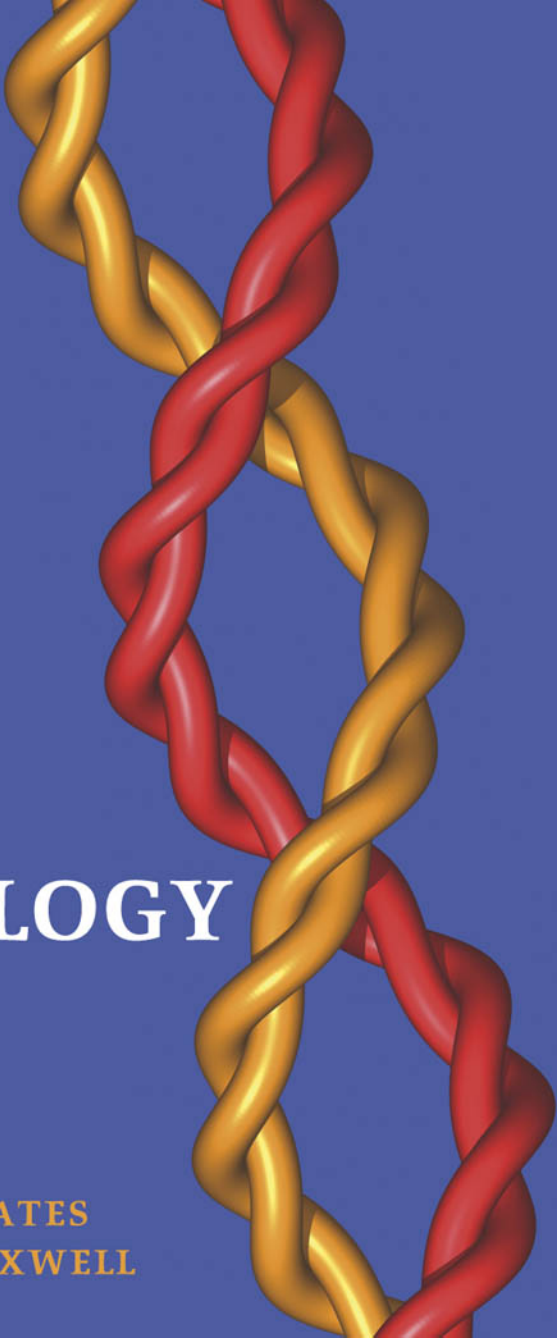
OXFORD
BIOSCIENCE



DNA TOPOLOGY

ANDREW D. BATES
ANTHONY MAXWELL

DNA Topology

This page intentionally left blank

DNA Topology

Andrew D Bates

School of Biological Sciences, University of Liverpool, UK

Anthony Maxwell

Department of Biological Chemistry, John Innes Centre, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan South Korea Poland Portugal
Singapore Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First Edition published 1993
Second Edition published 2005

A catalogue record for this title is available from the British Library
Library of Congress Cataloging-in-Publication Data
Bates, Andrew D.
DNA topology / Andrew D Bates, Anthony Maxwell. — 2nd ed.
p. ; cm.
Summary: "A key aspect of DNA is its ability to form a variety of structures, this book explains the origins and importance of such structures"—Provided by publisher.
Includes bibliographical references and index.
ISBN 0–19–856709–X (Hbk : alk. paper) — ISBN 0–19–850655–4 (Pbk : alk. paper)
1. DNA—Structure.
[DNLM: 1. DNA—physiology. 2. DNA—ultrastructure. QU 58.5 B329d 2005]
I. Maxwell, Anthony. II. Title.
QP624.5.S78B38 2005
572.8'633—dc22
2004024464
ISBN 0 19 856709 X (Hbk)
ISBN 0 19 850655 4 (Pbk)
10 9 8 7 6 5 4 3 2 1
Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by Biddles, King's Lynn

To

Marty Gellert and Steve Halford
for their guidance and wisdom,

and to

Sue and Sarah
for their forbearance.

This page intentionally left blank

PREFACE

This is effectively the second edition of *DNA Topology*, although not named as such. The original version was published in 1993, as part of the 'In Focus' series by IRL Press, since amalgamated into OUP. This series has long been discontinued and *DNA Topology* has been out of print for a number of years. However, many of our friends and colleagues have been complimentary about it, even to the extent of citing it fairly regularly in the literature, and we have been hassled over the years to write an update. Here, then, is a revised and expanded stand-alone version of *DNA Topology*.

When we started, it seemed that it would be a relatively straightforward matter of updating and developing the original. But it has turned out to be a much bigger job than we had anticipated and we overran the original target date for the revised manuscript by about 3 years! This was due to a combination of factors, including increased responsibilities and family commitments and, perhaps most importantly, not working in the same institution. The ability to pop down to the other end of the lab to hassle, whinge, or discuss was replaced by innumerable e-mails, sometimes mammoth telephone conversations and occasional trips between Liverpool and Norwich.

Of course, the subject has also moved on in 10 years, as has our perspective on it, and this has required significant additions and in some cases the adoption of a completely new approach. However our aim, as before, has been to make the concepts concerned with DNA topology approachable and interesting and to provide a bridge between the often superficial textbook discussions of the subject, and the in depth monographs and reviews that we cite as further reading. To do this we have tried to improve the clarity of the prose and the explanations, especially in response to specific comments from our colleagues.

Chapter 1 is somewhat expanded and updated and has benefited from very helpful input from Keith Fox and Stephen Neidle.

In Chapter 2, we have further developed the historical introduction to supercoiling, and included more material on the effect of solution conditions on the conformation and energetics of supercoiled DNA, which has been significantly developed in the literature in the last 10 years. This leads directly to consideration of the physical and theoretical background to the deformation (bending, twisting) of DNA and Monte Carlo simulations of supercoiled DNA molecules. All of this was absent from the first book, probably because we had

not taken the trouble to understand it properly, although it was a developing area at that time.

A colleague who read the original manuscript suggested that we should invite most readers to simply staple together the pages of Chapter 3 (DNA on surfaces), so as not to be put off by its apparent complexity. While we appreciate the problem, the book would be incomplete without it, and we hope that the further clarifications (new figures, text boxes, etc.) will make these ideas accessible. We have updated the story of the DNA in the nucleosome to include the most recent structural information, for which discussions with Andrew Travers were extremely helpful, and have described the application of the surface frame to DNA loops and genome-wide sequence analysis.

Chapter 4, knots and catenanes, has been considerably updated and revised, particularly in relation to the classification and nomenclature of knots and catenanes, which we felt was somewhat confusing in the original. We now introduce the 'ideal' description, which we feel is easier to understand, and we are indebted to Andrzej Stasiak for his advice on this.

Chapter 5, dealing with the DNA topoisomerases, has been made as current as possible, although rapid advances are being made in our understanding of enzyme structure and mechanism and this chapter will probably date rather quickly. Nevertheless, we anticipate that the main principles will remain valid. One major difference between this and the previous version is the inclusion of several crystal structures, which have been of enormous value in advancing our understanding of topoisomerase structure/function.

Chapter 6, on the biological consequences of DNA topology, has been revised, reorganized and expanded, and aims to exemplify the importance of DNA topology in biology. There is almost no limit to the amount of material we could have included here, so we hope the examples we have chosen serve to illustrate the principles.

To complement the black and white, two-dimensional figures in the book, we have developed a 'DNA Topology' website (http://www.oup.com/uk/bioscience/dnatopology), which includes more colourful versions of figures, three-dimensional representations of DNA and topoisomerase structures, useful links and other material. We hope this will prove to be a useful resource.

Of course this book would not have been possible without the help of many people. We would like to thank OUP for all their support and particularly their patience; key individuals were Kerstin Demata, Melissa Dixon, Abbie Headon, Liz Owen, Anita Petrie, and Ian Sherman. Among our academic colleagues we would particularly like to thank Keith Fox, Stephen Neidle, Marshall Stark, Andrzej Stasiak, and Andrew Travers for reading and making

helpful comments on parts of the manuscript. Thanks also to these and many others for permission to include their figures in the text, and to Liz Rawcliffe for her work on the website.

Andy Bates
University of Liverpool

Tony Maxwell
John Innes Centre, Norwich

This page intentionally left blank

PREFACE (1993)

'I only took the regular course.'
'What was that?' inquired Alice.
'Reeling and writhing, of course, to begin with', the Mock Turtle replied;

Lewis Carroll, *Alice's Adventures in Wonderland*

The Mock Turtle's reply, 'reeling and writhing' on the one hand gives a flavour of some of the subject matter of this book and, on the other, describes a common reaction of students when faced with the concepts of DNA topology. Since the description of the DNA double helix in 1953, the importance of DNA and the basic features of its structure have been widely appreciated by both scientists and non-scientists alike. However, it has more recently become clear that there are many deformations of this model structure that have important biological consequences. Prominent amongst these are the 'topological' deformations: supercoiling, knotting, and catenation. Unfortunately, a full appreciation of these aspects of DNA structure requires the grasp of concepts that can prove difficult for both students and more advanced researchers. We have written this book with the aim of explaining these ideas simply to allow a wider appreciation of DNA topology. We begin with a basic account of DNA structure before going on to cover DNA supercoiling, the definitions of linking number, twist and writhe and the free energy associated with supercoiling. The rather more complex description of DNA lying on a curved surface and its application to the nucleosome is then considered, followed by the phenomena of knotting and catenation. The final chapters deal with the enzymes that alter DNA topology (DNA topoisomerases) and, most importantly, the biological significance of the topological aspects of DNA structure.

We are grateful to Steve Halford for his careful reading of the manuscript and for many helpful comments and suggestions. We acknowledge Geoff Turnock, Dave Weiner, and Chris Willmott for their comments and Mike Dampier for useful discussions. We also thank the authors and publishers who have allowed the reproduction of figures from previous publications. A.D.B. is a Wellcome Trust Postdoctoral Research Fellow; A.M. is a Lister Institute Jenner Fellow.

Andy Bates
Tony Maxwell

This page intentionally left blank

CONTENTS

ABBREVIATIONS AND SYMBOLS

a	persistence length
A	adenine
Å	Angstroms
ARS	autonomously replicating sequences
ATP	adenosine triphosphate
att	attachment site
B	bending force constant
bp	base pairs
C	cytosine
C	twisting force constant
Ca	catenane number
CAP	catabolite activator protein
CRP	cAMP-receptor protein
d	effective diameter
ΔLk	linking difference
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide
Φ	DNA winding number
FIS	factor for inversion stimulation
G	free energy
G	guanine
γ	superhelix winding angle
h	helical repeat
h_s	surface-related helical repeat
h_t	twist-related helical repeat
h°	helical repeat under standard conditions
IHF	integration host factor
Int	bacteriophage lambda integration protein
k	Boltzmann constant
K	elastic constant
kb	kilobase pairs
kDa	kilodaltons
Kn	knot number

λ	bacteriophage lambda
Lk	linking number
Lk_m	Lk of most abundant relaxed topoisomer
Lk°	average linking number of relaxed DNA
mRNA	messenger RNA
n	number of superhelical turns
N	length of DNA in base pairs
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
R	purine
R	gas constant
res	site of recombination by resolvase
RNA	ribonucleic acid
σ	specific linking difference
SLk	surface linking number
STw	surface twist
T	absolute temperature
T	thymine
TBP	TATA-box binding protein
Tw	twist
ω	angular displacement
Wr	writhe
Y	pyrimidine

1
DNA structure

1.1 Introduction

Of all biological molecules, DNA has perhaps most fired the imagination of scientists and non-scientists alike. This is probably due to the aesthetic appeal of the DNA double helix, a structure that is apparently simple but also profound in terms of its implications for biological function. The term 'DNA structure' encompasses the chemical and stereochemical details of DNA molecules. The principal purpose of this book is to focus on the higher-order structural features of DNA, namely supercoils, knots, and catenanes but before embarking on these topics, it is essential to consider the primary and secondary structural features: how DNA chains are put together and how they can form helical structures.

1.2 DNA structures

1.2.1 *The Watson–Crick model*

Although it is now rather a cliché, it is nevertheless true that the elucidation of the structure of the DNA double helix by James Watson and Francis Crick in 1953 was one of the most important scientific discoveries of the twentieth century (1). The impact of their model has been enormous, providing the foundation for modern molecular biology, and leading to an explosion of discoveries and new techniques in the last half-century. Not only did this structure satisfy the known physical and chemical properties of DNA, it also revealed how DNA can fulfil its biological functions. For example, the complementarity of the two strands of the DNA double helix provided a potential mechanism for it to be copied and the possibility of a defined sequence of bases along a DNA strand suggested how genetic information could be encoded.

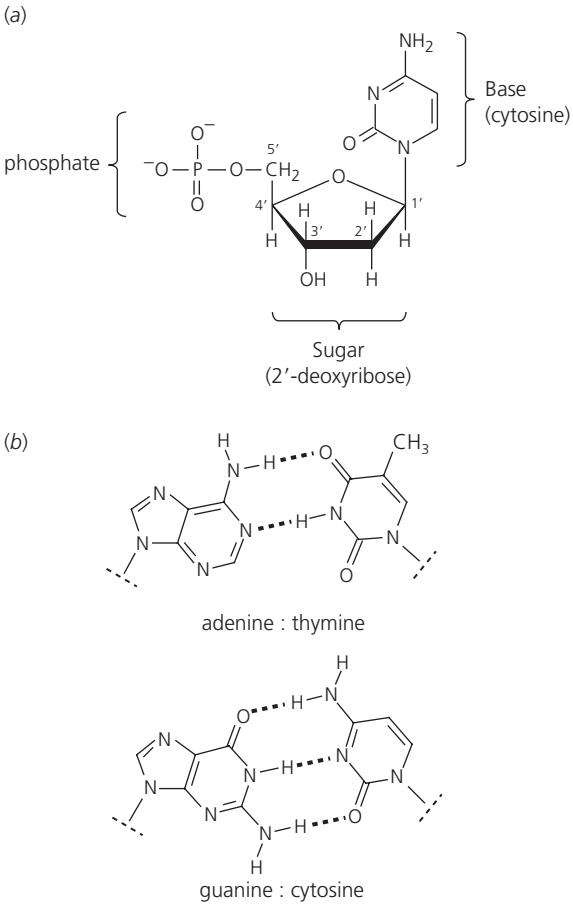


Figure 1.1 DNA components.
(*a*) A nucleotide, the repeat unit of DNA. (*b*) Base-pairing in DNA. Purine bases (adenine and guanine) pair with pyrimidine bases (thymine and cytosine); dotted lines indicate hydrogen bonds.

