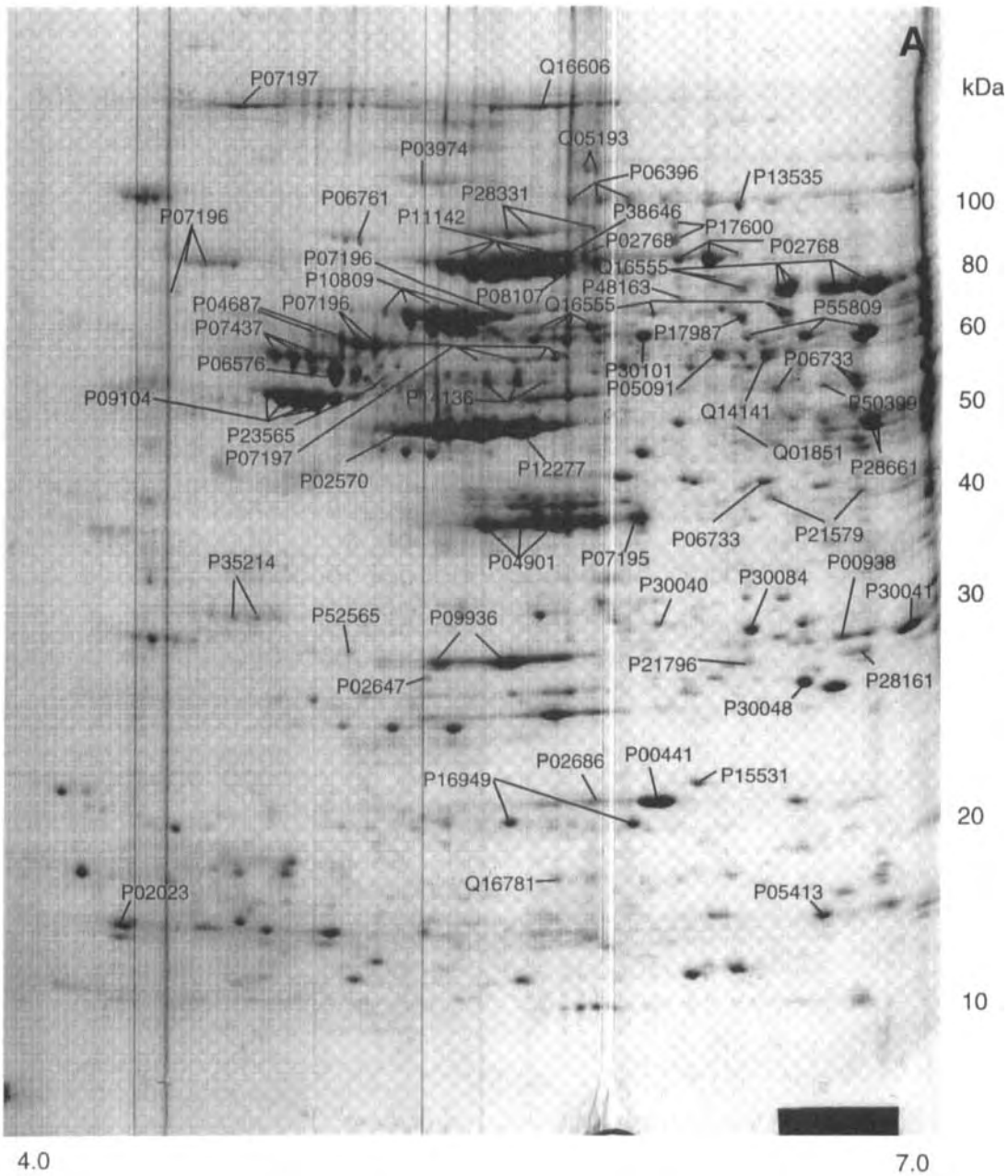
3 Results

Human brain extracts from the parietal region of cortex were separated by 2-D gels and the protein spots were visualized following staining with colloidal Coomassie blue. Figure 1 shows a representative example of separation by 2-D gels of brain proteins. The 2D ImageMaster software, following the staining with Coomassie blue, detected approximately 1000–1500 protein spots per gel. The protein spots were analyzed by MALDI-MS, following in-gel digestion with trypsin or endoproteinase Lys-C. The peptide masses were matched with the theoretical peptide masses of all proteins from all species of the SWISS-PROT database. Approximately 180 proteins were identified (Table 1). Table 1 gives the SWISS-PROT accession numbers, the abbreviated names and the full names of the proteins. It also includes the theoretical M_r and p*I* values and the approximate observed values, deduced from the 2-D gel. The confidence of identification is indicated by the number of matching and total peptides and the protein amino acid sequence coverage by the matching peptides. A minimal number of five matching peptides was used for protein search in the SWISS-PROT database. In most cases, a larger number of matching peptides was found (Table 1). Only in two cases, concerning proteins of low molecular mass, which usually deliver few peptides [16], the identification was based on four matching peptides. However, these were the major peptides and their amino acid sequence covered 35% or more of the sequence of the identified proteins. In Fig. 1, the proteins identified are labeled with their SWISS-PROT accession numbers. The protein search considered members from all species and therefore certain proteins were identified from other species, as their human homologs were probably not listed in the database. In many other cases, although a large number of peptides were recovered from the enzymatic digestion of the protein spots, probably sufficient for a confident identification, no protein could be assigned. This is most likely because neither the human nor a highly homologous counterpart from other species was included in the database.

4 Discussion

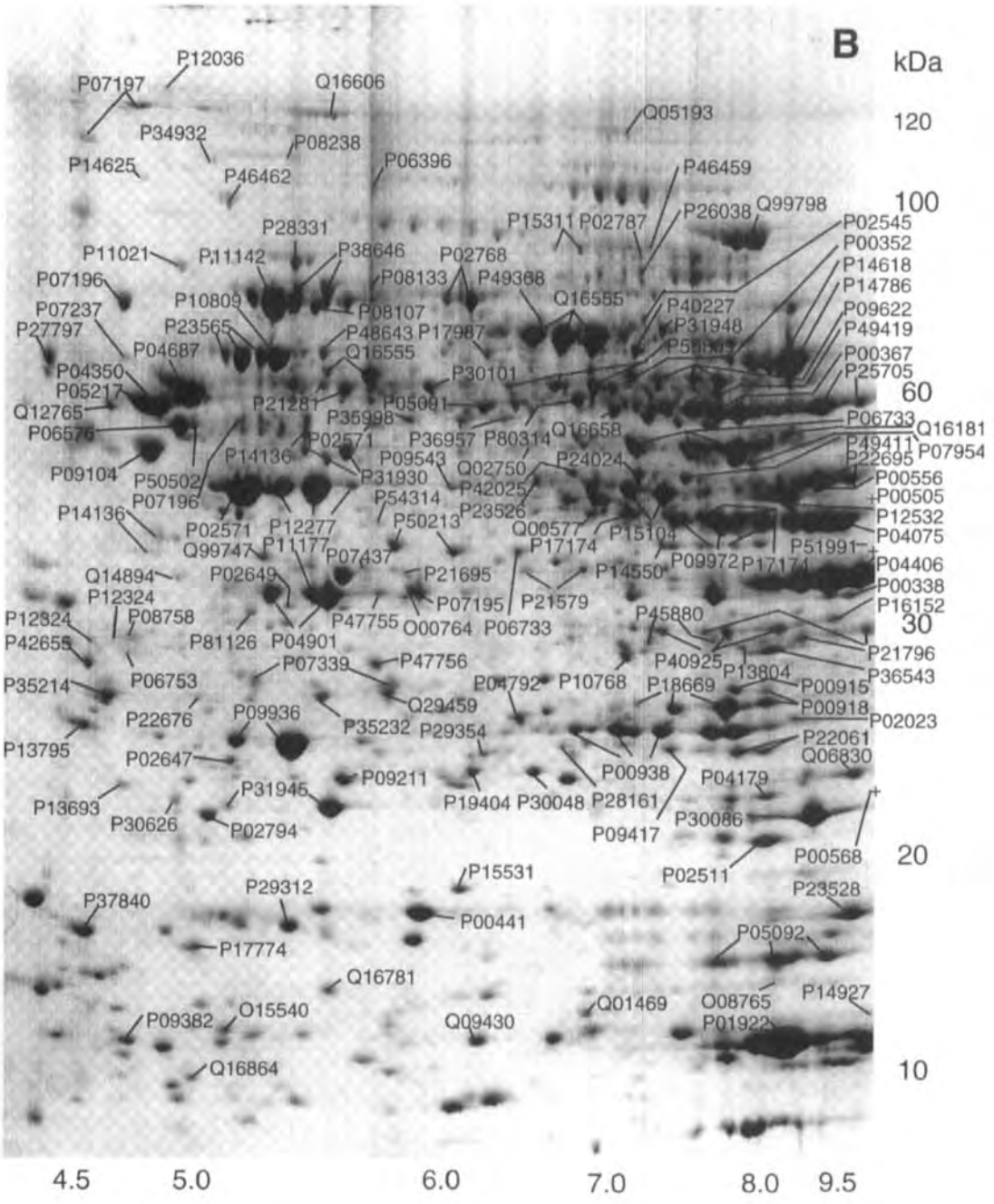
Two-dimensional electrophoresis is the principal approach to study differential protein expression in disorders of the central nervous system. However, a reference map to start with was missing; therefore, we constructed a 2-D database of human brain proteins. The database includes approximately 180 brain proteins, identified on the basis

of peptide mass matching [28]. The introduction of internal peptide standards to correct the measured peptide masses allowed the use of very narrow windows of mass tolerance (0.0075%) during protein search and this increased the confidence of identification. If the number of matching peptides is six or higher, out of a total of 20 peptides detected, the identification may be considered as unambiguous. Usually a high rate of homology among the corresponding proteins from the various mammalian species was observed. The inclusion of homologous mammalian proteins in the protein search list additionally increases the confidence of identification. Figure 2 shows



an example of protein search, where all matching proteins are glial fibrillary acidic protein from four mammalian species. For the human protein the highest number of matching peptides was detected.

The identified proteins (Table 1), which are simultaneously the major visible components of the brain tissue extract, can be classified into several groups. About 30 of them are structural proteins, such as tubulin chains, synaptoso-



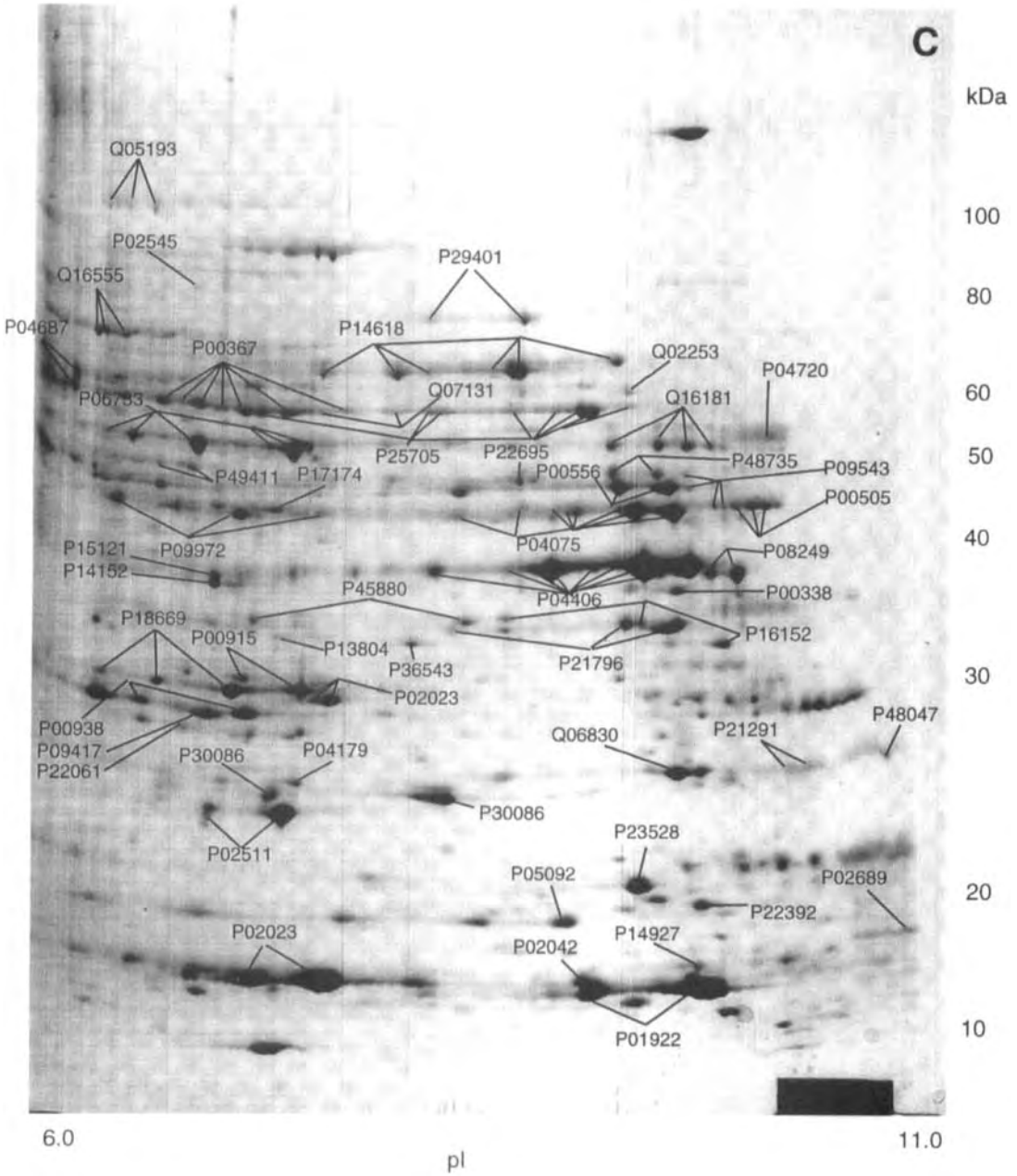


Figure 1. Two-dimensional map of human brain proteins. The proteins from the parietal lobe of cortex of a 63-year-old male individual, who died from coronary artery occlusion, were extracted and separated on (A) pH 4–7, (B) pH 3–10 nonlinear and (C) pH 6–11 IPG strips, followed by 9–16% SDS-polyacrylamide gels, as stated in Section 2.2. The gels were stained with Coomassie blue. The spots were analyzed by MALDI-MS. The proteins identified are designated with their SWISS-PROT accession numbers. The names of the proteins are listed in Table 1.

Table 1. Human brain proteins

Number	Abbr. Name	Protein Name	M_r		pI		MALDI-MS			Location
			Theor.	Observ.	Theor.	Observ.	Peptides		Sequence	(Figure)
							Matching	Total	Coverage (%)	
O00764	PDXK_HUMAN	Pyridoxine kinase (EC 2.7.1.35)	35307	31000	6.1	6.1	8	20	33	B
O08765	GEF2_RAT	Ganglioside expression factor 2	13715	14000	8.4	9.5	4	23	35	B
O15540	FABB_HUMAN	Fatty acid-binding protein, brain (B-FABP)	14862	15000	5.2	5.2	8	23	74	B
P00338	LDHM_HUMAN	L-lactate dehydrogenase (EC 1.1.1.27) chain M	36950	34000	8.3	9.6	9	23	36	B, C
P00352	DHAC_HUMAN	Aldehyde dehydrogenase, cytosolic (EC 1.2.1.3)	55323	62000	6.7	7	13	23	35	B
P00367	DHE3_HUMAN	Glutamate dehydrogenase 1	61701	50000	7.8	9	11	20	24	B, C
P00441	SODC_HUMAN	Superoxide dismutase (EC 1.15.1.1) (Cu-Zn)	16153	17000	6.1	6.1	5	23	72	A, B
P00505	AATM_HUMAN	Aspartate transaminase precursor, mitochondrial	47844	45000	9.6	9.9	9	23	23	B, C
P00556	PGK1_HUMAN	Phosphoglycerate kinase 1	44967	50000	8.1	8.5	10	23	31	B, C
P00558	PGK1_HUMAN	Phosphoglycerate kinase (EC 2.7.2.3)	44985	45000	8.1	9.8	13	23	43	C
P00568	KAD1_HUMAN	Adenylate kinase isoenzyme 1 (EC 2.7.4.3)	21777	22000	8.6	9.6	6	23	46	B
P00915	CAH1_HUMAN	Carbonic anhydrase I (EC 4.2.1.1)	28778	26000	7.1	8.2	11	20	54	B, C
P00918	CAH2_HUMAN	Carbonic anhydrase II (EC 4.2.1.1)	29153	25000	7.5	8.4	10	18	48	B
P00938	TPIS_HUMAN	Triose-phosphate isomerase (EC 5.3.1.1)	26937	27000	6.9	6.9	10	20	61	A, B, C
P01922	HBA_HUMAN	Hemoglobin alpha chain	15304	15000	9.2	9.3	5	23	42	B, C
P02023	HBB_HUMAN	Hemoblobin beta chain	15971	27000	7.3	8	10	20	71	A, B, C
P02042	HBD_HUMAN	Hemoglobin delta chain	16028	15000	8.2	9.2	7	23	65	C
P02511	CRAB_HUMAN	Alpha crystallin b chain	20146	21000	7.4	8.3	7	20	44	B, C
P02545	LAMA_HUMAN	Lamin A	74379	70000	7	7	7	15	15	B, C
P02570	ACTB_HUMAN	Actin beta	42051	40000	5.2	5.2	6	23	22	A
P02571	ACTG_HUMAN	Actin cytoplasmic 2 (gamma actin)	42107	40000	5.3	5.3	7	20	19	B
P02647	APA1_HUMAN	Apolipoprotein A-I (APO-AI)	30758	24000	5.6	5.2	7	18	22	A, B
P02686	MBP_HUMAN	Myelin basic protein	21536	18000	11.9	5.8	6	23	41	A
P02689	MYP2_HUMAN	Myelin P2 protein	15013	15000	10.6	10.6	5	23	50	C
P02768	ALBU_HUMAN	Serum albumin	71317	66000	6.2	6.4	11	20	20	A, B
P02787	TRFE_HUMAN	Serotransferrin precursor (siderophilin)	79280	85000	7.1	7.2	6	18	9	B
P02794	FRIH_HUMAN	Ferritin heavy chain	21252	22000	5.4	5.1	6	20	46	B
P03974	TERA_PIG	Transitional endoplasmic reticulum ATPase	89930	85000	5	5.3	6	22	13	A
P04075	ALFA_HUMAN	Fructose-biphosphate aldolase (EC 4.1.2.13)	39720	40000	8.1	9	10	17	36	B, C
P04179	SODM_HUMAN	Superoxide dismutase (EC 1.15.1.1) (Mn)	24877	22000	8.4	8.4	5	20	34	B, C
P04350	TBB5_HUMAN	Tubulin beta-5 chain	50055	48000	4.7	4.8	6	20	14	B
P04406	G3P2_HUMAN	Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)	36070	32000	8.7	8.7	7	21	29	B, C
P04687	TBA1_HUMAN	Tubulin alpha-1 chain	50809	50000	4.9	5	7	20	21	A, B, C
P04720	EF11_HUMAN	Translation elongation factor eEF-1 alpha-1 chain	50451	47000	9.7	10.2	6	23	16	C
P04901	GBB1_HUMAN	Guanine nucleotide-binding protein beta subunit 1	38151	31000	5.9	5.7	6	20	19	A, B
P05091	DHAM_HUMAN	Aldehyde dehydrogenase (ALDHI) (EC 1.2.1.3)	56858	50000	7	6.5	7	18	14	A, B
P05092	CYPH_HUMAN	Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase)	18097	15000	7.8	8.4	5	20	31	B, C
P05217	TBB2_HUMAN	Tubulin beta-2 chain	50255	50000	4.8	4.8	10	20	23	B
P05413	FABH_HUMAN	Fatty acid-binding protein, cardiac and skeletal muscle	14905	14000	6.8	6.7	5	23	53	A
P06396	GELS_HUMAN	Gelsolin	86043	95000	6.2	5.9	9	20	16	A, B
P06576	ATPB_HUMAN	ATP synthase beta chain (EC 3.6.1.34)	56524	48000	5.2	5	17	19	34	A, B
P06733	ENOA_HUMAN	Phosphopyruvate hydratase (EC 4.2.1.11) (alpha enolase)	47481	50000	7.4	6.8	10	23	41	A, B, C
P06753	TPM3_HUMAN	Tropomyosin α chain	32855	28000	4.5	4.8	5	18	18	B
P06761	GR78_RAT	78 kDa glucose regulated protein (grp 78)	72473	75000	4.9	5.2	8	19	22	A
P07195	LDHH_HUMAN	L-lactate dehydrogenase h chain (EC 1.1.1.27) (ldh-b)	36769	31000	6	6.1	8	19	25	A, B
P07196	NFL_HUMAN	Neurofilament triplet l protein (nf-l)	61665	66000	4.5	4.7	12	20	22	A, B
P07197	NFM_HUMAN	Neurofilament triplet M protein (NF-M)	102000	130000	4.7	4.7	15	20	19	A, B
P07237	PDI_HUMAN	Protein disulfide isomerase	57467	57000	4.7	4.7	8	20	17	B
P07339	CATD_HUMAN	Cathepsin D (EC 3.4.23.5)	45036	27000	6.5	5.3	7	20	14	B
P07437	TBB1_HUMAN	Tubulin beta-1 chain	50240	33000	4.6	5.8	8	20	19	A, B
P07954	FUMH_HUMAN	Fumarate hydratase, mitochondrial	54733	55000	9.4	8	8	23	20	B
P08107	HS71_HUMAN	Heat shock 70 kDa protein 1 (HSP70.1)	70294	65000	5.4	5.6	9	19	17	A, B
P08133	ANX6_HUMAN	Annexin VI (lipocortin VI)	76036	70000	5.4	5.8	13	20	19	B
P08238	HS9B_HUMAN	Heat shock protein HSP90-beta (HSP 90)	83453	105000	4.8	5.5	8	20	13	B
P08249	MDHM_MOUSE	Malate dehydrogenase, mitochondrial	36044	45000	8.8	9.8	7	23	32	C
P08758	ANX5_HUMAN	Annexin V (lipocortin V)	35840	29000	4.8	4.7	9	20	28	B
P09104	ENOG_HUMAN	Gamma enolase (EC 4.2.1.11)	47467	43000	4.8	4.8	10	21	27	A, B
P09211	GTP_HUMAN	Glutathione *S*-transferase P (gstp1-1) (EC 2.5.1.18)	23438	23000	5.4	5.8	6	17	36	B
P09382	LEG1_HUMAN	Galectin-1 (lactose-binding lectin 1)	14917	12000	5.2	4.9	5	20	45	B
P09417	DHPR_HUMAN	Dihydropteridine reductase (EC 1.6.99.	26015	22000	7.4	7.7	6	22	33	B, C
P09543	CN37_HUMAN	2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.2.15)	47947	46000	9.8	9.8	8	23	24	B, C
P09622	DLDH_HUMAN	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	54686	55000	7.7	8	7	17	13	B
P09936	UBL1_HUMAN	Ubiquitin carboxyl-terminal hydrolase isozyme L1 (EC 3.1.2.15)	25150	24000	5.3	5.5	7	19	57	A, B
P09972	ALFC_HUMAN	Fructose-bisphosphate aldolase C (EC 4.1.2.13)	39699	36000	6.8	8.1	12	19	40	B, C
P10768	ESTD_HUMAN	Esterase D (EC 3.1.1.1)	31955	28000	7	7.4	9	18	51	B
P10809	P60_HUMAN	Mitochondrial matrix protein P1 (hsp-60)	61187	58000	5.6	5.5	10	20	26	A, B
P11021	GR78_HUMAN	78 kDa glucose regulated protein (grp 78)	72185	77000	4.9	5	10	19	18	B
P11142	HS7C_HUMAN	Heat shock cognate 71 kDa protein	71082	70000	5.3	5.4	6	19	14	A, B
P11177	ODPB_HUMAN	Pyruvate dehydrogenase E1 component (EC 1.2.4.1) (pdhe1-b)	39536	31000	6.9	5.6	11	20	29	B
P12036	NFH_HUMAN	Neurofilament triplet H protein (NF-H)	111940	140000	6.1	5	10	20	11	B
P12277	KCRB_HUMAN	Creatine kinase beta chain (EC 2.7.3.2)	42902	40000	5.4	5.6	13	20	43	A, B
P12324	TPMN_HUMAN	Tropomyosin (TM30-NM)	29242	28000	4.6	4.7	6	18	25	B
P12532	KCRU_HUMAN	Creatine kinase, ubiquitous mitochondrial (EC 2.7.3.2)	47406	40000	8.3	8.4	10	20	29	B
P13535	MYSP_HUMAN	Myosin heavy chain	84015	80000	6.5	6.4	7	23	11	A
P13795	SN25_HUMAN	Synaptosomal associated protein 25 (snap-25)	23528	23000	4.5	4.6	7	20	33	B
P13804	ETFA_HUMAN	Electron transfer flavoprotein alpha chain precursor	35399	30000	8.4	7.6	8	23	33	B, C
P14136	GFAP_HUMAN	Glial fibrillary acidic protein (gfap)	49906	47000	5.3	5.5	13	20	29	A, B
P14152	MDHC_MOUSE	Malate dehydrogenase (EC 1.1.1.37), cytosolic	36625	34000	6.5	7.3	5	23	24	C
P14550	ALDX_HUMAN	Alcohol dehydrogenase (NADP+) (EC 1.1.1.2)	36760	35000	6.8	7.6	5	18	18	B

Table 1. continued

P14618	KPY1_HUMAN	Pyruvate kinase M1 (EC 2.7.1.40)	58280	55000	7.6	8.7	6	18	14	B, C
P14625	ENPL_HUMAN	Endoplasmin (94 kDa glucose regulated protein) (GRP94)	92696	100000	4.6	4.8	9	20	13	B
P14786	KPY1_HUMAN	Pyruvate kinase, M2 isozyme (EC 2.7.1.40)	58316	55000	7.8	8.5	9	20	20	B
P14927	UCR6_HUMAN	Ubiquinone-binding protein QP-C	13521	15000	9.6	9.7	6	23	54	B, C
P15104	GLNA_HUMAN	Glutamine synthetase (EC 6.3.1.2)	42664	39000	6.9	7.5	6	19	20	B
P15121	ALDR_HUMAN	Aldehyde reductase (EC 1.1.1.21)	36229	35000	7	7.3	7	23	29	C
P15311	EZRI_HUMAN	Ezrin (p81) (cytovillin)	69338	91000	6.2	7	8	18	12	B
P15531	NDKA_HUMAN	Nucleoside diphosphate kinase A (EC 2.7.4.6)	17308	18000	6.1	6.2	7	21	45	A, B
P16152	DHCA_HUMAN	Carbonyl reductase (NADPH) (EC 1.1.1.184)	30640	30000	8.3	8.6	8	23	46	B, C
P16949	STHM_HUMAN	Stathmin	17291	17000	5.9	5.6	6	23	38	A
P17174	AATC_HUMAN	Aspartate aminotransferase (transaminase A) (EC 2.6.1.1)	46334	37000	7.2	7.4	11	20	34	B, C
P17600	SYN1_HUMAN	Synapsins IA	74151	80000	10.6	6.1	8	23	13	A
P17774	GLMB_HUMAN	GLIA maturation factor β (GMF-β)	16742	15000	5	5.1	5	18	37	B, C
P17987	TCPA_HUMAN	T-complex protein 1, alpha subunit (TCP-1 alpha)	60818	58000	6	6.4	11	20	22	A, B
P18669	PMGB_HUMAN	Phosphoglycerate mutase, brain form (EC 5.4.2.1) (pgam-b)	28768	25000	7.2	8	7	20	35	B, C
P19404	NUHM_HUMAN	NADH-ubiquinone dehydrogenase 24 kDa subunit (EC 1.6.99.3)	27631	24000	8.1	6.5	6	20	26	B
P21281	VAT2_HUMAN	Vacuolar ATP synthase subunit B, brain isoform (EC 3.6.1.34)	56823	53000	5.6	5.8	8	20	16	B
P21291	CYSR_HUMAN	Cysteine-rich protein	21409	20000	8.5	10.3	5	18	43	C
P21579	SYT1_HUMAN	Synaptotagmin I (p65)	47884	32000	8.2	7	10	18	23	A, B
P21695	GPDA_HUMAN	Glycerol-3-phosphate dehydrogenase (EC 1.1.1.8)	38065	32000	6.1	6	5	20	15	B
P21796	POR1_HUMAN	Voltage-dependent anion-selective channel protein 1 (VDAC1)	30736	29000	9	10	8	15	45	A, B, C
P22061	PIMT_HUMAN	Protein-beta-aspartate methyltransferase (EC 2.1.1.77)	24642	24000	7.3	8.2	7	20	42	B, C
P22392	NDKB_HUMAN	Tumor metastasis inhibitor nm23-H2 (fragment)	19947	16000	10.4	9.8	6	23	36	C
P22676	CART_HUMAN	Calretinin (cr) (29 kd calbindin)	31615	27000	4.9	5	5	23	24	B
P22695	UCR2_HUMAN	Ubiquinol-cytochrome C reductase complex protein 2 (EC 1.10.2.2.)	48610	50000	9.1	9.4	7	17	18	B, C
P23526	SAHH_HUMAN	Adenosylhomocysteinase (EC 3.3.1.1)	48254	40000	6.4	6.8	8	18	20	B
P23528	COF1_HUMAN	Cofilin	18718	16000	8.2	9.4	7	22	50	B, C
P23565	AINX_RAT	Alpha-internexin (alpha-inx) (rat)	56252	56000	5	5.3	11	21	20	A, B
P25705	ATPA_HUMAN	ATP synthase alpha chain (EC 3.6.1.34)	59827	50000	9.9	9.8	14	20	31	B, C
P26038	MOES_HUMAN	Moesin (membrane-organizing extension spike protein)	67760	80000	6.3	7.2	9	23	17	B
P28161	GTM2_HUMAN	Glutathione S-transferase mu-2 (EC 2.5.1.18)	25768	22000	6.3	6.5	8	18	36	A, B
P28331	NUAM_HUMAN	NADH-ubiquinone oxidoreductase 75 kDa subunit (EC 1.6.99.3)	80548	80000	6	5.6	11	19	22	A, B
P28661	BH5_MOUSE	Gene H5 protein	43205	45000	6.7	6.8	9	23	30	A
P29401	TKT_HUMAN	Transketolase (EC 2.2.1.1)	68434	65000	7.8	8.3	9	23	21	C
P30040	ER31_HUMAN	Endoplasmic reticulum protein ERp29/ERp31	29297	28000	7.5	6.1	7	23	37	A
P30041	ULA6_HUMAN	Human mRNA for KIAA0106 gene, complete cds	25133	26000	6.3	6.7	5	23	28	A
P30048	TDXM_HUMAN	Thioredoxin-dependent peroxide reductase	28017	23000	7.8	6.8	5	18	21	A, B
P30084	ECHM_HUMAN	Enoyl-CoA hydratase	31807	26000	8	6.2	7	23	40	A
P30086	PBP_HUMAN	Phosphatidylethanolamine-binding protein	21026	22000	7.6	8.8	10	20	75	B, C
P30101	ER60_HUMAN	Protein disulfide isomerase er-60 (EC 5.3.4.1) (erp60)	57145	49000	6.3	6.2	10	20	18	A, B
P30626	SORC_HUMAN	Sorcin (CP-22)	21947	20000	5.3	5	5	20	23	B
P31930	UCR1_HUMAN	Ubiquinol-cytochrome-C reductase complex core protein I	53269	42000	6.3	5.8	11	20	30	B
P31945	NKFB_HUMAN	Natural killer cell enhancing factor B (NKEF-B)	22048	22500	5.8	5.7	5	20	40	B
P31948	IEFS_HUMAN	Transformation-sensitive protein ief ssp 3521	63226	70000	6.8	7	11	23	21	B
P34932	HS74_HUMAN	Heat shock 70 kDa protein 4 (HSP70RY)	79857	105000	5	5.2	8	20	13	B
P35214	143G_RAT	14-3-3 Protein gamma (kcip-1) (bovine)	28160	28000	4.6	4.7	8	20	28	B
P35232	PHB_HUMAN	Prohibitin	29842	27000	5.5	5.6	8	18	33	B
P35998	PRS7_HUMAN	26s protease regulatory subunit 7 (mss1 protein).	49002	53000	6.7	6	8	23	25	B
P36543	VATE_HUMAN	H+-transporting ATP synthase chain E, vacuolar (EC 3.6.1.34)	26185	29000	8.4	8.2	7	23	30	B, C
P36957	ODO2_HUMAN	Dihydrolipoamide succinyltransferase component E2	48951	48000	9.2	6.5	6	18	15	B
P37840	SYUA_HUMAN	Alpha-synuclein (non-a beta component of ad amyloid)	14451	18000	4.5	4.6	4	23	41	B
P38646	GR75_HUMAN	75 kDa glucose regulated protein (grp 75)	74018	69000	6.2	6.5	8	20	15	A, B
P40227	TCPZ_HUMAN	T-complex protein 1, zeta subunit (TCP-1 zeta)	58443	58000	6.7	7.4	7	20	13	B
P40925	MDHC_HUMAN	Malate dehydrogenase (EC 1.1.1.37)	36500	29000	7.4	8.1	6	20	21	B
P42024	ACTZ_HUMAN	α-Centractin (centrosome-associated actin homolog)	42700	40000	6.8	7.5	8	18	26	
P42025	ACTY_HUMAN	β-Centractin	42380	40000	6.4	7	6	18	15	B
P42655	143E_HUMAN	14-3-3 Protein epsilon (protein kinase C inhibitor protein-1)	29326	31000	4.5	4.5	13	20	49	B
P45880	POR2_HUMAN	Voltage-dependent anion-selective channel protein 2 (VDAC2)	38638	28000	6.7	7.5	8	20	30	B, C
P46459	NSF_HUMAN	Vesicular-fusion protein nsf	83115	85000	6.8	7.2	7	23	13	B
P46462	TERA_RAT	Transitional endoplasmic reticulum ATPase	89976	110000	5	5.2	10	20	14	B
P47755	CAZ2_HUMAN	F-actin capping protein alpha-2 subunit (CAPZ)	33156	31000	5.7	5.9	6	18	39	B
P47756	CAPB_HUMAN	F-actin capping protein β subunit (CAPZ)	30951	28000	5.8	5.9	5	18	22	B
P48047	ATPO_HUMAN	ATP synthase oligomycin sensitivity conferring protein	23376	24000	10.8	10.8	6	23	38	C
P48163	MAOX_HUMAN	Malate dehydrogenase	64679	65000	6	6	7	23	12	A
P48643	TCPE_HUMAN	T-complex protein 1, ε subunit (TCP-1-ε) (KIAA0098)	60088	58000	5.4	5.6	7	18	15	B
P48735	IDHP_HUMAN	Isocitrate dehydrogenase	51372	47000	9	9.6	10	23	29	C
P49368	TCPG_HUMAN	T-complex protein 1, γ subunit (cct-gammma)	60862	63000	6.6	6.8	12	20	20	B
P49411	EFTU_HUMAN	Elongation factor Tu (p43)	49852	40000	7.6	7.8	12	18	34	B, C
P49419	DHAX_HUMAN	Antiquitin	55844	60000	6.8	7.1	6	23	14	B
P50213	IDHA_HUMAN	Isocitrate dehydrogenase (NAD) (EC 1.1.1.41)	40022	35000	6.9	6.1	10	19	25	B
P50399	GDIB_HUMAN	Rab GDP dissociation inhibitor beta	51133	50000	6.3	6.7	7	22	27	A
P50502	HIP_HUMAN	hsc70-interacting protein	41469	55000	4.9	4.9	9	20	27	B
P51991	ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein a3	39946	42000	8.9	10	6	23	24	B
P52565	GDIR_HUMAN	Rho protein GDP-dissociation inhibitor 1 (IEF 8118)	23249	25000	4.8	4.8	6	23	30	A
P54314	GBB5_MOUSE	Guanine nucleotide-binding protein beta subunit 5	39504	36000	6	5.9	6	20	24	B
P55809	SCOT_HUMAN	Succinyl-CoA:3-ketoacid-coenzyme A transferase (EC 2.8.3.5)	56578	52000	7.4	7.1	9	20	26	A, B
P80314	TCPB_MOUSE	T-complex protein 1, beta subunit (TCP-1 beta)	57753	50000	6.4	7.1	8	20	22	B
P81126	SNAB_BOVIN	Beta-soluble NSF attachment protein (SNAP-β)	33875	30000	5.3	5.3	7	18	28	B
Q00577	PUR_HUMAN	Transcriptional activator protein PUR-alpha	35003	36000	6.4	7.2	6	18	27	B
Q01469	FABE_HUMAN	Fatty acid-binding protein (E-FABP)	15496	13000	7	7.1	5	18	31	B
Q01851	BR3A_HUMAN	Single-stranded-DNA-binding protein Pur alpha	35062	45000	6.7	6.3	5	23	11	A
Q02253	MMSA_RAT	Methylmalonate-semialdehyde dehydrogenase (acylating)	58226	55000	8.2	9.3	6	22	15	C
Q02750	MPk1_HUMAN	Dual specificity mitogen-activated protein kinase kinase 1	43636	50000	6.6	6.3	6	23	15	B
Q05193	DYN1_HUMAN	Dynamin-1	97745	115000	7.3	7.3	17	20	15	A, B, C
Q06830	TDX2_HUMAN	Enhancer protein (thioredoxin peroxidase 2)	22455	23000	8	9.4	7	23	47	B, C

Table 1. continued

Q07131	UDPG_HUMAN	UTP-glucose-1-phosphate uridylyltransferase, skeletal muscle	57100	50000	8.7	8.2	10	23	26	C
Q09430	PRO2_BOVIN	Profilin II (bovine)	12011	12000	9.3	6.5	5	14	41	B
Q12765	Y193_HUMAN	Hypothetical protein KIAA0193	39352	50000	4.4	4.7	5	18	16	B
Q14141	SEP2_HUMAN	Septin 2 homolog	58317	58000	6.3	6.5	6	21	14	A
Q16181	C10H_HUMAN	hCDC10 protein	40941	50000	9.3	9.3	9	23	31	B, C
Q16555	DPY2_HUMAN	Dihydropyrimidinase-related protein 2	62710	68000	6.3	6.5	13	23	32	A, B, C
Q16606	SPCN_HUMAN	Spectrin alpha chain, brain	118671	100000	5.7	5.7	9	23	11	A, B
Q16658	FASC_HUMAN	Fascin (actin binding protein)	55123	50000	7.2	7.4	8	18	16	B
Q16781	UBCC_HUMAN	Ubiquitin-conjugating enzyme (EC 6.3.2.-) E2	17183	15000	6.5	5.7	5	23	50	A, B
Q16864	VATF_HUMAN	Vacuolar ATP synthase subunit F (EC 3.6.1.34)	13349	11000	5.3	5.1	5	17	31	B
Q99747	SNAG_HUMAN	Gamma-soluble NSF attachment protein (SNAP-γ)	35066	35000	5.2	5.3	9	20	31	B
Q99798	ACON_HUMAN	Aconitate hydratase, mitochondrial	86346	90000	7.5	7.8	10	23	19	B

Proteins from the parietal region of cortex of human brain were extracted and separated by 2-D PAGE as stated under Section 2.2. The proteins were identified by MALDI-MS, following digestion with trypsin or endoproteinase Lys-C. The identified protein spots are indicated in Fig. 1 and are designated with their SWISS-PROT accession numbers. The theoretical and the approximate observed M_r and p*I* values, as well as the matching and total peptides and the protein amino acid sequence coverage by the matching peptides are given. The amino acid sequence coverage provides an indication of confidence of identification. Usually the protein with the highest sequence coverage was selected. The column "Location" indicates in which part of Fig. 1 the corresponding protein spot can be found.

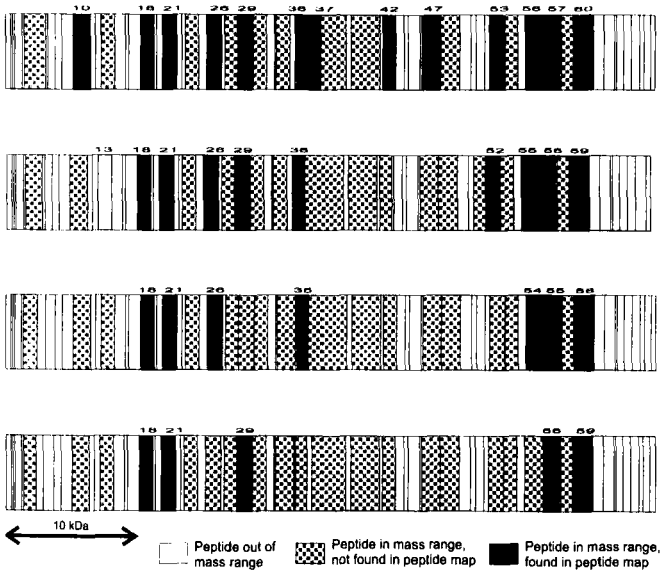


Figure 2. Matching proteins obtained from a protein search in the SWISS-PROT database, after introduction of the peptide masses determined by MALDI-MS derived from the tryptic digestion of a protein spot. All matches are glial fibrillary acidic proteins from mammals. The numbers of matching peptides for the protein from each species are indicated. The bars indicate the theoretical peptides of GFAP from each species generated by tryptic digestion. The width of each bar is proportional to the peptide mass. Black bars indicate the peptides found. For the human protein a higher number of matching peptides and higher amino acid sequence coverage were assigned.

mal proteins, neurofilaments and about 65 are enzymes or enzyme chains with a wide spectrum of catalytic activities. Many of the strong spots at the basic region of Figure 1B represent enzymes, mainly involved in sugar metabolism, such as glyceraldehyde 3-phosphate dehydrogenase (P04406), fructose-biphosphate aldolase (P04075), pyruvate kinase (P14618, P14786), phosphopyruvate hydratase (P06733) and many others. This can be attributed to the high levels of oxidative phosphorylation in brain tissue. The other classes comprise fewer proteins with various functions, such as heat-shock proteins (about 15, glucose-regulated proteins, T-complex protein chains), transporters, channels, cholesterol- and lipid-binding proteins, growth factors, enhancers, transcription factors, *etc.*. Eleven proteins are brain-specific, including neuron-specific structural and regulatory proteins. Approximately one half of the identified proteins are localized in the cytoplasma, one fourth in mitochondria, and one tenth in the nucleus (Fig. 3). For 46% of the proteins of Table 1, no information concerning their cellular localization existed in the SWISS-PROT database.

On the 2-D map, a large number of proteins are represented by multiple spots, as for example the proteins P00367, P00938, P04406, P18669, Q16555. The multiple spots may be the result of phosphorylation, glycosylation or other post-translational modification; to date, however, we do not know the precise reason(s) or biological signifi-

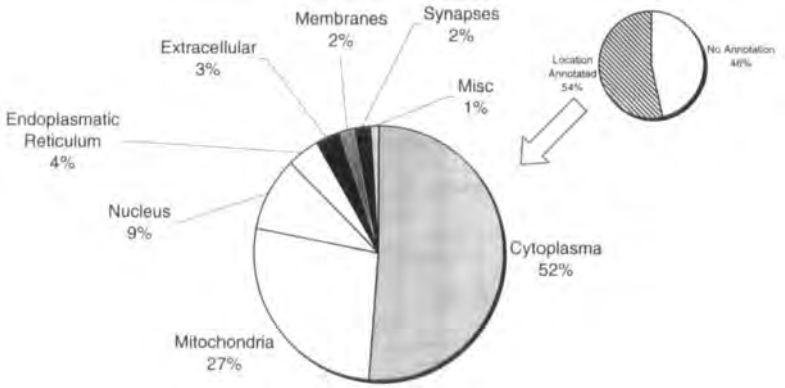


Figure 3. Cellular localization of brain proteins. The identified brain proteins (Table 1) were classified according to their cellular localization, based on annotations from the SWISS-PROT database. For 46% of the identified proteins no information on localization was available.

cance for the observed heterogeneities. Gel comparison revealed the presence of several minor, probably allelic, differences, mainly concerning the intensity of certain spots, when samples from the same brain region of other members of the control group were compared. Variations were observed, for example for tubulin chains, neurofilament proteins, enzyme chains and others. Differences were also observed in the 2-D maps of proteins from five brain regions of the same individual (frontal, parietal, temporal and occipital cortex and the cerebellum; data not shown). A more detailed description of the variations is in preparation.

In summary, we constructed a 2-D reference map of human brain proteins, including approximately 180 identified proteins. The major components represent structural proteins, heat-shock proteins, and enzymes with a wide spectrum of catalytic activities, and are mainly localized in the cytoplasma and mitochondria. To our knowledge, this is the first 2-D database of human brain proteins.

The authors at the University of Vienna are highly indebted to the Red Bull Company, Salzburg, Austria, for generous financial support. The technical assistance of J.-F. Juranville is appreciated.

Received October 28, 1998

5 References

[1] Hochstrasser, D. F., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 187–219.

[2] Anderson, N. L., Anderson, N. G., *Electrophoresis* 1998, *19*, 1853–1861.

[3] Golaz, O., Hughes, G. J., Frutiger, S., Paquet, N., Bairoch, A., Pasquali, C., Sanchez, J.-C., Tissot, J.-D., Appel, R. D., Walzer, C., Balant, L., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1223–1231.

[4] Yun, M., Wu, W., Hood, L., Harrington, M., *Electrophoresis* 1992, *13*, 1002–1013.

[5] Gravel, P., Sanchez, J.-C., Walzer, C., Golaz, O., Hochstrasser, D. F., Balant, L., Hughes, G. J., Garcia-Sevilla, J., Guimon, J., *Electrophoresis* 1995, *16*, 1152–1159.

[6] Hughes, G. J., Frutiger, S., Paquet, N., Pasquali, C., Sanchez, J.-C., Tissot, J.-D., Bairoch, A., Appel, R. D., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1216–1222.

[7] Evans, G., Wheeler, C. H., Corbett, J. M., Dunn, M. J., *Electrophoresis* 1998, *19*, 471–479.

[8] Corbett, J. M., Why, H. J., Wheeler, C. H., Richardson, P. J., Archard, L. C., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1998, *19*, 2031–2042.

[9] Tonella, L., Walsh, B. J., Sanchez, J.-C., Ou, K., Wilkins, M. R., Tyler, M., Frutiger, S., Gooley, A. A., Pescaru, I., Appel, R. D., Yan, J. X., Bairoch, A., Hoogland, C., Morch, F. S., Hughes, G. J., Williams, K. L., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 1960–1971.

[10] Link, A. J., Robinson, K., Church, G. M., *Electrophoresis* 1997, *18*, 1259–1313.

[11] Langen, H., Gray, C., Röder, D., Juranville, J.-F., Takacs, B., Fountoulakis, M., *Electrophoresis* 1997, *18*, 1184–1192.

[12] Fountoulakis, M., Langen H., Evers, S., Gray, C., Takacs, B., *Electrophoresis* 1997, *18*, 1193–1202.

[13] Fountoulakis, M., Juranville, J.-F., Berndt, P., *Electrophoresis* 1997, *18*, 2968–2977.

[14] Fountoulakis, M., Takacs, B., Langen, H., *Electrophoresis* 1998, *19*, 761–766.

[15] Fountoulakis, M., Langen, H., Gray, C. Takacs, B., *J. Chromatogr. A* 1998, *806*, 279–291.

[16] Fountoulakis, M., Juranville, J.-F., Röder, D., Evers, S., Berndt, P., Langen, H., *Electrophoresis* 1998, *19*, 1819–1827.

[17] Link, A. J., Hays, L. G., Carmack, E. B., Yates III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[18] Sanchez, J.-C., Golaz, O., Frutiger, S., Scaller, D., Appel, R. D., Bairoch, A., Hughes, G. J., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 556–565.

[19] Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., Payne, W. E., *Electrophoresis* 1997, *18*, 1347–1360.

[20] Antelmann, H., Bernhardt, J., Schmid, R., Mach, H., Volker, U., Hecker, M., *Electrophoresis* 1997, *18*, 1451–1463.

[21] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[22] Bairoch, A., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 93–132.

[23] Winner, K., Kuick, R., Thoraval, D., Hanash, S. M., *Electrophoresis* 1996, *17*, 1741–1751.

[24] Seidl, R., Greber, S., Schuller, E., Bernert, G., Cairns, N., Lubec, G., *Neurosci. Lett.* 1997, *235*, 137–140.

[25] Bradford, M., *Anal. Biochem.* 1976, *72*, 248–254.

[26] Langen, H., Röder, D., Juranville, J.-F., Fountoulakis, M., *Electrophoresis* 1997, *18*, 2085–2090.

[27] Fountoulakis, M., Langen, H., *Anal. Biochem.* 1997, *250*, 153–156.

[28] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 917–927

Giulia Friso
Lilian Wikström

Department of Cellular and Molecular Pharmacology, Astra Pain Discovery Unit Sweden, Astra Pain Control AB, Huddinge, Sweden

Analysis of proteins from membrane-enriched cerebellar preparations by two-dimensional gel electrophoresis and mass spectrometry

Two-dimensional polyacrylamide gel electrophoresis and mass spectrometry is a powerful combination for the separation of complex protein mixtures in biological samples and the subsequent identification of individual polypeptides. We have used this approach to construct a database of proteins of the porcine cerebellum, with emphasis on membrane-bound proteins, as part of our studies on the structure and function of the central nervous system. We compared the ability of different solubilization conditions (using zwitterionic and nonionic detergents; urea and thiourea) to improve the resolution of high molecular weight and hydrophobic proteins, and found the combination of 3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulfonate (CHAPS), Tris, thiourea and urea to give the best results in our experiments. As a marker membrane protein, the NR1 subunit of the *N*-methyl D-aspartate receptor, a 120 kDa hydrophobic protein, was identified using a monoclonal antibody in combination with Western blotting. Sodium chloride treatment of the membrane preparation prior to solubilization caused further enrichment of membrane proteins. Fifty-six spots were identified using matrix-assisted laser desorption/ionization time-of-flight and nanoelectrospray mass spectrometry.

Keywords: Membrane protein / Cerebellum / Two-dimensional polyacrylamide gel electrophoresis / Matrix-assisted laser desorption / ionization – time of flight EL 3346

1 Introduction

The use of IPG strips for isoelectric focusing during 2-DE has significantly improved the reproducibility of 2-D gels [1–3]. The strips permit loading of relative large amounts (micrograms or even milligrams) of protein, which makes subsequent identification of proteins easier [4, 5]. However, the selective loss of certain classes of proteins during conventional 2-DE is still a problem, membrane proteins normally being extensively under-represented. The low solubility of these hydrophobic proteins, especially those of high molecular weight, gives rise to protein aggregation, a major problem during sample application and IEF. Also, the identification of integral and peripheral membrane proteins by microsequencing techniques (Edman degradation, mass spectrometry) is a challenging task, mainly due to the small amount of protein available. Membrane proteins constitute a significant part of the cell's proteins. Key functions, such as the communication of a cell with its environment, are largely dependent on membrane proteins like receptors and ion channels, which therefore are important drug targets.

The aim of our study was to develop a protocol that allows us to monitor membrane proteins from the nervous system, and to generate a database over proteins present in the cerebellum. To date, few 2-D databases from neuronal tissues have been published [6–9], and to our knowledge there is no database of cerebellar proteins among them. We began our investigation on neuronal membrane proteins by applying different methods developed to improve the identification of hydrophobic proteins reported by Rabilloud *et al.* [10, 11] to a membrane preparation from the cerebellum [12]. We tried to further enrich the content of membrane-associated proteins in the cerebellar membrane preparation before 2-D analysis by comparing the effect of washing them with different agents (urea, Tris, NaCl). We chose the NR1 subunit of the *N*-methyl D-aspartate (NMDA) receptor [13, 14] (a ligand-gated ion channel) as a marker protein to follow the fate of large hydrophobic proteins because of its size (120 kDa) and hydrophobicity (three transmembrane domains). NR1 is relatively abundant in the cerebellum, and specific monoclonal antibodies are available. The results were monitored by 2-DE and one-dimensional gel electrophoresis in combination with Western blotting or silver staining.

Correspondence: Dr. Giulia Friso, Department of Cellular and Molecular Pharmacology, APDUS, Astra Pain Control AB, S-14157 Huddinge, Sweden
E-mail: giulia.friso@pain.se.astra.com
Fax: +46-8-553-25421

Abbreviations: ES-MS, electrospray mass spectrometry; **NMDA,** *N*-methyl D-aspartate; **SB 3-10,** caprylyl sulfobetaine

Protein identification was performed by mass spectrometry. Several examples show the advantage of using mass spectrometry to achieve protein identification of 2-D gel-separated proteins [15–17]. The high sensitivity of matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) [18] with delay extraction [19] and electrospray-nanospray mass spectrometry (ES-MS) [20, 21] allows the determination of proteins present in mixtures in low amounts (picomole, femtomoles), permitting identification of less abundant spots in the 2-D gels. For identification, proteins separated in 2-D gels were treated *in situ* with trypsin, the resulting peptide mixture was analyzed by MS, and the masses were compared with those theoretically expected using sequence database searches. This allowed the assignment of identities to 56 out of 74 spots.

2 Materials and methods

2.1 Reagents

The pH 3–10 linear IPG strips, pH 3–10 linear carrier ampholyte IPG buffers and the equipment for running the IPG gels (Dry-Strip kit, Multiphor II; EPS 3500 XL power supply), gel cassettes, the gel caster and the multiple gel unit (Iso-Dalt) for the second dimension and blotting were from Amersham Pharmacia Biotech (Uppsala, Sweden), as were the ECL reagents. Thiourea was from Fluka (Buchs, Switzerland). NP-40, Triton X-100, Tween-20, caprylyl sulfobetaine (SB 3–10) were from Sigma (St. Louis, MO, USA). Urea, CHAPS, DTT, prestained molecular weight markers (high molecular weight range), one-dimensional precast 10%T Tris-glycine gels, as well as the Mini-Protean II system used for running and blotting these gels, were from Bio-Rad (Richmond, CA, USA). Peroxidase-conjugated rabbit anti-mouse immunoglobulins were from Dako A/S (Glostrup, Denmark). Purified mouse monoclonal antibody raised against the NR1 subunit of the NMDA receptor was from Pharmingen (San Diego, CA, USA). Sequencing-grade trypsin for protein digestion was from Promega (Madison, WI, USA), and siliconized Eppendorf tubes used in protein identification were from Axygen (Union City, CA, USA). For peptide purification, Poros™ 50 R2 packing material from PerSeptive Biosystems (Framingham, MA, USA) was used. The MALDI-TOF (Voyager-DE STR) and electrospray quadrupole/orthogonal-acceleration time-of-flight (Q-Tof™) instruments were from PerSeptive BioSystems and Micromass (Manchester, UK), respectively.

2.2 Protein sample preparation

2.2.1 Cerebellar membrane preparation

Membrane preparation of pig cerebellum was performed as described [12]. The cerebellum from 6-month-old Piggham pigs were used. The membrane preparation was stored at –80°C until use. Protein determination was performed using a modified Lowry protein assay adapted for use in microtiter plates [22] using bovine serum albumin as a standard.

2.2.2 Enrichment in membrane proteins by salt and urea wash

Three pellets containing the same amount of cerebellum membrane preparation (around 3 mg each) were resuspended in solutions containing 50 mM Tris-HCl, pH 7.6, or 4.5 M urea in 1 mM Tris-HCl, pH 7.6, or 3 M NaCl in 1 mM Tris-HCl, pH 7.6, respectively. The samples were incubated for 30 min on ice and centrifugated at 100 000 *g* for 60 min at 4°C. The supernatants containing remaining soluble and peripheral proteins were discarded. The pellets were washed once more with 50 mM Tris-HCl, pH 7.6, and centrifugated at 100 000 *g* for another 60 min. The final pellets were then either directly solubilized and used for 2-DE or frozen in liquid nitrogen and stored at –80°C.

2.3 Two-dimensional gel electrophoresis

2.3.1 Sample preparation for isoelectric focusing

For isoelectric focusing we used the sample loading procedure recommended for reswelling of the IPGs [23, 24]. Individual 18 cm IPG strips, pH 3–10 linear, were rehydrated in 400 µL of sample solubilized in rehydration buffer containing zwitterionic detergents (CHAPS), nonionic detergents (NP-40) or the nondetergent sulfobetain SB 3–10 in the presence of urea or in a combination of urea and thiourea. Solubilization was performed for 30 min at 30°C. The samples were then centrifuged for 5 min at 12 000 rpm in a bench-top centrifuge to remove unsolubilized material, and the supernatant was loaded onto the IPGs for 2-DE analysis. We compared the efficiency in solubilizing the cerebellum high molecular weight proteins of four different rehydration buffers containing: (i) 8 M urea, 2% NP-40, 0.3% DTT, 2% carrier ampholytes; (ii) 7 M urea, 2 M thiourea, 4% NP-40, 0.3% DTT, 2% carrier ampholytes; (iii) 5 M urea, 2 M thiourea, 2 % SB 3–10, 0.3% DTT, 2% carrier ampholytes; (iv) 7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris, pH 9.6, 0.3% DTT, 2% carrier ampholytes. Sample loading and rehydration were allowed to proceed at room temperature for 24 h to ensure maximal diffusion of the proteins into the strips as previously described [23, 24]. Depending on the purpose of the gel (analytical or micropreparative), different quantities of proteins were loaded in the first dimension: 200–250 µg proteins for analytical and around 1.0 mg of protein for micropreparative gels. The isoelectric focusing

was done at 20°C. The running conditions were 300 V for 1 h, 500 V for 3 h and 3500 V until a total amount of 90 000 Vh was reached. Focused strips were either stored at −20°C, or directly used for the second dimension.

2.3.2 SDS-PAGE and gel staining

After isoelectric focusing, the strips were equilibrated for 15 min in a solution containing 65 mM DTT, 6 M urea, 30% glycerol, 2% SDS, 50 mM Tris-HCl, pH 8.8, and traces of bromophenol blue. A second equilibration step was also carried out for 15 min in the same solution except for DTT, which was replaced by 259 mM iodoacetamide for alkylation. For 2-DE, the IPG strips were loaded onto vertical SDS polyacrylamide gels (10%T, 2.1%C or 12.5%T, 2.1%C; 1.5 mm thick) and sealed in place with 1% agarose. Migration was performed at 100 V for 20 h at 20°C. After separation, the proteins were detected by silver stain for analytical gels (essentially as described [25] and modified [26]) by copper stain for mass spectrometric identification [27] or blotted onto PVDF membrane for immunodetection [28]. Gels were scanned with a Scan Jet 4C scanner (Hewlett Packard, Avondale, PA, USA) and images stored as TIFF (tag image file format) files. Matching of spots on preparative and analytical gels was done visually.

2.4 Immunoblotting

One- and two-dimensional gels were blotted onto PVDF membranes essentially as described [28]. Transfer of high molecular weight membrane proteins was enhanced by including SDS and by lowering the amount of methanol in the transfer buffer (25 mM Tris, 190 mM glycine, 15% methanol, 0.05% SDS). Immunodetection of NR1 was performed using monoclonal anti-NR1 antibodies and, as secondary antibodies, peroxidase-conjugated rabbit anti-mouse immunoglobulins. The result was visualized using the ECL reagents. The anti-NR1 antibody detects a single band at a relative molecular mass of 120 kDa, which is consistent with the calculated molecular weight of the protein.

2.5 Protein identification

Electrophoretically separated proteins were excised from the gel and analyzed (as described [29] and modified [26]). Briefly, gel slices were washed in water for 1 h and then in 30–40 % acetonitrile, 25 mM ammonium carbonate, pH 8.0 (1–3 times, for approximately 20 min) in presiliconized Eppendorf tubes and dehydrated in a Speed-Vac vacuum-dryer (for approximately 30 min). The gel pieces were reswollen in a trypsin solution containing 0.1–0.2 μg trypsin (typically 20 μL, 0.01 μg trypsin/μL) in 25 mM ammonium carbonate, pH 8.0. When necessary, ammonium carbonate buffer was added until the gel slices had recovered their original size; these were subsequently incubated at 37°C overnight. The protein fragments were extracted with 10–20 μL (depending on the size of the gel piece) of 50–75% acetonitrile, 5% TFA for 30 min.

2.6 Mass spectrometry

2.6.1 MALDI-TOF

The peptide mixtures from the tryptic digests (0.5 μL each) were crystallized in a matrix consisting of 100 mM α-cyano-4-hydrocynnamic acid prepared in 0.1% aqueous TFA. Molecular weight information of the peptides was obtained by using a MALDI-TOF mass spectrometer (Voyager-DE STR; PerSeptive Biosystems) equipped with a nitrogen laser and operating in reflector/delay extraction mode. All MALDI-MS spectra were internally calibrated using trypsin autodigestion peptides.

2.6.2 Nanoelectrospray-MS

The remaining peptide mixture from the extracted gel slices were purified using Poros 50 R2 beads in small Eppendorf vials as described [26]. The purified peptide mixture was extracted from the Poros beads with 50% methanol, 5% formic acid and the solution was loaded into a nanospray needle. The needle was transferred to the nanospray ion source mounted on the Q-Tof mass spectrometer (Micromass) and the sample analyzed by MS/MS.

2.7 Database searches

The most abundant ions in the MALDI spectra were directly used for database searches using the software developed in the UCSF Mass Spectrometry Facility to match known proteins [29]. Peptides from digestion were identified using the MS-Fit peptide mass fingerprinting search engine available on the Internet (http://prospector.ucsf.edu) to search genomic/proteomic databases as Genbank release 96 (http://www.ncbi.nlm.nih.gov/). The database search with MS-Fit was performed using the following values: all species, protein molecular mass range of 10–250 kDa, trypsin digest (one missed cleavage allowed), cysteines modified by carbamidomethylation, mass tolerance 30 ppm, oxidation of methionines, and acetylation of *N*-terminus of the protein. Ten to 15 molecular ions derived from the same MALDI spectra were used for protein identification. Only those matches with

scores higher than 70% were considered in this study. Some proteins with uncertain identities were further analyzed by nanoelectrospray MS to yield fragment ion-tag data. Database search with MS-Tag (http://prospector.ucsf.edu) were performed in non-error mode using the following values: all species, protein molecular mass range of 10–250 kDa, parent ion mass tolerance 1 Da, allowed fragment ions (type a, b, y, a-NH_3, b-NH_3, y-NH_3, and b-H_2O) and internal ions (trypsin digest; only one missed cleavage allowed).

3 Results and discussion

3.1 Two-dimensional gels of cerebellar membrane preparations

To test the potential advantages of different detergents and the chaotropic agents urea and thiourea in our system, equal amounts of cerebellar membrane preparation (3 mg protein) were solubilized in rehydration buffers containing different combinations of detergents and chaotropic agents, as described in Section 2.3.1. For each rehydration buffer, we tested the time required for complete solubilization by incubating the samples for 30, 60 or 90 min at 30°C. No significant differences were found between the three different incubation times and we concluded that 30 min are sufficient to achieve complete solubilization of the sample (data not shown). Two hundred and fifty μg of the cerebellar membrane preparations, differently solubilized, were then loaded onto pH 3–10 linear IPGs, and the protein mixture was separated by 2-DE as described above and detected by silver stain (see Fig. 1). When the extracts were solubilized in rehydration buffer containing NP-40 and urea, the resolution of gel spots, especially in the high molecular weight range, was quite poor (Fig. 1A). In contrast, when NP-40 was used in combination with urea plus thiourea, higher solubility and resolution was achieved (Fig. 1B). A similar effect was observed when NP-40 was substituted by the nonionic sulfobetain SB 3–10 (not shown). Pasquali *et al.* [30] have shown that the addition of 40 mM Tris, pH 9.6, to a rehydration buffer containing CHAPS, urea, and thiourea improves the solubilization of high molecular weight proteins. We used this method to solubilize the cerebellar membrane preparation (Fig. 1C). Our data showed that, indeed, the majority of proteins was efficiently resolved, with a clear improvement in the higher portion of the gel.

To monitor the presence of integral membrane proteins in our 2-DE system, we studied in detail the NR1 subunit of the NMDA receptor, a large hydrophobic protein present in the central nervous system. Cerebellar preparations were analyzed by immunoblotting by using a monoclonal antibody raised against the NR1 subunit. Interestingly, when the cerebellar proteins were solubilized in absence of thiourea (see Fig. 1A) and the proteins separated by 2-DE, transferred to a PVDF membrane and immunostained, no immunoreaction was found (data not shown).

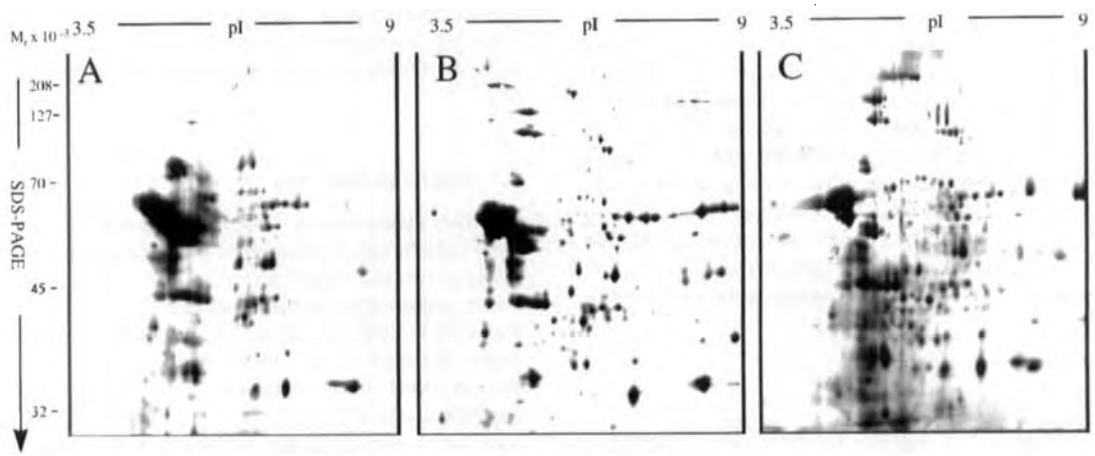


Figure 1. *Comparison of different solubilization conditions on porcine cerebellar membrane preparations by 2-DE.* An equal amount of sample (250 μg) was solubilized in (A) 8 M urea, 2% NP-40, 2% pH 3–10 carrier ampholytes, 0.3% DTT; (B) 7 M urea, 2 M thiourea, 4% NP-40, 0.3% DTT, 2% pH 3–10 carrier ampholytes; (C) 7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris, pH 9.6, 0.3% DTT, 2% pH 3–10 carrier ampholytes. The proteins were separated by their p*I* on 3–10 linear IPG strips and by their molecular weight on 10%T SDS-PAGE, and detected by silver staining. The panels represent only a portion of the original gels, containing the most significant differences (p*I* 3.5–9). Molecular weight markers are indicated on the left.

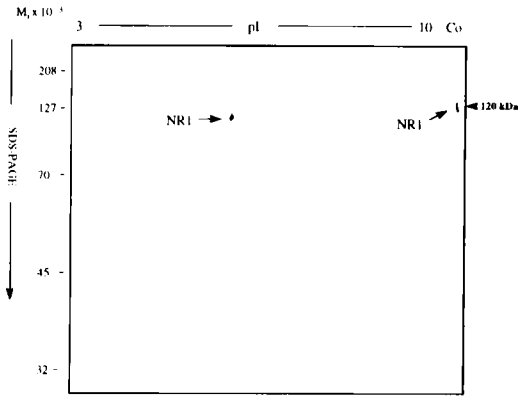


Figure 2. Detection of the NR1 subunit of the NMDA receptor from porcine cerebellar membrane preparation by Western blot on 2-DE. Cerebellar proteins separated by 2-DE were transferred onto a PVDF membrane for immunostaining. Immunodetection was performed by using a monoclonal antibody raised against the NR1 subunit of the NMDA receptor (1:1000). The arrow on the left side indicates the immunodetected spot, migrating at 120 kDa, in accordance with the predicted molecular weight of the NR1 subunit. On top of the same SDS-PAGE, 20 μg of solubilized cerebellum preparation was loaded separately and run in one dimension, as control (marked Co. in the figure). The arrow on the right side indicates an immunodetected band, corresponding to the NR1 subunit in the control sample and migrating with the same molecular weight as the spot identified by 2-DE.

However, after solubilization in presence of urea plus thiourea (independently of the detergent used), a spot at about 120 kDa, corresponding to the NR1 subunit of the NMDA receptor, was detected (Fig. 2). This demonstrated that the use of different solubilization methods reveals different protein populations, and that thiourea appeared to be a key factor for the detection of NR1 and probably for many other large hydrophobic proteins. In conclusion, best results were obtained when membranes were solubilized using CHAPS and Tris in combination with urea and thiourea, and these conditions were used in the following experiments.

3.2 Enrichment of membrane proteins in porcine cerebellar preparations

An alternative approach to facilitate the identification of proteins of low solubility (*e.g.*, large or hydrophobic) or of low abundance is to enrich the samples before analysis. It is well known that treatment of subcellular fractions with urea or salt followed by ultracentrifugation is an effective method to remove soluble proteins and peripherally attached membrane proteins from biological samples. To investigate if we could obtain an enrichment of large and integral membrane proteins in porcine cerebellum, the samples were treated with either 50 mM Tris-HCl, pH 7.6 (weak wash), 4.5 M urea (strong wash), or 3 M NaCl (strong wash) prior to solubilization, and the results were compared by immunoblotting and 2-DE. Figure 3 shows an immunoblot of one-dimensional SDS-PAGE of cerebellar membrane preparation after the three wash treatments described above. Immunostaining was performed using the anti-NR1 antibody. A significantly stronger immunoreaction was observed for the samples washed with a high concentration of urea and NaCl (Fig. 3, lane 2, 3) compared to the ones washed with 50 mM Tris-HCl (Fig. 3, lane 1). The same amount of protein was loaded in each lane. These experiments, which were performed repeatedly for different cerebellar membrane preparations, indicated a substantial enrichment in NR1 (and likely in other membrane proteins), in particular after NaCl wash. The poor detection of membrane-bound proteins by 2-DE is due to an under-representation of these proteins, mainly because of their low solubility, which causes protein aggregation during sample application and isoelectric focusing. An enrichment of membrane proteins in the sample is not sufficient *per se* to ensure an increased detection of this class of protein on the 2-D gels, but it is an alternative way to facilitate the study of membrane proteins.

To monitor the effects of NaCl wash on additional individual proteins, we analyzed these samples by 2-DE (Fig. 4). A 50 mM Tris-HCl wash (see Fig. 4A) did not cause significant alterations on the protein pattern compared to the untreated samples (data not shown). Changes could be observed after 4.5 M urea (see Fig. 4b) or 3 M NaCl (see Fig. 4C) treatment. Several proteins that could be clearly visualized by silver staining after a Tris wash were re-

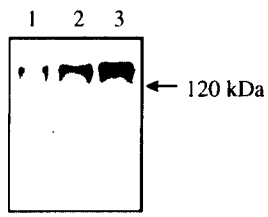


Figure 3. Effect of urea and salt treatments on the cerebellar membrane preparation monitored by Western blot. Cerebellar proteins (10 μg in each lane) were separated by one-dimensional SDS-PAGE and transferred onto a PVDF membrane. Immunodetection was performed by using a monoclonal antibody raised against the NR1 subunit of the NMDA receptor (1:1000). The arrow indicates the molecular mass of the immunodetected spot (120 kDa), consistent with the predicted molecular mass of the NR1 subunit.

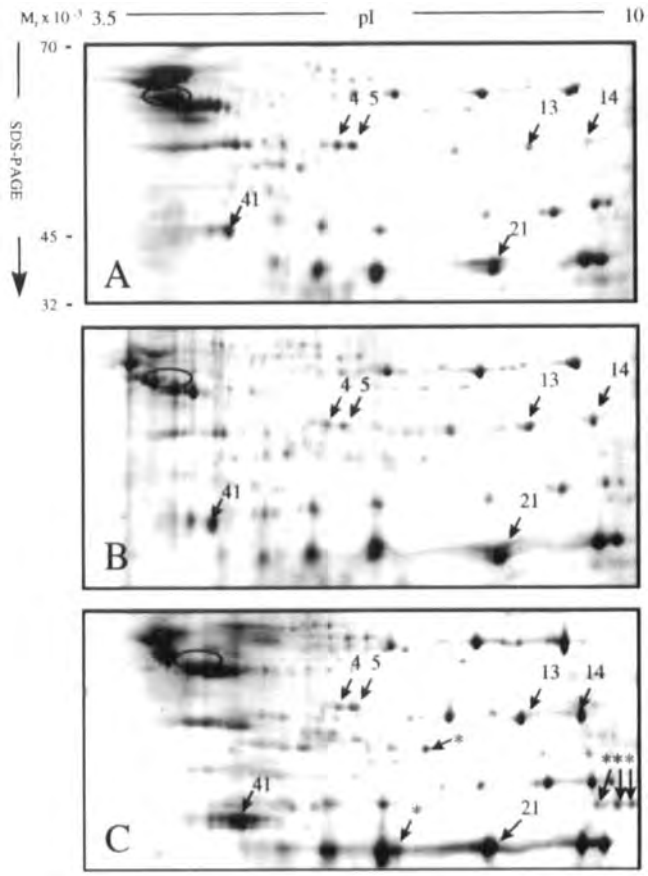


Figure 4. Effects of urea and salt treatments on cerebellar membrane preparations visualized by silver staining of 2-DE gels. Two-hundred and fifty μg of sample was pretreated with (A) 50 mM Tris-HCl, pH 7.6; (B) 4.5 M urea in 1 mM Tris-HCl, pH 7.6; or (C) 3 M NaCl in 1 mM Tris-HCl, pH 7.6, and analyzed by 2-DE. Samples were then solubilized in 7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris, pH 9.6, 0.3% DTT, 2% pH 3–10 carrier ampholytes, and loaded on pH 3–10 linear IPG strips for isoelectric focusing. In the second dimension proteins were separated on 12.5%T SDS-PAGE and detected by silver staining. The panels represent only a portion of the original 2-D gels. Indicated with arrows are spots with reduced (spot 4, 5 and circled area) or increased (spot 13, 14, 21) intensity. Spots marked with the asterisk (*) represent proteins that are specifically affected by NaCl wash. Molecular weight markers are indicated in the first panel of the figure. Arrows with numbers indicate spots that have been identified by mass spectrometry (see Table 1).

duced after a urea/NaCl wash (*e.g.*, spots 4 and 5 in Fig. 4). The protein in the circled area in Fig. 4A was identified as β-tubuline, a soluble protein, and was significantly reduced after the urea and NaCl wash (see Fig. 4B and C). Other protein spots became stronger (*e.g.*, spots 13, 14, and 21 in Fig. 4), indicating that they are present in higher concentrations. These differences seemed to be protein-specific, as NaCl wash was affecting a partially different population of proteins than the urea wash (see spots marked with an asterisk in Fig. 4C). These observations suggest that NaCl and urea washes of membranes can be used to achieve enrichment of specific proteins (probably integral membrane proteins) and to remove other proteins (probably soluble or peripheral membrane proteins). To investigate whether this conclusion is correct, we proceeded to identify individual proteins by mass spectrometric methods. Because of the significant enrichment of the NR1 subunit of the NMDA receptor after NaCl treatment (Fig. 3, lane 3), the following experiments were performed using this pretreatment.

3.3 Identification of cerebellar proteins from NaCl-washed fractions by mass spectrometry

The presence of a number of integral membrane proteins in our system was unambiguously established by performing the same experiments as above at a preparative scale to identify them by mass spectrometry. In the last decade mass spectrometry has emerged as the method of choice for protein identification. This approach can be used mainly in two ways: peptide fingerprinting and peptide sequencing to obtain fragment ion-tag data [15]. For peptide fingerprinting, the protein is cleaved by specific proteases (usually trypsin) into peptides [15, 31]. Molecular masses of such peptides are determined normally by either MALDI-TOF or by ES-MS. This information is used to search protein databases for protein identification. A second way to achieve protein identification is to obtain partial or complete sequence information by analyzing fragmentation spectra of peptides by tandem mass spec-

Table 1. Proteins identified from 2-D gel of pig cerebellum

Spot number/Protein identification	Species	Molecular mass (Da)	p*I*	Accession[a)]
1. ATP synthase α subunit	*Bos taurus*	59719.9	9.21	114402
2. ATP synthase α subunit	*Bos taurus*	59719.9	9.21	114402
3. ATP synthase α subunit	*Bos taurus*	59719.9	9.21	114402
4. Elongation factor tu	*Bos taurus*	49.398.5	6.72	1352352
5. Elongation factur tu	*Bos taurus*	49.398.5	6.72	1352352
6. Tubulin β-chain	*Sus scrofa*	49861.2	4.78	135490
7. Synaptojanin isoform beta lipid phosphatase	*Mus musculus*	127930.2	6.78	3241985
8. Glutamate receptor (GluR6)*)	*Mus musculus*	97808.2	6.44	423447
9. NADH dehydrogenase precursor 75 kDa subunit #)	*Homo sapiens*	79574.1	5.80	128826
10. KIAA0353 similar to myosin	*Homo sapiens*	150981	5.00	2224647
11. 79 KDa heat shock cognate protein*	*Bos taurus*	71239.9	5.49	123644
12. Leucocyte common antigen-related protein (LAR)	*Rattus norvegicus*	70949.76	6.51	693993
13. KIAA0445 protein fragment**)	*Homo sapiens*	148256.3	5.50	3413852
14. KIAA0445 protein fragment**	*Homo sapiens*	148256.3	5.50	3413852
15. Glyceraldehyde-3-phosphate dehydrogenase	*Sus scrofa*	35822.3	8.51	2506441
16. Glyceraldehyde-3-phosphate dehydrogenase #	*Sus scrofa*	35822.3	8.51	2506441
17. Glyceraldehyde-3-phosphate dehydrogenase	*Sus scrofa*	35822.3	8.51	2506441
18. Glyceraldehyde-3-phosphate dehydrogenase	*Sus scrofa*	35822.3	8.51	2506441
19. Voltage-dependent ion channel	*Homo sapiens*	31594.7	7.5	34612
20. Plasmalemmal porin brain-derived (BR1-VDAC)	*Bos taurus*	30825.8	8.83	1172552
21. Voltage-dependent anion channel 1 (VDAC1)	*Mus musculus*	30.755.6	8.62	130683
22. Voltage-dependent anion channel 1 (VDAC1) #	*Mus musculus*	30.755.6	8.62	130683
23. Voltage-dependent anion channel 1 (VDAC1) #	*Mus musculus*	30.755.6	8.62	130683
24. Porin (VDAC2)	*Homo sapiens*	38092.9	6.32	1172554
25. Porin (VDAC2) #	*Homo sapiens*	38092.9	6.32	1172554
26. Prohibitin	*Homo sapiens*	29820.2	5.57	464371
27. KIAA0336 protein fragment**	*Homo sapiens*	184659.4	5.07	2224613
28. Rieske iron-sulfur protein precursor	*Bos taurus*	29728.1	9.47	163044
29. α-B crystallin heat shock	*Rattus norvegicus*	20088.9	6.76	117388
30. M2-type pyruvate kinase family •	*Mus musculus*	57915.4	7.17	1363219
31. M2-type pyruvate kinase family • #	*Mus musculus*	57915.4	7.17	1363219
32. NADH dehydrogenase •	*Bos taurus*	23501.2	5.71	1364245
33. calpain 1 (calcium-activated neutral proteinase)	*Rabbit*	35274.7	6.45	115576
34. calpain 1 (calcium-activated neutral proteinase)	*Rabbit*	35274.7	6.45	115576
35. Tyrosine kinase receptor TY03 precursor fragment **	*Mus musculus*	96222.7	5.64	1717830
36. Phosphatidylinositol 3 kinase α subunit fragment **	*Homo sapiens*	83597.8	5.9	129387
37. Glutamate receptor β-2 fragment **	*Mouse*	100274.2	7.52	3287972
38. Tetratricopeptide repeat protein	*Homo sapiens*	55480.1	7.08	1688076
39. Rod cGMP phosphodiesterase β-subunit fragment **	*Bos taurus*	98311.4	5.22	116579
40. NMDAR receptor subunit 2A fragment **	*Rattus norvegicus*	165469.5	6.59	1072450
41. AHNAK nucleoprotein fragment **	*Homo sapiens*	180010.8	5.84	1351900
42. Gliar fibrillary acidic protein	*Homo sapiens*	49880.5	5.92	121135
43. Creatine kinase *	*Mus musculus*	42713.5	5.4	417208
44. Creatine kinase •	*Mus musculus*	42713.5	5.4	417208
45. Creatine kinase *	*Mus musculus*	42713.5	5.4	417208
46. Creatine kinase *	*Mus musculus*	42713.5	5.4	417208
47. β-tubulin	*Homo sapiens*	48880.1	4.7	2119276
48. α-tubulin	*Homo sapiens*	50157	4.86	2119276
49. Elongation factor 2 fragment **	*Rattus norvegicus*	95284.6	6.4	119176
50. Elongation factor 2 fragment **	*Rattus norvegicus*	95284.6	6.4	119176
51. KIAA0100 protein fragment **	*Homo sapiens*	237313.2	6.54	603949
52. F1 ATP synthase β-subunit precursor	*Bos taurus*	56283.8	5.15	114543
53. F1 ATP synthase β-subunit precursor	*Bos taurus*	56283.8	5.15	114543
54. F1 ATP synthase β-subunit precursor	*Rattus norvegicus*	51202.8	4.91	1374715

Table 1. continue

Spot number/Protein identification	Species	Molecular mass (Da)	p*I*	Accession[a)]
55. Gliar fibrillar acidic protein	*Bos taurus*	49452.7	5.3	2497271
56. F1 ATP synthase β-subunit precursor	*Bos taurus*	56283.8	5.15	114543

a) The accession code refers to the NCBI database
*) refers to protein family (when the information was not sufficient for assignment of a single protein)
**) refers to protein fragments. Molecular weight and p*I* values listed in the table refer to the entire protein
#) refers to proteins identified by sequence tag (listed in Table 2)

Table 2. Proteins identified by sequence tag from porcine cerebellum

Spot number/ Protein identification	Sequence tag	Mol. mass (m.u.)[a)]	a ions (m.u.)	b ions (m.u.)	b-H_2O ions (m.u.)	y ions (m.u.)
9. NADH dehydrogenase precursor	FASEIAGVDDLGTTGR	1608.78	120.08	148.8	288.13	175.12
			191.12	219.11	417.18	232.14
			278.15	306.15	530.26	333.19
			520.28	435.19	601.30	604.42
			747.40	548.27	658.32	719.47
				619.31		834.49
				676.33		933.59
						990.57
						1061.59
						1174.61
						1303.65
						1390.68
						1461.72
16. Glyceraldehyde-3-phosphate dehydrogenase	VPTPNVSVVDLTCR	1556.81	169.13	197.13	280.17	175.12
			270.18	298.18	590.33	335.15
			481.28	608.34	677.36	436.20
			580.35	695.37	990.53	549.28
			1093.63	1121.62	1103.61	664.31
				1382.70		763.38
						862.45
						949.48
						1048.55
						1162.59
						1259.64
						1360.69
						1457.74
22. Voltage-dependent anion channel 1 (VDAC1)	LTFDSSFSPNTGK	1400.66	187.14	215.14	197.13	147.11
			334.21	362.21	344.20	204.13
			536.27	477.23	459.22	305.18
			623.30	564.27	546.26	419.23
				651.30	633.29	516.28
				798.37	780.40	603.31
				885.40	867.39	750.38
						837.41
						924.44
						1039.47
						1186.54
						1287.59
23. Voltage-dependent anion channel 1 (VDAC1)	LTFDSSFSPNTGK	1400.66	187.14	215.14	197.13	147.11
			334.21	362.21	344.20	204.13
			536.27	477.23	459.22	419.23

Table 2. continue

Spot number/ Protein identification	Sequence tag	Mol. mass (m.u.)[a)]	a ions (m.u.)	b ions (m.u.)	b-H_2O ions (m.u.)	y ions (m.u.)
			1068.50	564.27	633.29	516.28
				651.30	867.39	603.31
				885.40		750.38
						837.41
						924.44
						1009.47
						1186.54
						1287.59
25 Porin (VDAC2)	LTFDTTFSPNTGK	1428.70	86.10	114.09	197.13	147.11
			187.14	215.14	459.22	204.13
			334.21	362.21	560.27	305.18
			449.24	477.23	661.32	419.23
			550.29	578.28		516.28
			651.34	679.33		603.31
			798.40	826.40		750.38
			982.49	913.43		851.43
						952.47
						1067.50
						1214.57
						1315.62
31. M2-type pyruvate kinase family	SLEAAHLAIDAGYR	1486.76	60.04	88.04	183.11	175.12
			302.17	201.12	312.16	395.20
			373.21	330.17	383.19	466.24
			444.25	401.20	591.29	581.27
			581.30	472.24	704.37	694.35
			694.39	609.30		765.39
				722.38		878.47
				793.42		1015.53
				906.50		1086.57
						1157.61
						1286.65
						1399.73

a) Molecular mass was calculated as monoisotopic masses (m.u., mass unit)

trometry (MS/MS) or by post-source decay on a MALDI-TOF. When tandem MS is performed, every molecular ion (representing a single peptide) can be selected and collisionally activated to produce individual fragment ions, which contain amino acid sequence information.

Both approaches allow a mixture of peptides to be analyzed. Seventy-four spots chosen from 2-D gels of NaCl-washed membranes (Table 1) – as well as nontreated membranes (data not shown) – were analyzed by MS using these two approaches. Protein spots were excised from the gels and treated with trypsin; the recovered peptide mixtures were analyzed by MALDI-TOF-MS and by ES-MS to yield peptide mass-fingerprint and fragment-ion data, respectively. Mass and sequence information thus obtained were used to search genome databases as described in Section 2.7. The most abundant ions in the spectra (10–15 molecular ions) were directly used for database searches using the program MS-Fit. For positive identification of a protein, a minimum of 70% of the used molecular ions was required to match. The identity of a number of spots was further investigated using sequence tag data obtained by nanoelectrospray MS/MS. Table 2 lists the sequence tags used for the identification of spots 9, 16, 22, 23, 25 and 31 with the program MS-Tag. For each sequence tag the corresponding observed a, b, b-H_2O and y ions used for the database search are indicated. Listed in Table 1 are 56 spots identified by spot numbers corresponding to the gel in Fig. 4C and 5. For the remaining 18 spots protein identification was not established, either because the ions in the MALDI spectra were not strong enough to have unambiguous identification through MS-Fit, or because the sequence tag did not match any translated nucleotide sequence in currently available databases. Since sequences of porcine proteins are not numerous in commonly available databases, most

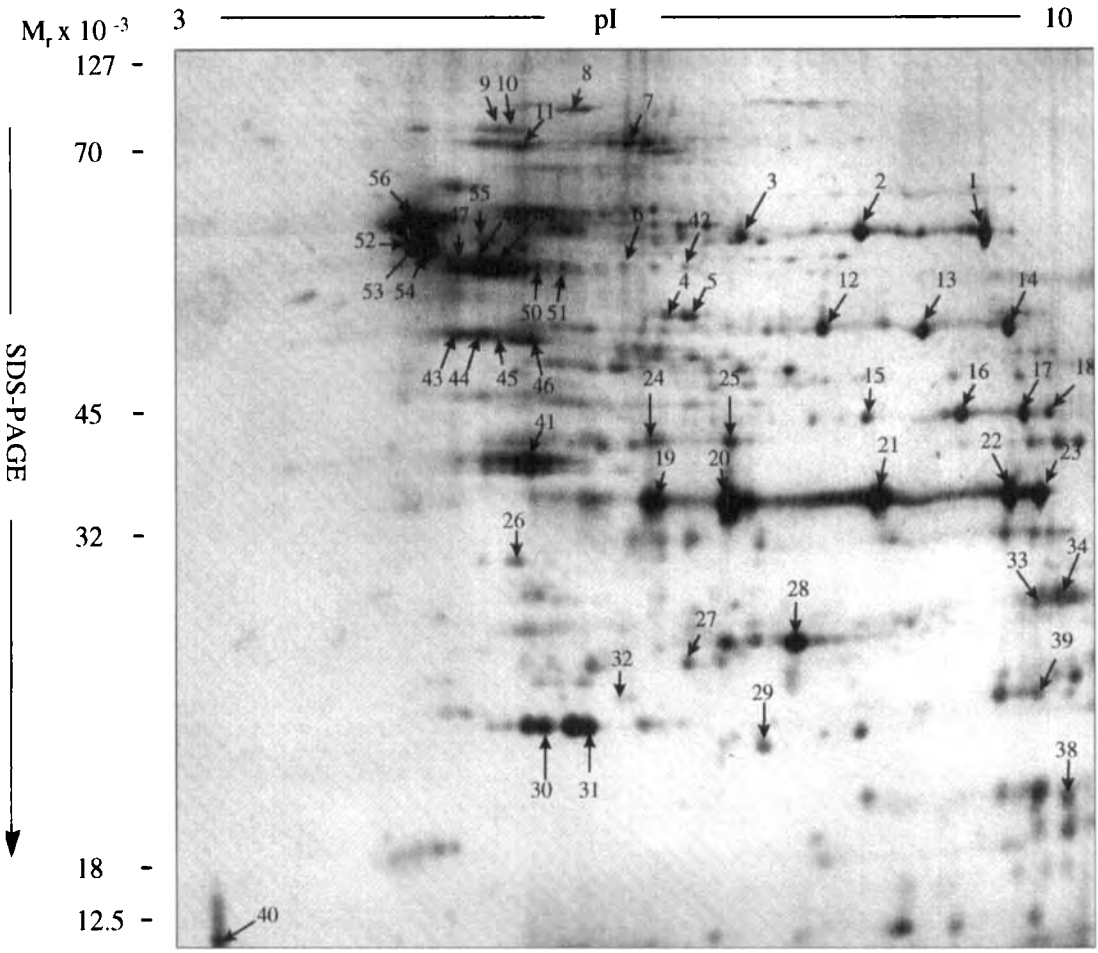


Figure 5. Analytical silver-stained 2-DE gel of porcine cerebellar membrane preparation after NaCl treatment. Two-hundred and fifty μg of sample was solubilized in 7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris, pH 9.6, 0.3% DTT, 2% pH 3–10 carrier ampholytes, and loaded on pH 3–10 linear IPG strips for isoelectric focusing. In the second dimension proteins were separated on a 12.5%T SDS-PAGE and detected by silver staining. Arrows indicate spots that have been identified by mass spectrometry (see Table 1).

of the proteins were identified in related species (most commonly *Homo sapiens* or *Bos taurus*). However, they all had high scores, possibly because of highly conserved sequences of such proteins among mammals.

In some cases (protein reported with an asterisk in Table 1) we could not determine the exact identity of the protein, but we could identify the protein family based on conserved homologous regions. In such cases we named the protein by their family name, without further specification. In other cases (*e.g.*, spot 1, 15, 43, 52 in Table 1) several spots were identified as one protein, probably representing modified forms of the same protein; however, this aspect was not further investigated in this study. A number of spots were identified as degradation fragments of larger proteins (indicated in Table 1 with two asterisks). Whether these fragments are generated *in situ* or during the sample preparation procedure is, as yet, unclear. In conclusion, 56 out of 74 spots analyzed could be identified from NaCl-washed cerebellar membranes.

4 Concluding remarks

In this study we presented a partial database of cerebellar proteins, which, to our knowledge, is the first to be reported. We combined different strategies to study membrane

proteins in particular. An enrichment for membrane proteins detected by 2-DE was achieved by pretreating the membrane preparation with high concentrations of NaCl, followed by solubilization in CHAPS and Tris in combination with urea and thiourea. This result was confirmed by the identification of protein spots by mass spectrometry. We found that spots increasing in intensity after the treatment were indeed membrane proteins (*e.g.*, spot 21 identified as a voltage-gated ion channel protein). Spots decreasing in intensity were soluble proteins (*e.g.*, spots 4, 5 identified as elongation factor tu, and spots in the circled area in Fig. 4A identified as β-tubuline). Fifty-six out of 74 spots chosen for MS analysis could be identified. Further development of this protein database is needed to establish a complete map of cerebellar proteins, which will become a useful tool to monitor experimentally induced changes of cerebellar proteins.

We thank Gunilla Brännström for providing the cerebellar preparations; we also thank Dr. Fredrik Nilsson, Dr. Ann-Christin Nyström and Helena Brockenhuus von Löwenhielm of the MS facility at Astra Hässle AB for their support and advice.

Received October 29, 1998

5 References

[1] Walsh, B. J., Gooley, A. A., Williams, K. L., Breit, S. N., *J. Leukocyte Biol.* 1995, *57*, 507–512.

[2] Herbert, B. R., Molloy, M. P., Jan, J. X., Gooley, A. A., Bryson, W. G., Williams, K. L., *Electrophoresis* 1997, *18*, 568–572.

[3] Cossio, G., Sanchez, J.-C., Wettestein, R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 548–552.

[4] Bjellqvist, B., Ek, K., Righetti, P. G., Giannazza, E., Görg, A., Westermeier, R., *J. Biochem. Biophys. Methods* 1982, *6*, 317–339.

[5] Görg, A., Postel, W., Weser, J., Gunther, S., Strahler, J. R., Hanash, S. M., Sommerlot, L., *Electrophoresis* 1987, *8*, 122–124.

[6] Yun, M., Wu, W., Hood, L., Harrington, M., *Electrophoresis* 1992, *13*, 1002–1013.

[7] Goldman, D., Merril, C., Ebert, M., *Clin. Chem.* 1980, *26*, 1371–1322.

[8] Hartinger, J., Stenius, K., Högemann, D., Jahn, R., *Anal. Biochem.* 1996, *240*, 126–133.

[9] Langen, H., Berndt, P., Röder, D., Cairns, N., Lubec, G., Fountoulakis, M., *Electrophoresis* 1999, *20*, 907–916.

[10] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[11] Vuillard, L., Marret, N., Rabilloud, T., *Electrophoresis* 1995, *16*, 295–297.

[12] Banks, M. D. G., Sandberg, M. P., Fowler, C. J., *Compl. Biochem. Physiol.* 1995, *111A 1*, 39–46.

[13] Monyer, H., Sprengel, R., Schoepfer, R., Herb, A., Higuchi, M., Lomeli, H., Burnashed, N., Sakmann, B., Seeburg, P. H., *Science* 1992, *256*, 1217–1221.

[14] Kutsuwada, T., Kashiwabuchi, N., Mori, H., Sakimura, K., Kushiya, E., Araki, K., Meguro, H., Masaki, H., Kumanishi, T., Arakawa, M., Mishina, M., *Nature* 1992, *358*, 36–41.

[15] Matsui, N. M., Smith, D. M., Clauser, K. R., Fichmann, J., Andrews, L. E., Sullivan, C. M., Burlingame, A. L., Epstain, L. B., *Electrophoresis* 1997, *18*, 409–417.

[16] Lindquist, J. A., Jensen, O. N., Mann, M., Hämmerling, J., *EMBO J.* 1998, *17*, 2189–2195.

[17] O'Connel, K. L., Stults, J. T., *Electrophoresis* 1997, *18*, 349–359.

[18] Karas, M., Hillenkamp, F., *Anal. Chem.* 1988, *60*, 2299.

[19] Tarach, E. J., Hines, W. M., Patterson, D. H., Juhasz, P., Falick, A. M., Vestal, M. L., Martin, S. A., *J. Protein Chem.* 1997, *16*, 363–369.

[20] Whitehouse, C. D., Dreyer, R. N., Yamashita, M., Fenn, J. B., *Anal. Chem.* 1985, *57*, 675–681.

[21] Mann, M., Wilm, M., *Trends Biochem. Sci.* 1995, *20*, 219–223.

[22] Harrington, C. R., *Anal. Biochem.* 1990, *186*, 285–287.

[23] Rabilloud, T., Vallette, C., Lawrence, J. J., *Electrophoresis* 1994, *15*, 1552–1558.

[24] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[25] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[26] Edvardsson, U., Alexandersson, M., Nilsson, F., Brockenhuus von Löwenhielm, H., Nyström, A. C., Ljung, B., Dahllöf, B., *Electrophoresis* 1998, *19*, 935–942.

[27] Lee, C., Levin, A., Branton, D., *Anal. Biochem.* 1987, *166*, 308–312.

[28] Barbato, R., Friso, G., Dalla Vecchia, F., Rigoni, F., Giacometti, G. M., *J. Cell. Biol.* 1992, *119*, 325–335.

[29] Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstain, L. B., Burlingame, A. L., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 5072–5076.

[30] Pasquali, C., Fialka, I., Huber, L. A., *Electrophoresis* 1997, *18*, 2573–2581.

[31] Cottrell, J. S., *Peptide Res.* 1994, *7*, 115–118, 120–124.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 928–934

Susi Greber[1]
Gert Lubec[1]
Nigel Cairns[2]
Michael Fountoulakis[3]

[1]University of Vienna, Department of Pediatrics, Vienna, Austria
[2]Institute of Psychiatry, Brian Bank, London, UK
[3]F. Hoffmann-La Roche, Basel, Switzerland

Decreased levels of synaptosomal associated protein 25 in the brain of patients with Down Syndrome and Alzheimer's disease

Synaptosomal associated protein 25 kDa (snap-25) is a widely distributed membrane-associated protein in the brain, mainly localized in nerve terminals. In nerve terminals, snap-25 participates in docking and/or fusion of synaptic vesicles with the plasmalemma, a process essential for synaptic vesicle exocytosis. Recent work suggests a role in brain development, forming presynaptic sites by regulating axonal outgrowth and nerve growth-induced neurite elongation. In Down syndrome (DS) brain, it is abnormally developed from early life, and brain pathology becomes even more pronounced when Alzheimer's disease (AD) develops in the fourth decade. This information led us to examine snap-25 in the brain of patients with DS and AD. We studied snap-25 and glial fibrillary acidic protein (GFAP) brain levels in five individual brain areas of 9 aged patients with DS, 9 patients with AD and 9 controls, applying two-dimensional gel electrophoresis. Decreased snap-25 levels were found in the five brain regions of the patients with DS and AD. Increased expression levels of GFAP were found in the frontal, parietal, temporal and occipital cortex regions of the DS and AD patients. Decreased snap-25 protein levels in the brain of DS and AD may reflect impaired synaptogenesis or represent neuronal loss. Findings of increased GFAP, a marker for neuronal loss, along with data from literature would support the notion of decreased snap-25 secondary to neuronal decay in both neurodegenerative disorders.

Keywords: Alzheimer's disease / Human brain proteins / Down syndrome / Glial fibrillary acidic protein / Synaptosomal associated protein 25 kDa (snap-25) / Two-dimensional electrophoresis
EL 3340

1 Introduction

Synaptosomal associated protein 25 kDa (snap-25) is a membrane-associated cytoplasmic, hydrophilic protein, 206 amino acids long. This phylogenetically highly conserved protein is widely distributed in the brain and localized mainly in nerve terminals. In the adult nervous system, it is palmitylated at one or more of its four closely spaced cysteines and it behaves as an integral membrane protein [1–3]. The concentration of snap-25 in nerve terminals led to the suggestion that it may participate in docking and/or fusion of synaptic vesicles with the plasmalemma. This process is essential for synaptic vesicle exocytosis [1] and there is now strong support for this protein function [4–6].

Correspondence: Michael Fountoulakis, F. Hoffmann - La Roche, Pharmaceutical Research-Gene Technology, Building 93-444, CH-4070 Basel, Switzerland
E-mail: michael.fountoulakis@roche.com
Fax: +41-61-691-9391

Abbreviations: AD, Alzheimer's disease; **DS**, Down syndrome; **GFAP**, glial fibrillary acidic protein; **snap-25**, synaptosomal associated protein 25

A challenging finding, independent of synaptic vesicle exocytosis, which stresses the role of snap-25 in brain development, was reported by Osten-Sand and co-workers [7]. The authors employed an antisense oligonucleotide approach to look into the function of snap-25 in developing neurons prior to synapse formation. Antisense oligonucleotides complementary to coding regions reduced snap-25 expression in cortical neurons in culture, resulting in decreased axonal outgrowth and nerve growth factor (NGF)-induced neurite elongation, thus indicating a reduction in the potential to form presynaptic sites. The relevance of this finding was confirmed by injecting antisense oligonucleotides in the eye of chick embryos; the consequences were impaired axonal growth of retinal neurons and reduction of the thickness of the internal plexiform neuronal layer. In Down syndrome (DS), brain weight is in the low normal range and the size of the cerebellum and brain stem may be reduced [8, 9]. Frontal-occipital length is shortened, probably secondary to reduced frontal lobe growth. There is a narrowing of temporal gyri in about one third of the cases [10] and the anterior commissure in adults with DS is reduced in cross-sectional area [11]. Nerve cell heterotopias in the white layers of the cerebellum indicate a disturbance of embryonic cell migration [12].

0173-0835/99/0405-0928 $17.50+.50/0

Abnormalities in morphology and number of dendritic spines including atrophy of the dendritic tree of the visual cortex, continuing into adulthood and becoming even more pronounced when Alzheimer's disease (AD) develops, have been described [13–16]. Other defects in brain histogenesis include poverty of granular cells throughout the cortex, decreased neuronal densities in layers II and IV of the occipital cortex [17] and diminution in the number of hypothalamic neurons [18]. There is also accumulating evidence for abnormalities of neuronal differentiation and abnormal migration in fetal and infant brains [8]. A consistent picture of the "wiring" of the brain in DS has not emerged yet.

In our studies on differential protein expression of brain proteins, by applying two-dimensional (2-D) gel electrophoresis, we observed decreased levels for snap-25 and increased levels for glial fibrillary acidic protein (GFAP) in the brain of patients suffering from DS and AD. Here we present a quantification of the proteins in the diseased states in comparison with the healthy state. Our findings may help to explain neuropathological features found in aged patients with DS.

2 Materials and methods

2.1 Brain samples

Post mortem brain samples were obtained from the MRC London Brain Bank for Neurodegenerative Diseases, Institute of Psychiatry. In all DS brains there was evidence of abundant beta A plaques and neurofibrillary tangles. The AD patients fulfilled the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD [19]. The histological diagnosis of AD was established and was consistent with the CERAD criteria [20] for a "definite" diagnosis of AD. The temporal, frontal, occipital, parietal cortex and cerebellum brain regions of karyotyped patients with DS (n = 9; 3 females, 6 males; 56.1 ± 7.1 years old), AD (n = 9; 6 females, 3 males; 72.3 ± 7.6 years old) and controls (n = 9; 5 females, 4 males; 72.6 ± 9.6 years old) [21], were used for the studies at the protein level. The controls were brains from individuals with no history of neurological or psychiatric illness. The major cause of death was bronchopneumonia in DS and AD patients and heart disease in controls. The *post mortem* interval of brain dissection in AD, DS, and controls was 34.1 ± 13.7, 30.6 ± 17.5 and 34.8 ± 15.0 h, respectively. Tissue samples were stored at −70°C and the freezing chain was never interrupted. The *post mortem* interval had a limited influence upon mRNA and protein products. Freezer storage time showed no effect [22].

2.2 Sample preparation

Brain tissue was suspended in 0.5 mL of sample buffer consisting of 40 mM Tris, 5 M urea (Merck, Darmstadt,Germany), 2 M thiourea, 4% CHAPS (Sigma, St.Louis, MO, USA) [23], 10 mM 1,4-dithioerythritol (Merck), 1 mM EDTA and a mixture of protease inhibitors, 1 mM PMSF, and 1 μg each of pepstatin A, chymostatin, leupeptin and antipain. The suspension was sonicated for approximately 30 s and centrifuged at 10 000 × g for 10 min to sediment cell debris. The supernatant was centrifuged further at 150 000 × g for 45 min to sediment undissolved material. The average protein concentration was 8 mg/mL.

2.3 Two-dimensional gel electrophoresis

The 2-D gel electrophoresis was performed essentially as reported [24]. Samples of approximately 1.5 mg were applied on pH 3–10 nonlinear immobilized pH-gradient (IPG) strips (Amersham Pharmacia Biotechnology, Uppsala, Sweden), at both the basic and acidic ends of the strips. The proteins were focused at 300 V for 1 h, after which the voltage was gradually increased to 3500 V within 6 h. Focusing was continued at 3500 V for 12 h and at 5000 V for 48 h. The proteins were then separated on 9–16% linear gradient polyacrylamide gels (with chemicals from Serva, Heidelberg, Germany and Bio-Rad, Hercules, CA, USA). After protein fixing with 40% methanol containing 5% phosphoric acid for 12 h, the gels were stained with colloidal Coomassie blue (Novex, San Diego, CA, USA) for 48 h. The molecular mass was determined by running standard protein markers at the right side of selected gels. The size markers (Gibco, Basel, Switzerland) covered the range of 10–200 kDa; the pI values were as given by the supplier of the IPG strips. The gels were destained with H_2O and scanned in a Molecular Dynamics personal densitometer. The images were processed using Adobe Photoshop and PowerPoint software. Protein spots were quantified using the ImageMaster 2D Elite software (Amersham Pharmacia Biotechnology).

2.4 Matrix-assisted laser desorption/ionization - mass spectroscopy

MALDI-MS analysis was performed as described [25] with minor modifications. Briefly, spots were excised, destained with 50% acetonitrile in 0.1 M ammonium bicarbonate, and dried in a Speedvac evaporator. The dried gel pieces were reswollen with 3 μL of 3 mM Tris-HCl, pH 8.8, containing 50 ng trypsin (Promega, Madison, WI, USA). After 15 min, 3 μL of H_2O were added and left at room temperature for 12 h. Two μL of 30% acetonitrile containing 0.1% trifluoroacetic acid were added, the content was vortexed, centrifuged for 3 min, and sonicated

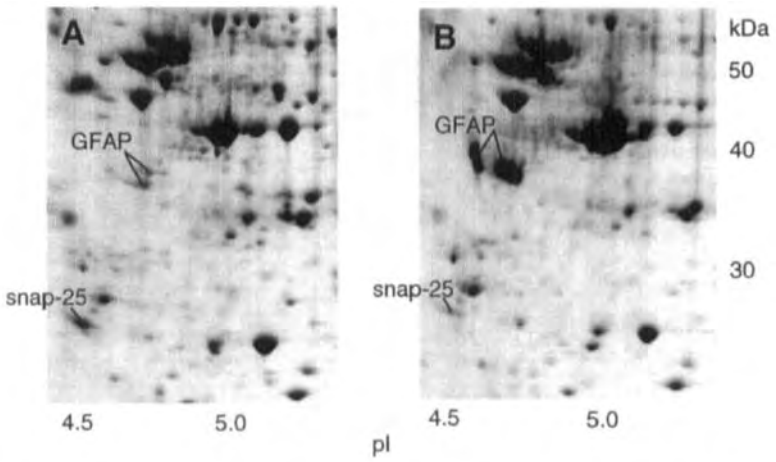


Figure 1. Partial two-dimensional gel images of human brain proteins from the parietal cortex lobe, from (A) a control, and (B) a patient with DS. The proteins were extracted and separated on 3–10 nonlinear IPG strips, followed by 9–16% SDS-polyacrylamide gels, as stated in Section 2.3. The gels were stained with Coomassie blue. The spots corresponding to snap-25 and GFAP are indicated.

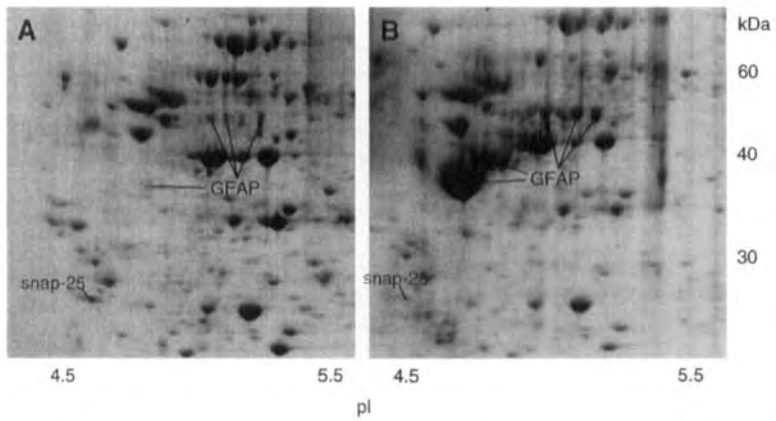


Figure 2. Partial two-dimensional gel images of human brain proteins from the parietal cortex lobe, from (A) a control, and (B) a patient with AD. The spots corresponding to snap-25 and GFAP are indicated. GFAP is represented by many spots located in two regions on the gel.

for 5 min. One μL was applied onto the dried matrix spot. The matrix solution consisted of 15 mg nitrocellulose (Bio-Rad) and 20 mg α-cyano 4 hydroxycinnamic acid (Sigma) dissolved in 1 mL acetone:isopropanol (1:1, v/v). The matrix solution (0.5 μL) was applied on the sample target. Specimen were analyzed in a time-of-flight Voyager Elite mass spectrometer (PerSeptive Biosystems, Cambridge, MA, USA) equipped with a reflectron. An accelerating voltage of 20 kV was used. Calibration was internal to the samples. The peptide masses were matched with the theoretical peptide masses of all proteins from all species of the SWISS-PROT database. For protein search, monoisotopic masses were used and a mass tolerance of 0.0075% was allowed.

3 Results

3.1 Quantification of snap-25 protein

The protein extracts from five brain regions of nine patients with DS, nine with AD and nine controls were separated by 2-D gel electrophoresis. The 2-D gels with proteins from the corresponding brain regions of the controls and the DS and AD patients were compared with each other. Several differences were observed, which could represent allelic differences or may be due to some irreproducibility in the separation of the brain proteins. Figure 1 shows the protein spots representing snap-25 in brain samples from a control (Fig. 1A) and a patient with Down Syndrome (Fig. 1B). Figure 2 shows the same protein in the brain sample from another control member (Fig. 2A) and a patient with Alzheimer's disease (Fig. 2B). The protein was unambiguously identified by MALDI-MS as synaptosomal associated protein 25 (SWISS-PROT accession number P13795). Homology to goldfish snap-25 a-protein (P36977) indicates that the human snap-25 protein is the a-form (data not shown). Snap-25 was in most cases represented by weak spots in all three (control, DS, and AD) groups. We quantified the spots representing snap-25 in the three groups in 2-D gels carrying proteins from the same brain region, using specific software. In

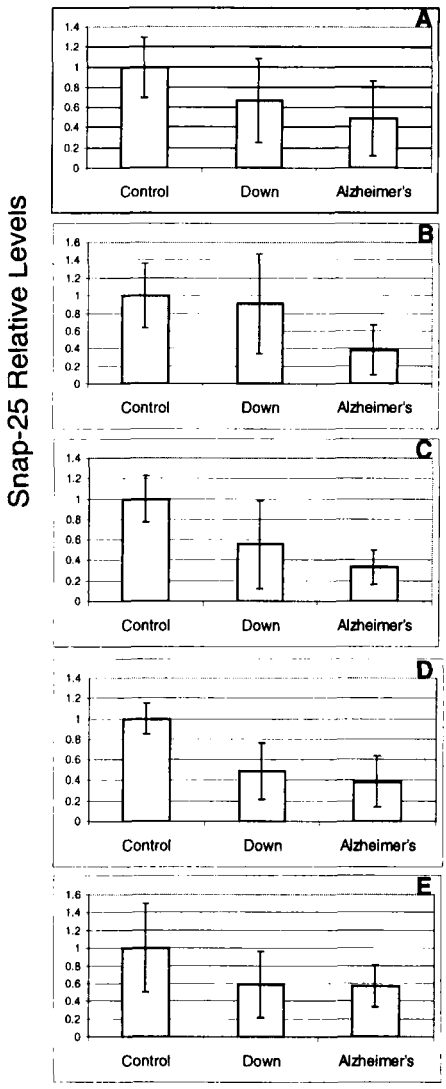


Figure 3. Quantification of snap-25 levels in the brain regions of patients with DS and AD. The proteins from (A) the frontal, (B) parietal, (C) temporal, and (D) occipital cortex lobes and (E) the cerebellum of the patients with DS and AD and the control group were separated on 2-D gels and visualized following staining with colloidal Coomassie blue. In partial gel images, including snap-25 and the neighboring proteins, the intensities of the spots representing snap-25 were quantified, compared to the total proteins present. The quantification was performed using the ImageMaster 2-D software. The mean values of snap-25 determined in the DS and AD groups were normalized to the mean value of the controls and the relative levels are indicated. The bars indicate the standard deviation of snap-25 levels in the three groups.

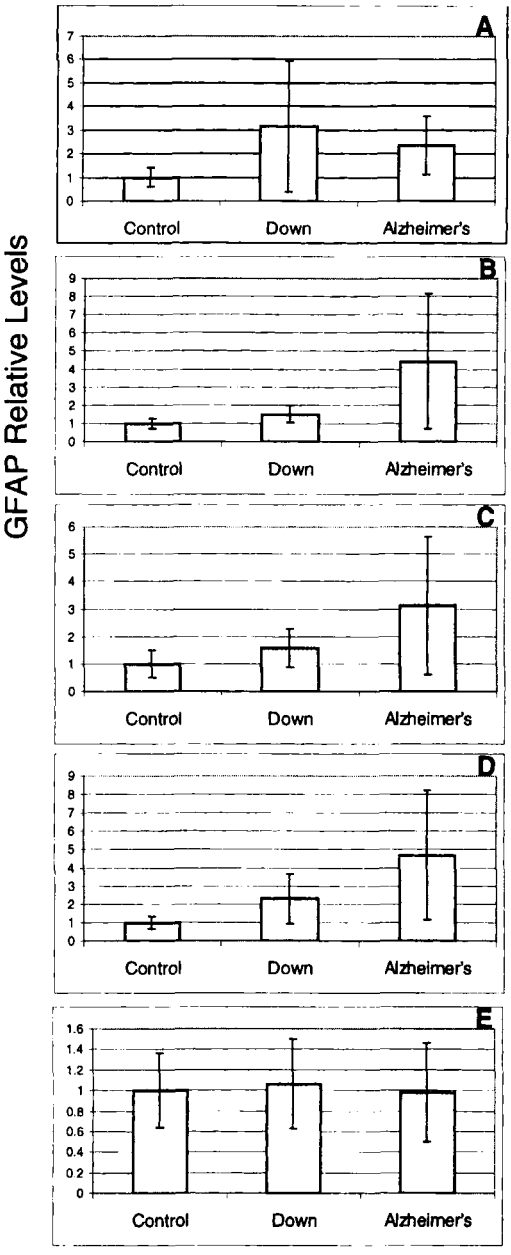


Figure 4. Quantification of GFAP levels in the brain regions of the patients with DS and AD. The spot quantification was performed as stated in the legend to Fig. 3. The relative protein levels in (A) the frontal, (B) parietal, (C) temporal, and (D) occipital cortex lobes and (E) the cerebellum of the patients with DS and AD are indicated, normalized to the mean value of the controls. The bars indicate the standard deviation of GFAP levels.

each group, snap-25 levels were determined in comparison with the levels of total proteins. The average snap-25 levels in the DS and AD groups were compared with the average protein levels determined in the control group (Fig. 3). Reduced levels of snap-25 were found in all brain regions of the DS and AD patients. In the patients with AD, the protein level reduction was even more evident in comparison with the DS patients.

3.2 Quantification of glial fibrillary acidic protein

GFAP was identified in the brain samples of the control, DS and AD groups. In the 2-D gel images of Figs. 1 and 2, GFAP is represented by weak spots in the control samples (Fig. 1A and 2A) and by large spots in the DS and AD samples (Fig. 1B and 2B). The protein was usually represented by many spots, located at two regions of the 2-D gels, migrating at about 35 and 50 kDa (Fig. 2A and 2B). The various spots were analyzed by MALDI-MS. The protein was unambiguously identified as GFAP (SWISS-PROT accession number P14136). Quantification of the spots representing GFAP revealed that in the frontal, parietal, temporal, and occipital lobes of cortex, the average levels were increased in the DS and AD patients, compared to the levels in the control group (Fig. 4). In cerebellum, no significant difference was observed in the three groups. The intensities of the spots representing GFAP showed a strong variation from group to group and from individual to individual. This resulted in increased values of standard deviation in the AD and DS groups (Fig. 4). Figure 5 shows examples of GFAP heterogeneity in six controls and six patients with AD. In general, more and larger GFAP spots were observed in the AD brain samples. It seems that in DS and AD brain, mainly the 35 kDa form increases, as seen in the examples of Fig. 5. In many patients, a 2- to 10-fold larger spot size was measured in the members of the DS and AD groups in comparison with the spot size in controls. The possible roles of the two GFAP forms and of their post-translational modifications in neurological and other disorders are currently under investigation.

4 Discussion

We studied differential protein expression in human brain in healthy and diseased individuals, suffering from DS and AD. In the DS and AD groups, we found reduced levels for snap-25 and increased levels for GFAP. Increased GFAP levels in the brain of patients with DS and AD and decreased levels for snap-25 in AD brain have been reported before [26, 27]. This study confirms those results, and in addition, it shows that snap-25 levels decrease in DS brain as well. Decreased snap-25 levels in the brain of aged DS patients can be due to increased neuronal loss or impaired synaptogenesis. Impaired synaptogenesis may appear either early in life or secondarily when AD lesions, tangles and plaques appear in DS brains beginning with the fourth decade of life [28]. However, our concomitant finding of increased GFAP, a marker for neuronal decay and brain damage, and decreased snap-25 would suggest that decreased snap-25 protein might well be

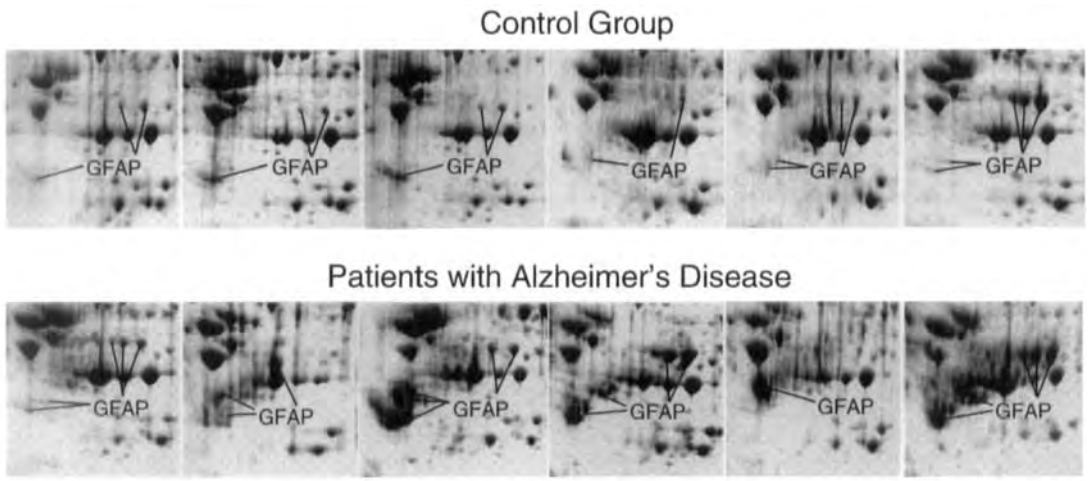


Figure 5. Partial two-dimensional images of brain protein samples from 6 controls and 6 patients with AD. The images show examples of glial fibrillary acidic protein levels in the brains of the two groups. In general, larger spots representing GFAP, particularly the 35 kDa form, were observed in the AD samples. The gels were prepared as stated in the legend to Fig. 1. The proteins were identified by MALDI-MS.

linked to neuronal loss. The report of Goodison and co-workers [29] describing decreased GFAP levels in brains of DS patients 15–45 years of age does not contradict our concept and results because AD pathology develops in the fourth decade of age. The authors studied GFAP-mRNA rather than protein levels. We do not agree, however, with their suggestion that trisomy 21 exerts a suppressive effect on GFAP gene expression.

Apart from its role for axonal growth and elongation, reduced snap-25 may be leading to deranged functions of exocytosis and neurotransmission known to occur in DS [8], including our study population [30–32]. Snap-25 is an integral constituent of the synaptic core complex, which mediates vesicle docking and membrane fusion and interacts with calcium channels [33–36]. Also, in the brains of patients with AD, decreased snap-25 may reflect secondary neuronal loss. This notion is supported by the report from Clinton and co-workers [37] showing that snap-25 and the relative synaptic index in cortical areas from AD brains are correlated. Shimohama and co-workers [27] confirmed the finding of decreased snap-25 in brains of patients with AD. However, Dessi *et al.* [38] did not find a correlation between snap-25 and markers of neuronal death. We are currently studying snap-25 expression in fetal brains of patients with DS in order to address the open question of the time course of the snap-25 protein decrease.

The work at the University of Vienna was supported by the Red Bull Company, Salzburg, Austria. We thank J.-F. Juranville for technical assistance.

Received October 31, 1998

5 References

[1] Oyler, G. A., Higgins, G. A., Hart, R. A., Rattenberg, E., Billingsley, M., Bloom, F. E., Wilson, M. C., *J. Cell. Biol.* 1989, *109*, 3039–3052.

[2] Bark, I. C., Hahn, K. M., Ryabinin, A. E., Wilson, M. C., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 1510–1514.

[3] Hess, D. T., Slater, T. M., Wilson, M. C., Pat Skene, J. H., *J. Neurosci.* 1992, *12*, 4634–4641.

[4] Soellner, T. , Whiteheart, W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., Rothman, J. E., *Nature* 1993, *362*, 318–323.

[5] Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino-de Laveto, P., DasGupta, B. R., Monecucco, C., *Nature* 1992, *359*, 832–835.

[6] Link, E., Edelmann, L., Chou, J. H., Binz, T., Yamasaki, S., Eisel, V., Baumert, M., Suthof, T. C., Niemann, H., Jahn, R., *Biochem. Biophys. Res. Commun.* 1992, *189*, 1017–1023.

[7] Osten-Sand, A., Catsicas, M., Staple, J. K., Jones, K. A., Ayala, G., Knowles, J., Grenningloh, G., Catsicas, S., *Nature* 1993, *364*, 445–448.

[8] Epstein, C. J., in: Scriver, C. R., Beaudet, A. L., Sly, W. S., Valle, D. (Eds.), *The Metabolic and Molecular Basis of Inherited Disease*, McGraw Hill, New York 1992, pp. 749–794.

[9] Crome, L., Cowie, V., Slater, E., *J. Ment. Defic. Res.* 1966, *10*, 69–77.

[10] Schmidt-Sidor, B., Wisniewski, K. E., Shepard, T. H., Sersen, A., *Clin. Neuropathol.* 1990, *9*, 181–195.

[11] Sylvester, P. E., *J. Ment. Def. Res.* 1986, *30*, 19–25.

[12] Rehder, H., in: Burgio, G. R., Fraccaro, M., Tiepolo, L., Wolf, U. (Eds.), *Trisomy 21: An International Symposium*, Springer Verlag, Berlin 1981, p. 57.

[13] Marin-Padilla, M., *J. Comp. Neurol.* 1976, *167*, 63–75

[14] Becker, L. E., Armstrong, D. L., Chan, F., *Ann. Neurol.* 1986, *20*, 520–532.

[15] Becker, L., Mito, T., Takashima, S., Onodera, K., in: Epstein, C. J. (Ed.), *The Morphogenesis of Down Syndrome*, Wiley-Liss, New York 1991, p. 133.

[16] Takashima, S., Ieshima, A., Nakamura, H., Becker, L. E., *Brain Dev.* 1989, *2*, 131–143

[17] Wisniewski, K. E., Laure-Kamionowska, M., Connell, F., Wen, G. Y., in: Epstein, C. J. (Ed.), *The Neurobiology of Down Syndrome*, Raven, New York 1986, p. 29.

[18] Wisniewski, K. E., Bobinski, M., in: Epstein, C. J. (Ed.), *The Morphogenesis of Down Syndrome*, Wiley-Liss, New York 1991, p. 153.

[19] Mirra, S. S., Heyman, A., McKeel, D., Sumi, S., Crain, B. J., *Neurology* 1991, *41*, 479–486.

[20] Tierney, M. C., Fisher, R. H., Lewis, A. J., Torzitto, M. L., Snow, W. G., Reid, D. W., Nieuwstraaten, P., Van Rooijen, L. A. A., Derks, H. J. G. M., Van Wijk, R., Bischop, A., *Neurology* 1988, *38*, 359–364.

[21] Seidl, R., Greber, S., Schuller, E., Bernert, G., Cairns, N., Lubec, G., *Neurosci Lett.* 1997, *235*, 137–140.

[22] Harrison, P. J., Heath, P. R., Eastwood, S. B., Burnet, P. W., McDonald, B., Pearson R. C., *Neuroscience Lett.* 1995, *200*, 151–154.

[23] Chevallet, M., Santoni, V., Poinas, A., Rouquiè, D., Fuchs, A., Kiefer, S., Rossignol, M., Lunardi, M., Garin, J., Rabilloud, T., *Electrophoresis* 1998, *19*, 1901–1909.

[24] Langen, H., Roeder, D., Juranville, J.-F., Fountoulakis, M., *Electrophoresis* 1997, *18*, 2085–2090.

[25] Fountoulakis, M., Langen, H., *Anal. Biochem.* 1997, *250*, 153–156.

[26] Jorgensen, O. S., Brooksbank, B. W., Balazs, R., *J. Neurol. Sci.* 1990, *98*, 63–79.

[27] Shimohama, S., Kamiya, S., Taniguchi, T., Akagawa, K., Kimura, J., *Biochem. Biophys. Res. Commun.* 1997, *236*, 239–242.

[28] Burger, P. C., Vogel, F. S., *Am. J. Pathol.* 1973, *73*, 457–476.

[29] Goodison, K. L., Parhad, I. M., White, C. L., Sima, A. A., Clark, A. W., *J. Neuropathol. Exp. Neurol.* 1993, *52*, 192–198.

[30] Risser, D., Lubec, G., Cairns, N., Herrera-Marschitz, M., *Life Sci.* 1991, *60*, 1231–1237.

[31] Baran, H., Cairns, N., Lubec, B., Lubec, G., *Life Sci.* 1996, *58*, 1891–1899.

[32] Schneider, C., Risser, D., Kirchner, L., Kitzmueller, E., Cairns, N., Prast, H., Singewald, N., Lubec, G., *Neurosci. Lett.* 1997, *222*, 183–186

[33] Sheng, Z.-H., Rettig, J., Cook, T., Catterall, W. A., *Nature* 1996, *379*, 451–454.

[34] Boschert, U., O'Shaughnessy, C., Dickinson, R., Tessari, M., Bendotti, C., Catsicas, S., Pich, E. M., *J. Comp. Neurol.* 1996, *367*, 177–193.

[35] O'Connor, V., Heuss, C., DeBello, W. M., Dresbach, T., Charlton, M.P., Hunt, J.H., Pellegrini, L. L., Hodel, A., Burger, M. M., Betz, H., Augustine, G. J., Schaefer, T., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 12186–12191.

[36] Linial, M., Ilouz, N., Parnas, H., *J. Physiol. London* 1997, *504*, 251–258.

[37] Clinton, J., Blackman, S. E., Royston, M. C., Roberts, G. W., *Neuroreport* 1994, *5*, 497–500.

[38] Dessi, F., Colle, M. A., Hauw, J. J., Duyckaerts, C., *Neuroreport* 1997, *8*, 3685–3689.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 935–942

Ulrika Edvardsson[1]
Maria Alexandersson[1]
Helena Brockenhuus von Löwenhielm[2]
Ann-Christin Nyström[2]
Bengt Ljung[3]
Fredrik Nilsson[2]
Björn Dahllöf[1]

[1]Cell Biology & Biochemistry
[2]Bioanalytical Chemistry
[3]Cardiovascular Pharmacology
Astra Hässle AB, Mölndal, Sweden

A proteome analysis of livers from obese (*ob/ob*) mice treated with the peroxisome proliferator WY14,643

The PPAR (peroxisome proliferator activated receptor) transcription factors are ligand-activated receptors which regulate genes involved in lipid metabolism and homeostasis. PPARα is preferentially expressed in the liver and PPARγ preferentially in adipose tissue. Activation of PPARα leads to peroxisome proliferation in rodents and increased β-oxidation of fatty acids. PPARγ-activation leads to adipocyte differentiation and improved insulin signaling of mature adipocytes. Both of these PPAR receptors are potential targets for treatment of dyslipidemia in man. Studies by others using a proteomics approach have characterized the effects of PPARα agonists in livers from lean healthy mice. However, we wanted to map the effects of a therapeutic dose of a PPARα agonist in a disease model of insulin resistance and diabetes, the obese diabetic *ob/ob* mouse, by proteomics. Therefore, *ob/ob* mice, which have highly elevated levels of plasma triglycerides, glucose and insulin, were treated for one week with WY14,643 (180 μmol/kg/day), a well-characterized selective PPARα agonist. Plasma triglycerides, glucose and insulin levels were determined and we found significant therapeutic effects on triglycerides and glucose levels. The liver protein compositions were investigated by high-resolution two-dimensional gel electrophoresis which showed that WY14,643 produced up-regulation of at least 16 spots. These were identified by mass spectrometry and 14 spots were found to be components of the peroxisomal fatty acid metabolism. Thus, WY14,643 at a therapeutic dose, caused induction of peroxisomal fatty acid β-oxidation in obese diabetic mice.

Keywords: Peroxisome proliferation / Two-dimensional polyacrylamide gel electrophoresis / Mass spectrometry / Peroxisome proliferator activated receptor EL 3408

1 Introduction

Peroxisome proliferator activated receptors (PPAR) are nuclear transcription factors that heterodimerize with RXRα and activate a multitude of genes involved in lipid metabolism [1–3]. There are three PPAR receptors known to date with different tissue distribution; PPARα is highly expressed in liver and kidney, PPARδ is ubiquitously expressed and PPARγ is preferentially expressed in adipose tissue [1, 4–6]. It has been shown that fatty acids and/or fatty acid metabolites are activators of PPARs [7–11] and it is possible to view PPARs as physiological sensors of intracellular lipid/fatty acid concentrations [1, 2, 9]. Activation of PPAR in rodent livers promotes increased fatty acid oxidation in peroxisomes and mitochondria, and induce proliferation of peroxisomes [4, 12, 13]. Also, several other proteins involved in lipid metabolism are regulated by PPARα, such as lipoprotein lipase, apolipoprotein AI, AII and CIII, and cytochrome P_{450}4A [1, 14, 15]. Fibrates, used in the clinic to treat severe hypertriglyceridemia, have been found to activate PPARα [4, 12, 16] and in man, major effects are thought to be through down-regulation of apolipoprotein CIII [2, 17, 18], an inhibitor of peripheral lipolysis. Peroxisome proliferation, as a result of PPARα activation, is considered a rodent-specific phenomenon and has not been convincingly detected in man [19].

Here, we have examined the effects in livers from obese diabetic mice (*ob/ob*) of the well-characterized highly selective PPARα agonist WY14,643 [4]. The methods employed were high-resolution two-dimensional electrophoresis (2-D PAGE) followed by protein identification using mass spectrometry (MS) and database searching. A previous study using the "proteomics approach" has characterized the liver effects of peroxisome proliferators [20]; however this was performed in lean, healthy B6C3F1 mice. Our results, which show that a therapeutic

Correspondence: Dr. Björn Dahllöf, Cell Biology and Biochemistry, Astra Hässle AB, S-43183 Mölndal, Sweden
E-mail: bjorn.dahllof@hassle.se.astra.com
Fax: +46-31-7763736

Abbreviations: ACO, acyl CoA oxidase; **CYP4A**, cytochrome P_{450}4A; **FABP**, fatty acid binding protein; **HMG CoA S**, HMG CoA synthase; **3KCT**, 3-ketoacyl-CoA thiolase; **PBE**, peroxisomal bifunctional enzyme; **PPAR**, peroxisome proliferator activated receptor

dose of WY14,643 causes induction of peroxisomal β-oxidation in livers from *ob/ob* mice, will constitute the basis for further proteomics studies on the effects of both PPARα and PPARγ agonists in livers from obese diabetic mice.

2 Materials and methods

2.1 Materials

The IsoDalt system, including gel cassettes and casting box, was from Hoefer (San Francisco, CA, USA). Equipment for isoelectric focusing (IEF), Multiphor II, Immobiline DryStrip Kit, Immobiline DryStrips (18 cm, pH 3–10 NL; nonlinear), IPG buffer 3–10 NL and Drystrip cover fluid were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). Iodoacetamide and sodium thiosulfate were from Sigma (St Louis, MO, USA). Trypsin was purchased from Promega (Madison, WI, USA). Silver nitrate, copper chloride, formaldehyde and sodium carbonate were from Merck (Darmstadt, Germany). NP-40 was from United States Biochemical (Cleveland, OH, USA). GelCode blue stain reagent was from Pierce (Rockford, IL, USA). Duracryl (30%, 0.65% Bis) and all other chemicals, unless otherwise stated, were electrophoresis grade (ESA Chelmsford, MA, USA). WY14,643 was from Sigma.

2.2 Animals and drug treatment

Obese *ob/ob*-mice (Umeå-strain) and lean litter mates were from Bomholtgård Breeding and Research Center, Denmark. Following one week of acclimatization in the animal quarters, the 7-week-old animals were divided into groups of five mice and one group of obese mice was treated orally once daily for one week with WY14,643 (180 μmol/kg/day). Lean mice (untreated) served as control for determination of physiologically normal levels of triglycerides, insulin and glucose and were not analyzed further in this study. On the last day of the experiment the mice were anesthetized with CO_2 and exsanguinated *via* a carotid artery. The blood was collected in EDTA vials. The apical end of the left lobe of the liver was rapidly removed and snap-frozen in liquid nitrogen less than 1 min after blood sampling. Liver samples were stored at –150°C.

2.3 Determination of plasma insulin, triglyceride and glucose levels

Insulin levels were determined using a rat insulin RIA kit (RI-13K Linco, St. Louis, MO, USA). Triglycerides and glucose levels were determined spectrophotometrically using a Cobas Mira plus (Hoffman la Roche, Basel, Switzerland) with Calibrator Human (07 3718 6; Roche) as calibrant. For triglycerides and glucose the enzymatic kits used were "Triglycerides/Glycerol Blanking" (450032; Boehringer Mannheim, Indianapolis, IN, USA) and "Glucose HK" (07 3672 4; Roche), respectively.

2.4 2-D PAGE sample preparation

The apical liver lobes were weighed and homogenized with a glass/Teflon homogenizer (five strokes at 400 rpm) in eight volumes of solubilizing solution (8 M Urea, 0.3% w/v DTT, 2% v/v NP-40 and 2% v/v IPG buffer 3–10 NL). To remove solid tissue, the homogenates were centrifuged at 100 000 × *g* for 30 min at 15°C. The supernatant was removed and frozen immediately at –70°C.

2.5 2-D PAGE – first dimension (IEF)

Immobiline DryStrips (18 cm, pH 3–10 NL) were used for isoelectric focusing. Prior to IEF, each strip was rehydrated in 400 μL of rehydration solution containing 4 μL (~ 100 μg) of solubilized liver protein for analytical gels and 10–15 μL (~ 250–375 μg) for preparative gels. The rehydration solution consisted of 8 M urea, 2% v/v NP-40, 0.3% w/v DTT and 2% v/v IPG buffer 3–10 NL. Rehydration was allowed to proceed overnight to ensure maximal diffusion of the proteins within the strip. The strips were then run under a layer of Drystrip cover fluid, at 20°C, in a Multiphor II unit using the Immobiline DryStrip Kit according to the manufacturer's instructions. Focusing was carried out at 0–500 V for 1 min, 500 V for 5 h, 500–3500 V for 5 h and 3500 V for ~ 12 h to reach a total of 50–60 kVh.

2.6 Equilibration of IEF gel strips

Following IEF the Drystrip cover fluid was poured off and the strips were equilibrated 2 × 15 min with gentle shaking [21]. The first equilibration solution contained 30% glycerol w/v, 6 M urea, 2% SDS, 50 mM Tris-HCl, pH 8.8, 65 mM DTT, and a trace of bromophenol blue as tracking dye. The second equilibration was carried out in the same solution, except that DTT was replaced by 260 mM iodoacetamide.

2.7 2-D PAGE – second dimension (SDS-PAGE)

The gels used in this study were continuous 14%T, 0.3%C gels (23 × 20 × 0.1 cm). After the equilibration, each IEF strip was drained on a filter paper and immersed in SDS running buffer (24 mM Tris base, 0.2 M glycine and 0.1% SDS) before it was sealed at the top of the second-dimensional gel with 1% agarose in SDS running buffer.

Electrophoresis was performed in the IsoDalt tank (Hoefer) at 100 V for ~ 19 h, until the tracking dye reached the anodic end of the gels.

2.8 Silver staining

Analytical gels were silver-stained according to Shevchenko *et al.* [22] with some modifications. The gels were fixed in 50% ethanol, 5% acetic acid for 1 h, washed in 50% ethanol for 30 min and additionally 60 min in water to remove the remaining acid. Thereafter, the gels were sensitized by a 1 min incubation in 0.02% sodium thiosulfate and rinsed with two changes of distilled water for 1 min each. After rinsing, the gels were incubated in 0.1% silver nitrate for 30 min. Following incubation, the gels were rinsed twice with distilled water for 1 min and then developed in 0.04% v/v formaldehyde, 2% w/v sodium carbonate. When the developer turned yellow (~ 30 s) it was discarded and replaced with fresh solution. When the desired intensity of staining was achieved (~ 3.5 min), the development was terminated by discarding the reagent, followed by washing with 5% acetic acid for 5–10 min. Finally, the gels were washed in water and stored in sealed plastic bags at 4°C.

2.9 Copper staining

To produce a negative image of colorless protein spots against a semi-opaque background some preparative gels were copper-stained according to Lee *et al.* [23]. Since the proteins are not permanently fixed within the gel, this method was used when further analysis of the proteins by MS was required. After electrophoresis the gels were rinsed in water for 3 min, thereafter immersed in 5% w/v $CuCl_2$ solution and gently shaken in the copper solution for 2–3 min until a pattern of clear protein spots against a semi-opaque whitish-blue background appeared. Protein spots for MS analysis were excised from the gels, transferred to Eppendorf tubes, and washed for ~ 2 h with 24 mM Tris base, 0.2 M glycine and then rinsed briefly with water.

2.10 Gelcode blue staining

Some preparative gels were stained with Gelcode blue stain reagent, which utilizes the colloidal properties of Coomassie G-250 for protein staining on polyacrylamide gels. As subsequent analysis of protein spots from these gels was performed, no fixation was done prior to staining. After electrophoresis the gels were rinsed with water 3 × 10 min to remove SDS, then incubated in Gelcode blue stain reagent for approximately 2 h followed by water equilibrium/enhancement for 1–2 h before excision of protein spots to be identified.

2.11 In-gel digestion and peptide extraction

In-gel digestion was performed according to Rosenfeld *et al.* [24] with minor modifications. The gel pieces were washed in water for 1 h, then in 30–40% acetonitrile, 25 mM ammonium carbonate, pH 8.0 (1–3 times, approximately 20 min) in pre-siliconized Eppendorf tubes (Axygen, Hayward, CA, USA) and finally dehydrated in a Speedvac vacuum evaporator (approximately 30 min). The gel pieces were reswollen with a trypsin solution containing 0.1–0.2 µg Promega modified trypsin (typically 20 µL, 0.01 µg trypsin/µL) in 25 mM ammonium carbonate, pH 8.0. If necessary, additional ammonium carbonate buffer was added until the gel pieces had recovered their original size. After incubation at 37°C overnight, the

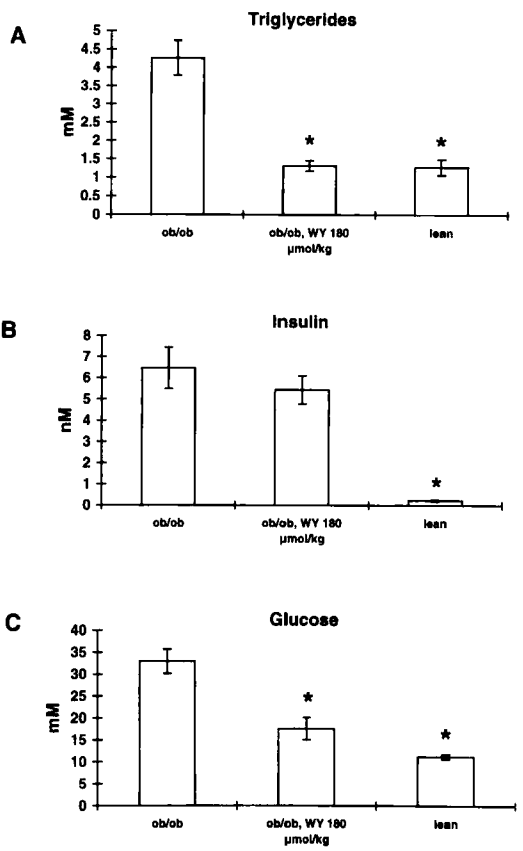


Figure 1. Therapeutic effects of WY14,643. Plasma levels of triglycerides, insulin and glucose in *ob/ob* mice after one week of treatment with WY14,643 (180 µmol/kg/day), in comparison to untreated obese controls and untreated lean litter mates. Values are mean ± SEM, $n = 5$ except for lean mice where $n = 4$. *, $P < 0.05$ using Student's *T*-test for each group compared with nontreated *ob/ob* mice.

lids of the tubes were carefully opened and condensed water was removed from under the lid. The protein fragments were extracted by addition of 10–20 μL (depending on the size of the gel piece) 50–75% v/v acetonitrile and 5% v/v TFA and 30 min incubation.

2.12 MALDI-TOF analysis

Peptide extracts, 0.5 μL, were mixed with 0.5 μL matrix solution (α-cyano-4-hydroxycinnamic acid; Hewlett-Packard, Böblingen, Germany) on the metal MALDI-TOF target and dried for 10 min. Internal mass calibration was performed using trypsin autodigestion products (842.51 Da and 2211.11 Da). The MALDI-TOF analysis were performed on a PerSeptive Biosystems STR mass spectrometer (Framingham, MA) in reflector mode.

2.13 Electrospray MS/MS analysis

Remaining peptide extracts were purified using Poros 50 R2 packing material (PerSeptive Biosystems). The extracts were diluted 10-fold in 1.6 mL Eppendorf tubes and a 10 μL Poros 50 R2 suspension was added (20 mg/mL in water). After one h incubation at room temperature the tubes were centrifuged and the supernatants removed. The beads were washed once with 0.1% v/v TFA in water, the washing solution was removed and the Poros material dried in a Speedvac vacuum evaporator. The peptides were released by addition of 8 μL 50% v/v methanol, 5% v/v formic acid. The samples were analyzed using a Perkin-Elmer SCIEX API-365 mass spectrometer fitted to a nanospray interface (Protana A/S, Odense, Denmark). Au/Pd-coated glass capillaries from the same supplier were used. The instrument was calibrated with polypropylene glycol according to the manufacturer's specifications.

2.14 Database searching

The proteins were identified by searching in Swiss-Prot and Trembl databases using MS-Fit and MS-Tag (Protein Prospector, UCSF, San Francisco, CA, USA). All searches were performed using a mass window between 1000–150 000 Da and included human, rat and mouse sequences. The search parameters allowed for oxidation of methionine, carboxyamidomethylation of cysteine and modification of glutamine to pyroglutamic acid. The following criteria were set for considering an identification as positive in MS-Fit database-searching: (i) at least three matching peptide masses, (ii) at least 60% of the measured masses must match the theoretical masses, and (iii) 50 ppm or better mass accuracy. Measured peptide masses could be excluded if their isotopic pattern were

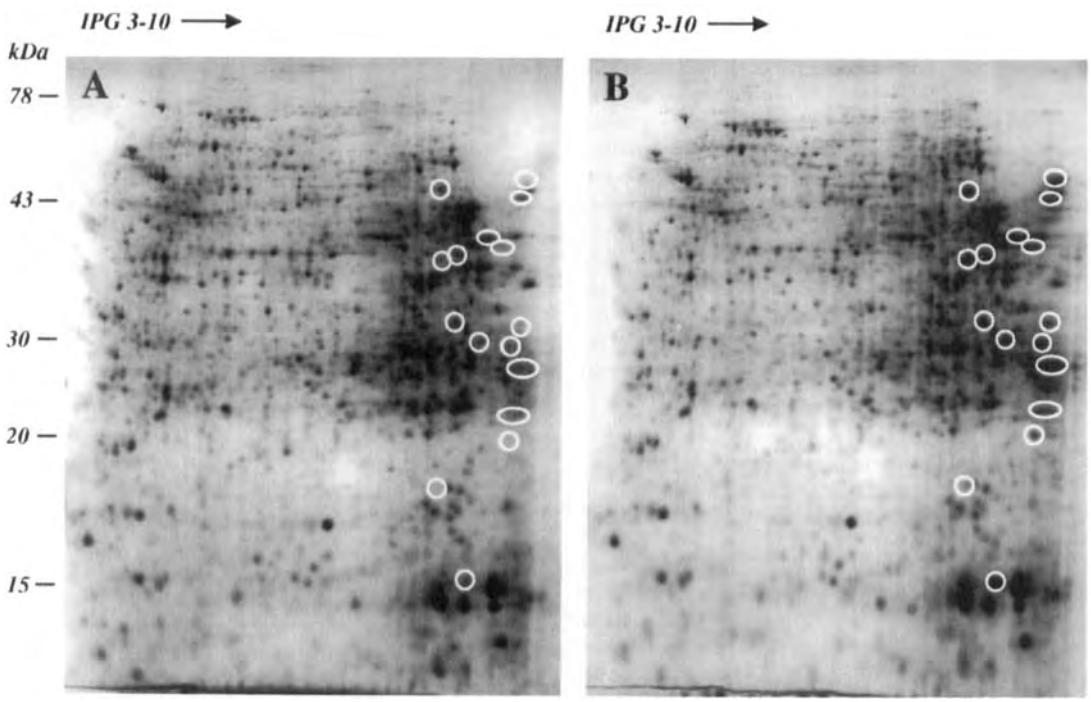


Figure 2. 2-D PAGE of pooled liver samples (A) from five untreated control mice and (B) from five WY14,643-treated mice. Circles indicate 16 spots that were detected by visual inspection to be up-regulated by treatment with WY14,643.

clearly atypical or if their masses corresponded to those of trypsin, adjacent and identified proteins. For MS-Tag searches, positive identifications required at least nine masses, no more than one unmatched fragment ion, 300 ppm or better mass accuracy for the parent ion, and 1500 ppm or better for the product ions.

3 Results

3.1 Drug treatment and therapeutic effects

Male *ob/ob* mice were treated for one week with 180 μmol/kg/day WY14,643. Plasma levels of triglycerides, insulin and glucose from lean, obese, and obese-treated mice were analyzed and it was found that the treatment had normalized triglycerides levels and reduced glucose close to the levels of the lean control mice (Fig. 1).

3.2 Two-dimensional gel electrophoresis

Liver samples homogenized individually in rehydration buffer were pooled (five in each group; obese and WY14,643-treated obese) and duplicate gels were run for both groups (as described in Sections 2.5 and 2.7). For detection, we used a highly sensitive monochromatic silver stain by which up to 1500 spots were detected in each gel. Figure 2A shows one gel with pooled liver samples from untreated mice and Fig. 2B one gel of pooled liver

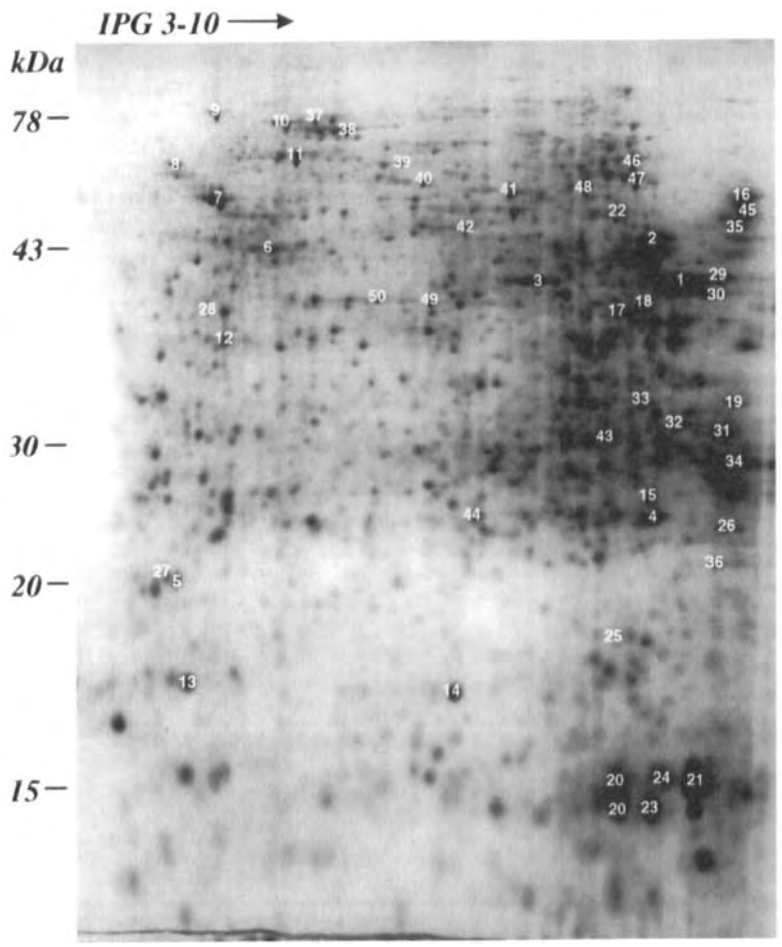


Figure 3. Master pattern with identified spots of liver proteins from *ob/ob* mice. The gel used for annotation is from the analysis of WY14,643-treated mice. Spot No. 1, 3-ketoacyl-CoA thiolase; 2, betaine-homocysteine methyltransferase; 3, arginase; 4, OSF-3; 5, major urinary protein; 6, cytoplasmic β-actin; 7, ATP synthase β-chain; 8, protein disulfide isomerase; 9, 78 kDa glucose-regulated protein; 10, heat shock protein 70; 11, heat shock protein 60; 12, regucalcin; 13, cytochrome b5; 14, Cu/Zn superoxide dismutase; 15, glutathione S-transferase P1; 16, peroxisomal bifunctional enzyme; 17, 3-ketoacyl-CoA thiolase; 18, 3-ketoacyl-CoA thiolase; 19, peroxisomal bifunctional enzyme; 20, hemoglobin β-chain; 21, IFABP; 22, HMG CoA synthase; 23, hemoglobin α-chain; 24, aFABP; 25, acyl CoA oxidase; 26, acyl CoA oxidase; 27, major urinary protein; 28, methionine adenosyltransferase; 29, aspartate aminotransferase; 30, 3-ketoacyl-CoA thiolase; 31, peroxisomal bifunctional enzyme; 32, 3-ketoacyl-CoA thiolase; 33, acyl CoA oxidase; 34, peroxisomal bifunctional enzyme; 35, peroxisomal bifunctional enzyme; 36, peroxisomal bifunctional enzyme; 37, heat shock protein 70; 38, albumin; 39, phospholipase C α-chain; 40, selenium-binding liver protein; 41, aldehyde dehydrogenase; 42, albumin fragments; 43, carbonic anhydrase III; 44, ATP synthase α-chain; 45, elongation factor TU/1-α; 46, catalase; 47, aldehyde dehydrogenase; 48, glutamate dehydrogenase; 49, HMG CoA synthase; 50, phosphoglycerate kinase.

samples from WY14,643-treated animals. By visual inspection, we found 16 spots in the basic region that were apparently up-regulated by the treatment with WY14,643. Next, we analyzed individual liver samples from five obese control animals and four WY14,643-treated obese animals on separate gels and this analysis confirmed the up-regulation by WY14,643 of these 16 spots (not shown). However, many more liver proteins may be regulated by treatment of obese mice with WY14,643, an issue that may be addressed in subsequent studies.

3.3 Identification

For identification of the 16 up-regulated spots and of abundant nonregulated proteins, spots were cut out of copper- or Coomassie-stained gels and subjected to MS, as described. Approximately 60% of the analyzed proteins could be identified by MALDI-TOF and database searching using MS-Fit. Failure to obtain positive identification was either due to insufficient amount of protein to obtain a good spectrum or absence of matching sequence in the database. In the latter case, the samples were reanalyzed by electrospray MS/MS and the results were used to search databases with MS tag. Altogether, we obtained identities of 50 protein spots which are compiled on a master gel using an image from WY14,643-treated obese mice (see Fig. 3). It was found that several spots were identified as the same protein, *e.g.*, peroxisomal bifunctional enzyme (spots 16, 19, 31, 34, 35 and 36 in Fig. 3), which is not unusual in two-dimensional electrophoresis. Of the 16 up-regulated spots, three were identified as acyl CoA oxidase (ACO), six as peroxisomal bifunctional enzyme (PBE) and five as 3-ketoacyl-CoA thiolase (3KCT). Finally, two other up-regulated spots were identified as HMG CoA synthase (HMG CoA S) and adipocyte fatty acid binding protein (aFABP). For details on the MS analysis and identification, see Table 1.

4 Discussion

Proteomics offers an unprejudiced possibility to study global changes in the expression profile of proteins from a given cell or tissue, to date not provided by other techniques. Obviously, for determination of levels of selected proteins, other assays are preferred. For a discussion of the proteomics technology and its applications, see [25]. Here we describe the liver effects in *ob/ob* mice of one antihyperlipidemic drug with PPARα activation properties. The approach that we have taken was largely inspired by the proteomics work of L. Andersson and collaborators (Large Scale Biology, Rockville, MD) who previously analyzed the liver effects of several peroxisome proliferators and PPARα-agonists in lean healthy B6C3F1-mice [20]. However, we were interested in mapping liver effects

Table 1. Up-regulated proteins identified by mass spectrometry

Proteins identified by MALDI-TOF and database searching using MS-Fit

Spot No.	Identity	Species	No. fragments	Matches	Coverage	δ M (ppm)	Accession No.	Database
1	3-Ketoacyl-CoA thiolase peroxisomal A/B[a)]	Rat	10	70%	26	30	P07871/P21775	Swiss-Prot
16	Enoyl-CoA hydratase	Rat	11	73%	17	38	P07896	Swiss-Prot
17	3-Ketoacyl-CoA thiolase peroxisomal A/B[a)]	Rat	10	60%	12	23	P07871/P21775	Swiss-Prot
18	3-Ketoacyl-CoA thiolase peroxisomal A/B[a)]	Rat	8	63%	16	23	P07871/P21775	Swiss-Prot
19	Enoyl-CoA hydratase	Rat	11	63%	10	36	P07896	Swiss-Prot
22	Hydroxymethylglutaryl-CoA synthase	Mouse	11	73%	17	22	P54869	Swiss-Prot
25	Peroxisomal acyl-CoA oxidase	Mouse	8	63%	13	36	O35616	Trembl
26	Peroxisomal acyl-CoA oxidase	Mouse	10	60%	10	26	O35616	Trembl
30	3-Ketoacyl-CoA thiolase peroxisomal A/B[a)]	Rat	6	83%	16	35	P07871/P21775	Swiss-Prot
32	3-Ketoacyl-CoA thiolase peroxisomal A/B[a)]	Rat	5	60%	10	23	P07871/P21775	Swiss-Prot
33	Peroxisomal acyl-CoA oxidase	Mouse	12	67%	15	37	O35616	Swiss-Prot
35	Enoyl-CoA hydratase	Human	10	60%	10	25	Q08426	Swiss-Prot

Proteins identified by MS/MS and database-searching using MS-Tag

Spot No.	Identity	Species	No. fragment ions	Matches	Search mode	δ M (ppm) fragments	Accession No.	Database
24	fatty acid-binding protein, adipocyte	Mouse	13	100.00%	Homology	1070	P04117	Swiss-Prot
31	Enoyl-CoA hydratase	Rat	9	100.00%	Identity	600	P07896	Swiss-Prot
34	Enoyl-CoA hydratase	Rat	11	100.00%	Identity	280	P07896	Swiss-Prot
36	Enoyl-CoA hydratase	Rat	9	100.00%	Identity	1500	P07896	Swiss-Prot

Both databases were updated January 17, 1999.

a) Isoforms which both meet the criteria for positive identification

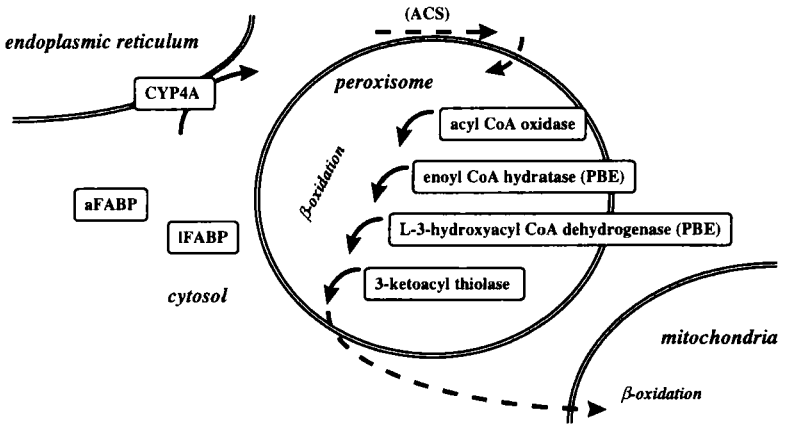


Figure 4. Metabolic flow chart of peroxisomal β-oxidation and some connected metabolic pathways. All boxed proteins were found to be up-regulated by WY14,643, whereas ACS has not yet been identified on these gels.

caused by both PPARα and PPARγ agonists in obese diabetic mice. Therefore, we started by analyzing the effects of a well-known PPARα activator using a therapeutic dose of WY14,643 in *ob/ob* mice.

Turning to the results, we have shown that major effects are found among the 1500 most abundant proteins in livers from *ob/ob* mice treated with WY14,643. The apparently up-regulated proteins identified were mapped on a metabolic flow chart and we found that three of the proteins detected (ACO, PBE; enoyl CoA hydratase and L-3-hydroxyacyl CoA dehydrogenase, and 3KCT) catalyze four out of five reactions in the peroxisomal fatty acid β-oxidation (Fig. 4). Missing is acyl CoA synthetase (ACS), a protein which we have not yet identified in our gels. Adipocyte FABP, a cytosolic protein which may facilitate the transport of fatty acids to peroxisomes and mitochondria, fits into this metabolic flow as a substrate provider. We also believe that liver FABP is up-regulated, since these spots turned red only in the treated samples. Such an "erratic" silver staining may possibly reflect a higher protein concentration. From previous experiments using Northern blotting, we have determined that mRNA for cytochrome P_{450}4A (CYP4A) is up-regulated by WY14,643 (data not shown). CYP4A metabolizes long chain fatty acids to dicarboxyl fatty acids, which are preferred substrates for peroxisomal β-oxidation. This enzyme is therefore included in the metabolic flow chart in Fig. 4. The enzymes identified, acyl CoA oxidase, peroxisomal bifunctional enzyme, 3-ketoacyl-CoA thiolase, and CYP4A, have PPAR responsive elements in their promotors and are known markers of PPARα activation, as well as of peroxisome proliferation [1, 26].

Taken together, the metabolic interpretation is consistent with the notion that WY14,643 induces peroxisomal fatty acid β-oxidation and the up-regulation of this metabolic pathway may be responsible for the therapeutic effect of WY14,643 on levels of triglycerides and glucose in plasma. Subsequent studies will focus on effects of other PPAR activators in this animal model of insulin resistance and diabetes.

We are grateful to Drs. G. Camejo, T. Clementz, J. Fryklund and N. Oakes for constructive criticism during the preparation of this manuscript. Dr. L. Anderson (Large Scale Biology) is acknowledged for valuable discussions during the course of this work.

Received October 23, 1998

5 References

[1] Wahli, W., Braissant, O., Desvergne, B., *Chem. Biol.* 1995, *2*, 261–266.

[2] Schoonjans, K., Staels, B., Auwerx, J., *J. Lipid Res.* 1996, *37*, 907–925.

[3] Spiegelman, B. M., *Diabetes* 1998, *47*, 507–514.

[4] Issemann, I., Green, S., *Nature* 1990, *347*, 645–650.

[5] Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I., Spiegelman, B. M., *Genes Devel.* 1994, *8*, 1224–1234.

[6] Braissant, O., Foufelle, F., Scotto, C., Dauca, M., Wahli, W., *Endocrinology* 1996, *137*, 354–366.

[7] Yu, K., Bayona, W., Kallen, C. B., Harding, H. P., Ravera, C. P., McMahon, G., Brown, M., Lazar, M. A., *J. Biol. Chem.* 1995, *270*, 23975–23983.

[8] Forman, B. M., Tontonoz, P., Chen, J., Brun, R. P., Spiegelman, B. M., Evans, R. M., *Cell* 1995, *83*, 803–812.

[9] Forman, B. M., Chen, J., Evans, R. M., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 4312–4317.

[10] Devchand, P. R., Keller, H., Peters, J. M., Vazquez, M., Gonzalez, F. J., Wahli, W., *Nature* 1996, *384*, 39–43.

[11] Kliewer, S. A., Lenhard, J. M., Willson, T. M., Patel, I., Morris, D. C., Lehmann, J. M., *Cell* 1995, *83*, 813–819.

[12] Issemann, I., Prince, R. A., Tugwood, J. D., Green, S., *J. Mol. Endocrinol.* 1993, *11*, 37–47.

[13] Frøyland, L., Madsen, L., Vaagenes, H., Totland, G. K., Auwerx, J., Kryvi, H., Staels, B., Berge, R. K., *J. Lipid Res.* 1997, *38*, 1851–1858.

[14] Schoonjans, K., Peinadoonsurbe, J., Lefebvre, A. M., Heyman, R. A., Briggs, M., Deeb, S., Staels, B., Auwerx, J., *EMBO J.* 1996, *15*, 5336–5348.

[15] Simpson, A., *Gen. Pharmacol.* 1997, *28*, 351–359.

[16] Peters, J. M., Hennuyer, N., Staels, B., Fruchart, J. C., Fievet, C., Gonzalez, F. J., Auwerx, J., *J. Biol. Chem.* 1997, *272*, 27307–27312.

[17] Staels, B., Vudac, N., Kosykh, V. A., Saladin, R., Fruchart, J. C., Dallongeville, J., Auwerx, J., *J. Clin. Invest.* 1995, *95*, 705–712.

[18] Hertz, R., Bisharashieban, J., Bar-Tana, J., *J. Biol. Chem.* 1995, *270*, 13470–13475.

[19] Ashby, J., Brady, A., Elcombe, C. R., Elliott, B. M., Ishmael, J., Odum, J., Tugwood, J. D., Kettle, S., Purchase, I. F., *Human Experim. Toxicol.* 1994, *13*, S1–117.

[20] Anderson, N. L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharmacol. 1996, 137, 75–89.*

[21] Görg, A., Boguth, G., Obermaier, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16*, 1079–1086.

[22] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[23] Lee, C., Levin, A., Branton, D., *Anal. Biochem.* 1987, *166*, 308–312.

[24] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173–179.

[25] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 1–243.

[26] Hijikata, M., Wen, J. K., Osumi, T., Hashimoto, T., *J. Biol. Chem.* 1990, *265*, 4600–4606.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 943–951

Frank A. Witzmann[1]
Carla D. Fultz[1]
Raymond A. Grant[2]
Linda S. Wright[3]
Steven E. Kornguth[3]
Frank L. Siegel[3]

[1]Molecular Anatomy Laboratory, Department of Biology, Indiana University-Purdue University, Columbus, IN, USA
[2]Corporate Research Division, The Procter & Gamble Company, Cincinnati, OH, USA
[3]The Waisman Center, University of Wisconsin, Madison, WI, USA

Regional protein alterations in rat kidneys induced by lead exposure

Lead is a potent neuro- and nephrotoxin in humans and a renal carcinogen in rats. Previous studies have detected lead-induced increases in the activities of specific detoxification enzymes in distinct kidney cell types preceding irreversible renal damage. While preferential susceptibility of the highly vascularized cortex to the effects of lead is clear, lead effects on the medullary region have remained unexplored. The present study was undertaken to investigate the extent to which regional renal protein expression differs and to determine which, if any, regionally distinct protein markers indicative of lead's renotoxic mechanism might be detected in kidney cortical and medullary cytosols. We examined protein expression in these two functionally and anatomically distinct regions, and identified several proteins that are differentially expressed in those regions and were significantly altered by lead. Kidney cytosols from rats injected with lead acetate (114 mg/kg, three consecutive daily injections) were separated by two-dimensional electrophoresis. Lead exposure significantly ($P<0.001$) altered the abundance (either ↑ or ↓) of 76 proteins in the cortex and only 13 in the medulla. Eleven of the proteins altered in the protein patterns were conclusively identified either by matrix-assisted laser desorption/ionization mass spectrometry / electrospray ionization-mass spectrometry (MALDI-MS/ESI-MS) analysis of peptide digests, immunological methods, or by gel matching. Several of the cortical proteins altered by lead were unchanged in the medulla while others underwent similar but lesser alterations. These observations reflect the complexity of lead's nephrotoxicity and endorse the application of proteomics in mechanistic studies as well as biomarker development in a variety of toxicologic paradigms.

Keywords: Cortex / Cytoplasm / Two-dimensional electrophoresis / Kidney / Lead / Medulla / Proteomics / Rat
EL 3312

1 Introduction

Exposure to lead is toxic to several organ systems, especially to kidney, where high levels of lead accumulate [1, 2]. In acute exposure, lead-induced pathobiological changes occur primarily in the proximal tubule, where damage to tubular structure and function is indicated by the appearance of nuclear inclusion bodies and mitochondrial swelling indicative of respiratory dysfunction [3–5]. We have previously reported lead-induced increases in the expression of specific glutathione *S*-transferase (GST) isoenzymes in distinct kidney cell types [5]. These increases in specific GST protein levels were paralleled by increased GST enzyme activity, and they preceded pathobiological changes in kidney [6]. The original reports from our laboratories were based upon HPLC analysis of GSTs and were more recently confirmed using two-dimensional electrophoresis (2-DE) [7]. The latter has the inherent ability to inform about the nature of other proteins whose expression is influenced by lead exposure.

Indicative of regional renal physiologic differences, histologic examination and enzyme assays confirm that the renal cortex and medulla are constitutively and biochemically different [8] and that lead accumulation [9] and lead-induced damage predominates in the cortical proximal tubular epithelium [1, 5]. Consequently, probing the effect of lead exposure on the expression of whole kidney homogenate/cell fraction protein patterns may not accurately reflect regiospecific differences in susceptibility to lead. This is particularly true if lead-induced protein alterations are regionally diametrical, hence minimizing or masking one another only to be overlooked in whole organ homogenate/cell fraction analysis. While preferential susceptibility of the highly vascularized cortex to the effects of lead is clear, lead effects on the medulla remain unexplored. The present study was undertaken to investigate the extent to which lead exposure alters regional renal protein expres-

Correspondence: Dr. Frank A. Witzmann, Molecular Anatomy Laboratory, Department of Biology, Indiana University Purdue University-Columbus, 4601 Central Avenue, Columbus, IN 47203, USA
E-mail: fwitzmann@iupui.edu
Fax: +812-348-7279

Abbreviations: **CMI**, charge modification index; **GST**, glutathione *S*-transferase; **MSN**, master spot number

sion and to determine which, if any, regionally distinct protein markers indicative of lead's renotoxic mechanism might be detected in kidney cortical and medullary cytosols.

2 Materials and methods

2.1 Reagents

Ultrapure electrophoretic reagents were obtained from Bio-Rad (Richmond, CA), Sigma (St. Louis, MO), BDH (Poole, UK) and National Diagnostics (Atlanta, GA). Sequence-grade trypsin was obtained from Roche Molecular Biochemicals (Indianapolis, IN). CHAPS (3-[3-cholamidopropyl)dimethylammonio]-1-propanesulfonate) and dithiothreitol were obtained from Calbiochem (La Jolla, CA) and *N*-isopropyl iodoacetamide from Molecular Probes (Eugene, OR). All other chemicals used were reagent grade. Hsc70, hsp70, and hsp90 antibodies were obtained from StressGen (Vancouver, BC, Canada), calbindin and calcineurin antibodies from Sigma, and anti-glutathione *S*-transferase P1 from Biotrin (Dublin, Ireland).

2.2 Animals and sample preparation

Sprague-Dawley male rats were bred and housed in our animal care facility and given access to Tekland Rodent Blocks and water *ad libitum*. Eight-week-old rats, housed four per cage, were injected with lead acetate: 114 mg/kg, three consecutive daily intraperitoneal (i.p.) injections; controls received physiological saline injections, i.p. Rats were sacrificed on day 4 by CO_2 asphyxiation. Kidneys were decapsulated, and renal cortex and medulla were dissected. Each tissue was homogenized in five volumes of ice-cold 20 mM Tris-HCl, pH 7.8, 2 mM EGTA, 10 mM EDTA, 2 mM DTT. Cytosols were prepared by centrifugation at 100 000 × *g* for 20 min and solubilized in one volume of a lysis buffer, pH 9.5, containing 9 M urea, 4% CHAPS, 1% DTT, and 2% carrier ampholytes, pH 8–10.5.

2.3 2-DE ISO-DALT electrophoresis

Sample proteins were resolved by 2-DE using the 20 × 25 cm ISO-DALT® 2-D gel system (20 gels per run) [10]. Ten μL of solubilized protein sample (approximately 175 μg) were applied to each isoelectric focusing (IEF) gel tube; gels were run for 25 000 Vh using a progressively increasing voltage protocol. A computer-controlled gradient casting system was used to prepare second-dimensional SDS gradient slab gels; acrylamide concentration varied linearly from 11 to 17%T. First-dimension IEF tube gels were loaded directly onto the slab gels without equilibration. Second-dimensional slab gels were run in groups of 20 in a DALT slab electrophoresis tank at 10°C for 18 h at 160 V. Following SDS electrophoresis, slab gels were stained for protein using a colloidal Coomassie Blue G-250 procedure [11] for approximately four days, after which equilibrium intensity was achieved.

2.4 Western blotting and immunological techniques

Protein patterns on replicate gels were electroblotted for immunological identification of calbindin, calcineurin (calmodulin-dependent protein phosphatase), GST, hsc70, hsp70, and hsp90. Proteins were transferred from gels slabs onto PVDF membranes (in 49 mM Tris – 39 mM glycine buffer with 0.04% SDS and 20% methanol, pH 9.2) using a semidry transfer cell for 37.5 Vh at room temperature. The PVDF membranes were then washed and blocked with several exchanges of 0.3% Tween-20 in PBS and incubated with primary antibody. Blots were washed with 0.3% Tween-20 in PBS, incubated with an alkaline phosphatase-conjugated secondary antibody and visualized using BCIP/NBT alkaline phosphatase substrate (Sigma *Fast*® tablets). Following visualization, all blots were rinsed with water and scanned as described below. They were then air-dried, blocked with 0.3% Tween-20 in PBS, rinsed with water, and stained with solubilized colloidal gold reagent [12].

2.5 Mass spectrometric protein identification

Other proteins whose expression was affected by lead (aldose reductase, α_2-microglobulin, aflatoxin B_1 aldehyde reductase, sorbitol dehydrogenase, transferrin, and transketolase) were excised from the gels and minced. Acrylamide pieces were rinsed with 50 mM ammonium bicarbonate/acetonitrile (1:1, v/v), dried, rehydrated in DTT, and alkylated with *N*-isopropyl iodoacetamide (100 mM in 50 mM ammonium bicarbonate). After removal of excess buffer, gel slices were dried, rehydrated in the presence of sequencing grade trypsin, and digested overnight at 37°C. Peptides were eluted using 5% formic acid/50% acetonitrile. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was done using a PerSeptive Biosystems Voyager RP-DE instrument (Perseptive Biosystems, Framingham, MA). Ion spray/nanoelectrospray MS and MS/MS measurements were performed on a Sciex API–III LC/MS/MS triple quadrupole instrument (Perkin-Elmer Biosystems and Sciex). MS results were analyzed using MassMap (Finnigan) and PeptideSearch (EMBL) programs [13].

2.6 Gel pattern analysis

Stained gels were optically scanned at 121 micron resolution using a CCD scanner and images were processed as

described [14]. Groupwise statistical comparisons were made to screen for protein alterations (Student T-test). Charge modification index (CMI), calculated as described [15], is a numerical description of the overall average number of charges added per protein molecule and is thus an excellent estimate of the degree to which a protein is chemically (*i.e.*, post-translationally) modified.

3 Results and discussion

An average of 727 protein spots were resolved and matched to the cortex cytosol reference pattern (Fig. 1) and 716 in the medulla cytosol (Fig. 2). Regional comparison of gene expression in normal, untreated rats appears in detail elsewhere [7]. Briefly, the abundance of 122 proteins differed significantly ($P<0.001$) between the two regions, with 82 proteins higher in cortex and 40 higher in medulla. Of these, 30 proteins were found to be unique to one region or the other (26 in cortex and four in medulla). Absent from the control cortical pattern, but detectable in medulla, was aldose reductase, while sorbitol dehydrogenase was more abundant in cortex. Aflatoxin B1 aldehyde reductase, hsp90, GSTP1 native and (–2) subunits, and transferrin were all more abundant in normal medul-

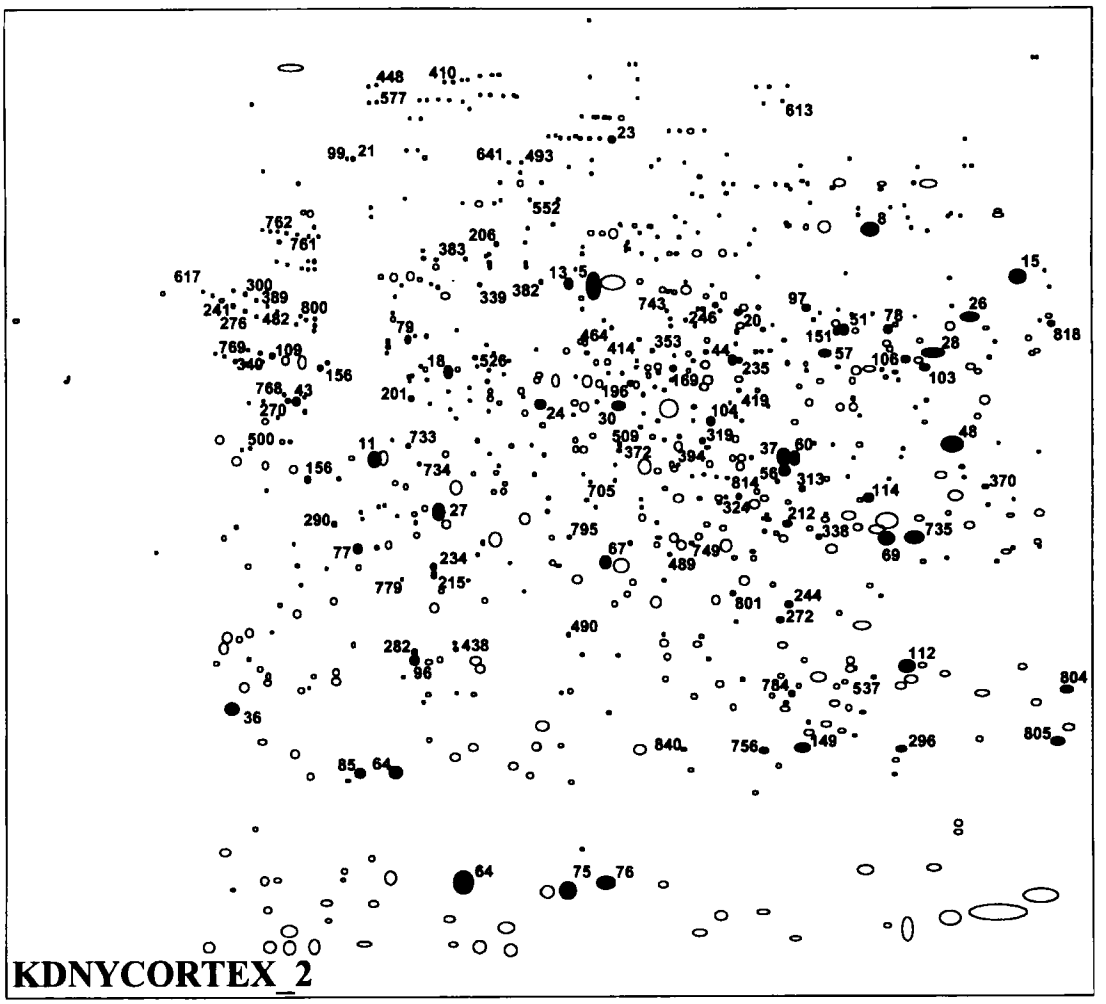


Figure 1. 2-DE map of rat kidney cortex cytosol master pattern highlighting the coordinate positions and master spot number (MSN) of 76 proteins whose abundance was altered by lead acetate exposure ($P<0.001$). Proteins with MSN 85, 151, 215, 241, 270, 276, 282, 290, 324, 370, 372, 382, 389, 394, 419, 438, 482, 489, 500, 526, 552, 577, 613, and 617 were detected only in controls while proteins with MSN 733, 734, 735, 749, 756, 768, 769, 779, 784, 795, 800, 801, 804, 814, and 818 were detected only in 2-D gels from the lead-exposed group.

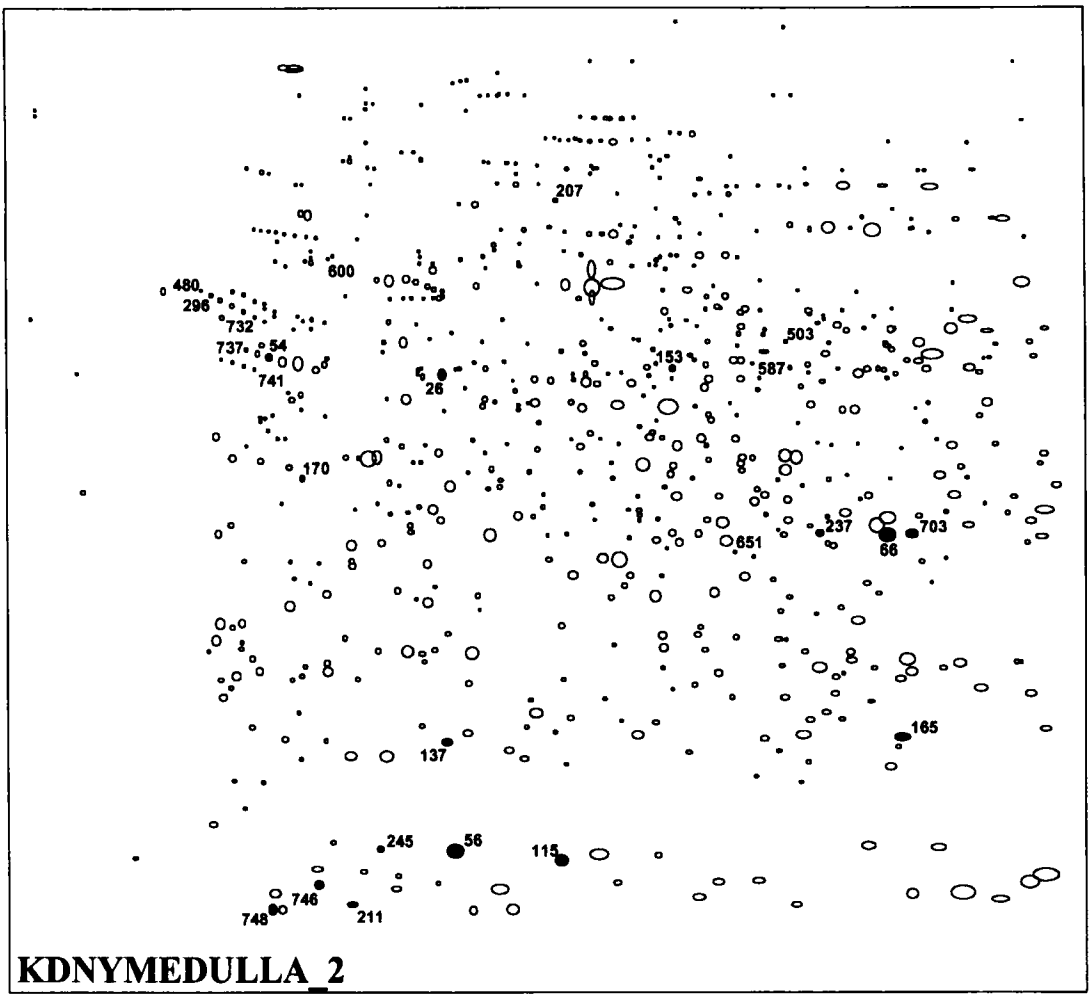


Figure 2. 2-DE map of rat kidney medulla cytosol master pattern highlighting the coordinate positions and MSN of 13 proteins whose abundance was altered by lead acetate exposure ($P<0.001$). Proteins with MSN 115, 211, 245, 296, 480, 600, and 732 were detected only in controls while proteins with MSN 737, 741, and 746 were detected only in 2-D gels from the lead-exposed group.

lary cytosol than cortex. Conversely, α_2-microglobulin, argininosuccinate synthase, calcineurin, calbindin, GST (–1), and sorbitol dehydrogenase were more abundant in the cortex.

Lead exposure significantly ($P<0.001$) altered the abundance of 78 proteins in the cortex (42 increased and 36 decreased; Table 1) and 16 proteins in the medulla (ten increased and six decreased; Table 2). Of the cortical proteins altered by lead exposure, 24 were rendered undetectable in the lead-exposed group while 15 previously undetectable proteins were observed. In the medulla, eight proteins seen in control rats were undetectable after lead exposure whereas three previously undetectable proteins appeared in the lead-exposed group. Tables 1 and 2 include the identified proteins found in the two kidney regions while Table 3 lists only the identified proteins, comparing the lead effect and indicating the method of identification for each. Notable comparative lead-induced alterations include major increases in aldose reductase, aflatoxin B_1 aldehyde reductase, and transketolase; major declines in α_2-microglobulin and in cortical argininosuccinate synthase but not medulla (where it was already significantly lower); cortical increase in hsp90 (but not me-

Table 1. Rat kidney cortex cytosol proteins altered by lead exposure

MSN[a)]	ID	Mean control abundance	Mean lead abundance	PROB[b)]	% of Control
5		78,810	57,250	0.009	73
8	Transferrin	48,860	32,600	0.0005	67
11		38,220	46,390	0.0002	121
13		10,310	6,685	0.0009	65
15	Transketolase	40,320	52,690	0.0008	131
18		11,980	9,000	0.0006	75
21	hsp90	4,857	7,548	0.008	155
23		5,879	3,812	0.00007	65
24		16,170	11,420	0.00008	71
26		25,820	18,370	0.0001	71
27		22,880	7,986	0.0003	35
28		36,090	27,770	0.0002	77
30		19,080	15,930	0.0008	83
36	Calbindin	21,430	12,910	0.0007	60
37		41,010	30,810	0.0002	75
43		6,733	1,174	0.00002	17
44		9,538	6,807	0.0003	71
48	Argininosuccinate synthase	69,860	15,380	0.00002	22
51	Calcineurin	8,703	3,340	0.0001	38
56		22,800	18,290	0.0001	80
57		9,309	7,164	0.001	77
60		19,410	11,310	0.00009	58
64	α_2-Microglobulin	69,570	5,977	0.00001	9
67		17,270	23,150	0.0002	134
69	Aflatoxin B1 aldehyde reductase	44,850	99,810	0.000002	223
75		31,240	2,358	0.00005	8
76		44,130	35,440	0.0003	80
77		7,868	4,821	0.0004	61
78	Calcineurin	7,718	5,503	0.0005	71
79		4,134	7,536	0.0006	182
85	α_2-Microglobulin	9,675	–	NA	NA
96		8,858	14,850	0.00007	168
97		5,111	4,435	0.001	87
99	hsp90	2,140	3,173	0.03	148
103		7,393	8,669	0.0009	117
104		6,249	4,652	0.002	74
106		6,571	4,099	0.00002	62
109		3,381	5,740	0.00004	170
112		35,430	28,470	0.0004	80
114	Sorbitol dehydrogenase	11,410	8,521	0.0005	75
120		3,934	2,484	0.00009	63
149	GSTP1	15,480	23,920	0.0003	155
156		3,568	975	0.00002	27
166		2,747	1,475	0.0004	54
169		3,259	9,337	0.000004	286
176		4,969	6,809	0.0008	137
196		2,541	4,586	0.0004	180
201		2,700	10,440	0.00004	387
206		1,261	523	0.00008	41
212		5,163	7,351	0.0006	142
234		3,173	6,039	0.00005	190
235		3,553	7,297	0.0001	205
244		4,194	3,541	0.0008	84
246		1,749	2,882	0.0005	165

Table 1. continued

MSN[a)]	ID	Mean control abundance	Mean lead abundance	PROB[b)]	% of Control
263		1,965	8,542	0.0009	435
272		3,613	2,028	0.0001	56
296	GSTP1	5,658	35,280	0,000005	624
300		1,129	354	0.00005	31
313		2,167	538	0.00003	25
319		1,632	9,301	0.000007	570
338		2,492	9,515	0.00001	382
339	hsp70	1,007	1,320	0.02	131
340		1,208	2,218	0.0005	184
353		1,326	2,147	0.0004	162
383		1,081	2,450	0.0005	227
410		415	134	0.00005	32
414		880	3,201	0.00003	364
448		374	177	0.0002	47
464		825	1,172	0.0003	142
490		730	2,107	0.0005	289
493		352	924	0.0007	262
509		839	665	0.0005	79
537		1,323	2,117	0.00004	160
641		284	702	0.0003	247
705		512	1,565	0.0004	306
735	Aldose reductase	–	37,465	NA	NA
743		374	1,812	0.0001	484
756	GSTP1	–	4,666	NA	NA
761		432	1,306	0.0002	302
762		673	1,385	0.0006	206
804		–	9,962	NA	NA
805		8,549	50,950	0.0001	596
840	GSTP1	–	946	NA	NA

a) MSN, Master spot number
b) PROB, probability, significance level
NA, not applicable

dulla where it was already elevated); cortical decline in transferrin but not medulla; decreased sorbitol deydrogenase, calbindin and calcineurin in cortex but not in medulla, where all were already significantly lower.

Cytosolic GSTP1 was resolved as three or four charge variants due to chemical charge modification of the native form [7]. The magnitude of this modification, calculated as CMI, is shown in Fig. 3. CMI in cortex is greater than in medulla but was altered by lead only in the cortex. Total GSTP1 abundance (sum of individual charge variant abundances) was the same in cortex and medulla from untreated rats. Lead exposure increased GSTP1 abundance in both regions and had a significantly preferential effect on cortical GSTP1 (3-fold increase) compared to medulla (2-fold).

Five principal findings arose from this study: (i) 30 proteins were found to be unique to either renal cortex or medulla; (ii) lead administration altered the expression of 10% of the detectable cortical proteins and 2% of the proteins detected in medulla; (iii) lead administration caused an almost equal number of increases and decreases in specific cortical proteins; (iv) the largest changes observed were in the abundances of α_2-microglobulin (down ninety percent in cortex and not detectable in medulla), aldose reductase (detectable in cortex only following lead administration and increased more than 20-fold in medulla after lead administration), GSTP, which increased sixfold in cortex after lead treatment, and aflatoxin B_1 aldehyde reductase (twofold increase in both cortex and medulla); and (v) lead administration altered the posttranslational modification of GSTP in renal cortex.

Among proteins that were detected in unexposed renal medulla, but not renal cortex, was aldose reductase. This finding is consistent with a previous report that aldose reductase mRNA levels in cortex are only one percent of

Table 2. Rat kidney medulla cytosol proteins altered by lead exposure

MSN[a)]	ID	Control abundance	Lead abundance	PROB[b)]	% of Control
26		10,780	7,255	0.0006	67
54		4,440	8,281	0.00009	187
66	Aflatoxin B1 aldehyde reductase	47,716	91,227	0.00004	191
115		13,601	–	NA	NA
137		7,069	5,324	0.0002	75
153		3,645	7,141	0.00005	196
165	GSTP1	11,076	27,784	0.00006	251
170		3,180	1,214	0.0003	38
207		1,482	670	0.00007	45
237		3,229	7,516	0.00005	233
245	α_2-microglobulin	4,014	–	NA	NA
503		581	1,092	0.0005	188
587		416	730	0.0005	175
651		863	2,020	0.0001	234
703	Aldose reductase	1,185	24,634	0.006	2079
748		–	6,178	NA	NA

a) MSN, Master spot number
b) PROB, probality, significance level
NA, not applicable

Table 3. Identification of proteins in rat kidney cortex and medulla cytosols

Identification	Cortex MSN[a)]	Cortex % of Control	Medulla MSN	Medulla % of Control	% Masses matched	% Coverage	Sequence from database
Actin, beta	11	121	13	97	–	–	HPGM[d)]
Actin, gamma	19	9	20	107	–	–	HPGM
Aflatoxin B1 aldehyde reductase	69	223	66	191	37	19	(R)FYAFNPLAGGLLTGR(Y)
Aldose reductase	735	NDC[b)]	703	2079	36	13	(K)MPLVGLGTWKSSPGQVK(E)
α_2-microglobulin	64	9	56	9	58	32	(K)NGETFQLMVLYGR(T)
α_2-microglobulin	85	NDL[c)]	245	NDL	58	32	(K)NGETFQLMVLYGR(T)
Argininosuccinate synthase	48	22	132	40	58	12	(K)QHGIPIPVTPK(S)
Calbindin	36	60	135	71	–	–	Ig[e)]
Calcineurin	51	38	164	90	–	–	Ig
Calcineurin	78	71	134	98	–	–	Ig
GSTP1	149	155	142	152	–	–	Ig
GSTP1	296	624	165	251	–	–	Ig
GSTP1	756	NDC	751	194	–	–	Ig
GSTP1	840	NDC	ND	ND	–	–	Ig
hsc70	12	84	2	103		–	Ig
hsc70	16	98	167	119	–	–	Ig
hsp70	339	131	62	114	–	–	Ig
hsp90	21	155	4	78	–	–	Ig
hsp90	99	148	34	88	–	–	Ig
Sorbitol dehydrogenase	114	75	225	82	54	26	(R)LENYPIPELGPNDVLLK(M)
Transferrin	8	67	8	75	58	12	(K)LPEGTTYEEYLGAEYLQAVGNIR(K)
Transketolase	15	131	23	111	56	18	(R)TSRPENAIIYSNNEDFQVGQAK(V)

a) MSN, Master spot number
b) NDC, not detected in controls
c) NDL, not detected in lead-exposed
d) HPGM, identified by homologous position, gel matching
e) Ig, identified immunologically
Coverage refers to the amino acids in identified peptide sequences as a % of all amino acids in the protein.

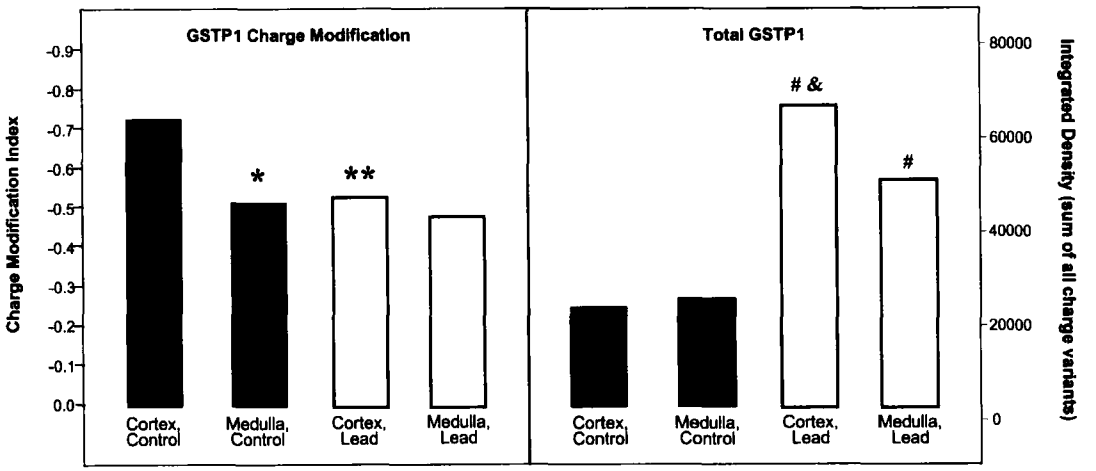


Figure 3. The effect of lead acetate on total GSTP1 abundance and GSTP1 CMI (*), mean (n=5) cortical GSTP1 is significantly different ($P < 0.05$) from medullary GSTP1 in controls. (**), only cortex GSTP1 CMI undergoes a significant decline with lead acetate exposure ($P < 0.05$). (#), Total GSTP1 abundance increased by lead, (&), preferentially in cortex.

those in medulla [16]. *The detection of aldose reductase* only in the medulla is consistent with published enzymatic data and recognized osmotic conditions in that region of the kidney where aldose reductase functions in the NADPH-dependent conversion of glucose to slowly diffusible sorbitol, which maintains osmotic balance during antidiuresis [17]. Medullary cells are exposed to far greater extracellular osmotic pressure than the isotonic cortical region and increases in sorbitol production serve to maintain the cellular integrity of renal medulla. It has been shown that aldose reductase decreases during diuresis, *while sorbitol dehydrogenase increases. During* antidiuresis, aldose reductase increases, while sorbitol dehydrogenase activity remains low [18]. Two possible mechanisms for aldose reductase induction by lead are suggested by the literature. Studies of aldose reductase promoter regions have identified a tonicity-responsive element; this would account for induction under conditions of antidiuresis which would follow dehydration [19]. However, following lead treatment, aldose reductase abundance increased (and sorbitol dehydrogenase abundance decreased) to a greater extent in renal cortex than in medulla, *where the highest interstitial osmolarity normally occurs.* Recent (unpublished) results from a separate study conducted in this laboratory indicate normal blood osmolarity in rodents exposed to lead as described here. It is thus unlikely that dehydration and increased cortical osmolarity underly the lead-associated alterations. An alternative explanation lies in a report that oxidative stress causes induction of aldose reductase in smooth muscle cells and that this confers protection against the cytotoxic effects of aldehyde products of lipid peroxidation [20]. *This may also explain the significant induction of the chemoprotective aflatoxin* B_1 aldehyde reductase, aldose reductase's relative in the aldo-keto reductase superfamily AKR1 [21]. Lead-related oxidative stress could thus account for reductase induction though we have no direct evidence to support this notion.

Oxidative stress has also been shown to induce transketolase [22]. Induction of transketolase, an enzyme which catalyzes a key step in the nonoxidative stage of the pentose phosphate pathway, may represent adaptive increase in the generation of NADPH to regenerate GSH under conditions of oxidative stress. An antioxidant response element (ARE) sequence has been identified in the promoter region of the GSTP gene and its induction by oxidative stress would be expected [23] and could explain the increased abundance of GSTP1. Because GSTP1 and aflatoxin B_1 aldehyde reductase belong to a set of genes similarly responsive to certain xenobiotics, their induction by lead may somehow be related [24]. Metabolic stress, in addition to osmotic stress, may explain the elevated constitutive expression of hsp90, a prominent cytosolic stress protein, in medulla.

Argininosuccinate synthase catalyzes the condensation reaction between citrulline and aspartate to form argininosuccinate (primarily in proximal tubule). Decreased argininosuccinate synthase abundace (4.5-fold) suggests impairment of mitochondrial activity as a consquence of lead-mediated mitochondrial dysfunction [25, 26]. The impairment of mitochondrial structure and function is known to play an important role in lead-related nephrotoxicity [1].

The dramatic decline in α_2-microglobulin may be the result of its increased excretion following lysosomal proteolysis [2]. Lead-related decreases in this protein stand in marked contrast to its previously reported accumulation by a variety of aromatic toxicants which bind to α_2-microglobulin, inhibiting its lysosomal degradation in the nephron [27].

4 Concluding remarks

The impact of lead on cellular calcium homeostasis has been demonstrated repeatedly but is not completely understood. Our observation of the nearly twofold decrease in levels of the calcium-binding protein calbindin and the calmodulin-dependent protein phosphatase calcineurin is similar to the inhibition by lead of 1,25-dihydroxyvitamin D3-stimulated synthesis of calcium-binding proteins in bone cells [27] and presumably reflects a perturbation of calcium homeostasis in the renal cortex. We have detected and quantified numerous regionally distinct protein markers indicative of lead's renotoxicity in kidney cortical and medullary cytosol. Some prominently altered proteins have been identified. Further investigation of these lead effects *via* additional systematic protein identification combined with a lead dose-response approach in all cell compartments as well as whole-tissue homogenates is underway.

The authors express their gratitude to the following for mass spectrometric protein identification: Mark Bauer, Angela Fieno, Tom Keough, Marty Lacey, and Yiping Sun, all of Procter & Gamble Company. This effort was supported in its entirety by AFOSR Grants F49620-96-1-0156 and F49620-96-1-0074.

Received September 1, 1998

5 References

[1] Nolan, C. V., Shaikh, Z. A., *Toxicology* 1992, *73*, 127–146.

[2] Logham-Adham, M., *Environ. Health Persp.* 1998, *105*, 928–937.

[3] Goyer, R. A., *Curr. Top. Pathol.* 1971, *55*, 147–176.

[4] Daggett, D. A., Obereley, T. D., Nelson, S. A., Wright, L. S., Kornguth, S. E., Siegel, F. L., *Toxicology* 1998, *128*, 191–206.

[5] Oberley, T. D., Friedman, A. L., Moser, R., Siegel, F. L., *Toxicol. Appl. Pharmacol.* 1995, *131*, 94–107.

[6] Moser, R., Oberley, T. D., Daggett, D. A., Friedman, A. L., Johnson, J. A., Siegel, F. L., *Toxicol. Appl. Pharmacol.* 1995, *131*, 85–93.

[7] Witzmann, F. A., Daggett, D. A., Fultz, C. D., Nelson, S. A., Wright, L. S., Kornguth, S. B., Siegel, F. L., *Electrophoresis* 1998, *19*, 2491–2497.

[8] Jacobson, H. R., *Am. J. Physiol.* 1981, *241*, 203–218.

[9] Tarloff, J. B., Goldstein, R. S., in: Hogson, E., Levi, P. E. (Eds.), *Introduction to Biochemical Toxicology*, Appleton and Lange, Norwalk, CT 1994, pp. 519–546.

[10] Anderson, N. L., *Two-Dimensional Electrophoresis: Operation of the ISO-DALT System*, Large Scale Biology Press, Washington DC 1991.

[11] Neuhoff V., Arnold, N., Taube, D., Ehrhardt, W. *Electrophoresis* 1988, *9*, 255–262.

[12] Fultz, C. D., Witzmann, F. A., *Anal. Biochem.* 1997, *251*, 288–291.

[13] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[14] Anderson, N. L., Esquer-Blasco, R., Anderson, N. G., in: Tyson, C. A., Frazier, J. M. (Eds.), *In Vitro Toxicity Indicators*, Academic Press, San Diego 1994, pp. 463–473.

[15] Anderson, N. L., Copple, D. C., Bendele, R. A., Probst, G. S., Richardson, F. C., *Fundam. Appl. Toxicol.* 1992, *18*, 570–580.

[16] Dorin, R. I., Shah, V. O., Kaplan, D. L., Vela, B. S., Zager, P. G., *Diabetologia* 1995, *38*, 46–54.

[17] Sands, J. M., Schrader, D. C., *J. Am. Soc. Nephrol.* 1990, *1*, 58–65.

[18] Martial, S., Price, S. R., Sands, J. M., *J. Amer. Soc. Nephrol.* 1995, *5*, 1971–1978.

[19] Daoudal, S., Tournaire, C., Halere, A., Veyssiere, G, Jean, C., *J. Biol. Chem.* 1997, *272*, 2615–2619.

[20] Spycher, S. E., Tabataba-Vakili, S., O'Donnell, V. B., Palomba, L., Azzi, A., *FASEB J.* 1997, *11*, 181–188.

[21] Jez, J. M., Flynn, T. G., Penning, T. M., *Biochem. Pharmacol.* 1997, *54*, 639–647.

[22] Salamon, C., Chervenak, M., Piatigorsky, J., Sax, C. M., *Genomics* 1998, *48*, 209–220.

[23] Wasserman, W. W., Fahl, W. E., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 5361–5366.

[23] Primiano, T., Gastel, J. A., Kensler, T. W., Sutter, T. R., *Carcinogenesis* 1996, *17*, 2297–2303.

[24] Reed, D. J., *Ann. Rev. Pharmacol. Toxicol.* 1990, *30*, 603–631.

[25] Chavez, E., Jay, D., Bravo, C., *J. Bioenerg. Biomembr.* 1987, *19*, 285–295.

[26] Hildebrand, H., Hartmann, E., Popp, A., Bomhard, E., *Arch. Toxicol.* 1997, *71*, 351–359.

[27] Pounds, J. G., *Neurotoxicology* 1984, *5*, 295–331.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 952–961

Jasminka Godovac-Zimmermann[1,2]
Vukic Soskic[2]
Slobodan Poznanovic[2]
Federico Brianza[2]

[1]Center for Molecular Medicine, University College London, UK
[2]Protein Laboratory, Institute for Molecular Biotechnology, Jena, Germany

Functional proteomics of signal transduction by membrane receptors

Functional proteomic methods have been developed and applied to the investigation of signal transduction systems involving platelet-derived growth factor (PDGF), endothelin and bradykinin receptors. Mouse fibroblast cells have been stimulated with PDGF or endothelin. Phosphorylation/dephosphorylation of several hundred proteins has been followed as a function of time following stimulation using 2-D gel electrophoresis and anti-phosphotyrosine or anti-phosphoserine antibodies. Up to 100 of these proteins showed strong changes in phosphorylation with minutes of receptor stimulation. Identification of some of these proteins by mass fingerprinting using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry and by partial peptide sequencing with ion trap electrospray mass spectrometry has identified proteins which were previously known to be associated with PDGF signaling, proteins which have been shown to be involved in other signaling pathways, but not PDGF and proteins not previously associated with signal transduction. Parallel to these studies, new methods for rapid, single-step isolation of peptide receptors using a peptide coupled to a $(dA)_{30}$ oligonucleotide have been developed and applied to mass spectrometric studies of post-translational modifications of the endothelin B and bradykinin B_2 receptors under *in vivo* conditions. Both receptors have been shown to undergo extensive phosphorylation as well as palmitoylation. The patterns of post-translational modifications are more complex than previously recognized and provide new indications of possible roles for these modifications in the regulation and response of these receptors.

Keywords: Signal transduction / Proteome / Endothelin / Bradykinin / Platelet-derived growth factor receptor
EL 3322

1 Introduction

One of the most characteristic features of eukaryotic cells is the presence of sophisticated signal transduction systems which regulate responses to the cellular environment and/or are involved in physiological control in higher organisms. The prototypic cellular signal transduction system involves a protein receptor in the cellular membrane which acts as a signal sensor through binding of effector molecules at the external surface of the cell and subsequently acts as a signal transducer. The signal response often involves changes in interactions of the receptor with cellular proteins (*e.g.*, G-proteins in the case of G-protein-coupled receptors), as well as post-translational modifications of the receptor itself, *e.g.*, changes in phosphorylation and/or acylation of the receptor. These initial signal responses are typically coupled to cascades of further responses that may involve large numbers of intracellular proteins and may culminate in changes in the expression of genetic information *via* altered protein synthesis. In many cases these processes also involve complex feedback systems, *e.g.*, the desensitization and resensitization of the cellular receptor itself [1–4].

Recent analyses suggest that only a small percentage (*ca.* 2 %, [5]) of diseases based on genetic defects are likely to be monogenomic. Similar indications that cellular systems are based on complex networks of interactions with potentially high levels of redundancy seem to be inherent in the large numbers of gene knockout experiments for which no obvious phenotypic characteristics are apparent [6, 7]. Furthermore, although the ultimate adaptation of a cellular system to its environment certainly involves expression of genetic information, it is clear that in many signal transduction systems complex responses involving large numbers of intracellular proteins occur without direct, immediate involvement of new protein synthesis. These considerations lead to the conclusion that many aspects of cellular systems will not easily be

Correspondence: Dr. Jasminka Godovac-Zimmermann, Centre for Molecular Medicine, University College London, 5 University Street, London WC1E 6JJ, UK
E-mail: j.godovac-zimmermann@ucl.ac.uk
Fax: +44-171-209-6211

Abbreviations: CHO, Chinese hamster ovary; **ET-1,** endothelin; **EMCS,** *N*-(ε-maleimidocaproyloxy)succinimide; **MKBK,** Met-Lys-bradykinin; **PDGF,** platelet-derived growth factor; **TOF,** time of flight

amenable to analysis by genomic methods and they emphasize that understanding of cellular signal transduction systems will require methods for sensitive, highly parallel analysis of large numbers or proteins, *i.e.* proteomics [8, 9].

The term proteomics seems in present vernacular to be synonymous with attempts to map all cellular proteins on 2-D gels. In attempting to establish proteomic methods for analysis of signal transduction systems, it did not seem sensible to identify many thousand proteins on 2-D gels prior to investigating the interesting cellular phenomena and we were therefore interested in establishing methods which would allow the analyses to concentrate on proteins which are more or less directly involved in the signal transduction system of interest. Since phosphorylation of proteins is known to be a common characteristic of signal transduction systems, following 2-D electrophoresis of total cellular proteins, we have used anti-phosphotyrosine and anti-phosphoserine antibodies to detect large numbers (several hundred) of phosphorylated cellular proteins. Since for signal transduction systems it is possible to initiate a specific cellular response at a defined time through stimulation of the membrane receptor of interest, large numbers of proteins involved in signal transduction pathways have been detected from the kinetics of phosphorylation/dephosphorylation following receptor stimulation. These proteins have been identified using a combination of mass fingerprinting *via* MALDI-TOF mass spectrometry and partial peptide sequencing using ion trap electrospray mass spectrometry. Examples are given for fibroblasts following stimulation with either platelet-derived growth factor (PDGF) or endothelin. The results which have been obtained suggest that the use of similar selective stimulation and detection strategies will in many cases be more efficient than trying to map all cellular proteins and leads us to suggest that the term "functional proteomics" is appropriate to characterize strategies based on selective excitation and observation of specific cellular networks of functionally interacting proteins.

In addition to identifying protein components of cellular interaction networks, in many cases it will be essential to identify post-translational modifications of individual proteins in detail. This is especially true for receptors of signal transduction systems where modifications such as phosphorylation or acylation of the receptor are known to play decisive roles in coupling to signaling pathways, desensitization, internalization and resensitization [1]. We have recently established new methods for single-step, very mild isolation of peptide membrane receptors that yield sufficient highly pure receptor for characterization of post-translational modifications by mass spectrometry. Results obtained for endothelin receptor from bovine lung and for bradykinin receptor expressed in transfected Chinese hamster ovary cells reveal patterns of post-translational modifications which are substantially more complex than previously suspected.

2 Materials and methods

2.1 Cell culture

Mouse fibroblasts (NIH 3T3 cells) were cultured in DMEM media supplemented with 10% fetal bovine serum, 100 units/mL of penicillin and 100 μg/mL of streptomycin in a water-saturated, 10% CO_2 atmosphere at 37°C in 75 cm^2 polystyrene petri dishes. Confluent cultures were made quiescent by extensive washing with PBS and subsequently by switching to DMEM media containing 0.5% fetal calf serum for 24 h. Chinese hamster ovary (CHO) cells transfected with rat bradykinin B_2 receptor cDNA [10] were grown in Ham's F12 medium containing 10% fetal calf serum, and 50 μg/mL of streptomycin in a water-saturated, 5% CO_2 atmosphere at 37°C. Cells grown to 80% confluency were harvested for membrane preparation by scraping in 100 mM NaCl, 10 mM Pipes at pH 6.8. The cells obtained were washed twice with the same buffer and subsequently resuspended in 2 mM EDTA, 10 mM Pipes at pH 6.8. All subsequent steps were carried out at 4°C. Cell rupture was performed by Ultra-Turrax (IKA-Labortechnik, Staufen, Germany) at 20 000 rpm for 2 min. Membranes were collected by centrifugation at 30 000 × *g* for 10 min and resuspended in 100 mM NaCl, 2 mM EDTA, 10 mM Pipes at pH 6.8 to a final protein concentration of 2 mg/mL. The membrane suspension was stored at –80°C.

2.2 PDGF induction

Quiescent NIH 3T3 cells were incubated with or without platelet derived growth factor B (PDGF-B; 50 ng/mL) for various periods of time at 37°C. After the induction time, the cells were washed twice with ice-cold PBS containing 30 mM sodium pyrophosphate, 50 mM NaF, 100 mM Na_3VO_4. The cells were lysed with 240 μL of buffer containing 0.3 % SDS, 200 mM DTT, 28 mM Tris-HCl, 22 mM Tris, 100 μM Na_3VO_4, 1 mM NaF, 2 mM EDTA, 1 mM sodium pyrophosphate, Complete^{tm} mini protease inhibitor cocktail and DNAse I/RNAse A cocktail (DNAse I, 25 μg/mL; RNAse A, 7 μg/mL; Boehringer Mannheim, Germany). After centrifugation at 14 000 rpm for 20 min at 4°C, the proteins in the supernatant were precipitated by addition of 80% v/v of cold acetone and maintenance on ice for 40 min. The precipitated proteins were collected by centrifugation at 14 000 rpm for 20 min at 4°C and subsequently were dissolved in 200 μL of IEF sample buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, 1% Triton X-100,

0.8% Pharmalyte 3–10, 1% DTT, 20 mM Tris and 5 mM Pefabloc. The dissolved proteins were stored at –80°C.

2.3 2-D electrophoresis and immunoblot analysis

Ready-to-use Immobiline DryStrips, pH 4–7, were reswollen overnight in 250 μL of protein solution (750 μg of proteins for Coomassie Brilliant Blue staining and 300 μg of proteins for immunoblotting). The IEF was carried out to a total of 70 kVh. Prior to SDS gel electrophoresis the gels were incubated in a solution of 10 mg/mL of DTT in equilibration buffer for 20 min and subsequently in a solution of 45 mg/mL of iodoacetamide in the same buffer for 20 min. SDS-PAGE was performed in 11.5% polyacrylamide gels (160 × 140 × 1 mm) at 11°C at a constant current of 50 mA for 3 h. Gels were stained with Coomassie Brilliant Blue or a Sigma Rapid Silver Staining Kit (Sigma, St. Louis, MO, USA). Proteins from SDS-PAGE were electroblotted onto a nitrocellulose membrane (Schleicher and Schuell, Dassel, Germany). The blot was incubated with anti-phosphoserine antibody at a 1:250 dilution or antiphosphotyrosine antibody at a 1:200 dilution in 3% BSA in TBS at room temperature for 2 h. The protein-antibody complexes were visualized with alkaline phosphatase-conjugate goat anti-mouse immunoglobulin G at a 1:5000 dilution and stained with nitrol blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (ready-made solution from Boehringer, Mannheim, Germany). Gels and blots were scanned using an OmniMedia Scanner 12cx (XRS, Torrance, CA, USA) and UMAX UC1260 (UMAX Data System, Hsinchu, Taiwan) ROC scanner. Image analysis and spot matching were performed using the software package from BioImage (Ann Arbor, MI, USA) on a Sun workstation or the NIH Image program available at the NIH web site (http://rsb.info.nih.gov/nih-image). For determination of the p*I* we used the values specified by the manufacturer of the p*I* strips (Pharmacia, Uppsala, Sweden).

2.4 Synthesis of a polyadenylated receptor ligands

Endothelin-1 and Met-Lys-bradykinin derivatized with *N*-(ε-maleimidocaproyloxy)succinimide (EMCS) at the ε-NH_2 group of Lys were prepared by a recently described synthesis [11]. The resultant *N*-(ε-maleimidocaproyloxy)-succinimide-endothelin-1 (EMC-ET) and *N*-(ε-maleimidocaproyloxy)succinimide-Met-Lys-bradykinin (EMC-MKBK) were purified on an Aquapore RP-300 (Applied Biosystems, Foster City, CA, USA) column using a gradient of 0–70% acetonitrile in 0.1% trifluoroacetic acid. $(dA)_{30}$-5′-SS-R was used for attachment of a 30-mer (dA) to EMC-ET and EMC-MKBK [11]. The resultant $(dA)_{30}$-5′-S-*N*-(ε-maleimidocaproyloxy)succinimide-endothelin and $(dA)_{30}$-5′-*S*-*N*-(ε-maleimidocaproyloxy)succinimide-Met-Lys-bradykinin were purified on a Sephasil C18 column (Pharmacia, Freiburg, Germany) by applying a linear gradient of 2–70% acetonitrile in 0.1 M triethylammonium acetate, 2 mM Bu_4NHSO_4.

2.5 Affinity purification of endothelin A, endothelin B and bradykinin B_2 receptor

Frozen bovine lung membranes containing endothelin receptors were thawed and suspended in two volumes of 20 mM potassium phosphate, pH 7.4, containing 0.40% digitonin, 0.25% CHAPS, 0.5 M NaCl, 20 mM EDTA, 1 μg/mL RNase, and protease inhibitors [12]. CHO membranes containing B_2 receptor (600 μL) were suspended in two volumes of 20 mM potassium phosphate, pH 7.4, containing 4 mM CHAPS, 500 mM NaCl, 20 mM EDTA. For either membrane preparation, the mixture was gently stirred at 4°C for 1 h and was then centrifuged at 100 000 × *g* for 1 h at 4°C. The supernatants were incubated with 0.2 nmol of $(dA)_{30}$-5′-S-EMC-ET or $(dA)_{30}$-5′-S-EMC-MKBK for 1 h at 4°C, 30 mg of oligo-(dT)-cellulose was added and the suspension was gently agitated at 4°C for 2 h. The oligo-(dT)-cellulose was pelleted by centrifugation (4°C, 1000 × *g*, 5 min) and packed into a micro column. The column was washed with 10 mL of buffer A at 4°C and afterwards eluted with 10 mM Tris-HCl, pH 7.4, 4 mM CHAPS, 1 mM EDTA. Fractions (50 μL each) containing receptor were identified by SDS-polyacrylamide gel electrophoresis and immunoblot analysis.

2.6 SDS-PAGE and in-gel tryptic protein digestion

Ten mL of each fraction from the oligo-(dT)-cellulose column were mixed with an equal volume of sample buffer containing 6% SDS, 10% v/v 2-mercaptoethanol, 20% v/v glycerol, 0.01% bromophenol blue, and 0.25 M Tris-HCl, pH 6.8, and brought to 95°C for 2 min. Electrophoresis was performed in 10 or 12.5% polyacrylamide gels in the presence of 0.1% SDS at a constant current of 40 mA for 1.5 h. The resolved proteins were visualized by Coomassie Blue staining. After visualization, the gel was destained with a solution of 25 mM ammonium bicarbonate / 50% acetonitrile. The band corresponding to either the endothelin receptor or the bradykinin B_2 receptor was excised and digested in the gel according to [13] as modified previously [11].

2.7 Mass spectrometric analysis

For MALDI mass spectrometry, samples were dissolved in 5 μL of 50% acetonitrile, 0.1% trifluoroacetic acid (TFA)

and sonicated for several minutes. Aliquots of 0.5 μL were applied to a target disk and allowed to air-dry. Subsequently, 0.3 μL of matrix solution (1% w/v of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% v/v TFA) was applied to the dried sample and again allowed to dry. Spectra were obtained using a Bruker Biflex MALDI-TOF mass spectrometer (Bremen, Germany). MS/MS analysis was carried out using a Finnigan MAT (San Jose, CA, USA) LCQ ion trap mass spectrometer coupled on-line with an HPLC system (Hewlett-Packard 1090; Avondale, PA, USA). For the interpretation of MS and MS/MS spectra of protein digests we used the Sherpa software [14], the MS-Fit, MS-Tag and MS-Product programs available at the UCSF web site (http://rafael.ucsf.edu/cgi-bin/msfit), the PepSearch programs at the EMBL web site (http://.www.mann.embl-heidelberg.de/Services/PeptideSearch/FR_peptideSearchForm.html), and the PepFrag program at the Rockefeller University web site (http/:prowl.rockefeller.edu/).

3 Results and discussion

3.1 Detection of proteins involved in PDGF and endothelin signal transduction pathways

NIH 3T3 fibroblast cells from mouse were stimulated with PDGF or endothelin, the reaction was terminated at various times and the total cellular proteins were separated by 2-D electrophoresis. Although we have presently limited our 2-D electrophoretic analysis to p*I* 4–7 and molecular size range of 15–120 kDa, about 3000 protein spots could be detected in gels stained with silver or Coomassie Blue (Fig. 1). This is in line with estimates that although the mouse genome may contain 50–100 000 proteins, a much smaller number is likely to be expressed and detected by 2-D gel electrophoresis in a particular cell type [15]. Although various proteome projects to obtain 2-D maps of the several thousand proteins in a given cell type are currently in progress, this is still a formidable task. We have therefore started our analysis of the responses to stimulation of the PDGF receptor by identifying proteins containing phosphoserine and phosphotyrosine using detection with anti-phosphotyrosine and anti-phosphoserine antibodies on immunoblots obtained for gels from unstimulated cells and for gels from cells with termination of stimulation at 3, 5, 10, and 20 min (Fig. 2). We observed at least 260 proteins containing phosphotyrosine; at least 44 proteins in the Tyr series showed strong intensity changes as a function of time after stimulation with PDGF. Parallel studies involving phosphorylation at Ser residues revealed that with these methods about 300 proteins could be observed and that about 50 of these proteins showed changes in phosphorylation following stimulation with PDGF (Soskic *et al.*, submitted).

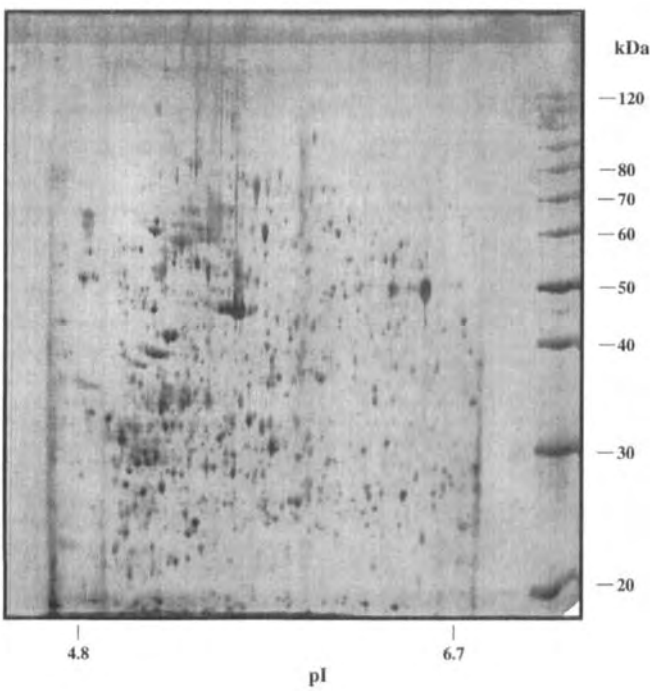


Figure 1. Coomassie blue-stained 2-D gel of total proteins from NIH3T3 mouse fibroblasts.

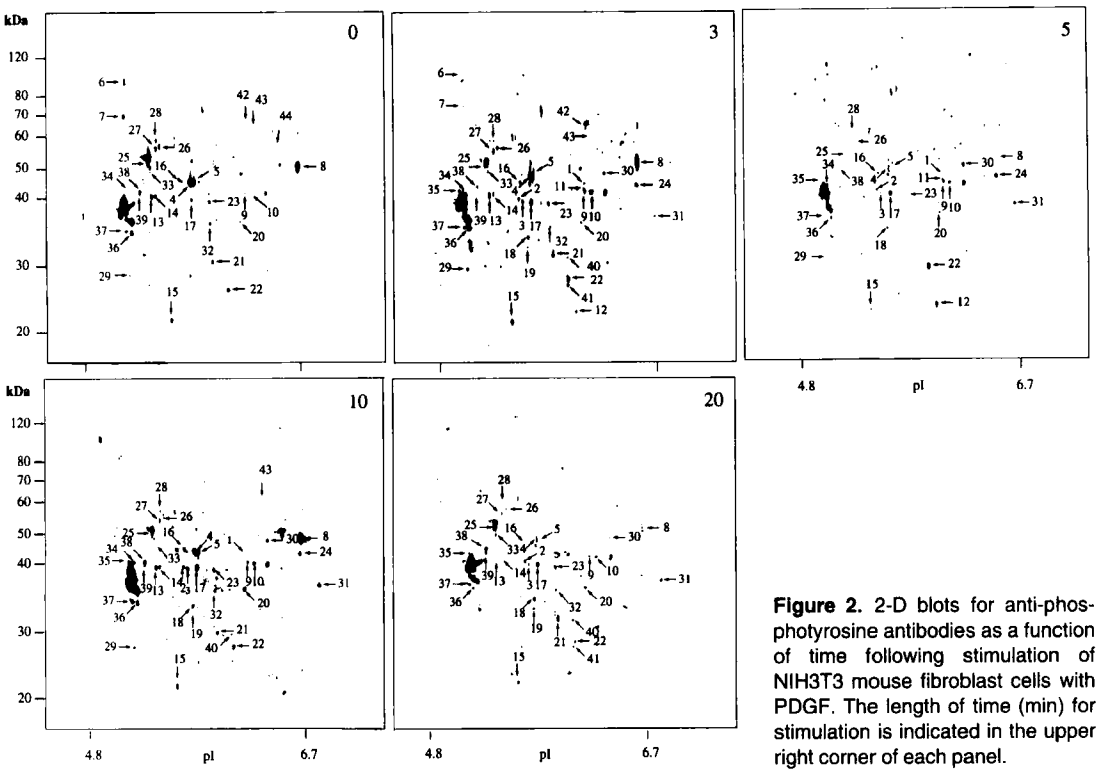


Figure 2. 2-D blots for anti-phosphotyrosine antibodies as a function of time following stimulation of NIH3T3 mouse fibroblast cells with PDGF. The length of time (min) for stimulation is indicated in the upper right corner of each panel.

Individual protein phosphorylations are readily followed as a function of time after stimulation, as indicated by the expanded plots in Fig. 3, showing blots covering a molecular size range of 35–60 kDa and p*I* 6.1–6.8. Observation of the intensities on the immunoblots for individual proteins following stimulation revealed a wide diversity of kinetic behavior (Fig. 3). Some proteins show moderate changes in intensity over the time course of the experiment (*e.g.*, spot s50 in Fig. 3), some are initially undetectable but rapidly show intense spots following stimulation (*e.g.*, s30), some show strong spots initially, but rapidly decrease in intensity (*e.g.*, s52), while other proteins show more gradual changes (*e.g.*, s20). Although such intensity changes might in some cases reflect multiple phosphorylation sites, and hence changes in the number of antibodies reacting with a single protein, it seems probably that in most cases the intensities in the blots will instead reflect the proportion of a protein which is present in phosphorylated form. For example, there were at least 13 proteins detected by the phosphotyrosine antibodies which were not observed in unstimulated cells, but which became observable 3 min after stimulation. For these proteins it seems likely that in the basal (unstimulated) state, virtually all of the protein is unphosphorylated, but is rapidly converted fully or partially to a phosphorylated form following stimulation with PDGF. Although it cannot be guaranteed that the presently used antibodies detect all proteins containing phosphotyrosine, it is clear that large numbers of phosphorylated proteins can be detected and that a surprisingly large number of these proteins show substantial changes in the intensities observed on the antibody blots. The unexpectedly large number of proteins which showed changes in phosphorylation of tyrosine and serine may be related to the diverse range of cellular and disease processes in which the PDGF signaling system seems to be involved (including, for example, cell growth regulation, differentiation, chemotaxis, tumor progression, wound healing, atherosclerosis, inflammatory joint disease, and fibrotic conditions) [16–20].

Initial experiments involving stimulation of NIH 3T3 fibroblast mouse cells with endothelin have also been carried out. In this case, of the approximately 260 proteins that could be detected with the phosphotyrosine antibodies, major changes in intensity on the immunoblots were observed for fewer proteins. Comparison of proteins in the same molecular weight and p*I* ranges revealed that for stimulation with endothelin (Fig. 4), different proteins and

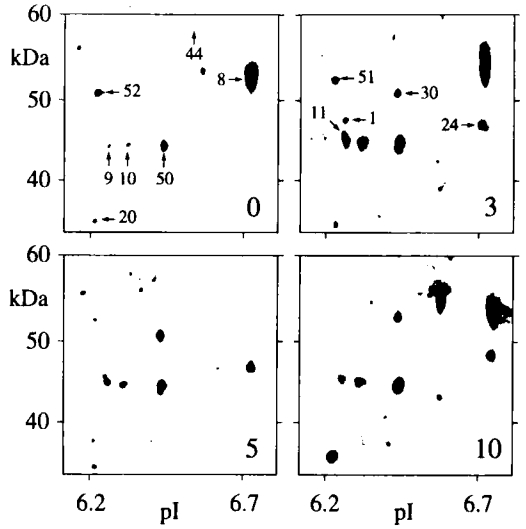


Figure 3. Expansion of 2-D blots for anti-phosphotyrosine antibodies in the ranges p*I* 6.1–6.8 and M_r 35 000–60 000 as a function of time (min, lower right on each panel) following stimulation of NIH3T3 mouse fibroblast cells with PDGF.

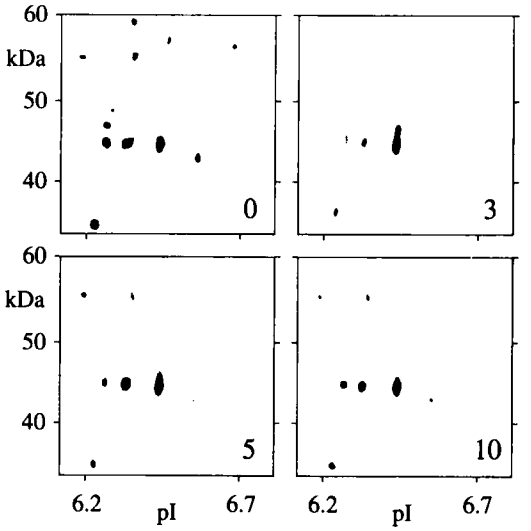


Figure 4. Expansion of 2-D blots for anti-phosphotyrosine antibodies in the ranges p*I* 6.1–6.8 and M_r 35 000–60 000 as a function of time (min, lower right on each panel) following stimulation of NIH3T3 mouse fibroblast cells with endothelin.

different kinetic patterns were detected (compare Figs. 3 and 4). These differences provide a strong indication that the changes in phosphorylated proteins that are detected as a function of time after stimulation indeed reflect specific cellular interaction networks connected with a specific receptor.

3.2 Identification of proteins with mass spectrometry

We have begun the process of identifying the several hundred phosphorylated proteins by concentrating on some of the proteins that showed large intensity changes as a function of time following stimulation. Protein spots detected on the immunoblots were excised from the corresponding gel and subjected to in-gel digestion with trypsin [13]. Even for those spots which were quite weak on the immunoblots, *e.g.*, spot s44 at 0 min (Fig. 3), adequate material to obtain large numbers of accurate masses of peptides and peptide fragments was available for mass fingerprinting by MALDI-TOF mass spectrometry. We estimate that the weakest spots used for the mass spectrometric analysis contained less than 1 picomol. As reported in more detail elsewhere (Soskic *et al.*, submitted; [21–24]), initial identification of proteins from mass fingerprints were performed using a maximum ±0.3 Da mass tolerance, a single trypsin missed cleavage, and no phosphorylation modification. We required that the peptides identified covered a minimum of 18–20% of the protein sequence. On average, about 6–12 peptide masses were matched. Once the protein was identified, manual corrections based on the theoretical protein digest were used to identify peptides with up to three missed tryptic site cleavages and phosphorylated peptides. The identity of the protein was then confirmed by subjecting selected peptides to MS/MS analysis using ion trap electrospray mass spectrometry of the unseparated peptide mixture previously analyzed by MALDI-TOF. Examples of the MALDI mass spectrometric data used to identify proteins showing large changes in phosphorylation are shown in Table 1.

Together with earlier experiments that identified other proteins showing major changes in phosphorylation following stimulation with PDGF (Soskic *et al.*, submitted), a picture of the type of information that can be gained by these types of experiments has now emerged. In particular, proteins which were previously known to be involved in the PDGF signaling system (*e.g.*, ERK 1 kinase, serine/threonine protein kinase akt, vimentin, Src cortactin substrate, protein tyrosine phosphatase SYP, and phospholipase C alpha), proteins which were previously known to be involved in other signaling systems, but not the PDGF system (*e.g.*, proto-oncogene tyrosine kinase fgr, phosphotyrosine phosphatase PTP-2, and protein serine/threonine kinase) and proteins which were not previously

Table 1. MALDI mass spectrometry analysis of proteins S42, S43 and S44

Start residue	End residue	Expected mass	Measured mass	Sequence
S42 Protein tyrosine phosphatase SYP (*Mus musculus*)				
6	35	3326.7	3326.9	WFHPNITGVEAENLLLTRGVDGSFLARPSK
245	268	2663.9	2664.6	QGFWEEFETLQQQECKLLYSR
326	351	3118.6	3118.1	FDSLTDLVEHYKKNPMVETLGTVLQLK
367	378	1514.7	1514.0	CVKYWPDEYALK
370	378	1184.6	1184.7	YWPDEYALK
517	531	1913.2	1912.4	FIYMAVQHYIETLQR
	6 peptides cover 18% (105/585 AAs)			
S43 Phospholipase C-alpha (*Mus musculus*)				
63	73	1192.3	1192.0	LAPEYEAAATR
94	103	1085.2	1085.0	YGVSGYPTLK
129	145	1888.1	1888.5	KQAGPASVPLRTEEEFK
147	160	1588.8	1588.3	FISDKDASVVGFFR
152	172	2374.6	2374.3	DASVVGFFRDLFSDGHSEFLK
183	193	1259.4	1259.2	FAHTNIESLVK
194	213	2303.5	2303.5	EYDDNGEGITIFRPLHLANK
194	217	2823.1	2822.7	EYDDNGEGITIFRPLHLANKFEDK
304	331	3051.4	3051.7	KTFSHELSDFSLESTTGEVPVVAIRTAK
335	343	1173.3	1173.6	FVMQEEFSR
347	362	1972.2	1971.6	ALEQFLQEYFDGNLKR
471	481	1398.5	1398.6	ELNDFISYLQR
	12 peptides cover 32% (164/504 AAs)			
S44 Protein serine/threonine kinase (*Homo sapiens*)				
87	96	1203.6	1203.3	TFHVDSPDER
114	145	3476.3	3475.8	APGEDPMDYKCGSPSDSSTTEEMEVAVSKARAK
147	156	1246.4	1246.4	VTMNDFDYLK
161	172	1347.3	1347.6	GTFGKVILVREK
209	224	1945.7	1946.2	HPFLTALKYAFQTHDR
438	467	3520.9	3520.8	YFDDEFTAQSITITPPDRYDSLGLLELDQR
	6 peptides cover 23% (111/481 AAs)			

known to be involved in signal transduction (plexin-like protein) have been identified. On the one hand, the observation of proteins already known to be related to PDGF signaling provides confirmation that experiments of this type are capable of detecting downstream responses following PDGF stimulation. On the other hand, the observation of unexpectedly large numbers of proteins which show changes in phosphorylation following PDGF stimulation together with the identification of a substantial number of proteins which were either previously unknown in PDGF signaling or in any signal transduction system confirms that experiments of the type reported here can provide a rich source of new information on the complex interaction networks involved in signal transduction. At present, our efforts are concentrated on the identification on 2-D gels of all proteins showing substantial changes in phosphorylation at tyrosine or serine residues following stimulation with PDGF or endothelin. The availability of such maps should greatly simplify, and perhaps supplant, current efforts by more conventional approaches to establish hierarchical relationships within and between different signaling pathways.

3.3 Post-translational modifications of endothelin and bradykinin receptors

Although a variety of experimental strategies can be envisaged for the establishment of hierarchical relationships in complex signaling networks, *e.g.*, use of inhibitors of specific pathways, gene knockout or knockdown methods, or mutation of selected pathway components, a special hierarchical position is occupied by the receptor that initiates all the downstream responses. Indeed, there is considerable evidence that post-translational modifications of the receptor itself are an important factor in the regulation of a signaling pathway. For this reason we

have recently been interested in the development of improved methods for characterization of post-translational modification of cellular receptors and in whether specific modifications of such receptors can be related to specific signaling responses. Due to their low cellular abundance and difficulty with purification, the direct chemical analysis of post-translational modifications of membrane receptors has been a notoriously difficult task. To date, post-translational modifications for only a few prototypical receptors, *i.e.*, rhodopsin and β-adrenergic receptor, have been investigated in detail [25–29] and much of the available data has been obtained under *in vitro* conditions [30]. Recent results have emphasized that there may be discrepancies between such data and mutational analyses obtained under *in vivo* conditions [31].

We have recently devised a new single-step, very mild "fishhook" strategy for the purification of peptide receptors which is based on methods used for isolation of mRNA. For example, a "fishhook" for the bradykinin B_2 receptor was constructed by linkage of a $(dA)_{30}$ oligonucleotide to ^{2}Lys of Met-Lys-bradykinin *via* a special bifunctional cross-linking reagent (insert, Fig. 5). The bradykinin receptor was expressed in transfected CHO cells and a highly pure receptor sample suitable for detailed analysis of post-translational modifications was then obtained as follows: (i) CHO cells at 80% confluency were harvested and cellular membranes isolated; (ii) the membranes were solubilized in CHAPS detergent and the bradykinin B_2 fishhook added to bind the bradykinin B_2 receptor; (iii) the fishhook-receptor complex was bound to an oligo-(dT)-cellulose column and washed under high salt concentrations that favor formation of a (dA:dT) double helix; (iv) the B_2 receptor was eluted under very low salt conditions that destabilize (dA:dT) helices, and (v) fractions containing the B_2 receptor were further purified by PAGE.

This approach has been applied to obtaining receptors with the post-translational modifications extant under *in vivo* conditions for both the bradykinin B_2 receptor from transfected CHO cells (Soskic *et al.*, submitted) and the endothelin receptor from bovine lungs [11] and should be applicable to other receptors.

Post-translational modifications of both the bradykinin B_2 receptor and the endothelin receptor have been analyzed by mass spectrometric methods. Following in-gel tryptic hydrolysis, virtually the entire sequence of these receptors could be scanned for post-translational modifications by detecting peptides with phosphorylation or palmitoylation using MALDI-TOF mass spectrometry. The identity of the peptide as well as the nature and site of the post-translational modification was then confirmed by ion trap electrospray mass spectrometry of the unseparated tryptic peptide mixture. An example is shown in Fig. 5 for the peptide ^{344}FRKKSREVYQAICRK^{358} from the bradykinin B_2 receptor. MALDI-TOF mass spectrometry showed an increase in mass of 80 Da, indicating that the peptide is phosphorylated at ^{348}Ser or ^{352}Tyr. Using ion trap ESI mass spectrometry, the masses of the *y* and *b* parent ions were consistent with monophosphorylation of the peptide and the *y* and *b* series of fragment ions *y*15, *y*14, *y*13, *y*12, *y*9, *y*8 and 7 as well as *b*15, *b*13, *b*12, *b*11, *b*9, *b*8 and *b*7 (Fig. 5) were sufficient to determine that ^{352}Tyr and not ^{348}Ser is the phosphorylation attachment site.

Analyses of the bradykinin B_2 and endothelin B receptors has revealed that the number of post-translational modifications is larger than previously anticipated. In the case of endothelin B receptor from bovine lung tissue, post-translational phosphorylation has been shown for seven serine and one tyrosine residues as well as palmitoylation at two cysteine residues (Fig. 6A). These modifications

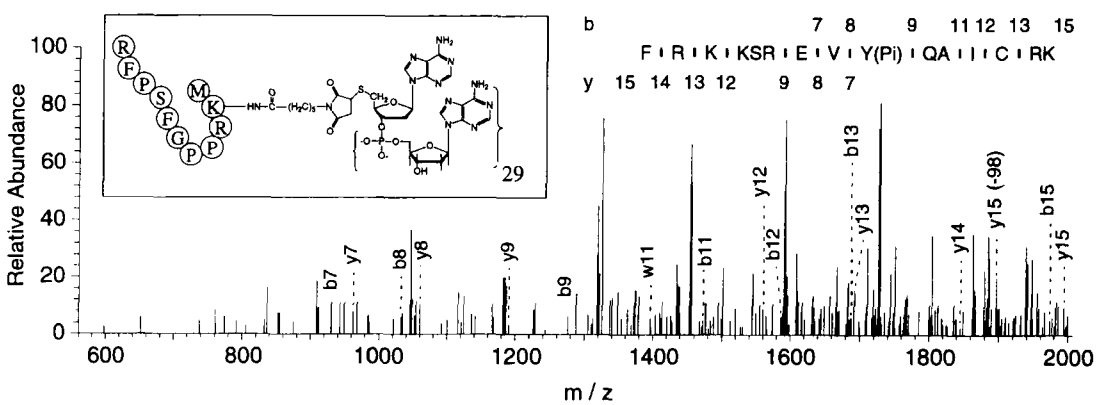


Figure 5. Ion trap electrospray MS/MS spectra of phosphopeptide ^{344}FRKKSREVY(Pi)QAICRK^{358} $(M+H)^+$ of bradykinin B_2 receptor demonstrating phosphorylation at ^{352}Tyr. Insert, structure of the Met-Lys-bradykinin fishhook.