DNA is made up of repeated units, nucleotides, which consist of three components: a sugar (2′-deoxyribose), phosphate, and one of four heterocyclic bases. There are two purine bases, adenine and guanine, and two pyrimidines, thymine and cytosine. The structures of these components are shown in *Figure 1.1*. The bases are attached to the 1′-position of the sugar and a DNA chain is made by joining the sugars by a 3′–5′ phosphodiester linkage, which connects the 3′-hydroxyl group of one sugar to the 5′-hydroxyl of the next. This polynucleotide chain constitutes the primary structure of DNA (*Figure 1.2a*).

The principal experimental method used to elucidate the secondary structure of DNA was X-ray diffraction of DNA fibres. This involves stretching a highly concentrated solution of DNA into a fibre, in which

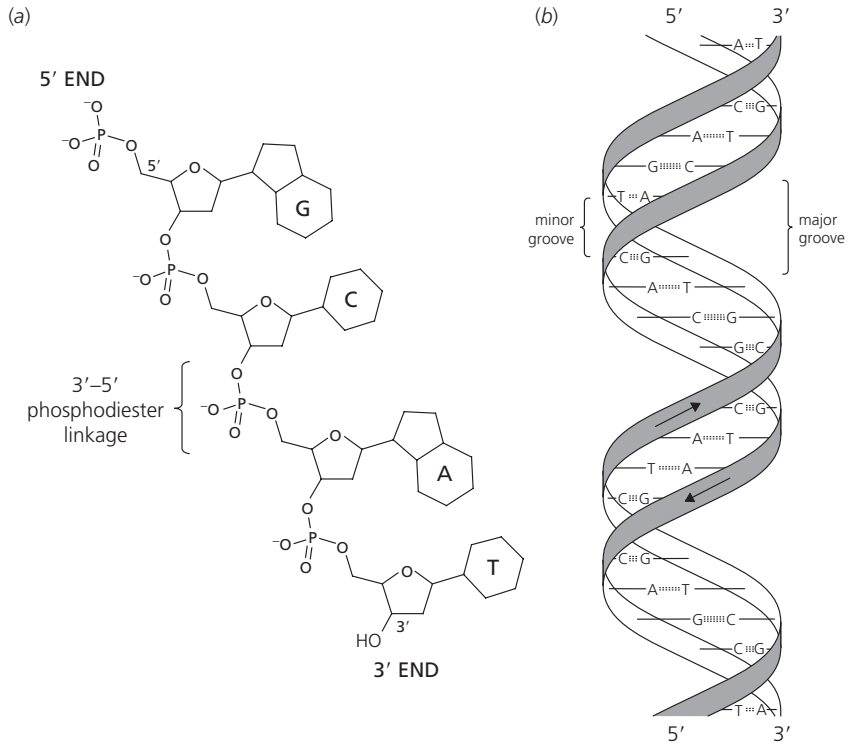


Figure 1.2 DNA structure. (*a*) A polynucleotide chain, showing the phosphodiester bonds that connect adjacent nucleotide units. (*b*) A schematic representation of the B-DNA double helix.

the DNA molecules become aligned with the fibre axis. The diffraction pattern of the mounted fibre is then recorded in an atmosphere of controlled humidity. Such experiments can provide only limited information but, coupled with chemical data and model building, they led to the discovery of the DNA double helix. The essence of double-stranded DNA is the specific pairing of the bases by hydrogen bonding (*Figure 1.1b*). Adenine pairs with thymine by means of two hydrogen bonds and guanine pairs with cytosine through three hydrogen bonds. Note that other base-pairing schemes are possible, apart from the classical 'Watson–Crick' scheme shown in *Figure 1.1b*. An example is 'Hoogsteen' base-pairing found in DNA triplexes (see Section 1.3.2); in addition, apparent mismatches (such as G–T and G–A) can occur, particularly in RNA structures.

The chemistry of the phosphodiester bond imparts a direction to the polynucleotide chain, from the free 5′- to the free 3′-end (5′ → 3′), and in duplex DNA the two strands run in opposite directions, that is, they are

antiparallel (*Figure 1.2*). The strands are coiled around each other in a right-handed fashion. (A right-handed helix observed along its helix axis spirals in a clockwise fashion away from the observer.) This produces a structure with a largely hydrophobic interior consisting of the planar DNA bases stacked on each other with the hydrophilic sugar–phosphate backbone on the outside. In the Watson–Crick model of DNA there are 10 base pairs (bp) for every turn of the helix and a repeat distance between each successive turn of 3.4 nm (34 Å). The width of the double helix is approximately 2 nm (20 Å) and the bases are perpendicular to the helix axis (*Figure 1.2b*). This form of the DNA double helix is known as the B-form, as distinct from the A-form, which was also detected by fibre diffraction experiments but under conditions of lower humidity (see Sections 1.2.2 and 1.2.3).

At this point it is pertinent to ask, 'why is DNA helical?' As we will see (Chapter 2, Section 2.2.1), the double helix is actually a major impediment to many of DNA's functions. It turns out that helices are common in biological macromolecules (e.g. the α-helix of proteins) and form as an almost inevitable consequence of the way in which asymmetric monomer units stack upon each other. In the case of DNA, each monomer (nucleotide) stacks onto another with a 'stagger' (*Figure 1.3*); you can think of this as stacking wedge-shaped blocks with a displacement each time a block is added (see the DNA Topology website: http://www.oup.com/uk/bioscience/dnatopology for further details). In real DNA, the conformation of the molecule is determined by the bond angles in the sugar–phosphate backbone, the hydrophobicity of the bases, the base-pairing between the two strands and the solution conditions. As we will see below, alterations in, for example, solution conditions and the sequence of the bases can have a profound effect on the overall conformation of the double helix (*Table 1.1*).

The DNA double helix is a relatively stable structure, although conditions such as high temperature or extremes of pH cause disruption of the helix into

Table 1.1 Structural features of DNA

DNA conformation	B	A	Z
Helix handedness	Right	Right	Left
bp per turn	10.5	11.0	12.0
Helix diameter	~2.0 nm (20 Å)	~2.6 nm (26 Å)	~1.8 nm (18 Å)
Sugar pucker	C2′-*endo*	C3′-*endo*	C2′-*endo* (pyr) C3′-*endo* (pur)
Major groove	Wide and deep	Narrow and deep	Flat
Minor groove	Narrow and deep	Wide and shallow	Narrow and deep

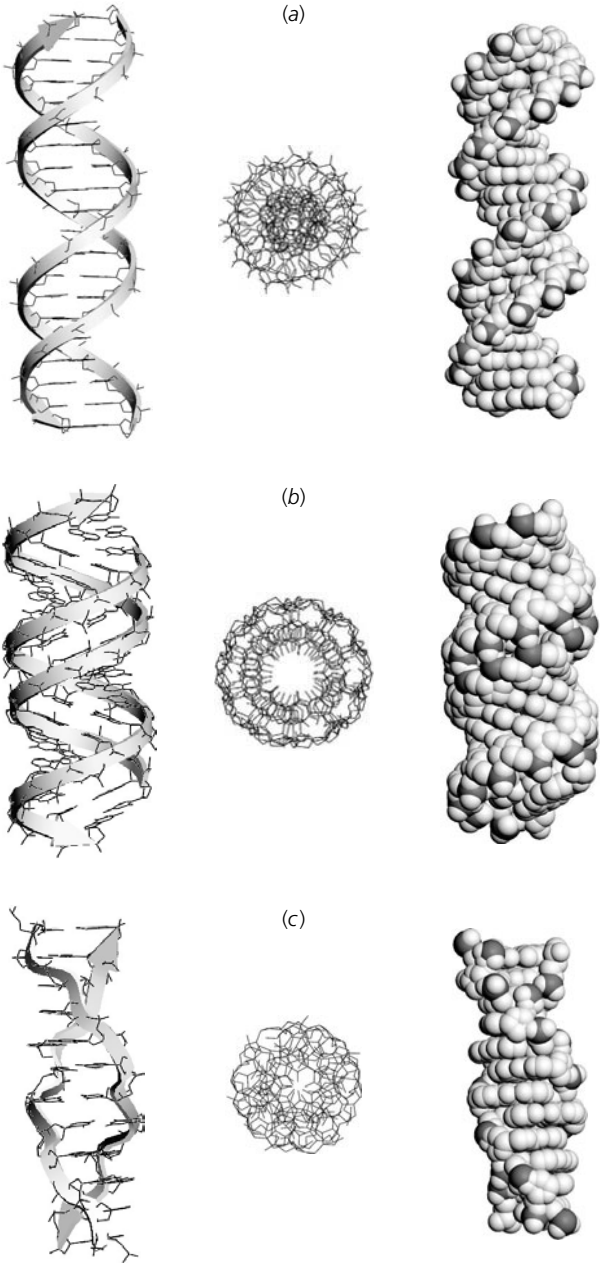


Figure 1.3 Comparison of the B-, A-, and Z-forms of DNA. (*a*) B-form; (*b*) A-form; (*c*) Z-form. In each case a ribbon representation is shown (left), a view down the helix axis (middle), and a space-filling representation (right) (adapted from ref. 58, courtesy of Prof. S. Neidle).

its single-strand components, a process known as denaturation. When this happens, the hydrogen bond donor and acceptor groups of the DNA bases become exposed. Transient denaturation of DNA occurs in the course of its biological functions and can be facilitated by proteins and, as described in Chapter 2, by DNA supercoiling.

The basic description of B-form DNA derived from fibre diffraction studies is sufficient as a working model for many purposes. However, we now know that DNA is not exactly a uniform structure. A number of biochemical and biophysical techniques, particularly X-ray diffraction studies of crystalline DNA, have shown that double-stranded DNA can adopt a variety of conformations. These include local variations in the helical parameters of B-form DNA, other helical species of DNA, and alternatives to the double-helical structure.

1.2.2 *B-form DNA*

B-form DNA is thought to represent the conformation of most DNA found in cells. The basic structural parameters were derived from X-ray diffraction data of DNA fibres at high humidity (e.g. 92%). The features that distinguish B-DNA (*Figure 1.3*) from other forms are the location of the base pairs on the helix axis, the near-perpendicular orientation of the base pairs relative to the helix axis and distinct major and minor grooves (the former in particular allowing easy access to the bases). Some of the helical parameters of B-form and other forms of DNA are given in *Table 1.1*. The helical repeat (base pairs per turn) of mixed-sequence B-form DNA has been determined by a variety of experimental methods and has been found to be close to 10.5 bp/turn; this value depends on the solution conditions, but is often taken as an average for B-DNA (see Chapter 2, Section 2.3.1). A number of other parameters are required for a complete description of the conformation of the double helix; see ref. 2 for a comprehensive description.

In 1980 the structure of a B-form DNA molecule was analysed at atomic resolution using X-ray diffraction of single crystals (3). The molecule used in these studies was the dodecamer d(CGCGAATTCGCG), which is self-complementary (i.e. two molecules can base pair with each other by hydrogen bonding) and forms a double helix of the type suggested by Watson and Crick. Although the overall features of the dodecamer structure conform closely to those expected of B-form DNA, the molecule shows local sequence-dependent variations. For example, the distance between base pairs varies from 0.314 to 0.356 nm, the average being about 0.33 nm (3.3 Å). The DNA base pairs are not all perpendicular to the helix axis and show 'propeller twist', where the purine and pyrimidine pair do not lie in the same plane but are twisted with respect

to each other like the blades of a propeller. In addition, the helix axis itself is not straight but is slightly curved. It is now known that these sequence-specific conformational variations are crucial for the interaction of DNA with other molecules (see Section 1.4), that is, DNA curvature is an important feature of DNA structure, particularly from the standpoint of its biological functions.

Aside from the local conformational variations in B-DNA that have emerged from the studies of the dodecamer and similar short DNA molecules of defined sequence, earlier work has described alternative B-like conformations, which are given distinct designations such as C and D; these species will not be discussed here (4, 5).

1.2.3 *A-form DNA*

A-form DNA was first identified from X-ray fibre diffraction studies at low humidity (75%) (6). More recently several short DNA molecules have been crystallized and shown to conform in general to the A-DNA parameters derived from fibre studies, these include duplexes of d(GTGTACAC) and d(GGGCGCCC) (7). Compared with B-DNA, the A-form helix is broader (2.6 nm in diameter), somewhat untwisted (11 bp/turn) and the bases are tilted and lie well off the helix axis (*Table 1.1*). The major groove is thin and deep while the minor groove is shallow (*Figure 1.3*). A key distinction between the A- and B-forms is the conformation of the ribose sugar (the ring 'pucker'), which is normally C3′-*endo* for A-DNA and C2′-*endo* for B-DNA. (Sugar pucker refers to the out-of-plane twisting of the furanose rings, in order to minimize non-bonded interactions between ring substituents (4).) The conformation of A-DNA closely resembles that of the double-helical form of RNA, A-RNA. It is not certain whether this form of DNA occurs *in vivo*, although the presence of A-DNA has been proposed at specific sites, such as promoter regions (8) and transcription factor binding sites (9). In addition, the crystal structures of some protein-DNA complexes reveal A-like conformations in the bound DNA, these include DNase I (10) and the TATA-box binding protein (TBP) (see Section 1.4.3; Chapter 6, Section 6.4.2) (11, 12).

1.2.4 *Z-form DNA*

The first DNA molecules to be analysed by X-ray crystallography were the self-complementary oligonucleotides d(CGCGCG) and d(CGCG) in 1979 and 1980 (13, 14). The resulting structures came as a great surprise as the double helices formed were left-handed. This conformation has become known as Z-DNA and it turned out that evidence for its occurrence had been in existence

since 1972 from spectroscopic studies of poly(dG–dC) (15). Apart from the difference in handedness and helical parameters when compared to A- and B-forms (*Table 1.1*), Z-form DNA is also distinguished by the zigzag path (hence 'Z'-DNA) of the sugar–phosphate backbone (*Figure 1.3*). Indeed, it perhaps makes more sense to think of Z-DNA as a helix of dinucleotide pairs, since successive base pairs are in alternating conformations. Other features are that Z-DNA essentially lacks a major groove and has a deep and narrow minor groove (*Table 1.1*; *Figure 1.3*). The Z conformation is characteristically, but not exclusively, found in DNA molecules with alternating GC sequences, and is favoured by specific conditions, notably high salt; the ease with which the Z conformation is formed depends on the particular sequence (16). An important issue is whether Z-DNA occurs in nature. At present this remains controversial, but there is some evidence supporting the existence of Z-DNA *in vivo* in both prokaryotes and eukaryotes (16). It is likely that the B-form represents the predominant conformation of DNA in cells and it is possible that left-handed DNA arises transiently during the course of cellular processes such as transcription (see Chapter 6, Section 6.5.1). In addition there is evidence that in yeast Z-DNA formation is involved in gene activation (17). In the context of DNA topology, the potential of sequences to form Z-DNA can have important consequences. For example, it can be shown that short segments of poly(dG–dC) in plasmids can be induced to form left-handed DNA at sufficient levels of negative supercoiling (18–21). This is one example of how the free energy of supercoiling can be used to stabilize DNA conformations that would be unfavourable in the absence of superhelical stress (see Chapter 2, Section 2.8.1; Chapter 6, Section 6.1.1).

1.3 **Alternative DNA structures**

Apart from the ability of DNA to adopt different helical forms, it can also be found as structures that depart from the familiar double helix. Some of these structures are stabilized by DNA supercoiling (e.g. cruciforms and H-DNA); others include the antiparallel and parallel quadruple helical structures proposed for chromosome telomeres, and the four-stranded structures proposed as recombination intermediates.

1.3.1 *Cruciforms and Holliday junctions*

Cruciforms arise as a consequence of intrastrand base-pairing in duplex DNA and consist of a pair of stem and loop or 'hairpin' structures. Their formation

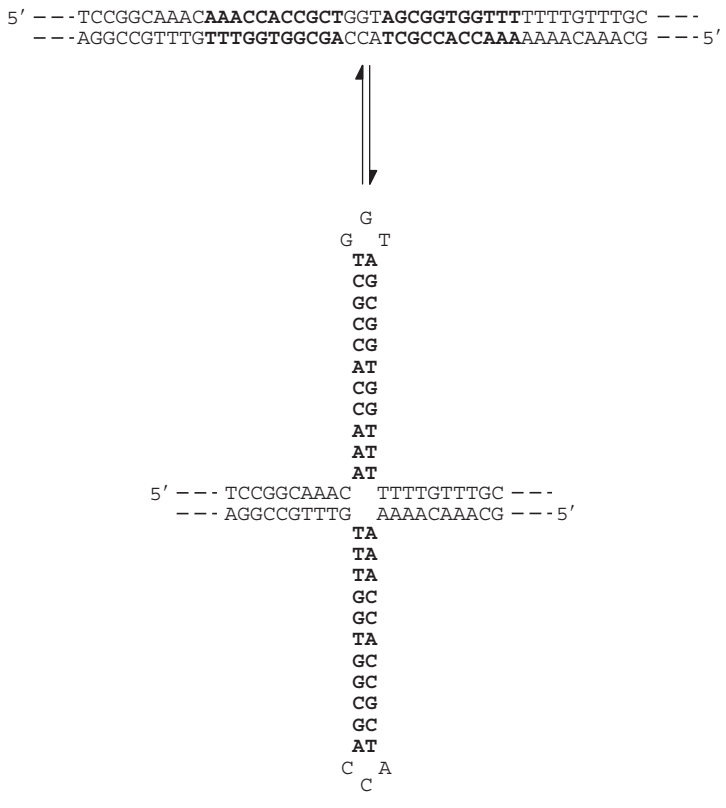


Figure 1.4 Formation of a cruciform from an 'inverted repeat' sequence in DNA. The sequence occurs in the bacterial plasmid pBR322; the 11 bp inverted repeat sequence is shown in bold.

requires the presence of inverted repeats (also termed 'palindromes') in double-stranded DNA, where a sequence is followed immediately, or soon after, by the same sequence in the opposite orientation. An example of such a sequence occurs in the bacterial plasmid pBR322 (*Figure 1.4*) and consists of an 11 bp inverted repeat spaced by 3 bp (22). This sequence can, in principle, form a cruciform with 11 intrastrand base pairs in the stems. Important questions concern the existence in nature and the biological significance of cruciform structures. It is clear that formation of a cruciform structure in linear B-DNA will be thermodynamically unfavourable due to the presence of unpaired bases. However, it is possible to show experimentally that cruciforms can form in negatively supercoiled closed-circular DNA *in vitro* (23). Here the unfavourable free energy associated with negative supercoiling is reduced by the formation of the cruciform (see Chapter 2, Section 2.8.1; Chapter 6, Section 6.1.1).

Evidence for the presence of cruciforms in such experiments relies on the preference of single-strand-specific reagents (e.g. S1 nuclease, bromoacetaldehyde) to act at the site of the inverted repeat sequence. Using these reagents it can be shown that increasing negative supercoiling promotes the extrusion (looping out) of the cruciform. Whether cruciforms actually exist *in vivo* and whether they are of any physiological significance are contentious issues; intracellular superhelical densities may be too low to extrude inverted repeat sequences efficiently (24). Kinetic studies of certain cruciforms have suggested that the extrusion process may be too slow for the cruciforms to be of physiological significance (25, 26). However, there are reports showing evidence of cruciform formation *in vivo* (27, 28), albeit in somewhat artificial situations. In general it is likely that palindromes are unstable in cells and it is known that long palindromes render bacteria inviable (29).

Despite the debatable physiological significance of cruciforms in DNA, they can be regarded as being equivalent to Holliday junctions: a four-armed intermediate formed between two DNA duplexes during homologous recombination (*Figure 1.5*). In this process, two homologous DNA double helices are aligned and, following cleavage of one strand of each helix, base-pairing is

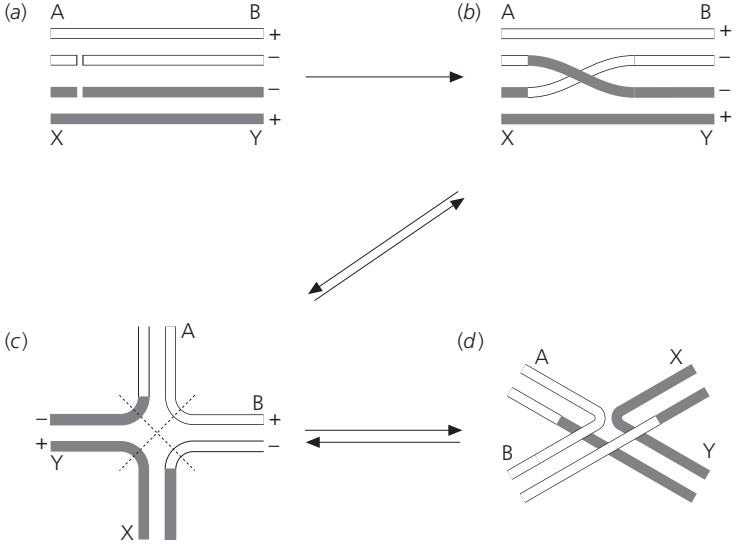


Figure 1.5 Formation of a Holliday junction during homologous recombination. Two double-stranded molecules (shown in different shades) are aligned and, after cleavage of one strand of each duplex (*a*), a process of strand exchange takes place (*b*). This intermediate may be redrawn so that it resembles a cruciform (*c*). Resolution of this intermediate can occur by cleavage and strand exchange across either diagonal in (*c*). Experiments have shown that, in the presence of magnesium ions, this intermediate probably adopts a scissor-like conformation (59) (*d*).

established with the intact strand of the other helix (30). Resealing of the DNA strands results in the formation of the Holliday junction, which may be redrawn so that it resembles a cruciform without the two unpaired DNA loops (*Figure 1.5c*). During recombination, subsequent cleavage of the second strand of each helix and resealing of the broken ends will generate recombinant DNA molecules; this process is known as resolution.

One way in which Holliday junctions have been studied is by the synthesis of four separate oligonucleotides which, when base-paired, form an artificial Holliday junction (31). In the presence of Mg^{2+} these synthetic junctions have been shown to adopt a 'scissor-like' conformation stabilized by base stacking across the four-way junction (*Figure 1.5d*). The identity of the bases at the four-way junction determines which arms of the structure are co-linear. Significant new information on the structure of Holliday junctions has been provided from crystal structures. First, the structures of Holliday junctions complexed with the protein RuvA have been determined (32, 33). RuvA is one of three proteins, RuvABC, known to be involved in the stabilization and processing of Holliday junction intermediates in *Escherichia coli* (30). These structures suggest that the protein-bound Holliday junction is planar with the four arms lying at 90° to each other (*Figure 1.6*). Branch migration (movement of the junction) in this structure is thought to occur as DNA is pumped out through the central cavity of the two opposing RuvB rings, driven by ATP hydrolysis. Resolution occurs following DNA cleavage by the RuvC protein. Second, the crystal structures of several four-way junctions have been determined (34, 35). These structures suggest that the naked Holliday junction is a right-handed

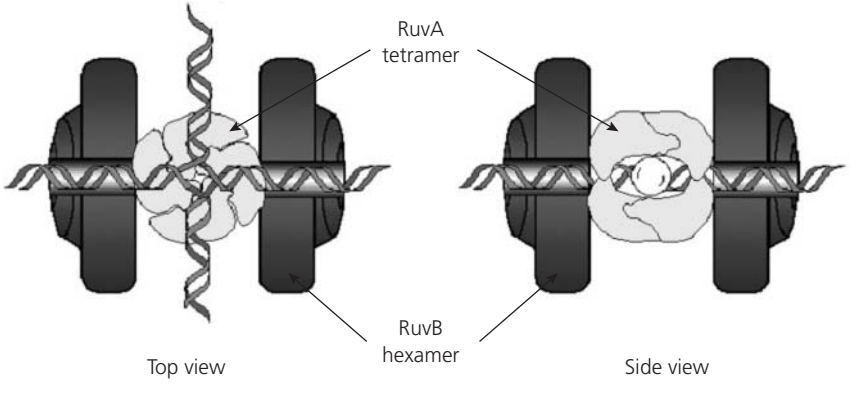


Figure 1.6 Model for the RuvAB–Holliday junction. RuvB hexamers (dark grey) surround the DNA on either side of the RuvA tetramer (light grey). DNA (ribbons) migrates through RuvA and out through RuvB, driven by ATP hydrolysis (60) (figure kindly provided by Dr. S.C. West; redrawn with permission from ref. 61; copyright (1998) Nature Publishing Group).

cross of pairwise coaxially stacked arms, similar to that shown in *Figure 1.5d*; it is likely that this structure is distorted during processing by Holliday junction-resolving proteins.

1.3.2 *DNA triplexes and H-DNA*

DNA triplexes represent an interesting alternative to the DNA double helix. In these structures a DNA duplex associates with another DNA single strand in either a parallel or antiparallel orientation (*Figure 1.7*) to form a triple-stranded structure. One manifestation of a triple helix is 'H-DNA', so called because of its requirement for protons, but it can also be thought of as 'hinged' DNA. H-DNA was first detected by the unusual sensitivity of certain plasmids to the nuclease S1 (36), attributable to the repeating DNA copolymer $(dT–dC)_n \cdot (dA–dG)_n$ (36, 37). Although the exact structure is not known, it is thought to consist of a triple-stranded and a single-stranded

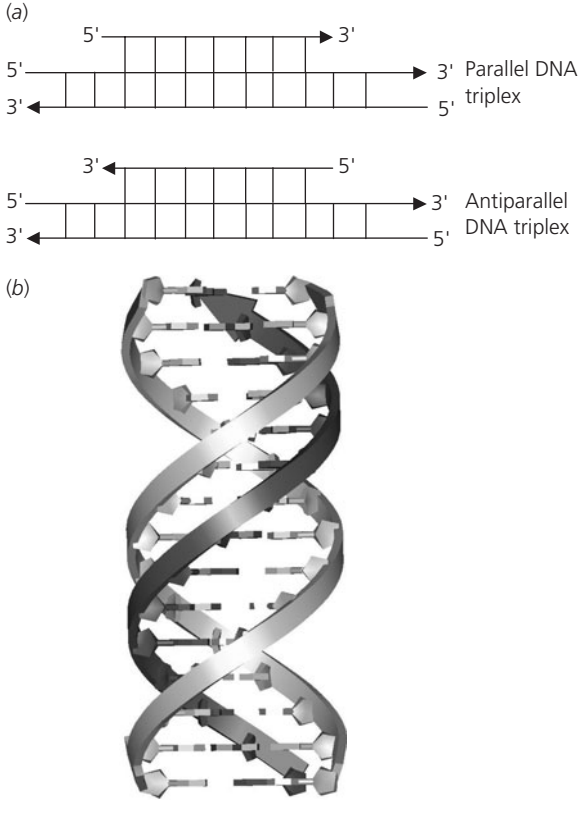


Figure 1.7 DNA triplexes. (*a*) Shows the strand orientations for parallel and antiparallel triplexes; (*b*) is a schematic view of a parallel DNA triplex, with the third (pyrimidine) strand in darker shading (figures (*a*) and (*b*) redrawn from ref. 58; courtesy of Prof. S. Neidle).

region. The essence of this structure is that a polypyrimidine strand dissociates from the Watson–Crick duplex and lies in the major groove of another section of duplex making non-Watson–Crick base pairs with the purine bases. The original polypurine partner to the polypyrimidine strand remains single stranded. This structural transition is favoured by negative supercoiling (due to the consequent loss of twist, see Chapter 2, Section 2.8.1 and Chapter 6, Section 6.1.1) and low pH (which favours the protonation required for alternative base pairs). H-DNA can therefore be regarded as an intramolecular DNA triplex; such structures have been known since the 1950s (38).

The essential features of a DNA triplex are as follows: They consist of either one purine and two pyrimidine strands (YR∗Y) or one pyrimidine and two purine strands (YR∗R) and are stabilized by Hoogsteen base pairs (*Figure 1.8*). Hoogsteen base pairs involve alternative hydrogen bond donor and acceptor partners to the standard Watson–Crick base pairs: A protonated C forms two hydrogen bonds to the N7 and O6 of G, and T forms hydrogen bonds to the N7 and 6-amino group of A (*Figure 1.8*). YR∗Y triplexes have a pyrimidine-rich third strand bound parallel to the duplex purine strand and include T·AT and C^+·GC triplets, whereas YR∗R triplexes have a purine-rich third strand bound in an antiparallel orientation and include G·GC, A·AT, and T·AT triplets (*Figure 1.8*). Parallel triplexes generally require low pH conditions (to protonate cytosines in the third strand), whereas antiparallel triplexes are pH-independent. Triplexes in DNA can be inter- or intramolecular and it is the latter which gives rise to the phenomenon of H-DNA. Intermolecular triplexes have recently been the subject of much attention, largely because of their therapeutic potential (39), stimulated by the realization that the third strand can be used to target a stretch of DNA (or RNA) in a sequence-specific manner.