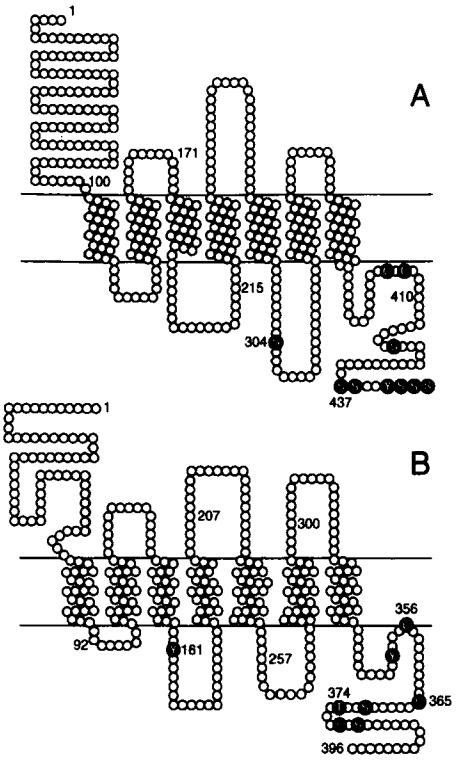


Figure 6. (A) Sites of post-translational modifications of the endothelin B receptor. (B) Sites of post-translational modifications of the bradykinin B_2 receptor. (C) Multiple forms of post-translational modifications detected by mass spectrometry for peptides ^{350}EVYQAICRK^{358} and ^{359}GGCMGESVQMENSMGTLRTSISVDR^{383} of bradykinin B_2 receptor (see Section 3.3). The residues which are phosphorylated (Y, S) or palmitoylated (C) are indicated by white letters on a black circle.

are clustered in the cytoplasmic carboxy-terminal domain of the receptor and may be involved with desensitization of the receptor *via* internalization [32]. The third cytoplasmic domain is also phosphorylated at ^{304}Ser. The observation that a single site Ser-Asn mutation at this residue leads to Hirschprung's disease [33–35] has led us to suggest that phosphorylation of ^{304}Ser is likely to be important to the signaling characteristics of this receptor [11, 36]. For the bradykinin B_2 receptor, post-translational phosphorylation has been found for 4 Ser, 1 Thr and 2 Tyr residues as well as palmitoylation at 1 Cys residue (Fig. 6B). These modifications are again clustered in the cytoplasmic carboxy-terminal domain of this receptor, where it has also been postulated that there is a causal relationship between these modifications and internalization of the B_2 receptor [32]. Phosphorylation has also been detected for residue ^{161}Tyr of the B_2 receptor (Fig. 6), which is part of the DRY motif that is highly conserved in G-protein-coupled receptors. For the B_2 receptor there is evidence for an important role of this residue in signaling [32], suggesting that phosphorylation of ^{161}Tyr may be an important regulator of signaling in the B_2 receptor and that phosphorylation of the DRY motif might be important for other G-protein-coupled receptors.

In the case of the bradykinin B_2 receptor, correlations between some post-translational modifications have been demonstrated (Fig. 6C). For example, phosphorylation of ^{352}Tyr and palmitoylation of ^{356}Cys appears to be mutually exclusive (Soskic *et al.*, submitted) and may constitute a control mechanism for receptor internalization [37, 38]. For the carboxy-terminal residues 359–383 of the B_2 receptor, a particularly intriguing set of post-translational phosphorylations has been observed for the basal (unstimulated) form of the receptor. In particular, of the four Ser and two Thr residues in this sequence, phosphorylation has been detected at all residues except ^{377}Thr, but only in the form of three very specific forms of diphosphorylation (Fig. 6C). The observations that only diphosphorylation apparently occurs (Soskic *et al.*, submitted), and that the sites are correlated, has important implications for attempts to analyze functional aspects of the bradykinin B_2 receptor – the present results suggest that in mutation experiments it is potentially possible to create novel phosphorylation patterns that do not exist in the natural receptor, with potentially aberrant physiological responses [30]. It is clear from the results which have recently been obtained for endothelin B and bradykinin B_2 receptors that the patterns of post-translational modifications for these G-protein-coupled receptors are much more complex than previously thought.

4 Concluding remarks

The results which have so far been obtained for signal transduction systems are very promising in that clear evidence has been obtained that under *in vivo* conditions: (i) complex networks of interactions can be observed through identification of large numbers of functionally related proteins; (ii) for signal transduction systems the

concept of functional proteomics provides a means to concentrate efforts on characteristics clearly associated with the functional system under study and to restrict the necessary analytical effort to manageable proportions, and (iii) detailed alterations in individual proteins are clearly accessible, *e.g.*, despite the low abundance levels and the complexity of the post-translational modifications of bradykinin or endothelin receptors, methods have been found to characterize the nature of these alterations. Although methodological improvements in the separation of individual proteins and in the sensitivity and rapidity of their subsequent identification/characterization can certainly be anticipated in the near future, the presently available analytical methodology is already clearly adequate for investigation of important cellular processes. Although present efforts are concentrated on identifying the protein components of such processes, it is apparent that, with appropriate ingenuity in the establishment and manipulation of the cellular systems, spatial, temporal and functional characteristics of the complex networks of cellular interactions will soon be the major focus of interest.

Received September 1, 1998

5 References

[1] Tae, H., Grossmann, M., Inhae, J., *J. Biol. Chem.* 1998, *273*, 17299–17302.

[2] Nill-Eubanks, D., Burnstein, E. S., Spaalding, T. A., Brauner-Osborn, H., Brann, M. R., *J. Biol. Chem.* 1996, *271*, 3058–3065.

[3] Nawes, B. E., Luttrell, L. M., Exum, S. T., Lefkowitz, R. J., *J. Biol. Chem.* 1994, *269*, 15776–15785.

[4] Krueger, K. M., Daaka, Y., Pitcher, A. A., Lefkowitz, R. J., *J. Biol. Chem.* 1997, *272*, 5–8.

[5] Strohman, R., *BioTechnology* 1994, *12*, 156–164.

[6] Miklos, G. L., Rubin, G. M., *Cell* 1996, *86*, 521–529.

[7] Williams, K. L., Hochstrasser, D. F., in: Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997, pp. 1–12.

[8] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[9] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *BioTechnology* 1996, *14*, 61–65.

[10] McEachern, A. E., Shelton, E. R., Bhakta, S., Obernolte, R., Bach, C., Zuppan, P., Fujisaki, J., Aldrich, R. W., Jarnagin, K., *Proc. Natl. Acad. Sci. USA*, 1991, *88*, 7724–7728.

[11] Roos, M., Soskic, V., Poznanovic, S., Godovac-Zimmermann, J., *J. Biol. Chem.* 1998, *273*, 924–931.

[12] Hick, S., Heidemann, I., Soskic, V., Müller-Esterl, W., Godovac-Zimmermann, J., *Eur. J. Biochem.* 1995, *234*, 251–257.

[13] Hellman, U., Wernstedt, C., Gonez, J., Heldin, C-H., *Anal. Biochem.* 1995, *224*, 451–455.

[14] Taylor, J. A., Walsh, K., Johnson, R. S., *Rapid. Commun. Mass. Spectrom.* 1996, *10*, 679–687.

[15] Wilkins, M. R., Gasteiger, E., Sanchez, J.-C., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 1505–1508.

[16] James, R., Bradshaw, R. A., *Annu. Rev. Biochem.* 1984, *53*, 259–292.

[17] Aaronson, S. A., *Science* 1991, *254*, 1146–1153.

[18] Raines, E. W., Bowen-Pope, D. F., Ross, R., in: Sporn, M. B., Roberts, A. B., (Eds.), *Peptide Growth Factors and Their Receptors*, Springer-Verlag, New York 1991, pp. 173–262.

[19] Ataliotis, P., Mercola, M., *Int. Rev. Cytol.* 1997, *172*, 95–127.

[20] Smith, E. A., Seldin, M. F., Martinez, L., Watson, M. L., Choudhury, G. G., Lalley, P. A., Pierce, J., Aaronson, S., Barker, J., Naylor, S. L., Sakaguchi, A. Y., *Proc. Natl. Acad. Sci. USA* 1991, *88*, 4811–4815.

[21] Henzel, W. J., Billeci, T. M., Stulis, J. T., Wong, S. C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

[22] Mann, M., Hojrup, P., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 338–345.

[23] Pappin, D. J. C., Hojrup, P., Bleasby, A., *J. Curr. Biol.* 1993, *3*, 327–332.

[24] James, P., Quadroni, M., Carafoli, E., Gonnet, G., *Biophys. Biochem. Res. Commun.* 1993, *195*, 58–64.

[25] Prado, G. N., Taylor, L., Polger, P., *J. Biol. Chem.* 1997, *272*, 14638–14642.

[26] Papac, D. I., Oatis, J. E., Crouch, R. K., Knapp, D. R., *Biochemistry* 1993, *32*, 5930–5934.

[27] Papac, D., Thornburg, K. R., Büllesbach, E. E., Crouch, R., Knapp, D. R., *J. Biol. Chem.* 1992, *267*, 16889–16894.

[28] Ohguro, H., Van Hooser, J. P., Milam, A. H., Palczweski, K., *J. Biol. Chem.* 1995, *270*, 14259–14262.

[29] Ohguro, H., Rudnicka-Nawro, M., Buczylko, J., Zaho, X., Taylor, J. A., Walsh, K. A., Palzweski, K., *J. Biol. Chem.* 1996, *271*, 5215–5224.

[30] Fredericks, Z. L., Pitcher, J. A., Lefkowitz, R. J., *J. Biol. Chem.* 1996, *271*, 13796–13803.

[31] Seibold, A., Januarym, B. G., Friedman, J., Hipkin, R. W., Clark, R. B., *J. Biol. Chem.* 1998, *273*, 7637–7642.

[32] Prado, G. N., Taylor, L., Polger, P., *J. Biol. Chem.* 1997, *272*, 14638–14642.

[33] Puffenberger, E. G., Hosoda, K., Washington, S. S., Nakao, K., deWit, D., Yanagisawa, M., Chakravati, A., *Cell* 1994, *79*, 1257–1266.

[34] Hosoda, K., Hammer, R. E., Richardson, J. A., Greenmstein-Baynash, A., Cheung, J. C., Giaid, A., Yanagisawa, M., *Cell* 1994, *79*, 1267–1276.

[35] Kusafuka, T., Wang, Y., Puri, P., *Hum. Mol. Genet.* 1996, *5*, 347–349.

[36] Auricchio, A., Casari, G., Staiano, A., Ballabio, A., *Hum. Mol. Genet.* 1996, *5*, 351–354.

[37] Moffet, S., Adam, L., Bonin, H., Loisel, T. P., Bouvier, M., Mouillac, B., *J. Biol. Chem.* 1996, *271*, 21490–21497.

[38] Moffet, S., Mouillac, B., Bonin, H., Bouvier, M., *EMBO J.* 1993, *12*, 349–356.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 962–968

Jiří Stulík[1]
Kamila Koupilová[1]
Lenka Hernychová[1]
Aleš Macela[1]
Václav Bláha[1]
Claudia Baaske[2]
Walter Kaffenberger[2]
Dirk van Beuningen[2]

[1]Purkyně Military Medical Academy, Institute of Radiobiology and Immunology, Hradec Králové, Czech Republic
[2]Federal Armed Forces Medical Academy, Institute of Radiobiology, Munich, Germany

Modulation of signal transduction pathways and global protein composition of macrophages by ionizing radiation

It is assumed that the exposure of cells to ionizing radiation modulates their signal transduction pathways, which then govern the early and late radiation-induced alterations in gene expression. In this study we tested the effects of low doses of X-irradiation on the cell signaling and global protein composition of an HL-60 human promyelocytic leukemia cell line differentiated along a macrophage-like cell pathway by 4β-phorbol-12-myristate-13-acetate (PMA). Using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by immunoblotting of anti-phosphotyrosine immunoprecipitates, we found radiation-induced changes in the level of phosphorylation of proteins with molecular masses of 45 and 48 kDa, but in the most intensively stained area, ranging from 54 to 60 kDa, no alterations were observed. When two-dimensional electrophoresis (2-DE) immunoblotting was applied, only proteins from this heavily stained region were visualized and in addition the evident differences in tyrosine phosphorylated protein patterns between nonirradiated and irradiated cells were found in this area. Furthermore, the immunostaining of extracellular signal-regulated kinase 2 (ERK2) which did not prove its tyrosine phosphorylation demonstrated the existence of several ERK2 charge isoforms showing differential expression after X-irradiation. Comparing the whole protein profiles we found after the simultaneous quantitation of 1000 matched spots two proteins whose expression was regulated in an opposite manner in nonirradiated and X-irradiated cells. The quantities of both spots showed increases or decreases by a factor of 2 or more between irradiated and nonirradiated samples and both these changes were statistically significant ($P < 0.05$).

Keywords: Macrophages / X-irradiation / Two-dimensional gel electrophoresis / Protein tyrosine phosphorylation / Extracellular signal-regulated kinase EL 3307

1 Introduction

The generally accepted response of cells to ionizing radiation is associated with the production of free reactive oxygen intermediates and subsequent oxidation of phospholipids and DNA. However, the mechanisms by which damage is sensed after irradiation have not been fully clarified. Recently the concept stressing the active process of tyrosine kinase signal transduction as a key cellular response to radiation-induced oxidative stress was suggested [1]. The cell membrane/cytoplasmic signaling should be triggered simultaneously from DNA damage and oxidized membrane lipids, thereby bypassing the control by normal ligands. The hydrolysis of oxidized lipids provides second messengers like arachidonic acid, diacylglycerol and ceramide, which in turn bind to and activate various kinases. The process of kinase activation can be even further strengthened by the inhibition of membrane-associated protein phosphatases *via* oxidation of their free SH-groups. The activated kinases phosphorylate transcription factors which are then transported to the nucleus where they bind to DNA regulatory sequences and specifically up- or downregulate gene expression. This radiation-modified gene expression underlies a variety of molecular steps involving growth arrest, DNA repair or apoptosis, which are responsible for radiosensitivity or radioresistance of afflicted cells [2].

In our study we focused on both the modulation of tyrosine kinase signal pathways and the global changes in protein expression levels after the exposure of monocyte/macrophage lineage cells to ionizing radiation. We worked with the HL-60 leukemic cell lines differentiated along the macrophage-like cell pathway as a macrophage reference model in studies devoted to humans (Fig. 1) [3]. Furthermore, low doses of X-ray radiation were used in order to have a clear correlation with the biological and clinical response to radiation therapy.

Correspondence: Dr. Jiří Stulík, Institute of Radiobiology and Immunology, Purkyně Military Medical Academy, Třebešská 1575, 50001 Hradec Králové, Czech Republic.
E-mail: jstulik@pmfhk.cz
Fax: +420-49-5513018

Abbreviations: ERK, extracellular signal-regulated kinase; **PMA**, 4β-phorbol-12-myristate-13-acetate

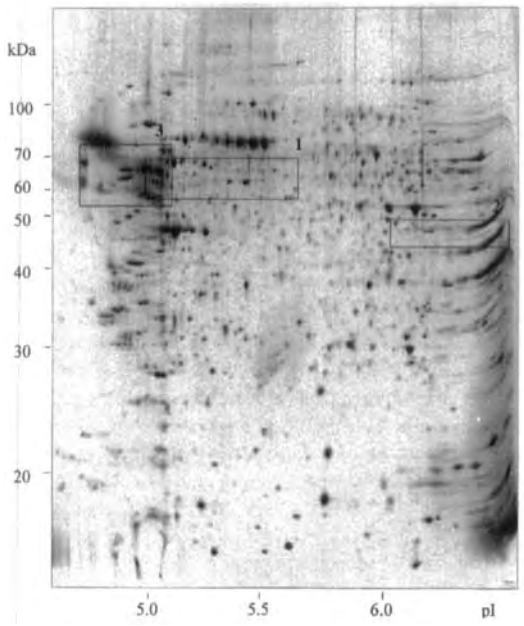


Figure 1. Reperesentative silver-stained protein pattern of PMA-treated cells.Three regions comprising (1) tyrosine phosphorylated proteins, (2) ERK2 isoforms, and (3) protein alterations after X-irradiation are boxed.

2 Materials and methods

2.1 Cell cultivation and sample preparation

The HL-60 cells (American Type Culture Collection, Rockville, MD, USA) were grown in RPMI 1640 medium (Boehringer, Mannheim, Germany) supplemented with 20 % heat-inactivated fetal calf serum (Boehringer Mannheim). The pH was maintained at 7.4 *via* supplementation of the culture medium with $NaHCO_3$. Cells were incubated for three days with 100 nM 4β-phorbol-12-myristate-13-acetate (PMA)-treated cells. On day 3 after PMA treatment the cells were X-ray irradiated with a single dose 2Gy of 240 kV X-rays (Isovolt 320/10; Seifert, Ahrensburg, Germany) and were harvested 2 min, 5 min, or 3 h after irradiation. As control samples, nonirradiated PMA-treated cells with or without additional 3 h cultivation were collected. The cells were washed twice in cold PBS containing 1 mM Na_3VO_4 and immediately lysed either in IEF buffer composed of 9 M urea, 2% Nonidet P-40, 3% CHAPS, 70 mM DTT, 2% Ampholines pH 9–11, protease inhibitor cocktail (Boehringer Mannheim) and phosphatase inhibitors (50 mM NaF, 0.1 mM Na_3VO_4) or in RIPA buffer containing 50 mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, protease inhibitor cocktail, 1 mM Na_3VO_4, 1 mM NaF. We included basic Ampholines in the IEF buffer to inhibit proteolytic activity. The determination of protein concentration was performed by a modified bicinchoninic acid protein assay kit [4]. For the study of the global X-ray-inducible protein alterations, established procedures for 2-DE separation and silver staining were used [5].

2.2 Electrophoresis

A commercial nonlinear immobilized pH 3–10 gradient (Pharmacia-Biotech, Uppsala, Sweden) was applied for the separation of proteins in the first dimension. The Immobiline strips, 18 cm length, were rehydrated overnight in 25 mL of urea/thiourea rehydration solution: 5 M urea, 2 M thiourea, 10 mM DTT, 0.4% carrier ampholytes (Pharmalytes 3–10), 0.5% Triton X-100, 2% sulfobetaines (SB 3–10) [6] and a trace of bromophenol blue. Using sample cups 150 μg protein was loaded on Immobiline strips. For the second dimension the proteins were separated on an SDS 9–16% linear polyacrylamide gradient. Silver-stained gels were scanned using a laser densitometer (4000 × 5000 pixels, 12 bits/pixel; Molecular Dynamics, Sunnyvale, CA, USA) and the 2-DE evaluation was performed using the Melanie II package (Bio-Rad, Richmond, CA, USA). Spots were quantitated in terms of their relative volume (% vol), *i.e.*, digitized staining intensity integrated over the area of an individual spot divided by the sum of integrated staining intensities of all spots and multiplied by 100. We used the unpaired Student T-test for determination of significant differences at the levels of $P < 0.05$. Immunoprecipitation of proteins phosphorylated on tyrosine residues were carried out according to an Upstate Biotechnology protocol using agarose-conjugated anti-phosphotyrosine monoclonal antibody, clone 4G10 (Upstate Biotechnology, Lake Placid, NY, USA). Each sample, prepared in duplicate, was resuspended in 40 μL SDS-PAGE sample buffer and denatured at 100°C for 5 min. The proteins were separated on 10% SDS-PAGE according to Laemmli and Favre [7]; in this case 15 μL of immunoprecipitates or 30 μg of RIPA buffer cell lysates were loaded. For 2-DE analysis 50 μL of immunoprecipitates or 250 μg of RIPA buffer cell lysates were diluted in a total volume of 350 μL of the rehydration solution (7 M urea, 2 M thiourea, 4% CHAPS, 0.5% Triton X-100, 40 mM DTE, 20 mM Tris, 0.4% Pharmalytes pH 3–10) and used for in-gel sample loading [8].

2.3 Immunoblotting

The detection of tyrosine-phosphorylated proteins and extracellular signal-regulated kinases (ERK) was done by immunoblotting experiments. Separated proteins were electrotransfered to 0.2 μm PVDF membrane (Boehringer

Mannheim) using Transblot Cell (Bio-Rad) at 0.1 mA, overnight. The membranes were blocked for 1 h at room temperature in 1% blocking solution (BM chemiluminescence blotting kit; Boehringer Mannheim) and then incubated overnight either with horseradish peroxidase-coupled anti-phosphotyrosine monoclonal antibodies, clone 4G10 (Upstate Biotechnology) diluted 1:1150 in 0.5% blocking solution or with mouse monoclonal anti-ERK antibodies (pan ERK, clone 16; Transduction Laboratories, Lexington, KY, USA; or monoclonal anti-ERK2 antibody, Santa Cruz Biotechnology, Santa Cruz, CA, USA) diluted 1:5000 and 1:400 in 0.5% blocking solution, respectively. For ERK detection the secondary horseradish peroxidase-conjugated anti-mouse antibody (Santa Cruz Biotechnology) diluted 1:1000 in 0.5% blocking solution was applied for 45 min. The bound antibodies were visualized with enhanced chemiluminescence (BM chemiluminescence blotting kit; Boehringer Mannheim) and exposed to X-ray film Bio Max Light-1 (Sigma, Prague, Czech Republic). The molecular weights and isoelectric points of individual proteins were determined using 2-D polypeptide standards (Bio-Rad).

3 Results and discussion

Previously, we observed by means of SDS-PAGE followed by immunostaining that in short time intervals after X-irradiation (2 and 5 min) the level of tyrosine phosphorylation of proteins with molecular masses ranging from 44 to 64 kDa was influenced in PMA-treated cells (unpublished observations). In this study we wanted to analyze the proteins phosphorylated on tyrosine residues in more detail; therefore, we applied 2-DE analysis in addition to one dimensional (1-D) protein separation. Since higher loading of proteins is required for 2-DE we first made an immunoprecipitation of tyrosine-phosphorylated proteins. Figure 2A shows the 1-D anti-phosphotyrosine immunoblot of immunoprecipitates collected from nonirradiated as well as irradiated PMA-treated cells. Compared with our earlier results, the radiation-induced changes in the level of phosphorylation of protein bands with molecular masses around 45 and 48 kDa (Fig. 2B) were found again. On the other hand, the area covering the proteins with molecular masses ranging from 54 to 60 kDa, which gave the strongest positive reaction, seemed not to be af-

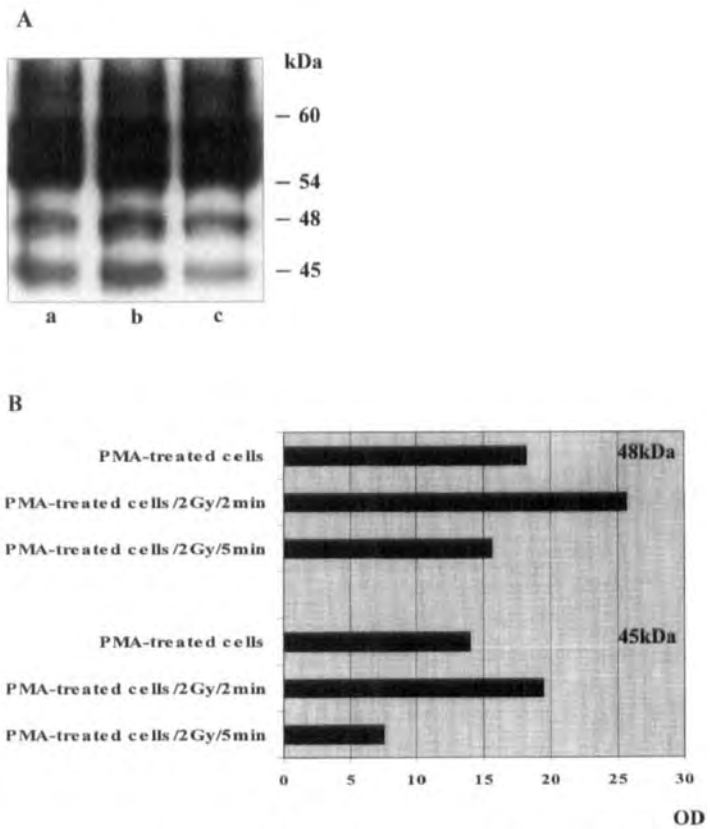


Figure 2. Quantitative changes of tyrosine-phosphorylated proteins in short time intervals after X-irradiation. (A) 1-D immunoblotting of anti-phosphotyrosine immunoprecipitates collected from (a) PMA-treated cells, or from X-irradiated PMA-treated cells incubated for an additional 2 (b) or 5 min (c). (B) Densitometry of tyrosine-phosphorylated proteins. The changes in OD of proteins of molecular masses of 48 kDa and 45 kDa 2 and 5 min after irradiation are shown.

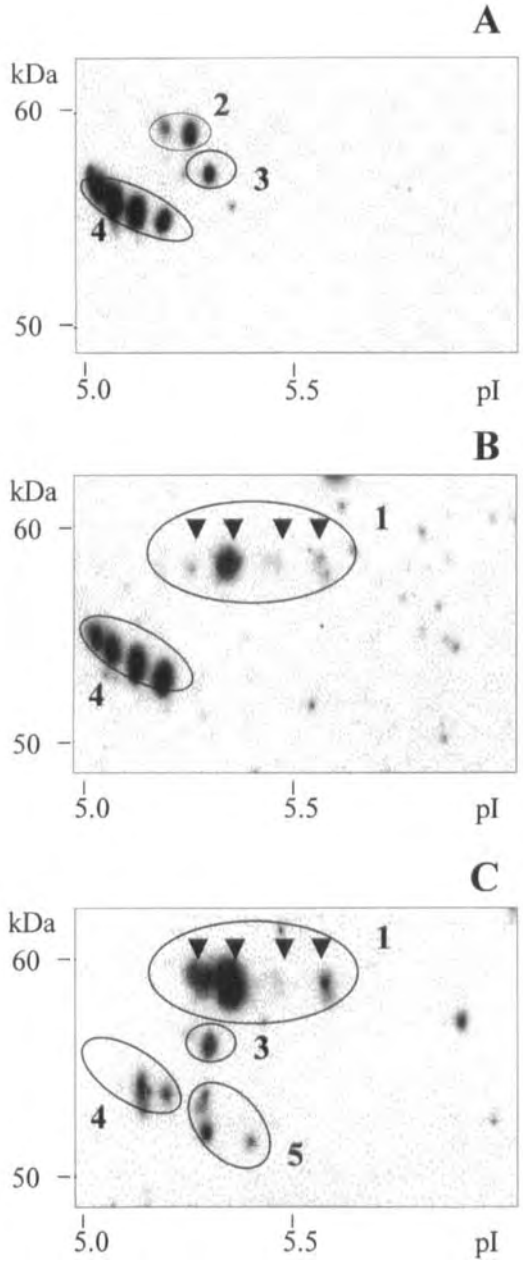


Figure 3. Detection of tyrosine-phosphorylated proteins. 2-DE immunoblotting of anti-phosphotyrosine immunoprecipitates collected from (A) PMA-treated cells, X-irradiated PMA-treated cells incubated for 2 (B) or 5 (C) min longer. Five regions numbered 1–5 are indicated.

fected by X-irradiation. However, when we performed 2-DE immunoblotting only proteins occurring in this intensively staining region were detected (Fig. 3) and, in contrast to 1-D separation, differentially phosphorylated protein patterns were found. The PVDF membranes were silver-stained after immunodetection so that the tyrosine-phosphorylated proteins on film could be matched with some of the silver-stained immunoprecipitated proteins and in this way their precise localization was determined.

Basically, the immunostained proteins can be divided into five groups. Group No. 1 (four spots, p*I* 5.25/molecular mass 60 kDa, 5.33/60 kDa, 5.46/60 kDa, 5.52/60 kDa) appeared 2 min after irradiation and the level of its tyrosine phosphorylation further increased 5 min after irradiation (Fig. 3B and C). Conversely, group No. 2 (two spots, 5.17/59 kDa, 5.25/59 kDa) occurred only in nonirradiated sample (Fig. 3A). Group No. 3 (1 spot, 5.28/57 kDa) was present in the nonirradiated sample 2 min after irradiation, was not immunostained, and appeared, again after a time interval of 5 min after irradiation. Finally, group No. 4 (four spots, 4.99/56.5 kDa, 5.04/56 kDa, 5.11/55 kDa, 5.17/54.5 kDa) was detected in nonirradiated sample; it persisted 2 min after irradiation but 5 min after irradiation two more acidic variants vanished and two more basic spots were only weakly detectable; on the contrary, the new phosphorylated spots occurring in group No. 5 (3 spots, 5.26/54 kDa, 5.28/53 kDa, 5.39/53 kDa) appeared (Fig. 3C).

We attempted to identify some of these tyrosine-phosphorylated proteins. We focused on ERKs that should be abundant in original HL-60 cells [9]. The 42 kDa ERK2 and 85 kDa ERK were immunostained but only in original RIPA lysates and not in anti-phosphotyrosine immunopre-

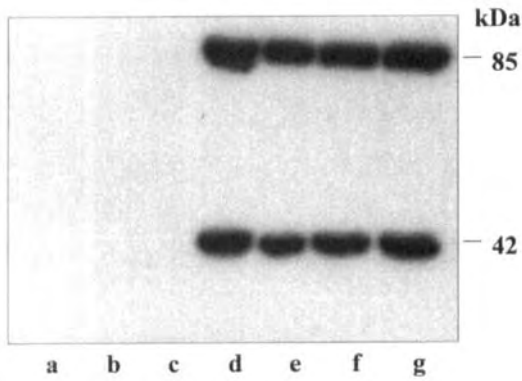


Figure 4. Immunodetection of ERK in anti-phosphotyrosine immunoprecipitates (a, b, c) or RIPA lysates (d, e, f, g). (a, e) PMA-treated cells; (b, f) PMA-treated cells, 2 min after irradiation; (c, g) PMA-treated cells, 5 min after irradiation and (d) RIPA lysate of nondifferentiated HL-60 cells.

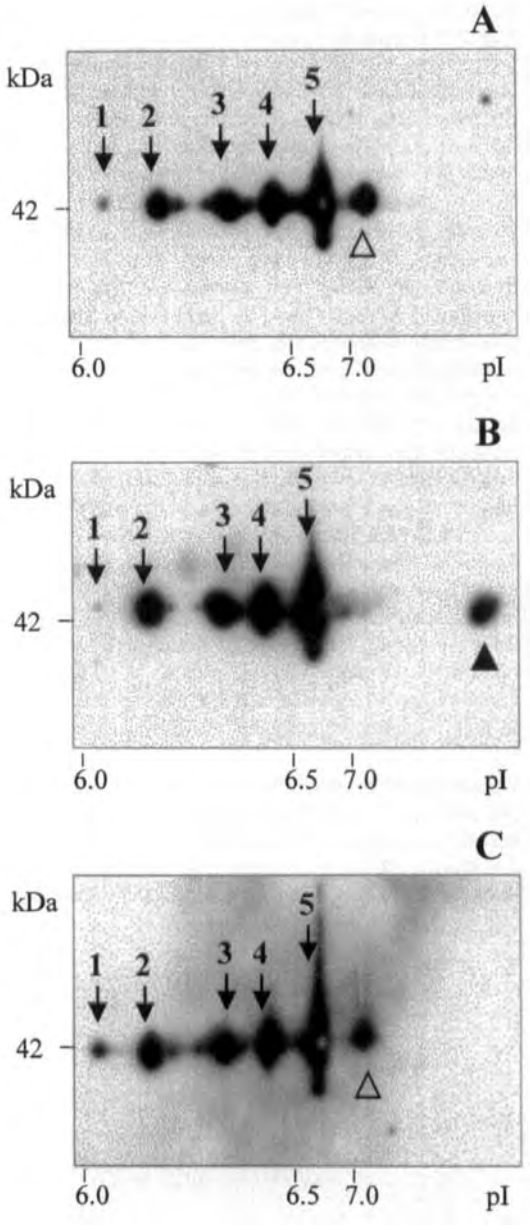
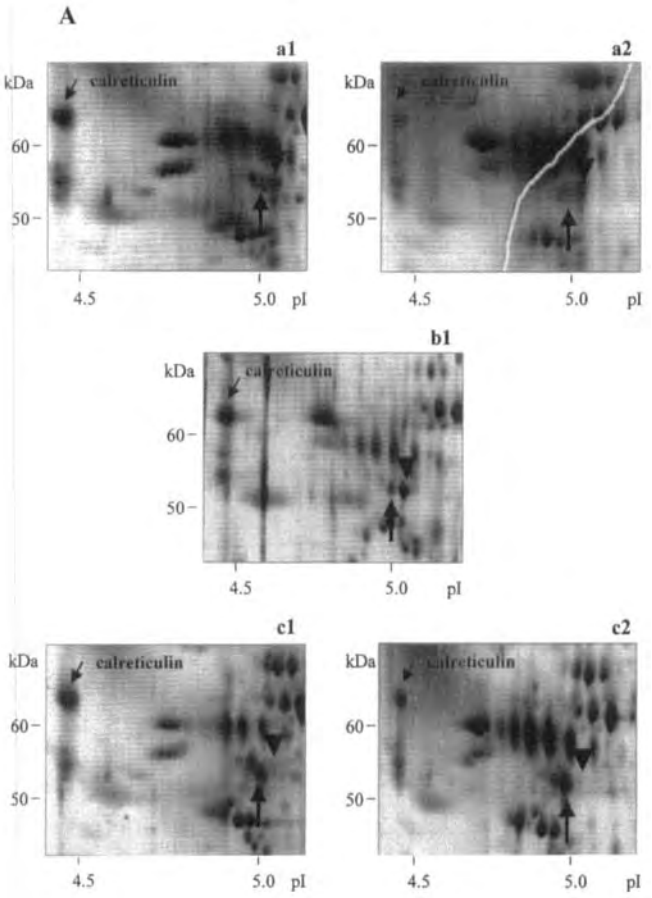


Figure 5. 2-DE immunoblotting of ERK2. The ERK2 kinase was detected using anti-ERK2 mAb and secondary horseradish peroxidase-labeled antibodies in (A) RIPA lysates of PMA-treated cells or in X-irradiated PMA-treated cells cultivated an additional 2 (B) or 5 (C) min. The positions of the main five isoforms of ERK2 (p*I* 6.05, 6.14, 6.26, 6.44 and 6.74/42 kDa) are denoted by arrows and Arabic numbers, the position of isoform detected only in time intervals A and C (7.08/42 kDa) is designated by an empty triangle and the position of isoform visualized only in time interval B (p*I* 7.84/42 kDa) is indicated by a triangle.

cipitates (Fig. 4). Since the activation of ERKs requires their phosphorylation on both tyrosine and threonine residues, our finding stands in contrast with the published data describing the maintenance of ERK2 activity during PMA-driven differentiation of HL-60 cells [10]. Furthermore, the phosphorylation of both regulatory residues should cause shifts in the mobility of the ERK on SDS-PAGE [11]. We detected only one ERK2 band on a 1-D immunoblot using monoclonal anti-ERK2 antibody (unpublished observation). When we examined the ERK2 protein pattern by 2-DE immunoblotting (Fig. 5) we also obtained only one row of spots differing in p*I* but with identical molecular mass. The existence of several ERK2 charge isoforms on 2-DE in addition to the possible single phosphorylation on threonine residue, is evidence for other, so far unknown, types of post-translational modification of ERK2. Some of them could be controlled by X-irradiation with respect to the shift of the most basic variant 2 min after irradiation (Fig. 5B).

We also looked for more global protein alterations induced by X-irradiation. The whole protein profile of nonirradiated PMA-treated cells was compared with X-irradiated PMA-treated cells cultivated for an additional 3 h by means of 2-DE. Two independently prepared cell lysates for each time interval including nonirradiated PMA-treated cells cultivated for additional 3 h were analyzed. A subset of about 1000 spots was matched among individual gels and statistically significant differences in protein levels between irradiated and nonirradiated samples were sought. We found two spots, probably isoforms, whose expression is regulated in an opposite manner in irradiated and nonirradiated samples (Fig. 6A). The more acidic isoform (5.05/53.5 kDa) is expressed at higher levels after irradiation and, conversely, the more basic variant (5.08/53.5 kDa) is supressed by irradiation. The quantities of both spots in terms of relative volumes (% vol) exhibit minimal variance in duplicates of analyzed samples and show increases or decreases by a factor of 2 or more between irradiated and nonirradiated samples (Fig. 6B). Both changes are statistically significant ($P < 0.05$). In conclusion, we have demonstrated the radiation-induced modulation of both signal transduction and protein expression in macrophage-like cells. The protein responsible for these alterations have not been identified so far, but this should follow in order to verify their possible functions within the cell which can aid in its survival.



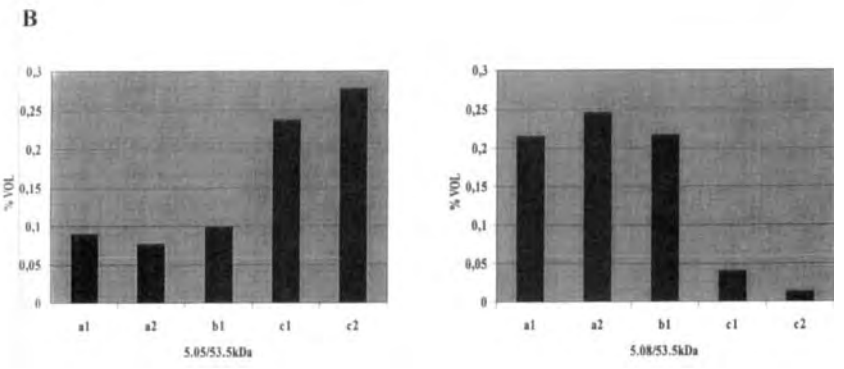


Figure 6. Protein alterations induced by X-irradiation. (A) The magnified regions of silver-stained 2-DE protein patterns of nonirradiated and irradiated PMA-treated cells. Tested time intervals are: (a1, a2) two independent samples of PMA-treated cells; (b1) PMA-treated cells, cultivated for an additional 3 h and (c1, c2) two independent samples of PMA-treated cells, X-irradiated and cultivated for an additional 3 h. Arrow indicates protein whose synthesis is induced by X-irradiation (5.05/53.5 kDa) and arrowhead designates protein whose expression is diminished after X-irradiation (5.08/53.5 kDa). (B) The significantly different abundances of 5.05/53.5 kDa and 5.08/53.5 kDa proteins in nonirradiated and irradiated PMA-treated cells expressed in the values of % vol ($P < 0.05$).

This work was supported by Linkage Grant from NATO (HTECH LG 960839) and by a grant from the Czech Ministry of Defense (MO66020397202). We thank J. Michaličková, A. Firychová and J. Žáková for excellent technical assistance.

Received September 1, 1998

4 References

[1] Schieven, G. L., Ledbetter, J. A., *Trends Endocrinol. Metab.* 1994, *5*, 383–388.

[2] Hallahan, D. E., *Seminars Radiat. Oncol.* 1996, *66*, 250–267.

[3] Collins, S. J., *J. Am. Soc. Hematol.* 1987, *70*, 1233–1244.

[4] Hill, H. D., Straka, J. G., *Anal. Biochem.* 1988, *170*, 203–208.

[5] Stulík, J., Bureš, J., Jandík, P., Langr, F., Kovářová, H., Macela, A., *Electrophoresis* 1997, *18*, 625–628.

[6] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[7] Laemmli, U. K., Favre, M., *J. Mol. Biol.* 1979, *80*, 575–599.

[8] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[9] Katagiri, K., Katagiri, T., Koyama, Y., Morikawa, M., Yamamoto, T., Yoshida, T., *J. Immunol.* 1991, *146*, 701–707.

[10] Meighan-Mantha, R. L., Wellstein, A.,. Riegel, A. T., *Exp. Cell. Res.* 1997, *234*, 321–328.

[11] Seger, R., Krebs, E. G., *FASEB J.* 1995, *9*, 726–735.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 969–970

Alison Pearce[1]
Clive N. Svendsen[2]

[1]ESA Analytical, St. Ives, Cambs, United Kingdom
[2]MRC Cambridge, Centre for Brain Repair, University Forvie Site, Cambridge, United Kingdom

Characterisation of stem cell expression using two-dimensional electrophoresis

Stem cells were isolated from foetal human brain tissue and induced to proliferate using the mitogens epidermal growth factor (EFG) and fibroblast growth factor (FGF). Under these conditions the dividing cells formed spheres which could be maintained in an undifferentiated state for extended periods of time. Following removal of the mitogen and addition of serum, the human stem cells rapidly began to differentiate. Samples from differentiated and undifferentiated cultures were lysed and analysed using two-dimensional (2-D) electrophoresis, a powerful technique for the separation and characterisation of proteins in complex mixtures. After 1 h post-differentiation, marked differences in protein expression could be observed between undifferentiated and differentiated stem cell samples.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Stem cells / Protein expression
EL 3351

The molecular mechanisms underlying the development of the mammalian central nervous system (CNS) are not as yet fully understood, due to the complexity of the mammalian brain. It is thought that dividing mammalian precursor cells are 'multi-potent', and give rise to specific types of neurones and glia at precise stages of development [1]. However, the developmental stages between a precursor cell and mature neurone are difficult to observe *in vivo*, and attempts to study these stages *in vitro* are hindered by the complex mix of neuronal cells present at various stages of development in primary embryonic CNS cultures. An alternative system has been described in which the various stages of CNS development can be studied under controlled culture conditions using embryonic stem cells [2, 3]. These stem cells represent a homogeneous source of starting tissue that can be followed through various stages of differentiation *in vitro*. However, as embryonic stem cells are totipotent, they can give rise to muscle, skin and bone in addition to neural tissue. Neural stem cells have also been isolated in culture and provide a convenient source of expandable tissue for studying the molecular mechanisms underlying neural development [3, 4].

Stem cells were isolated from human fetal brain (cortex) tissue and induced to proliferate using the two mitogens epidermal growth factor (EGF) and fibroblast growth factor (FGF). Under these conditions the dividing cells formed spheres which could be maintained in an undifferentiated state for extended periods of time [4]. Following removal of the mitogen and addition of serum, the human stem cells rapidly began to differentiate. We aimed to examine the differences between differentiated and undifferentiated stem cell cultures using two-dimensional (2-D) electrophoresis, and to determine the presence of novel stem cell proteins. 2-D electrophoresis is a powerful technique for the separation and characterisation of proteins in complex mixtures and can typically resolve 2000–3000 different protein spots [5, 6].

Stem cells were cultured in a serum-free growth medium containing DMEM/Hams (3/1), N2 supplement (Gibco, Grand Island, NY), and EGF/FGF-2, both at a concentration of 20 ng/mL. When the cells had reached a density of 5×10^5 cells/mL, 10 mL of the culture medium were removed and a cell lysate was prepared. A further 10 mL aliquot was removed and induced to differentiate by the addition of 100 μL serum to give a final serum concentration of 1%. The cells were cultured in the presence of serum for 1 h and a second lysate was prepared. High-resolution two-dimensional gel electrophoresis was performed according to the method of O'Farrell [7] with modifications as described by Garrels [8], using the ESA Investigator 2-D Electrophoresis System (St. Ives, UK) [8]. IEF was performed in precast carrier ampholyte tube gels (pH 3–10) and the second-dimensional electrophoretic separation on precast 10% polyacrylamide slab gels. Following electrophoresis, separated proteins were visualised by silver staining according to the method of Morrisey with modifications [9].

Correspondence: Dr. Alison Pearce, Genomic Solutions, Forge Close, Little End Road, Eaton Socon, St. Neots, Cambridgeshire PE19 3TP, United Kingdom
E-mail: alisonp@genomicsoutions.co.uk
Fax: +44-1480-471660

Abbreviations: CNS, central nervous system; **EGF**, epidermal growth factor; **FGF**, fibroblast growth factor; **MAP**, microtubule-associated protein

As expected, the expression of many stem proteins was induced after differentiation, and there were marked differences between the two protein expression patterns after only 1 h in the presence of serum. Several proteins appear to be native only to the undifferentiated stem cell culture. These may be novel stem cell proteins, or proteins that are expressed early in the development of stem cells, *e.g.*, nestin, an intermediate neurofilament expressed in precursor cells early in development, or microtubule-associated protein 2 (MAP-2c) [10]. Current work involves characterising the proteins native to the undifferentiated culture, and to examine their potential use as novel stem cell markers. The protein will be detected by Western blotting, and their identities confirmed by mass spectrometry. In addition to characterising proteins native to the undifferentiated culture, prior to serum addition, we hope to examine those proteins whose expression appears to be induced following serum addition. In order to exclude the possibility that these proteins arise as a result of the protein content in the serum aliquot, serum control gels will be run.

Received September 1, 1998

References

[1] McKay, R. D. G., *Cell*, 1989, *58*, 815–821.

[2] Svendsen, C. N., Rosser, A. E., *TINS* 1995, *8*, 465–467.

[3] Bain, G., Kitchens, D., Yao, M., Huettner, J. E., Gottlieb, D. I., *Dev. Biol.* 1995, *168*, 342–357.

[4] Svendsen, C. N., Caldwell, M. A., Shen, J., ter Borg, M. G., Rosser, A. E., Tyers, P., Karmiol, S., Dunnett, S. B., *Exp. Neur.* 1997, *148*, 135–146.

[5] Patton, W. F., Pluskal, M. G., Skea, W. M., Buecker, J. L., Lopez, M. F., Zimmermann, R., Belanger, L. M., Match, P. D., *BioTechniques* 1990, *8*, 518–527.

[6] Lopez, M. F., Barry, P., Sawlivich, W., Hines, T., Skea, W. M., *Appl. Theor. Electrophor.* 1994, *4*, 95–102.

[7] O'Farrell, P., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[8] *Investigator 2-D Electrophoresis System: User guide*, ESA Analytical Ltd, St. Ives, Cambs, UK 1990.

[9] Morrissey, J. H., *Anal. Biochem.* 1981, *117*, 307–310.

[10] Rosser, A. E., Tyers, P., ter Borg, M. G., Dunnet, S. B., Svendsen, C. N., *Dev. Brain Res.* 1997, *98*, 291–295.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 971–976

Claudia Bohring
Walter Krause

Clinic of Dermatology, Department of Andrology, Clinical Training Center of the European Academy of Andrology, Philipp University, Marburg, Germany

The characterization of human spermatozoa membrane proteins – surface antigens and immunological infertility

The aim of this study was to isolate and characterize highly enriched membrane proteins by two-dimensional (2-D) electrophoresis and to identify surface antigens binding sperm autoantibodies (SpAb). The presence of SpAb may reduce fertility by affecting sperm motility and acrosome reaction. The presence of the SpAb was shown to prevent sperm penetration of cervical mucus, to inhibit sperm-zona pellucida interaction, and to interfere with the sperm-egg fusion. The swim-up method was used to separate mature and motile sperm. Sperm membranes were obtained by hypoosmotic swelling, homogenization and sonication. Membranes were further isolated by differential centrifugation steps. The highly purified human sperm membrane proteins were separated by two-dimensional gel electrophoresis and electrotransferred to polyvinylidene difluoride (PVDF) membrane. The antigens were identified by bound SpAb, the sources of which were seminal plasma samples of infertile patients or of patients following vasectomy. Fourteen surface antigens were detected. Their identification may be (i) important for understanding the mechanism by which SpAb impair sperm fertilization capacity, (ii) suitable as a basis of new methods of fertility regulation, and (iii) helpful in developing reproducible and reliable methods for determinations of SpAb.

Keywords: Autoantibodies / Immunological infertility / Membrane proteins / Spermatozoa / Two-dimensional polyacrylamide gel electrophoresis EL 3394

1 Introduction

The characterization and identification of human sperm antigens are important for understanding the mechanism by which sperm autoantibodies (SpAb) may impair sperm fertilization capacity. In particular the interaction between antibodies and sperm surface antigens may be a mechanism causing immunological infertility. Knowledge of these processes is also important as a basis for new methods of fertility regulation. In addition, the development of reproducible and reliable methods for SpAb determinations requires the identification of particular antigens. There is no universally accepted method for antisperm antibody detection [1]. The occurrence of spontaneous autoantibodies to sperm in infertile men was first reported by Rümke [2] and Wilson [3] in 1954, independently of each other. In later contributions, it was shown that the occurrence of SpAb decreases fertility [4–6] and fecundity [7, 8]. SpAb can be detected in the seminal plasma [9], and bound to the sperm surface [10], in the blood serum of men and women [11], in oviduct fluid [12], cervical mucus [13], and the follicular fluid of women [14].

To date, different effects of SpAb on sperm functions were described. SpAb affected the sperm motility [15], the acrosome reaction [16, 17], and the penetration of cervical mucus by spermatozoa [18, 19]. Also, an inhibition of spermatozoa and zona pellucida interaction [20] and an interference with the sperm egg fusion was described [21, 22]. However, no proper methods exist to decide whether the SpAb of a single individual are of functional relevance and only a limited number of sperm antigens may be associated with immunological infertility [23]. In this study, sperm surface proteins were identified which bound SpAb of patients who consulted our clinic for investigation. Focus was placed on antibodies in the seminal plasma, because SpAb in sera which occur in a large proportion of infertile and fertile men are not related to infertility [24]. In the present investigation, we used the method according to Hinsch *et al.* [25], for the isolation of highly enriched membrane proteins, with some modifications. The sperm antigens were separated by 2-D electrophoresis techniques and exposed to seminal plasma samples containing SpAb and control samples.

Correspondence: Dr. Claudia Bohring, Clinic of Dermatology, Department of Andrology, Clinical Training Center of the European Academy of Andrology, Philipp University, Deutschhausstraße 9, D-35033 Marburg, Germany
E-mail: bohring@mailer.uni-marburg.de
Fax: +49-6421-282883

Abbreviations: IVF, *in vitro* fertilization; **MAR**, mixed antiglobulin reaction; **SpAb**, sperm autoantibody

2 Material and methods

2.1 Reagents

The IPG strips and the equipment for running the IPG gels (Multiphor II and DryStrip kit) as well as Pharmalyte (UpH 3–10), Ampholine (pH 3.5–10), CHAPS, DTT, SDS, Tween-20 and urea were from Pharmacia (Uppsala, Sweden). The 30% acrylamide/Bis solution, 35.5:1, 2.6%C and the 2-D SDS-PAGE standards as well as the prestained SDS-PAGE standards and the SDS-PAGE low range standards were from Bio-Rad (München, Germany). All other chemicals were analytical grade from Merck (Darmstadt, Germany). For the swim-up technique the IVF (*in vitro* fertilization) buffer was used (Medi Cult a/s, Copenhagen, Denmark). The indirect and the direct MAR (mixed antiglobulin reaction) tests were performed with SpermMAR IgG test from Ferti ProN.V (Sint – Martens – Latem, Belgium). The ECL (enhanced chemiluminescence) kit was from Amersham (Uppsala, Sweden).

2.2 Patients

Semen samples from 20 patients who were referred to our clinic for infertility or for control of semen following vasectomy investigations were obtained by masturbation after 3–6 days abstinence. The ejaculate was centrifuged for 15 min at 3000 × *g* at room temperature. The seminal plasma was then stored at –20°C until assayed. The unbound SpAb in seminal plasma were determined by means of indirect MAR test.

2.3 Donor sperm preparation

Semen samples from donors with normal sperm parameter [26] were obtained by masturbation after 3–6 days abstinence. Spermatozoa were then enriched by means of a swim-up preparation [27–29]. One mL of semen was diluted 1:5 with IVF buffer, followed by centrifugation at 300 × *g* at room temperature for 10 min. The pellet containing spermatozoa was resuspended in 5 mL IVF buffer and washed again. Two mL of IVF medium were then carefully placed above the final pellet, after which the tube was inclined to an angle of 45° and incubated for 1 h at 37°C. The supernatant containing the enriched fraction of motile spermatozoa was aspirated and centrifuged again for 10 min at 500 × *g* at room temperature. Finally, the cell concentration was adjusted to 20 × 10^6 mio/mL with IVF medium for the indirect MAR test. For the membrane preparation the final pellet was prepared as described in the following.

2.4 SpermMAR test

For the detection of sperm antibody the SpermMAR IgG test from Ferti ProN.V. was used. The direct SpermMAR test was performed according to the technical information provided by the manufacturer. The antibodies solubilized in the patients' seminal plasma samples were tested using the indirect MAR test [30]. In this case the soluble sperm antibodies in patients' seminal plasma are detected by using healthy donor spermatozoa as the antigen. The SpAb bound to the donor spermatozoa, which will react positively in a subsequent MAR test. For the MAR test, IgG-coated latex particles are attached; afterwards a monospecific antihuman IgG antiserum is added to the SpAb on the sperm surface. One hundred motile spermatozoa are evaluated and the percentage of spermatozoa with attached latex particles represents the test results in percent. The test is considered positive (by WHO standards) when > 10% of the motile sperms are agglutinated with latex particles. Therefore we used only seminal plasma samples with an indirect MAR test result higher than 15%.

2.5 Preparation of membrane proteins

The final pellet of the donor sperm preparation as described above was diluted 1:9 with hypoosmotic medium modified according to Jeyendran *et al.* [31]. The spermatozoa were swollen in the hypoosmotic medium for 2 h in a water bath (37°C), after which the solution was transferred to ice. First the membranes were stripped with a homogenizer (Ultraturrax; 1200 U/min; 2 min, 4°C). The suspension was sonicated (sonication bath) and centrifuged at 4000 × *g* (4°C, 15 min) to remove the cell debris and unbroken sperm. The supernatant was centrifuged at 10 000 × *g* (4°C, 10 min) and the resultant supernatant underwent a further ultracentrifugation at 100 000 × *g* (4°C, 2 h). Aliquots of each purification step were collected. The final pellet containing the membrane proteins was resuspended and dissolved in IEF buffer consisting of 9 M urea, 20 mM DTT, 2% CHAPS, 0.4% Pharmalyte and 0.4% Ampholine for 2-D electrophoresis. For the one-dimensional electrophoresis the proteins were dissolved in PBS buffer (pH 7.4). Good solution was obtained with sonication in a bath. The protein concentration of each step was determined using a BCA protein assay (Pierce, Rockford, IL).

2.6 Separation of membrane proteins

2.6.1 One-dimensional electrophoresis

All reagents for electrophoresis were of the highest purity obtainable. Aliquots of each purification step were collected and separated on a 4–20% SDS mini gel. Each slot was loaded with a volume of sample containing 50 µg protein. One-dimensional polyacrylamide gel electrophoresis was performed according to Laemmli [32] in a 4–20% gra-

dient mini gel using the prestained low range standards from Sigma (Deisenhofen, Germany) and the SDS-PAGE low range standards (Bio-Rad). The proteins were stained with Coomassie blue.

2.6.2 Two-dimensional electrophoresis and Western blot

The highly enriched human sperm membrane proteins were separated by 2-D PAGE [33, 34]. Isoelectric focusing, as a first dimension, was performed on precast Dry-Strip IPG 4–7 L with Multiphor II (Pharmacia). Seventy-five μg membrane proteins solved in the above-described IEF buffer with a final volume of 75 μL was loaded in each sample cup and one strip of six strips was loaded with 10 μL 2-D SDS-PAGE standards (Bio-Rad). The IEF step was performed at 15°C at a final voltage of 3500 V and a total of 150 kVh. The equilibration was performed as described [35].

The second dimension was carried out on 12% SDS polyacrylamide gels (10 × 13 × 0.1 cm) with a maximum current of 20 mA and maximum voltage of 150 V. For the molecular mass determination, prestained SDS-PAGE standards and the SDS-PAGE low range standards (Bio-Rad) were used. One IPG gel with 2-D SDS-PAGE standards was run parallel to five SDS gels. The membrane protein patterns were visualized by silver staining [36]. Separated membrane proteins were electrotransferred semidry onto Immobilon-P PVDF sheets (0.8 mA/cm^2; Millipore, Eschborn, Germany), using a semidry electrophoretic apparatus (Pharmacia) according to the procedure described by Millipore [37]. The blots were exposed to seminal plasma samples from patients diluted 1:20 and 1:50, according to the procedure of Lee *et al.* [38] for specific probing. The secondary antibody, conjugated with peroxidase (AffiniPure Goat Anti-Human IgG, IgM, IgA (H + L); dianova, Hamburg, Germany) was used at a 1:10 000 dilution. The ECL kit (Amersham) was used to identify the immunoreactive antigenic proteins. The ECL procedure was performed according to the manufacturer's instructions.

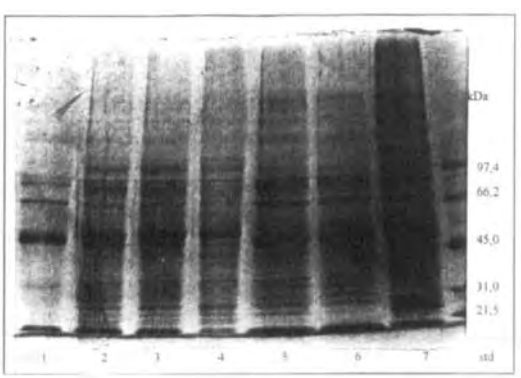


Figure 1. One-dimensional SDS-PAGE of sperm proteins on a 4–20% gradient mini gel; Coomassie staining; each slot was loaded with 50 μg protein. Lane (1) whole sperm protein; (2) whole sperm protein after the treatment with hypoosmotic medium; (3) pellet after 3000 × *g*; (4) supernatant after 10 000 × *g*; (5) pellet after 10 000 × *g*; (6) supernatant after 100 000 × *g*; (7) membrane proteins, pellet after 100 000 × *g*.

2.7 Gel scanning

Stained gels were scanned in a wet state using a ScanJet 6100CIT and the HP DeskScan II software (Hewlett-Packard, Greeley, CO, USA). The figures were drawn using Corel PHOTO PAINT 6.

3 Results

3.1 Membrane preparation and one-dimensional electrophoresis

Figure 1 illustrates the different protein purification steps after separation using one-dimensional SDS polyacrylamide gradient gel and visualization with Coomassie brilliant blue. The expression patterns of proteins from the whole spermatozoa extract (lane 1) differed clearly from the enriched membrane protein extract (lane 7). The good quality of the membranes was monitored by electron microscopy [24].

3.2 2-D gel electrophoresis and antigen detection

Two-dimensional electrophoresis offers resolution and separation of the proteins according to isoelectric points (p*I*) as well as molecular weights. Figure 2 shows a 2-D silver staining pattern of enriched sperm membrane proteins. The proteins had p*I* between 4 and 7 and an apparent molecular mass ranging between 10 and 110 kDa. The immunodetection of the sperm surface antigens by SpAb were compared with a 2-D-separated membrane pattern visualized with silver. All the recognized antigens were marked with circles in silver-stained gels, matched with immunoreactive spots in Western blots with patients' SpAb (Fig. 2). To date, 14 antigens were detected with seminal plasma antibodies in the 2-D Western blot. Some of the antigens were shared by many patients, while other antigens were detected only in individual cases. All seminal plasma samples containing SpAb showed immunoreactivity to at least three of the 14 surface antigens but no patient SpAb reacted with all of them. Immunoblots of two patients are exemplarily reported in Fig. 3. Table 1

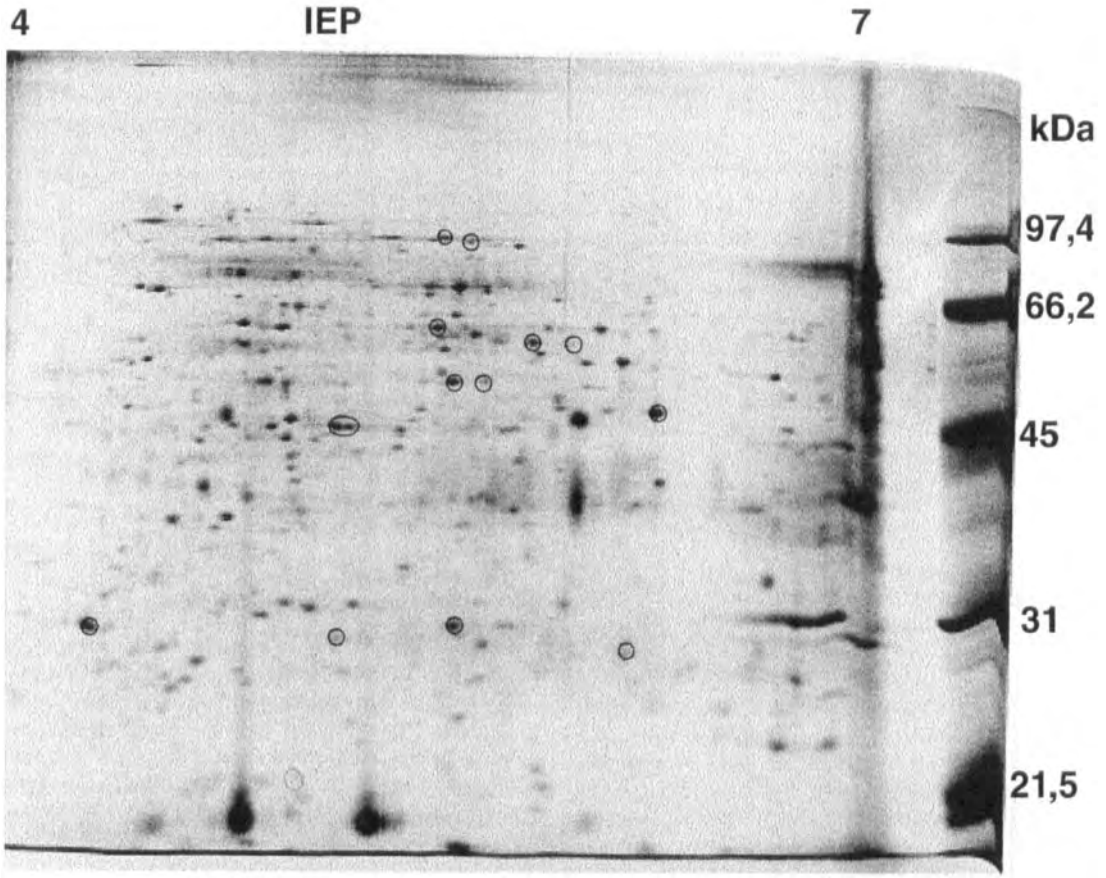


Figure 2. 2-D electrophoresis of membrane proteins from human spermatozoa. IEF with IPG (linear 4–7 pH gradient) was performed with 75 µg protein sample for the first-dimensional separation. The second dimension was carried out on a homogeneous running gel (12%T) and 4% stacking gel. Detection was done by silver staining. The circles indicate spots in silver-stained gel matching with immunoreactive spots in Western blots with patients' SpAb.

summarizes the distribution of positive reactions obtained by 2-D immunoblotting between separated sperm membrane antigens of different M_r/p*I* and antisperm antibodies present in seminal plasma from patients. The frequency of sperm membrane antigens binding with the sperm antibodies in the 20 patients by Western blot analysis is displayed in the last column. All patients (100%) had SpAb to the antigen with a molecular mass of 75.2 kDa and a p*I* of 5.4. Nearly all patients had SpAb against antigen 7 (95%), antigen 11 (90%) and antigen 12 (95%). The control samples showed no reaction with any of the membrane proteins of human sperm.

4 Discussion

In this study, membrane extracts of donor sperm were used as the antigen source for immunoblotting of soluble sperm autoantibodies of seminal plasma samples. The purpose of this study was to generate a map of human sperm membrane antigens immunologically detected by SpAb. To identify the antigens concerned, it is important to check the purity of the sperm membrane preparation. Sperm extracts obtained by previous investigators using detergent extraction were probably contaminated with acrosomal, cytoplasmatic and nuclear sperm proteins. Xu *et al.* [39] indicated that more than 85% of the proteins in the whole sperm extract originated from subcellular organelles and cytoplasm instead of the membrane. Also, the antigens detected from whole sperm preparations do not provide information on the antigens involved in immunological infertility processes, since nearly all men have SpAb against intracellular antigens, such as nuclear proteins [40]. Only the surface antigens may induce immunological infertility, therefore only the evidence of autoanti-

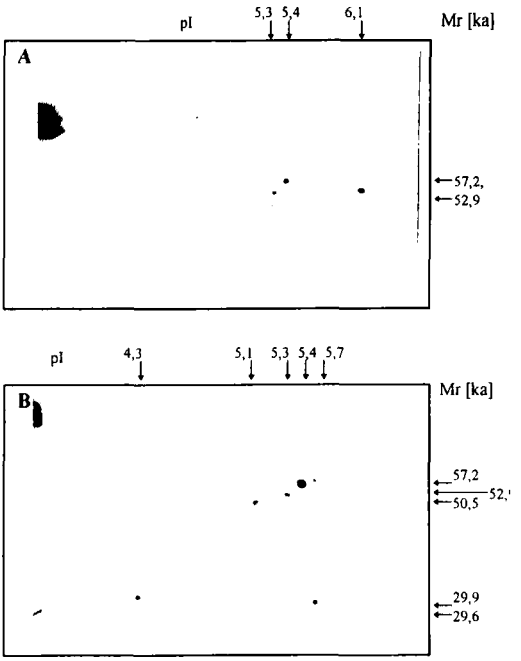


Figure 3. Two-dimensional Western blots of sperm membrane proteins against SpAb obtained by seminal plasma from two patients. (A) Patient 1, binding against antigens ag6, ag8, ag9; (B) patient 2 binding against antigens ag6, ag7, ag8, ag10, ag11, ag12.

Table 1. Molecular masses and isoelectric focusing points of membrane antigens binding with sperm antibodies in 20 patients by Western blot analysis

Antigen	M_r (kDa)	p/	No. of patient (%)
ag1	85.2	5.4	2(10)
ag2	84.1	5.50	3(15)
ag3	65.9	5.30	6(30)
ag4	64.0	5.70	2(25
ag5	63.3	5.80	17(85)
ag6	**57.2**	**5.40**	**20(100)**
ag7	**57.2**	**5.55**	**19(95)**
ag8	52.9	5.30	8(40)
ag9	52.9	6.10	4(20)
ag10	50.5	5.10	14(70)
ag11	**29.9**	**4.30**	**18(90)**
ag12	**29.6**	**5.40**	**19(95)**
ag13	28.9	5.05	16(80)
ag14	27.9	6.00	6(30)

Antigens that were reactive against 90% or more of seminal plasma samples are highlighted.

bodies binding to sperm surface antigens is an indicator of this disease. If we consider that most of the monoclonal antibodies produced against sperm do not react with sperm surface proteins, they are of no use in the detection of antigens involved in sperm functions. To improve the reproducibility of sperm membrane protein profiles we used the swim-up preparation technique for separation of motile mature spermatozoa, by which dead sperms and nonsperm cells are removed. The isolation of highly enriched human sperm membranes and the subsequent separation with 2-D electrophoresis made it possible to identify and characterize the membrane antigens. Two-dimensional electrophoresis employing immobilized pH gradient (IPG) gels produced well-reproducible sperm protein pattern.

The characterization of sperm surface molecules would be useful not only in understanding the complex processes of differentiation and maturation during spermatogenesis, but also for the role of membrane proteins in different sperm functions [41]. Further studies should determine which of these 14 antigens are related to sperm functions and which are important in the process of fertilization. Previous investigations supported the association of only a part of the antigens associated with immunological infertility [42]. Only the "fertility-related" antigens will be of value in diagnosing immunological infertility. This will also provide a basis for developing immunological methods of contraception. Our further investigations will focus on the relevance of certain surface antigens for sperm functions such as acrosome reaction, sperm motility and zona binding and on the amino acid sequencing of the antigens detected, especially the antigens ag6, ag7, ag11, and ag12.

Received November 7, 1998

5 References

[1] Dondero, F., Gandini, L., Lombardo, F., Salacone, P., Caponecchia, L., Lenzi, A., *Am. J. Reprod. Immunol.* 1997, *38*, 218–223.

[2] Rümke, P., *Vox. Sang.* 1954, *4*, 135–140.

[3] Wilson, L., *Proc. Soc. Exp. Biol. Med.* 1954, *85*, 652–654.

[4] Ansbacher, R., Keung-Yeung, K., Behrman, S. J., *Fertil. Steril.* 1973, *24*, 305–308.

[5] Menge, . C., Medley, N. E., Mangione, C. M., Dietrich, J. W., *Fertil. Steril.* 1982, *38*, 439–446.

[6] Acosta, A., van der Merve, J., Doncel, G., Kruger, T., Sayilgan, A., Franken, D., Kolm, P., *Fertil. Steril.* 1994, *62*, 826–833.

[7] Ingerslev, H. J., Ingerslev, M., *Fertil. Steril.* 1980, *33*, 514–520.

[8] Baker, H. W., Clarke, G. N., Hudson, B., McBain, J. C., McGowan, M. P., Pepperell, R. J., *Clin. Reprod. Fertil.K 1983, 2, 55–71.*

[9] Clarke, G. N., Stojanoff, A., Cauchie, M. N., Johnston, W. I. H., *Am. J. Reprod. Immunol. Microbiol.* 1985, *7*, 143–147.

[10] Anderson, D. J., Hill, L. A., *Am. J. Reprod. Immunol. Microbiol.* 1988, *17*, 22–30.

[11] Clarke, G. N., Elliott, P. J., Smaila, C., *Am. J. Reprod. Immunol.* 1985, *7*, 118–123.

[12] Wong, W. P., *Fertil. Steril.* 1979, *32*, 681–684.

[13] Clarke, G. N., *Am. J. Reprod. Immunol.* 198, *6*, 195–197.

[14] Kay, D. J., Boettcher, B., Yovich, J. L., Stanger, J. D., *Am. J. Reprod. Immunol. Microbiol.* 1985, *7*, 1133–1137.

[15] Barratt, C. R. L., Havelock, L. M., Harrison, I. E., Cooke, I. D., *Int. J. Androl.* 1989, *12*, 110–116.

[16] Tung, K. S. K., Okada, A., Yanagimachi, R., *Biol. Reprod.* 1980, *23*, 877–886.

[17] Bandoh, R., Yamano, S., Kamada, M., Daitoh, T., Aono, T., *Fertil. Steril.* 1992, *57*, 387–392.

[18] Jager, S., Kremer, J., Van Slochteren-Draaisma, T., *Int. J. Androl.* 1979, *2*, 117–130.

[19] Eggert-Kruse, W., Böckem-Hellwig, S., Doll, A., Rohr, G., Tilgen, W., Runnebaum, B., *Hum. Reprod.* 1993, *8*, 1025–1031.

[20] Mahony, M. C., Alexander, N. J., *Hum. Reprod.* 1991, *6*, 1426–1430.

[21] Clarke, G. N., Lopata, A., McBain, J. C., Baker, H. W. G., Johnston, W. I. H., *Am. J. Reprod. Immunol.* 1985, *8*, 62–66.

[22] Bronson, R. A., Cooper, G. W., Phillips, D. M., *Hum. Reprod.* 1989, *4*, 653–657.

[23] Snow, K., Ball, G. D., *Fertil. Steril.* 1992, *58*, 1011–1019.

[24] Shai, S., Naot, Y., *Fertil. Steril.* 1992, *58*, 593–598.

[25] Hinsch, K.-D., Hinsch, E., Aumüller, G., Tychowiecka, I., Schultz, G., Schill, W.-B., *Biol. Reprod.* 1992, *47*, 337–346.

[26] World Health Organization, *WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction*, Cambridge University Press 1992.

[27] Lopata, A., Patullo, M. J., Chang, A., James, B., *Fertil. Steril.* 1976, *72*, 677–684.

[28] Harris, S. J., Milligan, M. B., Masson, G. M., Dennis, K. J., *Fertil. Steril.* 1981, *36*, 219–221.

[29] Makler, A., Murillo, O., Huszar, G., Tarlatzis, B., DeCherney, A., Naftolin, F., *Int. J. Androl.* 1984, *7*, 61–70.

[30] Hinting, A., Vermeulen, L., Comhaire, F., *Fertil. Steril.* 1988, *49*, 1039–1044.

[31] Jeyendran, R. S., Van der Ven, H. H., Perez-Pelaez, M., Crabo, B. G., Zaneveld, L. J. D., *Reprod. Fertil.* 1984, *70*, 219–228.

[32] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[33] Görg, A., Postel, W., Weser, J., Günther, S., Strehler, J. R., Hanash, S. M., Somerlot, L., *Electrophoresis* 1987, *8*, 122–124.

[34] Righetti, P. G., Gianazza, E., Gelfi, C., Chiari, M., *Isoelectric Focusing: In-Gel Electrophoresis of Proteins*, Oxford University Press, New York 1990.

[35] Rabilloud, T., Vuillard, L., Gilly, C., Lawrence, J. J., *Cell. Mol. Biol.* 1994, *40*, 57–74.

[36] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[37] Millipore, *Immobilon-P Transfer Membrane User Guide*, Bedford, MA 1995.

[38] Lee, S.-L., Stevens, J., Wang, W. W., Lanzillo, J. J., *BioTechniques* 1994, *17*, 60–64.

[39] Xu, C., Rigney, D. R., Anderson, D. J., *J. Androl.* 1994, *15*, 595–602.

[40] Rosseaux-Prevost, R., DeAlmeida, M., Jouannet, P., Hublaus, P., Sautiere, P., Rosseaux, J., *Mol. Immunol.* 1992, *29*, 895–902.

[41] Naaby-Hansen, S., Flickinger, C. J., Herr, J. C., *Biol. Reprod.* 1997, *56*, 771–787.

[42] Bohring, C., Krause, W., *Int. J. Androl.* 1997, *Suppl. 20*, 23.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 977–983

Allan Christian Shaw[1,2]
Martin Røssel Larsen[3]
Peter Roepstorff[3]
Arne Holm[4]
Gunna Christiansen[1]
Svend Birkelund[1]

[1]Department of Medical Microbiology and Immunology, University of Aarhus, Denmark
[2]Department of Molecular and Structural Biology, University of Aarhus, Denmark
[3]Department of Molecular Biology, University of Odense, Denmark
[4]The Royal Veterinary and Agricultural University, Copenhagen, Denmark

Mapping and identification of HeLa cell proteins separated by immobilized pH-gradient two-dimensional gel electrophoresis and construction of a two-dimensional polyacrylamide gel electrophoresis database

The HeLa cell line, a human adenocarcinoma, is used in many research fields, since it can be infected with a wide range of viruses and intracellular bacteria. Therefore, the mapping of HeLa cell proteins is useful for the investigation of parasite host cell interactions. Because of the recent improvements of two-dimensional gel electrophoresis with immobilized pH gradients (IPG) compared to isoelectric focusing with carrier ampholytes, a highly reproducible method for examining global changes in HeLa cell protein expression due to different stimuli is now available. Therefore, we have initiated the mapping of [^{35}S]methionine/cysteine-labeled HeLa cell proteins with the 2-D PAGE (IPG)-system, using matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS) and *N*-terminal sequencing for protein identification. To date 21 proteins have been identified and mapped. In order to make these and future data accessible for interlaboratory comparison, we constructed a 2-D PAGE database on the World Wide Web.

Keywords: HeLa / Two-dimensional gel electrophoresis / Immobilized pH gradient / World Wide Web / Make2ddb
EL 3369

1 Introduction

The application of two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) for the separation of complex samples, such as cell lysate, is the best available method for looking for changes in protein expression on a large scale. In our laboratory we have a long-term interest in investigating changes in protein expression during interaction between the obligate intracellular bacterium *Chlamydia trachomatis* and its host cell [1, 2]. For this purpose we have used the HeLa cell, a human adenocarcinoma cell line which is suitable as a host cell for infection with a wide range of viruses and intracellular bacteria *in vitro*. HeLa cell proteins have previously been mapped by means of 2-D PAGE with carrier ampholytes [3]. However, an increasing number of investigators are now using 2-D PAGE with immobilized pH gradient (IPG) due to the higher resolution and reproducibility of this system. The standardization of this technique is, in part, a result of the commercially available IPG drystrips [4]. Furthermore, sample loading has been considerably eased by the introduction of reswelling cassettes, allowing direct loading of proteins onto the IPG strip [5]. In addition, the 2-D PAGE (IPG) system has a higher loading capacity, so that fewer preparative gel runs are required to obtain sufficient amounts of protein for identification. In combination with sensitive MALDI-TOF-MS techniques and because of the growing amount of proteins listed in databases, the 2-D PAGE (IPG) system is now a powerful tool in proteome analysis. The gel patterns from the two different 2-D PAGE systems are not easily matched, making comparison of old and new data questionable [6]. We have chosen to identify HeLa cell proteins by MALDI-MS or *N*-terminal sequencing instead of comparing gels to existing maps of HeLa cells or 2-D PAGE (IPG) of other human cell lines. In order to make these data available on the World Wide Web, we have constructed a parasite host cell interaction 2-D PAGE database (PHCI-2DPAGE) by installing the Make2ddb 1.1 software [7] from the SWISS-2DPAGE [8, 9] on a server (www.gram.au.dk).

Correspondence: Allan Christian Shaw, Department of Medical Microbiology and Immunology, The Bartholin Building, University of Aarhus, DK-8000 Aarhus C, Denmark
E-mail: shaw@medmicro.aau.dk
Fax: +45-8619-6128

Abbreviations: AC, accession number line; **ID,** identification line

2 Materials and methods

2.1 Cultivation and ^{35}S-labeling of HeLa cells

The HeLa 229 cell line was obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). The cells were cultivated in RPMI 1640 (Gibco BRL, Grand Island,

NY, USA) containing 10% fetal calf serum (FCS; Gibco BRL) and 10 μg/mL gentamicin at 37°C in a 5% CO_2 atmosphere. Semiconfluent monolayers were labeled for 6 h with 100 μCi/mL [^{35}S]methionine/cysteine (Promix, Amersham, England) in a methionine-free RPMI 1640 medium. The HeLa cells were subsequently washed in PBS and solubilized in a 9 M urea, 4% w/v 3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate (CHAPS), 40 mM Tris base, 65 mM dithioerythritol (DTE), 2% v/v Pharmalyte 3–10 (Pharmacia Biotech, Uppsala, Sweden) and sonicated. The amount of ^{35}S-incorporation was determined by TCA precipitation.

2.2 Two-dimensional gel electrophoresis

For isoelectric focusing, linear 18 cm immobilized pH 4–7 gradient drystrips (Pharmacia Biotech) were used. Using a reswelling tray (Pharmacia Biotech), strips were soaked overnight in 400 μL of the lysis solution, containing 3.5 × 10^5 cpm of labeled proteins. Rehydrated strips were run on a Multiphor II (Pharmacia) at 15°C for 1 h at 300 V, 1 h at 400 V, 1 h at 500 V, 3 h at 3500 V and 24 h at 5000 V. After the first-dimensional separation, strips were equilibrated in 6 M urea, 30% glycerol, 2% w/v SDS, 0.05 M Tris-HCl, pH 6.8, and 2% w/v DTE for 15 min. The strips were subsequently equilibrated for 15 min in buffer, in which DTE was replaced with 2.5% w/v iodoacetamide. In the second-dimensional run, the proteins were separated on 9–16% linear gradient polyacrylamide gels until the dye front had reached the bottom. The gels were fixed in a solution containing 10% acetic acid and 25% 2-propanol for 30 min, treated with Amplify (Amersham Pharmacia Biotech) for 30 min and vacuum-dried before exposure to Kodak Biomax-MR X-ray films. Preparative gels for Edman degradation and matrix-assisted laser desorption/ionization time of-flight mass spectrometry (MALDI-TOF-MS) were loaded with 4 mg protein per strip. Preparative gels were run for 72 h instead of 24 h in the last step of the focusing.

2.3 In-gel digestion of proteins

For visualization of HeLa cell proteins on preparative gels for identification by MALDI-MS, 2 × 10^6 cpm of labeled proteins from cells grown in parallel were added to the unlabeled samples. After separation by 2-D PAGE (IPG) the gels were vacuum-dried without prior fixation. In-gel digestion was carried out essentially as previously described [6, 10, 11]. Protein spots were excised from the gels, washed in 50 mM NH_4HCO_3/acetonitrile (60/40) and dried by vacuum centrifugation. Modified porcine trypsin (12 ng/mL sequencing grade; Promega, Madison, WI, USA) in digestion buffer (50 mM NH_4HCO_3) was added to the dry gel pieces and incubated on ice for 1 h for reswelling. After removing the supernatant, 20–40 μL digestion buffer without trypsin was added and the digestion was continued at 37°C for 4–18 h. The peptide mixture was extracted, vacuum-dried and redissolved in 10–20 μL of 5% formic acid prior to mass analysis.

2.4 Desalting and concentration of peptide mixtures

Custom chromatographic columns, consisting of 300 nL of Poros R2 material (20 or 50 μm bead size; PerSeptive Biosystems, Framingham, MA, USA) packed into a constricted GELoader pipette tip (Eppendorf, Hamburg, Germany), were used for desalting and concentration of the extracted peptide mixtures prior to mass spectrometric analysis [12]. A 10 mL syringe was used to force liquid through the column by applying gentle air pressure. Peptide mixtures were dissolved in 5% formic acid, loaded onto the column, and washed with 20 μL of 5% formic acid. For analyses by MALDI-MS, peptides were eluted with 0.4 μL matrix solution (15–20 g/L of α-cyano-4-hydroxycinnamic acid in 70% acetonitrile) and deposited directly onto the target.

2.5 MALDI-TOF-MS of digests

A Bruker REFLEX MALDI time-of-flight mass spectrometer (Bruker-Franzen Analytik GmbH, Bremen, Germany), equipped with the SCOUT source and variable detector bias gating, was employed for mass analysis of peptide mixtures in positive ion reflector mode. Ion acceleration voltage was 22 kV. Two MALDI sample preparation methods were used. In the first, the dried droplet method, analyte solution, 2% TFA, and the matrix solution (15–20 g/L α-cyano-4-hydroxycinnamic acid in 70% acetonitrile) were mixed in equal proportions (0.3 μL) on the target and dried. The dry crystalline deposit was rinsed carefully with a small volume of 0.1% TFA. In the second method, peptide mixtures were desalted by using custom microcolumns and directly deposited onto the MALDI probe by elution with matrix solution as described above.

2.6 *N*-terminal sequencing

For identification by Edman degradation, proteins were electrotransferred from preparative gels to PVDF membranes (Millipore, Bedford, MA, USA) according to Towbin *et al.* [13]. These membranes were then stained with Coomassie Brilliant Blue R-250, and spots from various positions in the 2-D pattern were excised and analyzed with an Applied Biosystems 494 protein sequencer (Perkin Elmer, Norwalk, CT, USA).

2.7 Database search for protein identification

Identification of peptide mass fingerprints was performed by searching the peptide mass map in a comprehensive nonredundant protein sequence database (NRDB, European Bioinformatics Institute, Hinxton, UK) using PeptideSearch software [14], which was further developed at EMBL. The protein identifications were evaluated using the "second pass search" feature of the software. Peptide Search and the database indexfile are available *via* the WWW at http//:www.mann.embl-heidelberg.de. Identification of proteins based on their *N*-terminal sequences was done using the FastA program from the GCG Package (Genetics Computer Group, Madison, WI, USA).

2.8 Software for construction of WWW-2D PAGE database

The Melanie 2-D software package for SUN Solaris (Bio-Rad, Richmond, CA) was used to annotate gel pictures. The Wisconsin Package Version 9.1 (GCG) for UNIX was used for extracting SWISS-PROT protein sequences. To construct the WWW database the Make2ddb package from the ExPASy server (http://expasy.hcuge.ch/ch2D/make2ddb.html) was used. For editing sequences the text editor "emacs" was employed.

3 Results and discussion

3.1 Identification of HeLa cell proteins with MALDI-TOF-MS

Four mg of HeLa cell protein and 2×10^6 cpm of [^{35}S]-labeled HeLa cell protein were separated by 2-D PAGE (IPG 4–7), resulting in high-resolution 2-D gel patterns. Proteins from different areas of the gels were excised and subjected to in-gel digestion. Analysis of 23 spots was done by MALDI-TOF-MS. Figure 1 shows the peptide mass fingerprint that identified TCPE_HUMAN (T-complex protein 1). By searching the database with 20 peptide masses, matches were found to the TCPE_HUMAN and to an "unknown" reading frame from a cDNA clone which was 100 % identical to TCPE_HUMAN (Table 1). Fourteen matches corresponded to the mouse homologue of the T-complex protein 1. Of the 23 spots on which mass spectrometry was performed, 19 proteins were identified with high confidence (Table 2). The high identification rate is presumably due to the relatively high abundance of the proteins identified.

3.2 Identification of proteins by *N*-terminal sequencing

Four mg of HeLa cell proteins were separated by 2-D PAGE (IPG 4–7), and the proteins were transferred to

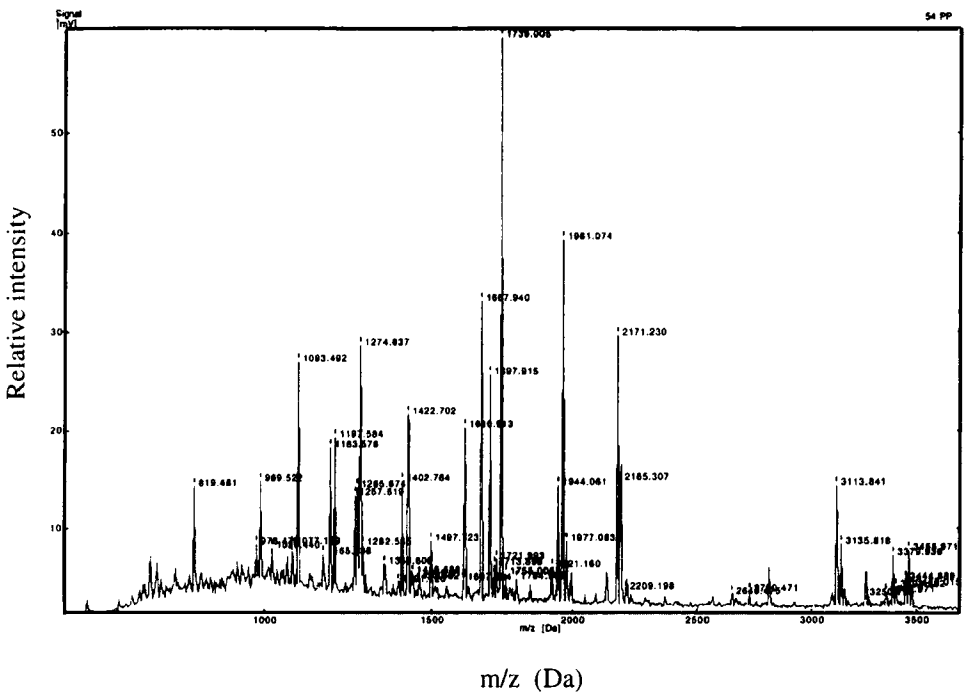


Figure 1. MALDI-TOF-MS spectrum of human T-complex protein. By means of PeptideSearch 20 peptides could be matched to the TCPE_HUMAN from the SWISS-PROT database.

Table 1. Database search of peptide masses that identified TCPE_HUMAN

Hits	Database I ac	M_r	Name
20	swissIP48643	59671.36	TCPE_HUMAN T-COMPLEX ROTEIN 1
20	tID43950	60122.89	HSKG1DD_1 GENE "KIAA0098"
14	swissIP80316	59624.42	TCPE_MOUSE T-COMPLEX PROTEIN 1
9	swissIQ13415	97359.20	ORC1_HUMAN ORIGIN RECOGNITION C
9	sptIQ13471	97350.18	Q13471 REPLICATION CONTROL PROTEIN
8	swissIP08428	58983.03	ATPA_XENLA ATP SYNTHASE ALPHA C

Table 2. HeLa proteins identified by MALDI-TOF-MS and *N*-terminal sequencing

Identification	M_r/p*I* calculated[a]	M_r/p*I* observed[b]	AC[c]	ID[d]	Name
M[e], E[f]	58.0*/5.24*	58.0/5.24	P10809	P60_HUMAN	Human mitochondrial HSP60
E	68.8*/5.51*	68.8/5.51	P38646	GR75_HUMAN	Human mitochondrial HSP70
M	70.9/5.37	67.2/5.40	Q13347	HS7C_HUMAN	Human heat shock cognate 71
M	59.7/5.45	57.3/5.50	P48643	TCPE_HUMAN	Human T-complex protein 1
M	42.6*/5.34	42.6/5.48	P12277	KCRB_HUMAN	Human creatine kinase, B-chain
E	49.1*/5.43	49.1/5.64	P31930	UCR1_HUMAN	Human cytochrome *c* reductase
M	36.5/5.38	37.9/5.52	Q13347	IF34_HUMAN	Human eukaryotic translation factor
M, E	24.6/4.50*	27.9/4.5	P24534	EF1B_HUMAN	Human elongation factor 1
M	16.7*/5.08	16.7/5.08	P10159	IF5A_HUMAN	Human initiation factor 5A
M	11.6*/4.82	11.6/4.85	P10599	THIO_HUMAN	Human thioredoxin
M	11.7/4.42	13.5/4.30	P42899	RLA2_HUMAN	Human 60S ribosomal protein P2
M	26.0/5.75	27.6/5.80	Q06323	IGUP_HUMAN	Interferon gamma up-regulated I-5111 protein precursor
M, E	54.3*/5.61	54.3/5.79	P30101	ER60_HUMAN	Human protein disulfide isomerase
M	14.6/5.34	12.8/5.00	P09382	LEG1_HUMAN	Human galectin-1
M	53.6/5.06	52.0/5.00	P08670	VIME_HUMAN	Human vimentin
M	50.6/4.97	50.2/5.00	P08779	K1CX_HUMAN	Human keratin, type I cytoskeletal 17
M, E		71.5/5.00			
M, E	70.3/4.98	71.5/4.97	P11021	GR78_HUMAN	Human GRP78
M, E		71.5/4.93			
M	21.9/4.91	22.3/4.83	P28072	PRCD_HUMAN	Human proteosome delta chain
M	38.6/6.64	37.0/6.42	P04083	ANX1_HUMAN	Human annexin I
M	53.1/5.83	53.8/6.07	P23381	SYW_HUMAN	Human tryptophanyl tRNA synthetase
M	23.2*/5.44	23.2/5.58	P09211	GTP_HUMAN	Human glutathione S-transferase

a) M_r/p*I* calculated: p*I* determined with pK_i of the amino acid residues in 9 M urea
b) M_r/p*I* observed: determined by using the Melanie II software
c) AC, SWISS-PROT accession number
d) ID, SWISS-PROT identification name
e) M, MALDI-MS (peptide mass tolerance: < 0.5 Da)
f) E, Edman degradation (see Table 3)
Asterisks (*), markers used for M_r/p*I* scale

PVDF membranes and stained with Coomassie blue. Although Edman degradation sequencing was done on 33 excised spots, sequences were obtained for only six protein spots. This was probably due to *N*-terminal blocking since a result was often not obtained even when analyzing strongly stained protein spots. A sequence length from 7 to 18 amino acids was obtained (Table 3). Database searching using the FastA program identified proteins unambiguously. Sequences were located either at the *N*-terminus of the protein or after a potential signal peptide. Four spots were identified by using both *N*-terminal sequencing and mass spectrometry, and both methods identified the same proteins (Table 2).

Table 3. *N*-terminal sequences and results of FastA database search

ID	Sequence: theoretical/experimental	Name
P60_HUMAN	AKDVKFG/ AKDVKFG	Human mitochondrial HSP60
GR75_HUMAN	ASEAIKG/ ASEAIKG	Human mitochondrial HSP70
UCR1_HUMAN	TATFAQALQ/ XATFAQALQ	Human cytochrome *c* reductase
EF1B_HUMAN	GFGDLKSPAG/ GFGDLKSPAG	Human elongation factor 1
ER60_HUMAN	SDVLELTDDN/ SDVLELTDDX	Human protein disulfide isomerase
GR78_HUMAN	EEEDKKEDVGTVVGIDLG/ EEEDKKEDVGTVVGIGLG	Human GRP78

'X' in the sequence: no interpretation could be made
ID, identification number in SWISS-PROT

3.3 Determination of molecular coordinates for proteins

Isoelectric points and molecular weights for the identified proteins were estimated with the "Compute pI/Mw tool" on the ExPASy server, which calculates the pK_i values for the amino residues in 9 M urea according to Bjellquist *et al.* [15, 16]. The proteins used to establish a pI/M_r scale are labeled with asterisks (*) in Fig. 2. The pI and M_r values were inserted using the landmark tool in the Melanie II software and gave good agreement between theoretical and observed coordinates for most of the identified proteins.

3.4 Construction of the parasite host cell interaction (PHCI) 2-D PAGE database using the Make2ddb software

The requirements to make a 2-D gel database with the Make2ddb software are a SUN computer with Solaris, and the Melanie II software. The first step in making the 2-D gel database was to detect the protein spots with the Melanie II software. Protein spots were then identified by peptide mass fingerprinting and/or *N*-terminal sequencing. They were then labeled with SWISS-PROT accession numbers using the labeling tool in the Melanie II software. In the next step, a data file containing a subset of the information from the SWISS-PROT data entry and the specific 2-D gel information was made. We generated this file by typing a file (2d.sw) containing the SWISS-PROT identification (ID) of each identified protein. The database entries were copied to the local disc with the command fetch @2d.sw. The identification line (ID), the accession number line (AC), and the description lines (DE) of the SWISS-PROT entry were extracted. This was done with the UNIX command "egrep".

```
egrep -h "^ID |^AC |^DE " *.sw > HELA01.dat
```

The extracted data were redirected to a data file with the same name as the Melanie 2-D gel image (HELA01) but with an extension [.dat] (HELA01.dat). The file was manually edited in emacs, each reference was separated by a line with //, and the specific data for the 2-D gels were added for each entry (2D). A masterline (MT) was added stating the name of the gel from which the spot had been identified, in this case HELA01.

The first line in the 2D entry indicates the name of the master gel (HELA01). This was followed by the experimentally determined isoelectric point and molecular weight for the protein, as exemplified by:

```
2D   -!- MASTER: HELA01;
2D   -!-   PI/MW=5.00/71500;
2D   -!-   PI/MW=4.97/71500;
2D   -!-   PI/MW=4.93/71500;
```

Therefore, if a protein with one accession number has many isoforms, all of these will be indicated. Furthermore, the entry can contain a serial number (spot 2-D number) that is a unique spot-identifier across all 2-D databases. The entry for one protein in the HELA01.dat file with relevant data is as follows:

```
//
ID   GR78_HUMAN;     PRELIMINARY      2DG.
AC   P11021;
DE   78 KD GLUCOSE REGULATED PROTEIN PRECURSOR (GRP 78)
(IMMUNOGLOBULIN DE HEAVY CHAIN BINDING PROTEIN) (BIP).
IM   HELA01.
```

```
2D   -!- MASTER: HELA01;
2D   -!-   PI/MW=5.00/71500;
2D   -!-   PI/MW=4.97/71500;
2D   -!-   PI/MW=4.93/71500;
DR   SWISS-PROT; P11021; GR78_HUMAN.
//
```

A database cross-reference (DR) is made for the protein to the SWISS-PROT, which is used as a link to other data for the protein. When the Melanie file (HELA01) and HELA01.dat files are generated, the new database site is made using the program Make2ddb. The PHCI-2DPAGE database is accessible at http://www.gram.au.dk.

4 Concluding remarks

The high identification rates of HeLa cell proteins obtained by MALDI-MS and *N*-terminal sequencing demonstrate the magnitude of human proteins existing in databases and their usefulness in proteomics. So far, this has enabled us to map and annotate 21 HeLa cell proteins.

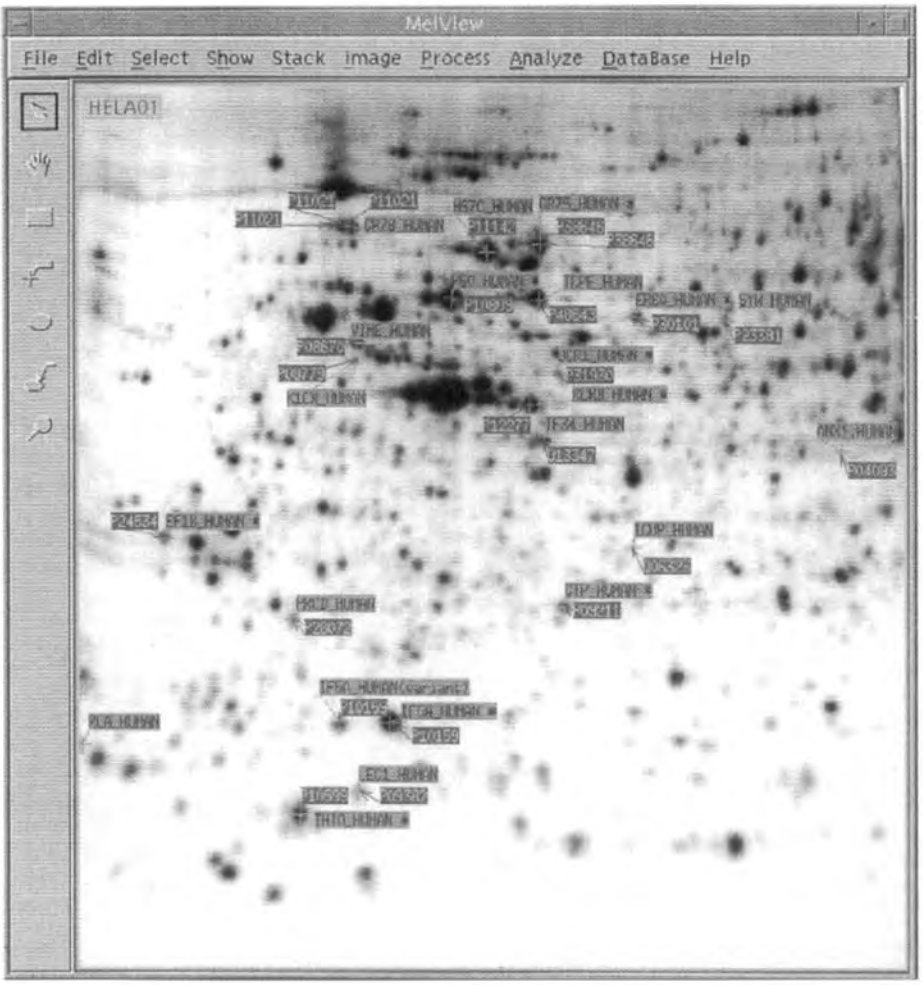


Figure 2. Two-dimensional gel of [^{35}S]methionine/cysteine-labeled HeLa cells, annotated in the Melanie II software with "landmarks": SWISS-PROT accessions names (text labels), SWISS-PROT accession numbers (numbered labels), asterisks (*) on labels indicate proteins with theoretical M_r/p*I* that were used for calculation of p*I* and M_r scales.

We used the pK_i in 9 M urea [11, 12] for selected marker proteins on the 2-D gels, which yielded a good agreement between theoretical and observed p*I* values obtained for the identified proteins. This gives a good opportunity to look for post-translational modifications of proteins, a growing field in proteomics. For example, the most basic variant of the human glucose binding protein GRP78 (annotated GR78_HUMAN) has a p*I* identical to that calculated, and the two acidic forms (P11021; Fig. 2) probably arise from phosphorylation events [17]. We are planning to include other data related to the mapped proteins such as changes in expression of proteins due to infection with parasites or treatment with interleukins. The widespread use of 2-D PAGE (IPG) has made interlaboratory comparisons of 2-D gels a manageable task [18]. Hopefully, a database of proteins from a human cell line used in a wide range of biological investigations will contribute to this standardization.

We are grateful to Karin Skovgaard and Inger Andersen for technical assistance. Financial support for this work was received from the Danish Health Research Council (grants 12-0850-1, 12-0150-1), the Danish Veterinary and Agricultural Research Council (grant 20-3503-1), Aarhus University Research Foundation, and the Danish Biotechnology Program.

Received October 25, 1998

5 References

[1] Lundemose, A. G., Birkelund, S., Larsen, P. M., Fey, S. J., Christiansen, C., *Infect. Immun.* 1990, *58*, 2478–2486.

[2] Birkelund, S., Bini, L., Palini, V., Sanchez-Campilo, M., Liberatori, S., Clausen, J. D., Østergaard, S., Holm, A., Christiansen, G., *Electrophoresis* 1997, *18*, 563–567.

[3] Bravo, R., Celis, J. E., *Clin. Chem.* 1982, *28*, 766–781.

[4] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[5] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[6] Nawrocki, A., Larsen, M. R., Podtelejnikov, A. V., Jensen, O. N., Mann, M., Roepstorff, P., Görg, A., Fey, S. J., Larsen, P. M., *Electrophoresis* 1998, *19*, 1024–1035.

[7] Hoogland, C., Sanchez, J.-C., Tonella, L., Hochstrasser, D. F., Appel, R. D., in: *From Genome to Proteome, 3rd Siena 2-D Electrophoresis Meeting* 1998, Abstract, p. 91.

[8] Appel, R. D., Bairoch, A., Sanchez, J.-C., Vargas, J. R., Golaz, O., Pasquali, C., Ravier, F., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 540–546.

[9] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[10] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173–179.

[11] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[12] Gobom, J., Nordhoff, E., Ekman, R., Roepstorff, P., *Int. J. Mass Spec. Ion Proc.* 1997, *169/170*, 153–163.

[13] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.

[14] Mann, M., Højrup, P., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 338–345.

[15] Bjellqvist, B., Hughes, G., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1023–1031.

[16] Bjellqvist, B., Basse, B., Olsen, E., Celis, J. E., *Electrophoresis* 1994, *15*, 529–539.

[17] Leustek, T., Amir-Shapira, D., Toledo, H., Brot, N., Weissbach, H., *Cell. Mol. Biol.* 1992, *38*, 1–10.

[18] Blomberg, A., Blomberg, L., Norbeck, J., Fey, S. J., Larsen, P. M., Larsen, M., Roepstorff, P., Degand, H., Boutry, M., Posch, A., Görg, A., *Electrophoresis* 1995, *16*, 1935–1945.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 984–993

Allan Christian Shaw[1,2]
Martin Røssel Larsen[3]
Peter Roepstorff[3]
Just Justesen[2]
Gunna Christiansen[1]
Svend Birkelund[1]

[1]Department of Medical Microbiology and Immunology
[2]Department of Molecular and Structural Biology, University of Aarhus, Denmark
[3]Department of Molecular Biology, University of Odense, Denmark

Mapping and identification of interferon gamma-regulated HeLa cell proteins separated by immobilized pH gradient two-dimensional gel electrophoresis

Interferon gamma (IFN-γ) is a potent immunomodulatory lymphokine, secreted by activated T-lymphocytes and NK-cells during the cellular immune response. Actions of IFN-γ are mediated through binding to the IFN-γ-receptor, present on most cells, and the subsequent activation of a great magnitude of IFN-γ responsive genes has been reported previously. Our goal is to identify and map IFN-γ-regulated HeLa cell proteins to the two-dimensional polyacrylamide gel electrophoresis with the immobilized pH gradient (IPG) two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) system. A semiconfluent layer of HeLa cells was grown on tissue culture plates, and changes in protein expression due to 100 U/mL IFN-γ were investigated at different periods after treatment, using pulse labeling with [^{35}S]methionine/cysteine in combination with 2-D PAGE (IPG). The identity of eight protein spots was elucidated by matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS), and several variants of the IFN-γ-inducible tryptophanyl-tRNA synthetase (hWRS) were detected by immunoblotting.

Keywords: Interferon gamma / Two-dimensional polyacrylamide gel electrophoresis / Immobilized pH gradient / HeLa / Pulse labeling
EL 3405

1 Introduction

IFN-γ is an important lymphokine, responsible for pleiotropic effects in the cell-mediated immune response. In contrast to IFN-α and β (type I interferon), which are produced in a wide range of cells upon virus infection, IFN-γ (type II interferon) is produced primarily by T-lymphocytes and natural killer cells [1, 2]. Actions of IFN-γ in the regulation of antigen-specific and nonspecific immunological functions involve the stimulation of IgG- class switching in B cells, increased expression of class I and II MHC complex antigens, antiproliferative effects, antiviral effects, and activation of both macrophage-dependent and independent killing of intracellular parasites [3]. Many activities of IFN-γ are believed to be mediated through IFN-γ-induced proteins, which are activated at the end of a signal transduction pathway, which has been described in some detail; IFN-γ acts through binding of a specific receptor on target cells. Upon ligand binding the receptor is activated by tyrosine phosphorylation through the JAK1 and JAK2 kinase that are associated with the cytoplasmic domain of the IFN-γ-receptor. The activated receptor complex subsequently catalyzes the tyrosine phosphorylation and activation of a group of cytoplasmic proteins, the STATs (signal transducers and activators of transduction), which, in turn, dimerize and enter the nucleus in order to bind and activate IFN-γ-specific DNA regulatory elements [4, 5].

The number of known IFN-γ-regulated genes has increased to over 200 genes, elucidating the magnitude of immunologic functions [6]. 2-D PAGE has been used as a powerful tool to investigate IFN-γ-regulated proteins in different cell lines in several investigations [7–11]. With this approach, differential expression of some IFN-γ-regulated proteins in different cell types [8] and synergic effects with other cytokines [12, 13] can be investigated in detail. Due to the recent improvements of 2-D PAGE (IPG) and the application of sensitive MALDI-MS for identification of protein spots, we now aim to localize and map IFN-γ-regulated proteins from a widely used cell line, using the 2-D PAGE (IPG) system. In this paper we report the mapping, identification and expression patterns of six different IFN-γ-regulated HeLa cell proteins.

Correspondence: Allan Christian Shaw, Department of Medical Microbiology and Immunology, The Bartholin Building, University of Aarhus, 8000 Aarhus C, Denmark
E-mail: shaw@medmicro.aau.dk
Fax: +45-8619-6128

Abbreviations: **hWRS**, human tryptophanyl-tRNA synthetase; **IFN-γ**, interferon gamma; **IGR**, interferon gamma-regulated; **K17**, cytoskeletal keratin 17; **OD**, optical density

2 Materials and methods

2.1 Sample preparation

HeLa 229 cells (human epitheloid carcinoma, cervix) were obtained from the American Type Culture Collection

(ATCC, Rockville, MD, USA) and cultivated in RPMI 1640 (Gibco BRL, Grand Island, NY, USA) containing 10 fetal calf serum (FCS) and 10 μg/mL gentamycin at 37°C in a 5% CO_2 atmosphere. Cells were grown to a semiconfluent monolayer. At the beginning of the pulse labeling experiments the medium was changed to a methionine-free RPMI 1640 medium (Life Technologies, Roskilde, Denmark) containing 10% FCS, 10 μg/mL gentamycin, 100 μCi/mL [^{35}S]methionine/cysteine (Promix, Amersham, UK) with or without addition of 100 U/mL human recombinant IFN-γ (Schering-Plough, Bloomfield, NJ, USA). Subsequently, cells were washed in PBS and loosened with a rubber policeman in a lysis buffer containing 9 M urea, 4% w/v 3-[(3-chloramidopropyl)dimethylammonium]-1-propane-sulfonate (CHAPS), 40 mM Tris base, 65 mM dithioerythritol (DTE), 2% v/v Pharmalyte 3–10; (Pharmacia Biotech, Uppsala, Sweden) and sonicated. Supernatants were stored at –70°C until used.

2.2 Two-dimensional gel electrophoresis

For isoelectric focusing, linear 18 cm immobilized pH 4–7 gradient drystrips (Pharmacia Biotech) were used. Each strip was soaked overnight, in 400 μL of the lysis solution containing the labeled proteins, in a reswelling tray (Pharmacia Biotech). Rehydrated strips were run on a Multiphor II (Pharmacia Biotech) at 15°C for 1 h at 300 V, 1 h at 400 V, 1 h at 500 V, 3 h at 3500 V and 24 h at 5000 V. After the first-dimensional separation, strips were equilibrated in 6 M urea, 30% glycerol, 2% w/v SDS, 0.05 M Tris-HCl, pH 6.8, and 2% w/v DTE for 15 min. The strips were subsequently equilibrated for another 15 min in a buffer in which DTE was replaced with 2.5% w/v iodoacetamide. A Protean II xi Multicell (Bio-Rad, Richmond, CA, USA) was used for the second-dimensional run. Proteins were separated on 9–16% linear gradient SDS-PAGE gels (18 cm × 20 cm × 1 mm) until the dye front had reached the bottom. The gels were fixed in a solution containing 10% acetic acid and 25% 2-propanol for 30 min, treated with Amplify (Amersham, Uppsala, Sweden) for 30 min and dried in a vacuum drier before exposure to Kodak Biomax-MR X-ray films for different periods of time.

2.3 Computer analysis

X-ray films were scanned on an HP Scanjet 3c/T. Changes in protein expression were investigated by means of the Melanie II software (Bio-Rad). The optical density volumes of regulated spots were measured. Actin, α-tubulin and β-tubulin were not found to be regulated by IFN-γ in this system and were therefore used to normalize the OD volumes (optical density integrated over the area) in order to obtain values for an increase or decrease in expression. In addition, the OD volumes of the protein spots were normalized to the total volume of protein spots on the gel.

2.4 Preparative 2-D PAGE

IPG strips for preparative 2-D PAGE were loaded with 1 mg or 4 mg unlabeled HeLa cell protein. For visualization of HeLa cell proteins on X-ray films, 2×10^6 cpm labeled proteins from cells grown in parallel with unlabeled cells were added to the sample preparations. In order to obtain sufficient isoelectric focusing of proteins, the last step of the isoelectric focusing was extended to 72 h. The preparative gels were vacuum-dried without prior fixation of the proteins. For identification by MALDI-TOF-MS, protein spots were excised from preparative 2-D gels with IFN-γ-untreated or treated HeLa cells. Protein spots were either excised from one gel loaded with 4 mg HeLa cell protein or pooled together from four gels, each loaded with 1 mg protein. To verify that the correct protein was removed, a fresh X-ray film was exposed after removal of the protein spots.

2.5 In-gel digestion

Digestion of protein in excised gel plugs (in-gel) was performed as described [14, 15] with minor modifications [16]. The excised gel plugs were washed in 50 mM NH_4HCO_3/acetonitrile (60/40) and dried by vacuum centrifugation. Modified porcine trypsin (12 ng/μL, sequencing grade; Promega Madison, WI, USA) in digestion buffer (50 mM NH_4HCO_3) was added to the dry gel pieces and incubated on ice for 1 h for reswelling. After removing the supernatant, 20–40 μL digestion buffer was added, and the digestion was continued at 37°C for 4–18 h. The peptide mixture was extracted with NH_4HCO_3/acetonitrile (40/60), dried in a vacuum centrifuge and redissolved in 10–20 μL of 5% formic acid prior to mass analysis.

2.6 Desalting and concentration of peptide mixtures

Custom chromatographic columns made from disposable laboratory plasticware were used for desalting and concentration of the extracted peptide mixtures prior to mass spectrometric analysis [17]. A column consisted of 300 nL of Poros R2 material (20 or 50 μm bead size; PerSeptive Biosystems, Framingham, MS, USA) packed into a constricted GELoader pipette tip (Eppendorf, Hamburg, Germany). In order to retain the column resin the narrow end of the GELoader tip was constricted by squeezing, *e.g.*, with a pair of heated tweezers. A 10 mL syringe was used to force liquid through the column by applying gentle air

pressure. Peptide mixtures were dissolved in 5% formic acid, loaded onto the column, and washed with 20 μL of 5% formic acid. For analyses by MALDI-TOF-MS, peptides were eluted with 0.4 μL matrix solution (15–20 g/L of α-cyano-4-hydroxycinnamic acid in 70% acetonitrile) and deposited directly onto the target.

2.7 Peptide mass mapping by MALDI-TOF-MS

A REFLEX MALDI time-of-flight mass spectrometer (Bruker-Franzen Analytik GmbH, Bremen, Germany) equipped with the SCOUT source and variable detector bias gating was employed for mass analysis of peptide mixtures in positive ion reflector mode. Ion acceleration voltage was 22 kV. Two MALDI sample preparation methods were used: (i) For coprecipitation of analyte and matrix according to the dried droplet method, the analyte solution, 2% TFA and the matrix solution (15–20 g/L α-cyano-4-hydroxycinnamic acid in 70% acetonitrile) were mixed in equal proportions (0.3 μL) on the target and dried. The dry crystalline deposit was rinsed carefully with a small volume of 0.1% TFA. (ii) Peptide mixtures were desalted by using custom microcolumns and directly deposited onto the MALDI probe by elution with matrix solution as described above.

2.8 Protein identification by database search

Protein identification was performed by searching the peptide mass map in a comprehensive nonredundant protein sequence database (nrdb, European Bioinformatics Institute, Hinxton, UK) using PeptideSearch software [18] further developed at EMBL. The protein identifications were evaluated using the "second pass search" feature of the software. PeptideSearch and the database index file are available *via* the WWW at http//:www.mann.embl-heidelberg.de.

2.9 Immunoblotting

Protein sample (0.5 mg), separated by 2-D PAGE, was electrotransferred to polyvinylidene difluoride (PVDF) Immobilon-P membranes (Millipore, Bedford, MA, USA) according to [19]. Immunodetections were carried out with a rabbit polyclonal antibody directed against human tryptophanyl-tRNA synthetase (anti-hWRS) diluted 1/4000 in a buffer containing 150 mM NaCl, 20 mM Tris, 0.2% gelatine, 0.05% Tween-20 (Bio-Rad) and 2% normal goat serum (Dako, Glostrup, Denmark). The secondary antibody used was alkaline phosphatase-conjugated goat anti-rabbit IgG (Bio-Rad) diluted 1:2000. Blots were stained with 5-bromo-4-chloro-3-indolylphosphate toluidium (BCIP)/nitroblue tetrazolium (NBT) (Bio-Rad).

3 Results and discussion

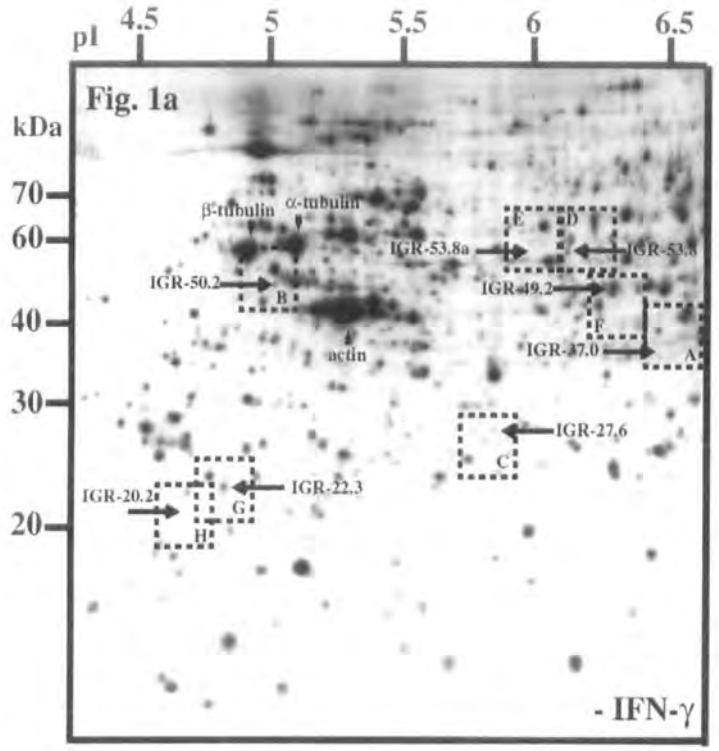
3.1 2-D PAGE (IPG) of [^{35}S]methionine/cysteine HeLa cell proteins

HeLa cells ere either untreated or treated in presence of 100 U/mL IFN-γ. To investigate changes in protein expression to different points after addition of IFN-γ, cells were labeled with [^{35}S]methionine/cysteine in the intervals 0–6 h, 6–12 h, 18–24 h and 42–48 h. Labeled proteins were harvested at the end of each interval and subjected to 2-D PAGE (IPG), yielding high-resolution 2-D gels, as exemplified in Fig. 1a and b. IPG strips in the range of pH 4–7 were used for better separation of acidic IFN-γ-regulated proteins and to reduce potential contamination of excised protein spots, coming from closely adjacent proteins, during identification.

3.2 Analysis of IFN-γ-specific changes in HeLa cell protein expression

Several proteins with altered expression due to IFN-γ treatment were observed (Fig. 1a and b). The criterion for up- or down-regulation of proteins due to IFN-γ was based on an evaluation of a minimum of two gels per experimental condition. 2-D gels with IFN-γ-treated cell lysates from each time interval were compared to IFN-γ-untreated controls gels from the same time interval in order to eliminate possible changes in protein expression due to culture time. Enlargements from total 2-D gel images show the identified IFN-γ-regulated proteins (designated IGR; molecular mass in kDa) and their degree of induction to different times after addition of IFN-γ (Fig. 2–H). OD volumes of proteins were normalized to OD volumes of either actin, α-tubulin, β-tubulin or to the total OD volume of proteins on the gel. When possible, the fold increase/suppression was calculated as follows: normalized OD volumes from IFN-γ-treated/normalized OD volumes from untreated. The values obtained from the different normalizations did not vary significantly, and their mean values are used as estimations for fold increase/suppression.