1.3.3 *Telomeres and DNA quadruplexes*

Telomeres are specialized nucleoprotein structures found at the ends of linear eukaryotic chromosomes (40). They are important in chromosome stability and are added to the chromosome by the enzyme telomerase. Telomere length is thought to play a role in tumour formation and ageing. Telomeres contain unusual DNA sequences characterized by G-rich and C-rich strands, with a single-stranded 3′-overhanging region. The repeating Gs in the single-stranded section can associate into tetrads (*Figure 1.9a*) that can form parallel or antiparallel DNA quadruplexes (41, 42). The sequence from the 3′ overhang of the telomere from the ciliate *Oxytricha nova* d(GGGGTTTTGGGG) has been studied by both X-ray crystallography and NMR (42) and shown

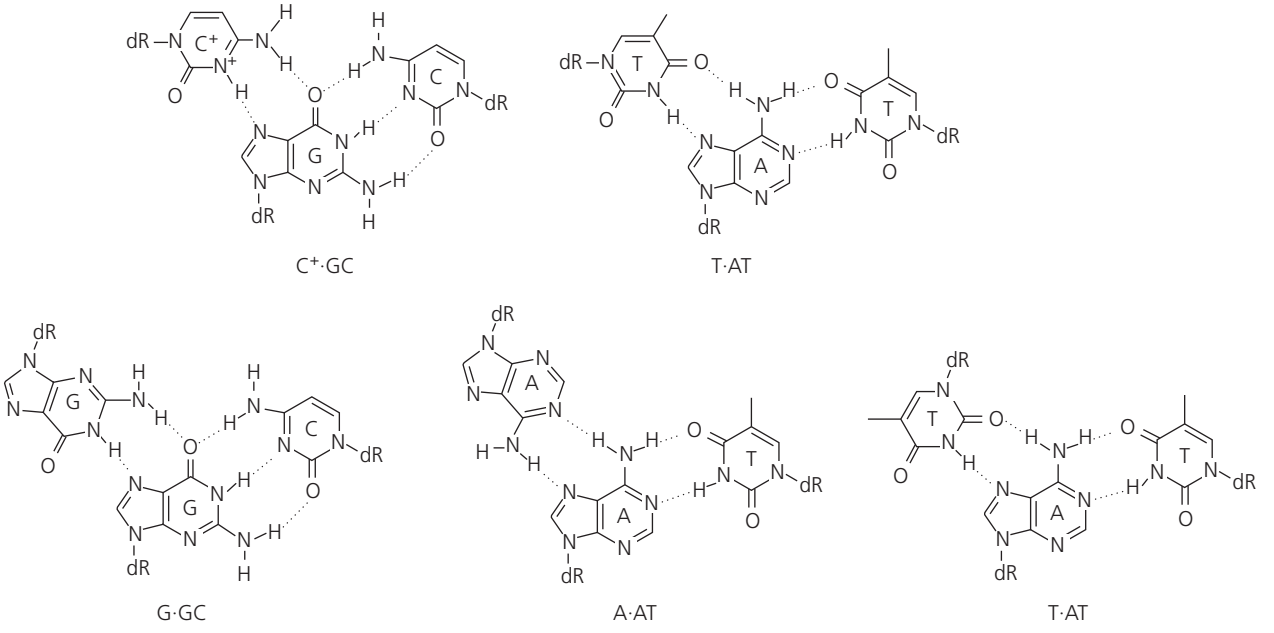


Figure 1.8 Examples of hydrogen bonding in DNA triplexes (reproduced from ref. 39).

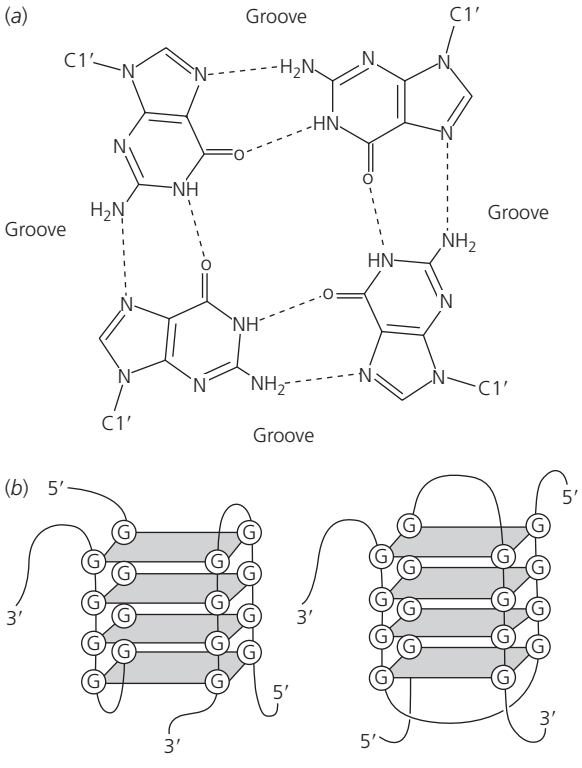


Figure 1.9 Quadruplex DNA. (*a*) Arrangement of hydrogen bonds in the guanine quartet. (*b*) Two possible folding arrangements for the quadruplex formed from two strands with the sequence d(GGGGTTTTGGGG) (figure from ref. 58, courtesy of Prof. S. Neidle).

to form a four-stranded structure (*Figure 1.9*). This oligonucleotide forms hairpins, two of which make the four-stranded structure. The Ts form the loops of the structure while the Gs are held together by hydrogen bonding with a K^+ or Na^+ ion located at the centre of the G quartets.

The counterpart of G-rich DNA, namely C-rich DNA, occurs both in telomeres and centromeres and has also been found to be able to form four-stranded structures known as i-motifs or i-tetraplexes (40, 41). In these structures two C-duplexes (each with one C-strand protonated) intercalate with each other. The presence of protonated C residues means that this structure is favoured by low pH and may not form *in vivo*. However, it has been proposed that DNA supercoiling may stabilize such structures.

1.4 Intrinsic curvature and DNA flexibility

It is apparent from the above sections that DNA can exhibit a high degree of conformational variability. One of the most striking manifestations of

this is the bending of the helix axis. This is observed in two related but distinct phenomena, intrinsic curvature of DNA and DNA flexibility. Intrinsic curvature refers to a bending of the DNA helix axis that is the preferred conformation of a particular DNA sequence. DNA flexibility refers to the ease with which certain DNA sequences can be bent, for example by being wrapped around a protein. A flexible DNA sequence will not necessarily possess any intrinsic curvature in free solution.

1.4.1 *Intrinsic curvature*

From crystallographic studies of DNA oligonucleotides, it seems that many DNA sequences may have a small degree of sequence-dependent bending (Section 1.2.2). However, significant examples of intrinsic curvature of DNA can be demonstrated by polyacrylamide gel electrophoresis. Curved DNA molecules show reduced electrophoretic mobilities when compared with straight molecules of the same size. Naïvely this can be correlated with the sieving effect of polyacrylamide gels, whereby curved molecules will pass through the pores less easily than their straight counterparts. Intrinsic curvature can arise in DNA sequences that have short runs of adenines periodically repeated with a spacing of ~10–11 bp (i.e. the helical repeat of B-form DNA). These runs of As (A-tracts) show a structure distinct from random sequence DNA and, when repeated in phase with the helical repeat, result in stable DNA curvature. Such sequences occur naturally, for example, in the kinetoplast DNA of trypanosomes, where the DNA molecules form small DNA circles (down to 700 bp in size) (43). Indeed it is possible to synthesize artificial DNA sequences with phased A-tracts and make DNA circles in the test tube as small as 105 bp (44).

A convincing demonstration of the effect of phased A-tracts in DNA curvature was given by the following experiment (45). A series of 423 bp DNA fragments was prepared from the kinetoplast DNA of trypanosomes such that they were circularly permuted, that is, they all had the same sequence but started and finished at different positions (e.g. ABCDE, BCDEA, CDEAB, etc.). Following polyacrylamide gel electrophoresis it was found that this series of molecules, although identical in molecular weight, differed significantly in their electrophoretic mobilities. These differences could be correlated with the position of a region containing phased A-tracts (*Figure 1.10*); when this region was close to the centre of a DNA fragment the electrophoretic mobility was a minimum, whereas when it was close to the end the mobility was a maximum (*Figure 1.10*). Such experiments identified the phased A-tracts as the source of intrinsic curvature in these DNA molecules. Other experiments have probed

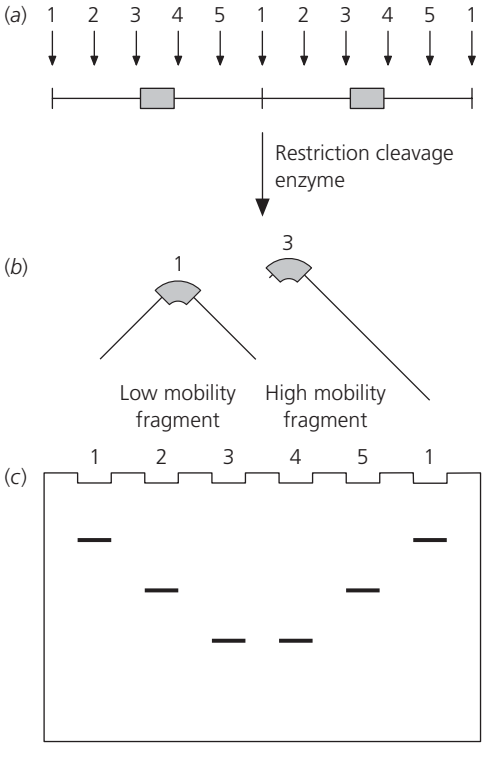


Figure 1.10 Demonstration of the effect of A-tracts in DNA curvature. DNA molecules containing phased A-tract sequences (shaded boxes) are cut with a series of restriction enzymes (1–5 in (*a*)) producing DNA fragments with the A-tracts at different positions (*b*). When analysed by polyacrylamide gel electrophoresis (*c*), the fragments show varying mobility depending on the position of the A-tract sequences.

the spacing of the A-tracts and concluded that curvature is maximal when the spacing is close to the helical repeat of B-DNA (46).

The origin of intrinsic curvature in DNA has been the subject of some controversy. The 'wedge' model suggests that A-tract bending is caused by kinking of the helix axis towards the minor groove in these regions. The 'junction' model suggests that bending is due to an abrupt change in the direction of the helix axis at the junction between the A-tract and random sequence B-DNA (47). Analysis of crystal structures of A-tract-containing DNA shows that such sequences tend to have a narrow minor groove and highly propeller-twisted base pairs, such that interstrand bifurcated hydrogen bonds are possible (48). This propeller twisting has been suggested to be important for the ability of such sequences to induce DNA curvature. Whatever the precise origin of a particular DNA bend it is clear that such features can have a profound effect on DNA–protein interactions (see below). In addition, the presence of an intrinsically curved segment of DNA in a supercoiled closed-circular DNA molecule can influence the position of loops in plectonemically supercoiled DNA (49, 50) (see Chapter 2).

1.4.2 *DNA flexibility*

DNA flexibility can be of two types: isotropic and anisotropic. Isotropic flexibility means that the DNA molecule can bend equally in all directions, whereas anisotropic flexibility means that the DNA has 'hinges' at which it can bend in a preferred direction. Flexibility is a function of the DNA sequence and there can be large differences in the relative ease with which different DNA sequences can be bent. DNA flexibility is relevant to DNA–protein interactions as many DNA-binding proteins bend DNA (see Section 1.4.3). In addition to bending flexibility, DNA also exhibits torsional flexibility, that is, flexibility resulting from a twisting motion about the DNA axis, leading to variations in the helical repeat of the DNA (Section 1.2.2). Both types of motion are important in the understanding of DNA supercoiling, and are considered further in Chapter 2. The physical origin of sequence-dependent DNA flexibility is complex and results from the chemical and stereochemical inhomogeneity of the DNA sequence.

1.4.3 *Protein-induced DNA bending*

DNA curvature and flexibility is particularly significant in relation to DNA-binding proteins. A striking example comes from studies of the *E. coli* cAMP receptor protein (CRP), also known as CAP (catabolite activator protein). CRP activates the transcription of a number of genes in response to carbon source limitation, which elevates cellular cAMP levels. This results in the formation of the CRP–cAMP complex, which binds to specific DNA sequences at the target promoters (51). In certain cases, CRP can also act as a repressor. A variety of biochemical data has suggested that the interaction of CRP with DNA induces bending of the DNA helix. This has been confirmed by the X-ray-derived structure of the CRP–DNA complex, which reveals a bend of $\sim 90^\circ$ in the DNA (52). Interestingly, the bend is not smooth, and is mostly attributable to two 40–45° kinks at symmetrically located pyrimidine–purine steps. Gartenburg and Crothers have also shown that changes in DNA sequence outside the consensus CRP binding site can change the binding affinity 10-fold, and alter the bending angle by up to 30° (53). Both these effects can be attributed to the influence of certain DNA sequences on the ease with which the DNA can be bent. It is thought that the bending of the DNA is involved in transcriptional activation by CRP, perhaps by facilitating the interaction of upstream sequences with RNA polymerase. Such an idea is supported by experiments showing that CRP binding sites can be replaced by A-tract sequences, which can impart

intrinsic curvature to DNA (see Section 1.4.1), resulting in transcriptional activation when the bend is in the same direction as that formed by the CRP complex (54).

The TATA-box binding protein (TBP) is a highly conserved eukaryotic transcription factor (see Chapter 6, Section 6.4.2). Crystal structures have shown that the complex of this protein with DNA involves a bend angle of ~100° (11, 12). The local DNA sequence can have a profound effect upon the bend angle and transcriptional activation by TBP. For example, substitution of the sequence TATAAAA by TAAAAAA dramatically reduces the bend angle and transcriptional activity (55). This effect can be attributed to the greater rigidity of the TAAAAAA sequence caused by the loss of the TpA step, which is proposed to be especially flexible.

The bacteriophage 434 repressor protein has been shown to bend its 14 bp operator DNA sequence (56). The central four bases of the operator sequence, d(ATAT), are not directly in contact with the protein. If these bases are replaced by the sequence d(AGCT) the repressor binding affinity is weakened by a factor of 50. Replacement by d(AAAA) increases repressor affinity fivefold. These changes in affinity may be correlated with the ease of flexure of the central four base pairs, although other factors, such as an alteration in DNA twist, are also thought to be involved (see Chapter 6, Section 6.4.2).

DNA flexibility has a profound effect on the binding of proteins that wrap DNA in the DNA–protein complex. A good example of this is the nucleosome where 146 bp of DNA are wrapped in 1.8 turns around the histone protein octamer (57). Although the nucleosome, by its very nature, can be regarded as non-specific in terms of DNA sequence, experiments have shown that there are preferences for certain sequence motifs within the 146 bp region. These observations are likely to reflect the anisotropic flexibility of certain DNA sequences (see Chapter 3, Section 3.5).

1.5 Conclusions

DNA is often thought of as being rather a rigid, uniform structure (perhaps even a bit boring!), which serves as a passive repository of genetic information. However, experimental work has revealed great structural diversity in DNA, in terms of different double-helical forms and other conformational states, many of which have important biological consequences. As the following chapters will show, when considered in terms of its higher-order structure, DNA displays a further diverse range of conformational variations.

1.6 Further Reading

Blackburn, G.M. and Gait, M.J. (1996). *Nucleic acids in chemistry and biology*. Oxford University Press, Oxford.

Dickerson, R.E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906–1926.

Hagerman, P.J. (1992). Straightening out the bends in curved DNA. *Biochim. Biophys. Acta* **1131**, 125–132.

Hartmann, B. and Lavery, R. (1996). DNA structural forms. *Q. Rev. Biophys.* **29**, 309–368.

Nature, 23 January 2003, pp. 395–453. A series of articles celebrating the 50th anniversary of the DNA double helix.

Neidle, S. (1999). *Oxford handbook of nucleic acid structure*. Oxford University Press, Oxford.

Neidle, S. (2002). *Nucleic acid structure and recognition*. Oxford University Press, Oxford.

Ohyama, T. (2001). Intrinsic DNA bends: an organizer of local chromatin structure for transcription. *Bioessays* **23**, 708–715.

Shafer, R.H. (1998). Stability and structure of model DNA triplexes and quadruplexes and their interactions with small ligands. *Prog. Nucleic Acid Res. Mol. Biol.* **59**, 55–94.

Sinden, R.R. (1994). *DNA structure and function*. Academic Press, London.

Ulyanov, N.B. and James, T.L. (1995). Statistical analysis of DNA duplex structural features. *Methods Enzymol.* **261**, 90–115.

1.7 References

1. Watson, J.D. and Crick, F.H.C. (1953). A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.
2. Dickerson, R.E. (1999). Helix structure and molecular recognition by B-DNA. In *Oxford handbook of nucleic acid structure*. Neidle, S. (ed.), Oxford University Press, Oxford, pp. 145–197.
3. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758.
4. Blackburn, G.M. and Gait, M.J. (1996). *Nucleic acids in chemistry and biology*, 2nd edn. Oxford University Press, Oxford.
5. Neidle, S. (1999). *Oxford handbook of nucleic acid structure*. Oxford University Press, Oxford.
6. Franklin, R.E. and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741.
7. Wahl, M.C. and Sundaralingam, M. (1999). A-DNA duplexes in the crystal. In *Oxford handbook of nucleic acid structure*. Neidle, S. (ed.), Oxford University Press, Oxford, pp. 117–144.

8. Shakked, Z., Rabinovich, D., Cruse, W.B.T., Egert, E., Kennard, O., Sala, G., Salisbury, S.A., and Viswamitra, M.A. (1981). Crystalline A-DNA: the x-ray analysis of the fragment d(G-G-T-A-T-A-C-C). *Proc. R. Soc. Lond.* **B213**, 479–487.
9. Fairall, L., Martin, S., and Rhodes, D. (1989). The DNA binding site of the *Xenopus* transcription factor IIA has a non-B-form structure. *EMBO J.* **8**, 1809–1817.
10. Weston, S.A., Lahm, A., and Suck, D. (1992). X-ray structure of the DNase I-d(GGTATACC)$_2$ complex at 2.3 Å resolution. *J. Mol. Biol.* **226**, 1237–1256.
11. Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.
12. Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527.
13. Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686.
14. Drew, H.R., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R.E. (1980). High-salt d(CpGpCpG), a left handed Z′ DNA double helix. *Nature* **286**, 567–573.
15. Pohl, F.M. and Jovin, T.M. (1972). Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG–dC). *J. Mol. Biol.* **67**, 375–396.
16. Herbert, A. and Rich, A. (1996). The biology of left-handed Z-DNA. *J. Biol. Chem.* **271**, 11595–11598.
17. Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P.O., and Zhao, K. (2001). Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309–318.
18. Klysik, J., Stirdivant, S.M., Larson, J.E., Hart, P.A., and Wells, R.D. (1981). Left-handed DNA in restriction fragments and a recombinant plasmid. *Nature* **290**, 672–677.
19. Singleton, C.K., Klysik, J., Stirdivant, S.M., and Wells, R.D. (1982). Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* **299**, 312–316.
20. Peck, L.J., Nordheim, A., Rich, A., and Wang, J.C. (1982). Flipping of cloned d(pCpG)$_n$ · d(pCpG)$_n$ DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proc. Natl. Acad. Sci. USA* **79**, 4560–4564.
21. Wang, J.C., Peck, L.J., and Becherer, K. (1982). DNA supercoiling and its effects on DNA structure and function. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 85–91.
22. Lilley, D.M.J. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc. Natl. Acad. Sci. USA* **77**, 6468–6472.
23. Murchie, A.I.H. and Lilley, D.M.J. (1992). Supercoiled DNA and cruciform structures. *Meth. Enzymol.* **211**, 158–180.
24. Lilley, D.M.J. and Hallam, L.R. (1984). Thermodynamics of the ColE1 cruciform. Comparisons between probing and topological experiments using single topoisomers. *J. Mol. Biol.* **180**, 179–200.
25. Courey, A.J. and Wang, J.C. (1983). Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* **33**, 817–829.

26. Gellert, M., O'Dea, M.H., and Mizuuchi, K. (1983). Slow cruciform transitions in palindromic DNA. *Proc. Natl. Acad. Sci. USA* **80**, 5545–5549.
27. Dayn, A., Malkhosyan, S., and Mirkin, S.M. (1992). Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res.* **20**, 5991–5997.
28. Allers, T. and Leach, D.R.F. (1995). DNA palindromes adopt a methylation-resistant conformation that is consistent with DNA cruciform or hairpin formation *in vivo*. *J. Mol. Biol.* **252**, 70–85.
29. Leach, D.R.F. (1994). Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays* **16**, 893–900.
30. West, S.C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213–244.
31. Lilley, D.M.J. and Clegg, R.M. (1993). The structure of the four-way junction in DNA. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 299–328.
32. Hargreaves, D., Rice, D.W., Sedelnikova, S.E., Artymuik, P.J., Lloyd, R.G., and Rafferty, J.B. (1998). Crystal structure of *E. coli* RuvA with bound DNA Holliday junction at 6 Å resolution. *Nature Struct. Biol.* **5**, 441–446.
33. Roe, S.M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I.R., and Pearl, L.H. (1998). Crystal structure of an octameric RuvA-Holliday junction complex. *Mol. Cell.* **2**, 361–372.
34. Nowakowski, J., Shim, P.J., Prasad, G.S., Stout, C.D., and Joyce, G.F. (1999). Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10–23 DNA enzyme. *Nature Struct. Biol.* **6**, 151–156.
35. Ortiz-Lombardía, M., González, A., Eritja, R., Aymaní, J., Azorín, F., and Coll, M. (1999). Crystal structure of a DNA Holliday junction. *Nature Struct. Biol.* **6**, 913–917.
36. Hentschel, C.C. (1982). Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. *Nature* **295**, 714–716.
37. Johnson, B.H. (1988). The S1-sensitive form of $d(C\text{-}T)_n \cdot d(A\text{-}G)_n$: chemical evidence for a 3-stranded structure in plasmids. *Science* **241**, 1800–1804.
38. Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995). Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95.
39. Gowers, D.M. and Fox, K.R. (1999). Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.* **27**, 1569–1577.
40. Rhodes, D. and Giraldo, R. (1995). Telomere structure and function. *Curr. Opin. Struct. Biol.* **5**, 311–322.
41. Gilbert, D.E. and Feigon, J. (1999). Multistranded DNA structures. *Curr. Opin. Struct. Biol.* **9**, 305–314.
42. Neidle, S. and Parkinson, G.N. (2003). The structure of telomeric DNA. *Curr. Opin. Struct. Biol.* **13**, 275–283.
43. Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982). A bent helix in kinetoplast DNA. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 279–283.
44. Ulanovsky, L., Bodner, M., Trifonov, E.N., and Choder, M. (1986). Curved DNA: design, synthesis and circularisation. *Proc. Natl. Acad. Sci. USA* **83**, 862–866.

45. Wu, H.-M. and Crothers, D.M. (1984). The locus of sequence directed and protein-induced DNA bending. *Nature* **308**, 509–513.
46. Hagerman, P. (1985). Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry* **24**, 7033–7037.
47. Crothers, D.M., Haran, T.E., and Nadeau, J.G. (1990). Intrinsically bent DNA. *J. Biol. Chem.* **265**, 7093–7096.
48. Crothers, D.M. and Shakked, Z. (1999). DNA bending by adenine-thymine tracts. In *Oxford handbook of nucleic acid structure.* Neidle, S. (ed.), Oxford University Press, Oxford, pp. 455–470.
49. Laundon, C.H. and Griffith, J.D. (1988). Curved DNA segments can uniquely orient the topology of supertwisted DNA. *Cell* **52**, 545–549.
50. Yang, Y., Westcott, T.P., Pederson, S.C., Tobias, I., and Olson, W.K. (1995). Effects of localized bending on DNA supercoiling. *Trends Biochem. Sci.* **20**, 313–319.
51. de Crombrugghe, B., Busby, S., and Buc, H. (1984). Cyclic AMP receptor protein: role in transcription activation. *Science* **224**, 831–838.
52. Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253**, 1001–1007.
53. Gartenberg, M.R. and Crothers, D.M. (1988). DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* **333**, 824–829.
54. Gartenberg, M.R. and Crothers, D.M. (1991). Synthetic DNA bending sequences increase the rate of in vitro transcription initiation at the *Escherichia coli* lac promoter. *J. Mol. Biol.* **219**, 217–230.
55. Starr, D.B., Hoopes, B.C., and Hawley, D.K. (1995). DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* **250**, 434–446.
56. Koudelka, G.B., Harbury, P., Harrison, S.C., and Ptashne, M. (1988). DNA twisting and the affinity of 434 operator for bacteriophage 434 repressor. *Proc. Natl. Acad. Sci. USA* **85**, 4633–4637.
57. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
58. Neidle, S. (2002). *Nucleic acid structure and recognition.* Oxford University Press, Oxford.
59. Duckett, D.R., Murchie, A.I.H., Diekmann, S., von Kitzing, E., Kemper, B., and Lilley, D.M.J. (1988). The structure of the Holliday junction and its resolution. *Cell* **55**, 79–89.
60. West, S.C. (1998). RuvA gets X-rayed on Holliday. *Cell* **94**, 699–701.
61. van Gool, A.J., Shah, R., Mézard, C., and West, S.C. (1998). Functional interactions between the Holliday junction resolvase and the branch-migration motor of *Escherichia coli. EMBO J.* **17**, 1838–1845.

This page intentionally left blank

2
DNA supercoiling

2.1 Introduction

In the first chapter, we described the classical double-helical structure of DNA, and the local alterations of that structure that can occur as a consequence of specific nucleotide sequences. We hope you now have a good feel for these concepts, and we will move on to the main subject of this book, the supercoiling of DNA, a global alteration of structure that arises directly from the double-helical nature of the molecule.

2.2 Historical perspective

2.2.1 *The problem of the double helix*

The double-helical structure of DNA is commonly seen as a rather elegant structure adapted, with its complementary strands, to the purpose of DNA replication. However, as we have seen, the helical nature of DNA is probably not an evolved feature in itself, but an inevitable consequence of the formation of a polymer of asymmetric subunits (Chapter 1, Section 1.2.1). If we start to consider the process of replication, we can see a possible problem. In fact it turns out that DNA replication would work much more conveniently with a pair of complementary straight strands (like a ladder), an alternative that was considered and dismissed early on (1).

DNA is a very long molecule in comparison to its diameter, and in order for it to be replicated the two strands must be separated to act as templates for the new daughter strands. However, pulling apart the two strands of the double helix requires the helix to rotate ahead of the separating strands, that is, ahead

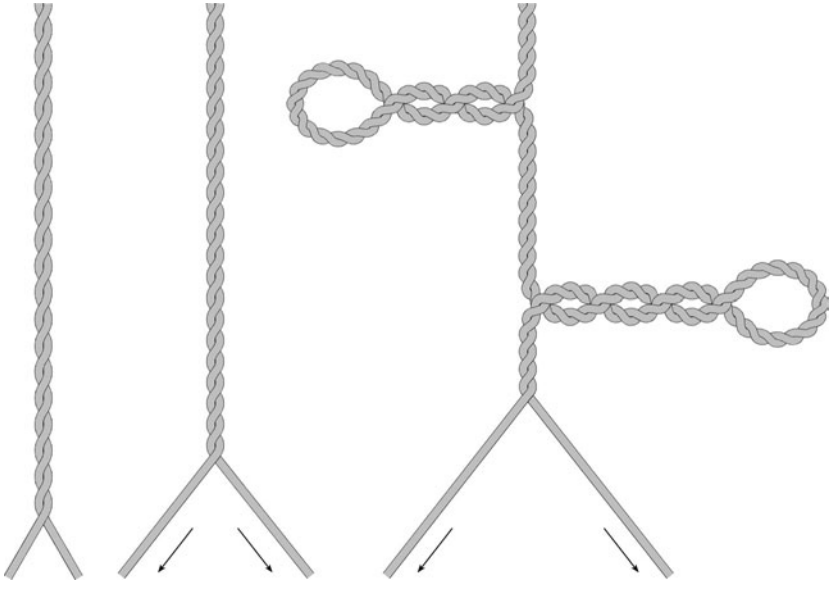


Figure 2.1 Pulling apart a long twisted rope: an analogy of DNA replication. An attempt to pull apart the strands of a twisted rope results in increased twisting of the strands and ultimately in the coiling of the rope upon itself.

of the replication fork. We can imagine what will happen if we think of a more familiar analogy. Imagine trying to pull apart the strands of a long twisted rope. The twisting of the rope will tighten ahead of the separating strands, and eventually the rope will tend to coil upon itself (*Figure 2.1*). If the whole rope cannot rotate to relieve the stress that builds up (and friction will generally prevent it), this process will soon stop the further separation of the strands. This coiling and tangling of the rope (or DNA) is the process of supercoiling. Supercoiling is the inevitable result of trying to manipulate the double-helical structure of DNA.