We estimated a 6-fold increase in incorporation of [^{35}S]methionine/cysteine into IGR-37.0 (p*I* 6.4/37.0 kDa) upon 6 h treatment with IFN-γ (Fig. 2A). The induction declined to a 2-fold increase at 48 h. Another protein, IGR-50.2 (p*I* 5/50.1 kDa) also reached maximum induction early in the time-course study, showing an approximately 3-fold increase in protein expression at 12 h IFN-γ-treatment (Fig. 2B). At 24 h the up-regulation of IGR-50.2 declined and no induction was detectable 48 h after IFN-γ treatment (Fig. 2B). Furthermore, a protein with the same molecular mass, but to the acidic side of IGR-50.2, was induced in a similar fashion (open arrow, Fig. 2B). The



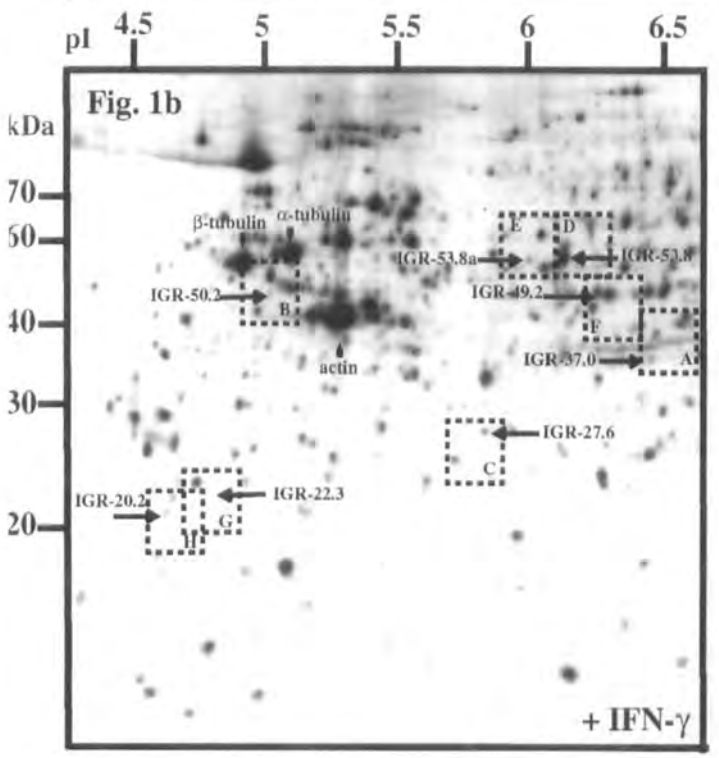


Figure 1: Total 2-D PAGE images of [^{35}S]methionine/cysteine-labeled HeLa cell proteins (a) untreated or (b) treated with 100 U/mL IFN-γ for 48 h. Identified regulated proteins are designated IGR. Quadrants are enlarged in Fig. 2. Actin, β-tubulin and α-tubulin are indicated by arrowheads.

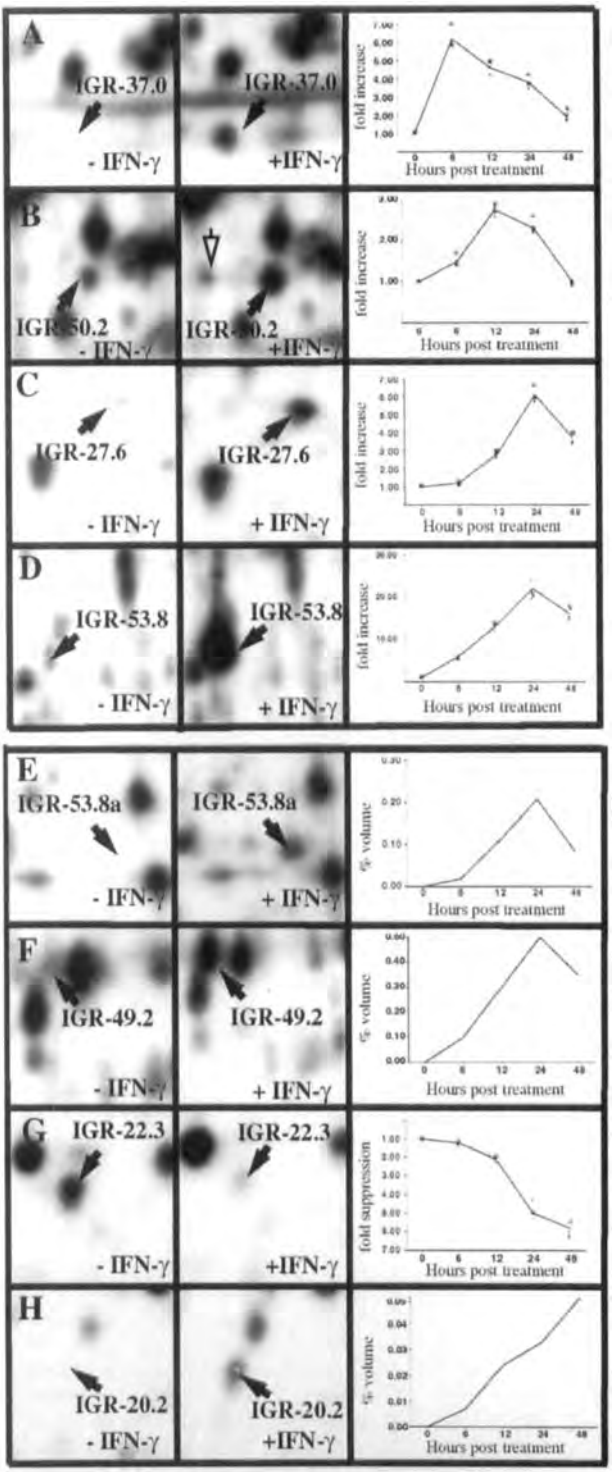


Figure 2: Enlargements of quadrants from Fig. 1a and b, showing the proteins at the time of their maximal regulation by IFN-γ. Graphs show the fold increase/suppression to 6, 12, 24 and 48 h after IFN-γ treatment. Intensity of protein spots were normalized to OD volumes of either (o) actin, (+) α-tubulin, (x) β-tubulin or (@) to the total OD volume on the gel (see Fig. 1 for positions on the gels). The means of these values are shown as lines. IGR-20.2, IGR-53.8a and IGR-49.2 was not measurable on 2-D gels with untreated cells and their expression was estimated by normalizing to the total volume of spots on 2-D gels with IFN-γ-treated cells.

induction of IGR-27.6 (p/ 5.8/27.6 kDa) was not significant before 12 h after IFN-γ treatment, where we estimated a 3-fold increase in [^{35}S]methionine/cysteine incorporation. At 24 h post IFN-γ treatment, a 6-fold increase was measured, which declined to an estimated 4-fold increase at 48 h (Fig. 2C). IGR-53.8 (p/ 6.1/53.8 kDa) was the most significantly up-regulated protein observed, peaking, with an estimated 22-fold increase, upon 24 h IFN-γ treatment (Fig. 2D). IGR-53.8a (p/ 6/53.8 kDa) showed a similar expression pattern in the time-course study and had the same molecular mass as IGR-53.8. However, IGR-53.8a had a slightly more acidic p/, indicating that this protein might be a post-translational modification of IGR-53.8 (Fig. 2E). In proximity to these proteins, IGR-49.2 (p/ 6.2/49.2 kDa) was strongly up-regulated and reached its peak at 24 h after addition of IFN-γ (Fig. 2F). Similar to IGR-53.8 and IGR-53.8a the induction of IGR-49.2 was reduced, but still significant at 48 h.

In the acidic end of the gels two proteins, in close proximity to each other, IGR-22.3 (p/ 4.8/22.3 kDa) and IGR-20.2 (p/ 4.7/20.2 kDa) were reciprocally regulated by IFN-γ (Fig. 1a and b). For IGR-22.3, a 2-fold suppression was estimated but not until 12 h of IFN-γ treatment. The effect of treatment was at its maximum at 48 h, showing a 6-fold suppression (Fig. 2G). IGR-20.2 was weakly induced at 6 h and the increase in expression due to IFN-γ continued through the entire 48 h period, showing the opposite protein expression pattern as IGR-22.3 (Fig. 2H). In contrast to IGR-22.3, we were not able to detect any expression of IGR-20.2 in untreated cells, indicating that this protein is not constitutively expressed in HeLa cells, but specifically induced upon IFN-γ treatment.

3.3 Identification of IFN-γ-regulated proteins

The identification of HeLa cell proteins regulated by IFN-γ was primarily performed by peptide mass mapping by MALDI mass spectrometry [20]. The identification of IGR-53.8 is given as an example in Fig. 3. The peptide mass fingerprints, shown in Fig. 3A, were obtained by analyzing IGR-53.8 excised from a gel loaded with 4 mg HeLa protein from IFN-γ-treated cells. Using Peptide Search 18 measured peptide masses matched the human tryptophanyl-tRNA synthetase (hWRS) and covered 50% of the sequence, 13 peptide masses matched the bovine WRS, and 9–10 peptide masses matched the rabbit and mouse WRS (Fig. 3B and C). The remaining matches, which were ranked significantly lower, were to unrelated proteins; these had molecular masses and isoelectric points that were not comparable with the one estimated for IGR-53.8. A total of 11 proteins were analyzed by this method and seven were positively matched to an entry in the protein sequence database. The remaining nonidentified proteins are probably proteins that are not contained in the protein sequence database. Typically, 2–5% of a mixture of peptides generated by in-gel proteolytic digestion was used for MALDI analysis, except when the extracted dried peptide mixture had to be desalted and concentrated on a microcolumn for ultimate sensitivity. Traces of contaminating human keratin could only be detected when analyzing very low-abundant proteins. The corresponding peptides were matched to the keratin amino acid sequence; they were used to calibrate each spectrum to obtain a high mass accuracy, which is crucial for making unambiguous identifications [21].

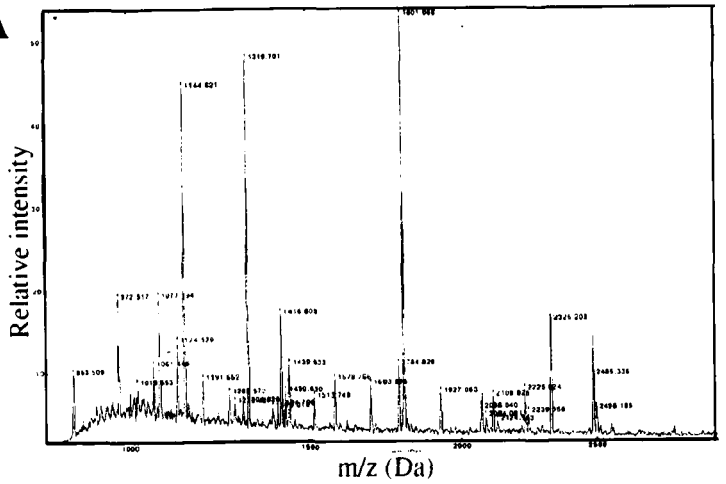
3.4 Identification of IGR-37.0 as lipocortin I (annexin I)

Lipocortin I is an inhibitor of inflammation and mediates effects of glucocorticoids, a group of hormones, that are involved in inhibition of T-lymphocyte proliferation and in their production of various cytokines, including IFN-γ [22, 23]. Although a constitutive expression of lipocortin I has been observed in other cancer cell lines [13], it is, to our knowledge, a novel finding that lipocortin I can be up-regulated by IFN-γ. The IFN-γ-induced up-regulation of lipocortin I in a cancer cell line is intriguing, as it has recently been shown that lipocortin I alone is an efficient inhibitor of an antigen-specific CD4+ T-cell line and thereby also an inhibitor of IFN-γ secretion [24]. Recent investigations provide evidence that treatment with IFN-γ includes up-regulation of proteins that are involved in apoptosis of certain cell types including HeLa cells [25, 26]. In agreement with this, we observed significant changes in morphology in addition to decreased proliferation microscopically when cells were treated 24 h or more with 100 U IFN-γ (data not shown). Our results suggest that the induction of lipocortin I is predominant early after IFN-γ treatment. (Fig. 2A). The up-regulation of lipocortin I in response to IFN-γ *in vivo* might serve as a safeguard mechanism that prevents T-cells from producing levels of IFN-γ that are apoptotic for its target cells. The IFN-γ-dependent up-regulation of lipocortin I in certain cell types could therefore be part of a negative feedback loop, which regulates the production of IFN-γ.

3.5 Identification of IGR-50.2 as cytoskeletal keratin 17

IGR-50.2 was found to be identical to the intermediate filament protein cytoskeletal keratin 17 (K17), so far the only IFN-γ-inducible cytokeratin [27]. K17 is up-regulated in stressed epithelial tissues such as inflammatory reactions in the skin of psoriasis patients, where the IFN-γ levels are high and in the development of abnormal scarring in contractile epithelial tissues [28]. We observed a protein

A



B

MPNSEPASLLELFNSIATQGELVRSLKAGNASKDEIDSAVKMLVSLKMSYKAAAGEDYKADCPFGNPAPTSNHGPDA
TEAEEDFVDPWTVQCSSAKGYDYDKLIVRFGSSKIDKELINRIERATGQRPHHFLRRGIFFSHRDMNQVLDAYENK
KPFYLYTGRGPSSEAMHVGHLIPFIFTKWLQDVFNVPLVIQMTDDEKYLWKDLTLDQAYSYAVENAKDIIACGF
DINKTFIFSDLDYMGMSSGFYKNVVKIQKHVTFNQVKGIFGFTDSDCIGKISFPAIQAAPSFSNSFPQIFRDRTDIC
CLIPCAIDQDPYFRMTRDVAPRIGYPKPALLHSTFFPALQGAQTKMSASDPNSSIFLTDTAKQIKTKVNKHAFSGGR
DTIEEHRQFGGNCDVDVSFMYLTFFLEDDDKLEQIRKDYTSGAMLTGELKKALIEVLQPLIAEHQARRKEVTDEIVK
EFYTPRKLSFDFQ

C

Index	No. of matched peptides	Acc. No	Mass (Da)	Identification
239867	18	tIM61715	53087.54	HSWRSAA_1 gene: "WARS", product
51591	18	swissIP23381	53165.66	SYW_HUMAN TRYPTOPHANYL-TRNA SYN
233610	17	tIX62570	53259.77	HSIFP53_1 product: "IFP53". H.
243714	13	tIX62113	52139.44	BTWTRLIG_1 Bovine mRNA for tryp
51587	13	swissIP17248	53729.16	SYW_BOVIN TRYPTOPHANYL-TRNA SYN
213	10	gnewIP23612	53799.38	SYW_RABIT TRYPTOPHANYL-TRNA SYN
51597	9	swissIP23612	53843.46	SYW_RABIT TRYPTOPHANYL-TRNA SYN
51593	9	swissIP32921	54283.01	SYW_MOUSE TRYPTOPHANYL-TRNA SYN
263787	9	tIX69656	53566.17	MMWRSA_1 gene: "WRS", product:
244488	8	tIM33460	53900.64	OCERFX_1 Rabbit eucaryotic rele
99441	7	sptIO00934	93334.16	AF006826ITGAF6626_1 product: "m
106375	7	sptIQ26607	17354.99	Q26607 TRYPTOPHANYL-TRNA SYNTHE
11763	7	swissIP51060	95633.35	CAPP_THESP PHOSPHOENOLPYRUVATE
87061	7	sptIQ18479	97378.68	Q18479 SIMILAR TO VAV PROTO-ONC
163896	7	sptIO33833	49841.67	O33833 BETA-FRUCTOSIDASE (EC 3.
86534	6	sptIQ17384	96025.72	Q17384 HUM-3 (FRAGMENT) //:trem
132630	6	sptIQ48156	92707.26	Q48156 SEROTYPE A CAPSULATION L
222682	6	sptnewIG274994	73378.91	G2749947 DNA TOPOISOMERASE IV P
243566	6	tIL04797	82435.18	BTPIGR_1 product: "polymeric im

Search Conditions:
Number of peptides included in search: 33
Protein mass window: 0-1000 kDa
Mass accuracy: 0.150 Da
No. of peptides required for match: 6

Figure 3: (A) Monoisotopic peptide mass fingerprint spectrum of IGR-53.8. (B) Alignment of experimentally obtained peptide sequences with hWRS. The underlined sequences correspond to the masses from the peptide mass fingerprint in (A). Bold letters indicate sequences, which matched experimental masses more than one time. (C) Result from Peptide Search, showing the identification of IGR-53.8 as human tryptophanyl-tRNA synthetase.

to the acidic side of K17 with identical molecular mass. Although this is a low-abundant protein, it was still detectable on gels with untreated cells and the expression patterns were similar to K17 upon IFN-γ treatment (Fig. 1G, open arrow). Phosphorylation events have been shown to regulate the assembly and disassembly of intermediate filament monomers [29], but whether the spot represents a phosphorylated variant of K17, another post-translational modification, or a different protein remains to be elucidated.

3.6 Identification of IGR-27.6 as interferon up-regulated protein 5111 (IGUPI-5111)

IGR-27.6 was identified as IGUPI-5111 [9] or γ7a [8], a IFN-γ-inducible protein, which was previously found up-regulated in keratinocytes and AMA cells by 2-D PAGE with carrier ampholytes. The protein shows homology to cytoskeletal proteins. Recently, however, Realini *et al.* [30] suggested that IGUPI-5111 is an IFN-γ-inducible activator of the 26S proteasome. The expression reached a maximum with a 6-fold up-regulation at 24 h IFN-γ-treatment. This was in agreement with previous studies, done by using 2-D PAGE with carrier ampholytes, where a 6-fold increase of IGUPI-5111 in HeLa cells due to IFN-γ was estimated [30]. However, the concentration used in these experiments was 200 U/mL IFN-γ, and the changes were investigated 72 h after treatment [30]. A 2.7-fold increase of IGUPI-5111 was observed in keratinocytes 19 h after addition of 100 U/mL IFN-γ [9], indicating that the degree of regulation of this protein varies depending on the cell line used.

3.7 Identification of IGR-53.8, IGR-53.8a and IGR-49.2 as tryptophanyl-tRNA synthetase

IGR-53.8 (Fig. 3) and IGR-53.8a were matched to hWRS. This protein was designated γ2 in studies done by Celis *et al.* [8] and Flecner *et al.* [31]. hWRS is found heavily up-regulated in a wide range of human cell lines due to IFN-γ treatment [8, 32]. We estimated a maximum 22-fold up-regulation of synthesis of hWRS protein compared to the 42-fold increase in mRNA levels observed by Flecner *et al.* [32] (Fig. 2D). A plateau of expression levels of hWRS was reported for HeLa cells treated longer than 24 h with IFN-γ [32]. Differences in quantification between protein and mRNA levels probably reflect the delay of translation compared to transcription or changes in mRNA stability.

The existence of alternative forms of hWRS has been reported previously [33] and led us to perform immunoblotting with a polyclonal antibody directed against hWRS (anti-hWRS). In support of the results obtained by MALDI-MS, immunoblotting with anti-hWRS was able to recognize IGR-53.8 and IGR-53.8a as indicated in Fig. 4. In addition, anti-hWRS reacted weakly with two additional protein spots to the acidic side (Fig. 4, arrow 1 and 2) of IGR-53.8 and one to the basic side (Fig. 4, arrow 3) of IGR-53.8, suggesting extensive post-translational modification of hWRS. Another basic variant of hWRS colocated with IGR-49.2. This variant had a molecular mass in accordance with the theoretical value obtained from an alternatively spliced hWRS mRNA product, described by Tolstrup *et al.* [33], and which was designated as γ3 by Celis *et al.* [8]. All variants of hWRS were found on immunoblots from untreated cells, suggesting that the modifications are not depending on IFN-γ treatment. Although equivalent amounts of protein were loaded on the gels, the protein spots reacted more intensively with anti-hWRS on immunoblots with IFN-γ-treated cells compared to IFN-γ-untreated controls, suggesting that the expression of all hWRS variants is regulated by IFN-γ (Fig. 4A compared to 4C).

It has been proposed that WRS may have other functions in the mammalian cell, besides assuring tryptophanylated tRNA for protein synthesis [34]. For example, the bovine WRS was recently associated with kinase activity and is itself being autophosphorylated [35]. hWRS is also involved in the synthesis of the diadenosine oligophosphates (Ap_3A)upon treatment with IFN-γ, which, together with the IFN-γ-inducible 2–5A synthetase and the tumor suppressor Fhit, has been proposed to be part of a growth regulatory pathway [36].

3.8 Identification of IGR-20.1 and IGR-22.3 as proteasome subunits LMP2 and Y.

The proteasome is a large multi-catalytic protease complex involved in the degradation of ubiquitin-conjugated peptides destined for MHC (major histocompatibility complex) class I antigen presentation on the surface of eukaryotic cells [37]. IFN-γ modifies the peptidase activity of the proteasome by regulating pairs of proteasome subunits reciprocally [38, 39]. In agreement with our findings, LMP2 (PSMB9), which is a β-subunit of the proteasome, is up-regulated upon stimulation with IFN-γ, whereas the constitutively expressed β-subunit Y (PSMB6) is down-regulated [40]. LMP2 and Y have very high sequence homology, suggesting that these substitute for each other in the proteasome [40]. The expression patterns in our study suggest that, initially, LMP2 is up-regulated at a higher rate than the rate at which subunit Y is suppressed (Fig. 2 E, F). A slower decrease in expression of subunit Y would allow the buildup of LMP2 before the substitution with subunit Y is initiated in the proteasome. In addition to the LMP2/subunit Y pair, IFN-γ modulates the exchange of the β-subunits LMP7 (PSMB8) and MECL1 (PSMB10) with the β-subunits X (PSMB5) and Z (PSMB7), respectively [41]. The substitution of β-subunits has been suggested to alter the proteolytic specificity of the proteasome, thereby changing the composition of peptides destined for MHC presentation [42].

A common feature for most of the proteins investigated in this study is an apparent decrease in induction upon 48 h IFN-γ treatment. It was recently shown that IFN-γ-activated STAT1 is degraded by the ubiquitin-proteasome pathway in HeLa cells [43], which could be part of the explana-

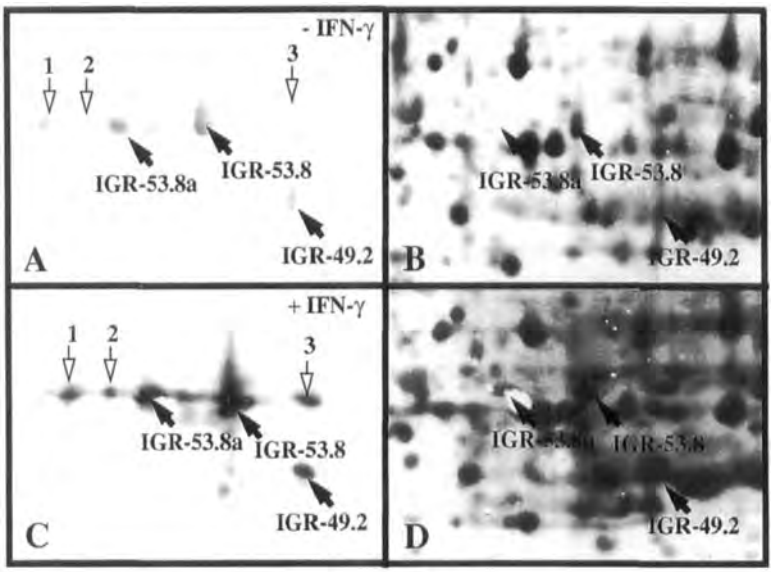


Figure 4: Immunostaining with a polyclonal antibody directed against tryptophanyl-tRNA synthetase (anti-hWRS) (A) untreated, (C) treated with 100 U/mL IFN-γ for 24 h. (B) and (D) show [^{35}S]methionine/cysteine-labeled proteins corresponding to proteins from (A) and (C), respectively. A protein colocating with IGR-49.2 is immunostained. IGR-53.8a and 1, 2, 3 (open arrows) are variants probably arising from post-translational modifications of hWRS.

Table 1: Fold increase/decrease

Spot	p*I*/M_r	Identification method	Maximum fold increase/ suppression	Time of maximum regulation (h)	Identity	SWISS PROT accession No.
IGR-37.1	6.4/37.0	MS	~ + 6	6 h	Lipocortin I	P04083
IGR-50.2	5/50.2	MS	~ + 3	12 h	Cytokeratin 17 (K17)	Q04495
IGR-27.6	5.8/27.6	MS	~ + 6	24 h	IGUPI-5111	Q06323
IGR-53.8	6.1/53.8	MS, IMB	~ + 22	24 h	Tryptophanyl-tRNA syntethase	P23381
IGR-53.8a	6/53.8	MS, IMB	(+)	24 h	Tryptophanyl-tRNA syntethase	P23381
IGR-49.2	6.2/49.2	IMB	(+)	24 h	Tryptophanyl-tRNA syntethase	P23381
IGR-22.3	4.8/22.3	MS	~ -6	48 h	Proteasome subunit Y	P28072
IGR-20.2	4.7/20.2	MS	(+)	48 h	Proteasome subunit LMP2	P28065

+, up-regulated; –, down-regulated; (+), up-regulated, (fold increase not determined). MS, MALDI-MS, IMB, immunoblotting

tion. Interestingly in this respect, the regulation of the proteasome subunits LMP2 did not seem to decline at 48 h, which might indicate that the signal transduction pathway responsible for activation of this gene is less dependent on activated STAT1.

4 Concluding remarks

We have used pulse labeling with [^{35}S]methionine/cysteine and 2-D PAGE (IPG) to locate and map IFN-γ-regulated HeLa cell proteins to the 2-D PAGE (IPG) system. The high loading capacity of IPG drystrips made it possible to make unambiguous identifications of IFN-γ-regulated HeLa cell proteins by sensitive MALDI-MS techniques. So far, we have mapped and identified lipcortin I, cytokeratin 17, IGUPI-5111, proteasome subunits LMP2/Y and several variants of the hWRS protein and investigated their expression patterns, as summarized in Table 1. We are planning to include data concerning the IFN-γ inducibility of identified proteins from this study, in addition to those currently under investigation, in a 2-D PAGE database containing HeLa cell proteins [44].

We are grateful to Karin Skovgaard and Inger Andersen for technical assistance. Financial support for this work was received from the Danish Health Research Council (grants 12-0850-1, 12-0150-1), the Danish Veterinary and Agricultural Research Council (grants 20-3503-1), Aarhus University Research Foundation, and The Danish Biotechnology Program.

Received October 29, 1998

5 References

[1] Vilcek, J., Gray, P. W., Rinderknecht, E., Sevastopolus, C. G., *Lymphokines* 1985, *11*, 1–32.

[2] Handa, K., Suzuki, R., Matsui, H., Shimizu, Y., Kumagai, K., *J. Immunol.* 1983, *130*, 988–992.

[3] De Maeyer, E., De Maeyer, G. J., *The Cytokine Handbook*, Academic Press Limited, London 1991, pp. 215–240.

[4] Darnell, J. E., Kerr, I. M., Stark, G. R., *Science* 1994, *264*, 1415–1421.

[5] Ihle, J. N., *Cell* 1996, *84*, 331–334.

[6] Boehm, U., Klamp, T., Groot, M., Howard, J. C., *Annu. Rev. Immunol.* 1997, *15*, 749–795.

[7] Weil, J., Epstein, C. J., Epstein, L. B., Sedmak, J. J., Sabran, J. L., Grossberg, S. E., *Nature* 1983, *301*, 437–439.

[8] Celis, J. E., Justesen, J., Madsen, P. S., Lovmand, J., Ratz, G. P., Celis, A., *Leukemia* 1987, *1*, 800–813.

[9] Honore, B., Letters, H., Madsen, P., Celis, J. E., *Evr. J. Biochem.* 1993, *18*, 421–430.

[10] Walsh, B. J., Gooley, A. A., Williams, K. L., Breit, S. N., *J. Leukocyt. Biol.* 1995, *57*, 507–513.

[11] Mansbridge, J. N., Nickoleff, B. J., Morhenn, V. B., *J. Invest. Dermatol.* 1987, *88*, 602–610.

[12] Beresini, M. H., Sugarman, B. J., Shepard, H. M., Epstein, L. B., *Electrophoresis* 1990, *11*, 232–241.

[13] Epstein, L. B., Smith, D. M., Matsui, N. M., Tran, H. M., Sullivan, C., Raineri, I., Burlingame, A. L., Clauser, K. R., Hall, S. C., Andrews, L. E., *Electrophoresis* 1996, *17*, 1655–1670.

[14] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173179.

[15] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[16] Nawrocki, A., Larsen, M. R., Podtelejnikov, A. V., Jensen, O. N., Mann, M., Roepstorff, P., Görg, A., Fey, S. J., Mose-Larsen, P., *Electrophoresis* 1998, *19*, 1024–1035.

[17] Gobom, J., Nordhoff, E., Ekman, R., Roepstorff, P., *Int. J. Mass Spectrom. Ion Proc.* 1997, *169/170*, 153–163.

[18] Mann, M., Højrup, P., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 338–345.

[19] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.

[20] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[21] Jensen, O. N., Podtelejnikov, A., Mann, M., *Anal. Chem.* 1997, *69*, 4741–4750.

[22] Miele, L., Cordella-Miele, E., Facchiano, A., Mukherjee, A. B., *Nature* 1988, *335*, 726–730.

[23] Reem, G. H., Yeh, N.-H., *Science* 1984, *225*, 429–430.

[24] Gold, R., Pepinsky, R. B., Zettl, U. K., Toyka, K. V., Hartung, H. P., *Neuroimmunol.* 1996, *69*, 157–164.

[25] Chin, Y. E., Kitagawa, M., Kuida, K., Flavell, R. A., Fu, X. Y., *Mol. Cell Biol.* 1997, *17*, 5238–5337.

[26] Deiss, L. P., Galinka, H., Berissi, H., Cohen, O., Kimchi, A., *EMBO J.* 1996, *15*, 3861–3870.

[27] Flohr, T., Buwitt, U., Bonnekoh, B., Decker, T., Bottger, E. C., *Eur. J. Immunol.* 1992, *22*, 975–979.

[28] Rao, K. S., Babu, K. K., Gupta, P. D., *Cell. Biol. Internat.* 1996, *20*, 261–274.

[29] Chou, Y. H., Rosevear, E., Goldman, R. D., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 1885–1889.

[30] Realini, C., Dubiel, W, Greg, Pratt, Ferrel, K., Rechsteiner, M., *J. Biol. Chem.* 1994, *269*, 20727–20732.

[31] Flecner, J., Rasmussen, H. H., Justesen, J., *Proc. Natl. Acad. Sci. USA* 1991, *88*, 11520–11524.

[32] Flecner, J., Martensen, P. M., Tolstrup, A. B., Kjeldgaard, N. O., Justesen, J., *Cytokine* 1995, *7*, 70–77.

[33] Tolstrup, A., Bejder, A., Flecner, J., Justesen, J., *J. Biol. Chem.* 1995, *270*, 397–403.

[34] Kisselev, L., Frolova, L., Haenni, A. L., *Trends Biochem. Sci.* 1993, *18*, 263–267.

[35] Paley, E. L., *Eur. J. Biochem.* 1997, *244*, 780–788.

[36] Vartanian, A., Navrovlyansky, A., Amchenkova, A., Turpaev, K., Kisselev, L., *FEBS Lett.* 1996, *381*, 32–34.

[37] York, I. A., Rock, K. L., *Annu. Rev. Immunol.* 1996, *14*, 369–396.

[38] Yang, Y., Waters, J. B., Früh, K., Petersen, P. A., *Proc. Natl. Acad. Sci. USA* 1992, *89*, 4928–4932.

[39] Gaczynska, M., Rock, K. L., Spies, T., Goldberg, A. L., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 9213–9217.

[40] Tanaka, K., *J. Leucocyt. Biol.* 1994, *56*, 571–575.

[41] Hisamatsu, H., Shimbara, N., Saito, Y., Kristensen, P., Hendil, K. B., Fujiwara, T., Takahashi, E., Tanahshi, N., Tamura, T., Ichihara, A., Tanaka, K., *J. Exp. Med.* 1996, *183*, 1807–1816.

[42] Aki, M., Shimbara, N., Takashina, M., Akiyama, K., Kagawa, S., Tamura, T., Tanahashi, N., Yoshimura, T., Tanaka, K., Ichihara, A., *J. Biochem.* 1994, *115*, 257–259.

[43] Kim, T. K., Maniatis, T., *Science* 1996, *273*, 1717–1719.

[44] Shaw, A. C., Larsen, M. R., Roepstorff, P., Holm, A., Christiansen, G., Birkelund, S., *Electrophoresis* 1999, 977–983.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 994–1000

Kathleen M. Champion[1]
David Arnott[2]
William J. Henzel[2]
Sam Hermes[1]
Stefanie Weikert[3]
John Stults[2]
Martin Vanderlaan[1]
Lynne Krummen[3]

[1]Department of Analytical Chemistry
[2]Department of Protein Chemistry
[3]Department of Cell Culture and Fermentation R & D, Genentech, South San Francisco, CA, USA

A two-dimensional protein map of Chinese hamster ovary cells

Chinese hamster ovary (CHO) cells are used extensively for the expression of biopharmaceutical protein products. As part of our effort to better understand CHO cell physiology and protein expression changes caused by modified culture conditions, we have begun to map CHO cell polypeptides. A parental cell line reference map was established using two-dimensional gel electrophoresis with immobilized pH gradients (pH 3–10) in the first dimension and a linear acrylamide gradient (9–18%T) in the second dimension. The map is composed of over 1000 silver-stained protein spots. Protein identification is proceeding using a combination of immunostaining, NH_2-terminal sequencing, and mass spectrometric analyses. Among the proteins so far identified are glucose-regulated protein 78 (GRP78), protein disulfide isomerase (PDI), galectin-1, and several heat-shock proteins. The goal is to generate a database which emphasizes those proteins most relevant to the use of CHO cells as a host for recombinant protein expression.

Keywords: Chinese hamster ovary cells / Two-dimensional gel electrophoresis / Protein identification / *N*-terminal sequencing / Proteome
EL 3343

1 Introduction

Chinese hamster ovary (CHO) cells are the host of choice for producing complex recombinant proteins. They are able to synthesize and secrete recombinant glycoproteins with carbohydrate structures similar to those found in normal human glycoproteins [1]. This is an important consideration in biopharmaceutical glycoprotein production since the carbohydrate moiety plays a critical role in biological activity and rate of plasma clearance [2, 3]. In addition to being able to modify recombinant glycoproteins correctly, CHO cells can produce them at relatively high levels in large-scale batch cultures. Cultivation of CHO cells in large-scale serum-free cultures has been made possible through development of media, as well as adaptive growth and genetic engineering of the cells. Exclusion of serum from production medium is preferred because of its expense, variability, and potential interference with recombinant protein recovery. Serum-free media for culturing CHO cells are normally composed of a basal medium supplemented with growth factors such as insulin. The requirement for exogenous insulin has been eliminated for certain cell lines genetically engineered to produce insulin-like growth factor I (IGF-1)I [4, 5]. Added growth factor requirements have also been bypassed in CHO cells expressing cloned transcription factor E2F-1 [6]. The effect of constitutive E2F-1 expression on intracellular protein levels was monitored using two-dimensional gel electrophoresis (2-DE), which showed that cyclin A was among the proteins up-regulated. These studies utilized a 2-DE reference map of CHO cellular proteins from cultures containing 10% fetal calf serum; the map contains some eight identified proteins [7]. Growth of CHO cells in culture has also been manipulated through reduction of temperature, which prolongs cell viability by delaying the onset of apoptosis [8]. It was shown that cells accumulate in the G1 phase of the cell cycle with an accompanying alteration in nucleotide pools.

Because it allows for the routine separation of 1000–2000 protein spots simultaneously, 2-DE is an appropriate technique for the analysis of multigene pathways controlling such complex cellular functions as cell cycle progression and protein export. Here we present a 2-DE map of a parental CHO cell line as a model for recombinant production cell lines. Several proteins that play a role in polypeptide folding, processing, and secretion have been identified.

Correspondence: Dr. Kathleen M. Champion, Department of Analytical Chemistry, Genentech, 1 DNA Way, MS #62, South San Francisco, CA, 94080, USA
E-mail: kmc@gene.com
Fax: +650-225-3554

Abbreviations: CHO, Chinese hamster ovary; **CID,** collision-induced dissociation; **GRP,** glucose-regulated protein; **MIF,** macrophage migration inhibitory factor; **PDI,** protein disulfide isomerase; **PPIase,** peptidyl prolyl *cis-trans* isomerase; **TOF,** time of flight

2 Materials and methods

2.1 Materials

Acrylamide was purchased as a 40% w/v solution with *N,N'*-methylenebisacrylamide (37.5:1) from Amresco

(Solon, OH). Carrier ampholytes (Pharmalyte pH 3–10) were from Amersham Pharmacia Biotech (Piscataway, NJ) and pH 8–10 carrier ampholytes (Biolyte 8–10) were from Bio-Rad (Hercules, CA). *N,N,N′, N′*-tetramethylethylenediamine (TEMED) and ammonium persulfate were obtained from Sigma (St. Louis, MO). Urea and agarose were from Bio-Rad. The following primary antibodies were used: anti-actin (Sigma), anti-Hsp47 monoclonal clone M16.10A1 (Stressgen, Victoria, BC, Canada), anti-Hsc70 polyclonal from rabbit (Stressgen), anti-calnexin polyclonal from rabbit (Stressgen), and anti-Grp94 polyclonal from rabbit (Stressgen). Secondary antibodies used in Western blotting were anti-rabbit IgG (whole molecule) peroxidase-conjugated from goat (Sigma) and anti-mouse IgG (whole molecule) peroxidase-conjugated from goat (Sigma).

2.2 Cell culture and sample preparation

CHO cells were derived from a homozygous dihydrofolate reductase minus (dhfr–) DUKX CHO host [9]. Cells were cultured in 2 L bioreactors (Aplicon, Foster City, CA) in a serum-free DMEM/F-12 based medium supplemented with protein hydrolysate and a proprietary nutrient mixture. The cultures were inoculated at 1.0×10^6 cells/mL and maintained at 37°C throughout the incubation. Glucose levels in the medium were maintained at greater than 3 g/L by a single glucose addition (5 g of glucose per liter of medium) on day 3. Dissolved oxygen was controlled on-line (60% of saturation) by sparging with oxygen and/or air. The pH was maintained at 7.2 by addition of Na_2CO_3 or CO_2. Cell viability was evaluated by measuring total and supernatant lactate dehydrogenase [10] and by trypan blue exclusion. Samples were collected during exponential growth when viability was at least 95 %. Cells were harvested by centrifugation ($1000 \times g$) for 10 min and washed with phosphate-buffered saline. Five $\times 10^7$ cells were solubilized directly in 1.25 mL of a solution composed of 9.0 M urea, 4% NP-40, 1% DTT, 4% Biolyte 8–10, pH 9.5. After homogenization with a Dounce device (20–30 strokes), the samples were centrifuged ($10\,000 \times g$) for 10 min. The supernatants were stored at –80°C until analysis. Protein was determined by a modified assay [11] using bovine serum albumin as standard.

2.3 Isoelectric focusing

For analytical gels, the equivalent of 1.5×10^5 cells (~ 120 μg protein) was combined with rehydration solution (8 M urea, 2% v/v CHAPS, 15 mM DTT, 0.8% Pharmalyte (pH 3–10) ampholytes, and 0.020 mg/mL Orange G dye) in a total volume of 380 μL. Eighteen cm pH 3–10 L immobilized pH gradient (IPG) gel strips (Amersham Pharmacia Biotech) were rehydrated overnight by placing the strips gel-side-down in sample-containing rehydration solution in an Immobiline DryStrip Reswelling Tray (Amersham Pharmacia Biotech) and covering with mineral oil. IEF was conducted using a Multiphor II unit with a DryStrip Kit (Amersham Pharmacia Biotech) at 20°C. For both analytical and preparative gels, the voltage was increased stepwise from 150 V (2 h) to 300 V (3 h), then to 1500 V (1 h), and finally to 3500 V, where it was maintained by the EPS 3500 XL power supply (Amersham Pharmacia Biotech) until 65 000 Vh was achieved.

2.4 Equilibration

Prior to second-dimensional SDS-PAGE, the IPG gel strips were washed for 10 min in equilibration solution (50 mM Tris-HCl, pH 8.5, 6 M urea, 2% w/v SDS, 30% w/v glycerol) containing 1% w/v DTT. This was followed by a 10 min wash in equilibration solution containing 5% w/v iodoacetamide. IPG gel strips were sealed into place on second-dimensional slab gels by applying 2 mL of molten (70°C) 0.5% w/v agarose in tank buffer containing 0.5% w/v SDS.

2.5 SDS-PAGE

Second-dimensional SDS-PAGE was carried out using the DALT system (Amersham Pharmacia Biotech) essentially as described by Anderson and Anderson [12]. Slab gels ($200 \times 250 \times 1.5$ mm) were poured using the Angelique computer-controlled gradient maker (Large Scale Biology, Rockville, MD) according to the manufacturer's instructions. Gels were cast so that the upper 5% was 9%T acrylamide and the remainder of the gel varied linearly from 9–18%T, 2.6%C, using bisacrylamide as the cross-linker. The tank buffer was composed of 200 mM glycine, 24 mM Tris, 0.1% w/v SDS, pH 8.6. For molecular weight calibration, 10 kDa protein ladder (Gibco BRL, Gaithersburg, MD) was combined with molten agarose and applied to the top of the slab gel adjacent to the IPG gel strip. Gels were run at 5°C overnight with voltage and current limits of 150 V and 600 mA, respectively. Upon completion of SDS-PAGE, the gels were fixed and silver-stained by a modification of the method described [13]. For preparative gels, approximately 350 μg of protein were loaded. Gels were stained with 0.2% Coomassie Brilliant Blue R-250 (Bio-Rad) in 2.5% v/v phosphoric acid, 50% ethanol. Destaining took place in several washes of 20% ethanol.

2.6 Protein sequencing

Preparatively loaded gels were blotted to polyvinylidene difluoride (PVDF) membranes (ProBlott; Applied Biosys-

tems, Foster City, CA) using a Semi-Phor Model TE77 semidry blotter by a modification of the method described [14]. Following blotting, proteins were detected by staining with 0.1% Coomassie Blue R-250 in 40% methanol, 1.0% acetic acid for 2 min and destaining in several washes of 50% methanol. The PVDF membranes were rinsed extensively with deionized water and permitted to air-dry. Protein spots of interest were excised from membranes and sequenced on models 494 CL and 494HT PE Applied Biosystems sequencers using 6 mm micro cartridges and equipped with an on-line capillary phenylthiohydantoin analyzer (model 140D). Peaks were integrated with Justice Innovation software using Nelson Analytical 760 interfaces. Sequence interpretation was performed on a DEC Alpha [15].

2.7 Cyanogen bromide cleavage

Proteins that were determined to be blocked were subjected to cleavage using cyanogen bromide. PVDF membranes were removed from the sequencer and wetted with 1 μL of methanol followed by 20 μL of 0.1 N HCl and a few crystals of CNBr. The reaction was carried out for 3 h at 45°C [16].

2.8 *In situ* digestion of proteins

Protein spots were excised from preparatively loaded gels and placed in siliconized sample tubes. *In situ* digests were performed using an ABiMED DigestPro robotics workstation employing a program optimized for sample recovery. Gel fragments were washed with 50% acetonitrile in 100 mM NH_4HCO_3 for 15 min and rinsed twice with acetonitrile (60 μL and 100 μL) for 10 min. Reduction of disulfide bonds was carried out in 30 μL of 50 mM DTT, 100 mM NH_4HCO_3 for 50 min at 50°C. Free sulfhydryls were then alkylated by incubation in 30 μL of 50 mM iodoacetamide, 100 mM NH_4HCO_3 for 15 min. After washing for 15 min with 30 μL of bicarbonate buffer, the gel fragments were dehydrated by two 20 min incubations with 100 μL acetonitrile. The gel fragments were then rehydrated with digestion buffer composed of 0.04 mg/mL of modified trypsin (Promega, Madison, WI) in 0.005 % zwittergent 3–16, 50 mM NH_4HCO_3. After 15 min the excess trypsin solution was removed and the volume replaced with 20 μL of 50 mM NH_4HCO_3. Trypsin digestion was allowed to proceed for 3 h at 37°C. Subsequent to digestion, the buffer was removed and gel fragments were extracted by addition of 15 μL of 0.1% TFA (30 min) and 10 μL of acetonitrile. This was followed by an additional extraction with 15 μL of acetonitrile for 25 min. The resulting digest mixtures were concentrated and desalted with a 1 × 10 mm reverse-phase trapping cartridge (Michrom BioResources, Auburn, CA). The digest extracts were diluted to 500 μL with 0.1% TFA and passed over the trapping cartridge. After washing with 50 μL of 0.1% TFA, peptides were eluted into a 0.5 mL microcentrifuge tube with 75% acetonitrile, 0.1% TFA in a volume of approximately 5 μL.

2.9 MALDI-TOF mass spectrometry

Samples were loaded as described previously [17, 18] in 0.5 μL aliquots directly to a MALDI target using a fast evaporation matrix, α-cyano-4-hydroxy-*trans*-cinnamic acid/nitrocellulose wetted with 0.5 μL of 0.1% TFA. Mass spectra were obtained using a Voyager DE-STR mass spectrometer (PerSeptive Biosystems, Framingham, MA) in reflectron mode, and with delayed extraction. External calibration was performed using Des-Arg1-bradykinin and ACTH (clip 18–31) standards adjacent to the samples.

2.10 Capillary LC-MS

Peptide mixtures were delivered to the microelectrospray ionization source of an ion trap mass spectrometer (LCQ; Finnigan MAT, San Jose, CA) by capillary HPLC. Reverse-phase microcapillary columns (15 cm × 100 mm ID) were constructed as described [17]. Solvents (solvent A: 2% acetonitrile, 0.1 M acetic acid; solvent B: 90% acetonitrile, 0.1 M acetic acid) were delivered by a model 140B dual syringe pump (Applied Biosystems). Flow rates of 100–400 μL/min from the pump were attenuated by a preinjector Tee union with a split ratio of approximately 130. Samples were loaded *via* 20 μL injection loop and washed onto the column at 3 μL/min with initial conditions of 2% solvent B. After loading, the sample loop was switched out of line with the column, and the flow rate was reduced to 0.75 μL per min (solvent was diverted to waste during this reequilibration). Peptides were eluted with a binary gradient of 2–30% solvent B in 10 min; 30–90% B in 5 min. The entire effluent from the column was sprayed from a liquid-junction microelectrospray ionization source described elsewhere [19]. Tandem mass spectrometry (MS/MS) was performed with automated data acquisition using the "triple play" experiment on the ion trap, in which ions detected in full mass range spectra were scanned with high resolution in a "zoom scan" and then subjected to collision-induced dissociation (CID). Automatic gain control was used in all experiments, with a maximum ionization time of 500 ms. CID was carried out using a 3 Da window, and a collision energy of 25% of maximum.

2.11 Protein identification using database searching algorithms

Gel-isolated proteins were identified by searching a protein sequence database using tryptic peptide molecular

masses either alone or in conjunction with CID spectra. The database searched was a compilation of several databases including SWISS-PROT, the Protein Identification Resource, and a translation of GenBank. The FRAGFIT program, developed at Genentech, was used to search the database with peptide masses determined by MALDI-TOF [20, 21]. CID spectra from LC-MS/MS experiments were matched against the database using the SEQUEST program developed by Eng and Yates [22].

2.12 Immunoblotting

Proteins were electrophoretically tranferred from 2-DE gels to PVDF membranes (Immobilon-P; Millipore, Bedford, MA) as described above (Section 2.6). The membranes were blocked with 3 % w/v BSA in Tris-buffered saline (10 mM Tris-HCl, pH 7.5, in 150 mM NaCl) and incubated with primary antibody at room temperature. Peroxidase-conjugated secondary antibody was used at 1:1000, and protein-antibody complexes were visualized using 4-chloro-1-naphthol in the color-forming reaction.

3 Results and discussion

Figure 1 shows the silver-stained 2-DE reference map of CHO cellular proteins in the pH range of 3–10. The cells were cultured under serum-free conditions similar to those used for recombinant protein production. The pro-

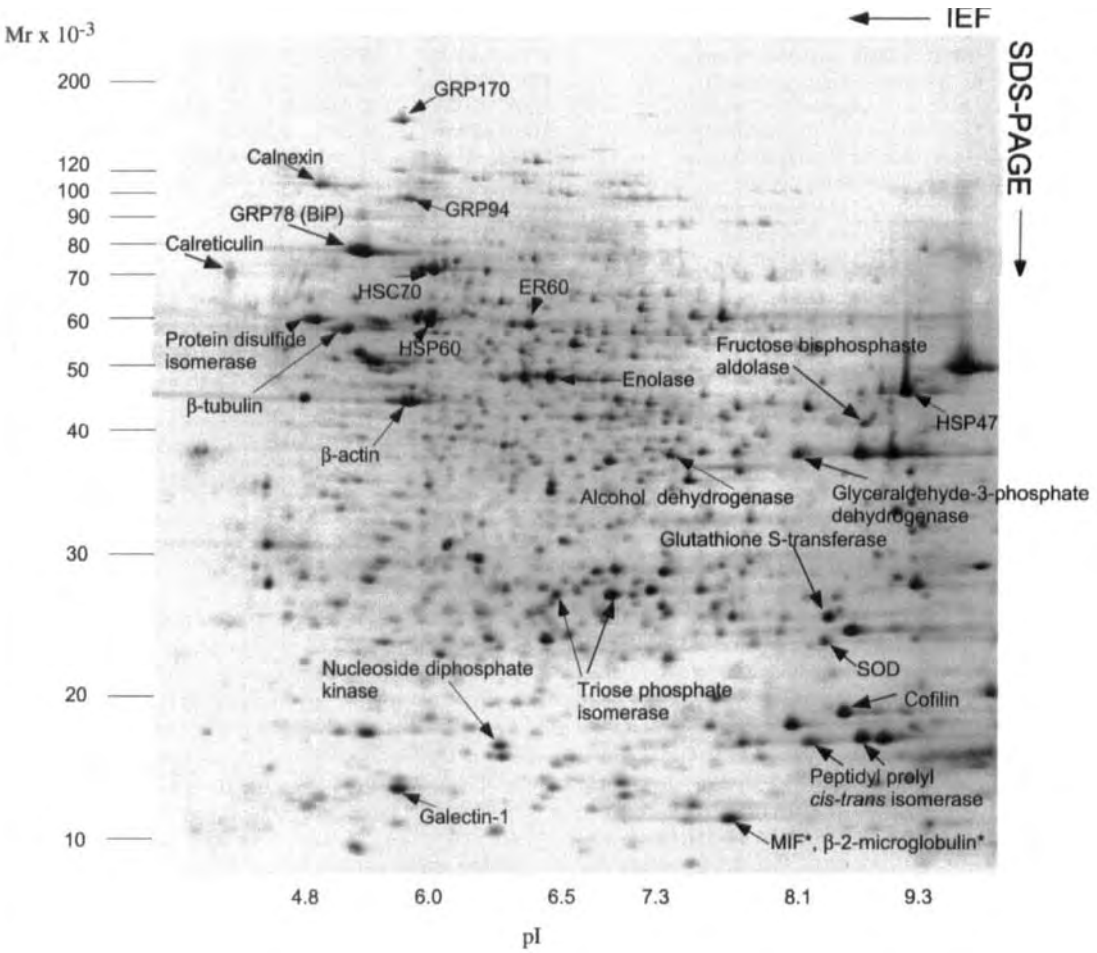


Figure 1. Silver-stained 2-DE pattern of Chinese hamster ovary cellular proteins. Approximate molecular weights are shown on the left. Approximate p*I*s are shown along the bottom. Proteins identified by primary structure analysis or immunoblotting are indicated. *Macrophage migration inhibitory factor (MIF) and β_2-microglobulin were found to comigrate on gels, as demonstrated by NH_2-terminal sequencing reactions in which the sequence mixture was deciphered.

Table 1. CHO cellular proteins (labeled in Fig. 1) identified as indicated[a)]

Gene name	Description	Accession No.	M_r	p*I*	Method(s) of identification
ACT	β-Actin	MMACTBR2_1	39 185	6.1	IB, MALDI
ALR	Alcohol dehydrogenase	ALDX_RAT	36 370	7.3	MS/MS
ALDA	Fructose biosphosphate aldolase a	ALFA_HUMAN	39 289	8.5	NS
B2M	β_2-Microglobulin	B2MG_RAT	11 636	7.6	NS
CANX	Calnexin, calcium-binding protein p93	CALX_MOUSE	67 254	4.5	IB
CALR	Calreticulin	CRTC_RAT	47 995	4.3	NS, MALDI
CFL	Cofilin, actin-depolymerizing factor	COF1_HUMAN	18 532	8.9	NS[b)]
ENO1	α-Enolase	ENOA_RAT	46 980	6.5	MALDI
Erp60	Protein disulfide isomerase ER60	ER60_HUMAN	56 800	6.3	MALDI and MS/MS
LGALS	Galectin-1, Chinese hamster	LEG1_CRIGR	14 802	5.7	MALDI
GRP78	GRP78, 78 kDa glucose-regulated protein	P_R53075	72 380	5.1	MALDI
GRP94	Glucose-regulated protein 94, HSP100	Y09136	92 514	4.8	IB
GRP170	Glucose-regulated protein 170	S68689	111 271	5.1	MALDI
GST	Glutathione *S*-transferase p	GTP_CRILO	23 507	8.1	NS
GAPD	Glyceraldehyde-3-phosphate dehydrogenase	G3P_MOUSE	35 679	8.6	NS
HSP47	GP46, colligin, serpin superfamily	HS47_MOUSE	46 589	9.3	IB, NS
HSP60	Mitochondrial matrix protein p1	P60_CRIGR	60 990	5.9	MALDI, NS
HSC70	Heat shock cognate 71 kDa protein	HS7C_CRIGR	70 804	6.1	IB, MALDI
MIF	Macrophage migration inhibitory factor	MIF_HUMAN	12 345	8.2	NS
NDK	Nucleoside diphosphate kinase a	NDKA_MOUSE	17 210	7.5	MALDI
PDI	Protein disulfide isomerase	PDI_HUMAN	55 294	4.7	MALDI
PPIA	Peptidyl-prolyl *cis-trans* isomerase a	CYPH_HUMAN	17 881	8.2	NS
SOD2	Superoxide dismutase (SOD)	SODM_HUMAN	22 204	7.3	NS
TPI	Triose phosphate isomerase	TRIS_HUMAN	26 538	6.5	NS
TUB	Tubulin β-chain	S18456	49 741	4.9	MALDI

a) Theoretical M_r and p*I* values are listed.
Protein analysis methods are IB, immunoblot; MALDI, peptide mass fingerprinting using MALDI-TOF analysis; MS/MS, tandem mass spectrometry; and NS, NH_2-terminal sequencing.

b) CNBr cleavage fragment was sequenced.

teins identified in this map (Table 1) represent several cellular compartments, including the endoplasmic reticulum (ER), cytosol, and mitochondria. The resident ER polypeptides which have been identified include the following: glucose-regulated protein 78 (GRP78), GRP94, GRP170, heat-shock protein 47 (HSP47), protein disulfide isomerase (PDI), ER60, calnexin, and calreticulin. All of these proteins have been implicated in the early folding of nascent secretory polypeptides [23, 24]. One of the most dominant ER proteins in the pattern is GRP78, also known as immunoglobulin binding protein or BiP [25]. We identified this protein through peptide mass fingerprinting using a list of twenty peptide masses to search the protein sequence database. Seventeen of the experimentally obtained peptide masses matched theoretical masses for this highly conserved protein, and two of the remaining peptide masses could be accounted for as contaminating keratin peaks. Therefore, the match for GRP78 was made with a high degree of confidence, supported by its 2-DE map position with expected p*I* and M_r. The corresponding protein in a previously published 2-DE map of CHO cellular proteins was assigned as HSP90 on the basis of immunoblotting results [7], and is probably an inappropriate protein spot assignment.

A number of cytosolic proteins have been identified, including two that play a role in proper protein folding: HSC70 and peptidyl prolyl *cis-trans* isomerase (PPIase). The identity of HSC70, established in our map through peptide mass fingerprinting, was confirmed by Western analysis using a primary antibody which specifically recognizes the constitutive form of the protein (not shown). Our designation of HSC70 is in good agreement with the 2-DE map published earlier [7]. Through NH_2-terminal sequencing we have identified two isoforms of PPIase, an enzyme which catalyzes the *cis-trans* isomerization of proline imidic peptide bonds in oligopeptides and has been demonstrated to expedite refolding of several proteins *in vitro* [26, 27]. PPIase is identical to cyclophilin, the major intracellular receptor for cyclosporin A, an immunosuppressive drug [28, 29]. Other cytosolic proteins whose identities are known include nucleoside diphosphate kinase, alcohol dehydrogenase, glutathione-*S*-transferase, and galectin-1. Galectin-1 has been analyzed on

2-DE gels previously [30] and shown to be expressed in CHO cells [31]. Although galectin-1, a β-galactoside-binding lectin, was initially isolated from cytosolic fractions, its localization to the cell surface has also been demonstrated [32]. Four enzymes which catalyze key reactions in glycolysis have been identified: fructose bisphosphate aldolase, triose phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, and α-enolase.

We identified macrophage migration inhibitory factor (MIF) in our CHO cell 2-DE map through NH_2-terminal amino acid sequencing and alignment with the human MIF sequence. MIF, the first known lymphokine, is expressed in many cell types [33, 34] and is involved in numerous immunological and hormonal processes, including insulin resistance in adipose tissue [35, 36]. During NH_2-terminal sequencing of MIF, a sequence mixture with a minor component was observed. This indicated that an additional protein, lower in abundance than MIF, comigrated on the gel. Upon further analysis through 21 cycles of Edman degradation sequencing and subsequent database matching, β_2-microglobulin was identified as the polypeptide unresolved from MIF. Their comigration on gels is not unexpected, based upon the expected M_r (11 363) and p*I* (7.6) of mature β_2-microglobulin and the observed migration of the major isoform of MIF from several cell types [34]. The presence of multiple proteins comigrating as single spots on 2-DE gels has been clearly demonstrated on large format gels by Klose and Kobalz [37]. Our results demonstrate that it is possible to simultaneously determine the NH_2-terminal sequences of a mixture of two proteins from a single spot on a 2-DE gel by employing high sensitivity techniques. Sequence mixtures were also observed while determining the NH_2-terminal sequences of triosephosphate isomerase and PPIase. The minor, contaminating species remain to be identified. Two mitochondrial proteins identified through NH_2-terminal sequencing are the molecular chaperone, Hsp60, and superoxide dismutase (SOD). So far we have found approximately 40% NH_2-terminal blockage in our sequencing efforts.

4 Concluding remarks

We have established a CHO cellular protein map using several techniques to identify over twenty proteins. During Edman degradation sequencing we were able to concomitantly determine the sequences of two polypeptides present in a single spot on a gel. The CHO reference map will be used as a tool in investigating physiological responses to modifications in culture condition, cell line improvement, and culture supernatant protein content. Our mapping efforts will continue with an emphasis on proteins that potentially affect growth of CHO cells and expression of recombinant proteins in CHO cells.

Received October 23, 1998

5 References

[1] Goochee, C. F., Gramer, M. J., Andersen, D. C., Bahr, J. C., Rasmussen, J. R., in: Todd, P., Subhas, K., Sikdar, K., Bier, M. (Eds.), *Frontiers in Bioprocessing II*, American Chemical Society, Washington DC 1992, pp. 198–240.

[2] Ashwell, G., Hartford, J., *Annu. Rev. Biochem.* 1982, *51*, 531–554.

[3] Jenkins, N., Curling, E. M. A., *Enzyme Microbiol. Technol.* 1994, *16*, 354–365.

[4] Pak, S. C. O., Hunt, S. M. N., Bridges, M. W., Sleigh, M. J., Gray, P. P., *Cytotechnology* 1996, *22*, 139–146.

[5] Hunt, S. M. N., Pak, S. C. O., Bridges, M. W., Gray, P. P., Sleigh, M. J., *Cytotechnology* 1997, *24*, 55–64.

[6] Lee, K. H., Sburlati, A., Renner, W. A., Bailey, J. E., *Biotech. Bioengineer.* 1996a, *50*, 273–279.

[7] Lee, K. H., Harrington, M. G., Bailey, J. E., *Biotech. Bioengineer.* 1996b, *50*, 336–340.

[8] Moore, A., Mercer, J., Dutina, G., Donahue, C. J., Bauer, K. D., Mather, J. P., Etcheverry, T., Ryll, T., *Cytotechnology* 1997, *23*, 47–54.

[9] Urlaub, G., Chasin, L. A., *Proc. Natl. Acad. Sci. USA* 1980, *77*, 4216–4220.

[10] Goergen, J. L., Marc, A., Engasser, J. M., *Cytotechnology* 1993, *11*, 189–195.

[11] Ramagli, L. S., Rodriguez, L. V., *Electrophoresis* 1985, *6*, 559–563.

[12] Anderson, N. G., Anderson, N. L., *Anal. Biochem.* 1978, *885*, 331–340.

[13] Guevara, J., Capetillo, S., Johnston, D. A., Martin, B. A., Ramagli, L. S., Rodriguez, L. V., *Electrophoresis* 1982, *3*, 197–205.

[14] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[15] Henzel, W. J., Rodriguez, H., Watanabe, C., *J. Chromatogr.* 1987, *404*, 41–52.

[16] Zalut, C., Henzel, W. J., Harris, W. H., *J. Biochem. Biophys. Methods* 1980, *3*, 11–30.

[17] Arnott, D., Henzel, W. J., Stults, J. T., *Electrophoresis* 1998, *19*, 968–980.

[18] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[19] Arnott, D., O'Connell, K. L., King, K. L., Stults, J. T., *Anal. Biochem.* 1997, *258*, 1–18.

[20] O'Connell, K. L., Stults, J. T., *Electrophoresis* 1997, *18*, 349–359.

[21] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

[22] Eng, J. K., McCormack, A. L., Yates III, J. R., *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989.

[23] Tatu, U., Helenius, A., *J. Cell Biol.* 1997, *136*, 555–565.

[24] Dierks, T., Volkmer, J., Schlenstedt, G., Jung, C., Sandholzer, U., Zachmann, K., Schlotterhose, P., Neifer, K., Schmidt, B., Zimmermann, R., *EMBO J.* 1996, *15*, 6931–6942.

[25] Munro, S., Pelham, H. R. B., *Cell* 1986, *46*, 291–300.

[26] Lang, K., Schmid, F. X., Fischer, G., *Nature* 1987, *329*, 268–270.

[27] Lang, K., Schmid, F. X., *Nature* 1988, *331*, 453–455.

[28] Takahashi, N.,. Hayano, T., Suzuki, M., *Nature* 1989, *337*, 473–475.

[29] Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., Schmid, F. X., *Nature* 1989, *337*, 476–478.

[30] Lutomski, D., Caron, M., Cornillot, J.-D., Bourin, P., Dupuy, C., Pontet, M., Bladier, D., Joubert-Caron, R., *Electrophoresis* 1996, *17*, 600–606.

[31] Cho, M., Cummings, R. D., *J. Biol. Chem.* 1995, *270*, 5198–5206.

[32] Cho, M., Cummings, R. D., *J. Biol. Chem.* 1995, *270*, 5207–5212.

[33] Tanaka, M., Maeda, K., Nakashima, K., *J. Biochem.* 1995, *117*, 554–559.

[34] Magi, B., Bini, L., Liberatori, S., Marzocchi, B., Raggiaschi, R., Arcuri, F., Tripodi, S. A., Cintorino, M., Tosi, P., Pallini, V., *Electrophoresis* 1998, *19*, 2010–2013.

[35] Wistow, G. J., Shaughnessy, M. P., Lee, D. C., Hodin, J., Zelenka, P. S., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 1271–1275.

[36] Waeber, G., Calandra, T., Rodui, R., Haefliger, J.-A., Bonny, C., Thompson, N., Thorens, B., Temler, E., Meinhardt, A., Bacher, M., Metz, C. N., Nicod, P., Bucala, R., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 4782–4787.

[37] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1001–1010

Aida Pitarch[1]
Mercedes Pardo[1]
Antonio Jiménez[2]
Jesús Pla[1]
Concha Gil[1]
Miguel Sánchez[1]
César Nombela[1]

[1]Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain
[2]Departamento de Medicina, Facultad de Medicina, Universidad de Salamanca, Spain

Two-dimensional gel electrophoresis as analytical tool for identifying *Candida albicans* immunogenic proteins

This paper reports the usefulness of two-dimensional gel electrophoresis followed by Western blotting with sera from patients with systemic candidiasis in the identification of the major *Candida albicans* antigens. In order to have different patterns of protein expression and subcellular localization, three types of protein preparations were obtained: cytoplasmic extracts, protoplast lysates and proteins secreted by protoplasts regenerating their cell wall. These proteins were separated by high-resolution two-dimensional electrophoresis using an immobilized pH gradient. Western blotting with sera from patients with systemic candidiasis allowed the detection of more than 18 immunoreactive proteins. Some of these proteins had different isoforms. All sera reacted with at least three *C. albicans* proteins and the most reactive serum detected up to eleven proteins. Some of these antigens, *e.g.*, enolase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), have been identified on the 2-D map. The most reactive proteins were enolase and a 34 kDa protein in the acidic part of the gel (p*I* 4–4.4) that was only detected in regenerating protoplast-secreted proteins. The identification of all these antigens would be useful for the development of diagnostic strategies.

Keywords: *Candida albicans* / Proteins / Two-dimensional polyacrylamide gel electrophoresis / Antigens / Immune response / Candidiasis / Immune sera EL 3360

1 Introduction

The dimorphic fungus *Candida albicans* is both a commensal and a human pathogen and is the most common cause of opportunistic fungal diseases. *C. albicans* infections have increased dramatically during the last two decades due to several predisposing factors such as immunosuppressive, cytotoxic and antibiotic treatments, long-term catheterization and longer survival of immunologically compromised individuals. Candidiasis can take many forms, ranging from asymptomatic colonization, through superficial mucocutaneous infections, to disseminated systemic disease, often with multiple organ involvement [1–3]. To understand the pathogenicity and clinical impact of *C. albicans*, it is important to have knowledge of the host-fungus interaction. Host defense mechanisms against *C. albicans* infections are highly complex and several studies have demonstrated the essential role of phagocytic cells, cell-mediated immunity, and even humoral immune responses in the resolution of *C. albicans* infections [4–6]. The importance of the cellular or humoral immune response seems to depend on the type of infection. The high incidence of mucosal candidiasis in individuals with deficiencies in cell-mediated immunity (CMI) emphasizes the important role of CMI in protection against *C. albicans* mucosal infections [7]. By contrast, antibodies probably play an important part in the defense of the host against disseminated candidiasis [8]. Despite this, the role of antibodies in resistance to candidiasis is still poorly understood and deeper insight into the humoral immune response to *C. albicans* would be useful not only for the development of diagnostic strategies but also for alternative forms of treatment of candidiasis [4]. Several immunogenic *C. albicans* proteins have already been identified: heat shock proteins (Hsp90, Hsp70), enolase, glycolytic enzymes and aspartyl proteinases [9–16]. All these works are analyzed in an excellent review [17]. Another important issue is the diagnosis of systemic candidiasis. Immunological tests are lacking in sufficient specificity and/or sensitivity, and do not permit a distinction between colonization and disseminated infection [18–20]. It would therefore be of great use to have simple and sensitive tests. In light of all the above, the detection and identification of *C. albicans* immunogenic proteins would be valuable.

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) appears to be a useful tool for the analysis of complex mixtures of proteins. This technique, originally described by O'Farrell [21] separates proteins on the basis of their differences in isoelectric point (p*I*) in the first dimension and apparent molecular mass in the second

Correspondence: Dr. Miguel Sánchez, Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense, 28040-Madrid, Spain
E-mail: gilmifa@eucmax.sim.ucm.es
Fax: +34-91-3941745

dimension. Over the last few years, with the advent of commercially manufactured immobilized pH gradients (IPGs) [22] this methodology has been greatly improved. After separation, protein identification can be achieved in many ways, such as through specific antibodies, by comparison of two-dimensional gel positions, Edman degradation, analysis of amino acid composition and, more recently, mass spectrometry. Owing to the need to identify *C. albicans* antigens, we have combined here 2-D PAGE with Western blotting with sera from patients with systemic candidiasis. With this approach, many immunoreactive proteins were detected and some were identified. In the present report, we show the usefulness of this integrated strategy for the identification of *C. albicans* antigens.

2 Materials and methods

2.1 Organism, growth conditions and sample preparation

C. albicans strain SC5314 (a clinical isolate) [23] was used because it is a reference strain used by many researchers and most virulence studies have been carried out with disruptant mutants in this genetic background [24]. This strain was maintained on solid YED medium (1% w/v D-glucose, 1% w/v Difco yeast extract and 2% w/v agar) and incubated at 30°C for at least 2 days. Three different types of samples were obtained: cytoplasmic extracts, protoplast lysates, and proteins secreted by protoplasts.

2.1.1 Cytoplasmic extracts

C. albicans cells were grown in liquid YED medium at 30°C up to an optical density of 4 at 600 nm. Yeast cells (10^9) were pelleted by centrifugation and washed once with ice-cold water. The cell pellet was resuspended in 200 μL of lysis buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 0.5 mM PMSF and 5 μg/mL of leupeptin, pepstatin and antipain). An equal volume of 0.5 mm diameter glass beads was added to the cell suspension. Cells were vortexed for 1 min and cooled on ice for 1 min. This procedure was repeated until at least 80% of the cells had been lysed as determined by phase-contrast microscopic examination. After the beads had settled, the cell extracts were transferred to other tubes and centrifuged for 15 min at 13 000 rpm. Supernatants were again transferred to other tubes, centrifuged once again and stored at –20°C.

2.1.2 Protoplast lysates

Cells were grown in liquid YED medium at 28°C up to an optical density of 4 at 600 nm, harvested and washed once with water. Cells ($1–2 \times 10^9$ cells/mL) were then incubated in a pretreatment solution (10 mM Tris-HCl, pH 9, 5 mM EDTA, 1% v/v 2-mercaptoethanol) at 28°C with shaking at 80 rpm for 30 min. After washing with 1 M sorbitol, they were resuspended in 1 M sorbitol up to a density of 5×10^8 cells/mL. Thirty μL/mL Glusulase® (Du Pont) was added. Cells were then incubated with gentle shaking until more than 90 were protoplasts [25], which were then pelleted by gentle centrifugation (600 × *g*). Protoplasts were washed three times with 1 M sorbitol by gently swirling liquid across the surface of the pellet, to eliminate any trace of Glusulase. The cell pellet was resuspended in 200 μL of lysis buffer, vortexed for 1 min and cooled on ice for 1 min. This suspension was centrifuged for 15 min at 13 000 rpm. Supernatants were again transferred to other tubes and centrifuged. The supernatants were kept at –20°C.

2.1.3 Proteins secreted by protoplasts

Proteins secreted by protoplasts were obtained as previously described [26]. Briefly, protoplasts obtained as above were induced to regenerate the cell walls in Lee medium [27] containing 1 M sorbitol as osmotic stabilizer at 28°C, with gentle shaking. After a 2 h incubation, cells were pelleted at 600 × *g* and the medium was filtered through a 0.22 μm pore size after the addition of protease inhibitors: 0.1 mM phenylmethylsulfonyl fluoride (PMSF; Fluka, Buchs, Switzerland), 2 μg/mL leupetin and antipain, and 1 μg/mL pepstatin A (Sigma, St. Louis, MO, USA). The medium was concentrated by ultrafiltration using a pore size of 10 000 Da (Diaflo; Amicon, Witten, Germany) diluted with water, concentrated a further three times and lyophilized. Cell lysis was controlled by quantitative determination of alkaline phosphatase [28]. Protein quantitation was performed by the Bradford assay [29].

2.2 Two-dimensional gel electrophoresis

Two-dimensional electrophoresis was performed using an IPG system, as described by Bjellqvist *et al.* [30]. Immobiline Dry Strips (pH 3–10 NL, 18 cm long; Pharmacia, Uppsala, Sweden) were rehydrated overnight in 8 M urea, 2% w/v CHAPS, 10 mM dithioerythritol (DTE), 0.8% v/v Resolytes (Pharmalyte pH 3–10; Pharmacia) and a trace of bromophenol blue. Samples (approximately 200 μg of protein) were resuspended in sample buffer (8 M urea, 4% w/v CHAPS, 40 mM Tris, 65 mM DTE, bromophenol blue) and loaded for the first dimension. Isoelectric focusing was carried out on a Multiphor II electrophoresis unit (Pharmacia) at 15°C with the following program: 0–500 V in 1 min, 500 V for 5 h, 500–3500 V in 5 h, and 3500 V for 12.5 h (total 56250 Vh). After focusing, IPG strips were

equilibrated for 12 min in 50 mM Tris-HCl, pH 6.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, 2% w/v DTE, and for 5 min in 50 mM Tris-HCl, pH 6.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, 2.5% w/v iodoacetamide, and a trace of bromophenol blue, and were loaded onto a slab gel with a 0.5% w/v agarose solution in Laemmli buffer [31]. The second-dimensional run (SDS-PAGE) was carried out on homogeneous running gels (10%T) without a stacking gel. Piperazine diacrylamide (1.6%C) was used as cross-linker. Electrophoresis was conducted at 40 mA/gel constant current for 5 h in a Protean II cell (Bio-Rad, Richmond, CA, USA). Analytical gels were silver-stained according to Merril *et al.* [32] with the Bio-Rad Silver Stain Kit. Electropherogram images were obtained with the GS-690 Imaging Densitometer and processed with the Melanie II software (Bio-Rad). M_r and p*I* values were estimated using internal 2-D SDS-PAGE standards (Bio-Rad).

2.3 Human sera samples

Six sera from patients with culture-confirmed disseminated *C. albicans* infection were used. Five of the patients had hematologic tumors: leukemias (sera number 2, 3 and 5) or mielodysplastic syndromes (sera number 1 and 6). One serum (number 4) was from a patient with a solid tumor. Four sera from healthy subjects were used as controls.

2.4 Immunoblot analysis

After the samples had been subjected to analytical 2-D PAGE, the gels were electroblotted onto nitrocellulose membranes in Towbin buffer at 50 mA overnight [33]. Blots were processed following standard protocols [34]. Monoclonal antibody A2C7 (ATCC) against *C. albicans* enolase (Eno; dilution 1:6000) was used for immunodetection. Polyclonal antibodies against *C. albicans* Hsp70 (dilution 1:10 000), *C. albicans* endoglucanase (Bgl2; dilution 1:5000) and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase (GAPDH; dilution 1:10 000) were used for immunodetection. Human sera were diluted 1:50 or 1:100. Immunoreactive spots were detected using horseradish peroxidase-labeled anti-mouse, anti-rabbit or anti-human IgGs (depending on the first antibody used) and an enhanced chemiluminescence detection system (ECL; Amersham, Uppsala, Sweden).

3 Results and discussion

This study is a preliminary step in the determination of the usefulness of 2-D PAGE for defining immunogenic proteins that induce antibody production in patients with systemic candidiasis. Below we report the results corresponding to the different phases involved.

3.1 Collection of different *C. albicans* protein preparations: cytoplasmic extracts, proteins secreted by protoplasts and protoplast lysates

In order to identify several *C. albicans* antigens, we obtained different *C. albicans* protein preparations. First, we collected a typical crude extract containing mainly cytoplasmic proteins (cytoplasmic extract). In this type of preparation, cell walls were removed by centrifugation, although some soluble cell wall proteins may also have been present. Since the yeast cell wall is the structure that mediates the initial interaction between the microorganism and the host, we employed a previously described approach used with *Saccharomyces cerevisiae* [26]. This is based on the analysis of proteins secreted by regenerating protoplasts. Initially, regenerating protoplasts secrete many of their cell wall components into the medium. In the case of *C. albicans*, cells were treated with Glusulase to obtain more than 90% protoplasts, and these were incubated under regenerating conditions for 2 h to allow the secretion of putative cell wall components. Because removal of the yeast cell wall is a stressful condition for the cells, crude extracts from cells without their cell wall were also obtained (protoplast lysates). This allowed us to compare soluble cytoplasmic proteins with proteins secreted by protoplasts from yeast cells under the same stressful conditions.

3.2 Two-dimensional protein maps of *C. albicans*

C. albicans proteins from the three different preparations were separated by high-resolution two-dimensional electrophoresis using an immobilized pH gradient [30]. Six gels (three different experiments carried out in duplicate) of each type of sample were run. Figures 1, 2 and 3 show the maps of these proteins, in the M_r 10–200 kDa and p*I* 3.5–9 experimental window, which were visualized by silver staining. Figure 1 shows the 2-D gel electrophoresis pattern of the soluble proteins of *C. albicans* SC5314. To our knowledge, this is the first *C. albicans* 2-D map obtained with IPGs. Using the Melanie II software, about 600–700 protein spots were detected in cytoplasmic extracts. The resolution was good and the technique was highly reproducible. Figure 2 shows the 2-D map of *C. albicans* protoplast lysates, in which about 600–700 protein spots were also detected. On comparing the maps of soluble proteins from exponentially growing cells with those from cells without cell walls, an important change in the protein pattern was observed. The most relevant finding was the higher expression of certain proteins (M_r range between 30–50 kDa) in the protoplast lysates.

Some of these were later identified as glycolytic enzymes (see below). The 2-D map of proteins secreted by protoplasts regenerating their cell walls is shown in Fig. 3; in this case about 350 proteins were detected.

In the early 1980s two-dimensional gel electrophoresis was used for separating *C. albicans* cytoplasmic proteins in order to investigate different protein expression during the morphological transition (yeast-hyphae) of this organism [35–38]. In these studies, proteins were radioactively labeled and about 200–400 spots were resolved. Finney *et al.* [38] reported the limitations of 2-D PAGE analysis, particularly in the detection of low-abundance regulatory proteins that might be critical to the establishment of alternative morphologies. However, at that time, apart from the problem involved in reproducing the technique, detected spots that were specific to each morphology could not be identified. More recently, Niimi *et al.* [39] have again addressed this issue using 2-D PAGE with carrier ampholytes but failed to detect polypeptides unique to either morphology. Fractionation of *C. albicans* preparations by affinity chromatography in order to detect DNA-binding proteins followed by SDS-PAGE revealed novel polypeptides preferentially synthesized in germ tube-forming cells. Here, using 2-D PAGE with IPGs, we show it is possible to detect more spots than in all previous studies carried out with *C. albicans* protein preparations and that resolution is clearly enhanced. Moreover, the possibility of identifying proteins directly from the silver-stained spots in the gel, by mass spectrometry [40], would allow *C. albicans* researchers to successfully resolve interesting approaches attempted previously but hampered by technical problems that could not be solved with the solutions then available.

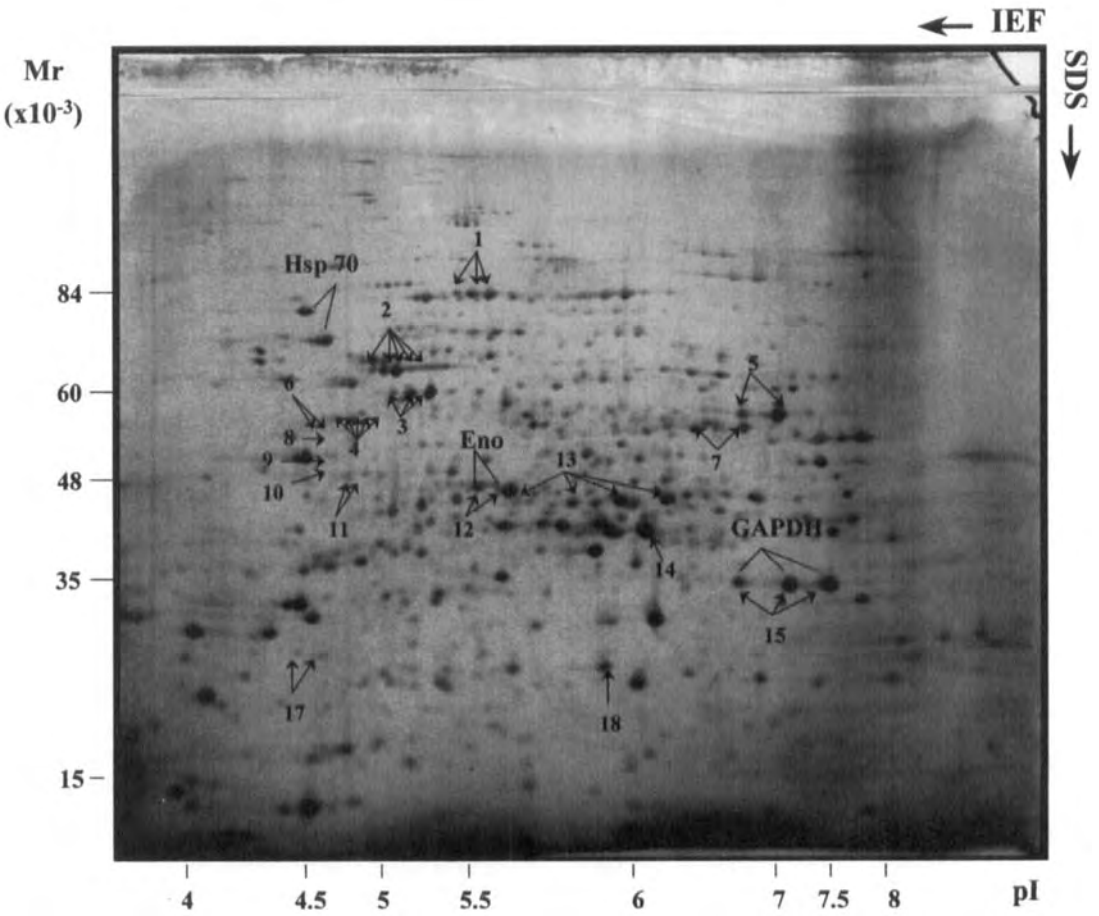


Figure 1. Two-dimensional silver-stained map of *C. albicans* cytoplasmic extracts. Eno, Hsp70 and GAPDH were identified by immunodetection. Arrows indicate antigenic proteins recognized with sera from patients with systemic candidiasis.

3.3 Immunodetection of *C. albicans* proteins

Some previously described *C. albicans* antigens include heat shock proteins and glycolytic enzymes [17]. Antibodies against Eno, Hsp70 and GAPDH were used for immunodetection. Monoclonal antibodies against Eno recognized several isoforms of enolase with slightly different p*I*s. A cluster of two spots at 48 kDa was immunodetected in cytoplasmic extracts. However, clusters of four and three spots at 48 kDa were detected in protoplast lysates and in protoplasts-secreted proteins, respectively. These were assigned to the corresponding clusters in the gels (Figs. 1–3). Summing up, these results showed that different Eno isoforms with different levels of expression were present in our protein preparations. Using a polyclonal antibody against Hsp70 (a protein with high homology with the *S. cerevisiae* SSA family), two spots were detected at 70 kDa and at 78 kDa in both cytoplasmic extracts and protoplast lysates which were assigned to the corresponding spots in the gels. In the protoplast-secreted proteins, this protein was immunodetected in the membrane but was undetectable in the silver-stained gel. We also assayed antibodies against *S. cerevisiae* GAPDH and one cluster of three and four spots was recognized by these antibodies in cell extracts and protoplast lysates, respectively. In the protoplasts-secreted proteins only two spots were recognized and their pattern rendered them easily recognizable in silver-stained gels. Finally, we used antibodies against *S. cerevisiae* Bgl2 (β-1,3-glucan transferase); this protein is strongly expressed in *S. cerevisiae* when protoplasts are regenerating their cell walls [26]. This protein was detected at 34 kDa and p*I* 4.3 in proteins

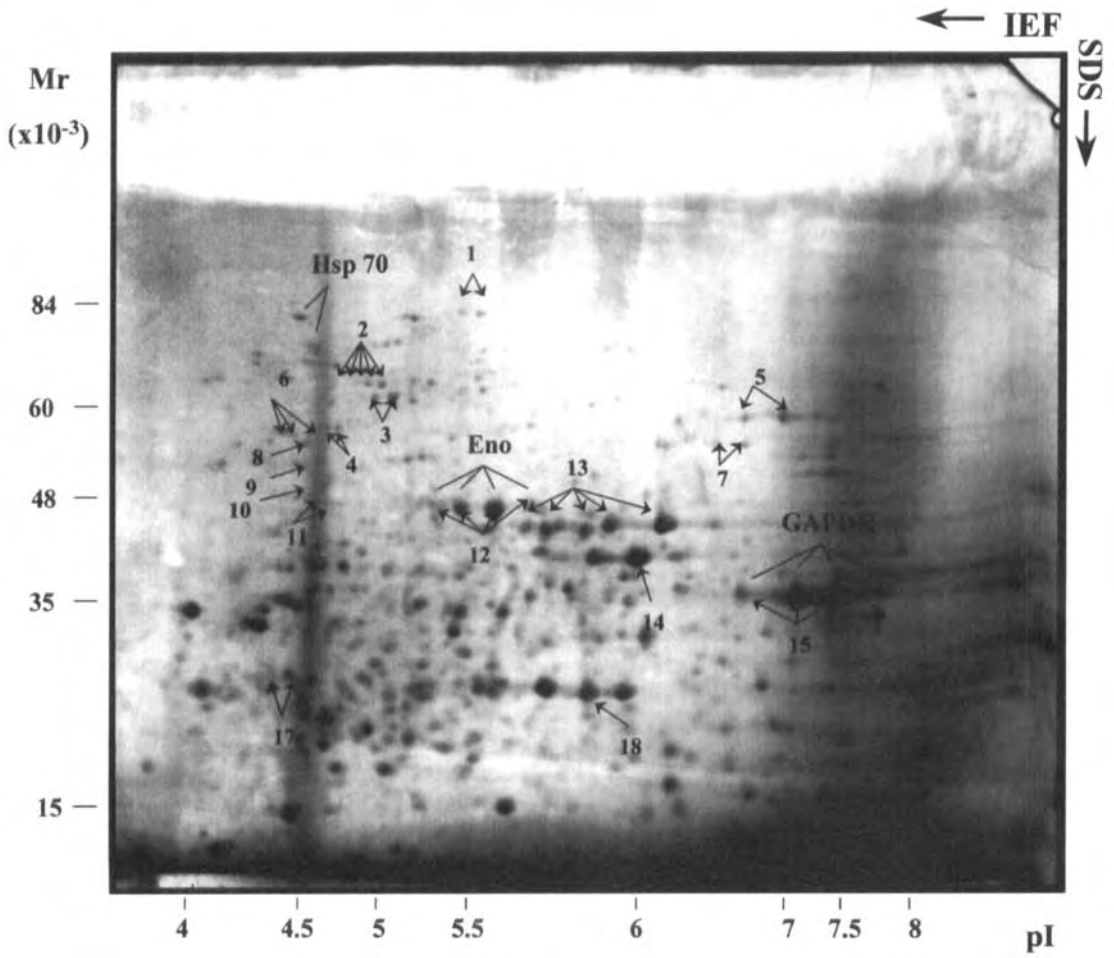


Figure 2. Two-dimensional silver-stained map of *C. albicans* protoplast lysates. Identified proteins are shown. Arrows indicate antigenic proteins recognized with sera from patients with systemic candidiasis.

secreted by protoplasts (Fig. 3). Bgl2 was undetectable by Western blotting in the other protein preparations (cytoplasmic extracts and protoplasts lysates). To summarize, Eno, GAPDH, Hsp70 and Bgl2 were identified with antibodies on the 2-D maps of different *C. albicans* protein preparations.

3.4 Detection of *C. albicans* proteins strongly reactive with sera from patients with systemic candidiasis

Six sera from patients with systemic candidiasis and four sera from healthy individuals were assayed. In preliminary assays using dot blots, tests were performed with sera from patients in order to determine which serum dilution was the most appropriate. Serum dilutions of 1:10–1:10 000 were immunoreactive by dot blot. However, when we used the sera for immunodetection in protein preparations separated by 2-D PAGE, better results were obtained using lower dilutions (1:50 or 1:100). All sera were therefore used at a dilution of 1:50. More than 18 different proteins were immunodetected (Tables 1, 2 and 3; and Figs. 1, 2 and 3), and important conclusions could be obtained from the different protein preparations used. When *C. albicans* cytoplasmic extracts were analyzed, several spots with a molecular mass ranging between 30 and 84 kDa were immunodetected (Table 1 summarizes these results.) Proteins with the same molecular mass and small differences in p*I* are considered to be isoforms. Thus, in this sample it was possible to immunodetect at

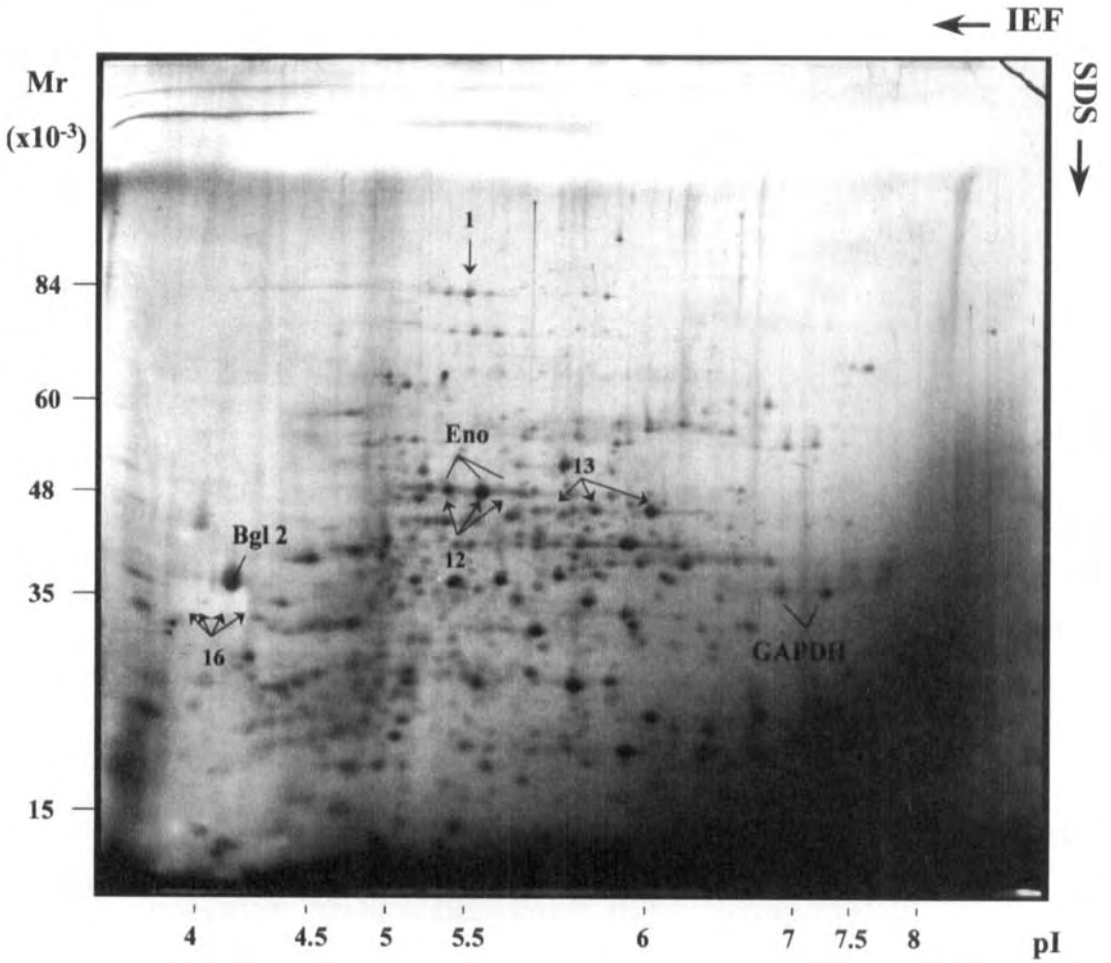


Figure 3. Two-dimensional silver-stained map of proteins secreted by *C. albicans* protoplasts undergoing cell wall regeneration. Identified proteins are shown. Arrows indicate antigenic proteins recognized in sera from patients with systemic candidiasis.

Table 1. *C. albicans* proteins of cytoplasmic extracts detected by sera from patients with systemic candidiasis

Number	Proteins M_r ($\times 10^{-3}$)	p*I*	Sera Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
1	84	5.4–5.7	+	–	±	–	–	+
2	67	4.7–5	+	±	–	+	–	–
3	60–61	4.9–5.1	+	–	–	–	±	–
4	57–58	4.5–4.7	+	±	–	±	–	–
5	58	6.7–7	–	±	–	–	–	–
6	56–57	4.2–4.5	+	–	–	+	–	–
7	56	6.4–6.7	–	±	–	–	–	–
8–10	51–55	4.3–4.4	+	–	–	–	–	–
11	48–50	4.5–4.8	+	–	+	+	–	+
12 Eno	48	5.6–5.8	+	+	+	±	+	+
13	44–46	5.6–6.2	–	+	+	+	–	+
14	42	6	–	–	±	–	–	–
15 GAPDH	35	6.7–7.5	–	±	–	–	–	±
16	34	4–4.4	–	–	–	–	–	–
17	30–32	4.2–4.4	–	±	±	–	–	+
18	30	5.9	–	–	–	–	–	+

Table 2. *C. albicans* proteins from protoplast lysates detected in sera from patients with systemic candidiasis

Number	Proteins M_r ($\times 10^{-3}$)	p*I*	Sera Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
1	84	5.4–5.7	±	–	±	–	–	+
2	67	4.7–5	+	±	±	±	–	–
3	60–61	4.9–5.1	+	–	–	–	+	–
4	57–58	4.5–4.7	+	–	–	–	–	–
5	58	6.7–7	–	±	–	–	–	–
6	56–57	4.2–4.5	+	–	–	+	–	–
7	56	6.4–6.7	–	–	–	–	–	–
8–10	51–55	4.3–4.4	+	–	–	–	–	–
11	48–50	4.5–4.8	+	–	+	+	–	+/–
12 Eno	48	5.6–5.8	+	+	+	+	+	+
13	44–46	5.6–6.2	–	+	+	+	–	+
14	42	6	–	±	±	–	–	–
15 GAPDH	35	6.7–7.5	–	±	–	–	–	+/–
16	34	4–4.4	–	–	–	–	–	–
17	30–32	4.2–4.4	–	+	–	–	–	+/–
18	30	5.9	–	–	–	–	–	+

Table 3. *C. albicans* protoplast-secreted proteins detected by sera from patients with systemic candidiasis

Number	Proteins M_r ($\times 10^{-3}$)	p*I*	Sera Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
1	84	5.4–5.7	–	–	±	–	–	–
12 Eno	48	5.6–5.8	+	+	+	±	±	+
13	44–46	5.6–6.2	–	±	+	+	–	±
16	34	4–4.4	+	+	+	+	±	+

least 17 different proteins (including Eno and GAPDH). All sera reacted with at least two different proteins and the most immunoreactive serum detected at least ten proteins (Fig. 4 shows the immunoblot with serum from the most immune reactive patient). All sera recognized Eno. Another highly immunogenic protein has a molecular mass of 49 kDa and a p*I* range of 4.5–4.8. When sera from uninfected subjects were assayed, a faint reaction

with Eno was observed (data not shown). This can be explained in terms of the notion that *C. albicans* is a commensal organism.

The same spots were immunodetected when protoplast lysates were used (Fig. 2). However, the intensity of the immunodetection (in the gel region corresponding to a molecular mass ranging between 30 and 55 kDa) was higher than when cytoplasmic extracts were used, a fact that simplified the immunodetection. Figure 5 shows a protoplast lysate immunoblot detected with serum from patient number 1. The least reactive protein preparation involved the proteins secreted by protoplasts; only four proteins were immunodetected. Again, Eno was immunodetected. A new protein that was not present in the other samples was recognized. This protein has a molecular mass of 34 kDa and a p*I* between 4–4.4 (Fig. 6). Angiolella *et al.* [41] have described a 34 kDa cell wall mannoprotein, which was recognized by a rabbit anti-*Candida* serum. Further research is necessary in order to know whether these proteins are related. The identification of this protein could be important and is currently ongoing.

In conclusion, this work is the first attempt at a large-scale identification of *C. albicans* antigens by combining 2-DE and Western blotting with sera from patients with systemic candidiasis. A similar approach is currently being used to identify *Haemophilus influenzae* immunogenic proteins [42]. We show the presence of at least 18 immunogenic proteins, some of them glycolytic enzymes. The identification of all these proteins is now underway. Although preliminary, the finding of the presence of different antibodies in the sera of patients and the study of different risk groups with systemic candidiasis would allow the identification of specific immunodominant antigens in order to design improved serologic tests for the diagnosis of candidiasis.

4 Concluding remarks

Two-dimensional gel electrophoresis of *C. albicans* protein preparations followed by Western blotting with sera from patients with systemic candidiasis provides a powerful tool for the identification of immunoreactive proteins. In the present study, we used different *C. albicans* protein preparations (cytoplasmic extracts, protoplast lysates and proteins secreted by protoplasts). The use of different samples with different patterns of protein expression reveals a useful approach for identifying more immunoreactive proteins as well as achieving better immunodetection. Protoplast lysates allowed easier immunodetection because some of the immunogenic proteins are highly expressed in this sample. Moreover, the use of a sample enriched in cell wall components also allowed the immunodetection of another antigen that was not present in the other samples. This antigen (M_r 34 and p*I* 4–4.4) was the most immunoreactive protein but, unlike Eno, it was not detected in sera from healthy subjects. The identification of all the immunodetected *C. albicans* proteins in this work is currently being addressed at our laboratory. Similar studies with patients with systemic candidiasis belonging to different risk groups (*i.e.*, diabetics, patients with AIDS, with solid or hematologic tumors) could help in understanding the role of the humoral response in *C. albicans* infections. The overall view of the different *Candida* antibodies that are being produced in a patient could be

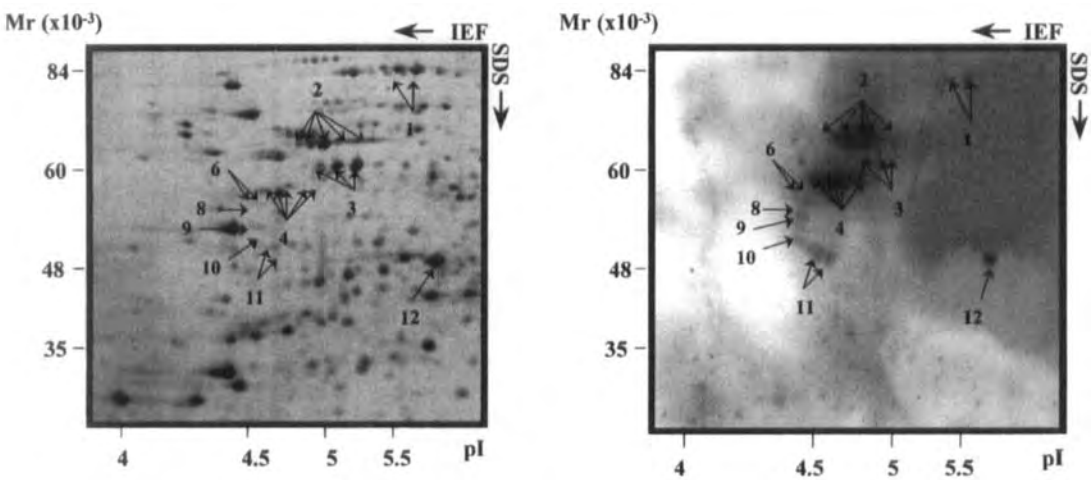


Figure 4. Left: Detail of the two-dimensional silver-stained electropherograms of *C. albicans* cytoplasmic extracts. Arrows indicate spots matching with immunoreactive spots in Western blots with serum from a patient with systemic candidiasis (number 1). Right: Western blots of similar electropherograms with serum from patient number 1.

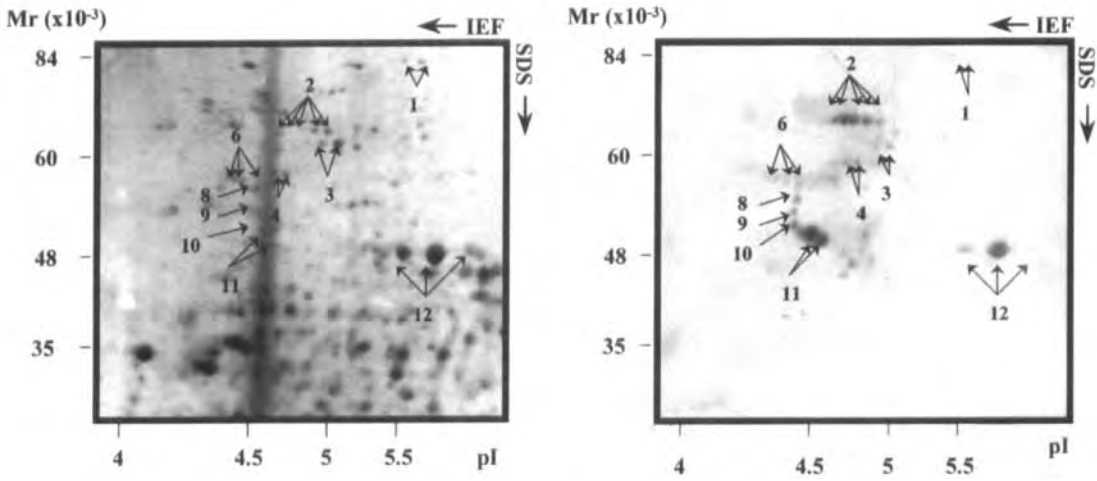


Figure 5. Left: Detail of the two-dimensional silver-stained electropherograms of *C. albicans* protoplast lysates. Arrows indicate spots matching with immunoreactive spots in Western blots with serum from a patient with systemic candidiasis (number 1). Right: Western blots of similar electropherograms with serum from patient number 1.

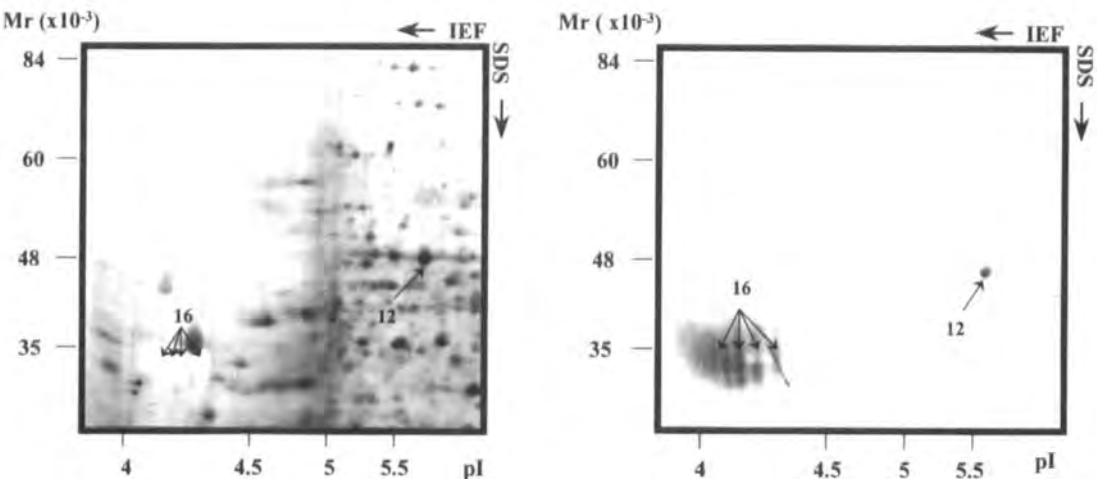


Figure 6. Left: Detail of the two-dimensional silver-stained electropherograms of *C. albicans* protoplasts-secreted proteins. Arrows indicate spots matching with immunoreactive spots in Western blot with serum from a patient with systemic candidiasis (number 1). Right: Western blots of similar electropherograms with serum from patient number 1.

useful for the diagnosis as well as monitoring the evolution of the disease.

We thank the colleagues who provided us with the antibodies used here for protein identification: J. P. Martínez and M. L. Gil (anti-GAPDH), A. Cassone, R. La Valle and C. Bromuro (anti-Hsp70), and P. Sullivan (anti-Bg12). We also thank W. H. Goessens and J. Pontón for supplying several reactives and M. Molina for encouragement and critical discussions for improvement of the manuscript. We are grateful to Nicholas Skinner for revising the English manuscript. This work was supported by grants SAF96-1540 and FIS 97/0047-01. A. Pitarch and M. Pardo were the recipients of fellowships from Fundación Ramón Areces and from the Ministerio de Educación y Ciencia of Spain, respectively.

Received November 16, 1998

5 References

[1] Anaissie, E. J., *Clin. Infect. Dis.* 1992, *14*, 43–53.

[2] Fox, J. L., *ASM News* 1993, *10*, 515–518.

[3] Dixon, D. M., McNeil, M. M., Cohen, M. L., Gellin, B. G., La, M. J., *Public Health Rep.* 1996, *111*, 226–235.

[4] Casadevall, A., *Infect. Immunol* 1995, *63*, 4211–4218.

[5] Fukazawa, Y., Cassone, A., Bistoni, F., Howard, D. H., Kagaya, K., Murphy, J. W., Cenci, E., Lane, T. E., Mencacci, A., Puccetti, P., *J. Med. Vet. Mycol.* 1994, *32*, 123–131.

[6] Levitz, S. M., *Clin. Infect. Dis.* 1992, *14*, S37–S42.

[7] Fidel, Jr., P. L., Sobel, J. D., *Clin. Microbiol. Rev.* 1996, *9*, 335–348.

[8] Matthews, R., Burnie, J., *Trends Microbiol.* 1996, *4*, 354–358.

[9] Alloush, H. M., Lopez-Ribot, J. L., Masten, B. J., Chaffin, W. L., *Microbiology* 1997, *143*, 321–330.

[10] Matthews, R., Burnie, J., *FEMS Microbiol. Lett.* 1989, *51*, 25–30.

[11] La Valle, R., Bromuro, C., Ranucci, L., Muller, H. M., Crisanti, A., Cassone, A., *Infect. Immunol.* 1995, *63*, 4039–4045.

[12] Gil-Navarro, I., Gil, M. L., Casanova, M., O'Connor, J. E., Martinez, J. P., Gozalbo, D., *J. Bacteriol.* 1997, *179*, 4992–4999.

[13] Macdonald, F., Odds, F. C., *JAMA* 1980, *243*, 2409–2411.

[14] Strockbine, N. A., Largen, M. T., Zweibel, S. M., Buckley, H. R., *Infect. Immunol.* 1984, *43*, 715–721.

[15] Shen, H. D., Choo, K. B., Lee, H. H., Hsieh, J. C., Lin, W. L., Lee, W. R., Han, S. H., *Clin. Exp. Allergy* 1991, *21*, 675–681.

[16] Swoboda, R. K., Bertram, G., Hollander, H., Greenspan, D., Greenspan, J. S., Gow, N. A., Gooday, G. W., Brown, A. J., *Infect. Immunol.* 1993, *61*, 4263–4271.

[17] Martinez, J. P., Gil, M. L., Lopez-Ribot, J. L., Chaffin, W. L., *Clin. Microbiol. Rev.* 1998, *11*, 121–141.

[18] Jones, J. M., *Clin. Microbiol. Rev.* 1990, *3*, 32–45.

[19] van Deventer, A. J., van Vliet, H. J., Hop, W. C., Goessens, W. H., *J. Clin. Microbiol.* 1994, *32*, 17–23.

[20] Mitsutake, K., Miyazaki, T., Tashiro, T., Yamamoto, Y., Kakeya, H., Otsubo, T., Kawamura, S., Hossain, M. A., Noda, T., Hirakata, Y., Kohno, S., *J. Clin. Microbiol.* 1996, *34*, 1918–1921.

[21] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[22] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer-Verlag, Berlin 1997.

[23] Gillum, A. M., Tsay, E. Y. H., Kirsch, D. R., *Mol. Gen. Genet.* 1984, *198*, 179–182.

[24] Pla, J., Gil, C., Monteoliva, L., Navarro-García, F., Sánchez, M., Nombela, C., *Yeast* 1996, *12*, 1677–1702.

[25] Gil, C., Pomés, R., Nombela, C., *J. Gen. Microbiol.* 1988, *134*, 1587–1595.

[26] Pardo, M., Monteoliva, L., Pla, J., Sánchez, M., Gil, C., Nombela, C., *Yeast* 1999, in press.

[27] Soll, D. R., Mitchell, L. H., *J. Cell Biol.* 1983, *96*, 486–493.

[28] Cabib, E., Duran, A., *J. Bacteriol.* 1975, *124*, 1604–1606.

[29] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[30] Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.

[31] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[32] Merril, C. R., Goldman, D., Van Keuren, M. L., *Electrophoresis* 1982, *3*, 17–21.

[33] Towbin, H., Staehelin, T., Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.

[34] Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. (Eds.), *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, New York 1995.

[35] Manning, M., Mitchell, T. G., *J. Bacteriol.* 1980, *144*, 258–273.

[36] Manning, M., Mitchell, T. G., *Infect. Immun.* 1980, *30*, 484–495.

[37] Brown, L. A., Chaffin, W. L., *Can. J. Microbiol.* 1981, *27*, 580–585.

[38] Finney, R., Langtimm, C. J., Soll, D. R., *Mycopathologia* 1985, *91*, 3–15.

[39] Niimi, M., Shepherd, M. G., Monk, B. C., *Arch. Microbiol.* 1996, *166*, 260–268.

[40] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[41] Angiolella, L., Simonetti, N., Cassone, A., *J. Antimicrob. Chemother.* 1994, *33*, 1137–1146.

[42] Marzocchi, B., Bini, L., Liberatori, S., Raggiaschi, R., Cellesi, C., Pallini, V., Rossoloini, A., *3rd Siena 2-D Electrophoresis Meeting* 1998, Abstract.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1011–1016

Michael T. Maicher
Arno Tiedtke

Institute for General Zoology and Genetics, University of Münster, Münster, Germany

Biochemical analysis of membrane proteins from an early maturation stage of phagosomes

We used an improved technique for pulse-chase labeling of phagosomes using custom-made magnetic microparticles. With the help of a permanent magnet we purified both newly formed, nascent and early matured (*i.e.*, 5-min-old) condensed phagosomes in high amounts. The protein patterns of membrane proteins of newly formed phagosomes and 5-min-old condensed ones were compared by two-dimensional (2-D) electrophoresis. The protein patterns allowed the detection of protein spots that changed in abundance between these two stages. Three protein spots abundant in condensed phagosomes only and one spot well-stained in both stages were collected from ten preparative Coomassie brilliant blue-stained 2-D gels. Following microdigestion, selected purified oligopeptides were sequenced by Edman degradation. While the oligopeptide sequences of proteins from two spots showed high homology to an already sequenced 25 kDa calcium binding protein, the other two showed no significant homology to protein sequences available in sequence databases. Presently polymerase chain reaction (PCR) and cloning experiments are set up to reveal the cDNAs of these proteins in order to study their function by knock-out and gene replacement experiments.

Keywords: Phagosome / Two-dimensional polyacrylamide gel electrophoresis / Oligopeptide sequences / *Tetrahymena thermophila* EL 3358

1 Introduction

Phagocytosis is a special form of endocytosis performed by eukaryotic cells to internalize large particles. In mammalia, specialized cells, primarily cells of the immune system like macrophages and polymorphonuclear leukocytes are engaged in phagocytosis [1]. Protozoa, *e.g.* ciliates, use phagocytosis mainly to take up nutrients like bacteria or other food particles [2]. In general, phagosomes are formed by receptor- and f-actin-dependent processes [1]. In ciliates, the membrane of the nascent and newly formed phagosome is provided by discoidal vesicles which fuse with the cytopharyngeal membrane of the cytostome. Along their maturation pathway the newly formed phagosomes undergo a series of interactions with other vesicles of the endocytotic and secretory pathway, which are accompanied by changes in pH, protein content, and protein composition [3]. For *Paramecium* it has been shown that newly formed phagosomes are acidified by fusion with a distinct class of vesicles, the so-called acidosomes [4]. Acidification of phagosomes by fusion with vesicles containing proton-pumping ATPases of the vacuolar type is necessary to activate lysosomal enzymes in the phagolysosome, the final compartment of the phagocytotic pathway, where biogenic macromolecules are digested [5–7]. Following retrieval of lysosomal enzymes and membranes, old phagosomes eventually fuse with the membrane of the cytoproct, the site where residual materials of the old phagosomes are egested [3, 4]. Although the order of steps of vesicle interactions involved in phagosome biogenesis and maturation are well-described at the cytological level, neither the molecular composition of phagosomes and the interacting vesicles nor the molecular mechanism that regulate these fusion events are known. To obtain more information on the molecular details of the phagocytotic pathway we began to study the age-dependent changes in phagosomal protein patterns in the ciliate *Tetrahymena thermophila*. To purify phagosomes of different maturation stages in high amounts, we have set up a technique for pulse-chase labeling of phagosomes, using custom-made colloidal iron-starch magnetic particles of about 1 μm in size [8]. Feeding *Tetrahymena* cells for 30 s with these magnetic particles is sufficient to label a single, nascent phagosome per cell in more than 90% of the cell population. The first phagosomal maturation stage, the condensed phagosome, is purified from cells that are kept for a further chase period of 4.5 min in fresh culture medium. Both types, the newly formed and condensed phagosomes, are separated from the cell homogenate by the use of permanent magnets.

Correspondence: Prof. Dr. Arno Tiedtke, Institute for General Zoology and Genetics, University of Münster, Schlossplatz 5, D-48149 Münster, Germany
E-mail: tiedtke@nwz.uni-muenster.de
Fax: +49-251-83-24723

In this study we analyzed the protein patterns of purified newly formed and condensed phagosomes. Their membrane proteins, enriched in the detergent phase by the Triton X-114 partition assay and separated by two-dimensional gel electrophoresis, show distinct patterns. A number of protein spots appear to be specific for a particular stage of maturation. Our purification technique allowed us to collect phagosomal membrane proteins in microgram amounts from preparative 2-D gels and, in addition, to sequence selected oligopeptides from some of these proteins.

2 Material and methods

2.1 Cell culture and medium

For the isolation of phagosomes we used the *Tetrahymena thermophila* strain SB281 which is blocked in exocytosis of mucocysts [9]. The complex medium (PPYS), in which the cells were grown, contained 1% w/v proteose peptone, 0.1% w/v yeast extract (both from Difco, Detroit, MI, USA), and 0.003% v/v Sequestrene^{R} (Ciba-Geigy, Greensboro, NJ, USA). The cells were grown at 30°C, 100 rpm, on a gyrotory shaker in Fernbach flasks (1.2 L), containing 400 mL culture medium [10]. For the isolation of phagosomes, the cells were grown to the late logarithmic phase of growth (0.5–1 × 10^6 cells per mL).

2.2 Labeling and purification of phagosomes

For labeling of phagosomes we prepared colloidal iron-starch magnetic microparticles of about 1 μm in diameter as described by Vosskühler and Tiedtke [8]. $FeCl_2$ and $FeCl_3$ were obtained from Fluka (Deisenhofen, Germany) and starch from Biomol (Hamburg, Germany). A cell culture of 6.4 L (10^6 cells per mL) was sedimented in portions of 4 × 100 mL in an oil test centrifuge at 380 × *g* and resuspended in 4 × 25 mL of fresh culture medium. To obtain phagosomes of a defined age of maturation the cells were fed for 30 s with a 10% v/v suspension of magnetic particles. During this short pulse period more than 90% of the cells formed a single phagosome. To inhibit further membrane fusion events, the cells were quickly cooled down to 7°C in an ice bath (–15°C, containing NaCl) and sedimented (380 × *g* for 1 min); the supernatant containing noningested magnetic particles was aspirated. Newly formed phagosomes were directly purified with the help of a permanent magnet from the homogenate of these cells. To obtain maturation stages of newly formed, labeled phagosomes, another aliquot of the sedimented cells was resuspended in fresh culture medium (warmed up to 30°C) and kept on a gyrotory shaker for the appropriate chase time. The condensed phagosomes, as an early maturation stage, were purified from the homogenized cells after a chase of 4.5 min. The cells were homogenized in ice-cold homogenization buffer (10 mM Tris-HCl, pH 7.4, 200 mM sucrose, 1 mM EDTA, 0.8 mM Pefabloc, 2 mM leupeptin) with an Ultra-turrax homogenizer (IKA, S25KR-18G; Staufen, Germany). Progress of homogenization was monitored by phase-contrast microscopy. The homogenate was layered on top of a sucrose cushion (10 mM Tris-HCl, pH 7.4, 1 M sucrose, 1 mM EDTA) of 1.5 cm height in a 600 mL beaker. The magnetically labeled phagosomes were pulled down through the sucrose layer to the bottom of the beaker with a strong permanent magnet. The supernatant was discarded and the phagosomes were washed twice with homogenization buffer to remove nonmagnetic contaminants. They were resuspended in a small amount of homogenization buffer and transferred into an Eppendorf tube for further washing steps (380 × *g*, 3 min). The phagosome preparations were regularly checked for contamination using isocitrate dehydrogenase as a marker enzyme for cytosol and mitochondria, as described by Vosskühler and Tiedtke [8]. Only uncontaminated phagosomal preparations were used for further experiments. The purified phagosomes were sedimented and frozen in liquid nitrogen and stored at –70°C. The purification procedure is schematically shown in Fig. 1.

2.3 Separation of phagosomal membrane proteins by the Triton X-114 assay

The purified phagosomes obtained from 6.4 L cell culture (about 6 × 10^8 phagosomes) were solubilized in 3 mL lysis buffer (10 mM Tris-HCL, pH 7.4, 1% v/v Triton X-114, 0.8 mM Pefabloc, 2 mM leupeptin) by three freeze-thaw cycles in liquid nitrogen and then stored at 4°C for a further 30 min. Magnetic particles were removed from the solubilized phagosomal proteins by sedimentation (50 min, 4°C, 20 000 × *g*). The protein concentration was determined according to the BCA protocol (Pierce, Rockford, IL, USA) using BSA as a protein standard. Membrane proteins and soluble proteins of the supernatant were separated by warming up the Triton X-114 lysate to 30°C followed by centrifugation (20°C, 5 min, 20 000 × *g*) to separate the detergent phase from the aqueous phase.