This problematic aspect of DNA was discussed by James Watson and Francis Crick in two papers in 1953 very soon after their description of the structure (1, 2), and possible solutions to the problem were considered in detail by Max Delbrück and Gunther Stent (3). Among the solutions considered was the straightforward one that the whole DNA helix rotates along its axis ahead of the separating strands. This would be fairly easy to envisage if the DNA were a reasonably rigid linear rod, less easy if the long DNA helix were bent and coiled in the cell or nucleus, and in fact is impossible since, as we now know, DNA in cells is either circular (see Section 2.2.2), or constrained from rotating by being tethered at intervals to organizing structures in the cell

(see Section 2.5; Chapter 6, Section 6.1.1). One of the alternatives considered (which subsequently turned out to be the correct one), was that one or both DNA strands are transiently broken and rejoined ahead of the replication fork to allow local rotation to occur at the break point (4). We now know that this breaking and rejoining is carried out by the DNA topoisomerases, enzymes that manipulate the supercoiling of DNA; the details of these enzymes will be discussed in Chapter 5. So, although the early workers on DNA structure did not explicitly formulate the idea of DNA supercoiling, they were very quickly aware that the double helix structure imposed important constraints on replication.

2.2.2 *Closed-circular DNA and supercoiling*

In the early 1960s, a number of groups were investigating the DNA molecules of viruses, in particular the DNA tumour virus, polyoma. DNA from polyoma virus consistently fractionated into two components, I and II, during sedimentation analysis, a technique that separates molecules according to size and compactness. Both components were shown to consist of double-stranded DNA of the same molecular weight. However, the major component (I) had a higher sedimentation coefficient (i.e. was more compact), and was unusually resistant to denaturation (the melting of base pairing between the strands to give single-stranded DNA; see Chapter 1, Section 1.2.1) on heating or exposure to high pH. Furthermore, following denaturation, the single strands did not separate from each other. This led to the suggestion (5, 6) that component I consisted of 'circular base-paired duplex molecules without chain ends', and component II was the linear form of the same molecule. In other words, the DNA molecules in component I consist of two antiparallel circular DNA single strands winding helically around one another, that is, the ends of each strand of a linear double-stranded molecule are joined in the conventional 5′ to 3′ manner to form a double-stranded circle (*Figure 2.2*). This configuration is now known as closed-circular DNA.

This interpretation was supported when electron micrographs of polyoma DNA showed mostly circular molecules. However, a problem arose with the discovery that a single break in one of the two strands of component I, caused by the endonuclease DNase I, converted it directly to component II. How could the cleavage of one strand of a closed-circular duplex convert it to a linear molecule? Subsequently, electron micrographs of the component II molecules formed in this way showed them also to be circular. So, how could two circular DNA molecules, I and II, distinguished only by the breakage of one backbone phosphodiester bond, have such different properties?

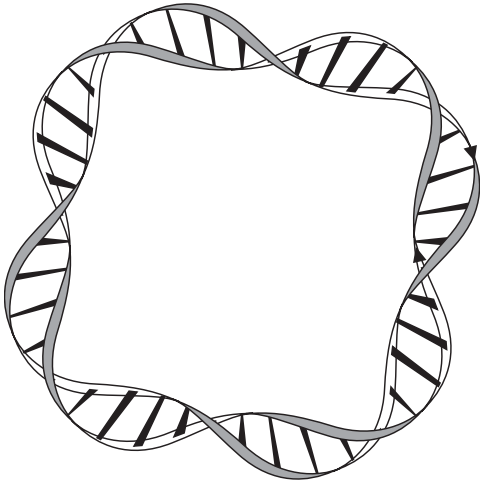


Figure 2.2 Closed-circular DNA. A schematic representation of a closed-circular DNA molecule. Each of the single stands of a double-stranded DNA double helix is covalently closed via a phosphodiester bond to form a circular double-stranded molecule.

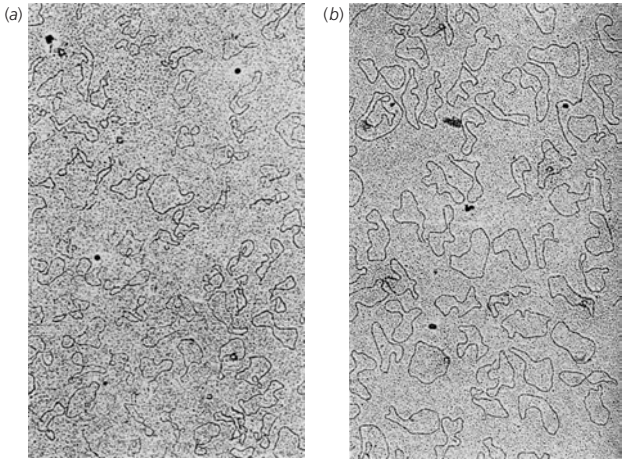


Figure 2.3 Polyoma component I and component II DNA. Electron micrographs of DNA isolated from the polyoma virus. (*a*) Component I. (*b*) Component II (reproduced from ref. 7).

The clue required to solve this conundrum came from the electron micrographs. Micrographs of the component I molecules showed many crossings of the DNA double strands (*Figure 2.3a*), whereas the component II molecules were mainly open rings (*Figure 2.3b*). Such crossings had been tentatively ascribed to protein cross-links, but Jerome Vinograd and his colleagues (7) suggested that they could result from a 'twisted circular form' of the DNA, which explained the more compact nature of component I. Such a conformation

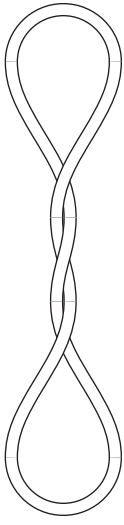


Figure 2.4 A tubing model of supercoiled DNA. A length of rubber tubing closed by a connector will adopt a conformation similar to that shown if the ends are twisted relative to each other before closure. If a double helix is imagined drawn on the tubing, this is quite a good model for the behaviour of DNA.

would result if, conceptually, before the joining of the ends of a linear duplex into a closed-circular molecule, one end were twisted relative to the other, thus introducing some strain into the molecule. This is analogous to the behaviour of any length of elastic material; such a relative twisting of the ends tends to be manifested as a coiling of the material upon itself, exactly equivalent to the twisted rope example in Section 2.2.1. One of the best models of this behaviour of DNA is a length of rubber tubing. If the unconstrained tubing represents the double helix of a linear DNA, then a relative twisting of the ends, followed by the closure of the tubing into a circle with a connector, will result, when external constraints are released, in something that looks like *Figure 2.4*. Such a coiling of the DNA helix upon itself is the literal meaning of the term supercoiling; that is, a higher-order coiling of the DNA helix. Furthermore, the supercoiling is locked into the system. If the two DNA strands are joined covalently across the original break 5′–3′, the elastic strain that results in superhelicity cannot be released without breaking one or both strands. This breakage corresponds in the tubing model to a reversal of the twisting at the break point. The DNA molecule may be geometrically contorted in a variety of ways, as we will see later, but the basic strained state of the molecule cannot be changed without strand breakage.[1]

Now we can see how the cleavage of one phosphodiester bond of the polyoma supercoiled DNA (component I) leads directly to the open-circular,

[1] Actually, this is a slight oversimplification, but it will do for now. See Section 2.5.3 for a more comprehensive version.

unconstrained component II, since the broken strand can rotate about the intact strand to dissipate the torsional stress. Component II is normally known as 'open-circular', or 'nicked-circular' DNA referring to this single break in one strand. The closure of a linear DNA into a planar circle without further constraint leads to 'relaxed' closed-circular DNA. Although nicked-circular and relaxed DNAs have a similar conformation, they differ chemically by the presence or absence of one or more single-strand breaks. The terms form I and II are still used from time to time, along with form III for linear DNA, to describe the different conformations of DNA as explained above, but these terms are really obsolete, and should be replaced by 'supercoiled', 'nicked' or 'open-circular', and 'linear'.

If you have an intuitive feel for these properties of rubber tubing, and by analogy, of DNA, then read on. If not, make a rubber tubing model and try the exercise above. This may seem to be labouring a small point, but without a firm understanding of these principles, your ability to follow the remainder of this discussion will rapidly deteriorate!

It may not have escaped your attention that the superhelical turns described may be of either handedness, corresponding to twisting in the direction tending to overtwist or untwist the DNA helix. We have already said that the right-handed DNA helix rotates in a clockwise (right-handed) direction away from an observer looking along the axis (Chapter 1, Section 1.2.1). Overtwisting of the helix corresponds to anticlockwise (left-handed) twisting of the end of the helix nearest the observer; untwisting of the helix corresponds to the opposite motion. (This is perhaps counter-intuitive and is worth verifying for yourself using a model.) The two opposite forms of supercoiling introduced in this way are designated positive (for overtwisting of the DNA helix) and negative (untwisting the helix). Vinograd correctly deduced the handedness of the supercoiling of polyoma DNA to be negative, although the supercoiling that tends to build up ahead of a replication fork (Section 2.2.1) is clearly positive.

Thus, with the powerful but (in retrospect) simple insight that a closed-circular DNA molecule could exist in a 'twisted circular' or supercoiled form, Vinograd and his colleagues neatly explained the properties of polyoma virus DNA and analogous results pertaining to other circular DNA molecules. However, they opened up a whole new field of study, since behind this straightforward idea are concealed many theoretical and practical ramifications. These range from the mysteries of linking number, twist and writhe (see below) to the processes required to untangle (often literally!) the biological consequences of these unexpected topological properties of the DNA molecule, such as the problem of unwinding during replication that we have already mentioned.

2.3 A quantitative measure of DNA supercoiling

As we said earlier, the elastic stress associated with a supercoiled closed-circular DNA molecule cannot be relieved without breaking one or both backbone strands, although the molecule can be geometrically deformed by external forces (but see Section 2.5.3). Can we devise a quantitative measure of the supercoiling of a DNA molecule?

2.3.1 *Linking number*

If a linear double-stranded DNA molecule is closed into a circle (*Figure 2.2*), by the formation of 5′–3′ phosphodiester bonds to close each strand, then the strands will be linked together a number of times corresponding to the number of double-helical turns in the original linear molecule. This number, which must be an integer, is known as the linking number of the two strands, or the linking number of the closed-circular DNA molecule, abbreviated as Lk. Earlier literature uses α, referred to as the topological winding number; this term is identical to Lk.

We can demonstrate this subtle concept by modelling DNA as a length of ribbon, as illustrated in *Figure 2.5*, where the two edges of the ribbon represent the single strands of DNA. Initially there are no helical turns, and the ribbon is closed into a simple circle (*Figure 2.5a*). This models a hypothetical situation where the two strands are totally unwound and the two ribbon edges are not linked. If we now cut the ribbon (*Figure 2.5b*), and rejoin it after making a full 360° twist in one end, we get a model of a closed-circular DNA with one double-helical turn (*Figure 2.5c*). (Remember that although a 180° twist will allow rejoining of the ribbon, in the case of DNA, this will place two 3′ and two 5′ ends together, and so is not allowed.) We can demonstrate the linking of the two strands (edges) by cutting longitudinally around the ribbon (*Figure 2.5c*), which is analogous to the separation of the two strands, to yield two circles linked together once (*Figure 2.5d*); that is with a linking number of 1 (Lk = 1). It is well worth making such a model to convince yourself of this phenomenon; the easiest materials to use are a long strip of paper and tape. In the same way, you can show that the introduction of n turns into the ribbon, corresponding to n double-helical DNA turns, leads to two circles linked n times. If n gets too large, the model becomes confusing, but colouring one edge initially will help you keep track.

Of course, a given length of real DNA has an inherent number of double-helical turns by virtue of its structure. This number is the length of the DNA, N, in base pairs (bp; usually thousands for natural DNAs) divided by the number

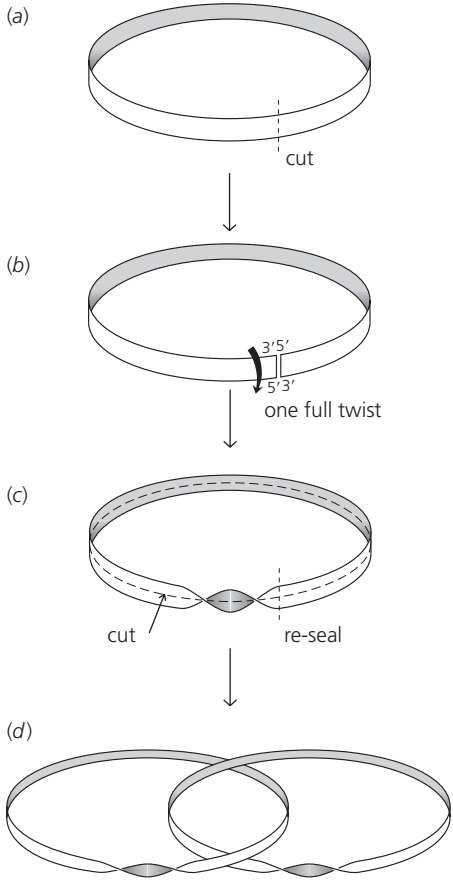


Figure 2.5 Demonstration of linking number (Lk), using a length of ribbon or paper. (*a*) A length of ribbon is taped into a circle. Each edge of the ribbon represents one strand of DNA. The edges (strands) are initially unlinked. (*b*) The ribbon is cut and a 360° turn is introduced into the ribbon, mimicking one turn of a DNA double helix. (*c*) The ribbon is rejoined, then cut longitudinally, to model the separation of the strands. (*d*) This produces two ribbons (single strands), linked together once.

of base pairs per turn of helix, h, which is dependent on the conditions, although a standard value, $h°$, is defined under standard conditions (0.2 M NaCl, pH 7, 37 °C; (8)) and is often taken to be 10.5 bp/turn (see Chapter 1, Section 1.2.2).[2] The value of N/h will not in general be an integer, so when the DNA is bent into a simple, planar circle, the strand ends will not line up precisely, although the slight twisting required to join the ends is relatively insignificant over thousands of base pairs. Hence a DNA molecule joined into a circle with the minimum of torsional stress will have a linking number that is the closest integer to N/h. We will refer to this 'standard' linking number as

[2] This variable is replaced by γ in some more recent literature, since h is also used for the base-pair rise (base-pair spacing) in double-helical DNA. We will stick with h to ensure consistency with most of the literature.