2.4 Two-dimensional gel electrophoresis

The detergent phase containing enriched phagosomal membrane proteins was precipitated with 3 M TCA, the precipitate was sedimented (4°C, 5 min, 20 000 × *g*), washed three times with diethylether, mixed, centrifuged again (4°C, 5 min, 20 000 × *g*) and stored at room temperature until the diethylether was evaporated. The protein pellet was solubilized by sonification (three times, 3 min each) in 275 μL IEF sample buffer (5 M urea, 2 M

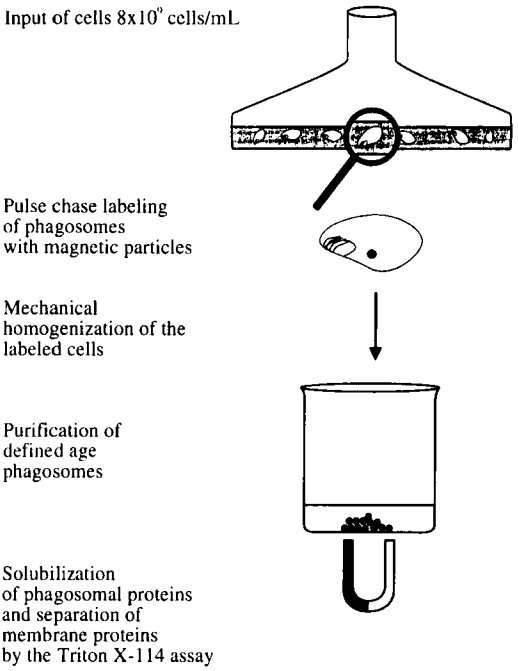


Figure 1. Illustration of the technical procedure for enrichment of phagosomes of defined age.

thiourea, 4% w/v CHAPS, 0.8% v/v Ampholine, 65 mM DTT) at room temperature [11] and centrifuged for 5 min at 20 000 × *g* to remove insoluble material. For isoelectric focusing the Pharmacia Biotech Multiphor II system with immobilized pH gradient gels (Pharmacia, Uppsala, Sweden) was used according to the method of Görg *et al.* [12, 13]. Immobiline dry strips (11 cm, linear gradient pH 4–7) were directly rehydrated with the sample solution for 6 h, according to the method of Sanchez *et al.* [14], using the Pharmacia rehydration chamber. After rehydration the gel strips were focused in the Multiphor II unit for 16 h (300 V for 4 h, 1150 V for 5 h, 3500 V for 7 h), cooled to 15°C during the run. Focused IEF gel strips were used directly for the second dimension (SDS-PAGE). For the second dimension, IEF gel strips were equilibrated two times for 10 min in equilibration buffer (6 M urea, 150 mM Tris-HCl, pH 8.8, 30% v/v glycerol, 10 mM DTT, 2% w/v SDS containing trace amounts of bromophenol blue). The IEF gel strips were loaded on top of the SDS gels (5% stacking gel; 10% separating gel). SDS-PAGE was run at 200 V for 2 h 45 min under cooling condition (4°C).

2.5 *N*-terminal amino acid sequencing

After SDS-PAGE, analytical gels were stained with silver salt [15]. Isoelectric point and molecular weight calibration of protein spots was performed by coelectrophoresis with the Mark 12 protein standard (Novex, San Diego, CA, USA) and the Carbamylate calibration kit (Pharmacia). The preparative gels were fixed and stained for 1 h in Coomassie brilliant blue R-350 and destained in destaining solution (670 mL bidistilled water, 250 mL ethanol p.a., 80 mL acetic acid p.a.). Gels were stored at 4°C in destaining solution. Selected protein spots of ten polyacrylamide gels were pooled in Eppendorf tubes for in-gel digestion with endoprotease Lys C. Purified HPLC fractions of oligopeptides were sequenced by Edman degradation on a gas phase sequenator (ABI 492 from TopLab GmbH, Martinsried, Germany). For the database search of peptide sequences we used PATTERNPROT at http://pbil.ibcp.fr/NPSA/npsa_pattern.html.

3 Results and discussion

We applied our improved technique for the labeling of phagosomes using custom-made magnetic microparticles [8]. We purified both newly formed, nascent, and early matured, *i.e.* 5-min-old, condensed phagosomes in high amounts. We obtained 0.53 mg of total phagosomal proteins from the nascent type and 0.8 mg protein of the condensed type from 6.4×10^9 labeled cells. These high yields are due to a high synchrony in phagosome formation and subsequent maturation, since 90 percent of the cells had formed a single phagosome within 30 s. Our system is simple and the yields high, and these features emphasize our *Tetrahymena* system as a useful model for studies on biochemical characterization of phagosomes. The 2-D separation of phagosomal membrane proteins is shown in Fig. 2. We used the Triton X-114 partition assay and stained the proteins of the detergent phase with silver. There were about 150 spots from both nascent and condensed phagosomes. They range in apparent molecular masses from 10 000 to 110 000 kDa and in p*I* values from pH 4 to 7. The differences between the two phagosomal preparations show both loss and gain of proteins, indicated by arrows in Fig. 2. Similar results have been seen in *Paramecium* [16].

We have aimed at a molecular characterization of phagosomes. First, we studied membrane proteins which change in amount between the different ages of phagosome maturation. The electropherogram of a Coomassie blue-stained preparative 2-D gel (Fig. 3) showed a similar number and distribution of protein spots from condensed phagosomes as in Fig. 2B, although some spots were stained less intensively. We collected three spots with low molecular masses numbered 1–3 in Fig. 1. These were better stained in the condensed phagosomes than in the nascentones (see Fig. 2A, B). In addition, we punched out protein 4, which was well-stained in both ages of

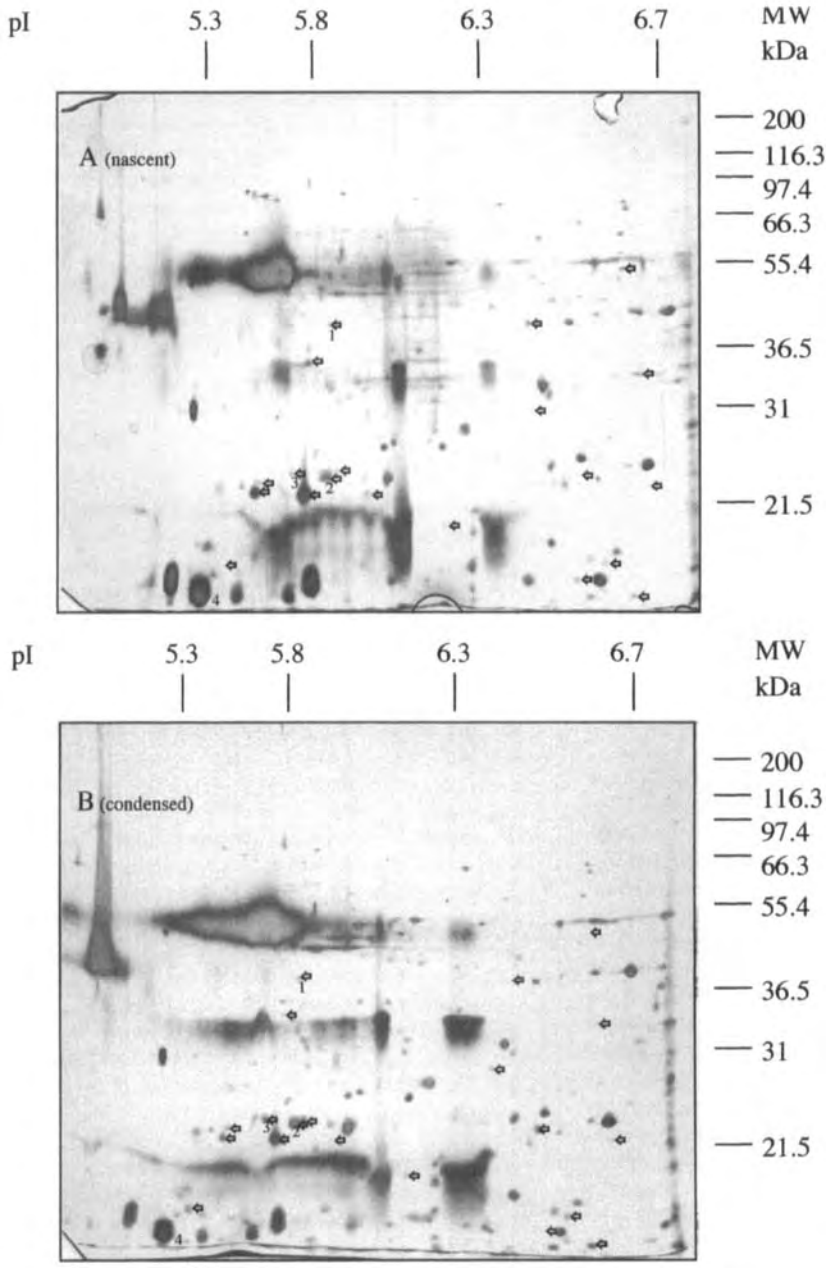


Figure 2. Silver-stained two-dimensional gel electropherograms of membrane proteins obtained from purified (A) Nascent and (B) condensed phagosomes. After solubilization with Triton X-114 and separation of the membrane phase from the aqueous phase, membrane proteins were precipitated with TCA and washed with diethylether. In the first dimension (IEF) proteins were separated on linear immobilized pH gradient gels (pH 4–7; 11 cm) and in the second dimension on a 10% SDS polyacrylamide gel. Open arrows point to protein spots that changed in abundance during the transition from nascent to condensed phagosomes. Spots 1–3 were considered specific for condensed phagosomes, while spot 4 was found well-stained in nascent and in condensed ones. The same spots were punched out from the preparative Coomassie brilliant blue-stained gels (see Fig. 3) used for microdigestion and amino acid sequencing.

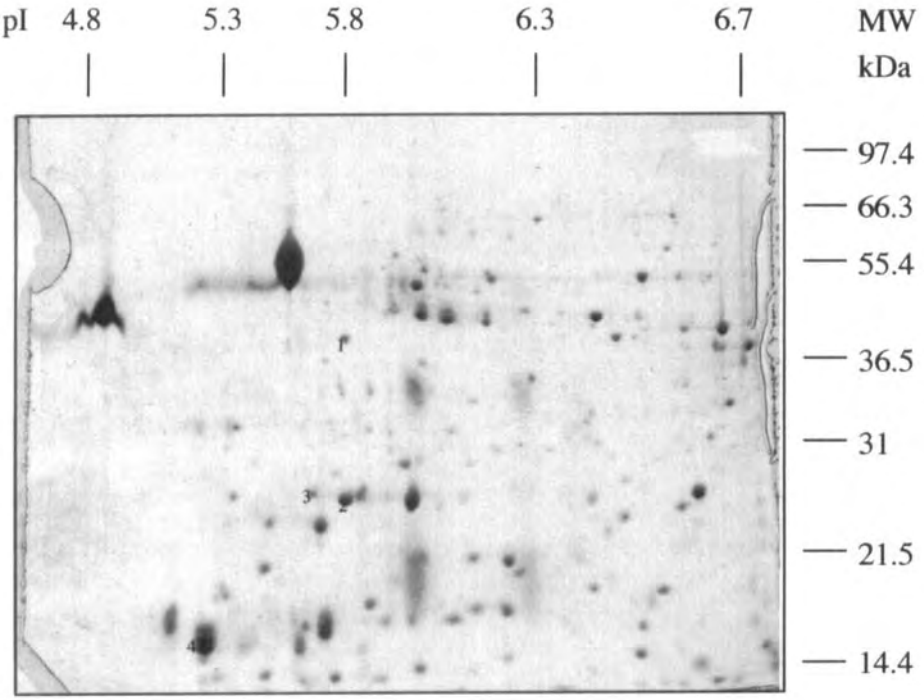


Figure 3. Coomassie blue-stained 2-D electropherogram of membrane proteins from condensed phagosomes. The proteins separated in the Triton X-114 detergent phase were precipitated with TCA and washed with diethylether. They were separated in the first dimension (IEF) on linear immobilized pH gradient gels (pH 4–7; 11 cm). In the second dimension the focused proteins were separated on 10% SDS polyacrylamide gels. Spots 1–3 were considered specific for condensed phagosomes, while spot 4 was found well-stained in both ages of phagosomal development. All four spots were used for microdigestion and amino acid sequencing.

Table 1. Sequences obtained from Edman degradation

Spot No.	M_r kDa	p*I*	Amino acid sequences[a)]
1	40	5.9	LAQTGLNIQQPYST QIGDNVFIGPNSYI NPNVEATIIDREF
2	26	5.9	LDVARRLFK (Q)SLVF
3	25	5.7	GIRVEKQ FN(S)FFAPSSD(D)IK
4	18	5.1	LIQDYK AIYN(D)AIEIK TLNLNDLFPTLK DGVDLSINLIK

a) Amino acids in parentheses are ambiguous

phagosome development. The oligopeptide sequences obtained from ten Coomassie brilliant blue-stained spots of each of the selected four protein spots are shown in Table 1. The database search for the sequences of spots 2 and 3 revealed identity of the sequences with those of the cDNA of a 25 kDa calcium binding protein of *Tetrahymaen thermophila* reported by Takemasa and co-workers [17]. They showed by immunofluorescence that the 25 kDa calcium binding protein occurred in the cilia. Further studies may reveal the function of this protein on phagosomes and explain the reasons for the various p*I* on our 2-D gels. The oligopeptide sequences obtained from spots 1 and 4 did not show any significant homology to known proteins, a finding that is not surprising, since only little is known about the molecular details of the phagosomal pathway. We are presently setting up PCR and cloning experiments to reveal the cDNAs of these proteins in order to identify the primary structure of these proteins. In addition, we will obtain sequence information about the up- and downstream sequences flanking the coding regions of these genes to obtain the tools for knock-out as well as gene replacement studies recently developed for *Tetrahymena thermophila* [18].

This study was supported by a grant of the Deutsche Forschungsgemeinschaft (SFB 310) to A. Tiedtke. We wish to thank all members of the laboratory of A. Tiedtke for their assistance and valuable discussion. We thank Prof. Dr. Leif Rasmussen, Odense University, for critically reading the manuscript.

Received November 8, 1998

4 References

[1] Brown, E. J., *BioEssays* 1995, *17*, 109–117.

[2] Nilsson, J. R., in: Levandowsky, M., Hutner, S. H. (Eds.), *Biochemistry and Physiology of Protozoa*, Academic Press, New York 1979, Vol. 2, pp. 339–379.

[3] Smith-Somerville, H. E., *Eur. J. Cell Biol.* 1989, *49*, 48–54.

[4] Allen, R. D., Fok, A. K., *J. Cell Biol.* 1983, *97*, 566–570.

[5] Mellman, I., Fuchs, R., Helenius, A., *Annu. Rev. Biochem.* 1986, *55*, 663–700.

[6] Desjardins, M., *Trends Cell Biol.* 1995, *5*, 183–186.

[7] Ishida, M., Aihara, M. S., Allen, R. D., Fok, A. K., *Protoplasma* 1997, *196*, *12–20*.

[8] Vosskühler, C., Tiedtke, A., *J. Euk. Microbiol.* 1993, *40*, 556–562.

[9] Maihle, N. J., Satir, B. H., *J. Cell. Sci.* 1985, *78*, 49–65.

[10] Tiedtke, A., *Comp. Biochem. Physiol.* 1983, *75B*, 239–243.

[11] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–326.

[12] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.

[13] Görg, A., Postel, W., Günther, S., Friedrich, C., *Electrophoresis* 1988, *9*, 57–59.

[14] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[15] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[16] Allen, R. D., Bala, N. P., Ali, R. F., Nishida, D. M., Aihara, M. S., Ishida, M., Fok, A. K., *J. Cell Biol.* 1995, *108*, 1263–1274.

[17] Takemasa, T., Ohnishi, K., Kobayashi, T., Takagi, T., Konishi, K., Watanabe, Y., *J. Biol. Chem.* 1989, *264*, 19293–19301.

[18] Hai, B., Gorovsky, M. A., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 1310–1315.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, refer to: *Electrophoresis* 1999, *20*, 1017–1026

Raymond Joubert-Caron[1]
Jean Feuillard[1]
Sylvie Kohanna[1]
Florence Poirier[1]
Jean-Pierre Le Caër[2]
Marino Schuhmacher[3]
Georg W. Bornkamm[3]
Axel Polack[3]
Michel Caron[1]
Dominique Bladier[1]
Martine Raphaël[1]

[1]Biochimie Cellulaire des Hémopathies Lymphoïdes et des Vascularites, UFR Léonard de Vinci, Bobigny Cedex, France
[2]Laboratoire de Neurobiologie, Ecole Supérieure de Physique Chimie de la Ville de Paris, Paris Cedex 5, France
[1]Institut für Klinische Molekuläre Biologie and Tumor-Genetik, München, GFF, Germany

A computer-assisted two-dimensional gel electrophoresis approach for studying the variations in protein expression related to an induced functional repression of NFκB in lymphoblastoid cell lines

Strategies are needed for conclusive interpretation of two-dimensional gel electrophoresis (2-D PAGE) maps in order to identify pertinent differences in protein expression during regulation of the transcription of discrete sets of genes. The model used in this study was a human lymphoblastoid cell line in which a functional repression of the transcription factors NFκB was obtained by induction of overexpression of IκBα, a physiological inhibitor of NFκB. The analytical methodology used relies on the comparison of two sets of 2-D PAGE maps for detecting differences in protein expression between samples overexpressing or not overexpressing IκBα. The analysis was based on a combination of an automatic computerized analysis, constituting an actual aid for deciding, and of an interactive visual validation, corresponding to the interpretation of computer propositions. This strategy is proposed as a rapid way to detect potential variations in protein expression applicable to any biological model. In this study, correspondence analysis data made it possible to discrimate between the samples overexpressing or not overexpressing IκBα, and pointed out some of the potential meaningful spots characterizing the samples in which NFκB was active. Then, after visual validation of the computer data, 53 polypeptides were considered to be different in the two classes of gels. Five polypeptides were specifically found in both samples overexpressing IκBα. The overexpression of IκB also induced a lower expression of 11 polypeptides. Finally, 15 polypeptides were only expressed in samples in which IκBα was not overexpressed and, consequently, in which NFκB factors were active. Thus, these polypeptides are candidates for further analysis as putative target gene products of NFκB.

Keywords: Two-dimensional gel electrophoresis / Lymphoblastoid cells / Transcription factors NFκB, IκB
EL 3334

1 Introduction

To understand how the chronic activation of the transcription factors NFκB in cells infected by the Epstein Barr virus (EBV) [1] may play a role in cell transformation, it would be of interest to identify the NFκB target gene products. For this purpose the two-dimensional gel electrophoresis (2-D PAGE) approach might be convenient. 2-D PAGE has become the primary tool for analysis of protein mixtures due to its high resolution and sensitivity. Since its introduction in 1975 [2], there have been many modifications of the technique [3]. The most important have been, first, the introduction of immobilized pH gradients [4–6] and, more recently, the development of third-generation image analysis software [7, 8]. Nevertheless, the 2-D PAGE basic principle has remained the same: separation of the components of a complex protein extract first by electrofocusing (IEF), which separates proteins according to their p*I* and then by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), which separates according to size/mass. 2-D PAGE provides a global view of the state of the protein expression, since many hundreds of polypeptides can be visualized at once after staining. Protein maps of various mammalian tissues and cell lines have been established [9–12]. Most of these maps are accessible by consulting 2-D PAGE databases on the World Wide Web (WWW) [13–23]. To determine differences in protein expression between two samples,

Correspondence: Raymonds Joubert-Caron, PhD., Biochimie Cellulaire des Hémopathies Lymphoïdes et des Vascularites, UFR Léonard de Vinci, 74 rue Marcel Cachin, F-93017 Bobigny Cedex, France
E-mail: caron@smbh.univ-paris13.fr
Fax: +33-1-4838-7777

Abbreviations: CA, correspondence analysis; **EBV**, Epstein Barr virus; **ID**, group identification number

the methodology used relies on the comparison of at least two maps. However, all qualitative or quantitative variations in protein expression visualized on maps should be representative of a biological phenomenon. Unfortunately, no two gels are exactly identical due to the versatility of the different technical steps, even when paying attention to details. The screening of numerous gels for each biological condition might minimize the part of the technical inhomogeneities. However, running numerous gels to analyze one sample is expensive, time-consuming, and inapplicable when the biological material is rare. Therefore, the aim of this study was to test, using only two gels per sample, the usefulness of statistical tools (correspondence and differential analysis) implemented in Melanie II software [7, 8]. The analysis was based on a combination of an automatic computerized analysis, constituting an actual decision-making aid, and of an interactive visual validation, corresponding to the interpretation of computer propositions. The biological model was a cell system in which a functional repression of NFκB complexes in the cell cytoplasm was induced by an overexpression of the NFκB inhibitor named IκBα. The results obtained made it possible to discriminate between the samples overexpressing or not overexpressing IκBα, and highlighted potential target gene products of NFκB factors that deserve further investigation.

2 Materials and methods

2.1 Apparatus and chemicals

The equipment for IEF and horizontal SDS-PAGE (Multiphor II electrophoresis chamber, reswelling tray, Immobiline strip tray, Multidrive XL 3500 programmable and EPS-500 power supplies) as well as the linear Immobiline dry strips (3 mm wide and 176 ± 4 mm long), pH 3–10, Excel gel XL SDS gradient 12–14%, Excel gel buffer strips, dithiothreitol (DTT), β-mercaptoethanol, carrier ampholyte mixture (Pharmalyte 3–10) and low molecular weight calibration kit were from Amersham Pharmacia-Biotech (Orsay, France). The GS-700 densitometer and Melanie II software release 2.2, as well as the urea and p*I* calibration markers, were from Bio-Rad (Ivry-sur-Seine, France). All other chemicals were purchased from Sigma and Fluka (Saint-Quentin Fallavier, France).

2.2 Cell culture model systems

In an effort to achieve a more reliable and objective identification of NFκB target genes in EBV-infected cells, we have subcloned the NFκB inhibitor IκBα cDNA mutated on serines 32 and 36 (IκB 32/36 A) into a mini EBV vector containing the doxycycline regulatory sequences [24, 25]. This construct was stably transfected in a classical lymphoblastoid cell line named PRI. Cells were grown in RPMI 1640 supplemented with 20% heat-treated fetal calf serum, 100 UI/mL streptomycin B, hygromycin (240 μg/mL) and 200 mM glutamine (medium A), in humidified atmosphere of 6% CO_2 in air, at 37°C for 3 days. Then, the cells were centrifuged at 900 × *g*, and seeded at 10^6 cells/mL in the same medium supplemented with 10% FCS instead of 20%, 120 μg/mL hygromycin instead of 240 μg/mL (medium B), in 75 cm^2 flasks (Costar, Strasbourg, France). For overexpression of IκBα, the cells were grown in medium B containing 2 μg/mL vibramycin, and the cells were collected after 48 h.

2.3 Protein extraction

Cells were centrifuged at 2000 rpm, washed in ice-cold PBS 4 × 10 min, resuspended in PBS, and then counted. The number of cells was adjusted to 10^8 cells/mL. Soluble proteins were extracted with buffer containing 50 mM Tris-HCl, pH 7.4, 10 mM EDTA, 100 mM lactose, 65 mM DTT, 1.5 mM phenylmethylsulfonyl fluoride (PMSF), and one tablet of anti-proteases (Pierce, Rockford, IL, USA) for 10 mL of buffer. Cells were homogenized with a potter in ice-cold buffer. After 1 h in ice, the homogenate was centrifuged at 40 000 × *g* for 1 h 30 min at 4°C. The supernatant was collected and the protein concentration was determined by the Bradford method [26]. Then, the supernatant was supplemented with 7 M urea, 2 M thiourea, 2% w/v CHAPS, and 32 mM DTT, and aliquots were stored at –20°C until use. For this study two independent sets of cultures were extracted at intervals of one week. Four samples were obtained: [1+IkB] and [2+IkB] correspond to cells overexpressing IκBα, extracted the first and the second week, respectively; [1–IkB] and [2–IkB] correspond to cells without induction of IκBα overexpression, extracted the first and the second week, respectively.

2.4 Two-dimensional gel electrophoresis

All manipulations were performed under a laminar flux, wearing gloves and mask, to prevent any contamination. The protein samples (25 μL of soluble extract) were mixed with 425 μL of 8 M urea, 1 M thiourea, 0.5% v/v Triton X-100, 0.8% v/v Pharmalytes, 15 mM DTT, 2% w/v CHAPS and a trace of bromophenol blue; next, they were in-gel applicated during the reswelling of dry immobilized pH gradient (IPG) strips, using a reswelling tray. IPG strips were covered with 2 mL of silicon oil to avoid evaporation. Strips were then left at room temperature overnight. The rehydrated IPG strips containing protein samples were soaked on a wet piece of Whatman paper, then positioned into the grooves of the strip tray on the Multi-

phor II electrophoresis chamber, as described by the manufacturer. Focusing was carried out for a total of 50 000 Vh. Immediately after IEF, IPG strips were equilibrated for 2 × 30 min under gentle shaking in 50 mM Tris-HCl, pH 8.6, containing 6 M urea, 1% w/v SDS, 65 mM DTT, 30% v/v glycerol, and a trace of bromophenol blue. Iodoacetamide (53 mM) was added to the second equilibration solution instead of DTT. IPG strips not used immediately were frozen at –20°C prior to equilibration. 12–14% Excel gels, (24 cm wide, 18 cm long), were used for second-dimensional electrophoresis. The electrode buffer was contained in acrylamide SDS buffer strips. The equilibrated IPG strip was transferred to a horizontal SDS gel. Fifteen μL of calibration markers were applied at each end of the IPG strip. Electrophoresis was carried out at 40 V for 40 min; then the voltage was increased to 1000 V, 45 mA for 20 min. The IPG strip was removed to prevent precipitation of proteins, and the cathodic buffer strips was moved forward to cover the area of the removed IPG strip. Electrophoresis was continued for 3 h 45 min at 1000 V, 45 mA. The gels were silver stained according to [27].

2.5 Gel analysis

Silver-stained gels were scanned using a fluorescent densitometer which generated digitalized images, linked to an Apple Power Macintosh computer. The computer image analysis, as well as the statistical data analysis, was carried out using Melanie II software release 2.2, allowing automatic detection and quantification of protein spots, as well as the resizing, alignment, and matching between our different bidimensional electrophoresis gels and the maps stored in the SWISS-2D-PAGE database [10, 28]. Spot volume of each protein was computed by considering spot area and spot optical density. The quantification of each spot was expressed as percent volume (% V), where % V = spot volume/Σ volumes of all spots resolved in the gel [7]. The p*I* and M_r were calibrated for each gel using the positions of identified proteins (β-actin major spot: p*I* 5.26/M_r 41 900; galectin 1: p*I* 5.1/M_r 14 500; ubiquitin: p*I* 7.16/M_r 8500) or p*I* and M_r calibration markers as described [12, 29]. Statistical data analysis was performed on groups of features in sets of gels. Variations, differences, and similarities were computed. Classes of gels were automatically created by correspondence analysis (CA) [8, 30]. Briefly, this statistical method based on data variation and their standard deviations creates a projection onto a factorial space of small dimensionality, in which gels and their significant spots appear as points [8]. Therefore, it is easy to visually determine which gels are similar and which are the most significant proteins in these gels. The Student T-test, performed on two given classes of gels, permits the selection of groups of proteins whose T values for one class are larger than the chosen threshold for the other class. This leads to the separation of two classes with a T-value percent of confidence [31]. This test is meaningful when more than two gels per class are analyzed [7].

3 Results

3.1 Reproducibility of protein extraction

In order to determine the reproducibility of the extraction step, we performed two sets of experiments at intervals of one week. Soluble protein extracts were prepared as described in Section 2.3. The amount of soluble proteins extracted from 10^8 cells was determined by the Bradford method. No significant differences were observed between the protein amount extracted from the cells overexpressing or not overexpressing IκBα (Student T-test: $p = 0.40$ and P variance = 0.017; Table 1). As the reproducibility of the extraction was excellent, we ran four gels on 3–10 IPG strips corresponding to the four extracts [1+IkB], [2+IkB], [1–IkB], and [2–IkB].

3.2 Establishing an automatic classification of the gels

The four silver-stained gels were digitalized and processed with Melanie II software. After the background was eliminated from the image, the features were automatically detected using a Laplacian algorithm [8]. Some manual adjustments were made to improve automatic detection. They consisted of elimination of molecular markers detected as features, and of separation of some adjacent spots detected as unique by the computer. The protein maps obtained are shown on Figure 1. The numbers of spots quantified were 438 for [1+IkB], 500 for [2+IkB], 406 for [1–IkB], and 612 for [2–IkB]. The molecular mass range covered was from approximately slightly more than 110 kDa to about 7 kDa. Resolved proteins spanned a p*I* from about 4.0 to about 8.0. Polypeptides with p*I* > 8 were poorly focalized in the pH range 3–10 used for running IEF.

Table 1. Amount of soluble proteins (mg/mL) extracted from 10^8 cells overexpressing (+IkB) or not overexpressing (–IkB) IκB α, in two independent sets of experiment

Cells	Experiment sets		
	No. 1	No. 2	Mean value
+IkB	4.85	4.60	4.72 ± 0.17
–IkB	4.22	4.70	4.46 ± 0.33

Determined by the Bradford method

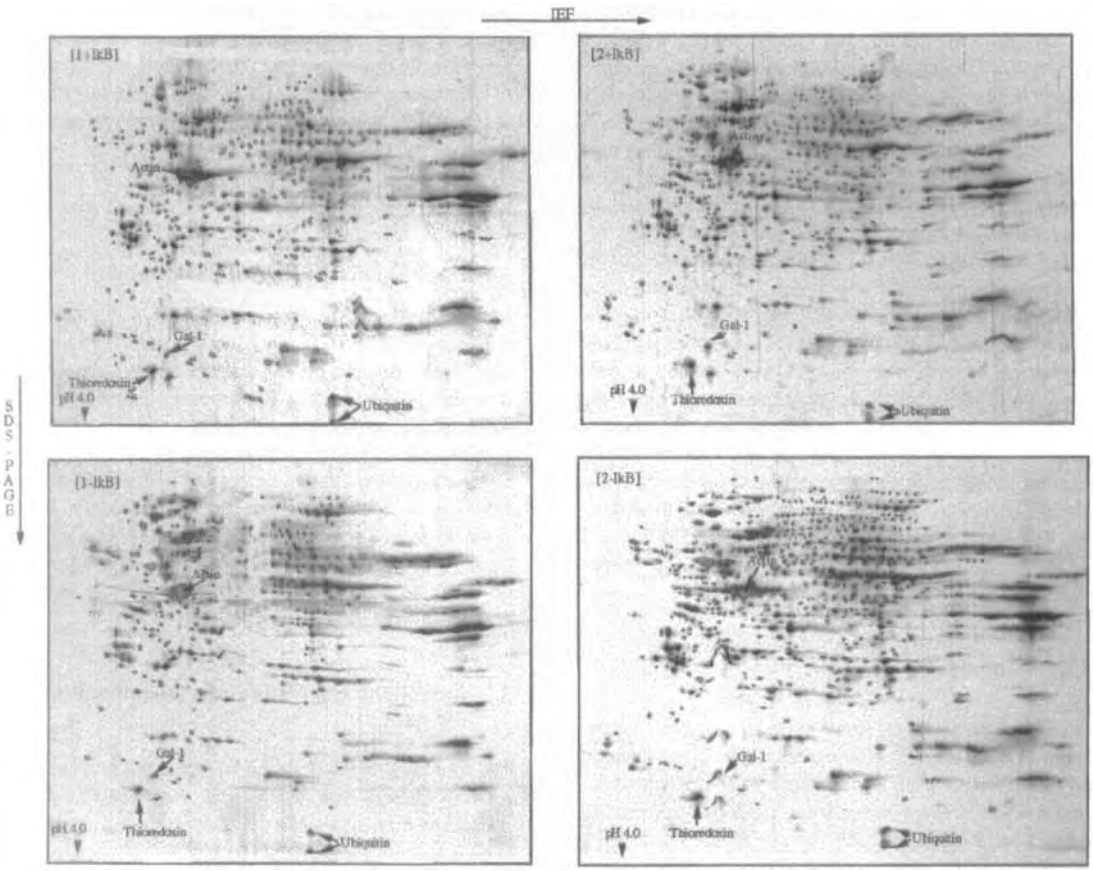


Figure 1. Silver-stained analytical 2-D gels containing soluble proteins of PRI lymphoblastoid cell line overexpressing [+IkB] or not overexpressing IκBα [–IkB]. Two sets of gels were run, corresponding to two independent experiments [1+IkB], [2+IkB], [1–IkB], and [2–IkB]. Soluble proteins from 2.5×10^6 cells were first separated by isoelectric point on 18 cm strips over a pH gradient of 3–10, and then by their relative mobility in 12–14% polyacrylamide with SDS. Gels were stained, scanned with a fluorescent densitometer, and image analysis was done using Melanie II software (release 2.2). Coordinates of identified proteins indicated by arrows on the maps: p*I*/M_r:β-actin, 5.26/41 800; galectin 1, 5.1/14 500; thioredoxin, 4.85/12 420; ubiquitin, 7.16/8500.

For aligning and matching of the four gels, control points were created at specific positions on the gel images. The matching between two gels resulted in an association of corresponding spots forming pairs. Each gel being used alternatively as a reference gel (RG), groups were automatically created when matching the other gels with the RG. Depending on the gel used as RG, the number of groups varied. When [2–IkB] was used as RG, 496 groups were generated, while only 439 were generated using [2+IkB] as RG. This difference was related to the absolute number of spots detected in the RG.

The visual task of comparing gels on the basis of the number of spots was critical and even impossible. Then, to emphasize without *a priori* the natural formation of classes among the four gels, a computerized classification of [1+IkB], [1–IkB], [2+IkB] and [2–IkB] gels was done by CA implemented in Melanie II. Figure 2 shows the graphic representation of CA obtained. There was clear evidence of two classes of gels in the projection into factorial space, namely class A composed of samples [1+IkB] and [2+IkB], and class B composed of [1–Ikb] and [2–IkB]. The dispersion of the points demonstrates that the gels belonging to the same class were not exactly identical as a consequence of the technical process. However, these defects did not affect the correct discrimination between the samples on the basis of IκBα overexpression. The most important spots representative of a

particular class of gels are also visualized on Fig. 2. Table 2 is the result of sorting the spots by their contributions to the first factorial axis, that is, the spots considered the most significant by the computer. Further, we examined the spots that the computer has deemed representative. Indeed, not all were pertinent. Some, such as No. ID 208, 7, or 283, were spots with a high volume value related to poor focalization. Conversely, spots No. ID 273, 25, 20 and 90 were expressed differently in the samples (Fig. 3). For example, spots 20 and 25 were clearly representative

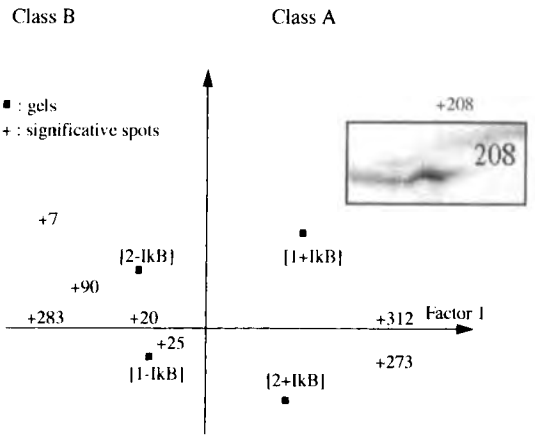


Figure 2. Correspondence analysis graphical results for the 4 gels represented by their names, and the 8 most significant spots represented by No. ID in [2+IkB] gel and "+" sign. The simultaneous projection of gels and spots into a 2-D factorial space allows dividing the gels in two classes, namely ([1–IkB], [2–IkB]), and ([1+IkB], [2+IkB]). Spot No. ID 208 is far from the others. Further interactive visual examination confirms that it is irrelevant to characterize the gel [1+IkB].

Table 2. Correspondence analysis data of [1+IkB], [2+IkB], [1–IkB], and [2–IkB] gels

Group ID	Contribution (%)
208	4.24
273	4.11
312	3.02
7	2.56
283	2.35
25	2.22
20	2.22
90	2.18

The eight most important spots (volume values), listed according to their absolute contribution to the first factorial (in percent) axis. The group identification number (ID) corresponds to the number provided by the reference gel chosen [2+IkB].

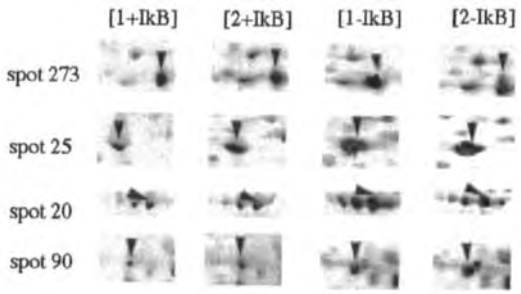


Figure 3. Montage showing the spots (arrows) that were detected by correspondence analysis as expressed differently in the two classes of gels. ID numbers are those in the [2+IkB] gel. The experimental coordinates (p*I*/M_r) determined on [2+IkB] for spot No. ID 273, 25, 20, 90 are: 5.07/47 768, 5.10/79 781, 7.23/84 116, 6.46/57 793, respectively.

of the gels of class B, namely [1–IkB] and [2–IkB]. Of the eight spots highlighted by the computerized classification, only half were actually meaningful for our purpose.

3.3 Variations in protein expression related to overexpression of IκBα

In a second step, the data of volume values obtained for the four gels were statistically analyzed using differential analysis (DA). DA is a simple method of quantifying variations in spot values. Actin was chosen as an internal reference for protein expression. The volume values of the major β-actin spot (No. ID 387 in reference gel [2–IkB]), as well as the relative abundance of this polypeptide, were close for all gels, as estimated by the value of the %Vol of this spot, *i.e.*, 2.49, 2.18, 2.30, and 2.02 for gels of class A and B, respectively. Then, after manual validation of the computer data, 53 polypeptides were considered as different in the two classes of gels (Tables 3, 4). Of these, 15 polypeptides were only expressed in samples in which IκBα was not overexpressed and, consequently, in which NFκB factors were active (Table 3). The overexpression of IκBα also induced a lower expression of 11 polypeptides (Table 4). Finally, only five polypeptides were specifically found in both samples overexpressing IκBα. There was a good correlation between the protein expression variations determined using the volume or using the %Vol values (data not shown). However, light variations in abundance of some given polypeptides were pointed out using the %Vol values. For example, an increased expression was specifically detected for four polypeptides in the samples corresponding to the gels of class A, namely the IκBα-overexpressing samples (Fig. 4A). Moreover, a polypeptde (coordinates: p*I* 4.56/ M_r 37 618) specifically detected in [2+IkB] gel yielded

0.03% of the total proteins expressed on the gel (Fig. 4B). The p*I* and M_r values of this spot were close to the theoretical p*I*/M_r (4.57/35 600) of IκBα protein indexed as P 25963 in the SWISS-PROT database. One of the best ways to study the variation in protein expression among our four gels was to produce synthetic gels by merging features from the studied gels. Therefore, two synthetic gels were created. The first one (S1) was produced by merging the groups common to [1+IkB] and [2+IkB] gels, and the second one (S2) was produced by merging the groups common to [1–IkB] and [2–IkB] gels. S1 and S2 gels were matched, and then groups were generated. By numeric inversion of spot selection, the spots that were not automatically paired were highlighted and numbered (Fig. 5). Among them, we found the 16 spots reported in Table 3 that were previously highlighted by the differential volume analysis. Figure 6 shows some of these representative spots that were expressed only in samples without IκBα overexpression.

4 Discussion

The aim of this study was to examine, using only two gels per sample, the usefulness of statistical tools implemented in Melanie II software [7, 8] to evidence variations in protein expression related to an inhibition of NFκB complexes in lymphoblastoid cells. The strategy applied was based on a combination of an automatic computerized analysis, constituting an actual aid for deciding, and of an interactive visual validation, corresponding to the interpretation of the computer propositions. This is a rapid way to detect objectively potential variations in protein expression before addressing the question of characterization of the variant polypeptides.

Table 3. Proteins expressed differently in the two classes of gels using [2–IkB] as reference gel

Group ID[a]	*T* value[b]	[1+IkB](A) Vol. value[c]	[2+IkB](A) Vol. value	[1–IkB](B) Vol. value	[2–IkB](B) Vol. value	p*I*	M_r
1*)	0.882	0.000	0.000	0.019	0.009	5.34	113917
10*)	0.834	0.000	0.000	0.027	0.010	5.21	99761
23*)	0.852	0.000	0.000	0.010	0.025	5.11	94864
24	0.940	0.042	0.066	0.116	0.148	5.13	94864
29*)	0.916	0.000	0.000	0.020	0.037	4.92	95299
59*)	0.895	0.000	0.000	0.032	0.067	4.80	88167
66*)	0.838	0.000	0.000	0.227	0.165	6.53	85780
81	0.993	0.025	0.023	0.109	0.096	5.51	80090
85	0.946	0.003	0.010	0.021	0.025	5.47	79724
90	0.959	0.017	0.000	0.074	0.057	5.97	78637
110	0.969	0.006	0.000	0.031	0.041	5.53	74436
112*)	0.902	0.000	0.000	0.011	0.005	5.23	74096
118	0.816	0.014	0.042	0.084	0.188	5.11	71760
122	0.871	0.073	0.157	0.216	0.283	5.41	70782
131	0.901	0.020	0.028	0.091	0.159	5.45	69498
139*)	0.978	0.000	0.000	0.326	0.440	5.34	68865
184	0.842	0.043	0.096	0.132	0.200	5.07	64594
189	0.896	0.095	0.179	0.261	0.339	5.41	65006
190	0.876	0.040	0.031	0.049	0.061	5.36	64006
217*)	0.991	0.000	0.000	0.067	0.081	5.32	62846
231*)	0.881	0.000	0.000	0.028	0.063	5.24	61425
232*)	0.948	0.000	0.000	0.126	0.204	5.28	61145
269	0.807	0.000	0.000	0.057	0.018	4.34	56573
350	0.821	0.494	0.450	0.045	0.324	6.07	47768
451*)	0.971	0.000	0.000	0.025	0.017	5.16	41693
456*)	0.916	0.000	0.000	0.025	0.047	5.53	41693
500	0.952	0.025	0.000	0.041	0.066	5.12	37618
536	0.813	0.027	0.000	0.040	0.077	5.39	35610
594*)	0.952	0.000	0.000	0.027	0.043	5.20	29278
658*)	0.999	0.000	0.000	0.019	0.020	6.07	24226

a) ID number in [2–IkB] gel

b) The Student T-test was performed on the two classes of gels; the range of confidence chosen was > 80%

c) The volume of each spot was determined using (OD × area)

*) Asterisks highlight the spots specifically detected in gels of class B

Table 4. Proteins expressed differently in the two classes of gels using [2+IkB] as reference gel

Group ID[a)]	*T* value[b)]	[1+IkB](A) Vol. value[c)]	[2+IkB](A) Vol. value	[1–IkB](B) Vol. value	[2–IkB](B) Vol. value	p*I*	M_r
2	0.805	0.000	0.005	0.006	0.010	7.23	00107
18[d)]	0.943	0.009	0.016	0.048	0.071	6.88	85915
20	0.903	0.205	0.361	0.583	0.504	7.23	84116
21	0.940	0.064	0.090	0.250	0.183	6.98	85010
30	0.918	0.000	0.008	0.027	0.044	7.22	78525
40	0.951	0.042	0.042	0.107	0.146	6.65	70267
42*)	0.967	0.011	0.016	0.000	0.000	5.09	70640
43	0.991	0.012	0.017	0.049	0.045	6.54	69897
90	0.808	0.100	0.113	0.150	0.239	6.46	57793
91	0.842	0.043	0.096	0.132	0.200	5.02	57526
110	0.973	0.037	0.039	0.074	0.088	7.23	55696
112	0.882	0.126	0.137	0.208	0.299	7.09	55696
124	0.935	0.000	0.047	0.132	0.074	5.48	54676
148	0.986	0.066	0.056	0.236	0.200	7.41	51489
161	0.868	0.046	0.059	0.080	0.114	7.23	49621
164	0.898	0.000	0.030	0.056	0.062	7.07	49621
195	0.855	0.125	0.099	0.045	0.079	6.67	43600
197	0.887	0.494	0.450	0.161	0.324	6.92	43000
198*)	0.811	0.119	0.039	0.000	0.000	7.12	44005
229*)	0.887	0.012	0.025	0.000	0.000	7.57	42573
363*)	0.857	0.014	0.035	0.000	0.000	6.75	28121
387	0.967	0.020	0.029	0.000	0.000	9.01	26604
402*)	0.965	0.054	0.036	0.000	0.000	6.61	25402

a) ID number in [2–IkB] gel
b) The Student T-test was performed on the two classes of gels; the range of confidence chosen was > 80%
c) The volume of each spot was determined using (OD × area)
d) Italic characters point out spots showing a lower expression in gels of class A
*) Asterisks highlight the spots specifically detected in gels of class B

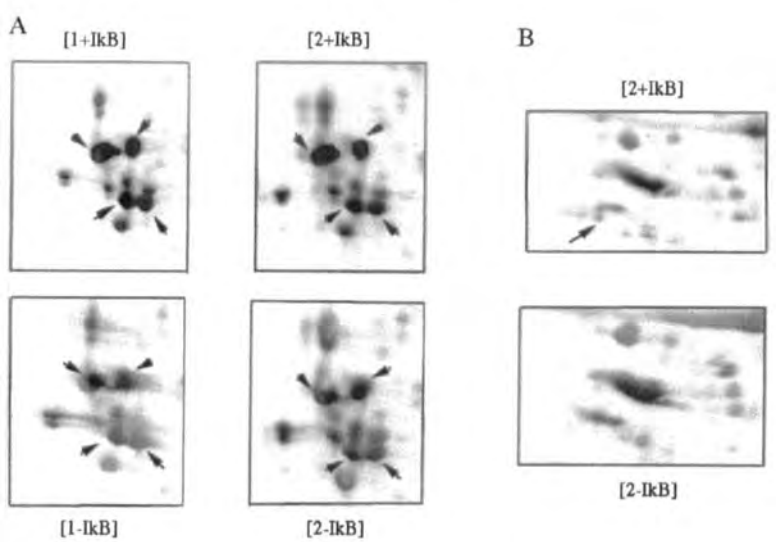


Figure 4. Enlargements of two regions of 2-D maps obtained with samples [1–IkB], [2–IkB], [1+IkB], and [2+IkB]. Variation in abundance was pointed out using the %Vol values. (A) An increased expression was specifically detected for 4 polypeptides in the IκBα-overexpressing samples. (B) A polypeptide (coordinates: p*I* 4.56, M_r 37 618) was specifically detected in [2+IkB]. The p*I* and M_r values of this spot were close to the theorical p*I*/M_r (4.57/35 600) of IκB protein indexed as P 25963 in SWISS-PROT database.

Our data demonstrate the usefulness of the correspondence and differential analysis for analyzing variations of protein expression. We first verified that the standardization of the protein extraction step was satisfying, and we demonstrated that no noticeable difference in protein content of the samples was generated in two sets of experiments. To detect differences between two samples, conventional 2-D PAGE methodology relies on the comparison of at least two different gels. Indeed, qualitative and *a fortiori* quantitative inter-gel comparisons of proteins separated by 2-D PAGE present a number of problems that may arise from mechanical handling of samples, protein loss during IEF or SDS-PAGE, variations in staining intensity, *etc.* So far, to circumvent the problems of reproducibility, investigators are faced with the run of many gels of the same sample, the use of alternative methods such as difference gel electrophoresis [32], or the use of normalization algorithm [33]. The emergence of third-generation software such as Melanie II release 2.2 has recently allowed the quantification of differences within groups of proteins and among gels through the use of statistical methods having complementary capabilities [7, 8]. Although constitutive proteins such as β-actin were found constantly expressed in the two experimental sets run, the number of spots varied from one sample to another, depending on the technical process used. Therefore, it was actually impossible to classify the four gels on the basis of the number of spots. To address the question of gel classification, CA was successfully employed in this work. It provided a clustering of the four gels on the basis of the overexpression of IκBα without *a priori*, and despite the presence of some technical defects. In this study the data of the CA was obtained with eight groups; however, the automatic discrimination between [IkB+] or [IkB–] samples remained valuable when a greater number of groups was displayed (data not shown). Four polypeptides highlighted by the computerized classification were actually meaningful for our purpose. In spite of the few

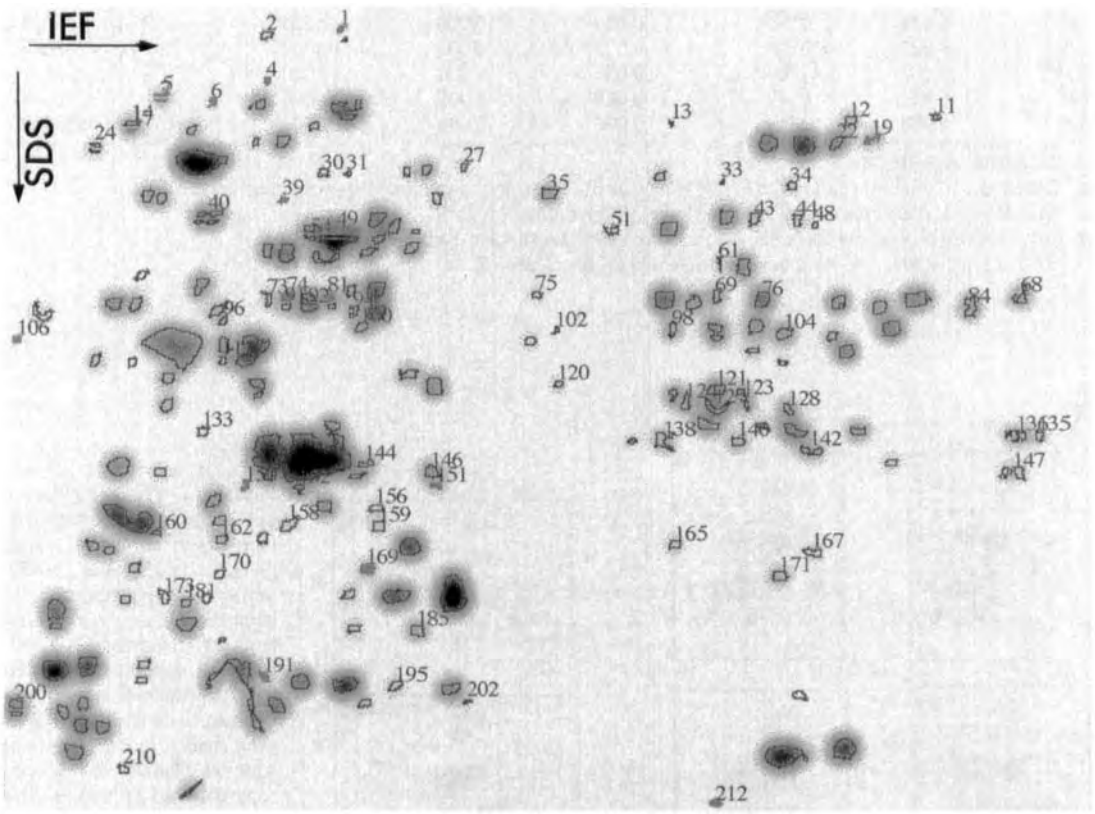


Figure 5. The synthetic map representing the spots highlighted by numeric inversion of selection of protein patterns produced by merging the groups common to four gels is displayed. The spots that were not automatically paired were highlighted and numbered. The 15 polypeptides occurring only in samples where IκBα was not overexpressed are shown in green. These spots are candidates for further analysis as putative target gene products of NFκB.

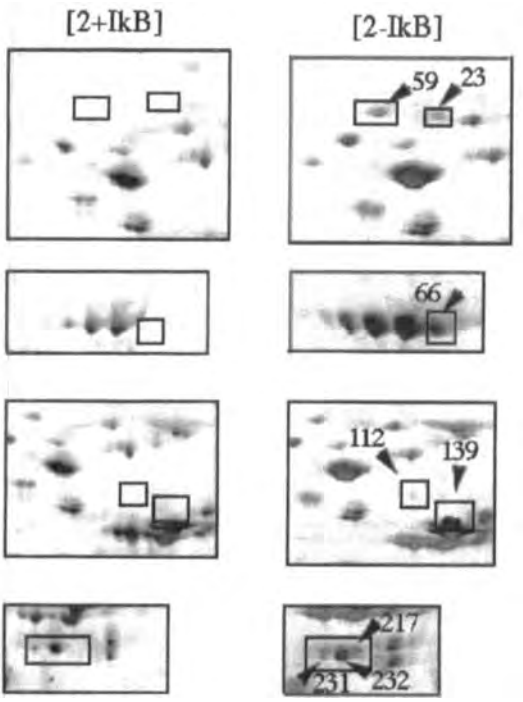


Figure 6. Enlargements of some regions of 2-D maps obtained with samples [2+IkB] and [2–IkB]. The displayed portions show 8 of the 15 polypeptides specifically expressed in samples where IκBα was not overexpressed. These spots are indicated by arrowheads in framed regions; the same regions framed in [2+IkB] appear empty of spots.

number of gels compared, the differential analysis with class set pointed out 15 polypeptides that were specifically detected in samples in which NFκB was not inhibited. From synthetic gels the same spots also emerged as specific of the samples without any IκBα overexpression. In the cell constructs employed in this study, *i.e.*, lymphoblastoid cells containing a mini EBV cassette [24, 25], NFκB complexes were always activated in absence of overexpression of IκBα. The overexpression of IκBα was controlled by tetracycline (vibramycin). Tetracycline operon was currently chosen as an efficient regulatory system in mammalian cells because it is well established that it has no side effects on eucaryotic cells [34, 35]. The overexpression of IκBα led to a dose-dependent inhibition of NFκB transcription activity (data not shown). Therefore, the proteins that were specifically detected in samples in which NFκB complexes were active are candidates as putative target gene products of NFκB. It will be necessary to proceed further in the identification of these polypeptides to reach a definite conclusion. With this object, further work including 2-D PAGE on narrow-range pH gradients and characterization of differently expressed polypeptides by mass spectrometry analysis is in progress.

This work was supported by grants from the Ministère de l'Education Nationale de la Recherche et de la Technologie (MENRT), from the Association pour la Recherche contre le Cancer (ARC, grant 9348) and from ANRS (Paris, France), grant 98017.

Received September 1, 1998

5 References

[1] Huen, D. S., Henderson, S. A., Croom-Carter, D. Rowe, M., *Oncogene* 1995, *10*, 549–560.

[2] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[3] Görg, A., *Nature* 1991, *349*, 545–546.

[4] Görg, A., Fawcett, J. S., Chrambach, A., *Adv. Electrophoresis* 1988, *2*, 1–43.

[5] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1994, *15*, 1205–1211.

[6] Righetti, P. G., *Immobilized pH gradients. Theory and Methodology*, Elsevier, Amsterdam 1990.

[7] Appel, R. D., Palagi, P. M., Walther, D., Vargas, J. R., Sanchez, J.-C., Ravier, F., Pasquali, C., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2724–2734.

[8] Appel, R. D., Vargas, J. R., Palagi, P. M., Walther, D., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 2735–2748.

[9] Klose, J., Feller, M., *Electrophoresis* 1981, *2*, 12–24.

[10] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.

[11] Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Gromov, P., Vorum, H., Vassilev, A., Baskin, Y., Lin, X., Celis, A., Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kjaergaard, I., Andersen, I., Puype, M., Van Damme, J., Vanderkerckhove, J., *Electrophoresis* 1994, *15*, 1349–1458.

[12] Lutomski, D., Caron, M., Cornillot, J.-D., Bourin, P., Dupuy, C., Pontet, M., Bladier, D., Joubert-Caron, R., *Electrophoresis* 1996, *17*, 600–606.

[13] URL: http://expasy.hcuge.ch/ch2d/

[14] URL: http://www.anl.gov/CMB/PMG/

[15] URL: http://WWW.LSBC.COM/2dmaps/patterns.htm

[16] URL: http://biobase.dk/cgi-bin/celis

[17] URL: http://www.harefield.nthames.nhs.uk/nhli/protein/

[18] URL: http://www.mdc-berlin.de/~emu/heart/

[19] URL: http://www-lecb.ncifcrf.gov/PDD/

[20] URL: http://iupucbio1.iupui.edu/franw/molan.htm

[21] URL: http://www.ludwig.edu.au/www/jpsl/jpslhome.html

[22] URL: http://www.tmig.or.jg/2D/2D_Home.html

[23] URL: http://www-smbh.univ-paris13.fr/biochemistry/biochimie.htm

[24] Polack, A., Hortnagel, K., Pajic, A., Christoph, B., Baier, B., Falk, M., Mautner, J., Geltinger, C., Bornkamm, G. W., Kempkes, B., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 10411–10416.

[25] Mucke, S., Polack, A., Pawlita, M., Zehnpfennig, D., Massoudi, N., Bohlen, H., Doerfler, W., Bornkamm, G., Diehl, V., Wolf, J., *Gene Ther.* 1997, *4*, 82–92.

[26] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 335–340.

[27] Heukeshoven, J., Dernick, R., *Electrophoresis* 1985, *6*, 103–112.

[28] Hoogland, C., Sanchez, J.-C., Tonella, L., Bairoch, A., Hochstrasser, D. F., Appel, R. D., *Nucleic Acids Res.* 1998, *26*, 332–333.

[29] Lutomski, D., Fouillit, M., Bourin, P., Mellottee, D., Denize, N., Pontet, M., Bladier, D., Caron, M., Joubert-Caron, R., *Glycobiology* 1997, *7*, 1193–1199.

[30] Pun, T., Hochstrasser, D. F., Appel, R. D., Funk, M., Villars-Augsburger, V., Pellegrini, C., *Appl. Theor. Electrophor.* 1988, *1*, 3–9.

[31] Armitage, P., Berry, G., *Statistical Methods and Biomedical Research*, Blackwell Scientific Publications, Oxford 1987.

[32] Unli, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071–2077.

[33] Merril, C. L., Creed, G. J., Joy, J., Olson, A. D., *Appl. Theor. Electrophor.* 1993, *3*, 329–333.

[34] Gossen, M., Bujard, H., *Proc. Natl. Acad. Sci. USA* 1992, *89*, 5547–5551.

[35] Furth, P., St. Onge, L., Böger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., Hennighausen, L., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 9302–9306.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1027–1032

Christina Nock[1,2]
Christine Gauss[1*]
Leonard C. Schalkwyk[1]
Joachim Klose[2]
Hans Lehrach[1]
Heinz Himmelbauer[1]

[1]Max-Planck-Institute for Molecular Genetics, Berlin-Dahlem, Germany
[2]Humboldt-Universität, Charité, Virchow-Klinikum, Institut für Humangenetik, Berlin, Germany

Technology development at the interface of proteome research and genomics: Mapping nonpolymorphic proteins on the physical map of mouse chromosomes

Data obtained from protein spots by peptide mass fingerprinting are used to identify the corresponding genes in sequence databases. The relevant cDNAs are obtained as clones from the Integrated Molecular Analysis of Genome Expression (I.M.A.G.E.) consortium. Mapping of I.M.A.G.E. clones is performed in two steps: first, cDNA clones are hybridized against a 10-hit genomic mouse bacterial artificial chromosome (BAC) library. Second, interspersed repetitive sequence polymerase chain reaction (IRS-PCR) using a single primer directed against the mouse B1 repeat element is performed on BACs. As each cDNA detects several BACs, and each individual BAC has a 50% chance to recover an IRS-PCR fragment, the majority of cDNAs produce at least a single IRS-PCR fragment. Individual IRS fragments are hybridized against high-density spotted filter grids containing the three-dimensional permutated pools of yeast artificial chromosome (YAC) library resources that are currently being used to construct a physical map of the mouse genome. IRS fragments that hybridize to YAC clones already placed into contigs immediately provide highly precise map positions. This technology therefore is able to draw links between proteins detected by 2-D gel electrophoresis and the corresponding gene loci in the mouse genome.

Keywords: Mouse / Proteome / Physical map / Map integration / Interspersed repetitive sequence polymerase chain reaction
EL 3407

1 Introduction

Protein extracts isolated from different tissues, developmental stages, or organ structures that have undergone pathological alterations will reveal differences in protein composition as observed by two-dimensional gel electrophoresis (2-DE) [1, 2]. Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) has become an established procedure to provide peptide mass fingerprints from protein spots excised from a gel that may allow the identification of a cognate DNA sequence in a database search [3, 4]. Once it is known from which gene a protein spot is derived, it is important to determine the chromosomal position to link the phenotype examined on 2-DE to a genomic region. Additionally, this will allow the integration with existing mouse maps in order to determine if a gene is a candidate for a genetically mapped mouse mutation. Knowledge of the map position in the mouse genome also allows one to predict the map location of the corresponding human homolog. With this information at hand, possible links to human disease loci can be explored.

Proteins that display strain-specific mobility variations or intensity differences are amenable to genetic mapping on segregating F2 or N2 populations. Nonpolymorphic proteins can not be subjected to this type of analysis and need to be brought into context with the genome maps *via* a DNA-based approach.

At the time of writing (Sept. 1998) 2000 different genes have been mapped in the mouse according to the UniGene data release [5] which corresponds to 2% of the genes in the mouse genome. At the same time, the mouse expressed sequence tag (EST) collection alone comprises 187 000 sequences that fall into 10 000 clusters [5]. The speed at which polypeptides are identified to be derived from novel genes clearly asks for a mapping technique that is able to match this pace. Here we present a simple strategy for high-precision gene mapping that exploits the current efforts to construct a yeast artificial chromosome (YAC)-based physical map of the mouse

Correspondence: Dr. Heinz Himmelbauer, Max-Planck-Institute for Molecular Genetics, Ihnestr. 73, D-14195 Berlin-Dahlem
E-mail: himmelbauer@pop3.mpimg-berlin-dahlem.mpg.de
Fax: +49-30-84131380

Abbreviations: **BAC**, bacterial artificial chromosome; **I.M.A.G.E.**, Integrated Molecular Analysis of Genome Expression; **IRS-PCR**, interspersed repetitive sequence polymerase chain reaction

* Current address: Toplab, Frauenhoferstr. 18A, 82152 Martinsried, Germany

genome at the Whitehead Institute [6] and in our unit. Mapping a gene onto YACs can be achieved using either a PCR- or a hybridization-based approach. PCR that employs gene-specific primers on purified YAC-DNA in the format of three-dimensional (3-D) pools [7, 10] has limitations in that a large number of PCR reactions with discrete pairs of primers have to be performed in order to comprehensively screen current YAC library resources. YAC identification by hybridization of a cDNA probe against YAC colony filters is difficult, mostly because little target DNA is retained on the membrane. These drawbacks can be avoided with the introduction of the interspersed repetitive sequence (IRS) PCR technique [8–11]. IRS-PCR is employed to amplify fragments flanked by repeat elements and can be conducted on any genomic DNA template including bacterial artificial chromosome (BAC) clones, YACs and YAC pools. In our set-up (Fig. 1), a cDNA is first used to identify corresponding genomic BAC clones by hybridization and to subsequently use the clones as a substrate for IRS-PCR. Fragments generated can then be hybridized against filter grids that contain the IRS-PCR amplification products of individual YAC pools. Using the appropriate software tools, YACs identified can be matched against the existing physical map framework [6, 12].

In the current study we chose to demonstrate the validity of the approach on two protein spots identified as the products of the mouse *Alb1* (Albumin 1) and *Nfl* (Neurofilament light peptide) loci, because the chromosomal map positions of both genes were well established through the work of other investigators [13–16].

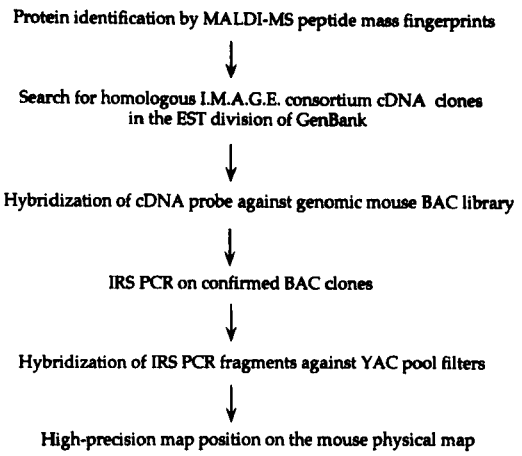


Figure 1. Flow diagram to describe the strategy for drawing links between nonpolymorphic proteins detected by 2-D gel electrophoresis and the corresponding gene loci on the mouse physical map.

2 Materials and methods

2.1 Experimental animals

Twelve-week-old female mice from the inbred strain C57BL/6 were obtained from The Jackson Laboratory (Bar Harbor, ME, USA).

2.2 Preparation of protein samples

Protein samples were prepared from single brains in three fractions: the buffer-soluble proteins, the urea/CHAPS-soluble proteins and the proteins remaining in the remaining pellet (DNAse-digested pellet suspension). The extraction procedure has been described in detail elsewhere [17]. The results presented here concern the buffer-soluble fraction.

2.3 2-D gel electrophoresis

Proteins were separated by using the large-gel 2-D electrophoresis technique recently described in detail [18, 19]. The gel format in this technique is 40 cm (IEF) × 30 cm (SDS-PAGE). Fifty μL (500 μg of protein sample) were loaded onto an IEF gel of 1.5 mm in diameter. SDS gels (1.5 mm) were stained with Coomassie blue according to Rosenfeld *et al.* ([20], "modified procedure").

2.4 Sample preparation for MALDI-MS

Coomassie blue-stained spots were destained with digestion buffer containing 40% acetonitrile. Protein spots were in-gel reduced, S-alkylated and in-gel digested with an excess of trypsin (6 h, 37°C) as described by Shevchenko *et al.* [21] except that the reduction was performed with tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) at room temperature. After digestion, a 0.3 μL aliquot from the 3–5 μL supernatant was used for MALDI-MS analysis.

2.5 Mass spectrometry

MALDI peptide mass fingerprint analyses were obtained on a Bruker Reflex II-MS (Bruker-Franzen Analytik, Bremen, Germany) equipped with a SCOUT multiprobe inlet and a gridless pulsed ion extraction source. Microcrystalline matrix surfaces for MALDI-MS containing nitrocellulose were made by the fast evaporation method according to Vorm *et al.* [22]. Aliquots of the tryptic digestion mixtures were mixed into a 0.5 μL droplet of 5% formic acid and rinsed with water as described by Jensen *et al.* [23].

2.6 Bioinformatics and protein spot nomenclature

The PeptideSearch software tool was used to identify proteins from peptide mass fingerprint data [3, 4, 24] at nrdb,

the nonredundant protein sequence database at the EBI (European Bioinformatics Institute, Hinxton, UK). Homology searches *via* blast or tblastn were carried out online as implemented on the NCBI web server [25]. Protein spots are named according to our Protein Standard Map [24, 26].

2.7 Clone libraries and filters

The mouse BAC library (B. Birren *et al.*, unpublished) was purchased from Research Genetics (Huntsville, AL, USA). Automation-assisted filter production and filter processing were as described [11]. A picking copy of the library was kept in-house. YAC pools from mouse YAC libraries 902, 903, 910, 917 [27] were produced in-house or obtained from HGMP Resource Center (Hinxton, UK). A detailed account on pool generation and the spotting procedure has either been presented [10] or will be given elsewhere (Schalkwyk *et al.*, unpublished). Integrated Molecular Analysis of Genome Expression (I.M.A.G.E.) clones were obtained from the RZPD Resource Center in Berlin, Germany [27].

2.8 IRS-PCR

Reactions with primer B1R (5′-AGTCCAGGACAGC-CAGGGCTAYACAGA-3′) were carried out either on BAC colony material or 1 ng of purified BAC-DNA in a buffer that contained 35 mM Tris base, 15 mM Tris-HCl, 0.1% Tween-20, 50 mM KCl, 1.5 mM $MgCl_2$, and 125 μM of each dNTP under conditions described previously [11].

2.9 Probe preparation and filter hybridization

Radioactive labeling was carried out at room temperature overnight using the random priming method [28]. BAC colony filters, YAC pool filters and Southern blots were hybridized in Church buffer [29] in a 65°C water bath overnight. Filters were washed in 2 × SSC/0.1% SDS and 0.1 × SSC/0.1% SDS for 15 min (BAC) or 60 min (YAC) each, at 65°C. Exposure to Kodak X-ray film was at room temperature or –80°C from overnight up to five days.

2.10 Miscellaneous techniques

BAC-DNA from liquid cultures was prepared on the PI-100S automatic plasmid isolation system (Kurabo, Osaka, Japan) according to the manufacturer's instructions. Restriction digests were done using standard protocols [30].

3 Results

3.1 Characterization of protein spots A4_199 and B2_017 and identification of corresponding cDNA and genomic clones

Mouse brain-derived protein spots A4_199 and B2_017 were retrieved from 2-D electrophoresis gels (Fig. 2) and the trypsin-digested peptides analyzed by MALDI-MS. The resulting peptide mass fingerprints were run against the nrdb database. At a sequence coverage of 29.9%, spot A4_199 was identified to be derived from the mouse albumin locus (*Alb1*). Protein spot B2_017 was identified

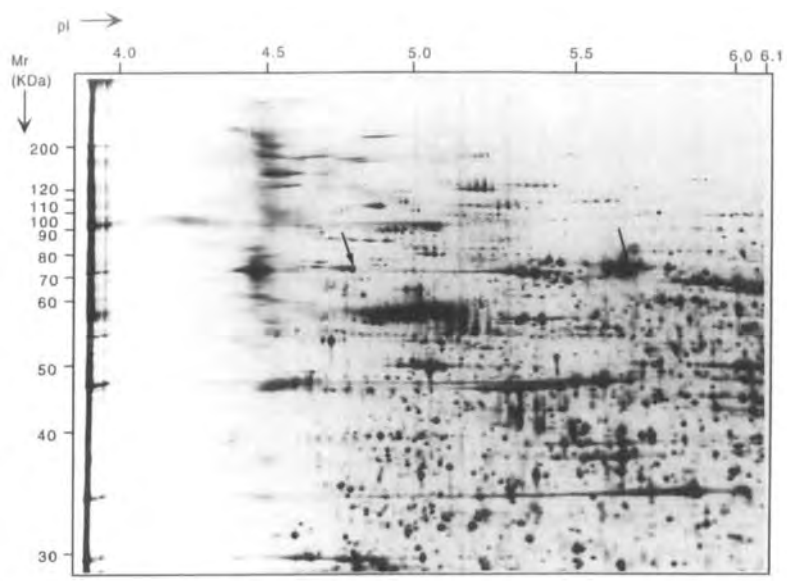


Figure 2. 2-D gel electrophoresis pattern of the mouse brain supernatant fraction. The complete protein standard map is shown on our homepage (URL: http://www.charite.de/humangenetik). The left arrow points to spot B2_017 (p*I* 4.79) and is identified as the neurofilament light peptide (*Nfl*; SWISS-PROT P08551). The arrow at p*I* 5.67 indicates spot A4_199, mouse serum albumin (*Alb1*; SWISS-PROT P07724).

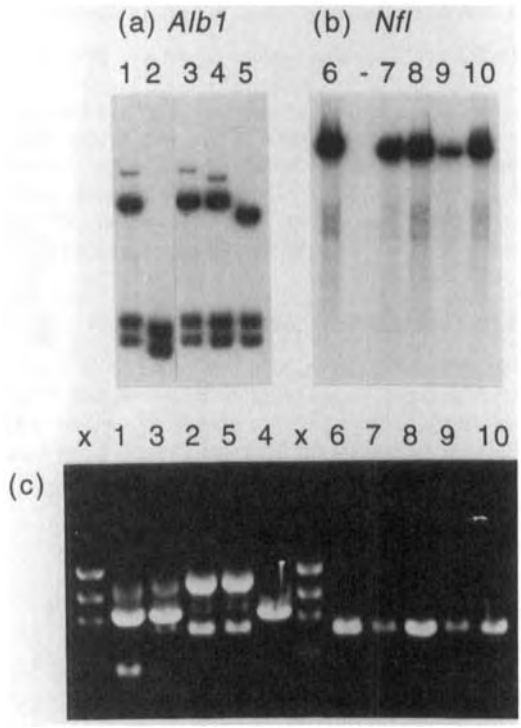


Figure 3. Characterization of mouse BAC clones from the *Alb1* and *Nfl* loci. (A) Restriction-digested DNA from BAC clones 1, 154b6; 2, 198g5; 3, 239h3; 4, 295n4; 5, 466j18 hybridized with the *Alb1* cDNA probe. (B) Restriction-digested DNA from BAC clones (6) 28e7, (7) 77g3, (8) 77m22, (9) 296n3, (10) 406k22 hybridized with the *Nfl* cDNA probe. (C) IRS-PCR performed on *Alb1*- and *Nfl*-containing BAC clones. (x) indicates the PhiX174/*Hae*III size standard.

as the product of the mouse neurofilament light peptide (*Nfl*) locus (sequence coverage 32.4%). The I.M.A.G.E. cDNA clones were identified by blast homology searches against the EST database [31] at NCBI [32]. For each *Alb1* and *Nfl*, a representative cDNA clone (I.M.A.G.E. clone 554173/IMAGp998h141333 for *Alb1* and I.M.A.G.E. clone 735764/IMAGp998f041002 for *Nfl*) was obtained from RZPD [27] and confirmed by sequence determination. Genomic clones from *Alb1* and *Nfl* were identified by hybridization of the respective cDNA probe against filter grids from a mouse BAC library (plates 1–504, corresponding to ten genome equivalents). The albumin cDNA detected 11 different BACs. Seven BAC clones were identified from the *Nfl* locus. Clones were confirmed by preparing Southern blots containing restricted clone DNA; these were subsequently probed with a labeled cDNA fragment from the *Alb1* or *Nfl* genes (Fig. 3a, b).

3.2 Mapping of mouse BAC clones on mouse YACs through IRS-PCR

To determine the position of *Alb1* and *Nfl* on the physical map, IRS-PCR was performed on the BAC clones identified (Fig. 3c). Four different IRS-PCR fragments could be distinguished at the *Alb1* locus, whereas a single IRS-PCR fragment common to all *Nfl*-derived BAC clones was observed. The fragments were hybridized against YAC pool filters that contained the IRS-PCR amplified pool products from libraries 910 and 917 (Whitehead I and II), and libraries 902 and 903 (ICRF I and II). The MBAC239h3-900 probe from the *Alb1* locus hybridized to 13 independent YAC clones. Integration with the Whitehead Institute YAC contig data accessible through the World Wide Web [6] showed that three of these clones (WIBRy917c07308, WIBRy917d01393, and WIBRy91703h434) had been placed into contig WC5.38 on mouse chromosome 5 (Fig. 4a). Each of these clones had been hit by microsatellite marker *D5Mit361*. The consensus map for mouse chromosome 5 assigns both *D5Mit361* and *Alb1* to Centimorgan position 50 on mouse chromosome 5 [33]. Our mapping result therefore is in excellent agreement with the existing *Alb1* gene mapping data.

Probe MBAC28e7-700 from the *Nfl* locus was hybridized against a YAC pool filter and detected a total of eleven different YAC clones. Only a single clone (WIBRy91705g417) provided a localization in contig WC14.25 on mouse chromosome 14 (Fig. 4b). However, five of the eleven YACs that were detected hybridized to probes derived from YAC clones included in contig WC14.25 (our unpublished results). Taken together, these data strongly point towards a location of *Nfl* in contig WC14.25. Included in this contig is marker *D14Mit28* which is placed at cM 28.5 on the consensus map [33]. This again perfectly matches the known position of *Nfl* at cM 28.5 on the consensus map.

4 Discussion

2-DE is a powerful tool in assessing changes in protein expression between different tissues, developmental stages and specimens derived from pathological conditions in comparison to control samples. Once a protein spot of interest has been identified to be derived from an unmapped gene, it is essential to know if this gene is a likely candidate gene for a locus of biomedical importance, for which a chromosomal assignment has been reported, but which in the absence of a suitable candidate gene had not yet been studied on the molecular level. In order to provide this mapping information, a method is required that is both fast, reliable, and highly precise. We

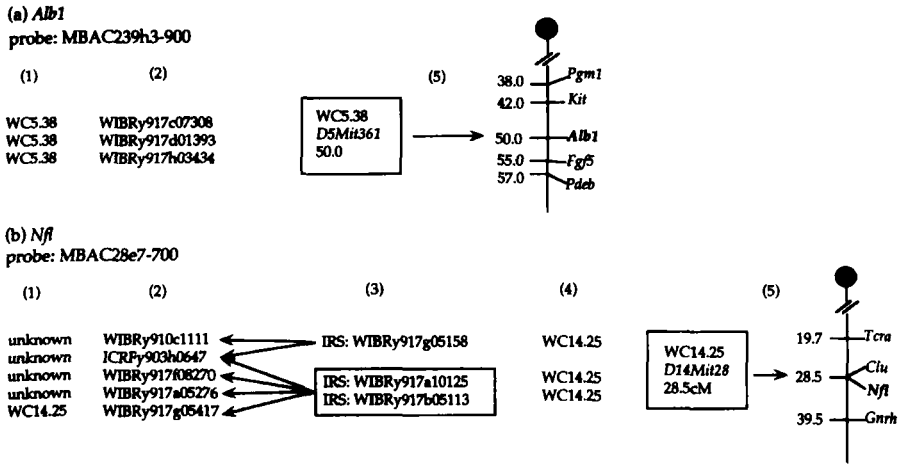


Figure 4. Mapping of (a) *Alb1* to mouse chromosome 5 and (b) *Nfl* to mouse chromosome 14. IRS-PCR probes amplified from the respective BAC clones from *Alb1* and *Nfl* are indicated. (1) Contig location according to Whitehead Institute data release [6]. (2) YAC clones detected by probes MBAC239h3-900 and MBAC28e7-700 relevant for the presented analysis. Each of the probes hybridized to additional YAC clones (see Section 3.2) which do not add further information to the analysis (not shown). (3) IRS fragments from indicated YAC clones that hybridize against YAC clones listed in column 2. These data are unpublished results of the Max-Planck-Institute mouse genome mapping project. (4) Physical map location of YAC clones in column 3 according to [6]. (5) Summary of the mapping result. Microsatellites *D5Mit361* and *D14Mit28* are located within the indicated contigs [6]. Centimorgan distances for microsatellite markers and genes are as displayed in the Chromosome Committee consensus maps [32]. *Pgm1*, phosphoglucomutase 1; *Kit*, kit oncogene; *Alb1*, albumin 1; *Fgf5*, fibroblast growth factor 5; *Pdeb*, phosphodiesterase, B subunit; *Tcra*, T-cell receptor alpha chain; *Clu* clusterin; *Nfl*, neurofilament, light polypeptide; *Gnrh*, gonadotropin-releasing hormone.

present such a approach exploiting the current efforts to produce a high-resolution physical map of the mouse genome and which is able to provide a link between protein identification and gene mapping.

For the purpose of demonstrating the feasibility of the proposed strategy, we chose the mouse genes *Alb1* and *Nfl* that had previously been mapped to mouse chromosomes 5 and 14, respectively. In both cases, experimental results obtained were in perfect agreement with the consensus maps of the relevant mouse chromosomes. In general, a mapping result is considered correct if a probe hits more than two independent YAC clones that have been mapped into the same contig. Our strategy makes use of ongoing efforts to generate physical YAC-based maps and, as such, can currently be applied to a number of organisms, including human. The map positions determined are highly precise at the resolution of a clone. The use of IRS-PCR provides both a cost-effective and efficient system for gene mapping, achieved by hybridization of cDNAs against BAC clones which can subsequently be used as the substrate for an IRS-PCR reaction. As each cDNA will hybridize against multiple BAC clones and the chance to recover an IRS fragment from a BAC is 50%, the majority of cDNAs will result in the recovery of at least a single IRS fragment from a cognate BAC.

A possible pitfall can be the cross-hybridization with sequences on unrelated BAC clones that are homologous to the cDNA probe, for instance pseudogenes. This problem can be detected as follows. First, an unusually high number of positive BAC clones in relation to the number of genome equivalents screened indicates the abundance of related sequences in the genome. Second, Southern blots can be prepared from both BAC clones and genomic DNA to compare whether the cDNA probe produces a similar pattern of bands in DNA of both sources. Third, the mapping position determined can be validated by comparison with mapping data (if available) of corresponding orthologs in other species with respect to known synteny relationships [34].

The proportion of spots resolved by 2-DE that can be assigned chromosomal map positions will directly reflect the number of proteins to which the corresponding genes can be identified, because this is the obvious prerequisite for

any mapping to be carried out. The mapping, as such, benefits from a complete-as-possible physical map. The YAC map of the mouse genome, though close to completion, is still actively being worked on. The full impact therefore will only be realized once the contig map is finished. At present, an unambiguous map position can be obtained for over 50% of the genes tested.

Taken together, while proteins showing polymorphisms on 2-D electrophoresis gels can be mapped by genetic linkage analysis, the strategy presented here offers a way to map nonpolymorphic proteins from 2-DE gels on the physical map of mouse chromosomes.

DNA sequencing was carried out at the MPI-MG service unit. This work was supported by the Max-Planck-Gesellschaft.

Received October 10, 1998

5 References

[1] Klose, J., *Humangenetik* 1975, *26*, 211–243.

[2] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[3] Mann, M., *Trends Biochem. Sci.* 1996, *21*, 494–495.

[4] Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., Mann, M., *Nature Genet.* 1998, *20*, 46–53.

[5] URL: http://www.ncbi.nlm.nih.gov/UniGene/Mm_Chr/

[6] URL: http://carbon.wi.mit.edu:8000/cgi-bin/mouse/index

[7] Green, E. D., Olson, M. V., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 1213–1217.

[8] Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D. H., Caskey, C. T., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 6686–6690.

[9] McCarthy, L., Hunter, K., Schalkwyk, L., Riba, L., Anson, S., Mott, R., Newell, W., Bruley, C., Bar, I., Ramu, E., Housman, D., Cox, R., Lehrach, H., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 5302–5306.

[10] Hunter, K. W., Ontiveros, S. D., Watson, M. L., Stanton, Jr., V. P., Gutierrez, P., Bhat, D., Rochelle, J., Graw, S., Ton, C., Schalling, M., Aburatani, H., Brown, S. D. M., Seldin, M. F., Housman, D. E., *Mamm. Genome* 1994, *5*, 597–607.

[11] Himmelbauer, H., Wedemeyer, N., Haaf, T., Wanker, E. E., Schalkwyk, L. C., Lehrach, H., *Mamm. Genome* 1998, *9*, 26–31.

[12] Mott, R., Grigoriev, A., Maier, E., Hoheisel, J., Lehrach, H., *Nucleic Acids Res.* 1993, *21*, 1965–1974.

[13] Chapman, V. M., *Mouse News Lett.* 1975, *53*, 61.

[14] Hearne, C. M., McAleer, M. A., Love, J. M., Aitman, T. J., Cornall, R. J., Ghosh, S., Knight, A. M., Prins, J. B., Todd, J. A., *Mamm. Genome* 1991, *1*, 273–282.

[15] Mattei, M. G., Duprey, P., Li, Z. L., Mattei, J. F., Paulin, D., *Biol. Cell* 1989, *67*, 235–237.

[16] D'Eustachio, P., Ingram, R. S., Tilghman, S. M., Ruddle, F. H., *Somatic Cell. Genet.* 1981, *7*, 289–294.

[17] Klose, J., in: Link, A. J. (Ed.), *Methods in Molecular Biology, 2-D Proteome Analysis Protocols*, Humana Press, Totowa 1998, chapter 9, in press.

[18] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[19] Klose, J., in: Link, A. J. (Ed.), *Methods in Molecular Biology, 2-D Proteome Analysis Protocols*, Humana Press, Totowa 1998, chapter 18, in press.

[20] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173–179.

[21] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[22] Vorm, O., Roepstorff, P., Mann, M., *Anal. Chem.* 1994, *66*, 3281–3287.

[23] Jensen, O., Podtelejnikov, A., Mann, M., *Rapid Commun. Mass Spectrom.* 1996, *10*, 1371–1378.

[24] Gauss, C., Kalkum, K., Löwe, M., Lehrach, H., Klose, J., *Electrophoresis* 1999, *20*, 575–600.

[25] URL: http://www.ncbi.nlm.nih.gov/

[26] URL: http://www.charite.de/humangenetik

[27] URL: http://www.rzpd.de/

[28] Feinberg, A. P., Vogelstein, B. A., *Anal. Biochem.* 1983, *132*, 6–13.

[29] Church, G. M., Gilbert, W., *Proc. Natl. Acad. Sci. USA* 1984, *81*, 1991–1995.

[30] Sambrook, J., Fritsch, E. F., Maniatis, T., *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989.

[31] Boguski, M. S., Lowe, T. M., Tolstoshev, C. M., *Nature Genet.* 1993, *4*, 332–333.

[32] http://www.ncbi.nlm.nih.gov/BLAST/

[33] http://www.informatics.jax.org/bin/ccr/

[34] DeBry, R. W., Seldin, M. F., *Genomics* 1996, *33*, 337–351.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1033–1038

Stanislav N. Naryzhny
Vera V. Levina
Elena Y. Varfolomeeva
Eugeny A. Drobchenko
Michael V. Filatov

Petersburg Nuclear Physics Institute of Russian Academy of Sciences, Gatchina, Leningrad district, Russia

Active dissociation of the fluorescent dye Hoechst 33342 from DNA in a living cell: Who could do it?

It is assumed that DNA in mammalian cells is a dynamic conformationally unstable system. This instability provides the cell with a mechanism for dissociating a large number of substances that bind tightly but not covalently to DNA. Among these is the fluorescent dye Hoechst 33342, which binds to DNA in the minor groove. We have selected cell lines with a high capability for active dissociation of Hoechst 33342. Comparative protein analysis of these lines by means of two-dimensional (2-D) electrophoresis was performed. Cell and nuclear proteins were analyzed from these and normal strains. A few proteins with significantly changed quantities have been found. The preliminary search of the 2-D database allowed us to identity some known and unknown cellular proteins that could participate in active dissociation of the dye from DNA.

Keywords: DNA clearing / Mammalian cells / Two-dimensional polyacrylamide gel electrophoresis / Hoechst 33342 EL 3404

1 Introduction

DNA in a eukaryotic cell is tightly packed in a highly organized structure, chromatin. On the one hand this structure allows a huge amount of genetic information to be stored, but on the other hand this information must be used. Numerous nuclear processes, including DNA replication, DNA recombination, DNA repair, and transcription, must be correctly directed and not interfere with each other. It seems to us that chromatin has a system of housekeeping that actively removes substances tightly bound to DNA from the chromatin and thus prepares it for carrying out all life-supporting functions. The first evidence of such a possibility was presented by Smith, Debenham and Watson [1–3] who described a mutant line of L cells resistant to Hoechst 33342. This bisbenzimidazole dye is a typical example of substances that are tightly but not covalenly bound to DNA. Using a cytofluorimetric technique we analyzed this phenomenon in more detail. This process is energy-dependent and can be suppressed by topoisomerase-2 inhibitors and DNA breaks [4]. Moreover, it was shown that step-by-step selection of cells resistant to the dye leads to strains that much more effectively remove the dye. Thus we can envisage a cell protein system that alters the DNA conformation at dye binding sites. We refer to this process of removing dye and other molecules from DNA as "DNA clearing" [5, 6].

Correspondence: Dr. Stanislav N. Naryzhny, Petersburg Nuclear Physics Institute of Russian Academy of Sciences, Gatchina, Leningrad district, 188350, Russia
E-mail: naryzhny@omrb.pnpi.spb.ru
Fax: +7-81271-323-03

Abbreviations: CHO, Chinese hamster fibroblasts; **2-ME,** 2-mercaptoethanol

Recent advances in 2-D PAGE, image analysis, protein sample handling, spot identification, together with algorithms for searching sequence databases, presents a great opportunity for the study of cellular proteins as an interrelated set of molecules, the proteome. Therefore, in a project to examine protein differences between cell lines differing in their ability to remove dye from DNA, we have conducted 2-D PAGE analysis of proteins from these lines.

2 Materials and methods

2.1 Materials

Reagents for electrophoresis, Triton X-100, and NP-40 were from Serva (Heidelberg, Germany), 2-mercaptoethanol (2-ME) from Ferak (Berlin, Germany), protease inhibitors set from Boehringer Mannheim (Mannheim, Germany). Ampholines were from LKB (Uppsala, Sweden), and Bio-Lytes were from Bio-Rad (Uppsala, Sweden). Cytochrome *c*, aldolase, and catalase from Pharmacia-LKB (Uppsala, Sweden) were used as markers. Hoechst 33342, novobiocin, elipticine, etoposide, and PMSF were from Sigma (St. Louis, MO, USA).

2.2 Cells

The strain of mouse L fibroblasts has been cultivated in the laboratory for more than 10 years. The strain of Chinese hamster fibroblasts (CHO AA8) was a gift from Prof. J. J. Huijmakers from Genetics Institute, Rotterdam, Holland. Cell cultures were grown in a mix of Eagle's basal and 199 medium with 10% bovine serum. The method of step-by-step selection with increasing concentration of Hoechst 33342 was used for production of a set of

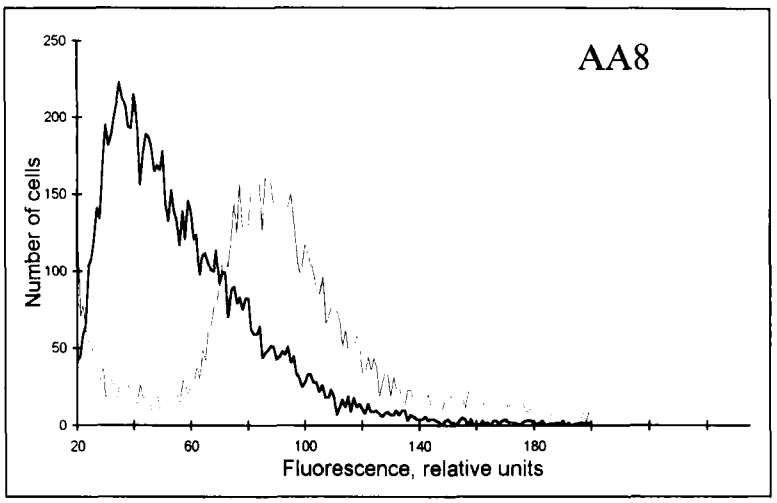
Hoechst 33342 resistant lines. A detailed description of the selection was published by [5, 6]. The highly resistant strains used here were AA8HoeR-7 in the case of AA8 cells and LHoeR-3 in the case of L cells. Staining with Hoechst 33342 was performed by adding the dye at a nontoxic concentration (usually 2 μg/mL) directly to the medium. The dye was removed by repeated changing of the medium. Fluorescence was measured by a Bruker flow cytofluorometer (Bremen, Germany) with a mercury lamp used as the light source.

2.3 Preparation of samples for electrophoresis

Cells were cultured and harvested. The cells were counted, washed in 10 mM phosphate buffer with 0.15 M NaCl, and centrifuged at 1000 × *g* for 5 min. Crude nuclei were obtained by treating the pellet with the same buffer and with 0.1% Triton X-100. The cell or nuclear pellets were dissolved in 8 M urea, 2% Triton X-100, 1% 2-ME, 2 % carrier ampholytes (Ampholines 5–8), 1 mg/mL Pefabloc® SC, 10 μg/mL leupeptin, 10 μg/mL pepstatin, 1 μg/mL aprotinin.

2.4 Two-dimensional gel electrophoresis

Electrophoretic procedures were carried out using the previously described devices and technology according to a modified version of the O'Farrell procedure [7, 8]. Samples for IEF or NEPHGE (cell or crude nuclear lysate) were loaded in 50 μL (1 million cells). IEF was performed



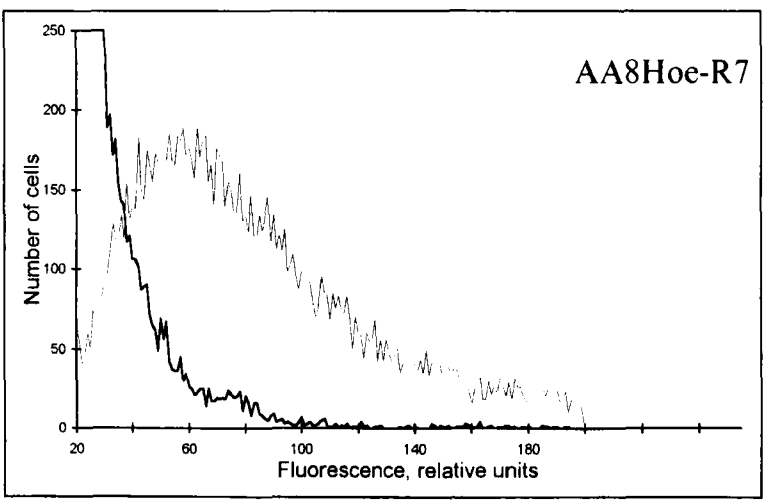


Figure 1. Elimination of Hoechst 33342 from DNA in normal (AA8) and Hoechst 33342-resistant (AA8HoeR-7) CHO cells. Flow cytofluorometric histograms of cells are represented. Intact cells in log phase of growth were stained *in vivo* with Hoechst 33342 (2 μg/mL) for 1 h at 37°C (light line) and then incubated in the dye-free growth medium (2 h, 37°C; bold line).

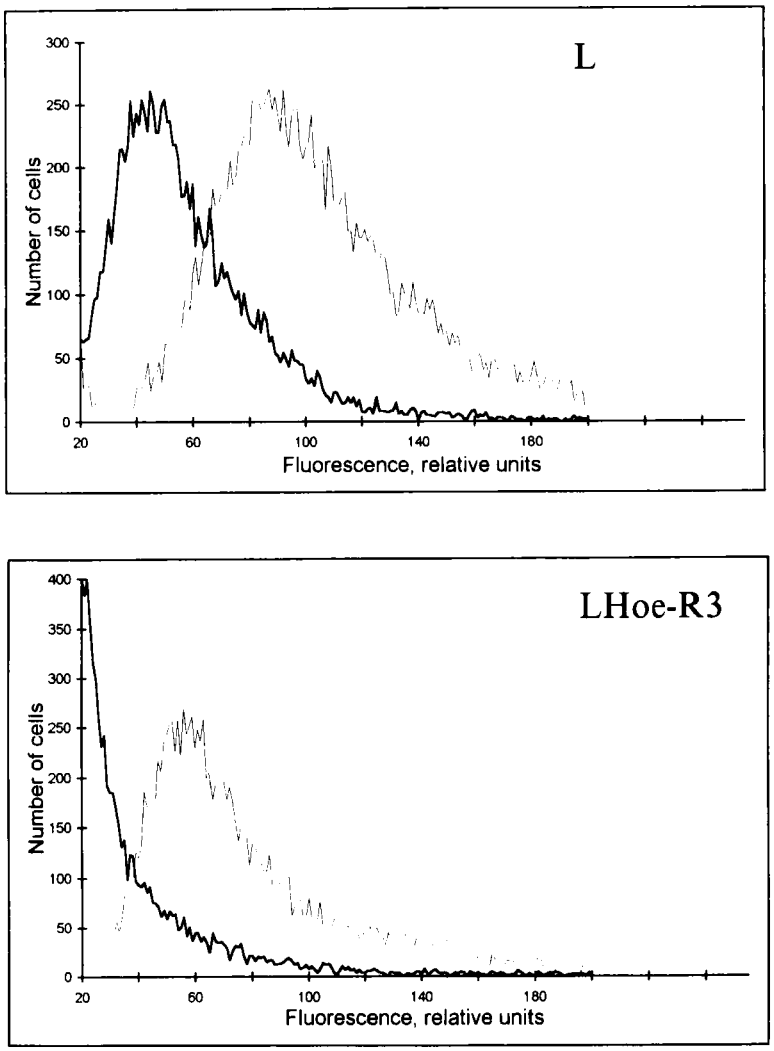


Figure 2. Elimination of Hoechst 33342 from DNA in normal (L) and Hoechst 33342-resistant (LHoeR-3) mouse cells. Flow cytofluorometric histograms of cells are represented. Intact cells in log phase of growth were stained *in vivo* with Hoechst 33342 (2 μg/mL) for 1 h at 37°C (light line) and then incubated in the dye-free growth medium (2 h, 37°C; bold line).

at 500 V for 18 h in 60 × 4 mm polyacrylamide (4%T, 5%C) rod gels containing 9 M urea, 0.1% NP-40, 2% Bio-Lyte (3–10). NEPHGE was performed according to O'Farrell [9] using 20 mM NaOH as cathode and 10 mM HEPES as anode buffer. The same technique and reagents as for IEF were used: 9 M urea, 0.1% NP-40, 0.8% Ampholine (5–8), 0.2% Bio-Lyte (3–10) in 60 × 4 mm polyacrylamide (4%T, 5%C) rod gels. Fifty μL of each sample was loaded onto each tube gel, covered with 10 μL 4 M urea in 1% 2-ME and run from acidic to basic end with incremental voltage (30 min at 300 V, 180 min at 600 V, 30 min at 900 V (total 2400 Vh). Following focusing, gels were equilibrated (20 min) in 5 mL of equilibration solution (60 mM Tris-HCl, pH 6.8, 1% SDS, 20 mM 2-ME, 10% glycerol) and applied in pairs to the second dimension (12% SDS-PAGE) with the aid of hot 1% agarose-containing equilibration solution and 0.001% bromophenol blue. Gels (9 × 14 cm) were run for 2 h at constant power at room temperature (100 V and 30 mA at start). Gels were stained with Coomassie Brilliant Blue or ammoniacal silver nitrate [10]. 2-D images were analyzed by means of Bio Image® Intelligent Quantifier (Ann Arbor, MI, USA).

3 Results and discussion

Step-by-step selection of cell lines with the ability to grow in an increased concentration of Hoechst 33342 leads to the appearance of cell mutants actively removing the dye

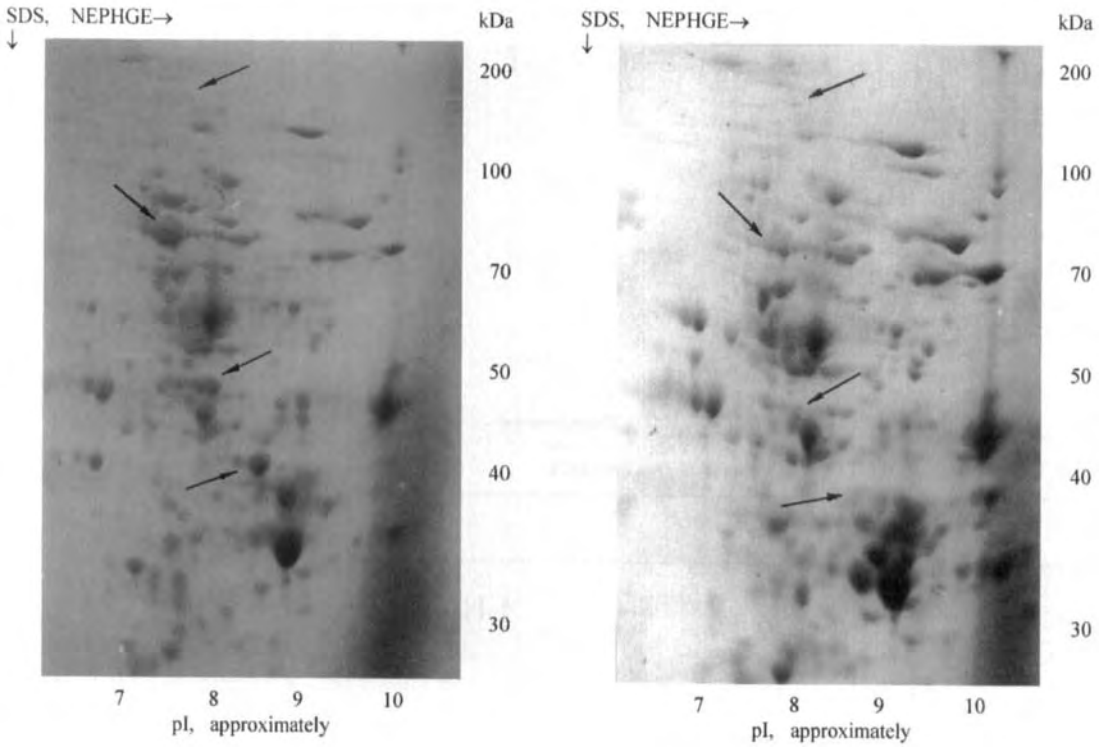


Figure 3. 2-D electrophoresis (NEPHGE, SDS-PAGE) of protein lysates from normal cells (AA8) and CHO cells resistant to Hoechst 33342 (AA8HoeR-7). NEPHGE was performed using 20 mM NaOH as cathode and 10 mM HEPES as anode buffer. The same technique and reagents as for IEF were used: 9 M urea, 0.1% NP-40, 0.8% Ampholine, pH 5–8, 0.2% Bio-Lyte, pH 3–10, in 60 × 4 mm polyacrylamide 4%T, 5 %C rod gels. Fifty μL of each sample was placed on each tube, covered by 10 μL of 4 M urea in 1% 2-ME and run from acidic to basic end with incremential voltage (30 min at 300 V, 180 min at 600 V, and 30 min at 900 V (total 2400 Vh). Second dimension: SDS-PAGE (12%). Staining with Coomassie blue. The main spots in which differences were revealed are marked by arrows.

from DNA [5, 6]. During the short incubation in a medium without the dye, the majority of bound Hoechst 33342 is effectively removed (Figs. 1, 2). As can be seen, cell lines permissive to Hoechst 33342 (AA8 and L, no growth at dye concentration more than 1 μg/mL) can remove no more than 50% of the dye in 2 h. Additional dissociation of the dye was not observed during a 24 h incubation period. At the same time, the cell lines resistant to Hoechst 33342 (AA8HoeR-7 and LHoeR-3, growth at the dye concentration of more than 30 μg/mL) can effectively remove almost all of the bound dye in 2 h. Thus the previously described phenomenon of the ability of cells to actively eliminate Hoechst 33342 from DNA [1] causes resistance to the dye's toxic action. Furthermore, the gradual change of this activity during cell selection points to the possibility of gene amplification, the typical phenomenon of permanent mammalian cell lines.

In order to clarify the possibility of differences in proteome compositions of cell lines actively removing the dye from DNA or not, protein samples were analyzed by 2-D gel electrophoresis. Matching of spots in the p*I* range from 4 to 7 (2-D PAGE with IEF in first direction; data not shown) has not shown clear differences in protein expression levels. Therefore, NEPHGE was used for better resolution of alkaline proteins. A set of runs was performed to compare different cell lines. First, one pair of strains (AA8 and AA8-HoeR-7) was analyzed. Visual inspection of gels stained with Coomassie blue (about 200 protein spots) was performed and proteins with different level of expression were marked (Fig. 3). Second, another pair of cell lines (L and LHoeR-3) was treated (Fig. 4). Next, two sets of marked polypeptides were compared and several of the same changes were found. Further quantitative analysis was performed using the Bio Image Intelligent Quantifier.

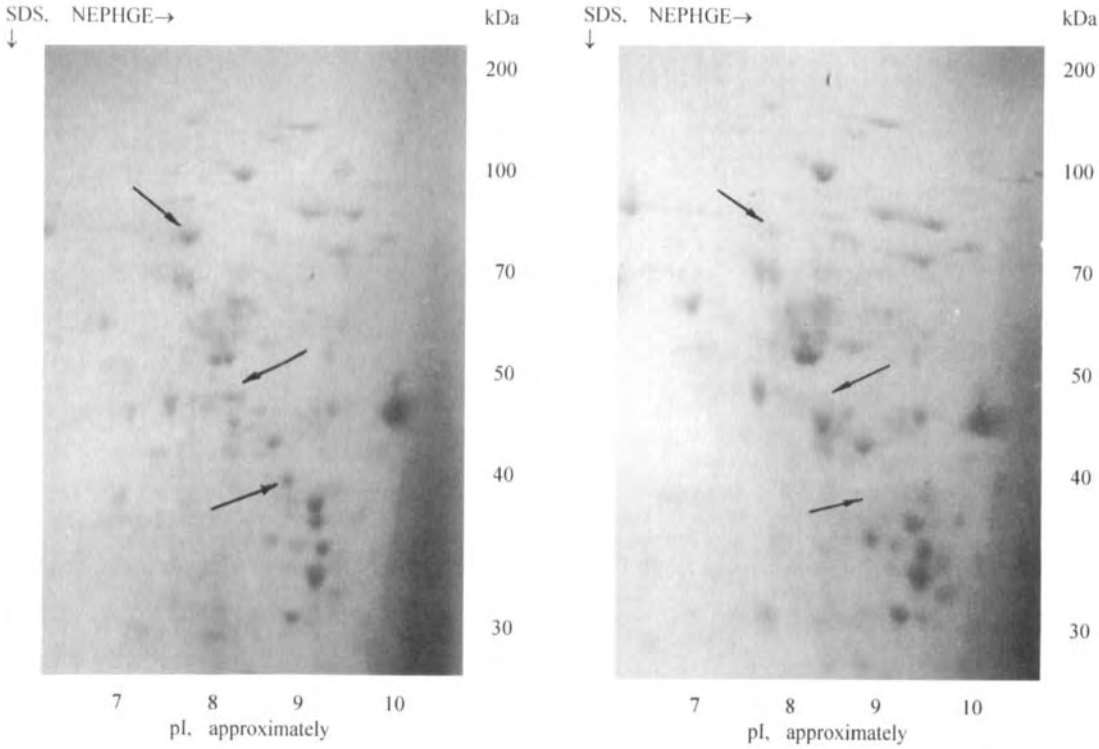


Figure 4. 2-D electrophoresis of protein lysates from normal cells (L) and mouse CHO cells resistant to Hoechst 33342 (LHoeR-3). Electrophoretic conditions are as in Fig. 3. The main spots in which differences were revealed are marked by arrows.

Such an analysis has revealed the most noticeable changes. Most surprising, they were a sharp reduction of the level of several proteins, p40 (40 kDa, p*I* 9), p48 (48 kDa, p*I* 8) and p80 (80 kDa, p*I* 7.5) in cells with a high ability to remove the dye. Another difference was in a splitting of a polypeptide with a molecular mass of 170 kDa (p*I* 7.5) in such a way that a more alkaline form appeared. Moreover, all these proteins have a nuclear localization (Fig. 5). Thus, these polypeptides could be related to DNA clearing. Preliminary identification of these polypeptides according to their positions on the 2-D map (SWISS-PROT) allow us to suggest than p40 could be a heat shock protein (HSP40) and p170 could belong to the helicase family.

In spite of insufficient information concerning the real mechanisms of DNA clearing, it is tempting to propose an explanation of this phenomenon. Taking into account all data obtained, there are many reasons to suppose that the origin of the cell's ability to remove the dye and other substances from DNA is conformational. DNA in eukaryotic cells is highly packed and organized. All genetic processes are accompanied by transient chromatin rearrangements. Thus we can imagine that DNA clearing is a manifestation of active conformational alterations of chromatin (remodeling) which "shakes itself" to be ready to carry out the necessary functions. Some proteins possibly involved in this process are seen on 2-D PAGE according to their significantly increased quantity in cell lines weakly removing Hoechst 33342 from DNA (p40, p48, p80; Figs. 3 and 4). We conclude that DNA clearing is a complicated process which depends on chromatin structure and is mediated through the action of different protein systems.

4 Concluding remarks

We selected cell lines with high ability for active dissociation of Hoechst 33342 from DNA. Proteome analysis by 2-D PAGE was used for a comparative study of these different cell lines. The data obtained allowed us to conclude that this process is carried out by – and depends on – certain proteins, most likely participating in topological organization of chromatin.

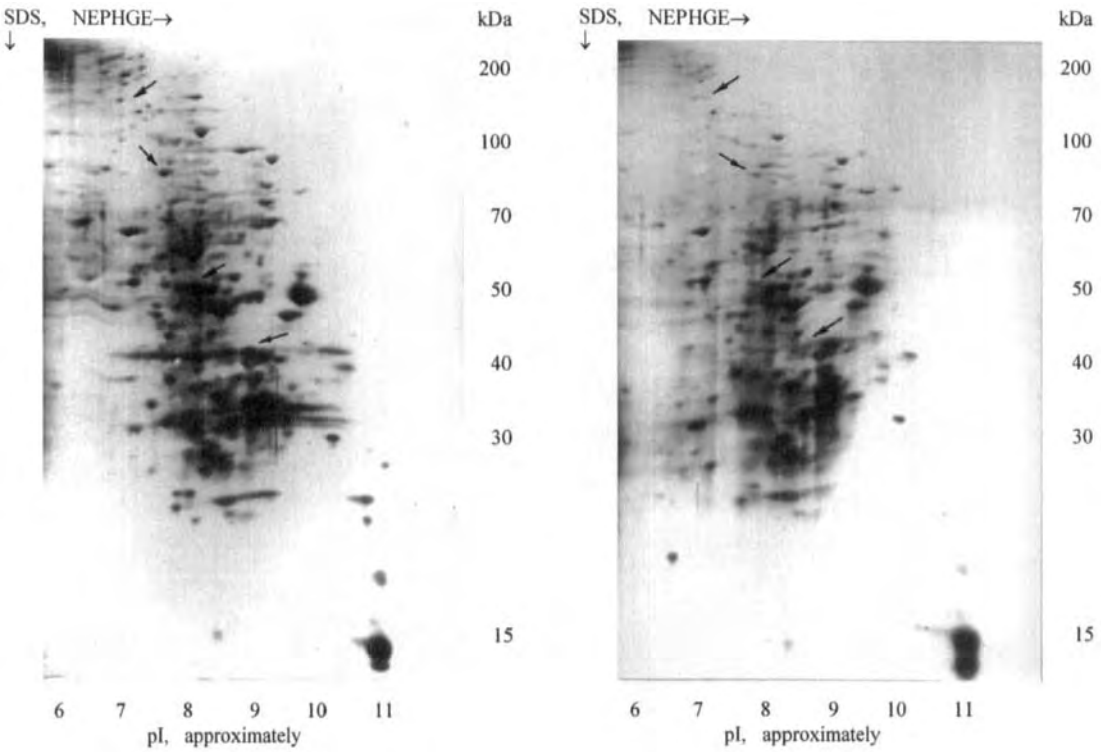


Figure 5. 2-D electrophoresis of crude nuclear proteins. Left, normal mouse cells (L); right, mouse cells resistant to Hoechst 33342 (LHoeR-3). Electrophoretic conditions are as in Fig. 3. Staining with silver. The main spots in which differences were revealed are marked by arrows.

We thank A. A. Zverev for the illustrations. This work was supported by a grant from the Russian Foundation of Fundamental Investigations.

Received September, 1998

5 References

[1] Debenham, P. G., Webb, M. B. T., *Somat. Cell Mol. Genet.* 1987, *13*, 21–32.

[2] Smith, P. J., Lacy, M., Debenham, P. G., Watson, J. V., *Carcinogenesis* 1988, *9*, 485–496.

[3] Smith, P. J., Debenham, P. G., Watson, J. V., *Mutation Res.* 1989, *217*, 169–172.

[4] Filatov, M. V., Varfolomeeva, E. Y., *Mutation Res.* 1995, *327*, 209–215.

[5] Levina, V. V., Varfolomeeva, E. Y., Filatov, M. V., *Biol. Membranes* (Russ.) 1998, *6*, 700–708.

[6] Levina, V. V., Varfolomeeva, E. Y., Filatov, M. V., *Cytometry* 1998, *9*, 64–68.

[7] Naryzhny, S. N., *Electrophoresis* 1997, *18*, 553–556.

[8] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[9] O'Farrell, P. Z., Goodmann, H. M., O'Farrell, P. H., *Cell* 1997, *12*, 1133–1141.

[10] Oakly, B. R., Kirsch, D. R., Morris, N. R., *Anal. Biochem.* 1980, *105*, 361–366.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing the article, please refer to: *Electrophoresis* 1999, *20*, 1039–1046

Ayodele A. Alaiya[1]
Bo Franzén[1]
Birgitta Moberger[2]
Claes Silfverswärd[3]
Stig Linder[4]
Gert Auer[1]

[1]*Unit of Cell and Molecular* Analysis, Department of Oncology-Pathology
[2]Department of Woman and Child Health, Division of Obstetrics and Gynecology
[3]Histopathology Unit, Department of Oncology-Pathology
[4]Radiumhemmets Research Laboratory, Department of Oncology-Pathology Cancer Center Karolinska, Karolinska Institute and Hospital, Stockholm, Sweden

Two-dimensional gel analysis of protein expression in ovarian tumors shows a low degree of intratumoral heterogeneity

The process of tumor progression leads to the emergence of multiple clones, and to the development of tumor heterogeneity. One approach to the study of the extent of such heterogeneity is to examine the expression of marker proteins in different tumor areas. Two-dimensional gel electrophoresis (2-DE) is a powerful tool for such studies, since the expression of a large number of polypeptide markers can be evaluated. In the present study, tumor cells were prepared from human ovarian tumors and analyzed by 2-DE and PDQUEST. As judged from the analysis of two different areas in each of nine ovarian tumors, the intratumoral variation in protein expression was low. In contrast, large differences were observed when the protein profiles of different tumors were compared. The differences in gene expression between pairs of malignant carcinomas were slightly larger than the differences observed between pairs of benign tumors. We conclude that 2-DE analysis of intratumoral heterogeneity in ovarian cancer tissue indicates a low degree of heterogeneity.

Keywords: Two-dimensional gel electrophoresis / Ovarian cancer / Tumor heterogeneity
EL 3316

1 Introduction

Tumors become heterogeneous as a consequence of tumor cell genetic instability [1]. In a tumor, a Darwinistic selection pressure operates to select the fastest growing and most invasive tumor clone. At any given time, a primary tumor will represent the sum of its competing clones. Clonal dominance has been shown to occur in some experimental models allowing a subpopulation of cells to overgrow the primary heterogeneous tumor and to metastasize [2]. From the literature, it is not clear whether tumors in general are extremely heterogeneous due to genetic instability, or relatively homogeneous due to overgrowth of dominant clones. Tumor heterogeneity has been studied by various methods. A number of studies investigated DNA content, karyotypes, or loss of heterozygosity at specific loci [3–5]. In other studies, the expression of various markers were investigated in different tumor areas [6–8]. We have previously used 2-DE analysis of tumor cell polypeptides to examine heterogeneity in gene expression between breast fibroadenomas and invasive ductal carcinomas [9]. By this approach we could compare the expression of a large number of gene products in tumors and were able to demonstrate a high degree of diversity of polypeptide expression in malignant breast tumors compared to benign fibroadenomas. Several large tumors allowed the examination of gene expression in different areas of the tumor for intratumoral heterogeneity. Somewhat surprisingly, similar patterns of protein expression were observed in different areas of the same breast tumor. This observation suggested that large tumors are relatively homogeneous with regard to protein expression.

In the present study we have examined tumor heterogeneity in ovarian carcinoma. These tumors are often large and different areas can be examined by 2-DE. Benign tumors are of sufficient size to be examined, and we could therefore study tumor heterogeneity during malignant progression. Similar to our previous study, we report that different areas of the same primary tumors show similar protein profiles. Intertumor heterogeneity was large, however, and showed relatively small differences between benign and malignant tumors.

2 Materials and methods

Twenty-six patients with ovarian tumors were analyzed. Fresh tumor tissues were collected shortly after resection at the central operation theater at the Karolinska hospital. All samples were first examined by a pathologist in order to obtain representative, viable, and nonnecrotic tumor tissue. One part of the tissue was used for preparation for 2-DE and the adjacent tissue was formalin-fixed and paraffin-embedded for histological characterization.

Correspondence: Ayodele A. Alaiya, Unit of Cell and Molecular Analysis, Department of Oncology-Pathology, Cancer Center Karolinska, Karolinska Institute and Hospital, S-17176 Stockholm, Sweden
E-mail: alaiya.ayodele@cck.ki.se
Fax: +46-8-331696

2.1 Sample preparation

All samples, described in Table 1, were obtained shortly after resection (within 40 min) and sample preparation and solubilization was performed essentially as previously described [11]. Briefly, tumor cells were collected from the cut surface of nonnecrotic tumor tissue by scraping with a sharp scalpel. Cells were collected in 2–5 mL of ice-cold RPMI-1640 medium containing 5% fetal calf serum and 0.2 mM phenylmethylsulfonyl fluoride/0.83 mM benzamidine. A metal filter with 0.5 mm pore size was used to remove tissue fragments and connective tissue. Cell suspensions were underlaid with 2 mL of ice-cold Percoll (54.7% in PBS) and centrifuged at 1000 × *g* for 10 min at +4°C. Cells at the interface were collected and washed twice with PBS. Parallel samples were prepared for histological and cytological examination. Extracts were then prepared for 2-DE analysis as described [11].

2.2 Characterization of formalin-fixed specimens

Histopathological characterization was carried out using hematoxylin-eosin-stained sections of formalin-fixed and paraffin-embedded specimens. Tumors were characterized according to size, lymph node status, and metastasis site. Histopathological typing was according to WHO criteria. The types of epithelial tumors included in the study are shown in Table 1.

2.3 Cytochemical analysis

Nuclear DNA content was assessed in all the specimens by image cytometric analysis of Feulgen-stained cells. Tumors with a single stemline in the normal diploid region (1.9c–2.1c) were classified as diploid and tumors with pronounced scattered DNA values exceeding the tetraploid region were classified as aneuploid [12].

2.4 2-D Electrophoresis

2-DE was performed by standard procedures as previously described [10]. Resolyte (2%, pH 4–8; BDH, Poole, UK) were used for the first-dimensional isoelectric focusing, 10–13% linear gradient SDS-polyacrylamide gels were used in the second dimension. Gels were stained with silver nitrate as described by [13].

2.5 Identification of polypeptides

The identification of various polypeptide maps in 2-DE patterns has been described previously [10]. Identification

Table 1. Clinical and histopathological characteristics of samples

No.	Sample	Diagnosis/histologic Type	Differentiation	Size(cm)	DNA	Metastasis	Age
1	OC-14A	Serous cystadenoma IA		14		-	75
2	OC-19	Serous cystadenoma IA		8		-	52
3	OC-26	Mucinous cystadenoma IIA		15	D	-	49
4	OC-34	Serous cystadenoma IA		15	D+T	-	65
5	OC-38	Serous cystadenoma IA		9		-	68
6	OC-66B	Serous cystadenoma IA					
7	OC-82	Cystadenofibroma					
8	OC-21A	Borderline mucinous IIB		19	D+T		
9	OC-39A	Borderline seropapillary IB	Implantation metastasis	13	D+T	Omentum	73
10	OC-46	Borderline seropapillary IB			D+T		84
11	OC-50	Borderline seropapillary IB	Noninfiltrative	10		-	44
12	OC-59A	Borderline mucinous IIB	Noninfiltrative	19	D+T	-	44
13	OC-68A	Borderline mucinous IIB	Borderline mucinous IIB	20	A	Omentum	54
14	OC-72B	Borderline serous					
15	OC-77A	Borderline serous					
16	OC-04B	Mixed tumor	Metastasizing	25	A	Omentum	80
17	OC-06	Clear cell tumor (IVC)	Medium/high differentiation adinoca	10	A	-	76
18	OC-07B	Sero papillary ADC(IC)	Low differentiation adenoca	8	A	Appendix	81
19	OC-08	Sero papillary ADC(IC)	Low differentiation adenoca	9	A	Omentum	55
20	OC-09B	Sero papillary ADC(IC)	Low differentiation adenoca	8	A	Omentum	78
21	OC-33B	Endometrioid Ca IIIC	Medium/high differentiation adenoca	16	D+T	-	51
22	OC-40L	Bil adenocarcinoma	Low differentiation adenoca		A	-	64
23	OC-43B	Bil seropapillary IC	Low differentiation adenoca	12	A	-	72
24	OC-45B	Endometrioid Ca IIIC	Low differentiation adenoca	7	A	Adnexia	51
25	OC-48B	Sero papillary IC	Medium/high differentiation adenoca	12	A	-	78
26	OC-84B	Clear cell tumor (IVC)	Medium/high differentiation adenoca	7	A	-	61

was performed by matching with published maps, exchanging samples with other investigators, or by coelectrophoresis of purified proteins, and by mass spectrometry. Proliferating cell nuclear antigen (PCNA) was identified by immunoblotting (PC10 mAB; Dakopatt, Glostrup, Denmark) using a semi-dry system (Multiphor II Nova Blot; Pharmacia Biotech AB, Uppsala, Sweden) and ECL detection (Amersham, Aylesbury, UK).

2.6 Gel scanning and image analysis

2-DE wet gels were scanned after silver staining at 100 μm resolution (12 bits/pixel) using a Molecular Dy-

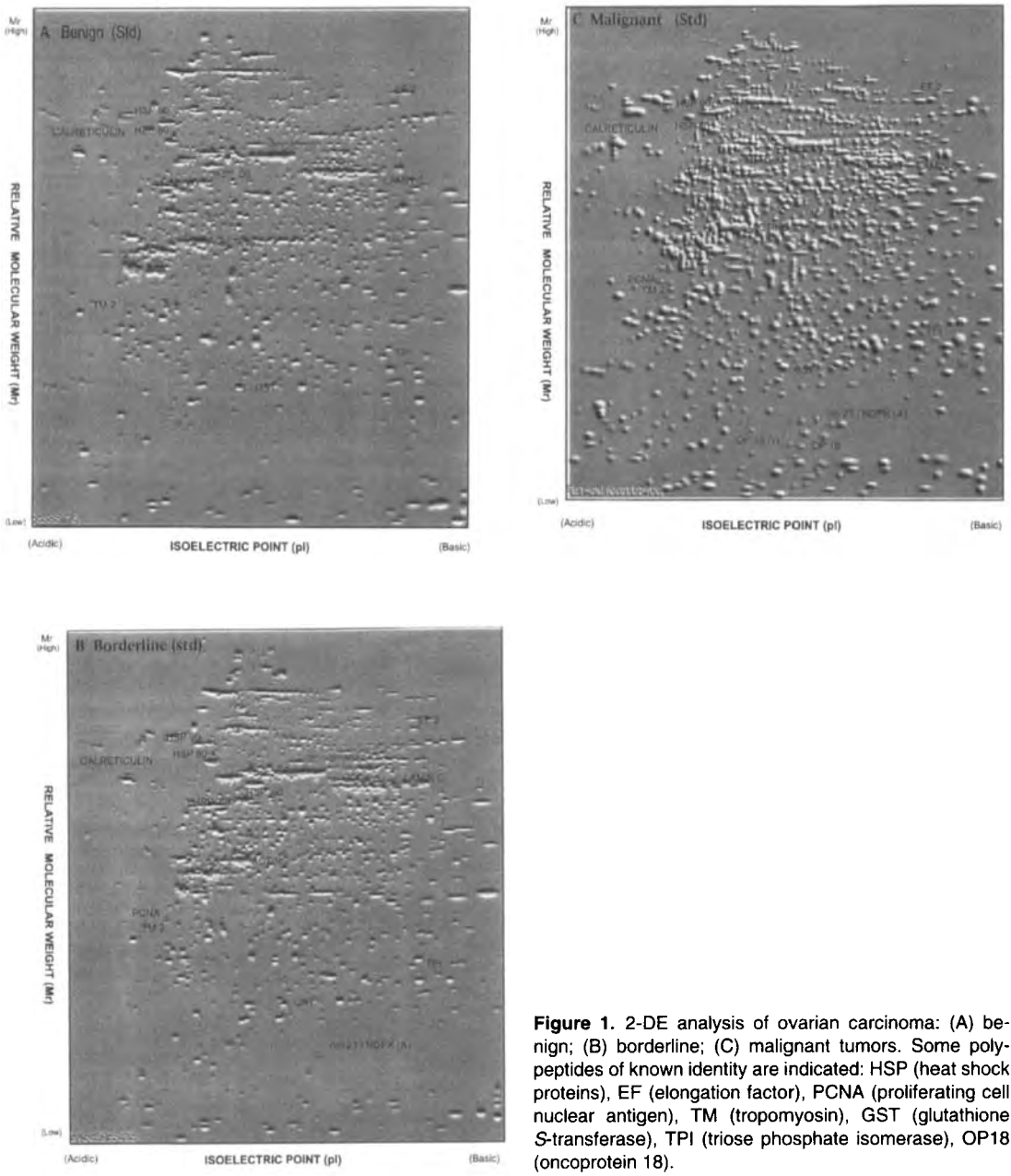


Figure 1. 2-DE analysis of ovarian carcinoma: (A) benign; (B) borderline; (C) malignant tumors. Some polypeptides of known identity are indicated: HSP (heat shock proteins), EF (elongation factor), PCNA (proliferating cell nuclear antigen), TM (tropomyosin), GST (glutathione *S*-transferase), TPI (triose phosphate isomerase), OP18 (oncoprotein 18).

namics (Palo Alto, CA, USA) laser densitometer. Data was analyzed using PDQuest™ software on a SUN SPARC 5 station [14] purchased from Pharmacia Biotech. Background was subtracted, peaks for the protein spots located and counted. The individual polypeptide quantities were expressed as PPM of the total integrated optical density. The total spot counts and the total optical density is directly related to the total protein concentration. An "identification reference pattern" was constructed ("matchset"). Gel matching was performed with image analysis software PDQuest. At least two or more gels were matched together by picking the best gel as the reference gel and subsequently matching each gel in the matchset to the reference gel. Several known proteins were used as landmarks to facilitate rapid and better gel matching. One general match set was made containing all the cases for effective intertumor comparisons, while three separate matchsets were made for benign, borderline and malignant tumors. Several additional matchsets were made for pairwise comparisons of preparations from two separate areas of the same lesion and preparations from the same sample run at two separate occasions.

2.7 Sample selection for pairwise analysis

Pairwise comparison of the intensity of multiple spots was performed using the PDQuest program as described previously [9]. Because the total optical density is directly related to the total protein concentration, minor differences in gel loading and silver staining may affect sample comparisons. Therefore, the selection of samples for pairwise matching and correlation analysis were based on the following criteria: (i) very similar total spot counts and similar total optical densities for each pair of gels, (ii) a gel segment of 75 × 50 mm in the molecular mass region of 35–70 kDa and p*I* range between 5–6.5, and (iii) no differences greater than 4%, neither in total spot numbers nor in total optical densities for any pair of gels.

3 Results

3.1 Analysis of the degree of similarity in polypeptide expression by calculation of correlation coefficients

Cells were prepared from fresh ovarian surgical specimens and washed free of serum proteins. Extracts were prepared and analyzed by 2-DE for their expression of multiple polypeptides. Histological and cytological controls of all cases were performed to ensure sample representation. Between 700–1100 spots/gel were obtained after silver staining (Fig. 1A–C). A sample set of approximately 400 gel spots was quantified from each gel, and used in pairwise comparisons with other gels. The reproducibility of the analysis and the degree of heterogeneity in polypeptide expression was evaluated by comparing the intensity of all matched spots between 2-DE gels and calculating the correlation coefficient [9]. A correlation coefficient was obtained by comparisons of the optical densities of the same spots in two gels. A correlation coefficient of one shows that the two gels being compared are identical with regard to matched spots. An important aspect of this work is reproducibility. A number of samples were analyzed on more than one occasion and the difference in the protein profiles from the resulting gels was determined. The average correlation coefficient of the total optical densities of all matched spots was 0.90 (n = 8 gel pairs; examples of the analysis of two pairs of gels are shown in Fig. 2A, B).

3.2 Intratumoral variations

For nine tumors, more than one sample could be obtained. The various samples were collected from distinct, separate sites. Examples of the analysis of two tumors are presented in Fig. 2C and D (showing correlation coefficients of 0.89 and 0.87, respectively). The average cor-

Table 2. Correlation analysis of 2-DE of preparations from same sample on more than one occasion

No.	Sample	Diagnosis	Spots/ segment	Matched spots(n)	Matched spots(%)	Correlation coefficient(r)	DNA ploidy
1	OC66B	Benign IA	200/200	194	97	0.89	D
2	OC26	Benign IA	241/242	236	98	0.89	D
3	OC82	Benign IA	147/147	147	100	0.91	D
4	OC72B	Borderline IB	238/244	234	97	0.90	D/T
5	OC21A	Borderline IIB	336/348	336	98	0.89	D/T
6	OC04B	Malignant IC	387/387	387	100	0.94	A
7	OC07B	Malignant IC	418/405	400	97	0.89	A
8	OC40L	Malignant IC	206/206	206	100	0.88	A

Mean = 0.90

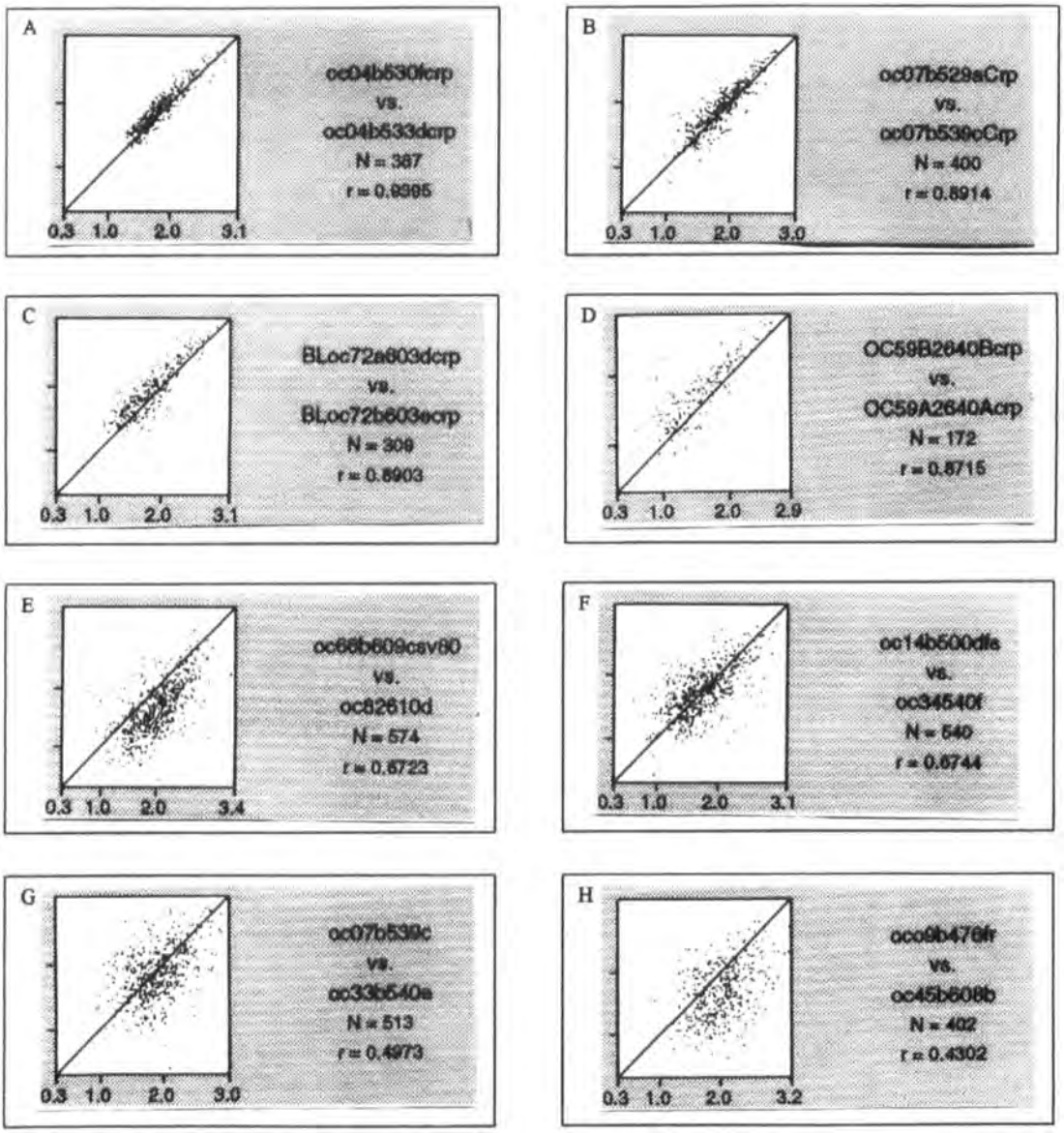


Figure 2. Scatterplots showing the correlation of expression levels of matched proteins in pairs of 2-DE gels (where each small dot represents the expression of one polypeptide in one gel). (A) and (B) methodological reproducibility of two samples (the same samples analyzed at two different occasions); (C) and (D) intratumor variation in tumors; (E) and (F) intertumor variations among benign tumors; (G) and (H) intertumor variations among malignant tumors.

relation coefficient was 0.85 for the nine pairs of gels representing two different areas of the same tumor.

3.3 Intertumor variations

When tumors from different individuals were compared, we observed a larger degree of heterogeneity using pairwise comparisons. The results of the analysis of two benign tumors and two malignant tumors are shown in Fig. 2E, F and 2G, H, respectively. Seven benign tumors, one mucinous and six serous cystadenomas, were examined. Ten pairwise comparisons were made and the average correlation coefficient was 0.61 (0.54–0.72). Eight border-

line tumors (5 serous and 3 mucinous) were examined. All but one tumor showed both diploid and tetraploid (D/T) DNA distributions (Table 1). The mean correlation coefficient from 13 gel pairs was 0.57. Eleven invasive carcinomas (6 seropapillary adenocarcinomas, 2 clear cell tumors, 2 endometrioid tumors and 1 mixed tumor) were examined. All tumors were aneuploid with varying degrees of differentiation, except one endometrioid tumor (OC-33B) with both diploid and tetraploid (D/T) DNA distribution. Comparisons between pairs of aneuploid ovarian carcinomas showed an average correlation coefficient of 0.54 (n = 35 gel pairs). Pairs of low differentiated tumors showed a larger degree of heterogeneity (r = 0.53) than pairs of highly differentiated tumors (r = 0.59). This difference was not statistically significant. Similarly, no significant difference in heterogeneity was observed among poorly differentiated and highly differentiated aneuploid carcinomas.

3.4 Analysis of the degree of similarity in polypeptide expression by calculation of the fraction of matched spots

In 2-DE analyses, a varying number of spots will not be present on one of the gels in a pair, and can therefore not be matched. The fraction of matched spots was found to be lower when pairs of gels from malignant tumors were compared (52%) as opposed to borderline (59%) and benign tumors (64%). Each tumor was assigned with an index of the fraction of matched spots and correlation coefficient (Fig. 3). A remarkable similarity between the values of the correlation coefficient of matched spots and the fraction of matched spots was observed (Fig. 4). Both parameters were lower in malignant tumors compared to benign.

4 Discussion

We have analyzed both qualitative and quantitative electrophoretic data to assess tumor heterogeneity in ovarian carcinomas. Our results showed similar patterns of gene expression in separate areas of the same tumor. The average correlation coefficient from nine pairs of samples was 0.85, which is only slightly lower than the methodological variation (showing an average correlation coefficient of 0.90). In a previous study [9], we showed that the protein profiles from different areas of two breast carcinomas were very similar (correlation coefficient of 0.81); consistent with the present data. The finding of similar patterns of gene expression in different areas of carcinomas is of general interest, since it implies that ovarian tumors are more homogeneous than might have been expected. A possible explanation for this homogeneity could be that tumors that have reached a size making them

Table 3. Correlation analysis of 2-DE of preparations from separate areas A and B of the same lesion

No.	Sample	Diagnosis	Spots/ segment	Matched spots(n)	Matched spots(%)	Correlation coefficient(r)	DNA ploidy
1	OC14AB	Benign IA	218/221	206	94	0.80	D
2	OC66AB	Benign IA	245/238	238	98	0.89	D
3	OC72AB	Borderline IB	309/309	309	100	0.89	D/T
4	OC39AB	Borderline IB	290/286	285	99	0.84	D/T
5	OC59AB	Borderline IB	172/173	172	100	0.87	D/T
6	OC68AB	Borderline IB	254/246	238	95	0.81	D/T
7	OC48AB	Malignant IC	200/200	200	100	0.86	A
8	OC84AB	Malignant IC	130/130	130	100	0.82	A
9	OC45AB	Malignant IC	256/234	234	96	0.87	A

Mean = 0.85

Table 4. Correlation analysis of 2-DE of same and different sample preparations

Sample	Correlation coefficient(r)	± SD	% CV	N
Preparations from same sample on more than one occasion	0.90	0.02	2.1	8 pairs
Preparations from 2 separate areas, A and B, of the same lesion	0.85	0.03	3.9	9 pairs
Group correlation among benign cases	0.61	0.07	11	10 pairs
Group correlation among borderline cases	0.57	0.10	17	13 pairs
Group correlation among malignant cases	0.54	0.05	10	35 pairs

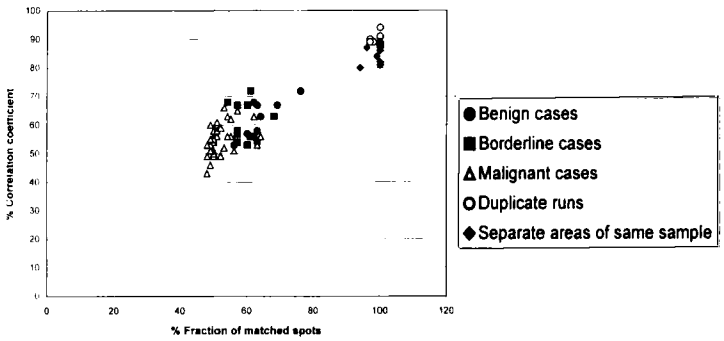


Figure 3. Graphical presentation of intertumor and intratumor variations showing the fraction of matched spots and correlation coefficients of matched spots. Benign tumors: black rings; borderline tumors: grey squares; malignant tumors: open triangles; duplicate runs: open rings; separate areas of same sample: black diamonds.

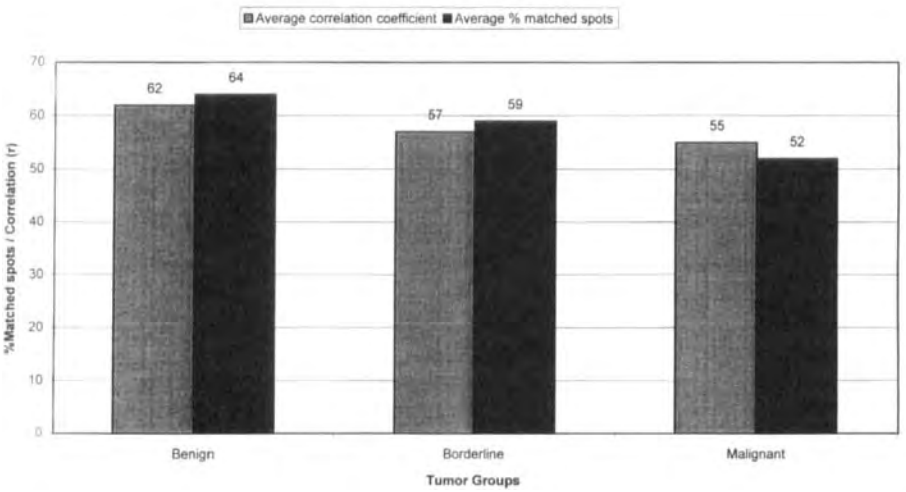


Figure 4. Relatedness of the fraction of matched spots and correlation coefficients of matched spots. Note that both the average number of matched spots and the average correlation coefficient is higher in benign compared to malignant tumors, showing that both parameters of homogeneity decrease during tumor progression.

amenable to 2-DE analysis have become relatively homogeneous due to overgrowth by a dominating clone. There are several reports in the literature of such overgrowth [2, 5, 15].

A number of investigators have presented evidence favoring large tumor heterogeneity. Orndal *et al.* [3] detected cytogenetic heterogeneity in about 29% of soft tumors with more than one cell clone. Tumors have been found to be composed of mixed subpopulations of both diploid and aneuploid clones in breast tumors [16], ovarian tumors [17] and in bladder carcinomas [18]. The degree of intertumor heterogeneity was substantial, compared with intratumoral variations. With regard to the correlation coefficient of matched spots, an *r*-value of 0.50 is similar to that observed when two different tumor types (lung and colon) are examined. Our results showed that malignant ovarian tumors are relatively more heterogeneous than benign tumors with average correlation coefficients of 0.54 and 0.61, respectively. There was no significant variation among pairs of different histologic subtypes compared with pairs of gels from tumors with the same histology. The significance of this finding is unclear.

Tumor heterogeneity has a significant clinical implication in the choice of therapy. Kern *et al.* [19] examined the relationship between tumor heterogeneity and drug resistance. He reported that the majority of primary ovarian carcinomas and their corresponding metastases have similar resistance properties against cisplatin and doxorubicin. Our previous result on correlation between primary tumors (0.81) and corresponding metastasis (0.79) indicat-

ed that both have similar dominating cell populations [9]. Most malignant solid tumors are heterogeneous as reflected by differences in tumor morphology, genomic instability, degree of invasiveness as well as sensitivity to therapy. Previous studies have used immunohistological, antigenic expression changes and DNA ploidy to study heterogeneity in various malignant tumors. Polypeptide expression profile may be an effective tool to study tumor heterogeneity and may help to establish a more precise diagnosis and provide useful information to therapy sensitivity.

We are grateful to Elina Eriksson, Ann Ohlsson, Kicki Svanholm, Birgitta Sundelin, Britt-Mari Graf and Inga Maurin for technical assistance. This work was supported by the Swedish Cancer Society (Cancerfonden) and the Cancer Society in Stockholm.

Received September 1, 1998

5 References

[1] Nicolson, G. L., *Cancer Res.* 1987, *47*, 1473–1487.

[2] Kerbel, R. S., Waghorne, C., Kroczak, B., Lagarde, A., Breitman, M. L., *Cancer Surveys* 1988, *7*, 597–630.

[3] Orndal, C., Rydholm, A., Willen, H., Mitelman, F., Mandahl, N., *Cancer Genet. Cytogenet.* 1994, *78*, 127–137.

[4] Fuji, H., Marsh, C., Cairns, P., Sidransky, D., Gabrielson, E., *Cancer Res.* 1996, *56*, 1493–1497.

[5] Ponten, F., Berg, C., Ahmadian, A., Ren, Z. P., Nister, M., Lundeberg, J., Uhlen, M., Ponten, J., *Oncogene* 1997, *15*, 1059–1067.

[6] Boyer, C. M., Borowitz, M. J., McCarty, K. S. J., Kinney, R. B., Everitt, L., Dawson, D. V., Ring, D., Bast, R. C. J., *Int. J. Cancer* 1989, *43*, 55–60.

[7] Magi-Galluzzi, C., Xu, X., Hlatky, L., Hahnfeldt, P., Kaplan, I., Hsiao, P., Chang, C., Loda, M., *Modern Pathol.* 1997, *10*, 839–845.

[8] Cormier, J. N., Hijazi, Y. M., Abati, A., Fetsch, P., Bettinotti, M., Steinberg, S. M., Rosenberg, S. A., Marincola, F. M., *Int. J. Cancer* 1998, *75*, 517–524.

[9] Franzén, B., Auer, G., Alaiya, A. A., Eriksson, E., Uryu, K., Hirano, T., Okuzawa, K., Kato, H., Linder, S., *Int. J. Cancer* 1996a, *69*, 408–414.

[10] Franzén, B., Linder, S., Alaiya, A. A., Eriksson, E., Uryu, K., Hirano, T., Okuzawa, K., Auer, G., *Brit. J. Cancer* 1996b, *73*, 1632–1638.

[11] Franzén, B., Okuzawa, K., Linder, S., Kato, H., Auer, G., *Electrophoresis* 1993, *14*, 382–390.

[12] Auer, G., Caspersson, T., Wallgren, A., *Anal. Quant. Cytol.* 1980, *2*, 161–165.

[13] Rabilloud, T., Vuillard, L., Gilly, C., Lawrence, J.-J., *Cell Mol. Biol.* 1994, *40*, 57–75.

[14] Garrels, J. I., *J. Biol. Chem.* 1989, *264*, 5269–5282.

[15] Molenaar, W. M., van den Berg, E., Veth, R. P., Dijkhuizen, T., de Vries, E. G., *Genes, Chromosomes Cancer* 1994, *10*, 66–70.

[16] Schvimer, M., Lash, R. H., Katzin, W. E., *Cytometry* 1995, *22*, 292–296.

[17] Zangwill, B. C., Balsara, G., Dunton, C., Varello, M., Rebane, B. A., Hernandez, E., Atkinson, B. F., *Cancer* 1993, *71*, 2261–2267.

[18] Sasaki, K., Hamano, K., Kinjo, M., Hara, S., *Oncology* 1992, *49*, 219–222.

[19] Kern, D. H., *Cancer J. Sci. Am.* 1998, *4*, 41–45.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1047–1054

Jiří Stulík[1]
Jan Österreicher[1]
Kamila Koupilová[1]
Jiří Knížek[1]
Aleš Macela[1]
Jan Bureš[2]
Pavel Jandík[3]
František Langr[4]
Karel Dědič[4]
Peter R. Jungblut[5]

[1]Institute for Radiobiology and Immunology, Purkyně Military Medical Academy, Hradec Králové, Czech Republic
[2]Diagnostic Centre, University Hospital, Hradec Králové, Czech Republic
[3]Department of Surgery, University Hospital, Hradec Králové, Czech Republic
[4]Department of Pathology, University Hospital, Hradec Králové, Czech Republic
[5]Max-Planck-Institute for Infection Biology, Protein Analysis Unit, Berlin, Germany

The analysis of S100A9 and S100A8 expression in matched sets of macroscopically normal colon mucosa and colorectal carcinoma: The S100A9 and S100A8 positive cells underlie and invade tumor mass

The expression of calcium-binding protein S100A9 was investigated in 23 matched sets of colorectal carcinoma and normal colon mucosa using two-dimensional gel electrophoresis. We found that, from a group of 23 patients, the level of S100A9 protein, in comparison with matched normal colon mucosa, was significantly increased in malignant tissues of 16 patients (70%). Furthermore, an additional protein, identified by matrix-assisted laser desorption/ionization - mass spectrometry (MALDI-MS) as S100A8, exhibited an increased expression in the same specimens of malignant tissues as the S100A9 protein. The immunohistological analysis revealed the accumulation of S100A9 positive cells, macrophages and polymorphonuclear leukocytes along the invasive margin of colorectal carcinoma. The S100A8 protein was found to be produced in the same location. The possible participation of both proteins and, especially, its heterodimeric complex calprotectin in colorectal carcinoma regression could be taken into account.

Keywords: Two-dimensional gel electrophoresis / S100 protein family / Colon cancer EL 3319

The urgent need for very early and specific detection of colorectal carcinoma reflects the current situation regarding both the incidence of this disease and the approaches used for its diagnosis. As for the former, colorectal carcinoma, among other malignant diseases, occupies the first position in the incidence as well as mortality in men in the Czech Republic. Globally, this type of cancer is the third most common malignancy in the world [1]. The routine screening for colorectal carcinoma which involves colonoscopic examination and fecal occult blood testing is not very effective; hence, colorectal carcinoma is often diagnosed in an advanced stage with bad prognosis for successful treatment [2]. The most powerful strategies exploited for the identification of new useful markers of colorectal tumorigenesis are based either on searching for differentially expressed genes at the RNA level [3] or on the subtractive analyses of protein patterns of normal and transformed cells [4]. Recently, we used subtractive 2-DE to compare protein patterns of normal, preneoplastic, and neoplastic colon mucosa in order to detect the proteins whose expression is associated with malignant transformation of colon mucosa. With this approach, S100A9 protein, a member fo the S100 calcium-binding protein family, was found to be overexpressed in preneoplastic and neoplastic colon mucosa [5]. The S100A9 protein, formerly called calgranulin B, MRP 14, or LI heavy chain, is a protein of about 13 kDa that can occur in three different charge isoforms depending on the level of its phosphorylation [6]. This protein is found predominantly in the cytosol but it can be also expressed on the cell surface or even secreted into the surroundings. The best characterized intracellular function proposed for S100A9 is that of inhibition of casein kinase II, resulting in the regulation of normal cellular transcription and translation. The possible extracellular functions assigned to S100A9 protein include chemotactic activity on the one hand and cytotoxic/cytostatic activities against bacteria, fungi, and tumor cells on the other hand [7]. The S100A9's suppressive effect on cell growth can also be achieved with purified S100A8 protein, formerly MRP8, or with the S100A9/S100A8 heterocomplex, called calprotectin [8]. The aim of this study was to confirm and extend our recent findings concerning the differential expression of S100A9 protein between normal and neoplastic colon mucosa comparing the sets of human colorectal carcinoma with those of normal colon mucosa. Furthermore, we followed

Correspondence: Dr. Jiří Stulík, Institute for Radiobiology and Immunology, Purkyně Military Medical Academy, Třebešská 1575, 50001 Hradec Králové, Czech Republic.
E-mail: jstulik@pmfhk.cz
Fax: +0420-49-551-30-18

the immunohistochemical distribution of S100A9 protein in routine biopsies of colorectal carcinoma with sufficiently wide perilesional mucosa to find the cell source of S100A9 protein and the location of its accumulation in tumor tissue.

The samples analyzed by 2-DE consisted of 23 matched sets of colorectal carcinomas and normal colon mucosa. The sample size was large enough to exclude that the observed difference in the proportion of positive cases is caused by chance. Tissue specimens were obtained within 30 min after surgical resection. Nonnecrotic tumor tissues and control mucosa samples that were taken from each patient at 5–10 cm from tumor mass were immediately homogenized in two volumes of 9.0 M urea lysis buffer (9.0 M urea, 2.0% Triton X-100, 3.0% CHAPS, 70 mM DTT and 2% carrier ampholytes), centrifuged at 15 000 × *g* for 5 min and subsequently stored at –80°C. The protein concentration was determined by the modified bicinchoninic acid (BCA) method [9]. Established procedures for 2-DE separation and silver staining were used [5]. For the first dimension a nonlinear immobilized pH 3–10 gradient (Pharmacia-Biotech, Uppsala, Sweden) was utilized. The IPG strips were swollen by rehydration buffer containing 2 M thiourea, 7 M urea, 4% CHAPS, 0.5% Triton X-100, 0.4% Pharmalytes 3–10 and 10 mM DTT [10]. A protein amount of 200 µg was loaded on Immobiline strips. For the second dimension the proteins were separated by SDS-PAGE in a 9–16% linear polyacrylamide gradient. The isoelectric points and molecular weights of individual proteins were approximated using polypeptide 2-D SDS-PAGE standards (Bio-Rad, Richmond, CA, USA). Silver-stained gels were scanned using a laser densitometer (4000 × 5000 pixels, 12 bits/pixel; Molecular Dynamics, Palo Alto, USA) which generated 30 megabyte images. This scanner was linked to a SunSparc s-s20 workstation (Sun Microsystems, Mountain View, CA, USA). The 2-DE image computer analysis was carried out using the Melanie II package (Bio-Rad). The S100A9 and S100A8 proteins were identified by peptide mass fingerprinting. Amido black-stained spots on nitrocellulose were digested by trypsin (S. Lamer *et al.*, in preparation). The peptide mix was measured together with α-cyano-4-hydroxycinnamic acid by MALDI-MS (Voyager Elite; Perseptive, Framing-

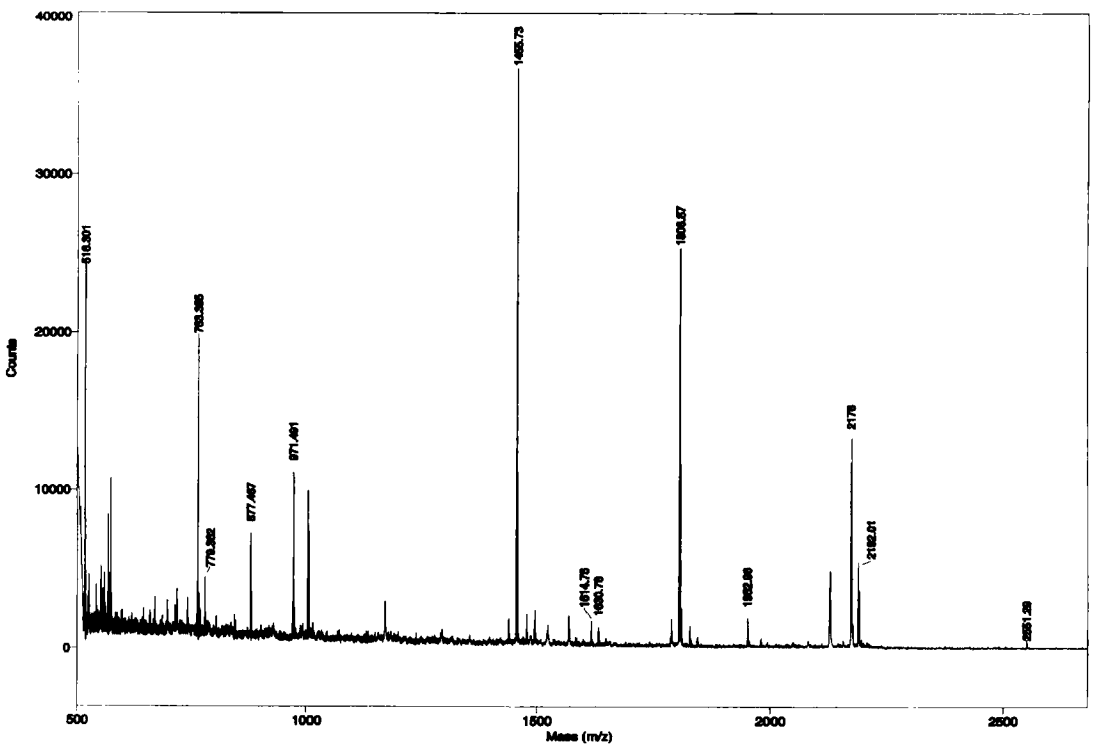


Figure 1. Peptide mass fingerprint of spot 1 (tentative S100A9) obtained by MALDI-MS. Peaks marked with mass values correspond to peptides from S100A9.

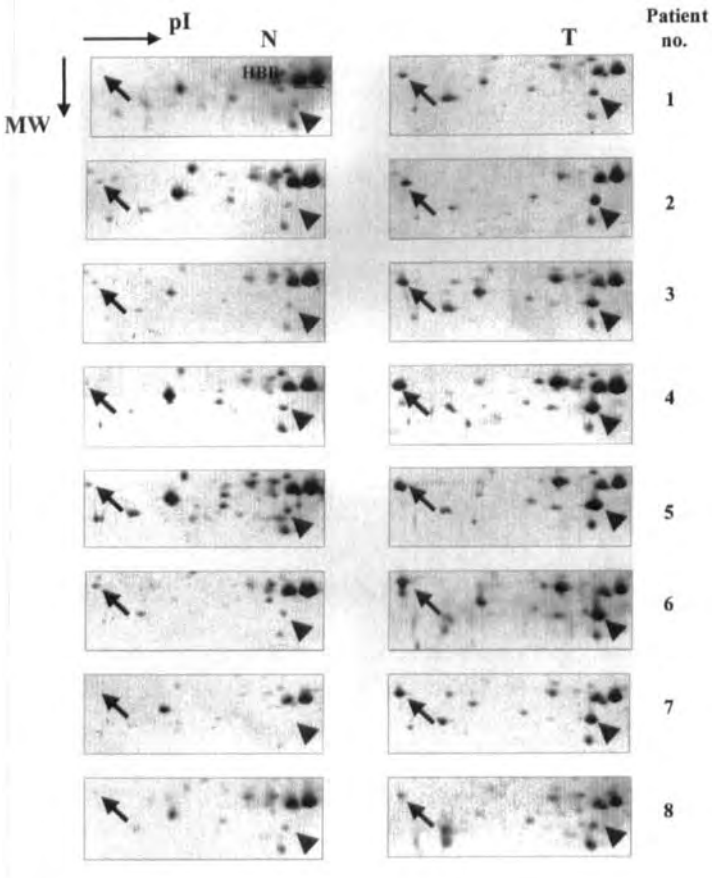


Figure 2. Silver-stained regions of 2-DE maps of matched normal and malignant tissues of 8 selected patients. N, normal control tissue; T, colorectal carcinoma. (1) Colon adenocarcinoma Dukes C; (2) colon adenocarcinoma Duke's B; (3) colon adenocarcinoma Duke's B; (4) colon adenocarcinoma Duke's C; (5) colon adenocarcinoma Duke's B; (6) colon adenocarcinoma Duke's C; (7) colon adenocarcinoma Duke's B; (8) colon adenocarcinoma Duke's B. Arrows indicate S100A9 position; arrowheads indicate the expected location of S100A8. HBB, hemoglobin β-chain.

ham, MA, USA). The resulting peptide mass fingerprint was double-internally calibrated (P. R. Jungblut *et al.*, in preparation) and searched in the NCBI sequence database with the help of MS-FIT (http://prospector.ucsf.edu/ucsfbin/msfit.cgi). For immunohistochemical studies, 4 μm tissue sections were cut from 16 formalin-fixed, paraffin-embedded blocks and examined for S100A9, S100A8 and CD68 expression using the standard peroxidase technique. After blocking for 20 min, tissue sections were incubated with individual mouse monoclonal antibodies (anti-human S100A8, anti-human S100A9 from BMA Biomedicals AG, Basel, Switzerland; and anti-human CD68 from Zymed Laboratories, San Francisco, CA, USA) for 24 h at 4°C, and washed three times in phosphate-buffered saline (PBS, pH 7.2). The slides were then incubated with horseradish peroxidase-coupled, anti-mouse antibody (Sigma, Prague, Czech Republic) for 45 min at 37°C. Excess antibody was washed off with PBS. Finally, 0.05% 3,3-diaminobenzidine tetrahydrochloride chromogen solution (Sigma, Prague, Czech Republic) in PBS containing 0.02% hydrogen peroxide was added for 10 min, at room temperature, to visualize the formed antigen-antibody complex. For the detection of polymorphonuclear leukocytes (PMN), chloroacetate esterase staining was used. The staining intensity was scored using computerized image analysis (Lucia M software, Version 3.00; Laboratory Imaging, Prague, Czech Republic). The density of S100A9 and S100A8 staining was measured for each antibody in six different locations of each biopsy (area, 10 000 pixels; original magnification × 125) selected from the central part of the tumor and the margin of tumor. The acquired values were used for statistical analysis. Negative staining controls were obtained by incubating with normal mouse serum purchased from Dako. Comparison between two characteristics was assessed using a nonpaired Student T-test. Relationships were considered statistically significant when $p < 0.05$. The samples of colorectal carcinomas for both electrophoretic as well as immunohistochemical studies were collected from various locations in the colon (right, transverse, and

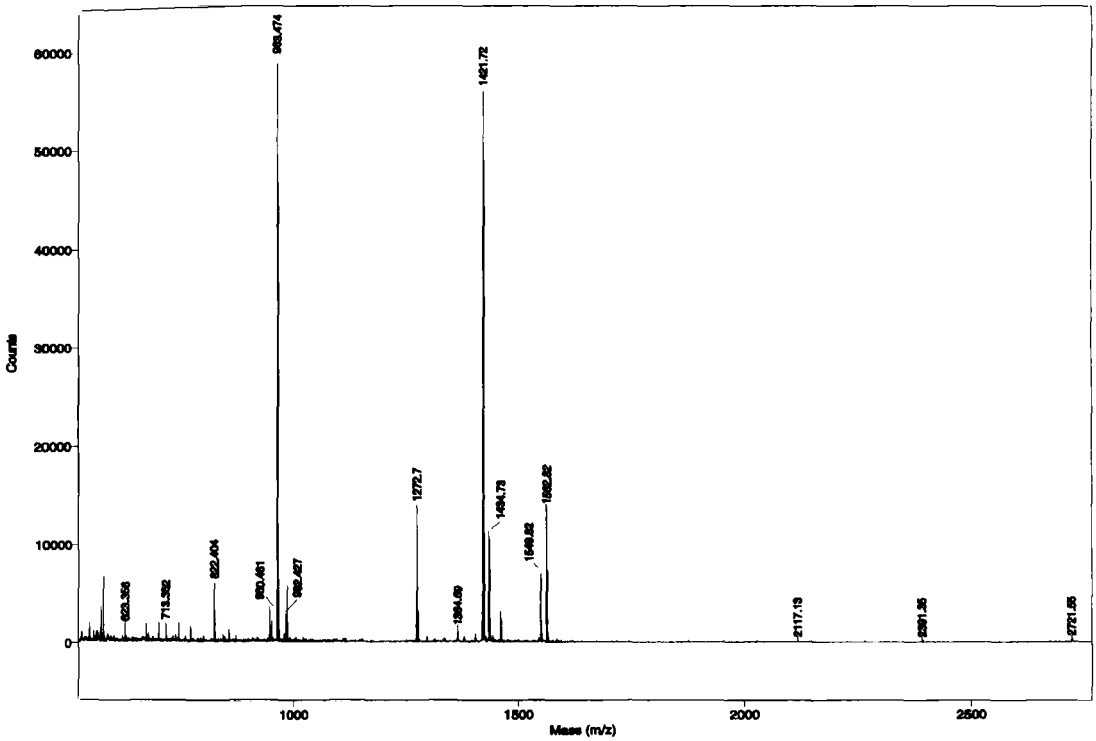
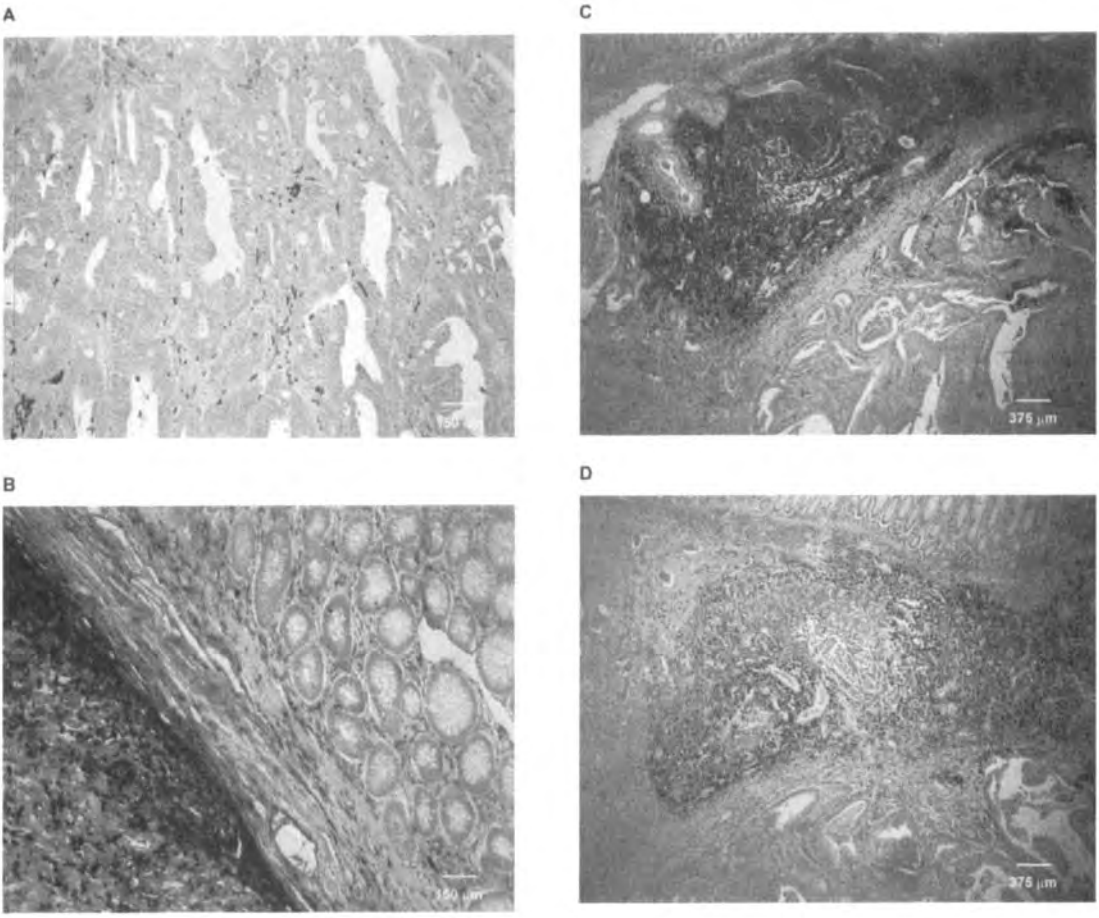


Figure 3. Peptide mass fingerprint of spot 2 (tentative S100A8) obtained by MALDI-MS. Peaks marked with mass values correspond to peptides from S100A8.

left colon) and were staged according to Duke's classification. None of the patients in this group was receiving any drug therapy.

In our previous study we demonstrated, compared to normal colon mucosa, the upregulation of S100A9 protein in 12 of the 15 analyzed samples of colorectal carcinomas. However, the following 2-DE analyses of chronically inflamed colon mucosa also documented the high S100A9 expression in patients with Crohn's disease and ulcerative colitis [5]. Hence, the specificity of tumor-associated S100A9 expression needed further investigation. In this study we examined the S100A9 expression in 23 specimens of identical cancer/normal pairs. Furthermore, to obtain more proteins on 2-DE maps, thiourea-containing buffer [10] that improves protein solubilization was applied. However, the utilization of this new type of rehydration buffer changed our 2-DE patterns of colon mucosa. We were not able to overlap precisley these new 2-DE gels with the images of previous gels [5] and, additionally, it was also impossible to compare our results with the reference maps of colon carcinoma cells available on the Internet (ExPASy molecular biology server or 2-DE gel protein database of colon carcinoma cell line LIM 1215). Therefore, we used MALDI-MS to verify the S100A9 position on our 2-DE. The peptide mass fingerprint of the spot, which occurred in the expected position of S100A9, matched with 13 (all major peaks; Fig. 1) of 19 masses and a sequence coverage of 75% with calcium-binding protein S100A9. Comparing 2-DE protein patterns of "macroscopically" normal tumor-adjacent tissues with neoplastic colon mucosa we observed that from the group of 23 patients the level of S100A9 protein was significantly increased in malignant tissues collected from 16 patients (70%; Fig. 2). There was no relationship between the disease severity and the S100A9 overexpression because the S100A9 positive samples were nearly evenly distributed among groups B ($n = 7$ cases), C ($n = 5$ cases) and D ($n = 4$ cases) of Duke's classification. In the remaining seven patients, either similar levels of S100A9 protein (5 tumors: 4 were in Duke's B and 1 in Duke's C) were detected in matched normal/malignant pairs, or (in 2 patients; tumors in Duke's D stage) an even higher amount of S100A9 protein in normal tumor-adjacent tissue compared to matched malignant tissue was found (unpublished observation). These findings are in contrast



with our previous results describing the lack of or only weak S100A9 expression in normal colon mucosa [5]. However, this discrepancy could be explained by the potential difference between control samples that were used before, and which were collected from patients who were examined for gastrointestinal bleeding but were found to have normal colonic mucosa, and macroscopically normal tumor-adjacent tissue used as control material now. Furthermore, there was an additional protein whose level changed in malignant tissues in the same way as the S100A9 (Fig. 2). This protein had approximate values of both molecular mass (10 kDa) and p*I* (6.7) close to the parameters of the S100A8 protein. As was mentioned above, S100A8 forms a protein heterodimer, called calprotectin, with S100A9 [7], and fecal calprotectin was recently suggested as a new stool marker for diagnosis of colorectal carcinoma, which should be less erratic than bleeding [11, 12]. The identity of this spot was also proved by MALDI-MS. In this case, the resulting peptide mass fingerprint matched with 14 (all major peaks; Fig. 3) of 21 masses and a sequence coverage of 97% with calcium-binding protein S100A8.

Next, we followed the immunohistochemical distribution of S100A9 protein in formaline-fixed biopsy material in order to find any difference either in cell origin or S100A9 accumulation between neoplastic and inflammatory conditions. When we examined the biopsies containing only a central part of the tumor mass we observed a diffuse S100A9 staining that was dependent on the extent and composition of inflammatory infiltrate (Fig. 2a). Nearly the same picture was obtained for normal colon mucosa (Fig. 2b) and Crohn's disease with moderate disease activity (Fig. 2g). However, when the biopsies of colorectal carcinoma with wide perilesional mucosa was examined, then the high staining intensity of S100A9 protein along the invasive margin of carcinoma with some S100A9 positive cells infiltrating the tumor mass was found (Fig. 2b, c).

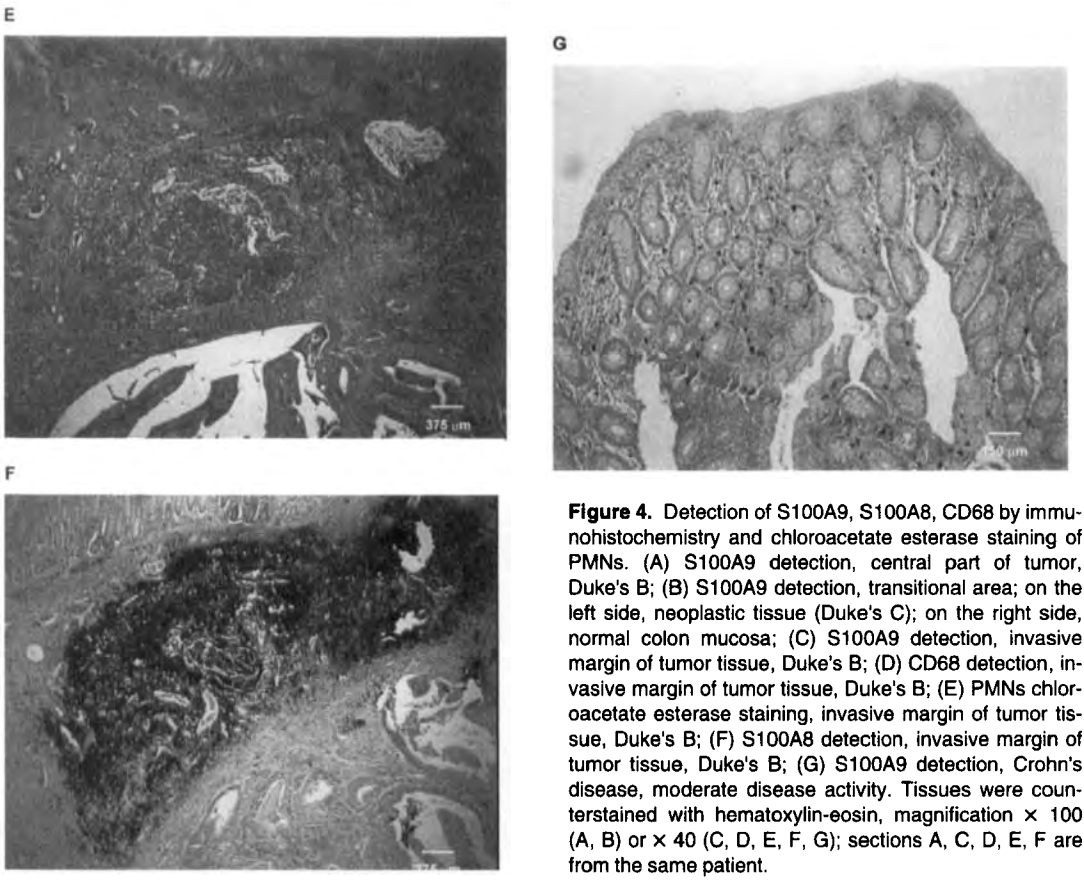


Figure 4. Detection of S100A9, S100A8, CD68 by immunohistochemistry and chloroacetate esterase staining of PMNs. (A) S100A9 detection, central part of tumor, Duke's B; (B) S100A9 detection, transitional area; on the left side, neoplastic tissue (Duke's C); on the right side, normal colon mucosa; (C) S100A9 detection, invasive margin of tumor tissue, Duke's B; (D) CD68 detection, invasive margin of tumor tissue, Duke's B; (E) PMNs chloroacetate esterase staining, invasive margin of tumor tissue, Duke's B; (F) S100A8 detection, invasive margin of tumor tissue, Duke's B; (G) S100A9 detection, Crohn's disease, moderate disease activity. Tissues were counterstained with hematoxylin-eosin, magnification × 100 (A, B) or × 40 (C, D, E, F, G); sections A, C, D, E, F are from the same patient.

This S100A9 accumulation was statistically significant for 15 of 16 examined cases (Fig. 5a). The biopsies involved 11 specimens that were analyzed in parallel by 2-DE and four discordant cases were detected between the two methods. These four specimens showed either similar expression of S100A9 in both normal tissue and carcinoma, or an even higher level in normal colon mucosa, but the tumor had a stronger marginal S100A9 expression than the normal mucosa by immunohistochemistry. As a possible explanation of this difference the lack of this boundary area in resected pieces of tumor tissue used for 2-DE analyses could be considered. We also determined the cell origin of S100A9 in neoplastic colon mucosa. It is known that this protein is abundant in circulating human neutrophils and monocytes [7]; therefore, we performed both immunohistochemical detection of CD68 antigen that should reside preferentially in cytoplasm of macrophage-lineage cells and chloroacetate esterase staining of polymorphonuclear leucocytes (PMNs). As can be seen from Fig. 2d and e, both types of cells occur to a different extent in the area with the maximal S100A9 positivity; hence, they all contribute to S100A9 production. Finally, we also searched for the S100A8 expression directly in bioptic specimens. Figure 2f shows that S100A8 co-localized with S100A9 in the marginal part of the tumor lesion. Likewise, in S100A9 accumulation in tumor tissue the statistical analysis showed a significant difference in the expression of S100A8 between the marginal and central part of colorectal carcinoma in 15 of 16 cases (Fig. 5b). The findings of S100A9 and S100A8 positive macrophage and PMNs which underlie and invade colorectal carcinoma are in accordance with the published data stressing the important role of these cells in immune reaction occurring along the invasive margin of colorectal carcinoma. This immune reaction comprises, on the one hand, the specific part in which macrophages activating T cells along an invasive margin participate [13] and, on the other hand, the nonspecific immune reaction aimed at the reduction of local tumor growth [14]. As for the latter, the production of S100A9 and S100A8 proteins by macro-

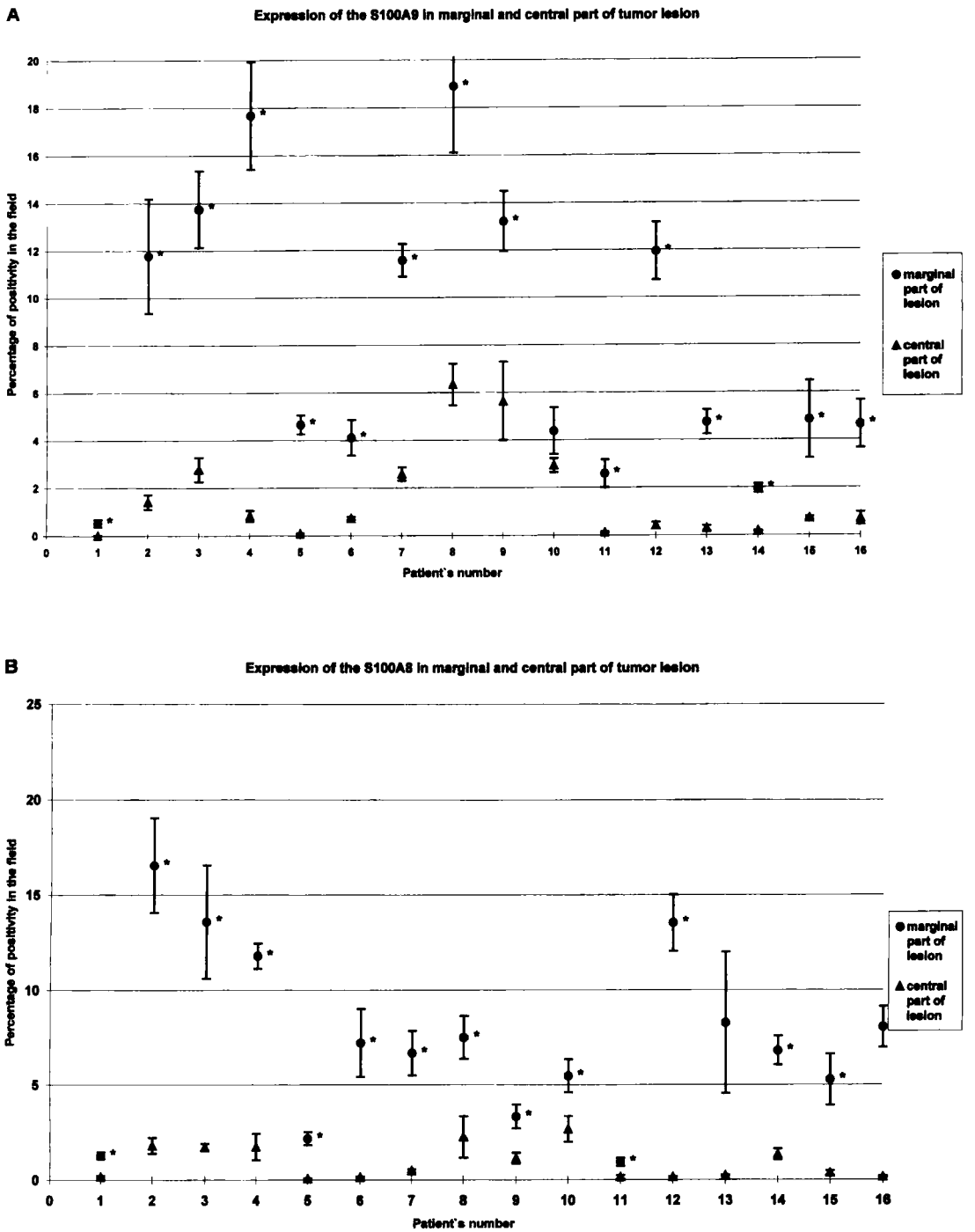


Figure 5. The S100A9 (A) and S100A8 (B) expression in marginal and central parts of tumor lesions. The dots are the mean, and the bars are the SD *, $p < 0.05$ compared with the central part of the tumor.

phages and PMNs can be decisive owing to the recently described apoptosis-inducing activity of calprotectin in tumor cells [15].

In conclusion, we extended our results describing the overexpression of S100A9 in neoplastic lesions of colon mucosa. Furthermore, we proved by means of an immunohistological analysis that S100A9 positive cells, macrophages, and PMNs accumulate along the invasive margin of colorectal carcinoma and infiltrate tumor tissue. This distribution stands in contrast with diffuse staining of S100A9 protein in normal and chronically inflamed colon mucosa. Finally, S100A8 protein is produced in the same location as S100A9, which can lead to calprotectin formation and possible apoptotic regression of tumor mass.

The authors thank Jana Michaličková, Alena Firychová, Jitka Žáková, Šárka Průchová and Stephanie Lamer for excellent technical assistance. This work was supported by grants from the Ministry of Health, Czech Republic (project No. 4075-3) and from the Grant Agency of the Czech Republic (project No. 310/98/P238).

Received September 1, 1998

References

[1] Klevit, J., Bruinvels, D. J., *Eur. J. Cancer* 1995, *31*, 1222–1225.

[2] Gilbert, J. A., Ahlquist, D. A., Mahoney, D. W., Zinsmeister, A. R., Rubin, J., Ellefson, R. D., *Scan. J. Gastroenterol.* 1996, *31*, 1001–1005.

[3] Van Belzen, N., Dinjens, W. N. M., Diesveld, M. P. G., Groen, N. A., Van der Made, A. C. J., Nozawa, Y., Vliestra, R., Trapman, J., Bosman, F., *Lab. Invest.* 1997, *77*, 85–92.

[4] Ji, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D., Simpson, R. J., *Electrophoresis* 1994, *15*, 613–619.

[5] Stulík, J., Kovarova, H., Macela, A., Bures, J., Jandík, P., Langer, F., Otto, A., Thiede, B., Jungblut, P., *Clin. Chim. Acta* 1997, *265*, 41–55.

[6] Guignard, F., Mauel, J., Markert, M., *Eur. J. Biochem.* 1996, *241*, 265–271.

[7] Hessian, P. A., Edgeworth, J., Hogg, N., *J. Leukoc. Biol.* 1993, *53*, 197–204.

[8] Yui, S., Mikami, S., Yamazaki, M., *J. Leukoc. Biol.* 1995, *58*, 650–658.

[9] Hill, H. D., Straka, J. G., *Anal. Biochem.* 1988, *170*, 203–208.

[10] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[11] Gilbert, J. A., Ahlquist, D. A., Mahoney, D. W., Zinsmeister, A. R., Rubin, J., Ellefson, R. D., *Scand. J. Gastroenterol.* 1996, *31*, 1001–1005.

[12] Kristinsson, J., Roseth, A., Fagerhol, M. K., Aadland, E., Schjonsby, H., Bormer, O. P., Raknerud, N., Nygaard, K., *Dis. Colon Rectum* 1998, *41*, 316–321.

[13] Ohtani, H., Naito, Y., Saito, K., Nagura, H., *Lab. Invest.* 1997, *77*, 231–241.

[14] Thomas, C., Nijenhuis, A. M., Dontje, B., Daemen, T., Scherphof, G. L., *J. Leukoc. Biol.* 1995, *57*, 617–623.

[15] Mikami, M., Yamazaki, M., Yui, S., *Microbiol. Immunol.* 1998, *3*, 211–221.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1055–1064

Roberta Melis
Ray White

Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

Characterization of colonic polyps by two-dimensional gel electrophoresis

To identify proteins that may specifically characterize colonic polyps we have investigated the abundance of numerous proteins in epithelial cells from 15 normal colon specimens and 13 colonic polyps, using two-dimensional gel analysis to detect possible differences in expression. Silver-stained digitized images of the gels were analyzed with Melanie II 2.1 software. We consistently detected more than 700 protein spots on each gel, and found that the intensity of 59 of them was significantly altered in polyp specimens (Wilcoxon test assuming $P \leq 0.05$). Immunostaining, microsequencing and mass spectrometry (matrix-assisted laser desorption/ionization – time of flight; MALDI-TOF) techniques were used to identify these proteins and selected others that did not show differential regulation. The expression of numatrin (nucleophosphine/B23), hsp 70, and hsp60 was increased in polyps; levels of fatty acid binding protein (L-FABP), 14-3-3 σ, citokeratin 20, cytochrome *c* oxidase polypeptide Va, Rho GDP-dissociation inhibitor (Rho GDI), and β- and γ-actins were decreased. Although the levels of expression of a given protein often varied among polyp specimens, it generally held true that the direction of variation (up or down) remained constant across the panel. We concluded that proteins showing constant differential regulation across all or most of the polyp specimens represent the most characteristic regulatory pathways in colon polyps, while more sporadic variations reflect characteristics of individual polyps.

Keywords: Two-dimensional gel electrophoresis / Colonic polyps / Differential protein expression
EL 3409

1 Introduction

More than 90% of colon cancers reflect no obvious hereditary pattern, and they affect most ethnicities and both genders at similar frequencies. Colon cancer is curable by surgical resection if identified early and prophylactic removal of polyps is advised [1]. Although individuals with a family history represent only a small proportion of all colon cancer patients, relatives of such patients carry a significantly increased risk for developing this disease [2]. Colonic tumors arise from the epithelial lining of the colon [3–5] and progress through a premalignant stage known as a polyp [6–8]. Some of the fundamental genetic changes that drive the development of colon cancer are believed to occur at the polyp stage; however, only a minority of polyps undergo malignant transformation. Cell lines established from cancerous colonic epithelium are available, but they are not considered adequate to model the intact epithelium because they have undergone changes and selection to facilitate long-term growth *in vitro*. To date, numerous investigations have analyzed colonic epithelial cells from normal and cancerous tissues, and from colon cancer cell lines, by two-dimensional electrophoresis [9–16]. The present study reports the application of the two-dimensional technology to identify global differences in expression of proteins extracted from normal mucosa as compared to polyp specimens as a means of detecting specific changes that characterize the polyp stage.

Correspondence: Dr. Roberta Melis, Huntsman Cancer Institute, University of Utah, 15 North 2030 East #7480, Salt Lake City, UT 84103, USA
E-mail: roberta.melis@hci.utah.edu
Fax: +801-585-3833

2 Materials and methods

2.1 Tissue specimens

Samples of normal colonic mucosa and polyps were obtained from the Departments of Surgery at University Hospital, the Veterans Administration Hospital, LDS Hospital and St. Mark's Hospital in Salt Lake City. Normal mucosal tissues, collected from patients undergoing resection of carcinomas of the colon, were recovered from the healthy margin most distal from the tumor. Polyps were obtained either from explorative colonoscopies or colectomies. Clinical information and description of the polyps are summarized in Table 1. All tissue samples were collected and stored in cold RPMI 1640 containing 15% fetal bovine serum, penicillin, streptomycin, fungizone, and gentamycin sulfate.

2.2 Isolation of colonic crypts and preparation of protein samples

Colonic crypts were isolated from the normal mucosa and from polyps following a procedure modified from White-

Table 1. Summary of pathology reports of the analyzed polyps

Patient age/ gender	Associated clinical finding	Procedure	Polyp No.	Size (cm)	Diagnosis
49/M	Colon cancer	Colectomy	1	0.6	Tubular adenoma
10/M	Familial polyposis syndrome	Colectomy	2/a	0.7	Adenomatous polyp
			2/b	0.5	Adenomatous polyp
53/F	Familial polyposis syndrome	Colectomy	3/a	0.5 × 0.7 × 0.7	Adenomatous polyp
			3/b	1.8 × 1.7 × 0.6	Adenomatous polyp
			3/c	1.3 × 1.1 × 1.1	Adenomatous polyp
58/M	Colonic polyp	Colonoscopy	4/a	0.8 × 0.6 × 0.2	Adenomatous polyp
			4/b	2 × 2.3 × 1.7	Tubulovillous adenoma
50/M	Colon cancer	Colonoscopy	5	6.3 × 5.6	Tubulovillous adenoma With invasive carcinoma
72/M	History of colonic polyp	Colonoscopy	6	2.5 × 2.0 × 1.0	Tubulovillous adenoma
59/M	None provided	Colonoscopy	7	5	Villous adenoma
79/M	History of colonic polyp	Colonoscopy	8	1.8 × 1.2 × 0.5	Tubulovillous adenoma
77/M	History of colonic polyp	Colonoscopy	9	2.2 × 1.2 × 0.9	Tubular adenoma

head [17]. Briefly, tissue samples were washed three times in PBS in the absence of calcium and magnesium, and incubated 30–60 min at room temperature in PBS containing 0.5 mM DTT and 3 mM EDTA. Fresh PBS replaced the solution, and the sample was vigorously shaken for 20–30 s to liberate the crypts from the stroma. The crypts were collected by centrifugation at 400 × *g* for 5 min. Aliquots of pelleted crypts were either frozen in dry-ice ethanol and stored at –80°C, or immediately resuspended in ten volumes of lysis buffer containing 9.5 M urea, 4% CHAPS, 65 mM DTT, complete protease inhibitors (Boehringer Mannheim, Indianapolis, IN) and 2% Pharmalyte pH 3–10 (Amersham Pharmacia Biotech, Piscataway, NJ). The cells were homogenized by passing them five times through a 26-$G^{1/2}$ needle. To complete the lysis, the samples were left at room temperature for 30 min. Cellular debris and insoluble materials were pelleted by centrifugation at 45 000 rpm for 15 min in a Beckman TLS55 bucket rotor (Beckman, Fullerton, CA). The protein concentration of each specimen was determined by the Bradford method [18] using Bio-Rad (Richmond, CA) color reagents. To identify which proteins might specifically characterize the colonic polyps, we compared the levels of proteins from epithelial cells from 15 normal colon specimens to 13 colonic polyps resolved by two-dimensional gel technology.

2.3 Analytical 2-D PAGE

The first dimension was performed following the protocol reported by Görg [19]. Precast Immobiline DryStrips, pH 4–7 linear, 180 × 3 × 0.5 mm (Amersham Pharmacia Biotech), were rehydrated with 100 µg of protein sample [20] overnight in a solution containing 8 M urea, 0.5 % CHAPS, 65 mM DTT, 0.8% Pharmalytes pH 3–10. Isoelectric focusing was carried out on a Multiphor II electrophoresis unit (Amersham Pharmacia Biotech) using a gradient mode, 500 V for 500 Vh, 500 V for 10 000 Vh and 3500 V for 60 000 Vh at 20°C. The first-dimension gels were equilibrated for 15 min in 6 M urea, 2% SDS, 50 mM Tris-HCl, pH 6.8, 30% v/v glycerol and 1% DTT, then for 15 min in the same solution except that DTT was replaced by 5% w/v iodoacetamine. The second dimension was

Table 2. Procedure for silver staining

Prefix I	50% methanol 7% acetic acid	overnight
Prefix II	10% methanol 10% acetic acid	40 min
Rinse	double distilled H_2O	2 × 10 min
Fix/ sensitize	0.05% glutaraldehyde (25%) 0.1% formaldehyde (37%) 0.5 M sodium acetate 20% ethanol	20 min
Rinse	double distilled H_2O	2 × 10 min
Staining		20 min
To prepare 150 mL solution of ammoniacal silver nitrate: 0.375 g of silver nitrate was dissolved in 10 mL of double distilled H_2O, which was slowly mixed into a solution containing 140 mL, 0.6 mL NH_4OH, 0.0360 g NaOH.		
Rinse	double distilled H_2O	4 × 4 min
Develop	2.5% Na_2CO_3 0.04% formaldehyde	
Stop	0.05 M EDTA 0.02% thimerosal	15 min

Staining was carried out in glass trays, using 150 mL of each different solution per gel, and 600 mL of double distilled H_2O per gel in the rinsing steps.

performed using precast 12–14% polyacrylamide ExcelGel SDS XL 245 × 180 × 0.5 mm (Amersham Pharmacia Biotech) on a Mutiphor II apparatus with precast anodic and cathodic SDS buffer strips (Amersham Pharmacia Biotech), for 1 h and 20 min at 300 V, and at 800 V, at 15°C, until the bromophenol blue reached the bottom of the gel. Resolved proteins were visualized by silver nitrate staining using a combination of Celis and Vari protocols [21, 22] (Table 2). After staining, the gels were separated from the polyester support and dried under vacuum between two sheets of cellulose (Amersham Pharmacia Biotech), for 2 h at 45°C. In the first dimension the pH gradient profile was established using the predicted values on 18 mm long IPG strips (pH 4–7). SDS-PAGE size markers (Amersham Life Science, Cleveland, OH) were used in the second dimension.

2.4 Preparative 2-D PAGE

The first and second dimensions of the preparative gels were performed as above except that 300–400 μg of protein was loaded in each gel.

2.5 Scanning and image analysis

The silver-stained 2-D gels were digitized using a computing densitometer Scanner 3 + (Millipore, Bedford, MA), and the images were imported in TIF format to Melanie II 2.1 image analysis software (Bio-Rad, Hercules, CA) for spot detection, quantitation and matching. The intensity of each protein spot was expressed as relative volume (volume of the spot divided by the total volume over the whole gel) to normalize for differences in gel loading and staining between different gels.

2.6 Identification of protein spots

Proteins were identified using a combination of methodologies that included two-dimensional immunoblotting, microsequencing (*N*-terminal sequence) and mass spectrometry (MALDI-TOF).

2.7 Immunoblotting and immunodetection

After two-dimensional electrophoresis the polyacrylamide gels were electroblotted onto PVDF Trans-blot membrane (Bio-Rad) according to Ji *et al.*l [10]. Immunodetection and antigen detection with enhanced chemiluminescence (ECL; Amersham Life Science) were performed according the manufacturer's protocols.

2.8 Antibodies

The different antibodies used included: monoclonal numatrin antibody (dilution 1:5000; kindly provided by Dr. Mark Olson); monoclonal anti-Hsp 70 (dilution 1:5000; Sigma Immunochemicals, St. Louis, MO); anti-α-tubulin (dilution 1:2000; ICN Biochemicals, Costa Mesa, CA); monoclonal anti-cytokeratin 18 (dilution 1:1000; Sigma); and monoclonal anti-cytokeratin 19 (dilution 1:100; Sigma). Secondary antibodies were murine anti-IgG1 and IG2a conjugated with horseradish peroxidase (HRP; dilution 1:20 000; Zymed Laboratories, San Francisco, CA), and HRP-conjugated rabbit anti-IgG (dilution 1:3000; Zymed).

2.9 Microsequencing

Protein spots from one or more 2-D gels were excised from Coomassie blue-stained PDVF membranes as described by Ji *et al.* [10] and sequenced by Edman degradation on an ABI Precise sequencer (Applied Biosystems, Perkin Elmer-Cetus Foster City, CA).

2.10 In-gel tryptic digestion

The 2-D gels were silver-stained following the original protocol of O'Connell [23], which does not interfere with mass spectrometric analysis. After staining, gels were dried between two cellulose sheets and stored at –20°C. Each spot of interest was excised from the dry gel and immersed in a drop of distilled water to separate the gel from the cellulose. The gel pieces were recovered, minced, and dehydrated in acetonitrile. Reduction, alkylation, trypsin digestion (Boehringer Mannheim), and extraction of trypsin-digested peptides from the gel were performed according to the protocol of Shevchenko [24].

2.11 Mass spectrometry (MALDI-TOF)

Each mixture of trypsin-digested peptides was dissolved in 5.0 μL of 5 % aqueous formic acid. About 0.5 μL of each peptide solution was placed onto a thin film of α-cyano-5-hydroxycinnamic acid deposited on the sample plate [25, 26]. For internal calibration, about 0.5 μL of peptide solution and 0.5 μL of standard peptide (ACTH 4–10 and ACTH 18–39, 100 fmol each) were placed onto a thin film of α-cyano-5-hydroxycinnamic acid. All spectra were collected on a MALDI-TOF mass spectrometer (Voyager-DE STR Biospectrometry Workstation; PerSeptive Biosystems, Framingham, MA) with a nitrogen laser operating at a wavelength of 337 nm. The instrument was operated at 20 kV accelerating voltage in the reflectron mode with delayed extraction. Under these conditions, the resolution was greater than 10 000 ppm. When internal calibration was used, the mass accuracy was greater than 30 ppm. With external calibration, the mass accuracy was better than 100 ppm.

2.12 Identification of proteins by mass fingerprinting and database searches

Mass fingerprinting of peptides was performed using either the ProteinProspector algorithm (URL: http://prospector.ucsf.edu/) or Profound (URL: http://prowl.rockefeller.edu/cgi-bin/ProFound), both available through the internet.

2.13 Statistical methods

P-values for differences between normal and polyp tissues were calculated for each of 700 spots using a nonparametric (Wilcoxon) test.

3 Results

3.1 Two-dimensional gel electrophoresis of colonic epithelial cells

Whole cell lysates from normal colonic mucosa and polyps, enriched in epithelial cells, were analyzed by isoelectric focusing (IEF) in the first dimension, then by SDS-PAGE in the second dimension. The protein profile of a colonic polyp is displayed in Fig. 1 in a pH range of 4–7 and M_r 200-10.

3.2 Protein expression level in normal colonic mucosa and polyps

More than 700 protein spots were consistently detected on each gel by this method. Using the 15 normal samples, we established a reference map of proteins present in normal colonic epithelium. Among the 15 2-D electropherograms from normal colonic mucosa, the one with the highest number of spots was chosen as a reference. All gels were then matched to this map. By comparing the 2-D protein pattern of polyp samples to the normal, we found that 59 (7.8%) proteins were significantly altered in terms of abundance (Wilcoxon test $P \leq 0.05$). Normal and polyp tissues, being from different patients, were analyzed as nonpaired samples. The differentially expressed spots are indicated on the reference gel (Fig. 2), and described in Table 3. Examples of relative expression levels of some of these proteins, across the panel of polyp samples, are represented in Fig. 3.

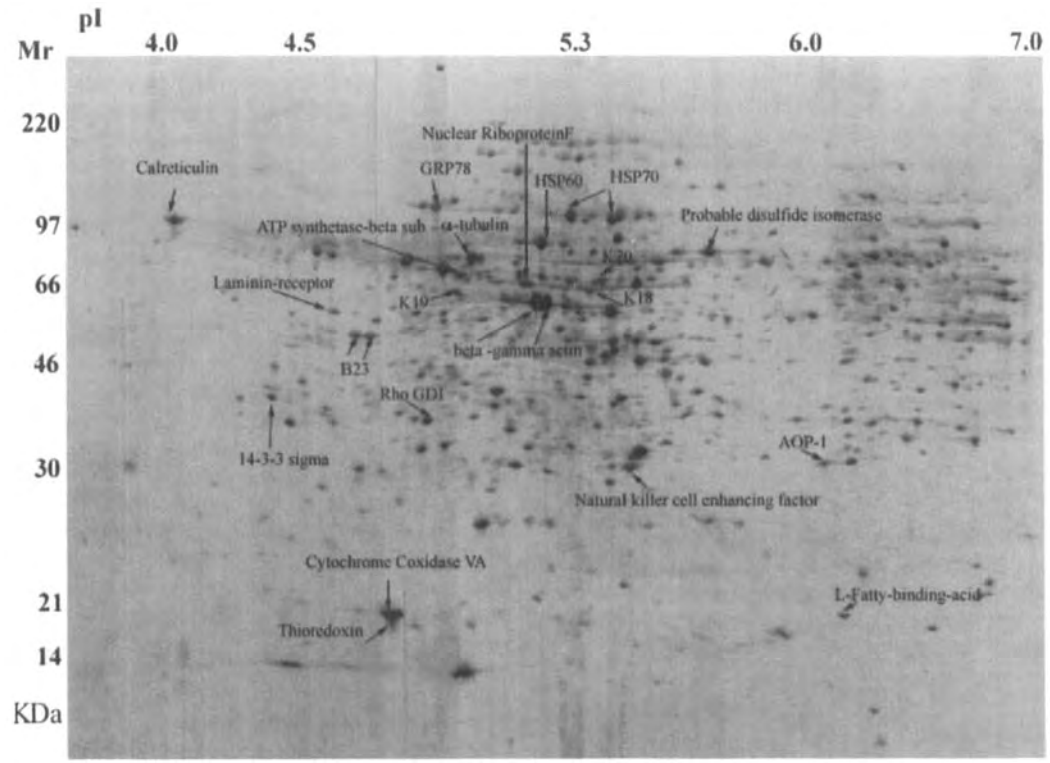


Figure 1. Silver-stained 2-D IPG gel (pH 4–7) of colonic-epithelial cells isolated from polyps. Spots for 23 proteins are identified.

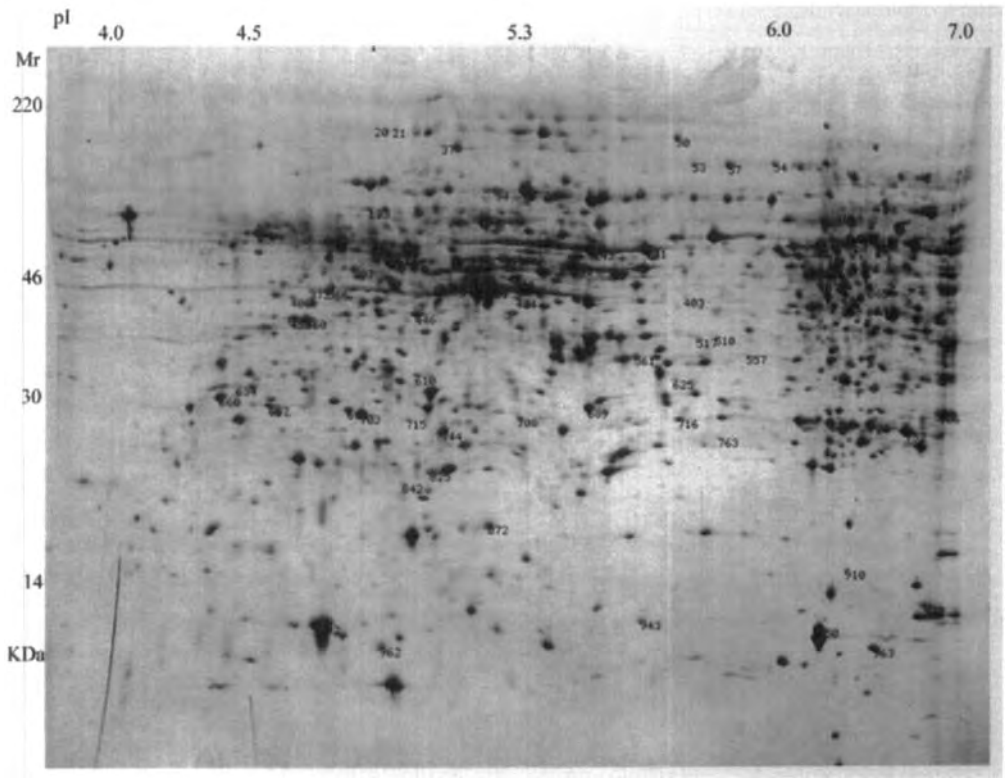


Figure 2. 2-D electropherogram (reference gel) of proteins from normal colonic mucosa image from Melanie software. Differentially expressed proteins are indicated by their spot numbers in our database.

3.3 Identification of protein spots from two-dimensional gels

From the 59 proteins spots that were found differentially expressed, eleven were identified by immunodetection, *N*-terminal sequencing and mass spectrometry, along with 12 arbitrarily chosen, nonaltered spots. The correct assignments of these identified proteins were confirmed by matching to reference gels in SWISS-2DPAGE. The identified protein spots are described in Table 4.

3.4 Proteins differentially regulated in polyp

Among the proteins that were up-regulated in polyps we identified the phosphoprotein numatrin (nucleophosmin/B23), and heat-shock proteins HSP70 and HSP60. The down-regulated proteins included fatty acid-binding protein (L-FABP), keratin 20, 14-3-3σ (stratifin), Rho GDP dissociation inhibition protein, cytochrome *c* oxidase polypeptide Va, and β- and γ-actins.

4 Discussion

To better understand development and progression of the earliest stage of colon cancer, we analyzed and compared specimens of colon epithelial cells isolated from normal and polyp tissue samples for their patterns of protein expression. To increase the sensitivity of protein detection, we first considered ^{35}S-metabolic labeling of the tissues as an alternative to silver staining, but several considerations made that technique useless for our experiments. For example, colonic tissue degenerates rapidly outside of the living host; within a few hours of incubation in culture media at 37°C, major changes occur in the tissue structure that prevent extraction of the crypts (unpublished observation). Labeling of isolated crypts is not recommended either, because data from the literature have shown that disruption of interaction between epithelial cells and matrix induces "anoikis", a specific form of apoptosis [27, 28]. Therefore we chose to observe general changes between normal and polyp samples, which could be attributed to changes in the cellular environment; these changes were quantitative rather than qualitative.

Table 3. Proteins differentially expressed in polyps

Protein identification	Spot number	M_r (kDa)	p*I*	Protein expression	Normal ($n = 15$) %Vol mean	(±) SD of the mean	Polyp ($n = 13$) %Vol mean	(±) SD of the mean	CP/CN	Wilcoxon $P \leq 0.05$	Detected in Normal	Polyp
NI[a)]	S-273	50	5.2	Down	0.363	0.053	0.101	0.014	0.3	0.000	15	12
NI	S-373	40	5.5	Down	0.036	0.007	0.003	0.002	0.1	0.000	13	2
NI	S-243	60	5.52	Down	0.899	0.300	0.269	0.042	0.3	0.000	15	13
NI	S-715	31	5.1	Up	0.052	0.008	0.125	0.015	2.4	0.000	13	13
NI	S-20	150	4.89	Up	0.015	0.005	0.083	0.008	5.6	0.000	7	13
NI	S-21	150	4.92	Up	0.022	0.005	0.077	0.007	3.5	0.000	9	13
NI	S-929	14.3	> 6.5[b)]	Down	1.628	0.215	0.479	0.097	0.3	0.000	15	11
NI	S-763	30	5.9	Up	0.023	0.004	0.051	0.005	2.2	0.000	12	13
NI	S-404	37	5.28	Up	0.018	0.006	0.052	0.006	2.8	0.001	7	12
NI	S-37	150	5.1	Up	0.011	0.006	0.060	0.009	5.3	0.001	4	11
Keratin 20	S-294	45	5.48	Down	0.729	0.097	0.329	0.025	0.5	0.001	15	13
NI	S-344	45.5	5.32	Down	0.245	0.027	0.116	0.019	0.5	0.001	15	12
NI	S-510	34	5.8	Up	0.023	0.006	0.085	0.015	3.7	0.001	10	12
NI	S-625	33	5.78	Up	0.043	0.007	0.119	0.017	2.8	0.001	14	13
Rho GDP	S-703	30	4.8	Down	0.443	0.026	0.287	0.024	0.6	0.002	15	13
NI	S-723	30	6.29	Down	0.230	0.024	0.117	0.022	0.5	0.003	14	10
NI	S-53	97	5.78	Up	0.040	0.012	0.097	0.018	2.4	0.003	12	13
NI	S-57	97	5.9	Up	0.075	0.015	0.155	0.021	2.1	0.004	13	13
14-3-3 sigma	S-660	30	4.49	Down	0.323	0.024	0.160	0.028	0.5	0.004	15	13
NI	S-962	12	4.9	Down	0.107	0.010	0.042	0.010	0.4	0.004	15	9
L-FAB	S-950	13	6.1	Down	0.154	0.033	0.009	0.005	0.1	0.005	11	3
NI	S-842	29	5.1	Up	0.060	0.009	0.101	0.007	1.7	0.005	15	13
NI	S-30	150	5.75	Up	0.054	0.010	0.106	0.016	2.0	0.005	13	12
NI	S-744	30	5.12	Up	0.189	0.017	0.294	0.025	1.6	0.007	15	13
NI	S-446	37	5	Down	0.265	0.018	0.171	0.020	0.6	0.008	15	13
Cytochrome oxydase	S-942	12	4.7	Down	1.950	0.171	1.127	0.128	0.6	0.008	15	13
NI	S-706	29	> 6.5	Down	0.353	0.046	0.147	0.022	0.4	0.010	15	11
NI	S-910	15	6.2	Down	0.030	0.007	0.005	0.004	0.2	0.012	11	2
NI	S-708	31	5.35	Down	0.116	0.017	0.058	0.009	0.5	0.012	15	12
NI	S-366	38	4.76	Down	0.117	0.018	0.049	0.011	0.4	0.015	15	9
B23/numatrin	S-460	37	4.7	Up	0.104	0.028	0.241	0.034	2.3	0.016	11	13
NI	S-716	30	5.8	Up	0.081	0.014	0.139	0.021	1.7	0.016	15	12
NI	S-610	33	5	Up	0.050	0.009	0.075	0.007	1.5	0.018	14	13
NI	S-634	30	4.56	Down	0.106	0.017	0.062	0.008	0.6	0.019	14	12
NI	S-403	40	5.81	Up	0.019	0.007	0.045	0.008	2.4	0.020	8	10
NI	S-963	12	6.3	Down	0.313	0.044	0.198	0.031	0.6	0.022	15	13
NI	S-54	97	6	Up	0.036	0.006	0.081	0.017	2.3	0.022	12	10
NI	S-963	12	6.3	Down	0.313	0.044	0.198	0.031	0.6	0.022	15	13
NI	S-825	29	5.15	Up	0.052	0.012	0.108	0.019	2.1	0.023	12	12
NI	S-693	30.5	4.8	Down	0.202	0.013	0.159	0.010	0.8	0.023	15	13
Actin beta	S-348	45	5.16	Down	1.558	0.133	1.338	0.047	0.9	0.023	15	13
Actin gamma	S-351	45	5.2	Down	1.538	0.129	1.196	0.069	0.8	0.026	15	13
Hsp 60	S-150	60	5.35	Up	0.758	0.062	1.004	0.071	1.3	0.027	15	13
NI	S-307	45	4.89	Up	0.043	0.012	0.077	0.014	1.8	0.028	10	11
NI	S-133	60	4.9	Up	0.026	0.006	0.064	0.013	2.4	0.028	10	11
Hps70	S-94	70	5.2	Up	0.087	0.008	0.132	0.017	1.5	0.030	15	13
NI	S-557	33	5.86	Up	0.023	0.008	0.054	0.010	2.4	0.030	8	10
NI	S-400	38	4.7	Up	0.019	0.005	0.041	0.009	2.2	0.033	8	11
NI	S-365	45	6.29	Down	0.095	0.021	0.026	0.009	0.3	0.038	12	7
NI	S-1	200	4.92	Up	0.062	0.009	0.103	0.015	1.7	0.039	14	12
NI	S-517	34	5.86	Up	0.036	0.009	0.061	0.008	1.7	0.040	12	11
NI	S-689	23	4.78	Down	0.362	0.043	0.232	0.024	0.6	0.041	15	13

Table 3. continued

Protein identification	Spot number	M_r (kDa)	p*I*	Protein expression	Normal ($n = 15$) %Vol mean	(±) SD of the mean	Polyp ($n = 13$) %Vol mean	(±) SD of the mean	CP/CN	Wilcoxon $P \leq 0.05$	Detected in Normal	Polyp
NI	S-561	37	5.6	Down	0.058	0.013	0.017	0.010	0.3	0.042	12	3
NI	S-687	31	4.6	Down	0.093	0.014	0.054	0.010	0.6	0.043	15	11
NI	S-943	12.6	5.8	Down	0.098	0.016	0.046	0.016	0.5	0.044	15	6
NI	S-872	21.5	5.2	Up	0.093	0.012	0.120	0.011	1.3	0.046	14	13
NI	S-241	60	5.64	Down	0.584	0.090	0.325	0.056	0.6	0.048	15	12
B23/numatrin	S-459	37	4.69	Up	0.091	0.025	0.206	0.033	2.3	0.050	12	12
NI	S-752	29	> 6.5	Down	0.209	0.037	0.094	0.018	0.4	0.050	13	11

a) NI, nonidentified spots

b) >, p*I* for basic protein can not be determined accurately above pH 6.5

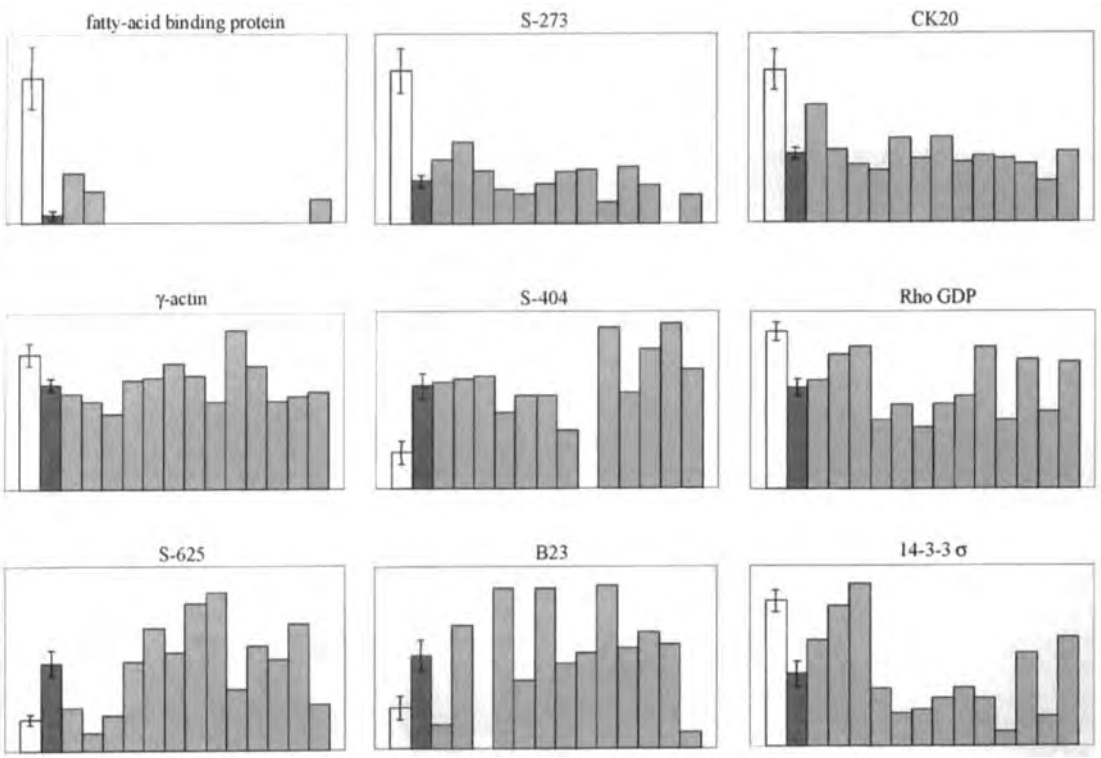


Figure 3. Heterogeneity of expression of selected proteins among different polyps. The light grey bars of each histogram indicate the level of designated proteins in each polyp specimen. White and the dark grey bars represent the mean of the normal and the mean of the polyps, respectively. Protein levels are expressed as relative volume on the *y*-axes. Unidentified protein spots are denoted by their database numbers.

Although the level of expression for individual genes varied among polyp specimens, it generally held true that the direction of variation (up or down) remained constant across the panel. We predict that proteins, which showed differential regulation across all or most of our panel, represent fundamental regulatory pathways in development of colon cancer, and that the proteins showing more sporadic variation reflect characteristics of individual polyps. Experiments using an expanded number of polyp tissues, as well as evaluations of pathologic reports concerning each sample, are in progress to confirm the results reported here.

Table 4. Colonic epithelial proteins identified from 2-D gel

Spot number	M_r (KDa)	pI	Protein description	Gene name	Method of identification
S-950	13	6.1	Liver fatty acid binding protein	L-FAEP	Ms[a] + Gm[b]
S-942	12	4.7	Cytochrome *c* oxidase va	COX5A	Mi[c] + Gm
S-94	70	5.2	Hps70	HSP70	Im[d] + Gm
S-703	30	4.8	Rho GDP-dissociation inhibition	RHO GDI 1	Ms + Gm
S-660	30	4.49	14-3-3 sigma	SFN-HME1	Ms + Gm
S-460	37	4.7	Numatrin/B23	NPM1	Ms + Im
S-459	37	4.69	Numatrin/B23	NPM1	Ms + Im
S-351	45	5.2	Actin gamma	ACTG	Ms + Gm
S-348	45	5.16	Actin beta	ACTB	Ms + Gm
S-294	45	5.48	Keratin cytoskeleton 20	KRT20	Ms + Gm
S-150	60	5.35	Hsp 60	HSP60	Mi + Gm
S-347	44	4.9	Keratin cytoskeleton 19	KRT19	Im + Gm
S-142	70	4.2	Calreticulin	CRTC	Ms + Gm
S-198	60	5.9	Probable disulfide isomerase	ER60	Mi + Gm
S-240	60	5	alpha Tubulin	TBA 1	Im + Gm
S-322	45	5.38	Keratin cytoskeleton 18	KRT18	Im + Gm
S-242	63	4.92	ATP synthetase-beta subunit	ATPB	Mi + Gm
S-958	12	4.8	Thioredoxin	TRX	Mi + Gm
S-310	46	5.06	Heterogeneous nuclear riboprotein F	NHRNP F	Ms + Gm
S-393	34	4.6	Laminin-receptor colon carcinoma L-binding protein	LAMR	Ms + Gm
S-85	80	4.9	78 kDa Glucose-regulated protein precursor	GRP 78	Ms + Gm
S-826	22	5.62	Natural killer cell enhancing factor	NKEF-B	Ms + Gm
S-808	23	6.1	Antioxidant protein 1 AOP-1 (MER 5 homolog)	AOP-1	Ms + Gm

a) Ms, mass spectrometry
b) Gm, gel matching
c) Mi, microsequencing
d) Im, immunoblotting

Among the identified proteins that showed altered expression in polyps, L-FABP, a member of a large multigene family [29], was the most strongly down-regulated protein. Fatty acids are the preferred energy source for normal colonic epithelial cells; L-FABP is known to be involved in their differentiation [30] and in migration of epithelial cells along the crypts [31, 32]. Lowered levels of L-FABP in colorectal cancers and colonic carcinomas have been reported in other studies [33, 34]. Another protein with decreased levels of expression in polyps was 14-3-3 σ, a member of the 14-3-3 family of proteins (the σ isoform is expressed by epithelial cells). This isoform, which is regulated by p53 in colonic cell lines [35], has been shown to harbor a pleckstrin homology (PH) domain and to enhance protein kinase C activity [36]. Decreased expression of this protein has also been observed in transformed keratinocytes [37] and in less differentiated bladder cell carcinomas [38]. Cytochrome *c* oxidase polypeptide Va, one the 13 subunits of cytochrome *c* oxidase, is reportedly decreased in colonic adenocarcinomas [39, 40]. GDI inhibits the GDP/GTP exchange among small GTP-binding proteins (G proteins) of the rho family, keeping them in the GDP-bound, inactive form [41]. Disruption of rho GDI in yeast does not induce apparent changes in phenotype, but overexpression results in growth inhibition [42]. Rho proteins are required for endogenous and EGF-induced migration of epithelial cells during the early phase of mucosal restitution [43]. Cytokeratin 20 is a marker for differentiation of colonic crypt cells [44, 45].

On the other hand, we found numatrin, HSP70, and HSP60 to be overexpressed in colonic polyps. Numatrin is a phosphoprotein involved in ribosome assembly [46] and it has some oncogenic properties [47]; overexpression of numatrin has been reported in highly proliferating cells [48], in transformed keratinocytes [37], and in mRNA from neoplastic colorectal mucosa [49]. HSP70 and HSP60 are often overexpressed in human colonic diseases and in colonic tissue cell lines [50–52]. Whether the proteins we have characterized here by differences in expression levels identify the primary culprits in cancer progression, or are merely consequential adjustments, remains to be determined. We predict that the majority of changes would fall in the latter category. However, when

the corresponding regulatory pathways have been elucidated, it should become possible to determine which altered protein patterns are specific to the tumorigenic process.

We are very grateful to Dr. Mark Olson for providing numatrin antibodies, to Dr. W. Zhuchun for performing the mass spectrometric analyses, to Robert Schackmann for performing the microsequencing, and to Alexander Tsodikov for assistance with statistical analysis. This work was supported by Huntsman Cancer Institute, Department of the Army grant No. DAMD17-94-J-4129 and by NCI grant No. 5P30 CA42014.

Received November 11, 1998

5 References

[1] Kewenter, J., *Lakartidningen* 1998, *95*, 2950–2952.

[2] Gerdes, H., Gitlin, J. S., Zimbalist, E., Urmacher, C., Lipkin, M., Winawer, S. J., *Cancer Res.* 1993, *53*, 279–282.

[3] Pretlow, T. P., O'Riordan, M. A., Pretlow, T. G., Stellato, T. A., *J. Cell Biochem. Suppl.* 1992, *166*, 55–62.

[4] Ilantzis, C., Stanners, C. P., *In Vitro Cell Dev. Biol. Anim.* 1997, *33*, 50–61.

[5] Bjerknes, M., Cheng, H., Hay, K., Gallinger, S., *Am. J. Pathol.* 1997, *150*, 833–839.

[6] Moss, S. F., Liu, T. C., Petrotos, A., Hsu, T. M., Gold, L. I., Holt, P. R., *Gastroenterology* 1996, *111*, 1425–1432.

[7] Moss, S. F., Scholes, J. V., Holt, P. R., *Dig. Dis. Sci.* 1996, *41*, 2238–2247.

[8] Roncucci, L., Modica, S., Pedroni, M., Tamassia, M. G., Ghidoni, M., Losi, L., Fante, R., Di Gregorio, C., Manenti, A., Gafa, L., Ponz de Leon, M., *Br. J. Cancer* 1998, *77*, 2343–2348.

[9] Izzo, R. S., Pellecchia, C., *Scand. J. Gastroenterol.* 1998, *33*, 191–194.

[10] Ji, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D., Simpson, R. J., *Electrophoresis* 1994, *15*, 391–405.

[11] Ji, H., Reid, G. E., Moritz, R. L., Eddes, J. S., Burgess, A. W., Simpson, R. J., *Electrophoresis* 1997, *18*, 605–613.

[12] Reymond, M. A., Sanchez, J.-C., Hughes, G. J., Gunther, K., Riese, J., Tortola, S., Peinado, M. A., Kirchner, T., Hohenberger, W., Hochstrasser, D. F., Kockerling, F., *Electrophoresis* 1997, *18*, 2842–2848.

[13] Szymczyk, P., Krajewska, W. M., Jakubik, J., Berner, A., Janczukowicz, J., Mikulska, U., Berner, J., Kilianska, Z. M., *Tumori* 1996, *82*, 376–381.

[14] Reid, G. E., Ji, H., Eddes, J. S., Moritz, R. L., Simpson, R. J., *Electrophoresis* 1995, *16*, 1120–1130.

[15] Hayward, I. P., Whitehead, R. H., Ward, L., Gianello, R., Dempsey, P., Bates, R., Burns, G. F., *Immunol. Cell Biol.* 1995, *73*, 249–257.

[16] Keesee, S. K., Meneghini, M. D., Szaro, R. P., Wu, Y. J., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 1913–1916.

[17] Whitehead, R. H., Brown, A., Bhathal, P. S., *In Vitro Cell Dev. Biol.* 1987, *23*, 436–442.

[18] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[19] Görg, A., *Biochem. Soc. Trans.* 1993, *21*, 130–132.

[20] Sanchez, J.-C., Rouge, V., Ravier, F., Tonella, L., Mossmayer, M., Wilkins, R. M., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[21] Celis, E. J., *Cell Biology, A Laboratory Handbook*, Academic Press Inc. London 1994, pp. 281–287.

[22] Vari, F., Bell, K., *Electrophoresis* 1996, *17*, 20–25.

[23] O'Connell, K. L., Stults, J. T., *Electrophoresis* 1997, *18*, 349–359.

[24] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[25] Vorm, O., Roepstorff, P., *Biol. Mass Spectrom.* 1994, *23*, 734–740.

[26] Zhuchun, W. (Sample Preparation for MALDI Mass Spectrometry using Pipette Tip Columns), paper presented at 46th Conference in Mass Spectrometry and Allied Topics, Orlando, Florida, 1998.

[27] Frisch, S. M., Francis, H., *J. Cell Biol.* 1994, *124*, 619–626.

[28] Strater, J., Wedding, U., Barth, T. F., Koretz, K., Elsing, C., Moller, P., *Gastroenterology* 1996, *110*, 1776–1784.

[29] Gordon, J. I., Lowe, J. B., *Chem. Phys. Lipids* 1985, *38*, 137–158.

[30] Joshi, S. S., Jackson, J. D., Sharp, J. G., *Cancer Detect. Prev.* 1985, *8*, 237–245.

[31] Wilson, A. J., Gibson, P. R., *Gastroenterology* 1997, *113*, 487–496.

[32] Huang, Y. C., Jessup, J. M., Forse, R. A., Flickner, S., Pleskow, D., Anastopoulos, H. T., Ritter, V., Blackburn, G. L., *Lipids* 1996, *31*, S313–S317.

[33] Davidson, N. O., Ifkovits, C. A., Skarosi, S. F., Hausman, A. M., Llor, X., Sitrin, M. D., Montag, A., Brasitus, T. A., *Lab. Invest.* 1993, *68*, 663–675.

[34] Sakai, Y., *Nippon Shokakibyo Gakkai Zasshi* 1990, *87*, 2594–2604.

[35] Hermeking, H., Lengauer, C., Polyak, K., He, T. C., Zhang, L., Thiagalingam, S., Kinzler, K. W., Vogelstein, B., *Mol. Cell* 1997, *1*, 3–11.

[36] Dellambra, E., Patrone, M., Sparatore, B., Negri, A., Ceciliani, F., Bondanza, S., Molina, F., Cancedda, F. D., De Luca, M., *J. Cell Sci.* 1995, *108*, 3569–3579.

[37] Celis, J. E., Olson, E., *Electrophoresis* 1994, *15*, 309–344.

[38] Ostergaard, M., Rasmussen, H. H., Nielsen, H. V., Vorum, H., Orntoft, T. F., Wolf, H., Celis, J. E., *Cancer Res.* 1997, *57*, 4111–4117.

[39] Barnard, J. A., Warwick, G., *Cell Growth Differ.* 1993, *4*, 495–501.

[40] Heerdt, B. G., Augenlicht, L. H., *J. Biol. Chem.* 1991, *266*, 19120–19126.

[41] Takai, Y., Kaibuchi, K., Sasaki, T., Tanaka, K., Shirataki, H., Nakanishi, H., *Princess Takamatsu Symp.* 1994, *24*, 338–350.

[42] Masuda, T., Tanaka, K., Nonaka, H., Yamochi, W., Maeda, A., Takai, Y., *J. Biol. Chem.* 1994, *269*, 19713–19718.

[43] Santos, M. F., McCormack, S. A., Guo, Z., Okolicany, J., Zheng, Y., Johnson, L. R., Tigyi, G., *J. Clin. Invest.* 1997, *100*, 216–225.

[44] Moll, R., Zimbelmann, R., Goldschmidt, M. D., Keith, M., Laufer, J., Kasper, M., Koch, P. J., Franke, W. W., *Differentiation* 1993, *53*, 75–93.

[45] Calnek, D., Quaroni, A., *Biochem. J.* 1992, *285*, 939–946.

[46] Yung, B. Y., Busch, H., Chan, P. K., *Biochim. Biophys. Acta* 1985, *826*, 167–173.

[47] Kondo, T., Minamino, N., Nagamura-Inoue, T., Matsumoto, M., Taniguchi, T., Tanaka, N., *Oncogene* 1997, *15*, 1275–1281.

[48] Feuerstein, N., Chan, P. K., Mond, J. J., *J. Biol. Chem.* 1988, *263*, 10608–10612.

[49] Nozawa, Y., Van Belzen, N., Van der Made, A. C., Dinjens, W. N., Bosman, F. T., *J. Pathol.* 1996, *178*, 48–52.

[50] Stulik, J., Bures, J., Jandik, P., Langr, F., Kovarova, H., Macela, A., *Electrophoresis* 1997, *18*, 625–628.

[51] Botzler, C., Schmidt, J., Luz, A., Jennen, L., Issels, R., Multhoff, G., *Int. J. Cancer* 1998, *77*, 942–948.

[52] Lu, X., Seligy, V. L., *Biochem. Biophys. Res. Commun.* 1992, *186*, 371–377.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1065–1074

Sarada C. Prasad
Viatcheslav A. Soldatenkov
Michael R. Kuettel
Peter J. Thraves
Xiaojun Zou
Anatoly Dritschilo

Department of Radiation Medicine, Division of Radiation Research, Vincent T. Lombardi Cancer Center, Georgetown University Medical Center, Washington DC, USA

Protein changes associated with ionizing radiation-induced apoptosis in human prostate epithelial tumor cells

Ionizing radiation (IR) is an important component in the therapy of localized prostate cancer. Identification of protein alterations during IR-induced apoptosis prostate cancer cells is an important step toward understanding the new metabolic status of the dying cell. In the present study, we report changes in protein profile that define the execution phase of the apoptotic response in the *in vitro* model of tumorigenic radiation-transformed SV40-immortalized human prostate epithelial cells (267B1-XR), induced to undergo programmed cell death by IR. We employed an approach that involves use of analytical two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) coupled with Western blotting with specific antisera. Our results point out that apoptotic cells experience significant reduction in the levels of the intermediate filament proteins, keratins-18, 19, vimentin and the associated 14-3-3 adapter proteins. At the same time, molecular chaperones such as glucose-regulated protein 94, calreticulin, calnexin, and protein disulfide isomerase exhibit marked accumulation in these dying cells. The present data indicate that apoptosis-associated processes in prostate epithelial cells include solubilization of the rigid intermediate filament network by specific proteolysis as well as increased levels of endoplasmic reticulum (ER) proteins with chaperone functions.

Keywords: Apoptosis of prostate cells / Ionizing radiation / Endoplasmic reticulum chaperones / Intermediate filament proteins / Two-dimensional polyacrylamide gel electrophoresis EL 3368

1 Introduction

Ionizing radiation (IR) is a major component in the treatment of localized prostatic carcinoma [1, 2] in addition to anti-androgen and cytotoxic therapies. The phenomenon of cell death by apoptosis is an active process, triggered by precise signals that induce crucial biochemical changes in target cells, including alterations in the pattern of protein expression and stability. These particular changes are expected to vary from one cell type to another, due to the fact that different stimuli have the potential to induce cell death by activating independent or overlapping metabolic pathways in the same cell [3]. Transcription factors, nonnuclear proteins such as members of the Bcl-family, and receptors mediating death signals have been held responsible for determining cell fate (reviewed in [4, 5]). Involvement of ICE-like proteases [6, 7] and that of other classes of proteases (TPCK- and benzyloxy carbonyl-sensitive) [8] in mediating limited proteolysis has also been recognized [8]. During the execution phase of apoptosis, limited proteolysis of actin-binding proteins (spectrin, fodrin, gelsolin, and focal adhesion kinase) and intermediate filaments (nuclear lamins A, B, and C, keratins 18, 19, and vimentin) appears to bring about the characteristic collapse of nuclear and cell shape and render the rigid framework of the epithelial cells soluble [9, 10]. The many other gene products participating in the cell death program include those maintaining cell structure, second messenger transduction cascades, cell cycle progression, and housekeeping endoplasmic reticulum (ER) and/or mitochondrial components (reviewed in [11, 12]). In particular, depletion of calcium (Ca) pools from the ER and elevation of intracellular Ca have been linked to induction of apoptosis mediated by various stimuli. As a consequence of the Ca imbalance, levels of resident ER proteins (molecular chaperones) have been found to be higher after treatment with diverse agents employed to alter normal ER functioning [13–16]. Instances of molecular chaperones influencing the organization and function of microtubules, microfilaments and intermediate filaments as well as protein breakdown have also been pointed out [19, 20]. These changes in proteins participating in multiple

Correspondence: Sarada Prasad, Ph.D., Department of Radiation Medicine, Division of Radiation Research, TRB-E204A, Georgetown University Medical Center, 3970 Reservoir Road NW, Washington, DC 20007-2197, USA
E-mail: prasads@gunet.georgetown.edu
Fax: +202-687-2221

Abbreviations: **267B1-XR**, radiation-transformed SV40-immortalized human prostate epithelial cells; **DAPI**, 4′,6′-diamidino-2-phenylindole; **ER**, endoplasmic reticulum; **grp**, glucose-regulated proteins; **hsp**, heat shock proteins; **IFAPs**, intermediate filament-associated proteins; **IFs**, intermediate filaments; **IR**, ionizing radiation; **PARP**, poly(ADP-ribose)polymerase; **PDI**, protein disulfide isomerase

biochemical pathways together define the new metabolic status of the dying cell.

Earlier studies from our laboratory, using the SV40-immortalized neonatal human prostate epithelial cells (267B1), demonstrated protein expression changes during ionizing radiation (IR)-induced neoplastic transformation [23–26]. These isogenic cells are unique in that their initial protein expression patterns, as well as changes associated with neoplastic transformation, are known. Consequently, further analysis of apoptosis-associated protein changes in these cells would corroborate our knowledge of carcinogenesis and apoptosis. We have recently established criteria of apoptosis using MCF-7 (estrogen receptor-positive), MDA-468 (estrogen receptor-negative) breast tumor cells and radiation-transformed 267B1 (267B1-XR) cells and pointed out the involvement of protein ubiquination, proteolysis of type I keratins, and vimentin, respectively [21, 22]. In the present study, we compare protein expression in the adherent (control) and nonadherent (apoptotic) 267B1-XR cells employing analytical 2-D PAGE combined with specific antibody blotting. We now report that intermediate filament proteins (keratins 18, 19 and vimentin) are underrepresented in the nonadherent cells representing cells undergoing apoptosis, while the ER resident proteins (calreticulin, calnexin, glucose-regulated protein94 – grp94 –, and protein disulfide isomerase (PDI) exhibit higher levels of accumulation. These results on the differential protein expression during apoptosis of 267B1-XR cells point to protein breakdown, protein synthesis, as well as post-translational modification, as necessary components of the apoptotic program. The evidence thus supports the premise that the cell death program is directly mediated by dysregulation of the expression/function of certain gene products.

2 Materials and methods

2.1 Cell culture and irradiation

IR-transformed 267B1-XR cells [23] were grown in P48F-medium supplemented with 2% heat-inactivated fetal bovine serum, 100 U/mL penicillin, with 100 μg/mL streptomycin and 2 mM L-glutamine. The gamma-ray source (^{137}Cs) was fixed at a dose of 2.37 Gy/min, and cells metabolically labeled or unlabeled were exposed to 6 Gy at room temperature, in air. Nonadherent cells were collected by media centrifugation at indicated time points subsequent to radiation treatments, while adherent cells and control nonirradiated cells were harvested by trypsinization. Cells were stored (at –70°C) as pellets of 2×10^6 cells for analysis. The proportion of adherent and nonadherent cells were counted at 24, 48 and 72 h after exposure to 6 Gy to determine the extent of cell death.

2.2 Metabolic labeling of cells in culture

Logarithmically growing 267B1-XR cells were preincubated in methionine-free culture medium for a period of 3–4 h and subsequently labeled with 50 μCi/mL of [^{35}S] methionine for 16 h in methionine-free medium [24]. Media was aspirated, cells were washed and replaced with fresh methionine-containing medium, and immediately subjected to IR exposure as indicated. Twenty-four hours after exposure to IR, the adherent and nonadherent cells were pelleted and stored frozen for 2-D gel analysis.

2.3 Analysis of apoptosis

For quantitation of apoptosis, adherent and nonadherent cells were pooled together and washed with PBS prior to fixation with 3.7% paraformaldehyde for 3–5 min. After rehydration in PBS, cells were stained with 4′,6′-diamidino-2-phenylindole (DAPI) (0.5 μg/mL, Sigma, St. Louis, MO) for 30 min in the dark and mounted with cover slips. Epifluorescence microscopy was performed using a Zeiss Photoscope II (Carl Zeiss, Jena, Germany) equipped with a Zeiss plan 4-X/0.65 NA objective and cells were photographed using a 35 mm camera [21]. Apoptotic cells exhibiting condensed and fragmented nuclei were countered in five to seven randomly selected fields. A minimum of 400–500 nuclei were examined in each case, and results were expressed as the number of apoptotic nuclei over the total number counted.

2.4 2-D PAGE

Control, adherent and nonadherent cell samples were analyzed by 2-D PAGE using the Iso-Dalt system (Hoefer Scientific, San Francisco, CA) [27]. The method involved minor modifications of O'Farrell's original technique to suit the Iso-Dalt equipment. IEF gel tubes were made up of 3.5% acrylamide, 9 M urea, 2% Ampholines (blended pH 3–10 and 4–8 in the ratio of 3:1), 2% NP-40, 0.03% ammonium persulfate, and 0.01% TEMED. Briefly, cell pellets (2×10^6) were suspended in 250 μL of 2-D PAGE sample buffer (9 M urea, 4% NP-40, 2% Ampholines 3–10 and 1% DTT) and 20 μL portions (either radioactive or unlabeled) were resolved by IEF for 20 000 Vh. After the first-dimensional separation, tube gels were layered on top of 10% polyacrylamide slab gels for SDS-PAGE resolution in the second dimension (100 V, overnight, 20°C). Slab gels were either stained with Commassie blue or transferred to Immobilon membranes for subsequent autoradiography and immunoblotting. Gel autoradiographs were scanned (Nikon f2.8 lens), and digitized by a combination of a 3/10 computer (Sun Microsystem, Mountain View, CA) interfaced to a scanner and digitizer (PS200 power supply, Photometrics, Tucson, AZ). Data analysis by computer-assisted comparison of the spots of interest

in the adherent and nonadherent cells was accomplished by using ELSIE-5 software, an updated version of ELSIE-4 previously described [24, 26].

2.5 Western immunoblotting

Samples of 50 μg protein from each of the total cell lysates, collected at various time points, were resolved by SDS-PAGE on 10% polyacrylamide gels and transblotted to Immobilon membranes (Millipore Corporation, Bedford, MA) as previously described [23–26]. Transfer buffer contained 25 mM Tris-HCl, 192 mM glycine, 0.037% w/v SDS and 10% methanol. Limited cleavage of the various protein substrates were analyzed by Western blotting with antisera to anti-poly(ADP-ribose)polymerase (1:2000; Boehringer Mannheim, Indianapolis, IN), topoisomerase-I (1:2000; Topogen, Columbus, OH), heat shock protein (hsp) 90 (1:1000; Transduction Laboratories, Lexington, KY), and hsp27 (2.0 μg/mL; Oncogene Research Products, Cambridge, MA). Proteins that appear new to the apoptotic cells were screened to test if they correspond to calreticulin (1:2500; gift from Dr. Ramsamooj), anti-proteasome subunits α and β (1:1000; Calbiochem, La Jolla, CA), anti 14-3-3 proteins (1.0 μg/mL), anti-hsp70 (1:500), anti-grp78 (1:500), anti-grp75 (1:500), anti-grp94 (1:500; Santa Cruz Biotechnology, Santa Cruz, CA), calnexin (1:1000), anti-hsp90 (1:1000) (Transduction Labs) anti-PDI (1:500; Stressgen, British Columbia, Canada), anti-keratin 8, 18 (cl. 8.13, 1:500), and anti-vimentin (1:1000; Sigma). Immunoblotting methods have been previously described [21–26]. Briefly, membranes were blocked with 5.0% bovine serum albumin in TBST (Tris-buffered saline with 0.1% Tween-20). After washing three times with TBST, membranes were incubated with primary antibodies for 3 h at room temperature. After three washes, membranes were incubated with appropriate anti-mouse/anti-rabbit/anti-goat/anti-human Ig-coupled alkaline phosphatase reagent (1:5000) for 1 h, and the immune reaction was detected by NBT/BCIP reagents from Promega (Madison, WI).