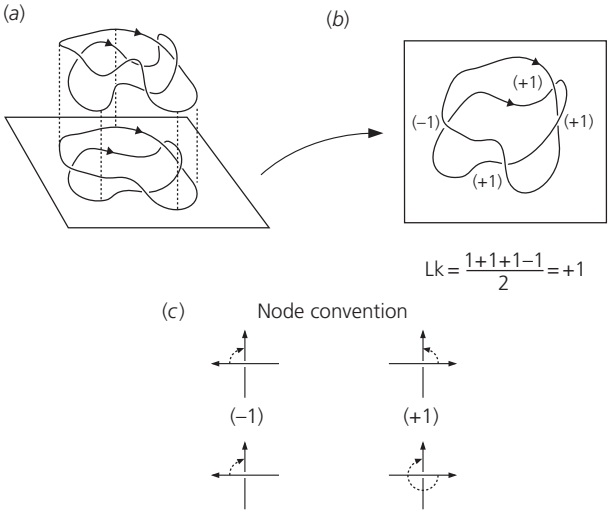


Figure 2.6 The definition of linking number (Lk). (*a*) Two closed curves are projected onto a plane (this can be in any direction). (*b*) The curves are each assigned a polarity, and the crossings of one curve over the other (nodes) are given a number (±1) according to either of the conventions in (*c*). Self-crossings of a curve do not count. The linking number is equal to half the sum of the node numbers.

Lk_m after Depew and Wang (9); hence:

$$Lk_m \approx \frac{N}{h} \approx \frac{N}{10.5} \tag{2.1}$$

The linking number of the strands of right-handed DNA is conventionally defined as positive. Thus, for example, for plasmid pBR322, with 4361 base pairs (N), in its relaxed form, Lk_m is +415 under conditions where the helical repeat (h) is 10.5 bp/turn.

Linking number turns out to be a general property of any two closed curves in three-dimensional space. Rigorously, the linking number may be determined using the number and handedness of the crossovers of the two curves when projected onto a plane. Projection onto a plane is illustrated in *Figure 2.6a*, and may be thought of as a two-dimensional representation of the three-dimensional curves viewed from any direction, except that the upper and lower strands must be distinguishable at the crossovers, which are called 'nodes'. The curves are each given a direction, and a value of +1 or −1 assigned to each crossing of one curve over the other, according to a rule involving the direction of movement required to align the upper with the lower strand with a rotation of less than 180°; clockwise is negative and counterclockwise positive (confusing, but illustrated in *Figure 2.6c*). An alternative rule is to consider only the clockwise rotation required to align the top with the bottom strand; in this case a rotation of $<180°$ corresponds to a negative node, and $>180°$ to a positive node (*Figure 2.6c*). The sum of the assigned values is halved to

give the linking number of the two curves (*Figure 2.6b*). This division by two makes sense, since a single link (e.g. *Figure 2.7a*) requires two crossings of the same handedness. Actually, it is fairly obvious that *Figure 2.6b* has a single link, and that *Figure 2.7a* represents a tidied up version of the same thing (albeit with the opposite handedness). This method works for any projection of two curves, and takes account of any fortuitous overlapping of the curves that does not constitute a link, for example, the left-hand half of *Figure 2.6b*: the signs of the two nodes cancel out. Likewise, this definition helps to show that linking number is unchanged without breakage of one of the curves. Any stretching or bending (technically, smooth deformation) of the curves may add or subtract nodes in a given two-dimensional projection, but only in self-cancelling positive and negative pairs, leaving Lk unchanged. The application of this method to a model DNA circle is shown in *Figure 2.7b*. The strands are shown running in the same direction, to ensure that the resulting Lk is positive. Of course, in a structural sense, the strands of DNA run antiparallel. This discrepancy arises because linking number is a mathematical property, defined independently of DNA, and the convention has been fixed to give right-handed DNA a positive Lk.

So, the linking number (Lk) is a measure of a fundamental property of a closed-circular DNA molecule. It has no meaning until both strands are sealed, and is unaffected by changes in the conformation of the DNA, as long as these do not involve the breakage of one or both strands. Linking number is thus a *topological* property of closed-circular DNA, and does not depend on the particular *geometry* of the DNA. DNA molecules differing only in linking number are known as topological isomers, or topoisomers.

2.3.2 *Supercoiling, linking difference, and specific linking difference*

We can now see that the process of introducing supercoiling into a DNA molecule (relative twisting or untwisting of the right-handed helix before closure of a linear molecule into a circle) reduces or increases the number of helical turns trapped by that closure and results in (necessarily integral) changes in the linking number of the final closed-circular species (*Figure 2.7c*). So, a change in supercoiling generated this way in a given molecule is accompanied by a change in the linking number. Specifically, the change in linking number from Lk_m, which corresponds to the most undistorted closed-circular (relaxed) topoisomer as discussed earlier, is a measure of the extent of supercoiling. This value, $Lk - Lk_m$, is known as the linking difference, or ΔLk, and may be positive or negative. Because the linking number of closed-circular

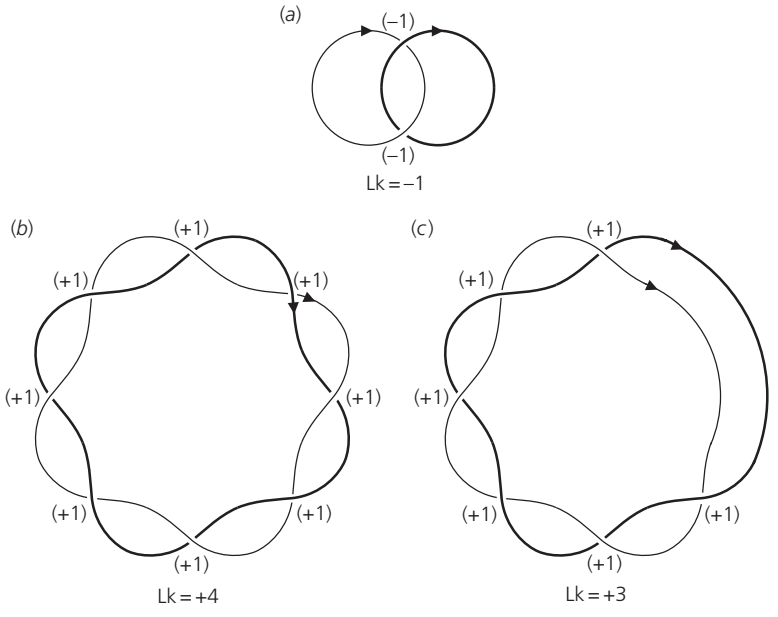


Figure 2.7 Linking number examples. (*a*) Two singly linked curves. In this case Lk = −1; the sign is determined by the directions assigned to the curves. (*b*) A hypothetical DNA molecule; the strands must be assumed to run in the same direction for the linking number to be positive. (*c*) If a DNA molecule is untwisted by one turn before closure into a circle, the linking number decreases by one.

B-DNA is defined as positive, twisting up of the helix before closure leads to an increase in linking number above Lk_m, that is a positive linking difference, and corresponds to positive supercoiling. Analogously, unwinding of the helix before closure yields a negative ΔLk and negative supercoiling.

Strictly speaking, the linking difference (ΔLk) of a supercoiled DNA molecule should be measured from the hypothetically unconstrained state corresponding to the closure of a linear molecule into a planar circle without *any* torsional strain. This is because the small twisting which may be required to form the Lk_m isomer (Section 2.3.1) should really be counted towards the total supercoiling of the molecule. This 'hypothetical' linking number, the number of double helical turns in the original linear molecule is exactly equal to N/h (cf. Equation 2.1), and need not be an integer. This quantity is known as Lk°:

$$Lk^\circ = \frac{N}{h} \tag{2.2}$$

Lk° is not a true linking number, since a linking number must be integral, but it serves as a reference point for the measurement of the level of supercoiling. Thus a better general definition of linking difference is:

$$\Delta Lk = Lk - Lk^\circ \tag{2.3}$$

For moderate supercoiling in plasmid-sized DNA molecules ($N >$ a few thousand base pairs), the distinction between Lk_m and Lk° is not too significant (the absolute difference between them is always ≤ 0.5); for example, Lk° for pBR322 is 415.3, if $h = 10.5$ bp/turn (Section 2.3.1). The physical relevance of Lk° will be discussed in Section 2.5.2.

As we have seen (Section 2.2.2), the introduction of supercoils into a DNA molecule corresponds to the introduction of torsional stress. A given ΔLk will produce more torsional stress in a small DNA molecule than in a large one; hence linking difference is commonly scaled to the size of the DNA circle, by dividing by Lk°,which is proportional to the number of bp, N, to give the specific linking difference (σ):

$$\sigma = \frac{Lk - Lk^\circ}{Lk^\circ} = \frac{\Delta Lk}{Lk^\circ} \tag{2.4}$$

This gives a measure of the extent of supercoiling that can be used for comparisons between DNA molecules. For example, natural closed-circular DNA molecules such as bacterial plasmids or the *Escherichia coli* chromosome, while varying widely in size, have when isolated a specific linking difference of around -0.06. Specific linking difference is often referred to as superhelix or superhelical density; however, these terms potentially confuse topological and geometrical properties, as we shall see, and should probably be avoided in a rigorous discussion.

2.4 Geometrical properties of closed-circular DNA

2.4.1 *Twist and writhe*

We now need a language for describing the geometrical conformations that may be adopted by closed-circular DNA in response to changes of supercoiling, or ΔLk. We have already seen that one way of imagining the generation of DNA molecules of varying Lk is to envisage a twisting up or untwisting of a linear DNA helix, followed by closure into a double-stranded circle. This suggests that one possible geometric consequence of a change in linking number is a

corresponding change in the winding of the DNA double helix itself. In the case of negatively supercoiled DNA, this corresponds to an untwisting of the helix, and an increase in the helical repeat, *h* (number of base pairs per turn).

On the other hand, using the rubber tubing model, we can see that a change in linking number from the most relaxed (Lk_m) topoisomer leads to a coiling of the tubing upon itself (*Figure 2.4*). This is the behaviour of DNA that first alerted Vinograd to the supercoiling phenomenon, as recounted above. In fact, these two behaviours are complementary to each other, and each may be defined and described. The twist (Tw) describes how the individual strands of DNA coil around one another (more rigorously, around the axis of the DNA helix). The writhe (Wr) is a measure of the coiling of the helix axis in space. Both are complex geometric functions, whose values need not be integral and may not in general be easily calculated, although, as we will show later in Section 2.4.2 and in Chapter 3, several special cases can be identified (see Box 2.1). However, the central important result for the study of DNA supercoiling is that the twist and writhe of a given DNA molecule must sum to the linking number. That is:

$$Lk = Tw + Wr \tag{2.5}$$

This implies that for a given closed-circular molecule, since Lk is invariant, any change in the twist of the DNA must be accompanied by an equal and opposite change in the writhe, and vice versa. Furthermore, any change in linking number manifests itself geometrically as a change in the twist and/or the writhe:

$$\Delta Lk = \Delta Tw + \Delta Wr \tag{2.6}$$

These results were originally proved as a piece of pure mathematics, without reference to DNA, by James White in 1969 (10), and were later independently developed by F. Brock Fuller (11, 12) as a rigorous examination of the results of Vinograd and his colleagues. More accessible discussions (for the biologist) of the technical aspects of this area are given in the references (13–15).

2.4.2 *The interconversion of twist and writhe*

The best way to get a feel for the properties of twist and writhe is again to consider the behaviour of DNA as modelled by rubber tubing. The first case to consider is when the tubing (or DNA) is closed into a planar circle (the DNA is relaxed; *Figure 2.9a*). It seems intuitive, and indeed it is the case, that the writhe

BOX 2.1 **TWIST AND WRITHE**

We can give slightly more detailed descriptions of the definitions of twist and writhe without involving too much mathematics:

Twist can be thought of as the sum of all the small rotations of one strand around the local helix axis, measured at many points along the axis. For example, one could sum all the rotations (measured in fractions of a turn) between one base pair and the next measured in a plane perpendicular to the helix axis at that point. A stricter mathematical description involves the integration along the length of the DNA axis of the instantaneous rate of rotation about the axis of the normal (perpendicular) vector joining the axis to one strand (*Figure 2.8*) (15).

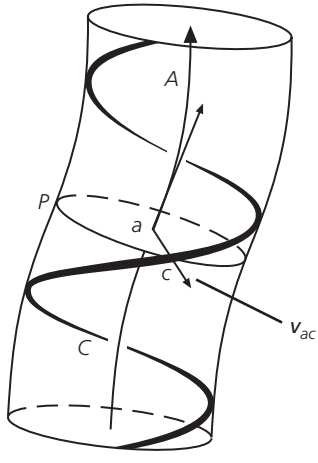


Figure 2.8 The definition of twist. The curve *C*, corresponding to one of the DNA strands, follows a helical path around the DNA axis, *A*. For any point *a* on *A*, curve *C* crosses the plane *P*, perpendicular to *A*, at point *c*. The twist is the integrated rate of rotation of vector $\boldsymbol{v}_{ac}$ in plane *P* as *a* traverses *A* (modified from ref. 15 with permission; copyright (1990) Cold Spring Harbor Laboratory Press).

Writhe can be defined in terms of nodes and crossings, related to the definition of linking number we gave above (Section 2.3.1). In any projection of the helix axis onto a plane (only one curve now), we can sum the +1 or −1 value of each of the nodes formed by the crossing of the helix axis over itself. This gives the directional writhe or two-dimensional writhe of that projection; there is no division by two in this case (16). The three-dimensional writhe of the original curve is the average of the two-dimensional writhe over all possible directions of projection. With a bit of thought it should be clear from this definition that the writhe is zero for curves that have a plane or centre of symmetry, since the two-dimensional writhe will be equal and opposite for pairs of symmetry-related directions. A curve that lies in one plane, like relaxed DNA, is the simplest zero-writhe case; another is a curve that lies on the surface of a sphere without self-intersection. This definition is useful in imagining what is happening, but can also be used to compute the writhe of a given curve or DNA, by computer simulation of projections in a large number of directions (17) (also see Section 2.7.3).