3 Results

3.1 Demonstration of IR-induced apoptosis in 267B1-XR cells

Treatment of 267B1-XR cells with X-rays resulted in detachment of cells from the monolayer, increasing with time after irradiation. At 24 h, the proportion of detached cells showing criteria of apoptosis was 8%. At 48 and 72 h after exposure to 6 Gy of IR, apoptotic cells were observed at frequencies as high as 18 and 28%. We scored for the presence of apoptotic nuclei in DAPI-stained cells exemplified by Figs. 1A and 1B, representing untreated control cells and apoptotic cells at 48 h post irradiation, respectively. Typical nuclear abnormalities, such as condensed chromatin, were noted in apoptotic cells (Fig. 1B).

3.2 Overview of the differentially expressed proteins in apoptotic cells

The effect on protein expression in apoptotic cells was investigated next. We performed 2-D PAGE of control and apoptotic 267B1-XR cells and compared their protein expression profiles upon colloidal Coomassie staining (Fig. 2). The overall protein profiles in control and apoptotic cells appear comparable; however, distinct differences in specific proteins characterize the metabolic states of control and apoptotic cells. These various spot differences can be grouped into two major classes: (i) proteins present in control cells and almost undetectable in nonadherent cells (series of spots marked in Fig. 2A; V, K-18 and K-19), and (ii) proteins with higher levels of accumulation in the nonadherent cells in comparison to control cells (spots marked 1–6 in Fig. 2B). Due to the complexity and the diversity of proteins exhibiting expression changes, we have taken a systematic approach to describe protein differences in a meaningful order. Primarily, our approach involved identification of the two classes of proteins marked in Fig. 2. We took advantage of the M_r/p*I* criteria of differentially expressed proteins together with

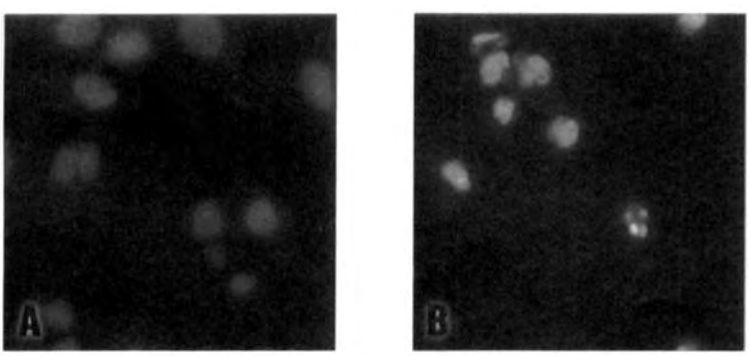


Figure 1. Demonstration of apoptosis: Fluorescence microscopy of DAPI-stained 267B1-XR cells (a) before and (b) 48 h after irradiation with 6 Gy. Control cells display uniformly stained nuclei, and approximately 20% of irradiated cells exhibit chromatin condensation and blebbing of the cell nuclei.

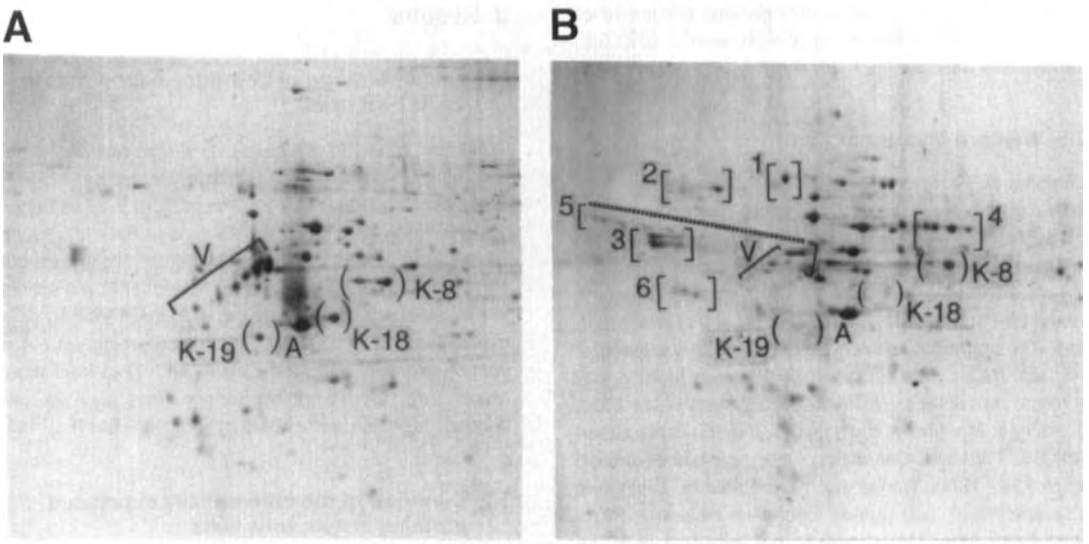


Figure 2. 2-D PAGE of control and nonadherent 267B1-XR cells: Coomassie blue-stained 2-D gels containing whole cell lysates of control/nonirradiated 267B1-XR cells or nonadherent cells (A, B) collected 48 h after treatment with 6 Gy of IR. Electrophoresis of cell lysates (200 μg protein) was performed using the Iso-Dalt equipment and methods. Proteins primarily present in control cells (A: V, vimentin; K-18, keratin 18; and K-19, keratin 19), and proteins missing as well as those showing significant increases in apoptotic cells (B: protein spots marked 1–6) are indicated. Actin (M_r = 43 kDa, p*I* = 5.3) is denoted by A. Protein identity was confirmed by immunoblot analysis with specific antisera.

clues provided by the 2-D databases [28] and assigned tentative identifications (Table 1). These assignments were confirmed by Western blotting of portions of the 2-D gels with specific antibodies in each case. Once the nature of the specific proteins was established, we reasoned to provide a plausible interpretation to associate the changes in particular proteins with the biochemical and morphological phenomena occurring during apoptosis. As can be noticed in the text that follows, differentially expressed proteins represented either qualitative differences (presence or absence), or quantitative differences (changes in amount). The latter were evaluated by densitometric analysis of the specific protein spots.

3.3 Proteins exhibiting significant decreases in apoptotic cells

A detailed look at the proteins of control cells (Fig. 2A) specifically points to proteins marked V (p*I* 4.9, M_r 54 kDa), K-18 (p*I* 5.4, M_r 48 kDa), and K-19 (p*I* 4.9, M_r 39 kDa), present in significant amounts, while they are almost undetectable in the nonadherent cells. At the same time, protein marked K-8 (p*I* 5.5, M_r 52 kDa) shows only a marginal difference in its intensity in control and apoptotic cells. These abundant proteins with their p*I*/M_r values and 2-D gel locations correspond to vimentin and keratins 8, 18, 19, respectively. The isoform variation of these intermediate filament (IF) proteins is due to phosphorylations at serine residues [29]. By virtue of the epithelial nature of 267B1-XR cells, they express keratins 8, 18 and 19 [26]. Keratins 18 and 19 have recently been shown to be susceptible to proteolytic cleavage by members of the cysteine protease family, particularly by caspase-6 [11, 12, 22]. These two type-I keratin family members exhibit caspase cleavage sites at VEVD representing P4–P1 motifs (at residue 258) of the protease recognition sites. Moreover, these common sequence motifs for recognition by caspases are shared by other members of the intermediate filament family of proteins, including several other type-I keratins and vimentin as well. Our recent reports documented that keratins 18, 19, and vimentin are candidate substrates of caspase-mediated proteolysis in the apoptotic 267B1-XR cells.

The ubiquitous family of 14-3-3 proteins that interact with keratin IF polypeptides are expressed in epithelial cells [11, 29–31]. We evaluated the status of 14-3-3 proteins in cells that lost attachment to the substratum while exhibiting criteria of apoptosis. To increase the sensitivity of detection of these less abundant proteins, we metabolically labeled the 267B1-XR cells prior to exposure to IR. Subsequently, we analyzed control and nonadherent cells col-

Table 1. Criteria of proteins exhibiting apoptosis-associated differential expression

Protein identity	M_r/pI	Cellular location-role/ apoptotic response
	Decreased in apoptotic cells	
Keratin 18[a)]	48/5.4	IF proteins/caspase substrate
Keratin 19[a)]	40/4.8	IF proteins/caspase substrate
Vimentin[a)]	65/5.0	IF proteins/caspase substrate
PARP[c)]	113/8.8	DNA repair/caspase substrate
Topo-I[c)]	101/7.5	DNA repair/caspase substrate
Tm-1[b, d)]	36/4.4	Actin stress fibers/disassembly
Tm-3[b, d)]	34/4.5	Actin stress fibers/disassembly
14-3-3[b, d)]	30/4.1–4.6	Adapter; isoform specific/proteolysis
	Increased in apoptotic cells	
grp94 (1)[a, b, d)]	94/4.8	ER chaperones/stress response
Calnexin (2)[a, b, d)]	100.0/4.4	ER chaperones/stress response
Calreticulin (3)[a, b, d)]	63/4.1	ER chaperones/stress response
PDI (5)[a, b, d)]	56/4.6	ER chaperones/stress response
	Not altered in apoptotic cells	
grp78[b)]	74/4.9	ER-chaperones/stress response
grp75[b)]	68/5.5	Mit-chaperones/stress response
hsp90[b, c)]	90/5.0	Cyt-chaperone/stress response
hsp70[b)]	66/5.1–5.3	Mit-chaperone/stress response
hsp27[c)]	27–30/4.6–5.5	Cyt-chaperone/stress response
Proteasome α[b)]	29/4.8	Prosomes-proteolysis/unknown
Proteasome β[b)]	36/5.4	Prosomes-proteolysis/unknown

Protein expression levels in control and apoptotic cells based on (indicated as superscript):

a) Coomassie blue stain coupled with Western analysis

b) Autoradiography of metabolically labeled cells followed by Western analysis

c) 1-D gel blots

d) Densitometric quantitation of the protein intensities in autoradiographs of control and apoptotic 267B1-XR cells.

For the proteins showing increase in apoptotic cells, the numbers (1–5 in parenthesis indicate the designation given to the corresponding protein spots in Figs. 2 and 4. Identities of each of the proteins were confirmed by Western analysis.

Abbreviations: IF, intermediate filaments; PARP, poly(ADP-ribose)polymerase; PDI, protein disulfide isomerase; grp, glucose-regulated protein; hsp, heat shock protein; topo-I, topoisomerase; Tm, tropomyosin isoforms; ER, endoplasmic reticulum; mit, mitochondrial; cyt, cytoplasmic

lected 24 h after IR exposure by 2-D PAGE. As can be seen in autoradiographs, proteins marked by double arrows in control cells (Fig. 3A) show an apoptosis-associated decrease in intensity (Fig. 3B). Figures 3C and 3D show Western blot analysis of specific areas of these 2-D gels; immunoreactive 14-3-3 proteins correspond to the spots indicated in the autoradiographs. Further, certain additional low molecular weight isoforms of 14-3-3 proteins are noticed in the apoptotic cells (marked in parenthesis; compare Fig. 3C and D). Recent studies indicate that IF proteins are also associated with several intermediate filament-associated proteins (IFAPs) [32], multicatalytic protease complexes called proteasomes [33] and heat shock proteins [34]. It has been suggested that the IF network serves the function of a docking matrix whre cellular processes can occur [32]. When portions of these 2-D gel blots were probed with antibodies to α- and β-subunits of the proteasome (Pr-α and Pr-β; shown in Fig. 3A), no detectable changes in the levels of these proteasome subunits were noted in apoptotic cells. At the same time, Tm-1 and Tm-3 members of the family of actin filament-binding proteins (Fig. 3A and 3B), show a significant decrease (85–90%) in the dying 267B1-XR cells, demonstrating deterioration of microfilament architecture.

Note that switching from Coomassie blue staining to metabolic labeling for detection of differential protein changes in the apoptotic cells allowed new insights. In particular, not all proteins that were detected by Coomassie blue staining were visible with equal intensity in the metabolically labeled protein profiles. We interpret this to mean that certain proteins were not translated *de novo* in signifi-

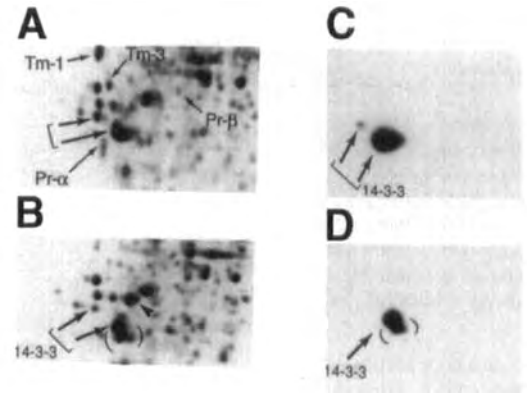


Figure 3. Status of 14-3-3 proteins in control and apoptotic 267B1-XR cells: Protein lysates (100 μg) spiked with 5×10^5 cpm of [^{35}S]methionine from control and nonadherent cells were subjected to 2-D PAGE, transferred to PVDF membrane, and exposed to X-ray film for 5–7 days at –70°C. Immunoblots were developed for the corresponding 2-D gel areas as described in Section 2.5. Positive spots on immunoblots were matched to autoradiograms. (A) and (B): autoradiographs of portions of 2-D gels; (C) and (D): Western blots of 2-D gel areas of control and nonadherent samples, respectively, with antisera to 14-3-3 proteins. Tm-1 and Tm-3, isoforms of tropomyosin; Pr-α and Pr-β, α- and β-subunits of proteasome, respectively. The identity of a protein spot marked by arrowhead in (B) is not known.

cant quantities during the period of 16 h pulse, and any possible chain elongations in this set of proteins were marginal. Therefore, certain proteins well detected by protein stain and not as well by metabolic labeling were due to accumulations of the endogenous amounts in the cell prior to metabolic labeling.

3.4 Proteins elevated in the apoptotic 267B1-XR cells

The set of proteins marked 1–6 represent candidates almost undetectable in control cells while showing higher levels in the nonadherent cells (Fig. 2B). Their p*I* and M_r criteria are shown in Table 1. Based on the 2-D gel database identifications, protein spots marked 1, 2 and 3 were tentatively identified as grp94 , calnexin and calreticulin, respectively. When metabolically labeled cells were analyzed to compare the levels of expression of these proteins (Fig. 4A and B), results obtained (see Fig. 2B) by Coomassie staining are emphasized. In particular, proteins marked 1, 2 and 3 show approximately 2-fold increases, while polypeptides 4, 5, 6 exhibit extensive isoform variation. These isoforms are better seen in the Coomassie blue-stained gels (Fig. 2B) than in the autoradiographs (Fig. 4B). We performed immunoblots of portions of these 2-D gels covering the areas of spots 1, 2, 3 and 5 with specific antibodies to grp94, calnexin, calreticulin, and PDI (Fig. 4C–F). These four proteins have a unique feature in common in that they represent members of the ER compartment. Furthermore, while proteins 4, 5, and 6 represent targets of extensive post-translational modification in the apoptotic cells, the nature of their modification is not known. We interpret this to mean that they are primarily glycosylated in addition to other possible modifications such as phosphorylation. Our interpretation is based on the fact that they are members of the ER compartment and the isoforms are rendered more acidic. Moreover, these extensively modified isoforms are not indicated at detectable levels in the autoradiographs showing [^{35}S]methionine incorporation into intact protein. The identities of protein spots marked 4 and 6 in Fig. 2B are not known.

The expression of the ER chaperones (grp78, grp94) is coordinatedly regulated, while that of mitochondrial chaperones (grp75, hsp70) is also known to be responsive to stress [36–38]. To clarify the status of these additional chaperones in the apoptotic 267B1-CR cells, we performed Western blotting with specific antibodies to grp78, grp75, and hsp70 (proteins marked 7, 8, and 9 in Fig. 4A correspond to immunoblots in Fig. 4G–I, respectively), while simultaneously comparing the autoradiographs of the control and apoptotic cells for the intensities of the corresponding spots. Our data revealed that relative levels of these chaperones were comparable in control and apoptotic cells.

3.5 Specificity of protein changes in apoptotic cells

While we report apoptosis-associated cleavage of IF proteins and increased accumulation of chaperones of the ER compartment, it is relevant to demonstrate the occurrence of protein changes that represent part, of the well-established phenomena of cell death. We therefore assessed ICE-mediated limited cleavage of enzymes of the DNA-repair pathway, namely poly(ADP-ribose) polymerase (PARP) and topoisomerase-I [6–7] in apoptotic 267B1-XR cells. Figures 5A and B show the characteristic generation of the 89 kDa and 72 kDa C-terminal fragments (lane 3 in each case) of the 113 kDa and 101 kDa native enzymes (lanes 1 and 2 in each case), respectively, as a result of limited cleavage by caspases. At the same time, the status of additional members of the heat shock family of proteins, namely hsp90 and hsp27, with chaperone functions remained unchanged (Fig. 5C and D).

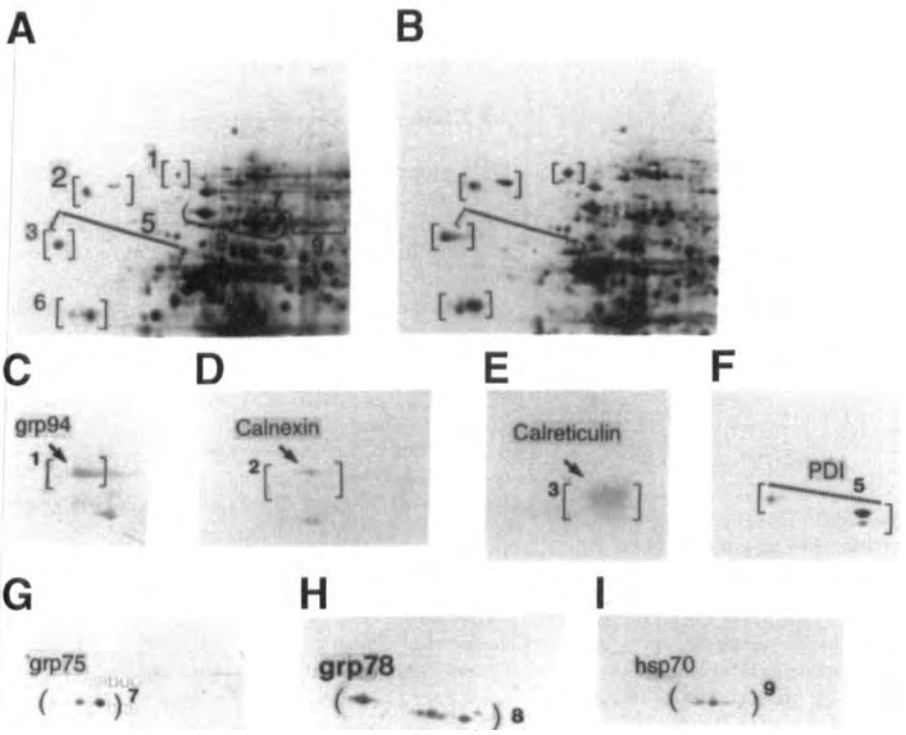


Figure 4. Differential protein expression during apoptosis: (A, B) Autoradiographs of portions of 2-D gels containing total cell lysates of control and apoptotic 267B1-XR cells, respectively. Immunoblots in (C)–(I) demonstrate the identities of proteins grp94 (protein 1); (D) calnexin (protein 2); (E) calreticulin (protein 3); (F) protein disulfide isomerase (protein 5); (G) grp75 (protein 7); (H) grp78 (protein 8); (I) and hsp70 (protein 9). Proteins 1–9 in (B) correspond to proteins thus indicated in (A) and (B) proteins marked 7, 8, and 9 in (A) represent locations of grp75, grp78, and hsp70, respectively, as judged by Western blotting. For proteins 1–3, 5, and 7–9, blots of the nonadherent cells were probed with specific antibodies.

4 Discussion

Knowledge of the identities of candidate proteins may reveal clinically useful targets of apoptosis, particularly if common mechanisms of action can be identified for these proteins. The overall response of prostate epithelial cells during IR-induced apoptotic cell death represents the composite of changes in protein breakdown and synthesis. Our present study addressing the status of protein expression in an *in vitro* prostate model system, defined to represent apoptosis due to IR, revealed more than ten proteins exhibiting significant quantitative and qualitative changes associated with known events during the execution phase of apoptosis. Furthermore, 2-D PAGE offered an excellent window to observe proteolytic breakdown of IF proteins and accumulation of ER-resident molecular chaperones simultaneously in the apoptotic cells.

4.1 Changes in cytoskeletal proteins

The IF-NM network provides the structural frame work of the cells. Keratin filaments predominate the epithelial cell phenotype. Acidic keratins – type I (K 10–20), and basic keratins – type II (K 1–9), type III (vimentin), and type V (lamins) proteins assemble, forming the IF network [32]. Lamin proteolysis by Mch-2/caspase-6 was well documented as a requirement for packaging of the condensed chromatin into apoptotic bodies. Recent studies reported fragmentation of type I keratins [11–12] in several *in vitro* model systems induced to undergo apoptosis. Our recent studies using the 267B1-XR cells pointed out that proteolysis of vimentin is a component of the apoptotic process as well. In the present study we observe confirmation of the apoptosis-associated breakdown of keratins 18, 19 and vimentin in the 267B1-XR cells.

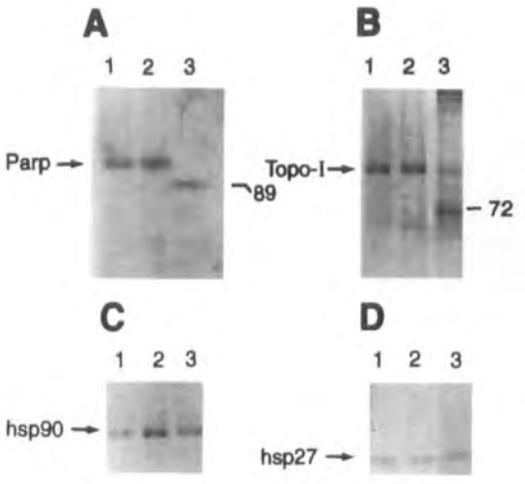


Figure 5. Caspase cleavage of PARP (Parp) and topoisomerase-I during apoptosis. Fifty μg protein from control and apoptotic cells were subjected to SDS-PAGE on 10% polyacrylamide gels, blotted to PVDF membranes, and immunoblotted with specific antibodies. Western blot analysis was carried out with (A) anti-PARP, (B) topo-I, (C) hsp90, and (D) hsp27 antibodies in control, adherent, and nonadherent (lanes 1–3, respectively) 267B1-XR cells 48 h after receiving a 6 Gy dose of IR.

Proteasomes, 14-3-3 proteins, and actin microfilaments have all been reported to be closely aligned with keratin intermediate filament network [11, 29]. The association of 14-3-3 proteins with keratin monomers is dependent on the phosphorylation status of keratins. It is of significance that certain isoforms of 14-3-3 proteins exhibit apoptosis-associated changes in the 267B1-XR prostate cell system used in the present study. Current knowledge of the participation of proteasomes during apoptosis is still being debated, either suggesting a direct role [36], or lack of involvement [35]. Our present data indicate no significant changes in the amounts of α- and β-proteasome subunits during apoptosis of the 267B1-XR cells. At the same time, it is well accepted that limited proteolysis of several actin-binding proteins assists in the disassembly of the IF-NM network of the apoptosing epithelial cells. Note the significance of our results: there is a marked decrease (85–90% in specific isoforms of tropomyosins (Tm-1 and Tm-3), a family of actin-modulating proteins, in the 267B1-XR cells undergoing programmed cell death (Fig. 3A, B). A decrease in the actin filament binding/modulating proteins with intact proteasome components, and altered isoform distribution of 14-3-3 proteins, thus points to the specificity of apoptotis-associated protein changes and is consistent with the fact that other actin-binding proteins such as spectrin, fodrin, and gelsolin, are attacked by caspases during apoptosis [7].

4.2 Changes in ER chaperones

ER is a major Ca storage site within the cell; the resident proteins of the ER assist in the maturation of secretory proteins destined for other subcellular compartments or for the extracellular space. These major molecular chaperones mediate insertion through the ER membrane, carry out post-translational modifications such as glycosylation, proteolytic trimming, and disulfide bond formation, and they assist in folding and assembly of secreted and integral membrane proteins.

A homologue of the hsp90 family members in the ER lumen, grp94, is inducible by various forms of stress [38]. The pair of chaperones, grp94 and grp78, appear as prototypes of a class of genes that are coordinatedly regulated by signal transduction pathways originating in the ER and traveling to the nucleus. Differential levels of induction of grp94 and grp78 were reported in response to thapsigargin, an inhibitor of Ca uptake into the ER, or by ethanol-induced stress [38, 40, 41]. In the present report, we identify grp94 and not grp78 as one of the proteins with elevated levels during apoptosis. Induction of grp75 has been reported within 30 min of receiving 0.25 Gy of IR and this response continued until 6 h post-irradiation. This particular event has been associated with radioresistance phenotype, and antisense oligos directed toward the initiation codon of PBP 74 sensitized HT29 cells to IR [18]. The constitutively expressed members of the highly conserved hsp70 family (hsp70, grp75) are not altered during mid-late phases of apoptosis execution in the present prostate epithelial cell system.

Calreticulin and calnexin, both members of the ER compartment, exhibit certain similarities in their primary structure and appear to function together in the post-translational processing of proteins [37]. Regulation of calreticulin expression has been shown to be mediated by different pathways in response to agents that disturb normal ER functioning [42, 43]. Calnexin is a highly conserved, 90 kDa membrane-bound lectin and a molecular chaperone that binds newly synthesized glycoproteins in the ER. Our present data points out that these two related proteins show increased accumulation in the apoptotic 267B1-XR cells. PDI, another ER protein, has both chaperone and isomerase/foldase activities. PDI acts catalytically to both form and reduce disulfide bonds as part of its isomerase activity [39]. A 2-fold increase in the accumulation of this protein is observed in the apoptotic 267B1-XR cells in the present report.

4.3 Cellular roles of the differentially expressed proteins

Breakdown of IF proteins and components of the actin cytoskeleton solubilize cell contents in preparation for the final disposal of the apoptotic epithelial cells [6–8, 11–12]. The chaperone induction is an ER stress response and is partly independent of the proteolytic responses occurring during the execution phase of apoptosis [13–16]. Increased transcription/translation including protein stabilization have been implicated as being responsible for the observed increases in their levels [36–38, 40–44]. Molecular chaperones and components of the cytoskeleton, both collections of polymeric structures, are often found in association [20]. In the context of apoptosis of epithelial cells, where specific proteolysis, disassembly of cytoarchitecture takes place in phases, any or all of the proposed roles of the molecular chaperones (preventing protein denaturation, dissociating protein aggregates, refolding monomers derived therefrom or directing their proteolytic destruction) [19, 20] could be operational. Chaperones appear to have different but cooperative roles in the formation and function of the eukaryotic cytoskeleton, and thus the concept of looking at them as indicators or biomarkers of cell death/injury is particularly valid. In conclusion, the results of the present study imply that solubilization of the cellular cytoskeleton by the action of cysteine proteases, and disturbance in calcium homeostatis, are probably significant factors in the nature of proteins exhibiting changes in intensities during the execution phase of apoptosis.

These studies were funded by an NIH grant CA45408 to A.D. 2-D PAGE was performed in the Shared Resource Facility of the Lombardi Cancer Center supported by an NCI-funded Cancer Center Support Grant (P30-CA51008). The authors thank Dr. Ramsamooj for providing antisera to calreticulin, Sheri Sharareh and Suneetha Menon for technical assistance, and Elaine North for preparation of the manuscript.

Received October 3, 1998

5 References

[1] Zeitman, A. L., *Semin. Radiat. Oncol. (USA)* 1998, *8*, 81–86.

[2] Dewey, W. C., Ling, C. C., Meyn, R. E., *Int. J. Radiat. Oncol. Biol. Phys.* 1995, *33*, 781–796.

[3] Metcalfe, A., Streuli, C., *BioEssays* 1997, *19*, 711–720.

[4] Muthukkumar, S., Nair, P., Sells, S. F., Maddiwar, N. G., Jacob, R. J., Rangnekar, V. M., *Mol. Cell Biol.* 1995, *15*, 6262–6272.

[5] Baudet, C., Perret, E., Delpech, B., Kaghad, M., Brachet, P., Wion, D., Caput, D., *Cell Death Different.* 1998, *5*, 116–125.

[6] Porter, A. G., Ng, P., Janicke, R. U., *BioEssays* 1997, *19*, 501–507.

[7] Rosen, A., Casiola, K. L., Rosen, L., *J. Cell Biochem.* 1997, *64*, 50–54.

[8] Sane, A.-T., Bertrand, R., *Cancer Res.* 1998, *58*, 3066–3072.

[9] Ku. N. O., Liao, J., Omary, M. B., *J. Biol. Chem.* 1997, *272*, 33197–33203.

[10] Caulin, C., Salvesen, G. S., Oshima, R. G., *J. Cell Biol.* 1997, *138*, 1379–1394.

[11] Chow, S. C., Oeters, I., Orrenius, S., *Exp. Cell Res.* 1995, *216*, 149–159.

[12] Steller, H., *Science* 1995, *267*, 1445–1449.

[13] Marin, M. C., Fermandez, A., Bick, R. J., Brisbay, S., Buja, L. M., Snuggs, M., McConkey, D. J., von Eschenbach, A. C., Keating, M. J., McDonnell, T. J., *Oncogene* 1996, *12*, 2259–2266.

[14] Liu, H., Bowes III, R. C., ven de Water, B., Sillence, C., Nagelkerke, J. F., Stevens, J. L., *J. Biol. Chem.* 1997, *272*, 21751–21759.

[15] Furuya, Y., Lundma, P., Short, A. D., Gill, D. L., Isaacs, J. T., *Cancer Res.* 1994, *54*, 6167–6175.

[16] Bian, X., Hughes, Jr., F. M., Huang, Y., Cidlowski, J. A., Putney, Jr., J. W., *Am. J. Physiol.* 1997, *272*, 1241–1249.

[17] Sadekova, S., Lehnert, S., Chow, T. Y. K., *Int. J. Rad. Biol.* 1997, *72*, 653–660.

[18] Hayes, S. A., Dice, J. F., *J. Cell Biol.* 1996, *132*, 255–258.

[19] Liang, P., MacRae, T. H., *J. Cell Sci.* 1997, *110*, 1431–1440.

[20] Kuettel, M. R., Thraves, P. J., Jung, M., Varghese, S. P., Prasad, S. C., Rhim, J. S., Dritschilo, A., *Cancer Res.* 1996, *56*, 5–10.

[21] Prasad, S. C., Thraves, P. J., Dritschilo, A., Kuettel, M. R., *Electrophoresis* 1997, *18*, 629–637.

[22] Prasad, S. C., Thraves, P. J., Kuettel, M. R., Dritschilo, A., *Crit. Rev. Oncol. Hematol.* 1998, *27*, 69–80.

[23] Prasad, S. C., Thraves, P. J., Dritschilo, A., Kuettel, M., *Prostate* 1998, *35*, 203–211.

[24] Soldatenkov, V., Prasad, S. C., Voloshin, Y., Dritschilo, A., *Cell Death Different.* 1998, *5*, 307–312.

[25] Prasad, S. C., Soldatenkov, V., Srinivasarao, G. Y., Dritschilo, A., *Int. J. Oncol.* 1998, *13*, 757–764.

[26] Prasad, S. C., Thraves, P. J., Kuettel, M. R., Srinivasarao, G. Y., Dritschilo, A., Soldatenkov, V., *Biochem. Biophys. Res. Commun.* 1998, *249*, 332–338.

[27] O'Farrell, P. Z., Goodman, H. M., O'Farrell, P. H., *Cell* 1997, *12*, 1133–1142.

[28] Celis, J. E., Rasmussen, H. H., Gromov, P., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Vorum, H., Kristensen, D. B., Østergaard, M., Haunse, A., Jensen, N. A., Celis, A., Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kjaegaard, I., Andersen, I., Puype, M., Van Damme, J., Vandekerckhove, J., *Electrophoresis* 1995, *16*, 2177–2240.

[29] Liao, J., Omary, M. B., *J. Cell Biol.* 1996, *133*, 345–357.

[30] Vincenz, C., Dixit, V. M., *J. Biol. Chem.* 1996, *271*, 20029–20034.

[31] Muslin, A. J., Tanner, J. W., Allen, P. M., Shaw, A. S., *Cell* 1996, *84*, 889–897.

[32] Chou, Y., Skalli, O., Goldman, R. D., *Curr. Opin. Cell Biol.* 1997, *9*, 49–54.

[33] Scherrer, K., Bay, F., *Prog. Nucleic Acid Res. Mol. Biol.* 1997, *49*, 1–64.

[34] Liao, J., Lowthert, L. A., Ghori, N., Omry, M. B., *J. Biol. Chem.* 1995, *270*, 915–922.

[35] Machiels, B. M., Henfling, M. E., Schutte, B., van Engeland, M., Broers, J. L., Ramaekers, F. C., *Eur. J. Cell Biol.* 1996, *70*, 250–259.

[36] Stefanelli, C., Bonavita, F., Stanic, I., Pignatti, C., Farruggia, G., Masotti, L., Guarnieri, C., Calderera, C. M., *Biochem. J.* 1998, *332*, 661–665.

[36] Rapoport, T. A., Rolls, M. M., Jungnickel, B., *Curr. Opin. Cell Biol.* 1996, *4*, 499–504.

[37] Herbert, D., Foellmer, B., Helenius, A., *Cold Spring Harbor Symp. Quant. Biol.* 1995, *60*, 405–415.

[38] Little, E., Ramakrishnan, M., Roy, B., Gazit, G., Lee, A. S., *Crit. Rev. Eukaryot. Gene Expr.* 1994, *4*, 1–18.

[39] Luz, J. M., Lennarz, W. J., *Experientia* 1996, *77*, 97–117.

[40] Miles, M. F., Wilke, N., Elliot, M., Tanner, W., Shah, S., *Am. Soc. Pharmacol. Exp. Ther.* 1994, *46*, 873–879.

[41] McCormick, T. S., McColl, K. S., Distelhorst, C. W., *J. Biol. Chem.* 1997, *272*, 6087–6092.

[42] Waser, M., Mesaeli, N., Spencer, C., Michalak, M., *J. Cell Biol.* 1997, *138*, 547–557.

[43] Ramsamooj, P., Notario, V., Dritschilo, A., *Cancer Res.* 1995, *55*, 3016–3021.

[44] Brostrom, C. O., Brostrom, M. A., *Prog. Nucleic Acid Res. Mol. Biol.* 1998, *58*, 79–125.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1075–1081

Jean-Philippe Charrier
Carole Tournel
Sandrine Michel
Pascal Dalbon
Michel Jolivet

Département Recherche et Développement, Unité Immunoessais, bioMérieux S.A. Marcy L'Etoile, France

Two-dimensional electrophoresis of prostate-specific antigen in sera of men with prostate cancer or benign prostate hyperplasia

Prostate-specific antigen (PSA), the main marker for prostate cancer (PCa), is released from the prostate into the blood stream at nanogram level and may increase in PCa and nonmalignant disease such as benign prostate hyperplasia (BPH). More recently, advantage was taken of PSA's ability to bind to protease inhibitors in serum in order to improve discrimination between PCa and BPH, using the free PSA to total PSA ratio. The understanding of this phenomenon at molecular level, which is still unknown, may promise new improvements in the field of diagnostics. For this purpose, we determined the pattern of PSA forms in PCa and BPH sera, using the high resolving power of two-dimensional electrophoresis (2-DE) in conjunction with the high sensitivity of chemiluminescence detection. Serum PSA differs drastically from seminal PSA: apart from complexed forms, serum PSA shows few cleaved forms. Moreover, 2-DE patterns from PCa are relatively homogeneous, whereas patterns from BPH may in some cases present a higher proportion of cleaved forms and in other cases present slightly more basic spots. We therefore demonstrated, for the first time, that an increase in the free to total PSA ratio in BPH cases may be due to cleaved PSA forms (which are enzymatically inactive and unable to bind inhibitors), or possibly related to basic free PSA, which may represent the zymogen forms.

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Prostate-specific antigen / Prostate cancer / Benign prostate hyperplasia / Diagnosis EL 3332

1 Introduction

Prostate-specific antigen (PSA) is the main marker for the diagnosis and monitoring of prostate cancer (PCa) [1, 2] which is the most commonly detected cancer, and the second leading cause of cancer deaths in men in Western countries [3]. PSA is a 30–34 kDa chymotrypsin-like serine protease [4], mainly produced by the epithelial cells of the prostate [5], and it consists of 237 amino acids and one *N*-linked carbohydrate chain [6] with various isoforms (p*I* from 6.8 to 7.5) [7]. It may display peptide chain cleavages, which lead to enzymatically inactive forms [8], but the fragments require reducing conditions to be dissociated because of five internal disulfide bonds [9]. PSA is found in seminal liquid at concentrations ranging from 0.5 to 5 mg/mL [10] and is released into the blood stream at a 1 000 000 times lower rate [11]. Its concentration in serum may increase drastically in cases of PCa and moderately in cases of nonmalignant disease such as benign prostate hyperplasia (BPH) [1, 2]. However, the overlapping zone between the different pathologies is responsible for a false positive rate of 65% and a false negative rate of 20–25% for the general prostate cancer population, and a false negative rate of 30–45% for patients with organ-confined – and potentially curable – prostate cancer [2, 12].

More recently, PSA in serum was shown to bind massively to protease inhibitors such as α_1-antichymotrypsin (ACT) and α_2-macroglobulin (α2M) [13, 14]. Moreover, the use of the free PSA to total PSA ratio – where total PSA means free PSA plus PSA bound to ACT (PSA-ACT) – was found to increase diagnostic specificity by 25–35% [14, 15].

However, the reasons for the pathology discrepancy of PSA complexation in serum are still unknown at the molecular level; an understanding of these bases, however, may promise new improvements in specificity and sensitivity. In order to attain this, we determined the pattern of PSA forms in the serum of patients with PCa or BPH, using the high-resolution power of two-dimensional electrophoresis (2-DE) in conjunction with the high sensitivity of chemiluminescence detection. As our objective was to observe PSA-ACT, free PSA and cleaved PSA on the

Correspondence: J.-P. Charrier, BioMérieux, 69280 Marcy L'Etoile, France
E-mail: jcharrier@ml.biomerieux.fr
Fax: +33-478872101

Abbreviations: **ACT**, α_1-antichymotrypsin; **BPH**, benign prostate hyperplasia; **DAB**, 3,3-diaminobenzidine; **α2M**, α_2-macroglobulin; **PCa**, prostate cancer; **PSA**, prostate-specific antigen; **PSA-ACT**, PSA bound to ACT

same image, we had to develop a possibility of detecting with an analytical resolution around ten picograms of PSA. This was achieved through the latest improvements of 2-DE by loading 100 µL of serum during the rehydration of the immobilized pH gradient (IPG) gel [16] and by using thiourea in addition to urea, CHAPS, DTT and Ampholine as sample buffer [17].

2 Materials and methods

2.1 Chemicals and materials

Total PSA and free PSA were determined using the TPSA Vidas immunoassay kit and the fPSA Vidas immunoassay kit, respectively (bioMérieux, Marcy L'Etoile, France). Immobiline Dry-Strip pH 3–10, 18 cm nonlinear, dry-Strip kit and Multiphor II used during the first dimension were from Pharmacia Biotech (Uppsala, Sweden). Second-dimensional vertical gels were cast into a multi-casting chamber (Bio-Rad, Hercules, CA) using acrylamide from Serva (Heidelberg, Germany) and 1,4-Bis(acryloyl)piperazine (PDA) from Bio-Rad. The Protean II xi system for SDS-PAGE, the Trans-blot system for protein blotting, the Fluor-S, the GS-700 densitometer, the Multi-Analyst, and the Melanie II software for membrane and gel analyses were also obtained from Bio-Rad. Western blot was achieved on Immobilon-P membrane (Millipore, Bedford, MA, USA). Peroxidase-labeled secondary antibody was from Jackson Immuno-Research Laboratories (West Grove, PA, USA). The peroxidase substrates, Super Signal Ultra for luminescence and ImmunoPure Metal Enhanced DAB (3,3′-diaminobenzidine) for colorimetric detection, were from Pierce (Rockford, IL, USA). All others chemicals were of the highest purity grade from Sigma (St. Louis, MO, USA) or Merck (Darmstadt, Germany), except carrier ampholytes Servalyt 4–9, which were from Serva.

2.2 Sample preparation

Human semen was allowed to liquefy for 1 h at room temperature and was then centrifuged at 10 000 × *g* for 20 min to remove spermatozoa. Supernatants were stored at –80°C before analysis. Ten µL of seminal plasma or 100 µL of serum were boiled for 5 min with 10 µL or 50 µL, respectively, of 10% SDS, 2.3% DTT. Next, sample volume was completed to 500 µL with a dry-strip rehydration solution (8.3 M urea, 2 M thiourea, 4% CHAPS, 100 mM DTT, 2% Servalyt 4–9, 1 mg/L Orange G). Thiourea, in addition to other chaotrops and detergents, first introduced in order to prevent the loss of membrane proteins [17], was used here to maintain the high resolution of 2-DE when loading large serum volumes on wide-range pH 3–10 nonlinear IPG.

2.3 IPG gel strip rehydration with sample

The dry IPG strips were laid on the sample solution previously loaded into a 20 cm long glass tube, covered with paraffin oil and allowed to rehydrate overnight. Rehydration in an individual glass tube was preferred to rehydration in a grooved reswelling chamber [18], to avoid cross contamination from proteins previously adsorbed on the methacrylate surface, due to the large serum load (data not shown).

2.4 2-DE running and electrotransfer

2-DE was run as described by Bjellqvist *et al.* [19] with some modifications. During the first dimension, a linear increase from 100 V to 3500 V within 8 h and a focus step at 6000 V for 80–100 kVh was performed. Homogeneous 12%T, 2.8%C gels were used for the second dimension. Gels were either silver stained [19] or transferred in a tank using Matsudaira buffer [20] to a PVDF membrane for one night at 1 A. Temperature was maintained at 15°C throughout the experiment.

2.5 Immunodetection of PSA

The membrane was incubated for 1 h with a blocking solution (15 mM Tris, 0.14 M NaCl, 0.5 mL/L Tween 20, pH 8.0, 5% nonfat milk) and subsequently processed for an additional hour with an anti-PSA monoclonal antibody diluted at 10 µg/mL in the blocking solution. This antibody was especially selected for its ability to detect all PSA forms and was shown to recognize an epitope on the *N*-terminal side of the molecule (submitted for publication). The membrane was washed with the blocking solution, and was incubated for 1 h with the peroxidase-labeled anti-mouse antibody diluted 1:5000 in blocking solution. After washing out the secondary antibody, the PSA spots were detected using a colorimetric substrate for seminal liquid or a chemiluminescent substrate for serum. Depending on the total PSA concentration, chemiluminescent detection was performed using the Fluor-S system for 1 min to 1 h using a high sensitivity threshold and a 2.8 aperture for the lens. For quantification purposes, the acquisition durations were elected to avoid any saturation of the images. PSA spots were later selected, quantified, and subtracted from the local background using the Multi-Analyst software.

2.6 Positioning of PSA spots

In order to allow a precise localization of PSA spots on 2-DE maps, all the proteins present on the membrane were stained with Amido Black [21] after specific probing. The specific colorimetric detection with DAB followed by

Amido Black staining led to nonambiguous positioning because of a green/black or a blue/black aspect of PSA *versus* other proteins spots, whereas, after specific revelation by chemiluminescence, the Amido Black-stained membrane needed to be scanned and stacked with the PSA chemiluminescent image using Melanie II software [22]. Subsequent alignment between membranes and gels from different samples was performed using previously identified proteins [23].

3 Results and discussion

3.1 Sensitivity of PSA detection

Using classical analytical 2-DE gels [19], it was not possible to observe all the complex, free, and cleaved forms of PSA in serum (data not shown). Indeed, considering that generally 10–30% of PSA is free and that only a lower proportion is cleaved, the 2–20 ng/mL overlapping zone between benign and malignant prostate diseases required detection of PSA in the 100–1000 fg range. By increasing the sample volume to 100 µL, we were theoretically able to move this range to 10–100 pg. However, the dramatic loss in protein recovery and resolution occurring during a micro-preparative 2-DE using a wide pH range as first dimension, made this approach generally unusable. The use of a narrow pH range was shown to solve this problem [24], but it multiplies gel and serum consumption. We therefore preferred to improve protein solubilization in 3–10 IPG by rehydrating gels directly with both the sample [16] and a solution containing thiourea, urea, and detergents [17].

In the serum, this approach led us to observe on the same image the PSA-ACT, the free PSA, and the cleaved PSA forms, up to around 10 pg (Fig. 1), and in a large PSA concentration range (Table 1). Notably, other authors described only one [25] or no [26] cleaved forms, in spite of pooling serum from patients with advanced PCa and performing additional PSA enrichments using affinity purification. Moreover, the sensitivity required to address PSA patterns in BPH serum or in the overlapping zone between pathologies had never been reached previously.

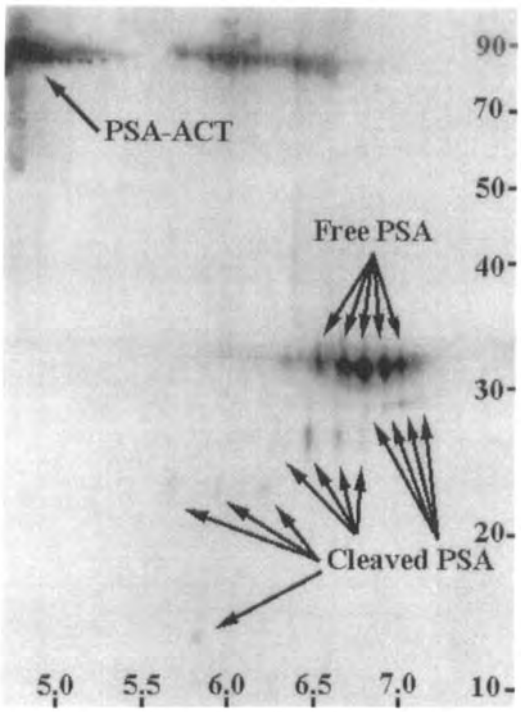


Figure 1. 2-DE Western blot of serum PSA forms from patient with PCa (total PSA: 4000 ng/mL; free PSA: 567 ng/mL). Detection by chemiluminescence. Molecular mass in kDa on the right, isoelectric points at the bottom.

3.2 Serum *versus* seminal liquid PSA forms

In seminal plasma from healthy donors, PSA was shown to present five mature and intact isoforms with a molecular mass around 32 kDa, and to contain numerous cleaved forms with a molecular mass ranging from 10–31 kDa (Fig. 2). This observation is in agreement with previous 2-DE observations [27, 28]. However, depending on individuals, the number and the amount of cleaved forms may vary but are still in significant quantity (data not shown). On the other hand, in PCa and BPH sera (Figs. 1 and 3), the proportion of cleaved forms remains very small. Note that cleaved forms detected us-

Table 1. Quantified PSA forms in patients with PCa or BPH

Samples	Patient 1	Patient 2	Patient 3	Patient 4
Pathology	PCa	PCa	BPH	BPH
Total PSA (ng/mL)	3120	713	3.9	5.7
Free PSA (ng/mL)	151	104	0.87	1.12
Cleaved PSA (ng/mL)	5.98	2.44	0.16	0.14
Free/total PSA (%)	4.8	14.5	22.3	19.6
Cleaved/free intact PSA (%)	4.12	2.4	23.4	13.8

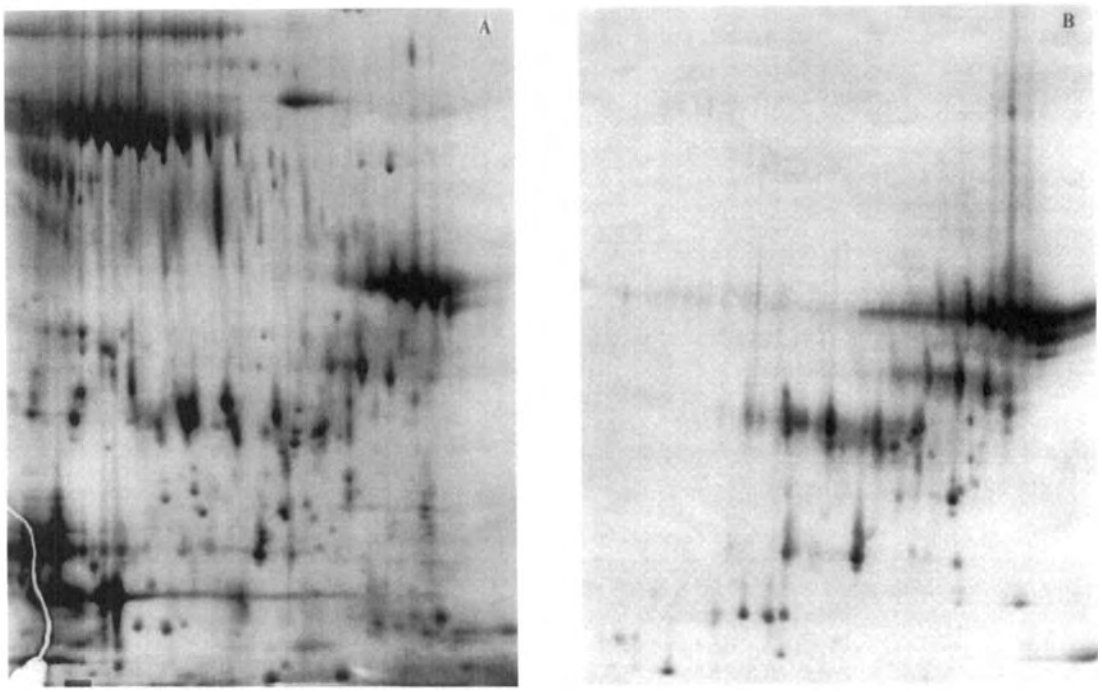


Figure 2. 2-DE of human seminal liquid. Detection using (A) silver staining or (B) Western blotting and PSA probing.

ing Western blot (Figs. 1, 2B and 3) are mainly intact in the *N*-terminal sequence because of antibody specifications and may not be considered as the full cleavage pattern. In other respects, spots of free intact PSA appear similar in both fluids and seem to be essentially composed of spots with comparable patterns to enzymatically active spots [29], whereas spots of PSA-ACT, only described in blood, are only visible on the acidic side of the serum pattern (Fig. 1).

3.3 PCa *versus* BPH PSA forms

When comparing serum from different patients, PSA-ACT leads to similar spots with only quantitative discrepancy (data not shown). In the same way, free intact PSA patterns from cancer patients are relatively homogeneous (Figs. 1, 3A, 3B). The three major spots on each image were estimated to be at 32 kDa and p*I* ranging from 6.6 to 6.9. On the other hand, free PSA from BPH may present more acidic spots, with a p*I* ranging from 6.35 to 6.55 (Fig. 3C), or more basic and heavy spots, with a p*I* ranging from 6.7 to 7.0 and a molecular mass estimated at 32.3 kDa (Fig. 3D). The more acidic spots may be related to a post-translational modification such as an abnormal phosphorylation or glycosylation in BPH tissues. The latter hypothesis may be supported by the higher capacity of PSA, in BPH serum, to bind to concanavalin A [30] and by an isoelectric shift following neuraminidase treatment [31], although these criteria were shown to be inappropriate in effectively distinguishing BPH from PCa [32, 33]. On the other hand, the more basic and heavy spots may be related to the zymogen form of PSA shown to present 3–7 additional amino acids in the *N*-terminal position [34]. Indeed, the pro-PSA, involving an arginine in –1, may present a 0.2 pH unit shift of its isoelectric point following a theoretical calculation [35]. However, the main difference between PCa and BPH serum seems to be the higher proportion of cleaved forms in BPH. Table 1 shows that the cleaved to free intact form ratio is over 13% in BPH cases whereas it is lower than 5% in PCa cases. Moreover, the diversity of cleaved forms suggests the action of various proteases and perhaps a difference in the kinetic or contact time depending on the pathology. Indeed, in comparison with PCa, PSA was shown to be entrapped longer in the cystic-like dilated BPH gland [36], which means PSA may remain in contact with prostate endogenous protease for a longer time in BPH cases before being released into the circulation.

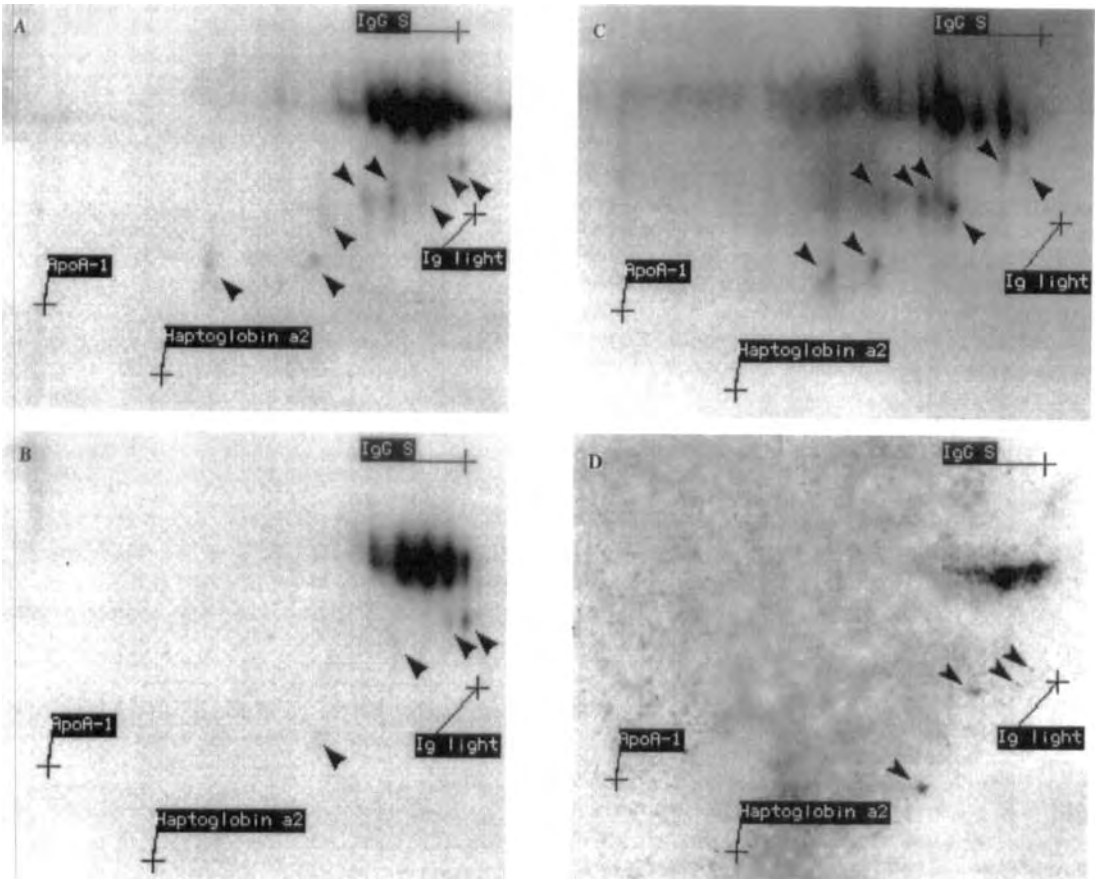


Figure 3. 2-DE of free and cleaved serum PSA forms for patients listed in Table 1. (A) PCa patient 1, (B) PCa patient 2, (C) BPH patient 3, (D) BPH patient 4. Detection by Western blot and chemiluminescence. The raw images, with 4096 levels of gray, are magnified to allow the observation of both cleaved and intact PSA forms. The gray scales are therefore not exactly comparable and the amount and proportion of PSA forms should be read in Table 1. Cleaved PSA forms are labeled with arrows, identified proteins are indicated with the following abbreviations: intermediate segment of the heavy chain of γ-immunoglobulin as IgG S, apolipoprotein A-1 as ApoA-1, haptoglobin α_2 chain as haptoglobin a2, immunoglobulin κ light chain as Ig light.

3.4 Improving diagnosis using inactive PSA forms

The main point in improving the discrimination between PCa and BPH was to better understand why the free PSA to total PSA ratio varies between pathologies. Data reported here show a higher proportion of cleaved or zymogen forms in cases of BPH, which means a higher proportion of forms that are enzymatically inactive and unable to bind to ACT. However, these forms are in much lower proportion than the free 32 kDa – potentially intact and mature – PSA forms. Further investigations will be required to know if these forms are really enzymatically active or not, as suggested by other authors [26]. However, taking into consideration previous studies, part of the free intact PSA may conceivably be active. Indeed, PSA seems to bind reversibly to ACT as shown by the *in vitro* dissociation of the complex and by the higher stability of pure PSA-ACT in the presence of excess ACT [37]. When it dissociates, ACT may lose its inhibitory property [38]; in blood however, an excess over PSA should not affect the association and dissociation rate. Nevertheless, *in vivo*, free PSA and PSA-ACT do not often seem to reach an equilibrium, probably due to the extremely low binding kinetics [39], or to the dependence of the PSA release rate on the prostate tissue origin [40], or possibly to the higher

proportion of ACT in PCa cells [41]. These equilibrium shift hypotheses may be supported by a decrease in free PSA, whereas total PSA remains almost stable in sera stored at 4°C [42].

We therefore assume that free PSA may be divided into three parts: (i) inactive forms (cleaved or zymogen), (ii) active forms in equilibrium with bound forms, and (iii) additional active forms due to an equilibrium shift. PSA derived from the second case seems to be the most important part and must be proportional to PSA-ACT concentrations. Free PSA concentrations may therefore present a higher correlation with PSA-ACT or total PSA concentrations than the inactive forms alone. For this reason, and in line with our initial data, the use of an inactive PSA / free PSA ratio may lead to more discriminating markers than the classically used free PSA / total PSA ratio.

4 Concluding remarks

Performance of analytical 2-DE despite micro-preparative loading onto IPG wide-range pH 3–10 gels, made it possible to demonstrate for the first time that the increase in the free PSA / total PSA ratio in some BPH cases may either be due to cleaved PSA or be related to more basic spots which may possibly be the zymogen forms. The implication of this observation for the diagnostic field will require further investigations on a larger number of samples, but we are now confident that the association of a marker which is mainly organ-specific, such as PSA, with post-translational events in relation with pathologies, will lead to more powerful cancer markers. Furthermore, in the field of prostate diseases, we hope that this strategy will lead to a higher detection rate of organ-confined and potentially curable PCa and a lower rate of unnecessary biopsies for BPH patients.

We wish to thank Drs. M. C. Pinatel and J. Passagot for providing us with human samples, Drs. N. Piga and G. Sibai's teams for providing us with the purified monoclonal antibody against PSA, and C. Micolaud for assistance with the translation.

Received September 1, 1998

5 References

[1] Stamey, T. A., Yang, N., Hay, A. R., McNeal, J. E., Freiha, F. S., Redwine, E., *N. Engl. J. Med.* 1987, *317*, 909–916.

[2] Oesterling, J. E., *J. Urol.* 1991, *145*, 907–923.

[3] Parker, S. L., Tong, T., Bolden, S., Wingo, P. A., *C. A. Cancer J. Clin.* 1997, *47*, 5–27.

[4] Akiyama, K., Nakamura, T., Iwanaga, S., Hara, M., *FEBS Lett.* 1987, *225*, 1894–1900.

[5] Wang, M. C., Valenzuela, L. A., Murphy, G. P., Chu, T. M., *Invest. Urol.* 1979, *17*, 159–163.

[6] Lundwall, A., Lilja, H., *FEBS Lett.* 1987, *214*, 317–322.

[7] Wang, M. C., Kuriyama, M., Papsidero, L. D., Loor, R. M., Valenzuela, L. A., Murphy, G. P., Chu, T. M., *Methods Cancer Res.* 1982, *19*, 179–197.

[8] Zhang, W.-M., Leinonen, J., Kalkkinen, N., Dowell, B., Stenman, U.-H., *Clin. Chem.* 1995, *41*, 1567–1573.

[9] Watt, K. W., Lee, P.-J., M'Timkulu, T., Chan, W.-P., Loor, R., *Proc. Natl. Acad. Sci. USA* 1986, *83*, 3166–3170.

[10] Sensabaugh, G. F., *J. Forensic Sci. 1978, 23, 106–115.*

[11] Papsidero, L. D., Wang, M. C., Valenzuela, L. A., Murphy, G. P., Chu, T. M., *Cancer Res.* 1980, *40*, 2428–2432.

[12] Oesterling, J. E., Jacobsen, S. J., Chute, C. G., Guess, H. A., Lieber, M. M., *Urol. Clin. North Am.* 1993, *20*, 671–680.

[13] Lilja, H., Christensson, A., Dahlén, U., Matikainen, M.-T., Nilsson, O., Pettersson, K., Lövgren, T., *Clin. Chem.* 1991, *37*, 1618–1625.

[14] Stenman, U.-H., Leinonen, J., Alfthan, H., Rannikko, S., Tuhkanen, K., Alfthan, O., *Cancer Res.* 1991, *51*, 222–226.

[15] Catalona, W. J., *Prostate* 1996, *7*, 64–69.

[16] Rabilloud, T., Valette, C., Lawrence, J.-J., *Electrophoresis* 1994, *15*, 1552–1558.

[17] Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., *Electrophoresis* 1997, *18*, 307–316.

[18] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[19] Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.

[20] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[21] Sanchez, J.-C., Ravier, F., Pasquali, C., Frutiger, S., Paquet, N., Bjellqvist, B., Hochstrasser, D. F., Hughes, G. J., *Electrophoresis* 1992, *13*, 715–717.

[22] Appel, R. D., Palagi, P. M., Walther, D., Vargas, J. R., Sanchez, J.-C., Ravier, F., Pasquali, C., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1998, *19*, 2724–2734.

[23] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., Hochstrasser, D. F., *Electrophroesis* 1995, *16*, 1131–1151.

[24] Bjellqvist, B., Sanchez, J.-C., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G. J., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1375–1378.

[25] Noldus, J., Chen, Z., Stamey, T. A., *J. Urol.* 1997, *158*, 1606–1609.

[26] Qian, Y., Sensibar, J. A., Zelner, D. J., Schaeffer, A. J., Finlay, J. A., Rittenhouse, H. G., Lee, C., *Clin. Chem.* 1997, *43*, 352–359.

[27] Lee, C., Tsai, Y., Sensibar, J. A., Oliver, L., Grayhack, J. T., *Prostate* 1986, *9*, 135–146.

[28] Huber, P. R., Schnell, Y., Hering, F., Rutishauer, G., *Scand. J. Urol. Nephrol.* 1987, *104*, 33–39.

[29] Lee, C., Keefer, M., Zhao, Z. W., Kroes, R., Berg, L., Liu, X. X., Sensibar, J. A., *J. Androl.* 1989, *10*, 432–438.

[30] Barak, M., Mecz, Y., Lurie, A., Gruener, N., *Oncology* 1989, *46*, 375–377.

[31] Huber, P. R., Schmid, H.-P., Mattarelli, G., Strittmatter, B., Van Steengrugge, G. J., Maurer, A., *Prostate* 1995, *27*, 212–219.

[32] Chan, D. W., Gao, Y.-M., *Clin. Chem.* 1991, *37*, 1133–1134.

[33] Jung, K., Lein, M., Henke, W., Schnorr, D., Loening, S. A., *Prostate* 1996, *29*, 65–66.

[34] Lövgren, J., Rajakoski, K., Karp, M., Lundwall, A., Lilja, H., *Biochem. Biophys. Res. Commun.* 1997, *238*, 549–555.

[35] Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1023–1031.

[36] Chen, Z., Chen, H., Stamey, T. A., *J. Urol.* 1997, *157*, 2166–2170.

[37] Pettersson, K., Piironen, T., Seppala, M., Liukkonen, L., Christensson, A., Matikainen, M.-T., Suonpaa, M., Lovgren, T., Lilja, H., *Clin. Chem.* 1995, *41*, 1480–1488.

[38] Potempa, J., Fedak, D., Dubin, A., Mast, A., Travis, J., *J. Biol. Chem.* 1991, *266*, 21482–21487.

[39] Huber, P. R., Mattarelli, G., Strittmatter, B., van Steenbrugge, G. J., Schmid, H. P., Maurer, A., *Prostate* 1995, *27*, 166–175.

[40] Hammerer, P. G., McNeal, J. E., Stamey, T. A., *J. Urol.* 1995, *153*, 111–114.

[41] Bjork, T., Bjartell, A., Abrahamsson, P. A., Hulkko, S., Di Sant'Agnese, A., Lilja, H., *Urology* 1994, *43*, 427–434.

[42] Arcangeli, C. G., Smith, D. S., Ratliff, T. L., Catalona, W. J., *J. Urol.* 1997, *158*, 2182–2187.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1082–1086

Eric J. Wagner[1,2]
Russ P. Carstens[1,3,4]
Mariano A. Garcia-Blanco[1,3,5]

[1]Department of Pharmacology and Cancer Biology,
[2]Program in Molecular Cancer Biology,
[3]Department of Medicine,
[4]Division of Nephrology,
[5]Department of Microbiology,
Duke University Medical Center,
Durham, NC, USA

A novel isoform ratio switch of the polypyrimidine tract binding protein

In this report we present evidence for a novel switch in the ratio of the two major isoforms of the polypyrimidine tract binding protein (PTB) in two related prostate cancer cell lines. The existence of different isoforms of PTB is thought to be the result of alternative splicing. We used UV cross-linking to identify a PTB doublet in the DT3 cell line, which is a rat prostate epithelial cancer line that is androgen-dependent and nonmetastatic. The AT3 cell line, a metastatic, androgen-independent cell line derived from the same tumor as the DT3 cells, was noted here to have a different isoform ratio of PTB. The two most prevalent isoforms of PTB were found to bind to an RNA probe containing a pyrimidine stretch. Western blot analysis demonstrated that these isoforms are indeed expressed differently in the two cell lines and that the observed binding is the result of this differential expression. These two cell lines are derived from the original Dunning prostate tumor, which is a model for studying tumor progression in the prostate. This ratio switch may be an important event in tumor progression in this model system of prostate cancer.

Keywords: Polypyrimidine tract binding protein / Alternative splicing / Isoforms / Prostate cancer
EL 3421

The polypyrimidine tract binding protein (PTB) was originally identified based on its ability to bind to pyrimidine-rich regions in an adenovirus major late splicing substrate [1]. Tryptic peptide analysis and cDNA sequencing revealed that human PTB existed in at least three different isoforms with predicted molecular masses of 57.2, 59.03, and 42.8 kDa ([2, 3]; unpublished results). These isoforms have been named PTB1, PTB2, and PTB3, respectively. Although the genomic sequence for PTB has not been published, it is believed that the isoforms are the result of alternative splicing [2, 3]. Western blot analysis demonstrates that human PTB will migrate as a doublet and that the identity of the doublet based on predicted molecular weight is PTB1 and PTB2; PTB3 is not typically detected by Western blot. In rat, PTB has also been seen to exist in multiple isoforms with a doublet having a migration pattern nearly equivalent to human PTB1 and PTB2 [4]. Figure 1 shows a comparison of rat PTB1 (Genbank Accession No. 112251) and human PTB1 and PTB2. Rat PTB1 and human PTB1 are 96% identical, with only a few changes, mostly to conserved amino acids. Although the rat equivalent to human PTB2 has not yet been cloned, a larger isoform has been identified (see results below).

Correspondence: M. Garcia-Blanco, Department of Pharmacology and Cancer Biology, Box 3686, Duke University Medical Center, Durham, NC 27710, USA
E-mail: garci001@mc.duke.edu
Fax: +919-613-8634

Abbreviation: PTB, polypyrimidine tract binding protein

The amino acid sequence of PTB predicts the existence of four weakly conserved RNA recognition motifs (RRMs) that have been implicated in the ability of PTB to oligomerize, bind to RNA, and possibly interact with other proteins [5].

The functions of PTB appear to be diverse. Initially, it was believed to be a necessary splicing factor that played a role in early spliceosome assembly at the branch point and 3′ splice sites; however, it was later observed to be nonessential for the splicing of many introns [1–3]. The research that followed implicated PTB as an alternative splicing factor. Work on the nonskeletal muscle exclusion of exon 7 of the rat β-tropomyosin mRNA [6] and later work on the smooth-muscle-specific repression of exon 3 of the rat α-tropomyosin transcript [7] suggested PTB as a protein factor necessary for repression of these exons. Tissue-specific modulation of PTB levels or differential expression of PTB isoforms may underlie specific regulation [8]. There does appear to be strong evidence for a neural-specific PTB-like protein, which may play a role in the regulation of both the *c-src* mRNA [9] and the $GABA_A$ receptor γ2 mRNA [10]. These data suggest a model, first proposed by Helfman and colleagues, that PTB is a general repressor active in alternative splicing [6]. Perez *et al.* [11] found that, in the case of α-tropomyosin, mutations within a PTB binding site immediately adjacent to the branchpoint sequence affected its ability to repress the inclusion of exon 3 in smooth muscle tissue. These data support the model proposed by Helfman suggesting that PTB could antagonize U2AF when PTB bindings sites are

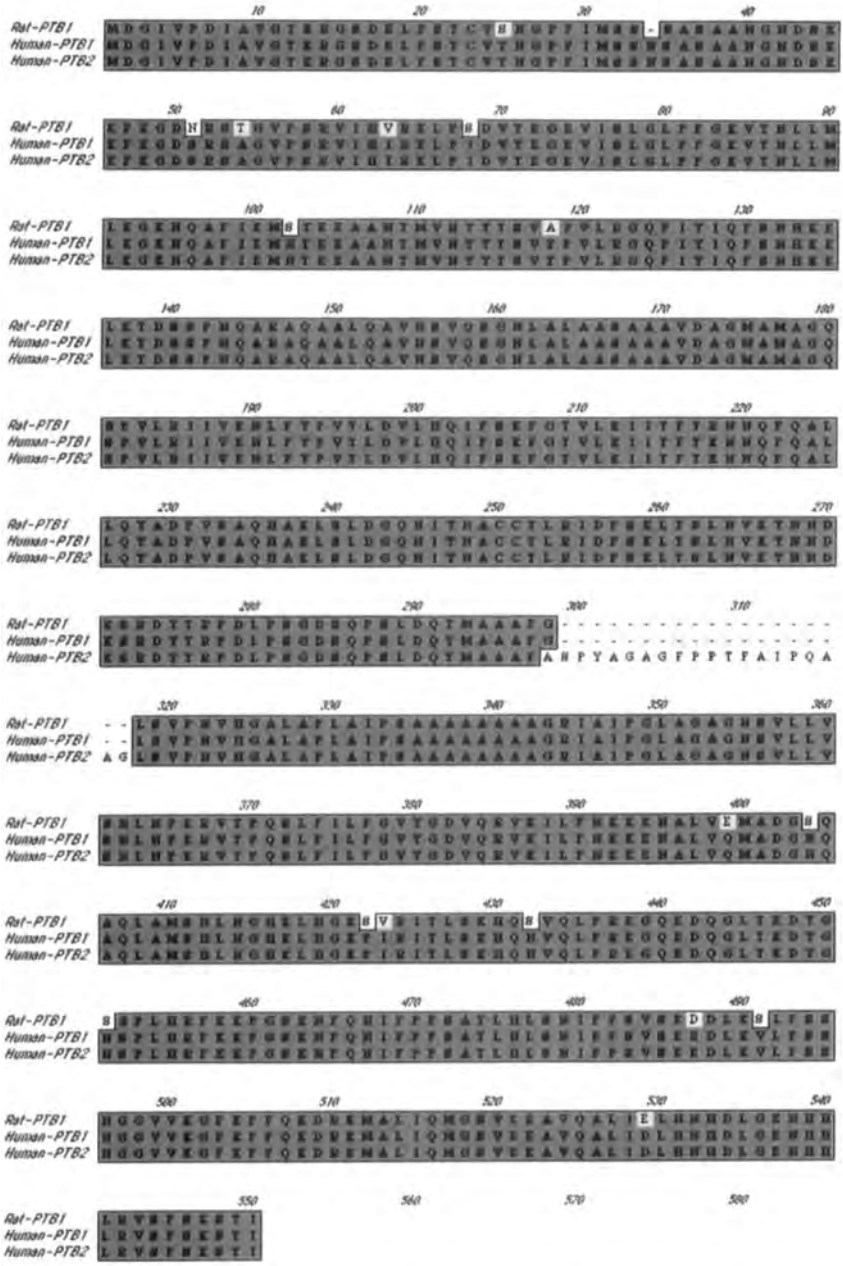


Figure 1. Alignment of amino acid sequence of human PTB1, human PTB2, and rat PTB1. The shaded regions are identical residues between at least two sequences. The white-boxed residues are nonidentical. Note the additional 19 amino acids in human-PBT2 which predict a protein with a molecular mass 1.83 kDa higher. Alignment was done using the Macvector program ClustalW Alignment function.

within the polypyrimidine tract at 3′ splice sites. In cases where PTB binding sites are outside this polypyrimidine tract, a more complex mechanism may be in operation.

A role for PTB in translation has been suggested, which appears separate from its role in splicing, for certain picornavirus messages which contain an internal ribosome entry site (IRES). Although these viruses require PTB for internal ribosome entry-dependent ranslation [12–18], cytoplasmic localization of PTB has not yet been documented. While there is evidence for tissue-specific forms of PTB or PTB-like proteins [9, 10] there is only one example of a change in PTB isoform expression. Recently, Zhang *et al.* [8] demonstrated that the isoform expression

changes as the mouse brain develops. In this report we also note a change in the relative expression of PTB isoforms that correlates with tumor progression in this prostate cancer model.

Our laboratory has been studying the alternative splicing of the fibroblast growth receptor 2 (FGFR2) pre-mRNA in cells derived from the Dunning rat tumor, a system that has been used to model tumor progression in the prostate. In DT3 cells, which are well differentiated and androgen-dependent, the FGFR2 mRNA includes the IIIb exon, which confers the resulting receptor isoform specificity for fibroblast growth factor 7 (FGF-7) also known as keratinocyte growth factor (KGF). AT3 cells were derived from the same tumor but they were harvested after multiple passages through castrated male rat hosts and female hosts. Their phenotype differs from the DT3 in that they are androgen-independent, poorly differentiated, and metastatic. In AT3 cells, alternative splicing of the FGFR2 transcript results in the inclusion of the IIIc exon. This switch gives the FGFR2 a high affinity for FGF-2, which is secreted by the AT3 cells, thus setting up a potential autocrine loop; these events may contribute to the aggressive phenotype observed *in vivo* [19–21]. Important *cis* elements of this RNA transcript have been identified and two of these *cis* elements were the basis of the RNA probe used in this work [22].

UV cross-linking: The RNA probe with the sequence 5′ GGG CGA AUU CCC AUG GAA AAA UGC CCA CAA UGU CAC UGU GGG CUG AUU UUU CCA UGU C 3′ was transcribed *in vitro* with [^{32}P]UTP (specific activity of 5000 Ci/mmole) using an *in vitro* transcription system (Ambion, Austin, TX) and gel purified. Each binding reaction contained 6 μL of nuclear extract (100 μg total protein), 95 mM KCl, 25 μg/mL heparin, 2 mM $MgCl_2$, and approximately 120 fmoles of RNA probe (total volume was 20 μL). All reaction components except probe were mixed on ice and then incubated at 30°C for 8 min, after which probes were added for another 15 min incubation at 30°C. After the incubation, reactions were crosslinked on ice using a Stratagene Stratalinker (La Jolla, CA) at an energy input of two pulses of 500 mJ each. RNAse A was then added at a final concentration of 100 μg/mL for 30 min at 37°C. An equivalent volume of 2 × SDS protein sample buffer was then added (100 mM Tris-HCl, pH 6.8, 200 mM dTT, 4% SDS, 20% glycerol, and 0.2% bromophenol blue). Samples were boiled for 4 min and analyzed on a 12.5% SDS-PAGE. The gel was dried for 1 h and exposed to both X-ray film and a phosphorimager screen.

Immunoprecipitation: 200 μL of a 1:1 v/v suspension of protein A-coated Sepharose beads (Pharmacia Biotech, Uppsala, Sweden) were prepared by incubation with either preimmune serum or PTB antiserum at 4°C for 1 h with rotation. Next, the beads were washed three times with 1 mL of IP buffer (100 mM KCl, 20 mM HEPES, 0.2 mM EDTA, 0.5 mM dTT, 1% Triton-X 100, and 0.5% NP-40, pH 7.9). After the RNAse A treatment of the above reactions (represented as "input" in panel 2B), aliquots were incubated with 20 μL of 1:1 v/v suspension of protein A beads at 4°C for 1 h with rotation in IP buffer. Beads were then washed three times with IP buffer containing 500 mM KCl. Finally, the beads were mixed with 20 μL of 1 × SDS sample buffer, boiled for 4 min and resolved alongside input on a 12.5% SDS-PAGE. Gels were dried and exposed to X-ray film and a phosphorimager screen.

Western blot: 25 μg of protein was mixed with 2 × SDS loading buffer, boiled for 4 min, and resolved on a 12.5% SDS-PAGE. Gels were transferred to a PVDF Immobilon-P membrane (Millipore, Bedford, MA) and blocked in a blocking solution (PBS, 5% nonfat dry milk, and 0.1% Tween-20) overnight, at 4°C, with rotation. Membranes were then incubated in blocking solution with a 1:1500 dilution of a PTB polyclonal antibody that had been Ig-purified (Intron, LLC, Durham, NC) from rabbit PTB antiserum (serum # 79108). Membranes were washed three times with PBS + 0.1% Tween-20 and then incubated with a horseradish peroxidase (HP)-conjugated α-rabbit secondary antibody. Blots were developed using an ECL kit as specified by the manufacturer (Amersham Life Sciences, Buckinghamshire, UK). Whole cell lysates were made by scraping equivalent amounts of cells from plates that were 80% confluent and incubating cells in lysis buffer (100 mM KCl, 20 mM HEPES, 0.2 mM EDTA, 0.5 mM dTT, 1% Triton-X 100, 0.5% NP-40, and 100 μg/mL PMSF) for 1 h on ice. Lysates were spun at 4°C at 20 000 *g* for 2 min to spin down particulate matter and then the supernatant was mixed with an equivalent volume of 2 × SDS sample buffer. The protein was analyzed in the same fashion as the nuclear extract except blots were probed using PTB antiserum (serum # 79113).

One step towards identifying the machinery responsible for regulation of this alternative splicing event is to exploit the identified *cis* elements to detect protein factors that may bind to them. Nuclear extracts of both the DT3 cells and AT3 cells were made and used to UV-cross-link an RNA probe that is composed of important *cis* elements of the FGFR2 primary transcript (Fig. 2A). As can be seen, the ratio fo the PTB doublet appeared to change when comparing the DT3 and AT3 lanes. The third lane is a cross-linking done with HeLa nuclear extracts as a control. The ratio of the PTB doublet, quantified with a phosphorimager, for the DT3 cell line was 1:1.9, with the faster migrating band being more prominent than the slower migrating band. Although the exact identity of the bands has not been determined, the predicted molecular weights

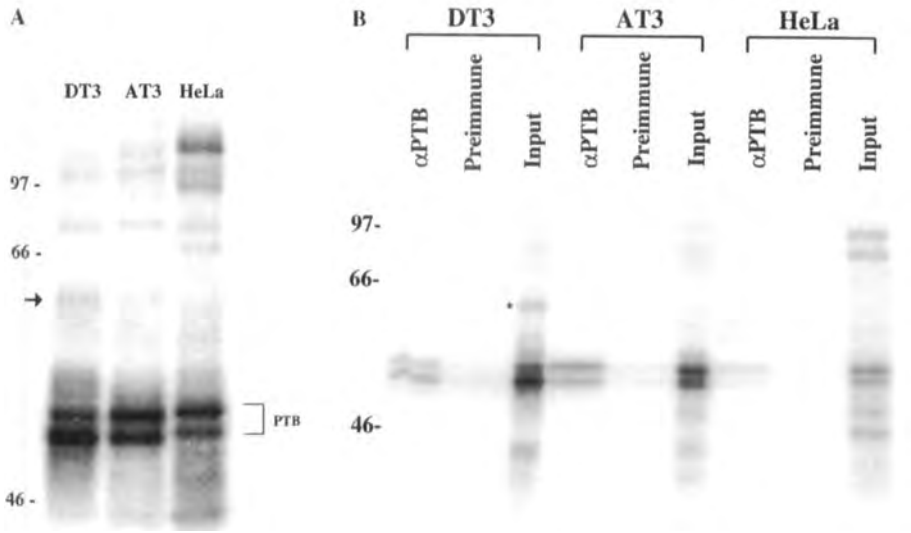


Figure 2. (A) UV cross-linking of proteins from nuclear extracts of DT3, AT3, and HeLa cells to an RNA probe. The UV-cross-linked proteins were resolved on a 12.5% SDS-PAGE and visualized using autoradiography. The bracketed region is the presumed PTB doublet and the arrowhead is a novel band in the DT3 nuclear extract. (B) Immunoprecipitation of the UV-cross-linked PTB doublet. UV cross-linking was performed as in (A); this is represented in the input lane for all three cases. Immunprecipitation was performed on all three cross-linking reactions using both preimmune serum and αPTB serum. The asterisk is next to the novel band in the DT3 extract that was also identified in (A). Molecular mass markers are given in kDa on the left of both figures.

suggest the faster migrating band to be PTB1 and the slower migrating band to be PTB2. The ratio in the AT3 line switched to approximately a 1:1 ratio (1.1:1, PTB2:PTB1), the control lane of HeLa nuclear extract showed the ratio of the doublet that is most commonly seen in previously published data having a profile nearly inverse that of DT3, and that is 1.8:1 (PTB2:PTB1). The slight difference in migration between rat and human PTB is expected; the predicted molecular mass of rat PTB1 is 56.9 kDa [4] and the human PTB1 is 57.2 kDa. The other bands seen on the gel represent other proteins that were UV-cross-linked to this RNA probe; their identity is unknown.

In order to confirm the identity of the PTB doublet, immunoprecipitation of the cross-linked proteins was performed using a PTB antiserum. PTB antiserum immunoprecipitated the cross-linked doublet whereas the preimmune serum did not (Fig. 2B). The background observed in the preimmune serum lanes was due to the Sepharose beads used in the experiment. This result has been repeated using monoclonal antibodies raised against PTB. Note that the overall cross-linking profile was similar between the AT3 and DT3 extracts, consistent with the notion that these cells are highly related and that any differences may be subtle. HeLa cells had a different set of proteins that crosslinked to the RNA probe. Overall, only one band migrating at around 62 kDa, cross-linked to the probe in the DT3 nuclear extract but not in the AT3 extract (labeled with an arrow in Fig. 2A and an asterisk in Fig. 2B). The identity of this cross-linked protein is unknown. Furthermore, there also appeared to be differences in the relative intensities of other cross-linked proteins. Equivalent amounts of nuclear extracts were used in each experiment, thus this may be the result of a slightly altered expression of proteins between the two extracts.

The simplest possibility for the observation is the change in the PTB cross-linking ratio is the result of a change in the levels of its isoforms relative to each other. To test this hypothesis, Western blot analysis was performed using a polyclonal antibody raised against PTB on the DT3, AT3, and HeLa nuclear extracts. As can be seen in Fig. 3A, there was a ratio switch of the PTB isoforms. To rule out artifacts due to nuclear extract preparation, whole cell lysates were made and PTB antiserum was used to identify PTB in a Western blot assay (Fig. 3B). Differential expression of PTB was confirmed in the whole cell lysates. Although densitometric quantification of the PTB isoforms is not as accurate as in that obtained with the phosphorimager, the ratio of the isoforms was at least consistent and potentially even more pronounced. The DT3 ratio of

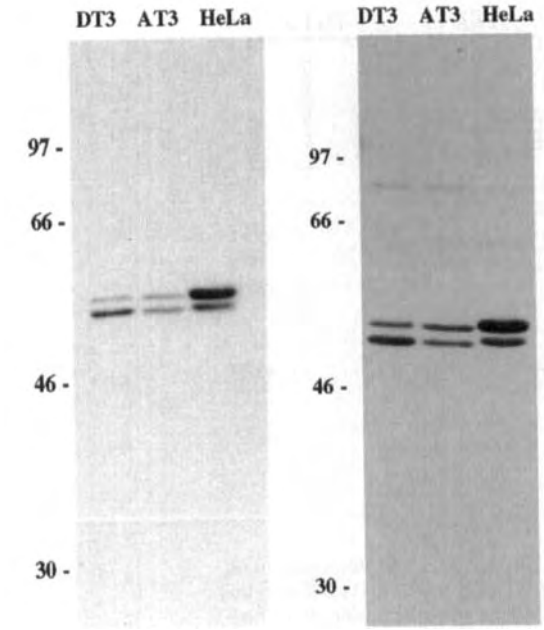


Figure 3. (A) Western blot of nuclear extracts from the DT3, AT3, and HeLa cell lines. Proteins were resolved on a 12.5% SDS-PAGE and the Western blot was performed using an Ig-purified polyclonal antibody to PTB. (B) Western blot of cell lysates made from the DT3, AT3, and HeLa cells. Proteins were resolved on a 12.5% SDS-PAGE and blotted using PTB antiserum. Molecular mass markers are given in kDa.

PTB2 to PTB1 was 1:2.1. The ratio of PTB2 to PTB1 in the AT3 lane was 1.2:1 and for HeLa it was over 3:1; the latter ratio may be slightly overestimated due to the nonlinearity of the film.

Little is known about the genetic events that are necessary for tumor progression and it is possible that alterations in the expression of proteins involved in RNA processing could have many important downstream effects. A current hypothesis is that in order for cancer cells to escape from the primary tumor site, they must undergo several major phenotypic changes. It is possible that these phenotypic changes arise from subtle changes in gene products, resulting in changes in the structures of proteins important in maintaining tissue integrity. The altered properties of these critical proteins could be brought about by changes in the alternative splicing machinery. In this report we noted changes in relative expression of PTB isoforms correlating with a change in metastatic potential. It is possible that changes in PTB isoforms could lead to changes in transcripts that are regulated by PTB. Understanding how PTB isoforms are regulated may shed light on the cause of its ratio switch and could also be informative in understanding the basic molecular mechanisms that contribute to tumor progression.

We thank Dr. David Helfman (Cold Spring Harbor Laboratories) for providing monoclonal antibodies used in experiments that support the data presented in this paper, as well as Wallace McKeehan (Texas A&M University) for his kind gift of the two rat cell lines used. We also thank L. Lindsey and other members of the Garcia-Blanco lab for the helpful discussions and A. Goldstrohm for reviewing the manuscript. This work was funded by a grant from the American Cancer Society to M.A.G.-B.

Received February 11, 1999

References

[1] Garcia-Blanco, M., Jamison, S., Sharp, P., *Genes Develop.* 1989, *3*, 1874–1886.

[2] Gil, A., Sharp, P., Jamison, S., Garcia-Blanco, M., *Genes Develop.* 1991, *5*, 1224–1236.

[3] Patton, J., Mayer, S., Tempts, P., Nadal-Ginard, B., *Genes Develop.* 1991, *5*, 1237–1251.

[4] Jansen-Durr, P., Boshart, M., Lupp, B., Bosserhoff, A., Frank, R., Shutz, G., *Nucleic Acids Res.* 1992, *20*, 1243–1249.

[5] Oh, Y., Bumsuk, H., Kim, Y., Lee, H., Lee, J., Song, O., Tsukiyama-Kohara, K., Kohara, M., Nomoto, A., Jang, S., *Biochem. J.* 1998, *331*, 169–175.

[6] Mulligan, G., Guo, W., Wormsley, S., Helfman, D., *J. Biol. Chem.* 1992, *267*, 25480–25487.

[7] Gooding, C., Roberts, G., Moreau, G., Nadal-Ginard, B., Smith, C., *EMBO J.* 1994, *13*, 3861–3872.

[8] Zhang, L., Weiquin, L., Grabowski, P., *RNA* 1999, *5*, 117–130.

[9] Chan, R., Black, D., *Mol. Cell. Biol.* 1997, *17*, 4667–4676.

[10] Ashiya, M., Grabowski, P., *RNA* 1997, *3*, 996–1015.

[11] Perez, I., Lin, C., McAfee, J., Patton, J., *RNA* 1997, *3*, 764–778.

[12] Luz, N., Beck, E., *FEBS Lett.* 1990, *269*, 311–314.

[13] Luz, N., Beck, E., *J. Virol.* 1991, *65*, 6486–6494.

[14] Jang, S. K., Wimmer, E., *Genes Develop.* 1990, *4*, 1560–1572.

[15] Hellen, C. U. T., Witherell, G. W., Schmid, M., Shin, S. H., Pestova, T. V., Gil, A., Wimmer, E., *Proc. Nat. Acad. Sci. USA* 1993, *90*, 7642–7646.

[16] Borman, A., Howell, M. T., Patton, J., Jackson, R., *J. Gen. Virol.* 1993, *74*, 1775–1788.

[17] Kaminski, A., Hunt, S., Patton, J., Jackson, R., *RNA* 1995, *1*, 924–938.

[18] Neipmann, M., *FEBS Lett.* 1996, *60*, 499–511.

[19] Yan, G., Fukabori, Y., McBride, G., Nikolaropolous, S., McKeehan, W., *Mol. Cell. Biol.* 1993, *13*, 4513–1422.

[20] Feng, S., Wang, F., Matsubara, A., Kan, M., McKeehan, W., *Cancer Res.* 1997, *57*, 5369–5378.

[21] Carstens, R., Eaton, J., Krigman, P., Walther, P., Garcia-Blanco, M., *Oncogene* 1997, *15*, 3059–3065.

[22] Carstens, R., McKeehan, W., Garcia-Blanco, M., *Mol. Cell. Biol.* 1998, *18*, 2205–2217.

From Genome to Proteome: *Advances in the Practice and Application of Proteomics*
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1087–1091

Theo M. Luider[1]
Johan M. Kros[2]
Peter A. E. Sillevis Smitt[1]
Martin J. van den Bent[1]
Charles J. Vecht[1]

[1]Dept.of Neuro-Oncology, Daniel den Hoed Cancer Center,
[2]Dept. of Pathology, University Hospital Rotterdam-Dijkzigt, Josephine Nefkens Building, Rotterdam, The Netherlands

Glial fibrillary acidic protein and its fragments discriminate astrocytoma from oligodendroglioma

In the last few years it has been shown that anaplastic oligodendrogliomas, in contrast to anaplastic astrocytomas, are responsive to a three drug regimen chemotherapy. The histologic criteria for the discrimination between oligodendrogliomas and astrocytomas are subject to substantial interobserver variability, particularly in anaplastic and mixed gliomas. In the present study a two-dimensional electrophoresis technique (2-DE) has been applied to glioma samples in an attempt to discriminate the glioma subtypes. It was found that the presence of glial fibrillary acidic protein (GFAP) fragments distinguishes oligodendroglioma from astrocytoma. One-dimensional (1-DE) immunoblots were compared with immunohistologically stained tissue sections in which various GFAP-positive cell types were seen. It is concluded that 2-DE and 1-DE GFAP immunoblotting provide accurate information for the reliable discrimination of anaplastic astrocytomas and oligodendrogliomas.

Keywords: Glioma / Astrocytoma / Oligodendroglioma / Glial fibrillary acidic protein / Two-dimensional electrophoresis
EL 3353

1 Introduction

Recently, it has been shown that anaplastic oligodendrogliomas, in contrast to anaplastic astrocytomas, are remarkably chemosensitive tumors. Systemic chemotherapy with procarbazine, CCNU and vincristine (PCV) in cases of recurrent anaplastic oligodendroglioma results in a response rate of 60–70% with a median response duration of about 18 months [1–3]. Obviously, this striking difference in therapy response [4, 5] necessitates accurate discrimination between oligodendrogliomas and astrocytomas. Unfortunately, the morphologic criteria used to differentiate between astrocytomas and oligodendrogliomas are subject to substantial interobserver variability [6–8]. Particularly in cases of anaplatic tumors, the classic morphologic criteria have often been lost. Characteristically, astrocytic tumor cells contain glial fibrillary acidic protein (GFAP)-positive intermediate filaments, while the majority of oligodendroglial tumor cells lack filaments and are immunonegative for GFAP. However, GFAP-positive cells may be encountered in oligodendrogliomas. These so-called transitional cells have round cell bodies with short cell processes, containing intermediate filaments, which are immunoreactive with anti-GFAP antibodies [9]. On the other hand, the cells in anaplastic astrocytomas may have lost the slender cell processes with intermediate filaments, while the immunoreactivity for GFAP has dropped dramatically, thus obscuring their astrocytic lineage. In the present study the value of GFAP and its fragments was evaluated for the discrimination of astrocytoma from oligodendroglioma. Protein spots found by 2-D electrophoresis were immunoblotted and the results were correlated with those of GFAP-immunohistochemistry.

Correspondence: Theo M. Luider, Ph.D., Department of Neuro-Oncology, Josephine Nefkens Building, P.O. Box 1738, 3000DR Rotterdam, The Netherlands
E-mail: luidertm@wxs.nl
Fax: +31-10-4088365

Abbreviations: GFAP, glial fibrillary acidic protein; **PCV**, procarbazine, CCNU, vincristine

2 Materials and methods

2.1 Tumor specimens and histological diagnosis

Six histologically proven oligodendrogliomas and six astrocytomas were used for this study. Tumor specimens were obtained at the time of surgery. The specimens were immediately snap-frozen and stored in liquid nitrogen, while histology was confirmed by making frozen sections of the same parts. Adjacent parts of the specimens were used for routine histology and immunohistochemistry. The histological diagnosis was made according to WHO criteria [10]. In addition, paraffin-embedded tissue sections of the respective tumor samples were immunostained with anti-GFAP monoclonal antibody (ICN Biomedicals, Aurora, OH, USA).

2.2 Two-dimensional electrophoresis

2.2.1 Sample preparation

Approximately 50 mg tumor tissue was homogenized in 350 µL sample buffer I (0.3% sodium dodecyl sulfate, 200 mM dithiothreitol, 28 mM Tris-HCl, 22 mM Tris base)

and incubated at 100°C for 5 min. The sample was placed on ice for 5 min. Thereafter, 50 µL sample buffer II (24 mM Tris base, 476 mM Tris-HCl, 50 mM $MgCl_2$, 1 mg/mL DNAse, 0.25 mg/mL RNAse A) was added and the sample was incubated for 10 min. The sample was centrifuged in an Eppendorf centrifuge (12 000 rpm) for 5 min. The supernatant was saved and 100 µL 50% trichloroacetic acid was added. The sample was incubated for 10 min or overnight. The suspension was centrifuged for 2 min (12 000 rpm) and the supernatant was discarded. The pellet was saved on ice and washed twice with 0.5 mL 95% acetone and centrifuged at 10 000 rpm for 5 min. The pellet was air-dried at room temperature and dissolved in 200 µL sample buffer mix (7.92 M urea, 0.06% SDS, 1.76% carrier ampholytes, 120 mM dithiothreitol, 3.2% Triton X-100, 22.4 mM Tris-HCl, 17.6 mM Tris base). The solution was incubated for 5 min with 800 µL acetone (100%). After centrifugation (10 000 rpm, 5 min) the pellet was resuspended into 200 µL sample buffer mix. For one-dimensional electrophoresis (1-DE) the samples were diluted with 50 mM Tris (pH 6.8), 33% glycerol, 10 mM dithiothreitol, and 0.001% bromophenol blue and then loaded without boiling.

2.2.2 Electrophoresis

The 2-DE was performed on an Investigator 2-D electrophoresis system (ESA, Chelmsford, MA, USA) as recommended by the manufacturer. All chemicals were supplied by ESA. The first dimension was performed at 18 000 Vh. The pH range of the carrier ampholytes was 3–10. The percentage of the duracryl was 10%. The gels were silver-stained according to Morrissey [11]. The molecular weight and isoelectric points were determined with a Carbamylyte calibration kit for 2-DE (Pharmacia-Amersham, Roosendaal, The Netherlands). Transfer blotting was performed in an IsoDalt system (Pharmacia) using 25 mM Tris base, 192 mM glycine, and 10% methanol. Proteins were transferred overnight at a constant current of 0.5 A at 4°C on Immobilon-P (Millipore, EttenLeur, The Netherlands).

2.2.3 Immunoblotting

Immobilon P membranes were incubated with sterilized low-fat milk for 1 h. Membranes were incubated for 2 h with mouse anti-GFAP monoclonal (ICN Biomedicals), diluted 1:200 in 20 mM Tris base, 320 mM NaCl, pH 7.5 (buffer A). Subsequently, the membranes were incubated for 1 h with alkaline phosphatase-conjugated rabbit anti-mouse immunoglobulin diluted 1:1000 (Dako, Glostrup, Denmark). The phosphatase activity was visualized with nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad, Veenendaal, The Netherlands). Between the incubation steps the membranes were thoroughly rinsed with buffer A as well as buffer A with 0.05% Tween-20 (ICN). Omitting the first antibody resulted in unstained samples on the Immobilon P membranes.

3 Results

Silver-stained 2-D images of the six astrocytomas and six oligodendrogliomas were compared by naked eye. The investigators were ignorant of the histologic diagnosis of the specimens. The majority of mass protein spots of the tumor samples showed overlap. A cluster of protein spots with a molecular mass (M_r) of 36 kDa (Fig. 1B, arrows 15–19) and two spots with slightly larger M_r (Fig. 1B, arrows 13 and 14) were found in the six samples of astrocytoma, but not seen in the oligodendroglioma specimens (Luider *et al.*, paper submitted). In control white matter, water-soluble fragments of 36 kDa were also seen. Since the M_r and p*I* of the protein spots (15–19 in Fig. 1) were in the range of glial fibrillary acidic protein fragments (GFAP) [12], the spots were immunoblotted with anti-GFAP. The spots numbered 1–19 were all GFAP-immunoreactive (Fig. 2). In addition, tissue sections of the tumors were immunostainedwith anti-GFAP monoclonal antibody. In astrocytoma, all tumor cells were immunopositive (Fig. 1A). Not only tumor cells with fibrillary phenotypes, but also the gemistocytic cells showed intense immunostaining (Fig. 1A). Part of the oligodendrogliomas (4/6) harbored GFAP-positive cells, phenotypically either typical oligodendroglial cells (so-called transitional cells), or gemistocytic cells of variable sizes (Fig. 1C). In four of six oligodendrogliomas and all six astrocytomas, at least two isoforms of 50 kDa GFAP molecules were found by 1-D immunoblotting (Fig. 3). In the normal white matter control, no 50 kDa GFAP was seen. In all astrocytomas, 36 kDa GFAP fragments were observed. Also, in normal white matter 36 kDa GFAP was seen. No 36 kDa GFAP fragments were present in any of the oligodendrogliomas. In two of six cases of oligodendroglioma, no GFAP immunoreactivity was seen (lanes 2 and 4).

4 Discussion

The histological delineation of oligodendroglioma and astrocytoma has gained importance because of the chemosensitivity of the former. Histologically, the distinction between the two gliomas depends largely on morphological characteristics. In the last decades, immunohistochemistry for GFAP has proven a valuable tool in the identification of astrocytic cells and astrocytic neoplasms [13, 14]. However, the identification of subsets of GFAP-positive cells in oligodendroglial tumors has invalidated the specificity of GFAP immunohistochemistry for astrocytomas [9].

In the meantime, various proteins have been proposed as candidate markers for oligodendroglial or astrocytic tumor lineage [15–19]. Unfortunately, none of these has emerged as a sufficiently reliable tumor marker so far.

The technique of 2-DE is a potential powerful tool in the identification of lineage-specific proteins in tumors. 2-DE has been successfully applied to detect proteins, which are specifically present in human cancers and melano-

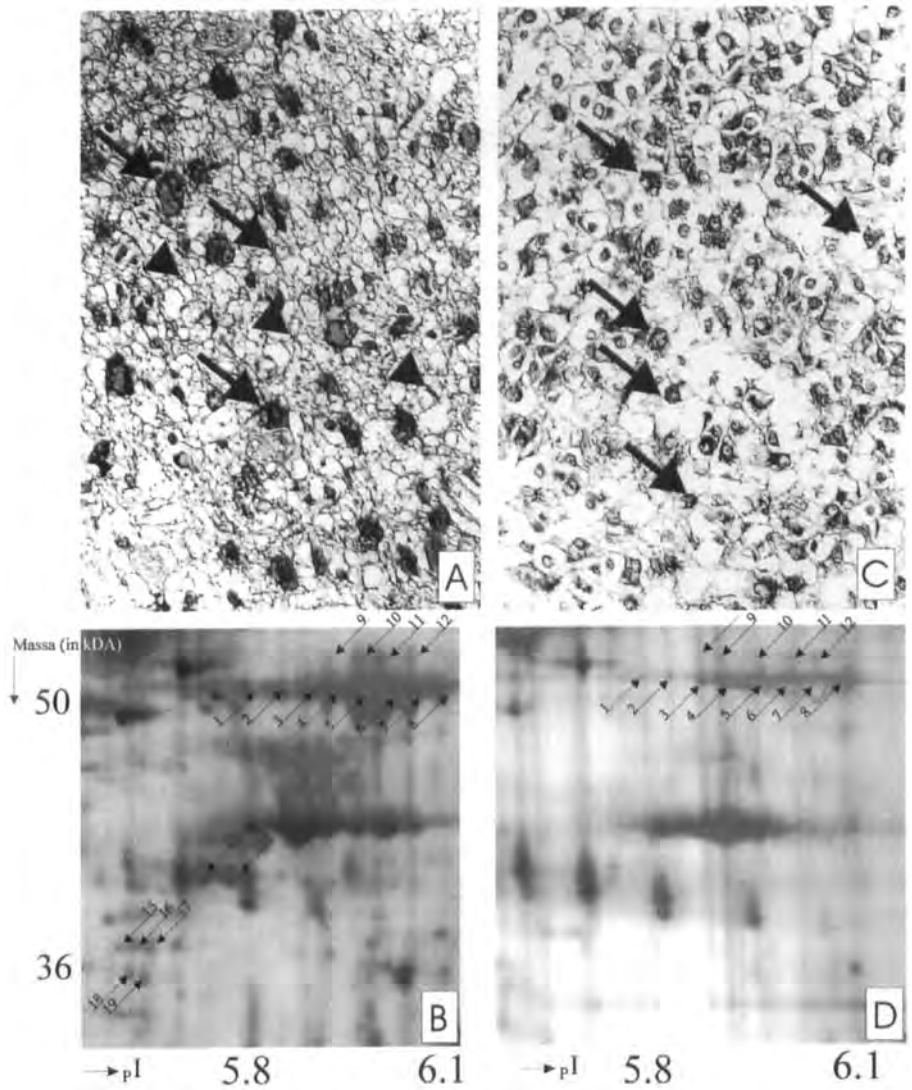


Figure 1. (A) Astrocytoma immunostained for anti-GFAP (× 40). Tumor cells with fibrillary phenotypes (arrowheads) are interspersed with cells with gemistocytic cells (arrows). Both cell types stain intensely for GFAP. (B) Corresponding silver-stained 2-DE gel. Spots of water-soluble 50 kDa GFAP isoforms (numbers 1–12) are present. Fragments of GFAP are numbered 13–19. (C) Oligodendroglioma immunostained for anti-GFAP (× 40). In this tumor, some GFAP-positive cells are present (arrows). These so-called transitional cells are morphologically characterized as oligodendroglial cells with a small GFAP-positive rim of cytoplasm, or as small gemistocytic cells. (D) Corresponding silver-stained 2-DE gels. Also in oligodendroglioma the 50 kDa GFAP isoforms are appreciated (arrows 1–12). No 36 kDa GFAP fragments are present in the oligodendroglioma.

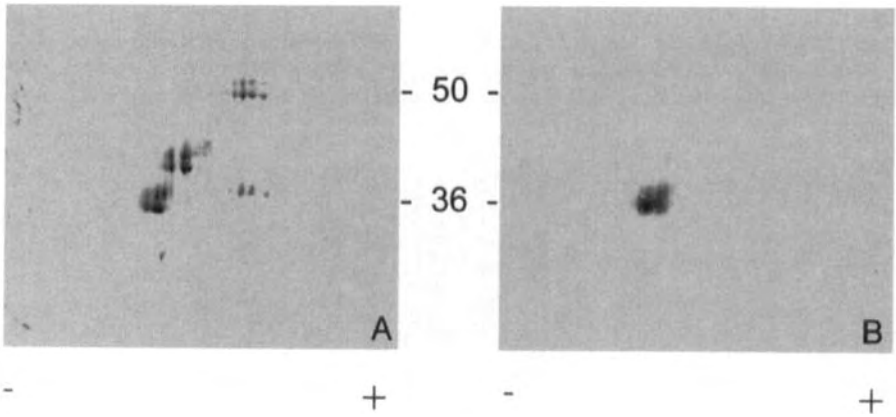


Figure 2. (A) 2-DE GFAP immunoblots of astrocytoma and (B) control white matter. In astrocytoma, a range of 50 kDa GFAP isoforms is observed. In addition, GFAP fragments with molecular masses between 36 and 50 kDa are present (panel A). In control white matter exclusively 36 kDa GFAP soluble fragments are found (panel B).

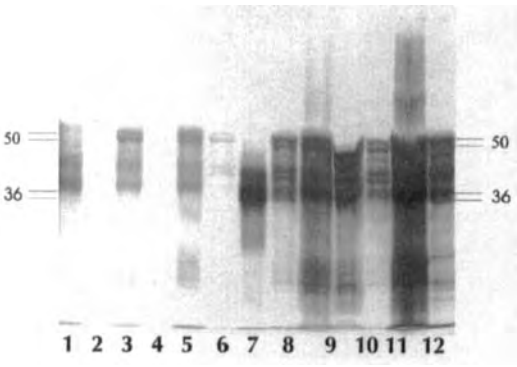


Figure 3. 1-DE GFAP immunoblots of 6 oligodendrogliomas (lanes 1–6) and 6 astrocytomas (lane 8–12). Lane (7) represents control white matter. While the 36 kDa GFAP fragments are present in astrocytoma and control white matter, they are not seen in oligodendroglioma samples.

mas [20–22]. To date, the approach has not been used for glial neoplasms. In the present study, GFAP-immunopositive cells were found both in astrocytomas and oligodendrogliomas. The 36 kDA GFAP fragments are specifically present in astrocytomas, not in oligodendrogliomas. Since the anti-GFAP antibodies react with all GFAP isoforms and their fragments, GFAP immunohistochemistry cannot be used to distinguish between the two glioma subtypes. The discriminative 36 kDa fragments can be traced by the technique of electrophoresis.

In astrocytomas and part of the oligodendrogliomas, twelve isoforms of GFAP of approximately 50 kDa were seen; these were water-soluble and immunoreactive for anti-GFAP. The mass weights of these molecules are in the range of the calculated molecular mass of the GFAP molecule (*viz*, 49.9 kDa using the DNA sequence of the gene for GFAP). The calculated p*I*, however, is 5.42. The p*I*s of the twelve GFAP molecules found in the tumors varied between 5.8–6.1. The shift towards more acidic p*I* of the fragments can be explained by the breakdown of the GFAP isoforms due to sample preparation, or metabolic turnover [12]. In the astrocytoma samples, a range of GFAP fragments with molecular masses varying between 36 and 50 kDa were found (Figs. 1B and D; Fig. 2A). Post-translational modification of the GFAP molecules or the existence of splice variants might account for this variation in molecular mass weights [12, 23]. The same factors could hypothetically play a role in the difference in water-solubility between the 50 kDa fragments found in the gliomas and those in control white matter. It is concluded that the clinically important differentiation of oligodendroglioma from astrocytoma can be reached by using the technique of 2-D electrophoresis. Specifically in astrocytomas, 36 kDa GFAP fragments are found. Future studies may correlate the specific fragmentation of GFAP found in gels with the various GFAP-positive cells seen in immunohistochemically stained tissue sections of oligodendrogliomas and astrocytomas.

The authors would like to thank Mr. L. Uitterwillighen for the preparation of the GFAP immunohistochemistry. This work was supported by The Trust Foundation, Erasmus University Rotterdam by B&L-systems, Maarssen, and by the Revolving Fund, Rotterdam.

Received November 2, 1998

5 References

[1] Cairncross, J. G., MacDonald, D. R., *Ann. Neurol.* 1988, *23*, 360–364.

[2] Cairncross, J. G., *Cancer Res.* 1994, *135*, 127–133.

[3] Van den Bent, M. J., Kros, J. M., Heimans, J. J., Pronk, L. C., van Groeningen, C. J., Krouwer, H. G. J., Taphoorn, M. J. B., Zonnenberg, B. A., Tijssen, C. C., Twijnstra, A., Punt, C. J. A., Boogerd, W., *Neurology*, in press.

[4] Kim, L., Hochberg, F. J., Thornton, A. F., Harsh, G. R., Patel, H., Finkelstein, D., Louis, D. N., *J. Neurosurg.* 1996, *85*, 602–607.

[5] Glass, J., Hochberg, F. H., Gruber, M. L., Louis, D. N., Smith, D., Rattner, B., *J. Neurosurg.* 1992, *76*, 741–745.

[6] Scott, C. B., Nelson, J. S., Farnan, N. C., Curran, W. J., Jr., Murray, K. J., Fischbach, A. J., Gaspar, L. E., Nelson, D. F., *Cancer* 1995, *76*, 307–313.

[7] Krouwer, H. G. J., Van Duinen, S. G., Kamphorst, W., van der Valk, P., Algra, A., *J. Neuro-Oncology* 1997, *33*, 223–238.

[8] Coons, S. W., Johnson, P. C., Scheithauer, B. W., Yates, A. J., Pearl, D. K., *Cancer* 1997, *79*, 1381–1393.

[9] Herpers, M. J., Budka, H., *Acta Neuropathol. (Berl.)* 1984, *64*, 265–272.

[10] Kleihues, P., Burger, P. C., Scheithauer, B. W. (Eds.), *Histological Typing of Tumours of the Central Nervous System. World Health Organization International Histological Classification of Tumours*, Springer Verlag, Berlin 1993.

[11] Morrissey, J. H., *Anal. Biochem.* 1981, *117*, 307–310.

[12] Bigbee, J. W., Bigner, D. D., Pegram, C., Eng, L. F., *J. Neurochem.* 1983, *40*, 460–467.

[13] Bignami, A., Dahl, D. J., *Histochem. Cytochem.* 1977, *25*, 466–469.

[14] Deck, J. H. N., Eng, L. F., Bigbee, J., Woodcock, S. M., *Acta Neuropathol. (Berl.)* 1978, *42*, 183–190.

[15] Nakagawa, Y., Perentes, E., Rubinstein, L. J., *Acta Neuropathol. (Berl.)* 1986, *72*, 15–22.

[16] Yokoo, H., Sasaki, A., Hirato, J., Nakazato, Y., *J. Neuropath. Exp. Neurology* 1996, *55*, 716–721.

[17] Bronstein, J. M., Popper, P., Micevych, P. E., Farber, D. B., *Neurology* 1996, *47*, 772–778.

[18] Golfinos, J. G., Norman, S. A., Coons, S. W., Norman, R. A., Ballecer, C., Scheck, A. C., *Clin. Cancer Res.* 1997, *3*, 799–804.

[19] Landry, C. F., Verity, M. A., Cherman, L., Kashima, T., Black, K., Yates, A., Campagnoni, A. T., *Cancer Res.* 1997, *57*, 4098–4104.

[20] Ji, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D., Simpson, R. J., *Electrophoresis* 1994, *15*, 391–405.

[21] Tang, N. E. M. L., Luyten, G. P. M., Mooy, C. M., Naus, N. C., de Jong, P. T. V. M., Luider, T. M., *Melanoma Res.* 1996, *6*, 411–418.

[22] Qian, Y., Sensibar, J. A., Zelner, D. J., Schaeffer, A. J., Finlay, J. A., Rittenhouse, H. G., Lee, C., *Clin. Chem.* 1997, *43*, 352–359.

[23] Noetzel, M. J. J., *Neurosci. Res.* 1990, *27*, 184–192.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1092–1097

Luca Musante[1]
Massimo Ulivi[3]
Giovanna Cutrona[2]
Nicholas Chiorazzi[3]
Silvio Roncella[2,4]
Giovanni Candiano[1]
Manlio Ferrarini[2,5]

[1]Department of Nephrology, Giannina Gaslini Children Hospital, Genova, Italy
[2]Clinical Immunology Division, National Institute for Cancer Research, IST, Genova, Italy
[3]Department of Medicine, North Shore University Hospital and NYU School of Medicine, Manhasset, New York, NY, USA
[4]Division of Histology and Anatomical Pathology, St. Andrea Hospital, La Spezia, Italy
[5]Department of Clinical and Experimental Oncology, University of Genova, Italy

Identification of HSP-60 as the specific antigen of IgM produced by BRG-lymphoma cells

In previous studies we described a patient with Burkitt's lymphoma and AIDS, whose cells recognized a molecule expressed by normal and malignant breast cells. In the present study, we identified this antigen by two-dimensional (2-D) electrophoresis and Western blotting using the antibody produced by lymphoma cells. The antigen so identified consisted of two clusters of spots with a molecular mass (M_r) of 60 and 50 kDa, respectively. Preparative immobilized pH gradient (IPG) was subsequently used to isolate the clusters of spots of higher molecular masses, from which peptide fragments of approximately 10 aa were separated on reverse-phase chromatography and sequenced. This procedure enabled the identification of the antigen recognized by the lymphoma cells as HSP-60. By means of serological analyses it was possible to identify the lower molecular mass cluster of spots as a molecule related to HSP-60. It is hypothesized that this molecule is a membrane form of HSP-60 that differs from HSP-60 in a COOH terminal portion.

Keywords: Apoptosis / Burkitt's lymphoma / Heat shock protein / Two-dimensional gel electrophoresis
EL 3352

1 Introduction

In a previous study, we described a patient with Burkitt's lymphoma and AIDS (BRG) whose malignant cells (BRG-M cells) expressed surface IgM molecules with specificity for an autoantigen of the epithelial cells of the breast ducts [1, 2]. Hybridomas, obtained by fusing the patient malignant cells with the appropriate myeloma cell partner *in vitro*, secreted IgM monoclonal antibodies (BRG-M mAb) which stained ductal cells in normal breast sections by immunocytochemistry and a number of cell lines from breast malignancies *in vitro* by immunofluorescence and flow cytometry. Co-culture of BRG-M cells with the cells from the cell lines that expressed this antigen, recognized by BRG-M mAb, caused specific apoptosis of BRG-M cells. Apoptosis could be prevented by providing the appropriate T cell help to the malignant BRG-M cells. These findings suggested a model in which antigen or autoantigen stimulation contributed to the pathogenesis of lymphoproliferation by focusing the appropriate T cell help on the surface of the proliferating malignant cells.

In the present study, we isolated and characterized the antigen recognized by BRG-M mAb. For this purpose, 2-D electrophoresis, peptide sequencing and serological tests with mAb were used. Collectively, the data demonstrate that the antigen recognized by BRG-M cells is heat shock protein (HSP)-60. In addition, the BRG-M mAb binds to another molecule, of lower M_r, which shares some antigenic determinants located in proximity of the *N*-terminal with the HSP-60 molecules. This HSP-like molecule is expressed at the surface of breast cells and is responsible for the induction of apoptosis of BRG-M cells.

2 Materials and methods

2.1 Chemicals

Carrier ampholytes Resolyte pH 4–8, 3.5–10 and DTT were purchased from BDH (Poole, UK); IPG dry strips, pH 3.5–10 NL gradient, Ampholine pH 4–6, Immobiline p*K* 3.6, 4.6, 6.2, 7.0, 8.5, and 9.3, Gelbond film, acrylamide IEF grade, CHAPS, Tris, and glycine were from Pharmacia (Uppsala, Sweden); urea, thiourea, iodacetamide, silver nitrate and sodium thiosulfate were from Fluka (Buchs, Switzerland); DTE and SDS were from Sigma (St. Louis, MO, USA) and PDA, 1,4–bis(acryloyl)pipera-

Correspondence: Dr. Luca Musante, Dipartimento di Nefrologia, Istituto Giannina Gaslini, Largo Gerolamo Gaslini, n. 5, 16247 Genova GE, Italy
E-mail: labnefro@tin.it
Fax: +39-010-395214

Abbreviation: PI, propidium iodide

zine from Bio-Rad (Hercules, CA, USA). Polyclonal goat antihuman HSP-60 antibodies (Abs) (K19, N20) and phosphatase-labeled second antibody (donkey anti-goat Ig) were obtained from Santa Cruz Biotechnology (Santa Cruz, CA USA). All chemicals for electrophoresis, of the purest grade, were obtained from Merck (Darmstadt, Germany).

2.2 Rehydratation of analytical IPG gel strips

Hydration of Immobiline dry strips in nonlinear pH 3–10 was performed according to a modification of the method of Bjellqvist *et al.* [3], in which 0.5 M thiourea was added to increase the resolution of the spots [4].

2.3 Rehydratation of preparative IPG gel strips

Preparative IPG, pH 4.0–5.5, was performed according to the protocol of Righetti [5]; the total amount of Immobiline species was used at a concentration of 30 mM and the β value of 3 mequiv. L^{-1} pH^{-1}. The sample was prepared with the same protocol as the analytical method but with an increased amount of protein loaded per strip.

2.4 Sample preparation and extraction protocol

MDA-MB-468 were maintained in RPMI 1640 (Seromed – Biochrom, Berlin, Germany). MCF-7 cells were cultured in Dulbecco's modified Eagle medium (DMEM, Seromed). Media were supplemented with 10% fetal calf serum (FCS, Seromed). Both cell lines were obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). Cells, growing tightly adherent to the plastic surfaces, were washed in the flasks five times in an excess of phosphate-buffered saline, subsequently scraped from the flask in lysate buffer (10 mM HEPES, 100 mM KCl, 2.5 mM $MgCl_2$, 1 m4M $CaCl_2$, 300 mM sorbitol) containing 1% of Triton X-100; the lysates were spun and the pellet was discharged. All extraction steps were performed at 4°C in presence of antiproteolytic agents 3 mM PMSF, 50 μg/mL TPCK, 0.7 μg/mL pepstatin and 1 μg/mL aprotinin. Fifty μg of total proteins were applied to each strip of nonlinear IPG, pH 3.5–10, as described in detail by Hochstrasser *et al.* [6] for silver staining. One hundred μg of protein were used for immunoblot detection. Preparative IPG, pH 4.5–6.0, were loaded with 2 mg of membrane proteins extracted according to the protocol detailed below. Briefly, washed cellular pellets were lysed in hypotonic shock buffer (10 mM Tris, pH 7.4) and spun at 3000 × *g* for 20 min. After washing three times in hypotonic shock buffer, the pellet containing the antigen was treated with lysate buffer containing 0.1% β-exyl-glucopyranoside. This mixture was sedimented by centrifugation at 10 000 × *g* for 30 min. The resulting pellet contained more than 95% of the antigen recognized by BRG-M cells. Membrane proteins were solubilized with lysate buffer containing 1% of Triton X-100. Cellular debris and nuclei were discarded after centrifugation at 30 000 × *g* for 15 min at 10°C in a Beckman TL 100 (Palo Alto, CA, USA). The presence of the antigen in all steps of enrichment was monitored in a 2-D PAGE Western immunoblot utilizing BRG-M mAb [2]. Protein concentration was determined using a modified Bradford assay as described by Ramagli and Rodriguez [7].

2.5 Staining and Western blot

Staining of proteins was performed with methyl trichloroacetate negative staining [8], followed by silver staining [9]. Transblot of proteins to Hybond nitrocellulose membrane (Amersham, Little Chalfont, UK) was done in a Nova Blot semidry system (Pharmacia Biotech, Uppsala, Sweden) utilizing a continuous buffer system with 48 mM Tris, 39 mM glycine, 0.035% SDS, and 20% methanol. The transfer of proteins was achieved at 1.5 mA/cm^2 for 3.5 h. The membrane was probed with purified BRG-mAb, followed by rabbit anti-human-IgM (Dako, Glostrup, Denmark) and goat anti-rabbit conjugated with alkaline phosphatase (AP; Southern Biotechnology Associates, Birmingham, AL, USA). The membrane was also probed with goat polyclonal IgG antibodies against HSP-60, N20 Ab, and K19 Ab (Santa Cruz Biotechnology) followed by a donkey anti-goat AP (Santa Cruz Biotechnology). Positive spots were revealed with Sigma Fast tablets (Sigma Aldrich, Milano, Italy).

2.6 Sequence of peptide

The quantity of 2 mg of total protein was loaded on preparative IPG gel strip. Gels of the second dimension were electroeluted onto polyvinylidene difluoride (PVDF) membrane according to Matsudaira [10] and stained with Coomassie blue to visualize protein spots for excision. The reserved-phase separation of tryptic digestion *in situ* of each single spot was resolved and analyzed using a Procise 494 protein sequencer apparatus from Applied Biosystems (Weiterstadt, Germany), according to the method of Wilkins *et al.* [11].

2.7 Co-culture

BRG-M cell lines originated from Burkitt's lymphoma in an AIDS patient [1]. BRG-M cells were resuspended at a concentration of 3×10^5/mL in RPMI 1640 supplemented with 10% FCS and co-cultured with MDA-MB-468 at the

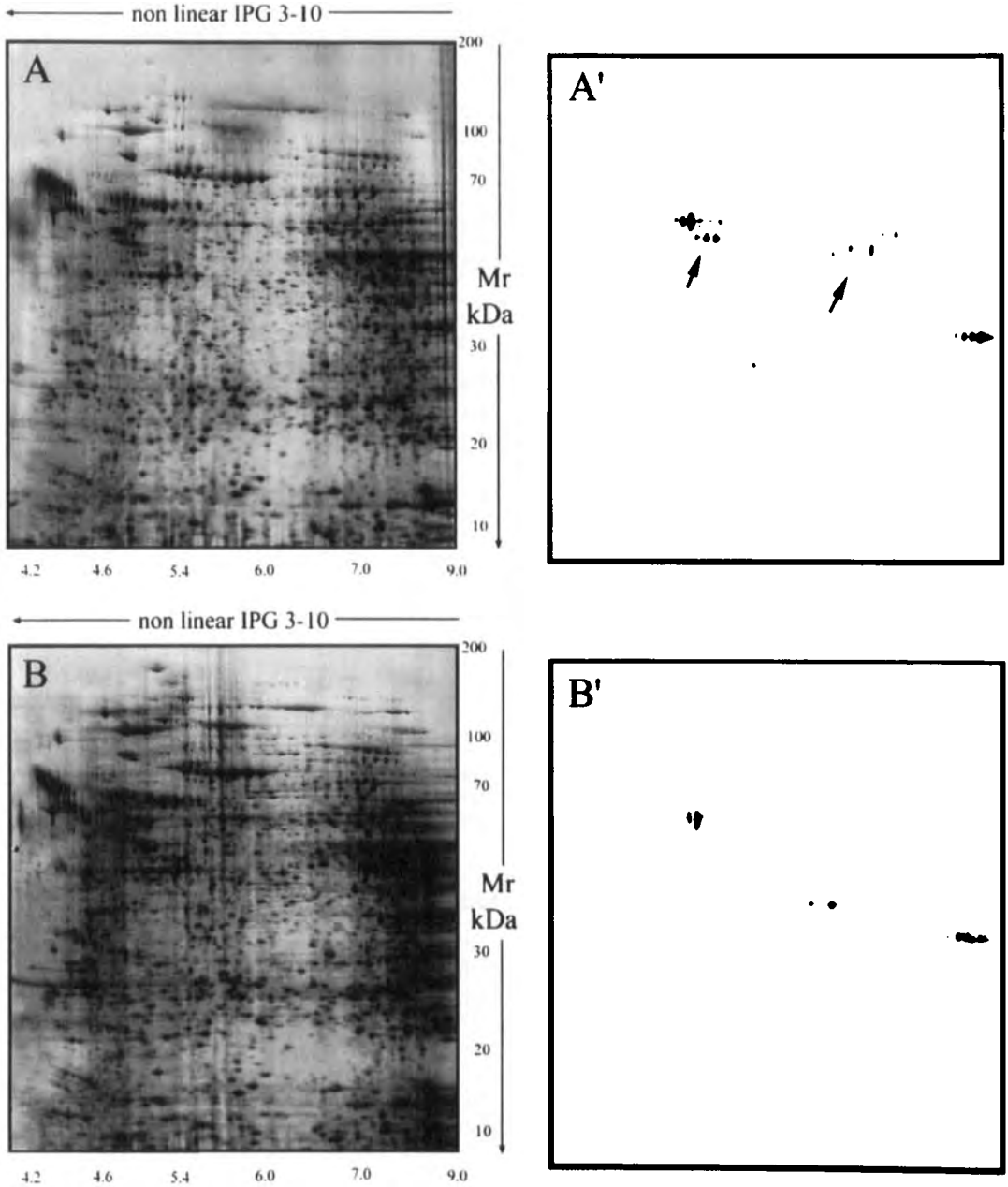


Figure 1. Computer-processed image of silver-stained electropherogram (A) of MDA-MB-468 and (B) of MCF-7. (A′) Immunoblotting of MDA-MB-468 cells and (B′) MCF-7 cells with BRG-M mAb.

optimal BRG-M/MDA-MB-468 ratio of 4:1 for 24 h [2]. To block the binding of surface IgM to the specific antigen, MDA-MB-468 cells were incubated with one of the goat anti-human HSP-60 Ab (K19 or N20) at a final concentration of 1 μg/5 × 10^5 cells for 30 min in the cold, before being exposed to BRG-M cells.

2.8 Propidium iodide (PI) staining

Apoptosis was measured by staining with a hypotonic propidium iodide fluorochrome solution (PI, Sigma Aldrich) and placed in the dark overnight at 4°C. The amount of DNA fragmentation was measured by flow cytometry (Becton Dickinson, Sunnyvale, CA, USA) [2].

3 Results and discussion

Cell lysates from MDA-MB-468 and from MCF-7 cell lines were electrophoresed in 2-D gels. The gels were subsequently silver-stained or blotted and identified with BRG-M mAb. The immunoblotting patterns obtained for each cell line in different experiments were compared. The spots of the proteins observed in both blots were discarded from the analysis and the spots consistently present in the MDA-MB-468 cells but not in the MCF-7 cells of the immunoblotting patterns were considered. This procedure yielded only two families of spots as putative candidates for molecules specifically reactive with BRG-M mAb (Fig. 1). Previous data demonstrated substantial differences between the immunoreactivity of MDA-MB-468 and MCF-7 with BRG-M mAb since MDA-MB-468 but not MCF-7 cells stained brightly with BRG-M mAb by surface immunofluorescence and flow cytometry. Moreover, in MDA-MB-468 cells there was an immunoreactive band of approximately 50 kDa by SDS-PAGE analysis which was consistently absent from MCF-7 cells [2]. Both lines, however, displayed another band which reacted with BRG-M mAb and had an apparent molecular mass (M_r) of 60 kDa as assesed by SDS-PAGE. Thus, the family of higher M_r spots shared by both cell lines probably represented this last polypeptide chain. In contrast, the spots seen in only MDA-MB-468 cells possibly represented the surface antigen detected by BRG-M mAb.

In subsequent experiments, preparative IPG was employed to isolate the spots characteristic of MDA-MB-468 cells starting from 2 mg of membrane protein-enriched preparation as depicted in Section 2.4. Seven identical IPG gels were run with the membrane protein-enriched preparation and one gel was employed to further assess the localization of the specific spots, whereas the spot-containing areas were excised from the remaining six gels, collected, and subjected to tryptic digestion. Peptide fragments were separated on reverse-phase chromatography and microsequenced. Three major peptide fragments of approximately 10 aa each were identified. These sequences were compared to those deposited in the EMBL database and identified as part of HSP-60. Figure 2 depicts the sequence of HSP-60; the fragments isolated by us (with 100% homology with the known fragment of the protein) are bold and underlined.

Next, we utilized commercially available antibody (Ab) against HSP-60 to further characterize the spots identified in this study. The epitopes recognized by these Abs formed one stretch of twenty amino acids located near the *N*-terminal (N20 Ab) and by another stretch of 19 aa (K19 Ab) located near the *C*-terminal, respectively. When employed in Western blot on IPG-separated proteins from MDA-MB-468 cells, it was clear that N20 Ab recognized both families of specific spots previously identified using BRG-M mAb (Fig. 3; also see Fig. 1 for comparison). In contrast, only the higher M_r family of spots was recognized by K19 Ab. The single family of spots present in MCF-7 cells was recognized by both N20 and K19 Abs (not shown). Collectively, these data indicate that BRG-M mAb recognized an epitope probably located somewhere near the *N*-terminal of the HSP-60 molecule. This epitope, like the one recognized by N20 Ab, was shared by the two families of peptides present in MDA-MB-468 cells.

Previously, we demonstrated that co-culture of MDA-MB-468 cells with BRG-M cells induced substantial apoptosis of the latter cells. This apoptotic process was initiated by the reaction between the specific antigen present on MDA-MB-468 cells and the surface IgM of BRG-M cells as documented by the inhibition of apoptosis caused by the addition of a monovalent Fab fragment to human IgM to the cultures [2]. Based upon these premises, we reasoned that N20 Ab could have caused inhibition of apoptosis of BRG-M cells seen in co-cultures with MDA-MB-

N20

```
MLRLPTVFRQ MRPVSRVLAP HLTRAYAKDV KFGADARALM LQGVDLLADA VAVTMGPKGR  60
TVIIEQSWGS PKVTKDGVTV AKSIDLKDKY KNIGAKLVQD VANNTNEEAG DGTTTATVLA 120
RSIAKEGFEK ISKGANPVEI RRGVMLAVDA VIAELKKQSK PVTTPEEIAQ VATISANGDK 180
EIGNIISDAM KKVGRKGVIT VKDGKTLNDE LEIIEGMKFD RGYISPYFIN TSKGQKCEFQ 240
DAYVLLSEKK ISSIQSIVPA LEIIEGMKFD RGYISPYFIN EALSTLVLNR LKVGLQVVAV 300
KAPGFGDNRK NQLKDMAIAT GGAVFGEEGL TLNLEDVQPH DLGKVGEVIV TKDDAMLLKG 360
KGDKAQIEKR IQEIIEQLDV TTSEYEKEKL NERLAKLSDG VAVLKVGGTS DVEVNEKKDR 420
VTDALNATRA AVEEGIVLGG GCALLRCIPA LDSLTPANED QKIGIEIIKR TLKIPAMTIA 480
KNAGVEGSLI VEKIMQSSSE VGYDAMAGDF VNMVEKGIID PTKVVRTALL DAAGVASLLT 540
TAEVVVTEIP KEEKDPGMGA MGGMGGGMGG GMF                              573
```

K19

Figure 2. Amino acid sequence alignment of human HSP-60. The three identified peptides of 10 aa each are indicated in bold and underlined. The two regions identified by commercial antibodies N20 and K19 are boxed. Human HSP-60 amino acid sequences were obtained from the EMBL Heidelberg Germany database.

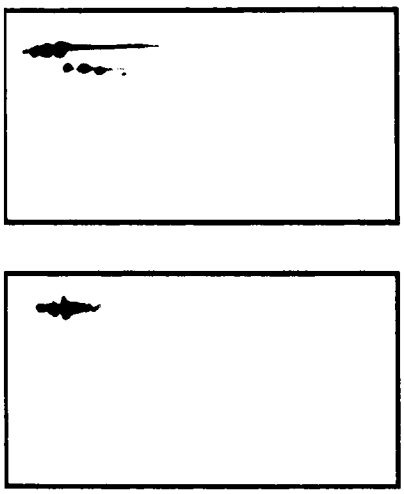


Figure 3. Immunodetection of HSP-60-like family in MDA-MB-468 cells by N20 antibody (top), or K19 antibody (bottom). K19 antibody bound the high molecular weight component only.

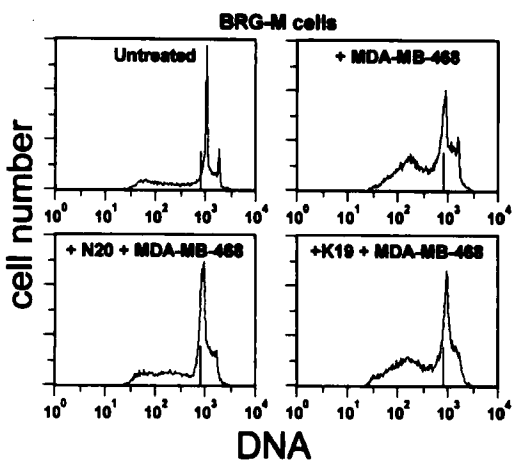


Figure 4. Inhibition of apoptosis of BRG-M cells by N20 antibody. MDA-MB-468 were incubated with N20 or K19 antibody (1 μg/5 × 10^5 cells) in the cold for 30 min and placed in culture. BRG-M cells were cultured alone, with MDA-MB-468 cells or with MDA-MB-468 cells pretreated with the anti-HSP-60 antibodies. After 24 h in culture the cells were stained with PI and analyzed for apoptosis by flow cytometry.

468 cells. Such inhibition may be related to the capacity of the Ab to interfere, by steric hindrance, with the reaction between surface IgM of BRG-M cells and its specific epitope on MDA-MB-468 cells. Indeed, N20 Ab (and not K19 Ab) caused a substantial inhibition of apoptosis of BRG-M cells *in vitro* (Fig. 4.).

4 Concluding remarks

Collectively, the above data demonstrate that IgM produced by BRG lymphoma cells recognize an epitope present on the HSP-60 molecules. This epitope is also expressed by an HSP-60-like molecule which shares a substantial portion of its *N*-terminal polypeptide sequence with HSP-60, as possibly also suggested by the cross-reactivity with N20 Ab. This HSP-60-like molecule is probably the one which is expressed on the cell surface of MDA-MB-468 cells and is partly responsible for the apoptotic process when co-cultured with BRG-M.

HSPs belong to a highly conserved family of protein that are normally expressed by the cells where they exert a variety of functions connected with intracytoplasmic transport and assembly of molecular complexes [12]. Expression of HSP may be increased following a number of events such as cell injury or activation of the cell cycle or of the apoptotic process. It is possible that, owing to their enhanced expression, HSPs become the target of an autoimmune response. Indeed, membrane expression of HSP has been observed following induction of apoptosis [13]. Moreover, production of autoantibodies has been described in pathological conditions characterized by tissue injury and apoptosis such as rheumatoid arthritis, diabetes mellitus or cardiomyopathy [14–16]. By inference, it is conceivable that both the extensive apoptosis that takes place in lymphoid tissues of patients with AIDS and the concomitant widespread tissue injury that occurs as a consequence of opportunistic infection may have induced autoantibody production to HSP-60 in the patient described here. Such an autoimmune process may have been facilitated by the general disregulation of the immune response that characterizes the AIDS process. Once the autoimmune process was initiated, it may have facilitated the expansion of one of the HSP-60 specific clones that either subsequently accumulated transforming mutations or had already been the subject of one or more transforming events prior to expansion.

This work was supported by grants from the Istituto Superiore di Sanità (ISS) (AIDS Project).

Received October 3, 1998

5 References

[1] Jain, R., Roncella, S., Hashimoto, S., Carbone, A., Francia Di Celle, P., Foà, R., Ferrarini, M., Chiorazzi, N., *J. Immunol.* 1994, *153*, 45–52.

[2] Roncella, S., Cutrona, G., Favre, A., Ulivi, M., Fais, F., Signorini, A., Grossi, C. E., Chiorazzi, N., Ferrarini, M., *Blood* 1996, *88*, 599–608.

[3] Bjelqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.

[4] Musante, L., Candiano, G., Ghiggeri, G. M., *J. Chromatogr. B* 1998, *705*, 351–356.

[5] Righetti, P. G., *Immobilized pH-Gradient: Theory and Methodology*, Elsevier Biomedical Press, Amsterdam 1990.

[6] Hochstrasser, D. F., Frutiger, S., Paquet, N., Bairoch, A., Ravier, F., Pasquali, C., Sanchez, J.-C., Tissot, J.-D., Bjellqvist, B., Vargas, J. R., Appel, R. D., Hughes, G. J., *Electrophoresis* 1992, *13*, 992–1001.

[7] Ramagli, L. S., Rodriguez, L. V., *Electrophoresis* 1985, *6*, 559–563.

[8] Candiano, G., Porotto, M., Lanciotti, M., Ghiggeri, G. M., *Anal. Biochem.* 1996, *243*, 245–248.

[9] Hochstrasser, D. F., Harrington, M. G., Hochstrasser, A. C., Miller, J. J., Merril, C. R., *Anal. Biochem.* 1998, *173*, 424–435.

[10] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[11] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, X., Gooley, A. A., Hughes, G. J., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *BioTechnology* 1996, *14*, 61–65.

[12] Craig, E. A., Weissman, J. S., Horwich, A. L., *Cell* 1994, *78*, 365–372.

[13] Poccia, F., Piselli, P., Vendetti, S., Bach, S., Amendola, A., Placido, R., Colizzi, V., *Immunology* 1996, *88*, 6–12.

[14] Portig, I., Pankuweit, S., Maisch, B., *J. Mol. Cell. Cardiol.* 1997, *29*, 2245–2251.

[15] Ozawa, Y., Kasuga, A., Nomaguchi, H., Maruyama, T., Kasatani, T., Shimada, A., Takei, I., Miyazaki, J.-I., Saruta, T., *J. Autoimmunity* 1996, *9*, 517–524.

[16] Krenn, V., Voilmers, H. P., von Landenberg, P., Schmaußer, B., Rupp, M., Roggenkamp, A., Müller-Hermelink, H. K., *Virchows Arch.* 1996, *427*, 511–518.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1098–1108

Paulo Costa[1]
Cédric Plonneau[1]
Guy Bauw[2]
Christian Dubos[1]
Nasser Bahrmann[1]
Antoine Kremer[1]
Jean-Marc Frigerio[1]
Christophe Plomion[1]

[1]INRA, Laboratoire de Génétique et Amélioration des arbres forestiers, Cestas, France
[2]Laboratory Genetics, University Gent, V.I.B. Flemish Interuniversity Institute for Biotechnology, Gent, Belgium

Separation and characterization of needle and xylem maritime pine proteins

Two-dimensional gel electrophoresis (2-DE) and image analysis are currently used for proteome analysis in maritime pine (*Pinus pinaster* Ait.). This study presents a database of expressed proteins extracted from needles and xylem, two important tissues for growth and wood formation. Electrophoresis was carried out by isoelectric focusing (IEF) in the first dimension and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in the second. Silver staining made it possible to detect an average of 900 and 600 spots on 2-DE gels from needles and xylem, respectively. A total of 28 xylem and 35 needle proteins were characterized by internal peptide microsequencing. Out of these 63 proteins, 57 (90%) could be identified based on amino acid similarity with known proteins, of which 24 (42%) have already been described in conifers. Overall comparison of both tissues indicated that 29% and 36% of the spots were specific to xylem and needles, respectively, while the other spots were of identical molecular weight and isoelectric point. The homology of spot location in 2-DE patterns was further validated by sequence analysis of proteins present in both tissues. A proteomic database of maritime pine is accessible on the internet (http://www.pierroton.inra.fr/genetics/2D/).

Keywords: *Pinus pinaster* / Database / Microsequencing / Protein / Two-dimensional polyacrylamide gel electrophoresis
EL 3306

1 Introduction

Two-dimensional polyacrylamide gel electrophoresis (2-DE), which separates polypeptides by their charge (p*I*) and relative molecular weight (M_r), is a powerful technique for analyzing complex mixtures of denatured proteins [1]. The main fields of application are: (i) the analysis of changes in protein expression, with the state of development and the tissue as well as environmental conditions as extracellular stimuli (*e.g.*, hormones, drugs, toxic agents), abiotic or biotic stresses [2], and (ii) the study of post-translational modifications. The analysis of the total protein complement in the genome has only recently become possible as a result of the reproducibility of the 2-D PAGE technology, advanced image analysis techniques allowing for 2-D gel comparison and protein quantification, and the routine identification of proteins excised from 2-D gels. This so-called "proteome" [3] analysis can nowadays be used for assigning a function to gene products and for providing physiological and biological explanations for differential protein expression. While proteomics is intensively studied in an increasing number of microorganisms, as well as in humans (see review, [4]), proteome analysis has not received as much attention in plants. The only plants that are studied at this molecular level are those for which large-scale gene sequencing projects, which facilitate protein identification, were begun (*i.e., Arabidopsis*, rice, and maize).

Here we report on the results of the first protein sequencing project for a forest tree species: maritime pine (*Pinus pinaster* Ait.). This conifer is native to southwestern Europe and is widely used in intensive monospecific stands that cover about four million hectars (mainly in France, Spain, Portugal, and Italy). Conifers are the largest group of industrial plantation species in the world. They include the largest (*Sequoiadendron giganteum*) and the oldest (*Pinus aristata*) living organisms on earth. Maritime pine is a diploid species and member of the large and important genus *Pinus* which includes more than 100 species of the Pinaceae family [5]. Recent genetic research on maritime pine have made use of proteins revealed by 2-D PAGE [6]. These include: (i) genetic mapping experiments [7–10]), (ii) population genetic studies [11–13], and (iii) the analysis of genome expression during abiotic stress [14]. Our objectives in the present study were threefold. First, to extend our knowledge in proteome analysis by characterizing proteins that accumulate in photosynthetic cells (needles) and in differentiating xylem. These tissues are important for growth and wood formation. Second, we set out to demonstrate that proteome re-

Correspondence: Dr. C. Plomion, INRA, Laboratoire de Génétique et Amélioration des arbres forestiers, BP45, 33610 Cestas, France
E-mail: plomion@pierroton.inra.fr
Fax: +33-557979088

Abbreviations: HSP, heat shock protein; **SAM**, *S*-adenosyl methionine

0173-0835/99/0405-1098 $17.50+.50/0

search can take place irrespective of the prior existence of DNA sequence information, and can help to identify gene products on the basis of similarity with other sequences in other plant or animal species. This is because no DNA sequence of maritime pine exists in public databases. And third, our aim was to construct a protein two-dimensional gel database of maritime pine. (These data can be accessed *via* the internet URL: http://www.pierroton.inra.fr/genetics/2D/).

2 Material and methods

2.1 Plant material

All plant material was taken from maritime pines in the Landes region. Fully expanded needles were sampled on the first shoot cycle of a two-year-old seedling. Needles were sampled once a six-week water stress was applied. Differentiating xylem was collected in July from a fast growing 22-year-old tree. Bark, phloem, and cambium were peeled from the stem, after which the differentiating xylem was removed with a knife. Following collection, the material was immediately frozen in liquid nitrogen and stored at –80 °C.

2.2 Analytical and preparative 2-D electrophoresis

Protein extraction, isoelectric focusing (IEF), and SDS-PAGE dimensions were performed as described in [6]. For analytical 2-D electrophoresis, the first dimension used acrylamide, purchased from Serva (Heidelberg, Germany) and Bis from Bio-Rad (Richmond, CA, USA), while the second dimension used acryl/Bis from Interchim (Montluçon, France). For the preparative 2-D electrophoresis, acrylamide was purchased from Serva and Bis from Bio-Rad for both the first and second dimension. *N,N,N′,N′*-Tetramethylethylenediamine (TEMED) and dithiothreitol (DTT) were purchased from Sigma (St. Louis, MO, USA), ammonium persulfate (APS) was from Promega (Madison, WI, USA), glycine and urea were from Merck (Darmstadt, Germany), and sodium dodecyl sulfate (SDS) was from Serva. Silver staining was performed according to Damerval *et al.* [15] with minor modifications as described by Costa *et al.* [14]. Dried gels of needle and xylem were visually compared on an illuminated box by superimposition as described by Bahrman and Petit [11]. In addition, coelectrophoresis was run to verify that spots of similar molecular weight and isoelectric point present in needle and xylem were actually identical. Tissue-specific percentages were deduced, after which the gels were compared by two independent readers.

Preparative 2-DE conditions were similar to the analytical 2-D electrophoresis procedure, with only few modifications: (i) A mix of 400 μL proteins, and urea-potassium carbonate-SDS (UKS) solution [16] (30 μL/mg of pellet) was loaded onto isoelectro focusing gels (3 mm in diameter); (ii) the 2-DE gels were 1.5 mm thick, with stacking gels; (iii) the gels were stained using the Amido Black procedure (according to [17]) prior to *in situ* digestion.

2.3 Internal amino acid sequence analysis

Microsequencing is a straightforward method for identifying the putative function of a protein. Since the *N*-terminal amino acid can be blocked for up to 70% of proteins [18], the characterization of spots either requires removing the blocking group or the internal microsequences to be performed. In our study, this problem was avoided by sequencing internal stretches of the protein. Two methods were used to obtain microsequences. In method 1 (microsequences obtained by the Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris, France), gel *in situ* protease cleavage and peptide sequencing were performed as described [19]. In method 2, gel pieces (or their parts) were lyophilized and treated as described [20, 21]. For both methods, an internal amino acid sequence was obtained by pooling 10–40 spots to reach a total of at least 100 pmol. Amino acid sequence determination was performed with a 473 Applied Biosystems protein sequencer. Microsequenced spots were named according to the first letter of the tissue from which they were collected: "N" for needle and "X" for xylem, followed by a spot ID within each tissue.

2.4 Electrospray ionization MS/MS

The protein band of interest was excised from the polyacrylamide gel and sliced in small pieces. These were rinsed several times with a solution of 0.1 M Tris, pH 8.5, to adjust the pH and were lyophilized in a Speedvac. To the dried gel pieces 2 μL of a 1 mg/mL trypsin solution, 50 μL of 0.1 M Tris, pH 8.5, and 5 μL of 0.1 M $CaCl_2$ were added. The swollen gel pieces were completely covered with the NH_4HCO_3 solution (approximately 0.1–0.15 mL in total) and incubated overnight at 37°C. To stop the digest and to extract the polypeptides from the gel pieces, 200 μL of 20% acetic acid was added. After incubating for 1 h the solution was transferred to another Eppendorf tube and the extraction of the gel pieces was repeated twice, once with 20% acetic acid, once with water. All extraction solutions were combined and a small amount of pre-wetted Poros R2 (Perseptive Biosystems, Framingham, MA, USA) was added. After 30 min of incubation the Poros beads were spun down (3 min 10 000 × *g*) and the

liquid taken off. The Poros beads were immediately washed with 50 μL of 1% acetic acid. At this point the samples were stored at –20 °C until analysis. After centrifugation and removal of the washing solution, Poros-bound material was eluted by addition of 5 μL of 50% methanol. One to 2 μL of the eluent was applied in a nanospray needle after incubating and centrifuging the beads for 30 min. Mass spectrometric analysis was performed on an Ion Trap Mass Spectrometer (LC-Q Finnigan, San Jose, CA, USA) equipped with a nanospray device.

2.5 Comparison of amino acid sequences to protein and DNA databases

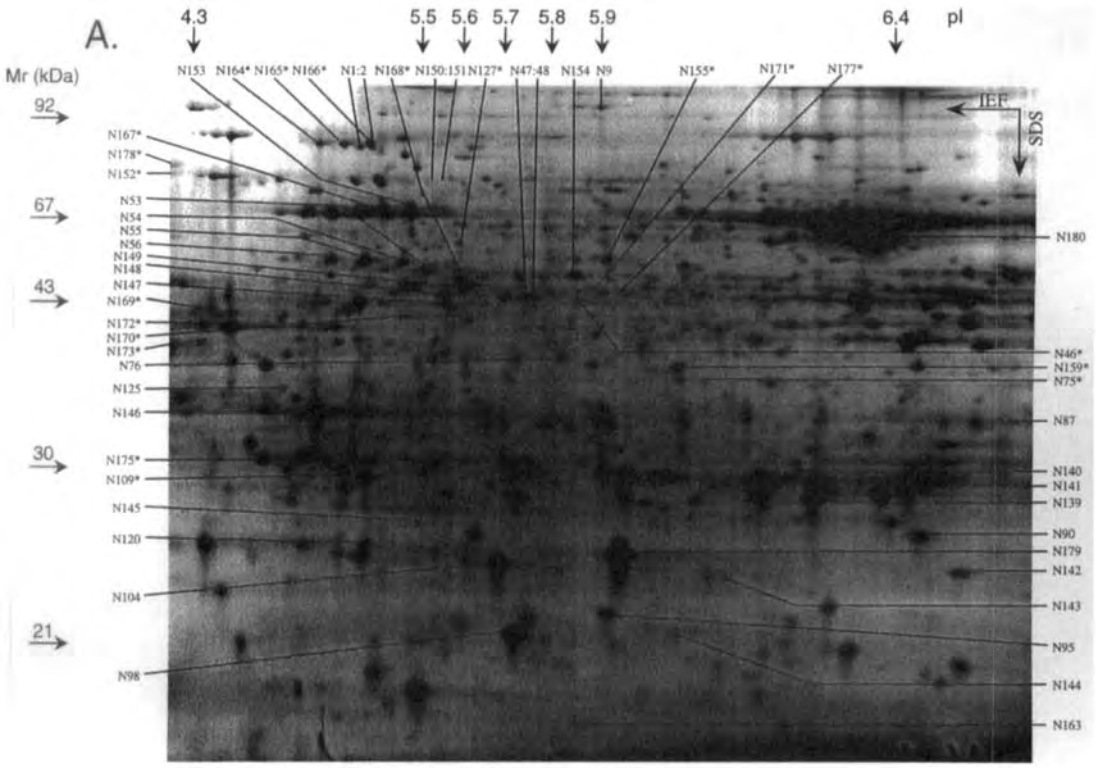
The amino acid sequences were compared to the sequences of five databases: (i) the OWL Database (release 30.2, containing 250514 entries and indexed June 11, 1998), a nonredundant composite of four publicity available primary sources: SWISS-PROT, PIR, GenBank (translation), and NRL-3D Brookhaven; (ii) the Nonredundant Nucleotide Sequence Databases (2501073) sequences nucleotide) containing all nonredundant GenBank, EMBL, DDBJ, PDB sequences; and (iii) the dbSTS, a database of Sequence Tagged Sites developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (release 98.07, containing 54841 entries and indexed July 19, 1998); (iv) dbEST, a database from NCBI containing detailed information and database search results for all publicly available EST sequences (release 98.07, containing 1718327 entries and indexed July 23, 1998); and (v) the *Pinus taeda* EST database (http://www.cbc.umn.edu/ResearchProjects/Pine/). The BLAST and FASTA programs [22, 23] were run using the default parameters proposed by the NCBI. When a similarity was found with ESTs or STSs, the corresponding sequence was retrieved and compared to the above database using the BLAST program.

3 Results and discussion

3.1 Comparison of needle and xylem 2-D gels

Maritime pine proteins were isolated from two tissues: needle and xylem, from which about 900 and 600 protein spots, respectively, were resolved on a two-dimensional gel (20 × 24 cm) with silver staining. Most polypeptides weighed between 20 and 100 kDa. Examples of 2-DE gels are shown in Fig. 1. Comparison of needle and xylem 2-DE patterns indicated that 29.5 and 36.3% of the spots were specific to xylem and needles, respectively. The other spots had identical M_r and p*I*, though they often differed in staining intensity. Comparison of protein accumulation from spots common to both tissues was not attempted since needle and xylem were sampled from different individuals in different experimental conditions. The genetic effect on spot intensity has been extensively described [24–29]. In four cases, the homology of spot location in 2-DE patterns was validated by comparing HPLC profiles and sequence analyzing the spots present in both tissues: N53 and X9 (actin), N125 and X46 (caffeoyl CoA *O*-methyltransferase), N98 and X47 (low molecular weight heat shock proteins, HSP), X1 and N166 (HSP70). Two other interesting cases of homology may be deduced from concordant genetic arguments: two xylem polypeptides, X16 and X17, with similar HPLC profiles and identical internal sequences, showed 100% similarity with glutamine synthetase (GS). Both spots were close together and could easily be visually identified in needle 2-D gels. In this photosynthetic tissue, GS spots showed positional polymorphism (monogenic and codominant mode of inheritance) in a segregating F2 family. This allelic variation allowed us to localize the needle protein in the linkage map of maritime pine. This protein marker was found to be colocated with the GS gene previously mapped using a PCR-based approach [30], demonstrating that both needle spots do correspond to two allelic products of GS. Such comparative analysis was performed in rice between embryo and endosperm [18], and extended between rice and *Arabidopsis* [31], where identical proteins presenting similar M_r and p*I* were reported.

All these examples seem to prove that needle and xylem protein spots characterized by identical M_r and p*I* are likely to have the same function. More data will be necessary for a definitive confirmation of this hypothesis. If true, pooling sequence information from several tissues (*e.g.*, root, needle, xylem, phloem, bud, pollen, megagametophyte), where common proteins often accumulate differentially [11], could partially solve the limitation caused by low resolution in preparative gels stained with Amido Black. The detection threshold of this technique is not as sensitive as that obtained with prolonged silver staining or radiolabeling generally used for staining analytical 2-D gels. Compared to these other techniques, only one-fifth of the spots can be readily detected with Amido Black staining. A similar ratio has been noted in rice [18]. This limitation could, however, be solved by the use of Coomassie Blue R-250, a more sensitive staining than Amido Black, compatible with protein sequencing (*e.g.*, [32]). The finding that most spots were common to xylem and needle confirmed the results of comparing needle, bud, and pollen proteins cited in Bahrman and Petit [11]. These authors interpreted the presence of the same proteins in well-differentiatied organs as "house-keeping proteins" (*e.g.*, actin HSP70, glutamine synthetase). Conversely, proteins accumulating only in particular tissues are most



likely to be present in specialized cells and to have specific functions, such as the needle proteins involved in photosynthesis.

3.2 Sequence similarity

Figure 1 shows the proteins analyzed on 2-D gels of needle and xylem. Table 1 reports the amino acid sequences and results of the similarity search. The position of microsequenced proteins on both needle and xylem silver-stained 2-D gels can be viewed on the internet (http://www.pierroton.inra.fr/genetics/2D/). A total of 87 amino acid sequences, comprising between 4 and 20 amino acids, were obtained for 28 xylem and 31 needle spots. Three other needle spots (N2, N47, and N151) were allelic forms [10] of microsequenced proteins (indicated in the first column of Table 1), and an additional needle spot (N180) could be easily assigned to the large subunit of rubisco (*rbc*L) based on its M_r and p*I* values, and on its extremely high expression in this photosynthetic tissue. Out of these 63 proteins, 57 (90.5%) were identified by amino acid similarity with known proteins in databases. This number also included the three allelic spots and *rbc*L. This proportion is greater than that obtained in similar projects for rice (58% [18], 53% [31], and 56% [33]), maize (48.5% [19]), and *Arabidopsis* (56.5% [31], 50% [32], and 63% [34]). This percentage may also be contrasted with the approximate 20–30% identification rate obtained in random sequencing cDNA projects (see review [35]).

Among the identified proteins, 25 (43.8%) were previously reported in conifers (indicated in the penultimate column of Table 1), 31 (54.3%) were identified as showing strong homology with proteins sequenced in other plant species, and one protein (X54) showed homology with a protein described in a microorganism. These results indicate that sequence similarities among proteins are common among widely distant plants (angiosperms *vs.* gymnosperms). Although the small stretches of determined amino acids can reveal high conservation, the result may also be different at the level of the complete sequence of the protein. Overall, there were small discrepancies between the apparent molecular masses M_r and the calculated mass. Finally, for six proteins (9.5%) no function could be assigned. The amino acid sequences obtained for these proteins may represent nonconserved regions of previously characterized proteins or new gene products. Identifying protein

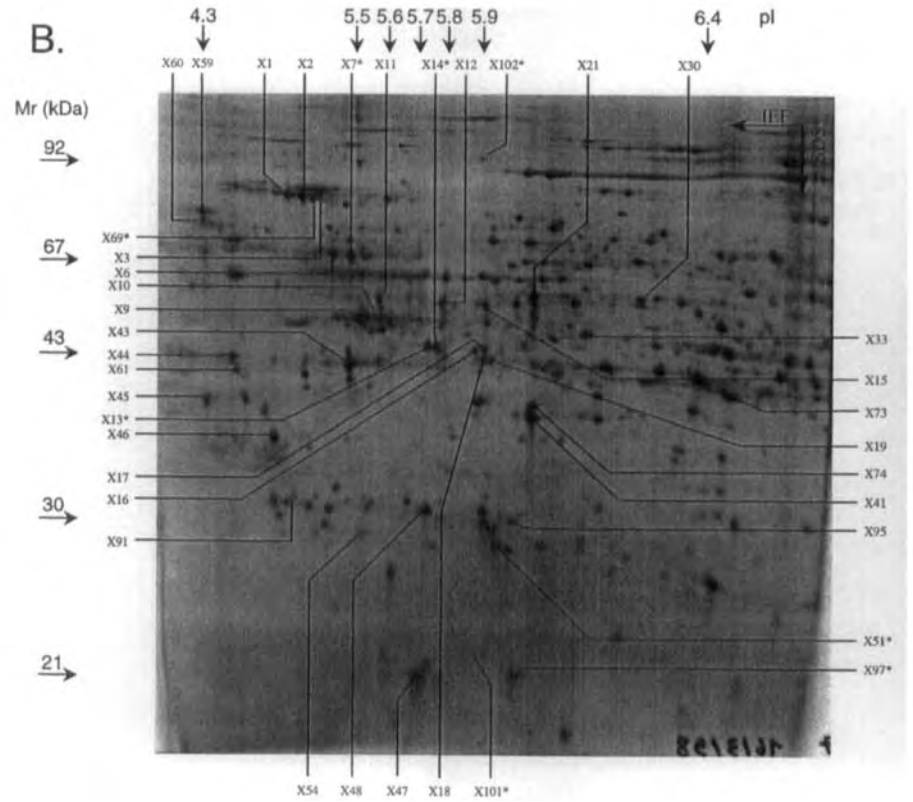


Figure 1. Needle and xylem proteins separated by 2-D PAGE and detected by silver staining. (A) Proteins extracted from needle. Spot denoted by an asterisk indicates that the spot was microsequenced from xylem proteins and indentified in a needle 2-D gel by comigration analysis. (B) Proteins extracted from xylem. Spot denoted by an asterisk indicates that the spot was microsequenced from needle proteins and identified in a xylem 2-D gel by comigration analysis.

function allowed us to group xylem and needle protein by metabolic pathways. These groups are described below. At this stage it must be pointed out that intensity variation of microsequenced spots is being studied. Costa *et al.* [14] and Costa *et al.* (submitted) have presented needle proteins responding to a progressive drought stress and Plomion *et al.* (in preparation) have described xylem proteins responding to gravitropism.

3.2.1 Proteins involved in carbon fixation

N148 corresponded to phosphoribulokinase, a key enzyme of the Calvin cycle found in the chloroplast of higher plants. N90, N87, N163, and N104 were found to be identical to ribulose bisphosphate carboxylase/oxygenase large chain precursor (*rbc*L), a key enzyme in carbon assimilation. Their M_r and p*I* values varied greatly and were largely different from the value reported for this enzyme (molecular mass 52.7 kDa). In fact, these spots have been shown previously to be the result of rubisco proteolysis in drought stress conditions [14]. The 52 kDa subunit of rubisco actually corresponds to spot N180. The sequences of N54, N149, N56, and N55 were identified as rubisco activase, a plastid enzyme that activates rubisco. These spots were also characterized by different M_r and p*I* and may be different proteins encoded by a multigene family (as previously reported [33]) or else proteolytic products of the same protein. Spots N146 and N179 were attributed to the 33 kDa and 23 kDa subunits of the oxygen evolving complex, respectively. This protein complex is located in the chloroplast thylakoid membrane, associated with the photosystem II complex. N139 was found to be similar to triose phosphate isomerase, an enzyme involved in glycolysis. X74 was similar to malate dehydrogenase. X6 was similar to uridine diphosphate (UDP)-glucose pyrophosphorylase, a cytosolic soluble protein responsible for the synthesis and pyrophosphorolysis of UDP-glucose.

Table 1. List of microsequenced maritime pine proteins (p*I* are indicated approximately) and identity scores

Spot ID[a] (allelic spot)	*Mr* (kDa)	p*I*	Method[b]	Tissue	Tissue specificity[c]	Internal sequence	Similar protein	Identity	Similarity	EC	Metabolic pathway	Accession of similar protein in OWL or EMBL[d]	Species	Calculated *M*, (kDa)
N148	37	5.6	MS	Needle	S	ANNFDLM*YEQVK	Phosphoribulokinase	11/12	12/12	2.7.1.19	Carbon fixation	P25697	*Arabidopsis thaliana*	47.395
			MS			FYGEVTQQM*LK	Phosphoribulokinase	11/11	11/11		Calvin cycle	P25697	*Arabidopsis thaliana*	
			MS			VSVLEM*DGQFDR	Phosphoribulokinase	12/12	12/12			P25697	*Arabidopsis thaliana*	
N90	22.0	6.5	1	Needle	S	DTDILAAFR	Rubisco LSU	9/9	9/9	4.1.1.39	Carbon fixation	P48711*	*Picea abies*	52.724
			1			VTPQPGVPPEEAGAA	Rubisco LSU	15/15	15/15			P48711*	*Picea abies*	
N87	28	6.7	1		S	WSPELAAACEIWK	Rubisco LSU	13/13	13/13			P48711*	*Picea abies*	
N163	15	6.0	1		S	DVTLGFVDLLR	Rubisco LSU	11/11	11/11			P26959*	*Cedrus deodora*	52.599
N104	21	5.7	2		S	VTPQPGVPPEEAGAA	Rubisco LSU	15/15	15/15			P48711*	*Picea abies*	
			2			DTDILAAFR	Rubisco LSU	9/9	9/9			P48711*	*Picea abies*	
N54	42	5.5	1	Needle	S	SFQCELVFAK	Rubisco activase	10/10	10/10		Carbon fixation	A23703	*Hordeum vulgare*	45.710
N149	38	5.6	2		S	GLAYDESDDQQDIR	Rubisco activase	12/14	12/14			A23703	*Hordeum vulgare*	
			2			YVSEAALGD	Rubisco activase	8/9	8/9			A23703	*Hordeum vulgare*	
N56	43	5.3	2		S	FY-AP	Rubisco activase	4/4	4/4			A23703	*Hordeum vulgare*	
			2			PLILGI	Rubisco activase	6/6	6/6			A23703	*Hordeum vulgare*	
N55	43	5.4	2		S	FYSAPTR	Rubisco activase	6/7	6/7			A23703	*Hordeum vulgare*	
			2			DNVFEL	*							
N146	29	5.3	2	Needle	S	GGSTGYDNAVALPAGGR	33 kD subunit of oxygen-evolving complex	17/17	17/17		Photosynthesis	P23321	*Arabidopsis thaliana*	33.000
			2			LTYDEIQSK	33 kD subunit of oxygen-evolving complex	9/9	9/9			P23321	*Arabidopsis thaliana*	
N179	22	5.9	MS		S	AYGEAANVFGAPK	23 kD subunit of oxygen-evolving complex	13/13	13/13		Photosynthesis	A60484*	*Pinus monticola (fragment)*	23.000
N139	24	5.9	2	Needle	C(X51)	TFDVCYEQL	Triose phosphate isomerase	8/9	9/9	5.3.1.1	Carbon fixation	P48496	*Spinacia oleracea*	34.409
			2			FFVGGNWK	Triose phosphate isomerase	8/8	8/8		Calvin cycle	P48496	*Spinacia oleracea*	
N145	23	5.6	1	Needle	S	YKDQGLEILAFPCNQFG	Glutathione peroxidase	17/17	17/17	1.11.1.9	Glutathione	P30708	*Nicotiana sylvestris*	25.986
N142	20	6.7	1		S	YKDQGLEILAFPCNQFG	Glutathione peroxidase	17/17	17/17		Metabolism	P30708	*Nicotiana sylvestris*	
N95	18	5.9	1	Needle	C(X101)	HAGDLGNIVAGSDGV	Cu/Zn superoxide dismutase	15/15	15/15	1.15.1.1		P24707*	*Pinus sylvestris (fragment)*	22.291
N125	30	5.2	1	Needle	C(X46)	VGGLIAYDN	Caffeoyl CoA-*O*-methyltransferase	8/9	9/9	2.1.1.104	Lignin biosynthesis	Z82982	*Nicotiana tabacum*	27.767
						IEISQIPVGDGVTLC	Caffeoyl CoA-*O*-methyltransferase	12/15	14/15			Z82982	*Nicotiana tabacum*	
N76	32	5.8	1	Needle	S	VNIGVNYGMLGNSYP	Glucan endo-1,3-beta glucosidase	10/15	12/15	3.2.1.39	Starch, sucrose and	P23431	*Nicotiana plumbaginifolia*	39.939
						TGPETERHFGLNPD	Glucan endo-1,3-beta glucosidase	12/14	13/14		cellulose	S82315	*Zea mays*	34.701
						TLHVVRHAQGFHNLA	*Arabidopsis* EST (AC002329)	11/15	15/15		metabolism			
N53	41	5.5	1	Needle	C(X9)	DLYGNIVLSGGSTMF	Actin	15/15	15/15			P24902*	*Pinus contorta* (fragment)	41.749
N1 (N2)	73	5.4	1	Needle	C(X69)	XGESWETPETGDEVE	Peptidylprolyl *cis-trans* isomerase	11/15	14/15	5.2.1.8		S55383	*Triticum aestivum*	61.876
N48 (N47)	36	5.7	1	Needle	C(X14)	DVNWPLGWPVGGYPG	Glutamine synthetase	15/15	15/15	6.3.1.2	Nitrogen metabolism	P52783*	*Pinus sylvestris*	39.526
N140	25	6.2	2	Needle	S	V/SSFNL+YVLTPAETV/L	*							
			2			AYGNLPFDQQLVLCE	*							
			2			IIT/GH/GPNF	*							
			MS			VENI/LVI/LGH	Carbonic anhydrase	8/8	8/8	4.2.1.1	Nitrogen	P40880	*Hordeum vulgare*	35.074
			MS			M*EVVACADSR	Carbonic anhydrase	8/10	8/10		metabolism	P40880	*Hordeum vulgare*	
N153	55	5.5	1	Needle	C(X7)	ESIASFQGVLDGKYD	ATP synthase beta-chain	14/15	15/15	3.6.1.34	Oxidative phosphorylation	F29685	*Hevea brasiliensis*	60.190

Table 1. continued

Spot ID[a] (allelic spot)	*Mr* (kDa)	*pI*	Method[b]	Tissue	Tissue specificity[c]	Internal sequence:	Similar protein	Identity	Similarity	EC	Metabolic pathway	Accession of similar protein in OWL or EMBL[d]	Species	Calculated M_r (kDa)
N120	22	5.4	1	Needle	S	SDTTAGSYAEALAEL	*Pinus taeda* EST (H75181) or	13/15	14/15					
							ATP synthase delta chain	10/15	13/15	3.6.1.34	Oxidative phosphorylation	P02758	*Pisum sativum*	27.592
						LTIDPSLIVT	*							
						TVDTQLAVAG	*							
N9	92	5.9	1	Needle	C(X102)	LIGSPPGYVGYTEGG	ATP-dependent protease	15/15	15/15			P35100	*Pisum sativum*	102.593
						AESEAGDASPLVTEV	ATP-dependent protease	12/15	15/15			P31541	*Lycopersicon esculentum*	102.439
N166	73	5.4	MS	Needle	C(X1)	VEIIANDQGNR	HSP70	11/11	11/11			U41652	*Sorghum bicolor*	70.000
N144	17	6.1	1	Needle	C(X97)	YFDPFSLDVWDPFQA	Low molecular weight HSP (class I)	13/15	14/15			S71768*	*Pseudotsuga menziesii*	18.171
N98	18	5.7	1	Needle	C(X47)	SFDPFSLDVWDPFQA	Low molecular weight HSP (class I)	13/15	15/15			S71768*	*Pseudotsuga menziesii*	18171
N143	21	6.2	1	Needle	S	KKIEEEAAAAGATAE	*Pinus taeda* EST (PTU59451)	12/15	14/15			U59451*	*Pinus taeda* (fragment)	*
							water stress-inducible protein Ip3-2							
							Prunus armeniaca cDNA (U93164)	9/15	13/15			U93164	*Prunus armeniaca*	21.218
							abscisic stress ripening protein							
N147	36	5.6	2	Needle	S	IIYD-VGQCDK	*							
						DF/IY/ER/KPELPR								
N141	25	6.6	2	Needle	S	LYGNLPF	*							
N150 (N151)	60	5.5	1	Needle	S	HEEQITQPSATNDEA	*							
X1	75	5.3	1	Xylem	C(N164)	SINPDEAVAYGA	HSP70	12/12	12/12			Y17053	*Arabidopsis thaliana*	71.067
						EQVFSTYSDNQP	HSP70	12/12	12/12			Y17053	*Arabidopsis thaliana*	
			1			VSAIR	*							
X2	75	5.3	1		C(N165)	SINPDEAVAYGAAVQ	HSP70	15/15	15/15			Y17053	*Arabidopsis thaliana*	
						EQVFSTYSDNQPGVL	HSP70	15/15	15/15			Y17053	*Arabidopsis thaliana*	
X3	75	5.4			C(N166)	TAGDTHLGGEDFDN	HSP70	14/14	14/14			Y17053	*Arabidopsis thaliana*	
X47	18	5.7	1	Xylem	C(N98)	ASMENGV	Low molecular weight HSP (class I)	7/7	7/7			P19036	*Arabidopsis thaliana*	17.454
X61	33	5.1	1	Xylem	C(N172)	HKIEEEVAAAAAVGG	*Pinus taeda* cDNA (PTU52865)	14/15	15/15			U52865*	*Pinus taeda*	14.361
							water stress-inducible protein:Ip3-1							
							Lycopersicon esculentum cDNA (Q08655)	13/15	15/15			Q08655	*Lycopersicon esculentum*	13.115
							abscisic stress ripening protein 1							
X6	56	5.4	1	Xylem	C(N167)	ATSDLLLVQSDLYTVEEGFV	UDP-glucose pyrophosphorylase	17/20	20/20	2.7.7.9	Starch and sucrose metabolism	AB013353	*Pyrus pyrifolia*	51.585
X9	41	5.5	1	Xylem	C(N53)	DLYGN	Actin	5/5	5/5			P24902*	*Pinus contorta* (fragment)	41.749
X10	38	5.6	1	Xylem	C(N168)	DLGERYGVSGFPTLK	Disulfide isomerase	14/15	15/15	5.3.4.1	Biosynthesis and degradation of glycoprotein	Y11209	*Nicotiana tabacum*	39.663
X59	61	4.9	1		C(N178)	NHDPPIVLAK	Disulfide isomerase	9/10	9/10			P52588	*Zea mays*	57.032
X11	44	5.6	1	Xylem	C(N127)	YLDENTIFH	SAM synthetase	9/9	9/9	2.5.1.6	Activated methyl cycle	P50300*	*Pinus banksiana*	43.119
			1			TNMVMVFGE	SAM synthetase	9/9	9/9			P50300*	*Pinus banksiana*	
X15	45	5.9			C(N155)	VLVNIEQQSPDIAQG	SAM synthetase	14/15	14/15			P50300*	*Pinus banksiana*	
X30	46	6.3	1		S	RPEEIGAGDQGHMFG	SAM synthetase	15/15	15/15			P50300*	*Pinus banksiana*	
						NFDFRPGMMTINLDL	SAM synthetase	15/15	15/15			P50300*	*Pinus banksiana*	
X12	51	5.8	1		S	VLVNIEQQSPDIAQG	SAM synthetase	14/15	14/15			P50300*	*Pinus banksiana*	
X21	45	6.1	1		S	LCDQISDAVLDACLT	SAM synthetase	14/15	14/15			P50300*	*Pinus banksiana*	
X16	37	5.9	1	Xylem	C(N46)	IIAEYIWIGGSGMDI	glutamine synthetase	13/15	15/15	6.3.1.2	Nitrogen	X74429*	*Pinus sylvestris*	39.480

Table 1. continued

Spot ID[a] (allelic spot)	*Mr* (kDa)	*pI*	Method[b]	Tissue	Tissue specificity[c]	Internal sequence:	Similar protein	Identity	Similarity	EC	Metabolic pathway	Accession of similar protein in OWL or EMBL[d]	Species	Calculated M_r (kDa)
X17	36	5.9	1		C(N171)	IIAEYIWIGGSGMDI	glutamine synthetase	13/15	15/15		metabolism	X74429*	*Pinus sylvestris*	
X18	35	5.9	1	Xylem	C(N177)	HVGGDMFDSVPSGQA	caffeic *O*-methyltransferase	11/15	15/15	2.1.1.68	Lignin biosynthesis	U70873*	*Pinus radiata*	41.816
X46	30	5.2	1	Xylem	C(N125)	VGGLI	caffeoyl CoA *O*-methyltransferase	5/5	5/5	2.1.1.104	Lignin biosynthesis	Z82982	*Nicotiana tabacum*	27.767
X33	38	6.2	1	Xylem	S	LVDEHHPCLYRQFIH	ACC oxydase	11/15	14/15		Ethylene	L42466*	*Picea glauca*	34.134
						QEMERLSTACQEGGF	ACC oxydase	7/15	12/15		biosynthesis	L42466*	*Picea glauca*	
X41	32	6.1	1	Xylem	C(N75)	LPANTLSFNELVALW	Isoflavone reductase	12/15	14/15	1.3.1.	Isoflavonoid metabolism	AF071477	*Pyrus communis*	35.000
X45	31	5.0	1	Xylem	C(N173)	TWVWHATDFSDGELK	*ran*-binding protein	12/15	13/15			U62742	*Arabidopsis thaliana*	24.694
X48	26	5.8	1	Xylem	C(N175)	DVFGHMGLNDKEIVA	Ascorbate peroxidase	13/15	14/15	1.11.1.11	Ascorbate and aldarate metabolism	X80036	*Arabidopsis thaliana*	27.430
X74	32	6.1	1	Xylem	C(N159)	LIDVDVPVVGGHAGI	Malate deshydrogenase	15/15	15/15	1.1.1.37	Carbon assimilation Krebs cycle	AF020273	*Medicago sativa*	43.133
X43	35	5.5	1	Xylem	C(N170)	EYGSIDVLEFGEFPV	*Pinus taeda* EST (AA556564 or AA556410) auxin-induced protein	12/15	14/15			AA556564*	*Pinus taeda*	*
X60	59	4.9	1	Xylem	C(N152)	EYFDGTLPQFRRSE	*							
X54	24	5.7	1	Xylem	C(N109)	GLAIAAISSNAVTH	*							
						QALDCVLSDQPLNFA	*							
						AYGAVCTPEFYLFK	*ficA* gene (transcription factor)	10/14	13/14			Y12363	*Azospirillum brasilense*	*
X19	34	5.9	1	Xylem	S	AVVESLPENGLLVGG	*							
X44	34	5.0	1	Xylem	C(N169)	VSEEEVGFLTGGADP	*							

a) Spot ID refer to Fig. 1.
b) See Section 2.
c) Indicates whether a spot is common (C) or specific (S) to needle (N) and xylem (X).
d) Accession from coniferous species are indicated with an asterisk.
– indicates that no unambiguous amino acid could be assigned.
+ between two amino acids indicates that they are present on this position in more or less the same amounts. We presume that there is microheterogeneity in the sequenced protein.
/ indicates that two amino acids are present, probably due to the presence of two different microsequences.
M* indicates oxidized methionine.

3.2.2 Water stress-inducible proteins

N143 and X61 corresponded to two members of a water stress-inducible *Pinus taeda* gene family [36], showing strong homology within a tomato protein TMA SN1 (Q08655), reported to be abscisic acid (ABA)-inducible and expressed during fruit ripening. N144, N98, and X47 corresponded to low molecular weight heat shock proteins (HSP, class I). The small HSP are large oligomeric proteins composed of smaller subunits of 15–30 kDa. They are present in all organisms, but are especially abundant in plants, where they play a prime role in the plant heat shock response [37], as well as in other abiotic stresses.

3.2.3 Chaperone proteins

X1, X2, X3, and N166 were identical to HSP70, a high molecular weight HSP that maintains proteins in an unfolded state, for example for translocation and antigen presentation [37]. X10 and X59 corresponded to disulfide isomerase, an enzyme involved in the rearrangement of both intrachain and interchain disulfide bonds to form the native structures of proteins. N1 was similar to peptidyl-prolyl *cis-trans* isomerase, which catalyzes the *cis-trans* isomerization of proline imidic peptide bonds in oligopeptides, and has been shown to accelerate the refolding of several proteins *in vitro*.

3.2.4 Oxygen radical scavenging enzyme

Three antioxydant enzymes that destroy oxygen radicals, which are normally produced within the cells and are toxic to biological systems, were identified. Spots N145 and N142 were identical to glutathione peroxidase, an enzyme which catalyzes the oxidation of glutathione in the presence of hydrogen peroxide to yield oxidized glutathione and water. Both microsequences were identical; the HPLC profiles of both proteins were similar. These spots could correspond (i) to allelic differences in the primary structure of the same protein, (ii) to two isoforms (*e.g.*, cytosolic and chloroplastic forms), or (iii) to proteolytic products of the same protein. N95 was found to be Cu/Zn superoxide dismutase, which destroys O_2^- radicals. X48 was similar to ascorbate peroxidase, which destroys hydrogen peroxide.

3.2.5 Lignification enzymes and associated proteins

N125, X46, and X18 corresponded to two lignification proteins. The first two spots were similar to caffeoyl CoA-*O*-methyltransferase (CCoA-OMT), which catalyzes the methylation of caffeoyl-CoA into feruloyl-CoA, a precursor of lignin. The third spot was similar to caffeic *O*-methyltransferase (COMT), which catalyzes the conversion of caffeic acid to ferulic acid, and the conversion of 5-hydroxyferulic acid to sinapic acid using *S*-adenosyl methionine (SAM) as the methyl group donor. SAM synthetase (SAM-S), which catalyzes the formation of SAM from methionine and ATP, corresponded to spots X11, X15, X12, X21, and X30. In plants, SAM is a precursor of ethylene.

3.2.6 Defense-related proteins

N76 was identified as glucan endo-1,3-beta glucosidase, which is thought to be an important plant defense-related product against fungal pathogens. X41 was similar to isoflavone reductase, an enzyme that reduces achiral isoflavones to chiral isoflavones. It is also involved in the biosynthesis of phytoalexin.

3.2.7 Nitrogen metabolism

Three spots were found to be similar to glutamine synthetase (N48, X16 and X17), an enzyme that catalyzes the ATP-dependent incorporation of ammonium to glutamate for glutamine biosynthesis. N140 was similar to carbonic anhydrase, a chloroplast-localized enzyme which catalyzes the reversible hydration of carbon dioxide.

3.2.8 Other functions

N120 was similar to ATP synthase (ATPase) delta chain chloroplast precursor and N153 was similar to ATPase beta chain mitochondrial precursor. ATPase is the major primary pump in ATP-generating pathways, playing a role in proton conduction throughout the membrane. The sequences of N9 were identified as ATP-dependent protease, an enzyme involved in the degradation of denaturated proteins in the chloroplast. N53 and X9 had similar HPLC profiles and identical sequences. They were similar to actin, a protein produced during the division and elongation of cells. X33 corresponded to ACC oxydase, the ethylene-forming enzyme. X45 was similar to a *ran*-binding protein, which interacts with the *ran* protein to transfer proteins from the cytoplasm to the nucleus. Finally, X54 was the only spot showing similarity with a protein described in a bacterial organism involved in the regulation of bacterial floculation.

3.3 Database

A proteomic database of maritime pine is accessible on the internet (http://www.pierroton.inra.fr/genetics/2D/). This database has scanned gels, clickable images, with characterized protein spots highlighted by hyperlinked

symbols. Individual protein entries are linked to other protein databases such as OWL by active cross-references. Besides sequence data, this database contains further information. For certain needle proteins, the location on a linkage map, the behavior in drought environment, and seasonal and genetic variations are provided. The response to gravitational stimuli is given for xylem proteins.

4 Concluding remarks

High-throughput automated sequencing is becoming a common procedure for identifying gene functions from partial cDNA sequences. Such major initiatives in gene discovery are underway in major crop species, yeast, humans, and laboratory animals (see http://www.hgmp.mrc.ac.uk/GenomeWeb/genome-db.html). Projects aiming at systematically sequencing proteins of plants and animals, revealed by 2-D PAGE, are less advanced. This is because protein identification through microsequencing is still expensive (about 100 times more expensive than DNA sequencing). In addition, fully mastering this technique is difficult and time-consuming. However, techniques such as peptide-mass fingerprinting and protein sequencing *via* tandem mass spectrometry should soon allow large-scale proteomics research. Compared to genomics projects, one of the main advantages of proteome studies is direct access to the expressed genome in a given organ or tissue.

The major value of our project is protein identification: 63 spots were analyzed, from which 90.5% were identified. About 44% of the identified sequences were common to those of other conifers. This first set of characterized proteins is preliminary, and continuation is planned using electrospray ionization MS for spot identification. Choice of proteins for sequencing will aim at (i) segregating protein loci localized in the maritime pine linkage map [10]; in a recent review, [38] 2-DE was classified as the best technique for obtaining a map of expressed sequences; (ii) proteins whose quantitative trait loci (QTL) controlling spot quantity (also called protein quantity loci: PQL [25]) colocalize with the QTL of traits of interest to sylviculture (*e.g.* drought resistance, wood quality). Characterization of spots within the above categories will allow proteins to be incorporated into a "candidate protein" approach, which aims to characterize quantitative trait loci. Conifer species are characterized by a large genome size [39]. For instance, the *C*-value for maritime pine was estimated to be 24 pg [40], making its genome 160 times larger than that of *Arabidopsis*. With such a large genome, classical RFLP analysis, as well as gene isolation (positional cloning), will be difficult. The study of colocation between "candidate proteins" and QTL could be an alternative to the characterization of QTLs; (iii) proteins differentially or constitutively expressed when the plant is faced with a series of abiotic stresses; and (iv) tissue-specific proteins that are also likely to be of great interest. For instance, wood (secondary xylem) is a unique tissue of woody plants and its formation is likely to involve genes expressed rarely or not at all in herbaceous plants.

We recently demonstrated that the quantification of proteins revealed by 2-D PAGE, together with clustering analysis, is a powerful technique for studying the simultaneous expression of many proteins in organisms submitted to abiotic stress (Costa *et al.*, submitted). Although a cluster image is visually attractive, its biological information *per se* is minimal without knowledge of protein functions. More protein characterization should make it possible to assign functions to unidentified proteins located in biologically meaningful clusters, *i.e.*, clusters corresponding to specialized pathway (*e.g.*, lignification) or comprising gene products involved in specific mechanisms (*e.g.*, detoxication). Finally, proteome analysis in maritime pine (*Pinus pinaster*) is likely to complement genomic research in a closely related conifer species, *Pinus taeda* [41]. Similarities with fully characterized cDNAs or ESTs, described in various conifer species, have been obtained for several protein spots. This will greatly facilitate cloning of maritime pine cDNAs encoding the identified proteins.

We wish to thank J. d'Alayer from the Institut Pasteur for the peptide sequences and helpful discussions. We are grateful to R. Petit for a critical review of the manuscript. P. Costa was supported by fellowship PRAXIS XXI / BD / 3845 / 94 from 'Sub-programa Ciência e Tecnologia / 2 º Quadro Comunitário de Apoio' (Portugal). This research was funded by INRA (AIP Génome), Région Aquitaine, and the European Union (FAIR- CT95 – 0781, FAIR CT98/3953). We also thank Michael Dever for correcting the manuscript and anonymous reviewers for their critical comments on a previous version of the manuscript.

Received September 1, 1998

5 References

[1] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[2] Damerval, C., Zivy, M., Granier, F., de Vienne, D., *Adv. Electrophoresis* 1988, *2*, 263–340.

[3] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[4] Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P., *Electrophoresis* 1997, *18*, 1217–1242.

[5] Farjon, A., in: Brill E. J., Backhuys, W. (Eds.), *Pines*, E. J. Brill, Leiden, The Netherlands 1984, pp. 1–220.

[6] Bahrman, N., Plomion, C., Petit, R. J., Kremer A., *Ann. Sci. For.* 1997, *54*, 225–236.

[7] Bahrman, N., Damerval, C., *Heredity* 1989, *63*, 267–274.

[8] Gerber, S., Rodolphe, F., Bahrman, N., Baradat, P., *Theor. Appl. Genet.* 1993, *85*, 521–528.

[9] Plomion, C., Bahrman, N., Durel, C.-E., O'Malley, D. M., *Heredity* 1995, *74*, 661–668.

[10] Plomion, C., Costa, P., Bahrman, N., *Silvae Genet.* 1997, *46*, 161–165.

[11] Bahrman, N., Petit, R., *J. Mol. Evol.* 1995, *41*, 231–237.

[12] Bahrman, N., Zivy, M., Damerval, C., Baradat, P., *Theor. Appl. Genet.* 1994, *88*, 407–411.

[13] Petit, R. J., Bahrman, N., Baradat, P., *Heredity* 1995, *75*, 382–389.

[14] Costa, P., Bahrman, N., Frigerio, J.-M., Kremer, A., Plomion, C., *Plant. Mol. Biol.* 1998, *38*, 587–596.

[15] Damerval, C., Le Guilloux, M., Blaisonneau, J., de Vienne, D., *Electrophoresis* 1987, *8*, 158–159.

[16] Damerval, C., de Vienne, D., Zivy, M., Thiellement, H., *Electrophoresis* 1986, *7*, 52–54.

[17] Chen, H., Chen, H., Bjerknes, M., *Anal. Biochem.* 1993, *212*, 295–296.

[18] Komatsu, S., Kajiwara, H., Hirano, H., *Theor. Appl. Genet.* 1993, *86*, 935–942.

[19] Touzet, P., Riccardi, F., Morin, C., Damerval, C., Huet, J.-C., Pernollet, J.-C., Zivy, M., de Vienne, D., *Theor. Appl. Genet.* 1996, *93*, 997–1005.

[20] Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B., Celis, J. E., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 7701–7705.

[21] Rasmussen, H. H., Van Damme, J., Bauw, G., Puype, M., Gesser, B., Celis, J. E., Vandekerckhove, J., in: Jornvall, H., Höög, J.-O., Gustavsson, A. M. (Eds.), *Methods in Protein Sequence Analysis*, Birkhäuser Verlag, Basel 1991, pp. 103–114.

[22] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* 1990, *215*, 403–410.

[23] Pearson, W. R., Lipman, D. J., *Proc. Natl. Acad. Sci. USA* 1988, *85*, 2444–2448.

[24] Thiellement, H., Bahrman, N., Colas des Francs, C., *Theor. Appl. Genet.* 1986, *73*, 246–251.

[25] Damerval, C., Maurice, A., Josse, J.-M., de Vienne, D., *Genetics* 1994, *137*, 289–301.

[26] de Vienne, D., Leonardi, A., Damerval, C., *Electrophoresis* 1988, *9*, 742–750.

[27] Leonardi, A., Damerval, C., Hébert, Y., Gallais, A., de Vienne, D., *Theor. Appl. Genet.* 1991, *82*, 552–560.

[28] Burstin, J., de Vienne, D., Dubreuil, P., Damerval, C., *Theor. Appl. Genet.* 1994, *89*, 943–950.

[29] Gerber, S., Lascoux, M., Kremer, A., *Silvae Genet.* 1997, *46*, 286–291.

[30] Plomion, C., Hurme, P., Frigerio, J.-M., Ridolphi, M., Pot, D., Pionneau, C., Avila, C., Gallardo, F., David, H., Neutlings, G., Campbell, M., Savolainen, O., Kremer, A., *Mol. Bred.* 1998, in press.

[31] Tsugita, A., Kamo, M., Kawakami, T., Ohki, Y., *Electrophoresis* 1996, *17*, 855–865.

[32] Santoni, V., Doumas, P., Rouquié, D., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouzé, P., Dehais, P., Sahoun, I., Barlier, I., Rossignol, M., *Plant J.* 1998, *16*, 633–641.

[33] Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N., Nozu, Y., *Electrophoresis* 1994, *15*, 708–720.

[34] Kamo, M., Kawakami, T., Miyatake, N., Tsugita, A., *Electrophoresis* 1995, *16*, 423–430.

[35] Yamamoto, K., Sasaki, T., *Plant. Mol. Biol.* 1997, *35*, 135–144.

[36] Padmanabhan, V., Dias, D. M. A. L., Newton, R. J., *Plant. Mol. Biol.* 1997, *35*, 801–807.

[37] Vierling, E., *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 1991, *42*, 579–620.

[38] de Vienne, D., Burstin, J., Gerber, S., Leonardi, A., Le Guilloux, M., Murigneux, A., Beckert, M., Bahrman, N., Damerval, C., Zivy, M., *Heredity* 1996, *76*, 166–177.

[39] Wakamiya, I., Newton, R. J., Johnston, J. S., Price, H. J., *Am. J. Bot.* 1993, *80*, 1235–1241.

[40] Ohri, D., Khoshoo, T. N., *Plant. Syst. Evol.* 1986, *153*, 119–131.

[41] Allona, I., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J., Riedl, J., Retzel, E., Campbell, M., Sederoff, R., Whetten, R. W., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 9693–9698.

From Genome to Proteome: Advances in the Practice and Application of Proteomics
Edited by Michael J. Dunn

When citing this article, please refer to: *Electrophoresis* 1999, *20*, 1109–1119

Anne Sallandrouze[1]
Mireille Faurobert[2]
Mohammed El Maataoui[1]
Henri Espagnac[1]

[1]Département de Biologie, Laboratoire de Biologie Végétale Expérimentale, Université d'Avignon et des Pays du Vaucluse, Faculté des Sciences, Avignon, France
[2]Institut National de la Recherche Agronomique, Station d'Amélioration Génétique des Fruits et Légumes, Avignon Cedex 9, France

Two-dimensional electrophoretic analysis of proteins associated with somatic embryogenesis development in *Cupressus sempervirens* L.

Two-dimensional gel electrophoretic analysis and histological studies were performed on somatic embryos in cypress. Embryogenic cultures were obtained from *in vitro* culture of immature seeds. On a modified Murashige and Skoog (MS) medium they showed an intense and repetitive cleavage polyembryogenesis phenomenon which maintained them in a continuous proliferating status instead of undergoing a complete embryogenic development. Only the addition of bovine serum albumin to the culture allowed somatic embryo development and maturation. Major histological differences were noticed between developing and nondeveloping embryogenic cultures. Attempts to find proteins that could be associated with developmental stages of somatic embryos have been achieved. Proteins were extracted and analyzed by two-dimensional electrophoresis from nondeveloping embryogenic cultures (S0) and from embryogenic cultures at three different stages of somatic embryo development: small size and rounded shape embryos (S1), increased size embryos with a well-developed suspensor (S2) and embryos with two well-separated cotyledons (S3). The results revealed some qualitative and quantitative protein variations between the two cultures. Some could be connected with the induction of pro-embryo differentiation whereas others should be more related to the mechanisms involved in somatic embryo development and maturation. Specific polypeptides associated with the presence of bovine serum albumin (BSA) in the medium have been detected.

Keywords: *Cupressus sempervirens* L. / Two-dimensional polyacrylamide gel electrophoresis / Somatic embryogenesis / Somatic embryo maturation / Protein markers EL 3357

1 Introduction

Somatic embryogenesis is a process analogous to zygotic embryogenesis, in which a single cell or a small group of vegetative (*i.e.*, somatic) cells are the precursors of the embryos [1]. This phenomenon can be divided into four major steps: (i) initiation of proembryogenic masses, (ii) proliferation of embryogenic cultures, (iii) maturation of somatic embryos, and (iv) regeneration of whole plants [2, 3]. Somatic embryogenesis is widely investigated in conifer biotechnology because, on one hand, it provides useful systems for plant propagation and genetic manipulations for reforestation programs and, on the other hand, it allows fundamental studies on embryo development [4–7]. The gymnosperm species have for a long time been considered to be recalcitrant with respect to the process of somatic embryogenesis [1]. It is only within the last ten years that somatic embryogenesis in conifers was demonstrated for the first time in *Picea abies* and *Larix decidua* [4]. Somatic embryogenesis in conifers has since been greatly improved and has been reported for a wide range of conifer genera and species [4, 5, 8]. However, whole plant regeneration still remains critical due to problems in full maturation of somatic embryos. Several factors including nutrient composition, osmotic pressure, and abscisic acid (ABA) may promote maturation (reviewed in [4]).

Cupressus sempervirens L. (common cypress) is a conifer species native to the Mediterranean region where it is used both as ornamental species, wind-breaker, and for high-quality wood production [9, 10]. For this species, Lambardi *et al.* [11] obtained embryogenic cultures from excised immature zygotic embryos but without any plant regeneration. Although the cultures were maintained for two years and subjected to different maturation promoting treatments (*i.e.*, ABA), they failed to produce mature somatic embryos. In our laboratory, somatic embryogenesis was investigated as a method of cypress multiplication.

Correspondence: Mireille Faurobert, Institut National de la Recherche Agronomique, Station d'Amélioration Génétique des Fruits et Légumes, Domaine Saint-Paul, F-84914 Avignon Cedex 9, France
E-mail: mireille.faurobert@avignon.inra.fr
Fax: +33-4-9031-6291

Abbreviations: ABA, abscisic acid; **AGPs**, arabinogalactan proteins; **DM**, development medium; **ESM**, embryonal suspensorial masses; **MM**, maintenance medium; **MS**, Murashige and Skoog medium

When an embryogenic cell line was maintained on the induction medium, the events of somatic embryo morphogenesis were blocked early. Attempts have been made to overcome this developmental deviation. Since *C. sempervirens* embryo cultures proved to be recalcitrant to ABA treatment for maturation [11], we tested various other components to promote somatic embryo differentiation and maturation. Some, such as bovine serum albumin (BSA), showed positive effects leading to the production of well-conformed cotyledonary embryos, able to regenerate whole plants.

Mastering somatic embryogenesis implies a better understanding of biochemical mechanisms involved in this phenomenon. Unlike angiosperm embryogenesis, little is known about gene regulation and protein patterns during conifer embryo development. However, it has been shown that the presence of specific proteins can be correlated with the morphology of the somatic embryos and that some proteins and glycoproteins excreted in culture medium can influence the morphology of the somatic embryos of *P. abies* and *Pinus caribaea* [12, 13]. Some of these proteins have been identified as being arabinogalactan proteins (AGPs), chitinase, and germins [14–16]. In gymnosperms, extracellular proteins therefore seem to play a crucial role during embryogenesis as it was reported for angiosperm [17].

The aim of this work was a better understanding, at the cytohistological and biochemical levels, of *C. sempervirens* somatic embryo development and maturation in the presence of BSA. Our approach has consisted, first, in investigating if biochemical and cytohistological differences could be detected between embryogenic cell line maintained on the induction medium or cultured in presence of BSA and, second, in characterizing specific protein markers associated with the various stages of somatic embryo development by means of two-dimensional electrophoresis.

2 Materials and methods

2.1 Plant material and *in vitro* culture

The embryogenic cell line used in this study was originally induced on an immature zygotic embryo isolated from developing female cones collected on an adult tree growing in a wind-break near Avignon (France). The initiation of somatic embryos was realized according to Lambardi *et al.* [11] with modifications: (i) female cones were collected in mid-February, one year after fertilization, and (ii) the induction medium was hormone-free and consisted of Murashige and Skoog (MS) salts supplemented with 15 g/L of both fructose and glucose, 7 g/L Bacto-agar, 4 g/L charcoal and 10 mL/L coconut water (Sigma, St. Louis, MO, USA). Once initiated, the cell line was separated from the parent explant and maintained by fragmentation and subculture, in an interval of three weeks, on fresh medium of the same composition (maintenance medium, MM). Somatic embryo development and maturation was induced by transferring culture portions on medium supplemented with 1 g/L filter sterilized Sigma BSA (development medium, DM). Under our conditions, this organic compound proved to promote development and maturation of somatic embryos when it replaces coconut water in the maintenance medium (El Maataoui, unpublished results). All cultures were maintained in the dark at 25°C. For this study, embryogenic cell line maintained on MM (S0) was compared to the embryogenic cell line cultured on DM at three different stages of somatic embryo development: small size and rounded shape embryos (S1 stage), increased size embryos with a well-developed suspensor (S2 stage), and embryos with two well-separated cotyledons (S3 stage).

2.2 Cytohistological investigations

During the culture period, samples collected from cultures on maintenance medium and BSA treatments were fixed in formalin-acetic acid-ethanol (FAA) and prepared for cytohistological studies as previously described [18].

2.3 Protein extraction

About 50 mg of embryogenic cultures at the four developmental stages described above were placed in liquid nitrogen and then stored at –80°C. The plant material was transferred to a pre-chilled mortar, frozen in liquid nitrogen and ground for 5 min into a fine powder. The powder was homogenized in 1 mL of ice-cold extraction buffer containing 10 mM Tris-HCl, pH 8.0, 50 mM EDTA, 250 mM NaCl, 14 mM β-mercaptoethanol, 1 mM phenylmethyl sulfonyl fluoride (PMSF), 0.1% v/v Triton X-100 and 2% w/v polyvinyl polypyrrolidone (PVPP). After centrifugation at 20 000 × *g* for 15 min at 4°C, proteins of the supernatant were allowed to precipitate at –20°C for at least 2 h after addition of four volumes of cold acetone containing 10% w/v trichloroacetic acid and 0.07% v/v β-mercaptoethanol. After centrifugation at 13 000 × *g* for 10 min at 4°C, the supernatant was discarded and the pellet was washed three times with 1 mL 0.07% v/v β-mercaptoethanol in 80% v/v cold acetone. The pellet was then dried and stored at –20°C. The protein concentrations of the samples were estimated using the Bio-Rad (Richmond, CA, USA) protein assay [19].

2.4 Electrophoretic separation and proteins staining

Two-dimensional polyacrylamide gel electrophoresis was modified, using the procedure described by Damerval *et al.* [20] and Faurobert [21]. The Hoefer Iso-Dalt system (San Francisco, CA, USA) was used. Protein pellets were resuspended in a UKS solution containing 9.5 M urea, 5 mM K_2CO_3, 1.25% w/v SDS, 0.5% w/v DTT, 6% v/v Triton X-100 and 5% v/v Pharmalyte 3–10. For isoelectric focusing the gel mixture was 4% w/v acrylamide, 9.2 M urea, 2% v/v Triton X-100 and 4% carrier ampholytes, consisting of a mixture of equal parts of Pharmalyte 3–10 and Pharmalyte 4–6.5. For all samples, 35 µg of proteins were loaded on the basic end of the rod gels. IEF was performed at 25°C for 35 kVh. The rod gels were then extruded and stored at –20°C. For the second dimension, ten slab gels were simultaneously polymerized with 8.5% w/v acrylamide, 0.5 M Tris-HCl, pH 8.8, 0.5% w/v sucrose and 0.15% SDS. The rod gels were deposited on the slab gels and proteins were separated according to the classic SDS polyacrylamide gel procedure overnight at constant 100 V at 14°C. The gels were silver-stained according to Morrissey [22]. For estimation of isoelectric points and apparent molecular mass, extracts were comigrated with Bio-Rad 2-D SDS-PAGE standards and Pharmacia 2-D electrophoresis carbamylyte™ (Uppsala, Sweden).

2.5 Data analysis

At least two reproducible gels were obtained for each sample. Gels were digitized with an Agfa Arcus II scanner at 300 dots per inch. Spots were detected and quantified with Bio-Rad Melanie II software on the basis of their relative volume, *i.e.*, the spot volume divided by the total volume over the whole set of gel spots. Gels comparisons were performed by picking out a reference gel and matching each gel to the reference gel. Synthetic gels were computed for each sample by matching gel replicates. Spot % volume was the mean of the % volume for each matched gel. Correspondence analysis was used to describe relationships between gels and to detect main sources of variation between samples without taking into account a possible *a priori* classification of the gels. Statistical analysis were performed with Statistica StatSoft software.

3 Results and discussion

3.1 Developmental and cytohistological observations

In the literature, the majority of studies devoted to biochemical and physiological approaches of somatic embryo maturation in conifers have been performed using genetically different embryogenic and nonembryogenic cell lines [23–27]. Here we used a single cell line, the behavior of which differed according to the medium composition: repetitive polyembryogenesis leading to an undifferentiated state on MM *versus* coordinated development and maturation on DM. The cell line selected in our study consisted of fast-growing translucent to white loose tissue (Fig. 1A) embedded in a mucilaginous substance. Isolated from the parent explant and transferred to MM, it proved to be highly prolific (the culture volume doubled approximately every two weeks) and exhibited a homogeneous composition (S0 stage). It was continually composed of long vacuolated groups of cells subtending aggregates of cells with a dense cytoplasm (Fig. 1B). These structures resembled the well-known embryonal suspensorial masses (ESM) described in other systems of somatic embryogenesis in conifers [13, 28, 29]. Cytohistological controls showed (Fig. 1B) that the two constituents of cypress ESM were concerned by cell dissociation and polyembryogenesis. On the embryonary poles, the peripheral cells accumulated starch, vacuolated, and separated from mother masses *via* a process of cell wall gelification (Fig. 1B). Subsequently, they behaved as embryogenic cells and generated other ESM (results not shown). The same phenomenon concerned the suspensorial structures.

Despite their embryogenic appearance the cypress ESM never gave rise to recognizable cotyledonary somatic embryos when maintained on MM. In the presence of BSA (DM) the colonies exhibited reduced growth and within five weeks they formed prolific pro-embryo populations (Fig. 1C) with similar morphology and synchronized development (S2 stage). Each pro-embryo consisted of a well-differentiated embryo head subtended by a long and vigorous suspensor. Cytohistological observations (Fig. 1D) revealed that the embryo heads were composed of densely packed embryogenic cells surrounded by a well-individualized protodermic layer. In the cell cytoplasm, numerous protein inclusions could be observed after naphthol blue B staining. They exhibited a greater abundance as compared to the S0 stage. The suspensor showed cohesive vacuolated cells, closely associated to the embryo head (Fig. 1D). According to these observations, at structural and morphological levels, the effect of BSA appeared to be a maintenance of cell cohesion during the early stages of proembryogenesis both in embryonic and suspensorial regions.

After two months on DM, the ESM formed globular and cotyledonary somatic embryos displaying morphological and structural characters similar to their zygotic counterparts. When transferred from DM onto a BSA-free medi-

um, the somatic embryos which reached the cotyledonary stage generated whole plants (not shown). Thus, the maintenance of cell to cell contact in embryonic region as well as in the suspensor at early developmental stages seems to be a prerequisite for differentiation and development of well-conformed somatic embryos in our system. This is in contrast with the situation in the other investigated conifers where the limiting factor is the maturation step, which is commonly stimulated by ABA and increased osmolarity [4, 7, 23]. In particular, both ABA and osmotic agents such as sucrose, polyethylene glycol and/or dextran promote deposition of lipid storage products and synthesis of a number of proteins, including storage proteins and late-embryogenesis-abundant (LEA) proteins (reviewed in [7]). The LEA proteins are believed to prevent embryos from damage desiccation and from precocious germination during seed development. In cypress, although pro-embryos are blocked at an early developmental stage above the maturation step, BSA may act, similarly, as an osmotic agent. In fact, as the most

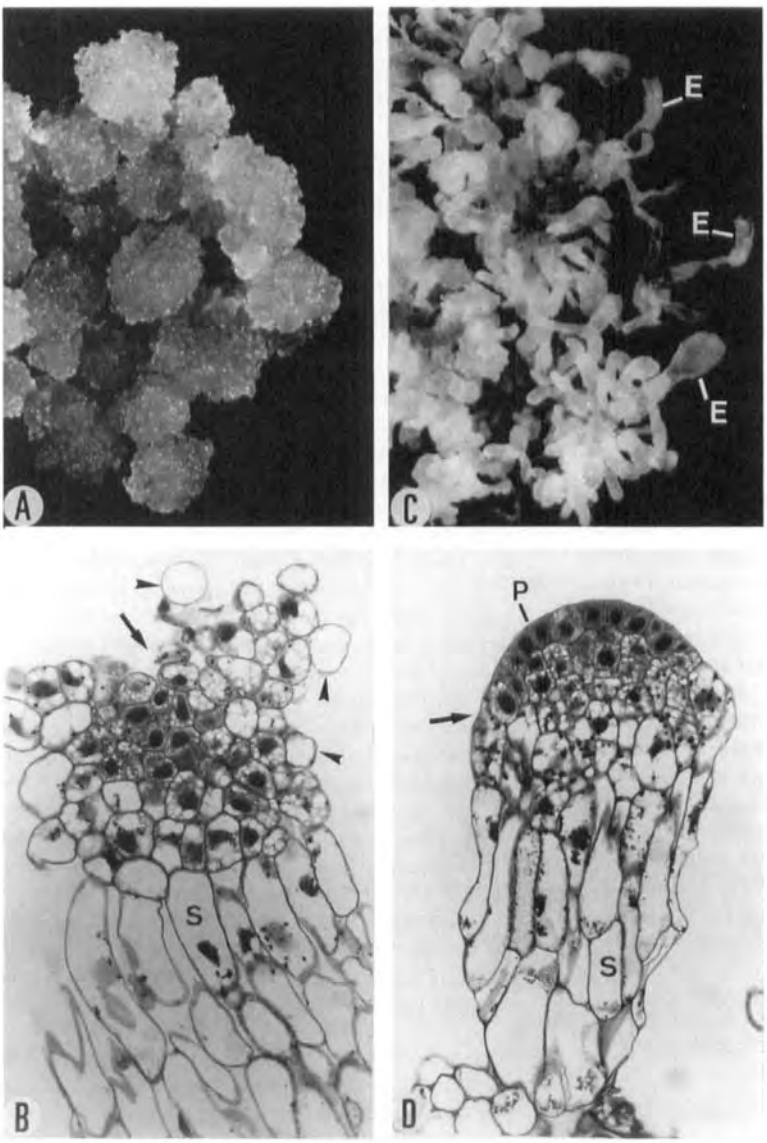


Figure 1. Morphological and cytohistological aspects of *Cupressus sempervirens* cell line on (A, B) maintenance medium and (C, D) BSA-supplemented medium. (A) An embryogenic colony showing a nodular structure. Each nodule corresponds to an embryonal suspensorial mass (ESM) which soon proliferates to produce other ESM. (B) Section of an ESM showing lack of a protodermic layer. Note (arrowheads) the vacuolization and the dissociation of cells constituting the embryo head (arrow). (C) An embryogenic colony after 5 weeks culture on BSA-containing medium. Numerous embryos (E) developed into polarized structures with a vigorous suspensor subtending an embryo head. (D) Section of a BSA-treated pro-embryo showing a protoderm (P) surrounding the embryo head (arrow). Note the cell cohesion both in the suspensor (S) and the embryonal region.

abundant protein in the circulatory system, serum albumin is known to contribute 80% to colloid osmotic blood pressure in higher vertebrates [30]. BSA may also play a metabolic role as an organic source of nitrogen. In *in vitro* systems the use of BSA is common in artificial culture media for microorganisms such as mycoplasmas [32]. Organic nitrogen sources have also been shown to be more effective in supporting somatic embryogenesis than inorganic ones in *Picea glauca* [31].

3.2 Protein content

Our 2-DE study showed that these cellular and morphological modifications were associated with considerable biochemical changes. One of the first major differences between the embryogenic cell line maintained on MM (S0) or cultured in presence of BSA was the increase in protein amount when somatic embryos differentiated and developed (Table 1). The protein concentration was particularly high (22.2 mg/g) when the somatic embryos were organized with two well-developed cotyledons. At this stage of development, a tenfold increase in protein content could be observed, as compared to the S0 stage. A possible explanation to these higher protein levels detected at the S2 and S3 stages may be an activation of metabolism in relation to the somatic embryos development. Filipecki and Przybecki [33] also reported that embryogenic lines had a considerably higher concentration of proteins than nonembryogenic ones in *Cucumis sativus* L. In the same way, they suggested that this difference could be connected with the presence in embryogenic cell lines of numerous small meristematic cells characterized by an intensive metabolism. The 3.7-fold difference in the amounts of protein between S2 and S3 stages may be partly due to the accumulation of storage proteins at the later developmental stage (S3). Since somatic embryos at the S3 stage appeared to be as highly hydrated as previous embryos stages, it is unlikely that the important increase in the protein content may be due to a partial dehydration of the sample. Similarly, a several-fold increase in storage protein concentration was noticed during somatic embryogenesis of *Gossypium hirsutum* [34].

3.3 Analysis of embryogenic cell line protein patterns

The polypeptide composition of embryogenic cultures on MM (S0) and DM (S1, S2, S3) was assessed by two-dimensional gel electrophoresis. The 2-D gels obtained for each sample exhibited several hundreds of spots with a wide distribution of apparent molecular mass (Fig. 2). The spots showed a symmetric shape indicating a good resolution in both electrophoresis dimensions. Replicate gels for the same sample showed a good reproducibility. For each gel, the studied zone was restricted between p*I* 4.0–6.5 and between 30 and 123 kDa for molecular mass. This zone corresponded to the part of the gel where the pH gradient was best stabilized. All the spots resolved within this zone were numbered and are illustrated in Fig. 3. Under our conditions, 244 reproducible spots could be detected in S0. In embryogenic cultures on DM an increase in spot numbers was related to somatic embryo development: 261 spots were observed in the S1 stage and 315 spots for the S2 stage. This increase was probably related to the differentiation of somatic embryos in more complex and specialized structures. In *Vitis rupestris*, it was shown that the protein number was higher in somatic embryos than in induced calli [35]. This was regarded as an intensification of metabolic activity in relation to the embryo morphogenesis. On the contrary, the S3 stage exhibited a reduced number of detected spots (only 157 spots) compared to the other developmental stages of cypress somatic embryos. As the same amount of proteins was loaded on each IEF gel, this deep decrease could be attributed to the relative abundance of all polypeptides as compared with those that were greatly accumulated in S3 (surrounded with a rectangle in Fig. 3). On the other hand, the lower polypeptide numbers observed at stage S3 could be explained by the onset of the maturation process that involved accumulation of major storage products and resulted in prevalence of a limited number of components as showed by Gianazza *et al.* [35] in *Vitis* mature somatic embryos.

The analysis of electrophoretic patterns revealed that S1 polypeptides profiles were close to the polypeptide pro-

Table 1. Evolution of protein content of cypress embryogenic cultures on MM or DM medium

Protein content (mg/g of fresh weight)			
Embryogenic cell line on MM	Embryogenic cell line on DM		
S0 stage	S1 stage	S2 stage	S3 stage
2.1	3.7	6	22.2

S1, S2, S3 stages correspond, respectively, to small size and rounded shape embryos, increased size embryos with a well-developed suspensor, embryos with two well-separated cotyledons.

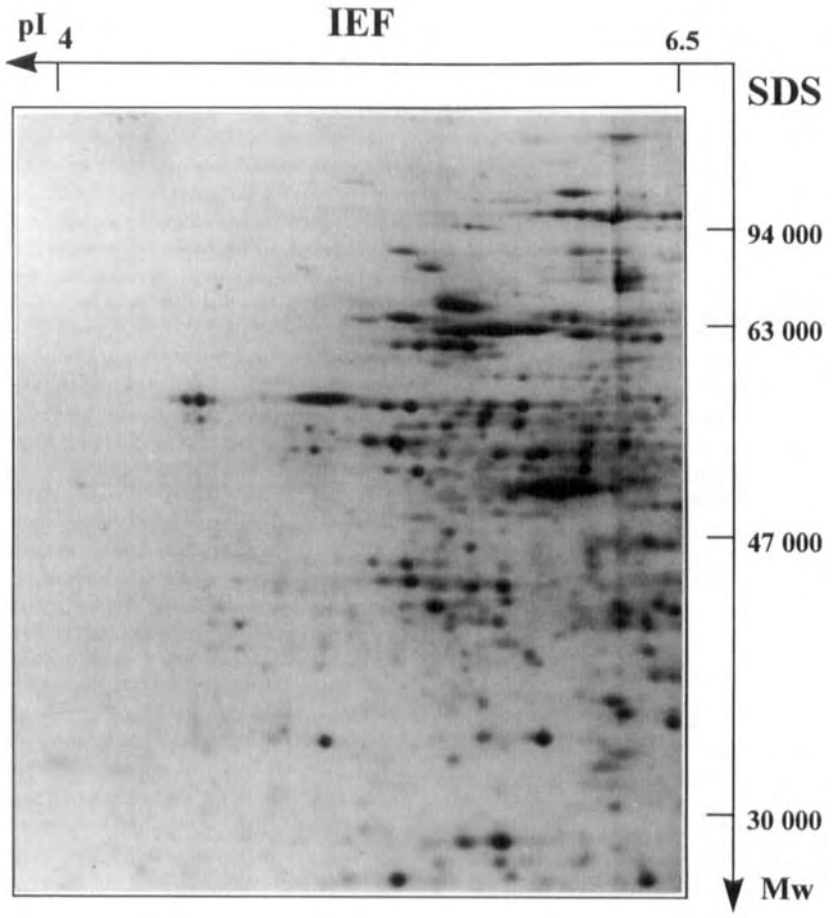


Figure 2. 2-DE gel of cypress somatic embryos at S2 stage. At this stage somatic embryos were well-conformed and the number of detected spots was maximal.

files observed in the control culture S0. Among the 244 polypeptides detected by 2-DE in the control cultures (S0), 200 could be retrieved in embryogenic cultures at the S1 stage, which represented 82% of total proteins present at the S0 stage. The volume of additional polypeptides was equivalent to 17% of the volume of the whole spots present at the S1 stage. This indicates that numerous polypeptides were common to well-formed and mis-shaped pro-embryos. However, the S2 map revealed a more complex pattern compared to the control culture (S0). Only 71% of the S0 polypeptide could be observed at the S2 stage and 31 additional polypeptides whose volume represented 41% of total spots volume could be noted. On the other hand, embryogenic cultures at the S3 stage showed protein patterns clearly modified compared to other embryogenic cultures: only 112 proteins were common to the embryogenic control cultures (S0) which corresponded to 46% of total protein content of SO cultures.

A comparison of samples was realized by means of correspondence analysis. The dataset was a matrix with % of volume of matching spots as rows, and gels as columns. The four synthetic gels were positioned in a space defined by three axes. The percentage of inertia associated with the three axes was 54.44%, 33.07% and 12.48%, respectively. Figure 4 reports the position of the samples in plane 1–2. The first axis clearly discriminates S3 from S0, S1 and S2. On this axis, S0 and S1 were close to each other while S2 and S3 were more distant. The second axis distinguishes between S2 and S0, S1 and S3. The same distribution in the space was obtained when repeats of gels were used instead of synthetic gels. Spots which exhibited a maximum % of volume for one stage of development were, as expected, positioned in the same area. Thus, results of correspondence analysis appeared to corroborate the observations described above which could be related to important physiological changes associated with development and maturation of somatic em-

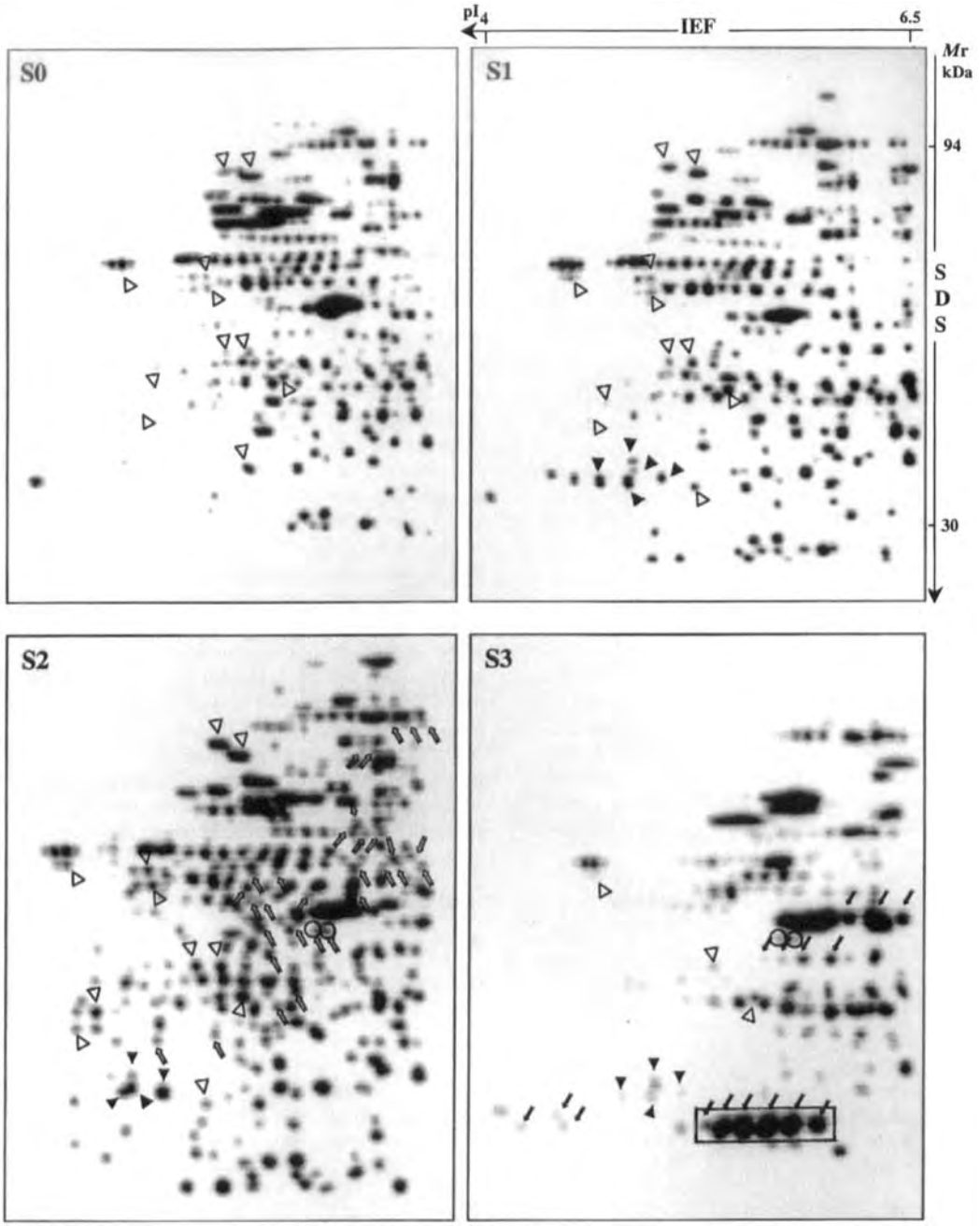


Figure 3. 2-DE synthetic gels of proteins from embryogenic cell line cultured without (Stage S0) or with BSA (Stages S1, S2, S3). S1, S2, S3 stages correspond to small size and rounded shape embryos, increased size embryos with a well-developed suspensor, and embryos with two well-separated cotyledons, respectively. S0 stage represents the embryogenic cell line cultured on maintenance medium. △, indicates polypeptides showing quantitative variations from stage S0 up to S3; ▲, represents polypeptides which appear at stage S1; ⇨, corresponds to polypeptides which appear at stage S2; ➡, indicates polypeptides specific to stage S3.

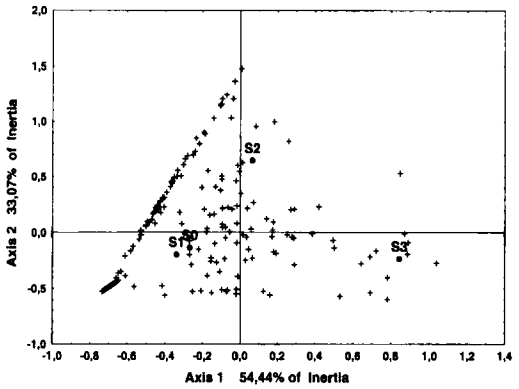


Figure 4. Correspondence analysis. Position on the plane 1–2 of the four synthetic gels corresponding to the S0, S1, S2, S3 developmental stages (S0, embryogenic cell line cultured on maintenance medium; S1, small size and rounded shape embryos; S2, increased size embryos with a well-developed suspensor; S3, embryos with two well-separated cotyledons) and position of the reference matched spots (indicated by +). The different developmental stages are well separated on axis 1.

bryos. In this connection, the more complex protein pattern observed at the S2 stage may be explained by a larger physiological gap between the two stages S0 and S2, in relation with a metabolism intensification. Similarly, in *V. rupestris*, Gianazza *et al.* [35] showed that callus protein patterns differed qualitatively from embryo-specific ones. In like manner, these authors suggested that this is related to an activation of cell division (massive protein and DNA synthesis). A drastic alteration in the polypeptide composition has also been reported during somatic embryogenesis in *Digitalis lanata* with the most pronounced modifications for the later stages of embryo morphogenesis [36]. On the contrary, the 2-D map of the S3 stage indicated a simplification and a specialization of the protein pattern as somatic embryos began to mature. Similar observations have been described in *V. rupestris* [35].

3.4 Characterization of specific protein markers associated with somatic embryo development

A detailed spot analysis allowed us to classify polypeptides from an embryogenic cell line maintained on MM medium and cultured in presence of BSA in four categories based on their differential accumulation: (i) constitutive polypeptides, whose amount remained almost unchanged; (ii) polypeptides, decreasing in concentration from the S0 stage up to the S3 stage; (iii) polypeptides increasing in abundance during somatic embryo development; and (iv) polypeptides lacking in the control culture, and appearing during somatic embryo differentiation and maturation (S1, S2, S3 stages). Among the polypeptides present in embryogenic cell line cultured on MM and DM, some characteristic quantitative variations were observed. The more representative ones are presented in Table 2. Two polypeptides (S0-21, S0-229) had a maximum level in control cultures (S0) and then decreased in their amount with the development of somatic embryos cultured on DM. The other ones showed, on the contrary, an increase with the differentiation and the development of somatic embryos up to the S2 stage. However, varia-

Table 2. Polypeptides showing quantitative variations during induction and development of cypress somatic embryos

	Spot number	Molecular mass (kDa)	p*I*	Polypeptide abundance (% Vol)[a] Embryogenic cell line on MM S0 stage	Embryogenic cell line on DM S1 stage	S2 stage	S3 stage
Polypeptide with a maximal	S0-21	79.75	5.44	1.07	0.86	0.75	0
volume level at S0 stage	S0-229	32.57	5.44	0.5	0.34	0.17	0
Polypeptides with a major	S0-93	57.79	4.64	0.06	0.19	0.2	0.05
increase between S0 and S1 stage	S0-175	39.62	5.67	0.44	0.66	0.69	0.46
Polypeptides with a major	S0-191	37.57	4.85	0.07	0.08	0.29	0
increase at S2 stage	S0-109	54.97	5.16	0.21	0.23	0.29	0
	S0-204	36.85	4.75	0.08	0.11	0.17	0
Polypeptides with a gradual	S0-152	43.61	5.29	0.15	0.23	0.34	0.21
increase from S0 to S3 stage	S0-90	57.35	5.16	0.18	0.24	0.29	0
	S0-150	43.61	5.44	0.22	0.37	0.52	0
	S0-22	80.66	5.27	0.33	0.55	0.69	0

S1, S2, S3 stages correspond, respectively, to small size and rounded shape embryos, increased size embryos with a well-developed suspensor, embryos with two well-separated cotyledons.

a) Data presented (% Vol) are mean from replicate experiments.

Table 3. Characteristics of new polypeptides detected at S1, S2 and S3 stages

	Spot number	Molecular mass (kDa)	pI	Polypeptide abundance (% Vol)[a]		
				S1 stage	S2 stage	S3 stage
Polypeptides which appear at S1 stage	S1-235	33.60	5.05	0.16	0.26	0.30
	S1-230	34.20	5.05	0.21	0.13	0
	S1-238	33.20	5.24	0.41	0.56	0.13
	S1-240	32.95	4.83	0.56	0	0.11
	S1-239	32.90	5.02	0.65	0.23	0.19
Polypeptides which appear at S2 stage	S2-23	87.25	6.38	0	0.07	0
	S2-36	81.57	6.18	0	0.07	0
	S2-126	57.92	6.42	0	0.10	0
	S2-21	87.25	6.57	0	0.11	0
	S2-22	87.25	6.47	o	0.11	0
	S2-35	81.57	6.26	0	0.12	0
	S2-157	52.60	5.76	0	0.13	0
	S2-129	56.62	5.74	0	0.13	0
	S2-82	64.70	6.20	0	0.15	0
	S2-128	56.80	5.84	0	0.15	0
	S2-156	53.08	5.71	0	0.16	0
	S2-152	53.45	6.00	0	0.17	0
	S2-259	37.34	5.21	0	0.18	0
	S2-105	60.42	6.49	0	0.19	0
	S2-256	37.94	5.55	0	0.19	0
	S2-120	58.11	6.58	0	0.20	0
	S2-79	65.02	6.30	0	0.20	0
	S2-178	49.12	6.04	0	0.21	0.10
	S2-81	64.86	6.11	0	0.22	0
	S2-74	66.50	6.14	0	0.23	0
	S2-237	40.64	5.85	0	0.25	0
	S2-174	49.45	5.77	0	0.25	0
	S2-108	59.84	6.38	0	0,26	0
	S2-176	49.00	6.07	0	0.26	0
	S2-177	49.23	5.91	0	0.26	0.11
	S2-122	57.92	6.31	0	0.29	0
	S2-219	42.15	5.94	0	0.30	0
	S2-191	46.28	5.81	0	0.34	0
	S2-143	54.80	5.70	0	0.37	0
	S2-149	53.57	6.11	0	0.58	0
	S2-123	57.92	6.12	0	0.66	0
Polypeptides specific to S3 stage	S3-145	31.92	4.40	0	0	0.10
	S3-96	46.41	5.75	0	0	0.12
	S3-153	31.59	4.17	0	0	0.14
	S3-152	31.59	4.44	0	0	0.15
	S3-95	46.20	5.86	0	0	0.22
	S3-93	46.31	5.98	0	0	0.50
	S3-92	46.31	6.07	0	0	0.52
	S3-151	31.63	5.38	0	0	0.93
	S3-89	51.89	6.45	0	0	1.17
	S3-88	52.01	6.07	0	0	1.87
	S3-149	31.63	6.07	0	0	2.14
	S3-150	31.63	5.49	0	0	3.34
	S3-146	31.63	5.92	0	0	4.34
	S3-148	31.63	5.63	0	0	4.86
	S3-147	31.63	5.77	0	0	4.89
	S3-87	51.54	6.29	0	0	6.33

These stages correspond, respectively, to small size and rounded shape embryos, increased size embryos with a well-developed suspensor, embryos with two well-separated cotyledons.

a) Values indicated represent the mean from replicate experiments.

tions were not similar according to the spots. Two of them (S0-93, S0-175) presented a more significant increase at the S1 stage than at the S2 stage whereas five others (polypeptides S0-150, S0-22, S0-152, S0-90 and S0-204) showed a gradual increase during the development of the somatic embryos. Polypeptides S0-109 and S0-191, on the other hand, were mainly expressed at the S2 stage. Similar changes in protein abundance have been reported in *D. lanata* embryogenic suspension culture [36]. According to these authors, these quantitative variations were correlated with a morphogenesis process. In a similar way, polypeptides increasing or decreasing in abundance during cypress somatic embryo differentiation and maturation could be associated with morphological modifications occurring during embryo development. The major difference which could be pointed out upon comparing electrophoretic patterns of an embryogenic cell line maintained on MM and cultured in presence of BSA was the presence of specific polypeptides at the S1, S2 and S3 stages (presented in Table 3 and Fig. 3). Five polypeptides were retained as specific of S1 in comparison with S0. These spots showed qualitative and quantitative variations from stages S1 to S3. Three of them (polypeptides S1-230, S1-239 and S1-240), mainly expressed in S1, should rather be connected with the induction of somatic embryo differentiation. Polypeptide S1-235, which showed a gradual increase with a maximal intensity at the S3 stage, should be linked with somatic embryo differentiation and maturation whereas polypeptide S1-238, having a maximal volume level at the S2 stage, should be more specific of somatic embryo differentiation. Thirty-one additional polypeptides showing abundance variations were observed at the S2 stage. Among them, only two were also present, but in smaller amount, at the S3 stage (polypeptides surrounded with a circle on Fig. 3). Sixteen polypeptides were specific to the somatic embryos with two well-separated cotyledons (S3 stage). Six of them (S3-146, S3-147, S3-148, S3-149, S3-150, S3-151) were accumulated to extremely high levels and represented 20% of the volume of the whole spots present at the S3 stage (Table 3). These S3-specific polypeptides had a molecular mass of 31.63 kDa and a p*I* range between 6.07–5.38. As similar proteins accumulated during cypress zygotic embryo development and maturation (Sallandrouze *et al.*, in preparation), they are assumed to correspond to storage proteins.

Protein markers associated with somatic embryogenesis were reported in other species for different embryogenic and nonembryogenic cell lines [33, 37–39]. In our case, protein markers were induced solely by medium composition modifications for a single cell line. Polypeptides synthesized at the S1 stage and a few polypeptides specific to the S2 stage could be considered as putative protein markers of the induction of pro-embryo differentiation. The nature and the function of these proteins in the induction process remains unknown to date. The main structural differences observed between cultures from the two treatments (with and without BSA) were: (i) lack of cell cohesion, (ii) absence of epidermis, and (iii) vacuolization and dissociation of the cells constituting the embryoïds of control culture (S0). Similar cellular events have been reported to occur in a developmentally blocked embryogenic cell line of *P. abies* [28, 29]. In this system, it was shown that changes in morphology and growth habit of the cell line were correlated with the absence of specific extracellular glycoproteins such as chitinase and AGPs [15, 16, 24, 40]. Direct implication of chitinase-like protein and AGPs in the formation of structurally conformed somatic embryos during the early stages of development has also been demonstrated in *Daucus carota* [41–43]. In this manner, the promoting effect reported here for BSA may be partly related to the presence of similar proteins which remain to be characterized.

4 Concluding remarks

In conclusion, adding BSA to the maintenance medium induced important biochemical changes associated with cellular and morphological modifications, leading to the differentiation of pro-embryos in mature somatic embryos. In particular, we have shown that (i) cell cohesion and epidermis formation were some requirements essential to a good development of somatic embryos, (ii) an increase in the complexity of protein pattern was connected with the differentiation and development of somatic embryos whereas the onset of the maturation process was characterized by a simplification and a specialization of the profile, in relation with accumulation at high levels of some sets of specific polypeptides, and (iii) identification of specific protein markers at the three developmental stages S1, S2 and S3 could be related to the production of well-conformed mature somatic embryos. Further studies are required to clarify the specific functions of the polypeptides characterized in embryogenic cultures at the S1 stage. Identification and cellular localization of these polypeptides may provide a better understanding of their involvement in the somatic embryogenesis process.

This work was supported by the European Community in the framework of AIR 3 CT 93 1675 Contract "Cypress: a flexible Mediterranean species for the protection of intensive farmland and for the production of high quality wood in marginal forest sites subject to fire risk".

Received November 11, 1998

5 References

[1] Ammirato, P. V., in: Evans, D. A., Sharp, W. R., Ammirato, P. V., Yamada, Y. (Eds.), *Handbook of Plant Cell Culture*, McMillan, New York 1983, Vol. 1, pp. 82–123.

[2] Goldberg, R. B., Barker, S. J., Perez-Grau, L., *Cell* 1989, *56*, 149–160.

[3] Thomas, T. L., *Plant Cell* 1993, *5*, 1401–1410.

[4] Tautorus, T. E., Fowke, L. C., Dunstan, D. I., *Can. J. Bot.* 1991, *69*, 1873–1899.

[5] Attree, S. M., Fowke, L. C., *Plant Cell Tissue Organ Cult.* 1993, *35*, 1–35.

[6] Bommineni, V. R., Chibbar, R. N., Dalta, R. S. S., Tsang, E. W. T., *Plant Cell Rep.* 1993, *13*, 17–23.

[7] Zimmerman, J. L., *Plant Cell* 1993, *5*, 1411–1423.

[8] Jain, S. M., Newton, R. J., Soltes, E. J., *Theor. Appl. Genet.* 1988, *76*, 501–506.

[9] Raddi, P., Panconesi, A., *Eur. J. For. Path.* 1981, *11*, 340–347.

[10] Xenopoulos, S. G., in: Panconesi, A. (Ed.), *Il Cypresso National Research Council (C.N.R.) of Italy* 1991, pp. 61–66.

[11] Lambardi, M., Harry, I. S., Menabeni, D., Thorpe, T. A., *Plant Cell Tissue Organ Cult.* 1995, *40*, 179–182.

[12] Egertsdotter, U., Mo, L. H., von Arnold, S., *Physiol. Plant.* 1993, *88*, 315–321.

[13] Domon, J. M., Meyer, Y., Faye, L., David, A., David, H., *Plant Physiol. Biochem.* 1994, *32*, 137–147.

[14] Domon, J. M., Dumas, B., Lainé, E., Meyer, Y., David, A., David, H., *Plant Physiol.* 1995, *108*, 141–148.

[15] Egertsdotter, U., von Arnold, S., *Physiol. Plant.* 1995, *93*, 334–345.

[16] Egertsdotter, U., von Arnold, S., *J. Exp. Bot.* 1998, *319*, 155–162.

[17] Van Engelen, F. A., De Vries, S. C., *Trends Genet.* 1992, *8*, 66–70.

[18] El Maataoui, M., Pichot, C., Alzubi, H., Grimaud, N., *Theor. Appl. Genet.* 1998, *96*, 776–779.

[19] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[20] Damerval, C., De Vienne, D., Zivy, M., Thiellement, H., *Electrophoresis* 1986, *7*, 52–54.

[21] Faurobert, M., *Electrophoresis* 1997, *18*, 170–173.

[22] Morrissey, J. M., *Anal. Biochem.* 1981, *117*, 307–310.

[23] Misra, S., Attree, S. M., Leal, I., Fowke, L. C., *Ann. Bot.* 1993, *71*, 11–22.

[24] Mo, L. H., Egertsdotter, U., von Arnold, S., *Ann. Bot.* 1996, *77*, 143–152.

[25] Sabala, I., Egertsdotter, U., von Fircks, H., von Arnold, S., *J. Plant Physiol.* 1996, *149*, 163–170.

[26] Kormutak, A., Vookova, B., *Biol. Plant.* 1997, *39*, 125–130.

[27] Find, J. I., Norgaard, J. V., Krogstrup, P., *J. Plant Physiol.* 1998, *152*, 510–517.

[28] Jalonen, P., von Arnold, S., *Plant Cell Rep.* 1991, *10*, 384–387.

[29] Egertsdotter, U., von Arnold, S., *J. Plant Physiol.* 1993, *141*, 222–229.

[30] Carter, D. C., Ho, J. X., *Adv. Protein Chem.* 1994, *45*, 153–203.

[31] Barrett, J. D., Park, Y. S., Bonga, J. M., *Plant Cell Rep.* 1997, *16*, 411–415.

[32] Chang, C. J., in: Whitcomb, R. F., Tully, J. C. (Eds.), *The Mycoplasmas V: Spiroplasmas, Acholeplasmas and Myocoplasmas of Plants and Arthropods*, Academic Press, San Diego 1988, pp. 208–241.

[33] Filipecki, M. K., Przybecki, Z., *Genet. Pol.* 1994, *35*, 1–9.

[34] Schoemaker, R. C., Christofferson, S. E., Galbraith, D. W., *Plant Cell Rep.* 1987, *6*, 12–15.

[35] Gianazza, E., De Ponti, P., Scienza, A., Villa, P., Martinelli, L., *Electrophoresis* 1992, *13*, 203–209.

[36] Reinbothe, C., Tewes, A., Reinbothe, S., *Plant Sci.* 1992, *82*, 47–58.

[37] De Vries, S. C., Booij, H., Meyerink, P., Huisman, G., Wilde, H. D., Thomas, T. L., Van Kammen, A., *Planta* 1988, *176*, 196–204.

[38] Hilbert, J. L., Dubois, T., Vasseur, J., *Plant Physiol. Biochem.* 1992, *30*, 733–741.

[39] Fellers, J. P., Guenzi, A. C., Porter, D. R., *J. Plant Physiol.* 1997, *151*, 201–208.

[40] Egertsdotter, U., Mo, L. H., von Arnold, S., *Physiol. Plant.* 1993, *88*, 315–321.

[41] De Jong, A. J., Cordewener, J., Lo Schiavo, F., Terzi, M., Vandekerckhove, J., Van Kammen, A., De Vries, S. C., *Plant Cell* 1992, *4*, 425–433.

[42] Kreuger, M., Van Holst, G. J., *Planta* 1993, *189*, 243–248.

[43] Toonen, M. A. J., Schmidt, E. D. L., Van Kammen, A., De Vries, S. C., *Planta* 1997, *203*, 188–195.

Index