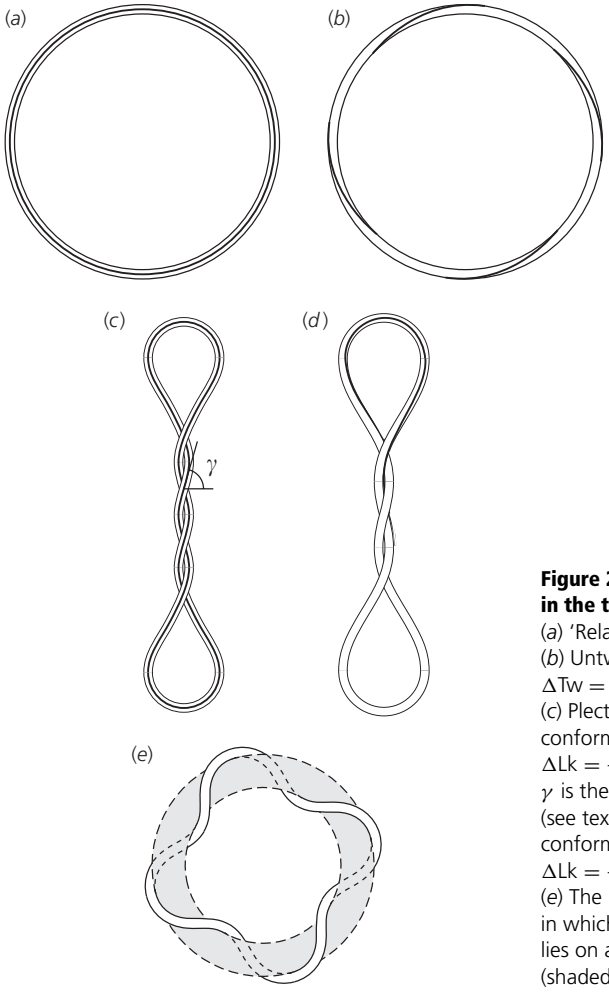


Figure 2.9 Twisting and writhing in the tubing model of DNA.
(*a*) 'Relaxed' DNA; $Lk = Lk_m$.
(*b*) Untwisted DNA: $\Delta Lk = -4$, $\Delta Tw = -4$, $Wr = 0$.
(*c*) Plectonemic, or interwound conformation of supercoiled DNA: $\Delta Lk = -4$, $\Delta Tw \approx 0$, $Wr \approx -4$. γ is the superhelix winding angle (see text). (*d*) The plectonemic conformation with lowest energy: $\Delta Lk = -4$, $\Delta Tw \approx -1$, $Wr \approx -3$. (*e*) The 'toroidal' winding of DNA, in which the axis of the DNA helix lies on a real or imaginary torus (shaded region).

of a plane circle is zero (see Box 2.1). So, in this case, from Equation (2.5), the twist of a DNA is equal to its linking number (i.e. when $Wr = 0$, $Lk = Tw$); that is, the twist is equal to the number of double-helical turns around the circle. This makes good sense (see Box 2.2).

At this point it is helpful to draw a line on the relaxed rubber tubing, as in *Figure 2.9a*, as a marker to follow changes in the twist of the tubing. If the tubing is broken and one end rotated four times in the untwisting sense (Section 2.2.2) before reclosure (the line must be continuous across the break to ensure integral turns), the tubing now models a DNA with $\Delta Lk = -4$. The model will now adopt the interwound, or 'plectonemic' conformation we

BOX 2.2 **Tw° AND Wr°**

The identity Lk = Tw is only normally true if the helix axis lies in a plane. This will not generally be the case for the most relaxed topoisomer with Lk = Lk_m; as seen in Section 2.3.1, the strand ends of a linear DNA will not normally line up precisely, since N/h will not usually be an integer, and the slight distortion required to close the circle may be manifested as a small writhe. It is therefore useful to define the term Tw°, which is the twist of the linear DNA before ring closure, or the theoretical twist of the nicked-circular(open-circular) form of the DNA. That is:

$$\mathrm{Tw}^\circ = \frac{N}{h} \tag{2.7}$$

This is identical to our previous definition of Lk° (Section 2.3.2). It is important to note, however, that Lk° was defined for a planar circle. If the unconstrained DNA does not lie in a plane, for example because it contains regions of intrinsic curvature (Chapter 1, Section 1.4.1), then Lk° may contain a contribution from the writhe corresponding to this lack of planarity. In this special case, Lk° ≠ Tw°; Lk° = Tw° + Wr° (cf. Equation (2.5)), where Wr° is the writhing contributed by the intrinsic structure of the DNA, rather than by any deformation due to supercoiling.

have seen before (*Figure 2.4*), in which the helix axis is considerably contorted. By deforming it, we can demonstrate two extreme cases. If the tubing is forced to lie in a plane, we can see that a twisting of the tubing is required to accomplish the movement. Wr = 0 again, and hence ΔTw = −4 (Equation (2.6)). These twists may be counted as the four left-handed rotations of the marker line around the tubing (*Figure 2.9b*). So the linking difference is manifested as an untwisting of the DNA double helix.

Alternatively, the molecule may be deformed to a conformation like that shown in *Figure 2.9c*. For a helix whose axis does not lie in a plane, the contributions of twist and writhe are not always easy to determine; however for a regular right-handed plectonemic superhelix, the writhe is equal to $-n \sin \gamma$, where n is the number of nodes (or superhelical turns; 4, in this case) and γ is the superhelix winding angle (the pitch angle of the plectonemic superhelix, illustrated in *Figure 2.9c*) (15). For an extended superhelix of this sort, where γ approaches 90°, Wr ≈ $-n$ ≈ −4. From the continuous presence of the marker line on the upper surface of the tubing, we can also see that the ΔTw of the tubing is close to zero; hence ΔLk = −4; ΔTw ≈ 0; ΔWr = Wr ≈ −4. The most stable conformation (*Figure 2.9d*) will probably involve changes in both twist and writhe, lying between the two extremes

shown in *Figure 2.9b,c*; the contribution of twist and writhe to this conformation will reflect the relative energy required to twist the tubing (or DNA helix; Tw) and to bend the tubing or helix axis (Wr) (see Section 2.7.1).

2.4.3 *Plectonemic and toroidal conformations*

The conformation considered up to now has been the plectonemic (or interwound) conformation in which the DNA helix is wound around another part of the same molecule in a higher-order helix (*Figure 2.9c,d*). The sense of this interwinding is right-handed in the case of negatively supercoiled DNA. However, we can imagine an alternative conformation, toroidal winding, so called because the axis of the DNA helix lies on the surface of an imaginary torus (like a doughnut; *Figure 2.9e*). This conformation most closely corresponds to the literal meaning of the term 'superhelix'; an untwisting of the DNA helix is manifested as a left-handed helix of a higher order wound around the torus (cf. the right-handed superhelix in the interwound form). Unfortunately, an attempt to produce toroidal winding with the rubber tubing model does not work very well in the absence of a real torus to constrain the DNA; writhing of the tubing axis is achieved with less-severe bending in an interwound superhelix. The available evidence suggests that the same is true for free supercoiled DNA both *in vitro* and *in vivo*. Measurements of supercoiled plasmid DNA molecules in electron micrographs (17–19) have revealed only interwound molecules and have suggested that moderate levels of ΔLk are partitioned into approximately 75% ΔWr and 25% ΔTw. Studies of the Int recombination reaction (see Chapter 6, Section 6.6.2.1) have concluded that plectonemic supercoiling is most likely to be the conformation present *in vivo*. These conclusions are supported by theoretical studies of the conformation and energetics of supercoiled DNA (see Section 2.7.4). Nevertheless, the toroidal conformation still merits consideration, since it is the model for the binding of DNA on many protein surfaces (see Section 2.5.7 and Chapter 3, Section 3.5).

Toroidal winding of DNA may be conveniently illustrated by considering a coiled telephone wire (*Figure 2.10a*). This has substantial toroidal writhe, but very little twist. Stretching out the telephone wire causes an interconversion of the writhe to twist (*Figure 2.10b*); the ultimate end-point is a highly twisted linear wire (Wr $\approx$ 0; *Figure 2.10c*). Although a telephone wire is not a model of a closed-circular DNA, it is still meaningful to consider its twist and writhe. Twist and writhe are geometric, rather than topological, properties and may be defined for sections of circular DNA or for linear DNA.[3]

[3] Actually, this is a slight oversimplification for writhe. See Chapter 3, Section 3.5.1 for more detail.

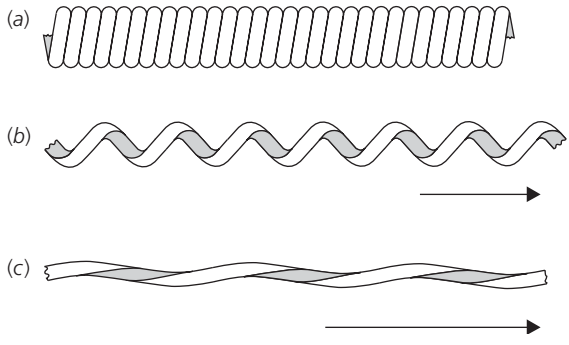


Figure 2.10 A telephone wire as a model of toroidal writhing of DNA. (*a*) The unstretched wire has a lot of toroidal writhe, but little twist. (*b*) As the wire is stretched, writhe is converted to twisting of the wire. (*c*) Eventually, the wire is almost straight (unwrithed), but highly twisted.

2.4.4 *Nomenclature*

We should say just a few words about the nomenclature used to describe the phenomenon of DNA supercoiling. The term supercoiling itself was coined to describe the coiling of the DNA helix axis upon itself; likewise superhelix (helix of helices). As already described, however, the topological extent of supercoiling (linking difference, ΔLk) is in general not only manifested as a 'supercoil', but also as an alteration in the DNA helix parameters (unwinding or untwisting). Expressions such as extent of supercoiling, number of supercoils, and supercoil or superhelix density could ambiguously refer either to changes in linking number or changes in the writhing component of ΔLk. Such terms are therefore probably best avoided in a technical discussion, although they are probably acceptable as vernacular descriptions of the overall process.

Before the introduction of the rigorously defined geometric parameters, twist and writhe, another nomenclature was devised for the description of supercoiled DNA by Vinograd and colleagues. The linking number of a closed-circular DNA, then known as α, was made up of the sum of β, the 'duplex winding number' and τ, the number of 'superhelical turns'; that is:

$$\alpha = \beta + \tau \tag{2.8}$$

These terms suffer from a similar imprecision to that described above. Some usages imply that β is equivalent to twist, and τ to writhe (20), whereas in others β seems most similar to Tw° (or Lk°), and τ to ΔLk (21). Therefore, although these terms appear in much of the early literature, it is important to try to recast the arguments in terms of twist and writhe. Although α = Lk, it is *not* safe to assume that β = Tw and τ = Wr. For a more detailed discussion of the relationship between these quantities, see Cozzarelli *et al.* (15).

2.5 Topology and geometry of real DNA

The previous discussion of topological and geometric properties of closed-circular rubber tubing has perhaps been rather abstract, so we will now turn to real DNA and consider some of the methods that can be used to analyse supercoiling.

Almost all DNAs in nature, from small bacterial plasmids to huge eukaryotic chromosomes, exhibit the properties we have described in the previous sections. Although, in a technical sense, linking number is only defined for closed-circular DNA such as plasmids, large chromosomes consist of loops of DNA anchored to a matrix. The anchorage points constrain the rotation of the DNA, and these loops, or domains, of up to several hundred kilobase pairs (kb) in length, can be considered as if they were closed-circular, with the same principles applying. Furthermore, most DNAs *in vivo* are negatively supercoiled, that is, they have a negative linking difference ($Lk < Lk^\circ$). The ubiquity of supercoiling of DNA *in vivo* will be discussed in more detail in Chapter 6, Section 6.1.1.

2.5.1 *Agarose gel electrophoresis of plasmids*

The molecules most commonly used in the investigation of the topological properties of DNA are bacterial plasmids of a few thousand base pairs in length, and the usual method of analysis is agarose gel electrophoresis (22, 23). Gel electrophoresis separates DNA molecules on the basis of size and compactness; smaller and/or more compact molecules will migrate more rapidly through the matrix of the gel under the influence of the electric field.

A typical negatively supercoiled plasmid isolated from a bacterium, for example the *E. coli* plasmid pBR322 ($N = 4361$ bp), separates into two bands on an agarose gel (*Figure 2.11a*). The band of higher mobility corresponds to negatively supercoiled molecules, having the compact, writhed structure modelled in *Figure 2.4*. The lower-mobility band comprises nicked-circular (open-circular) molecules, having the less compact planar circle conformation. Such nicked-circular molecules are not necessarily present *in vivo*, but may be formed by breakage of one strand of the closed-circular molecules during purification. These supercoiled and nicked bands are the same as components I and II identified by sedimentation analysis of polyoma virus DNA (see Section 2.2.2). The mobility of linear DNA of the same molecular weight varies with the conditions of electrophoresis; however, it is most often found to migrate between the nicked and supercoiled species.

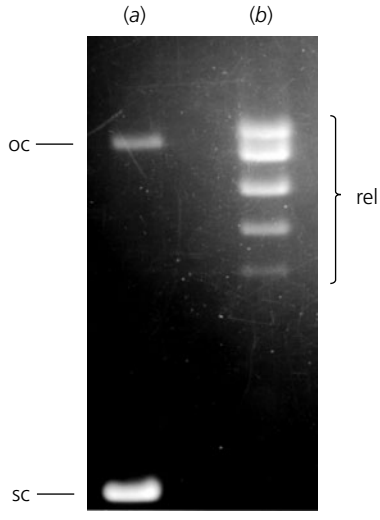


Figure 2.11 Agarose gel electrophoresis of closed-circular pBR322 DNA. (*a*) Supercoiled DNA isolated from bacteria commonly migrates as two bands on an agarose gel. sc: negatively supercoiled (closed-circular) pBR322 DNA, oc: open-circular (nicked) pBR322 DNA. (*b*) The same material as in (*a*), after treatment with a topoisomerase, consists of a distribution of topoisomers. rel: relaxed pBR322 DNA.

2.5.2 *Relaxation with topoisomerases*

Negatively supercoiled DNA has a higher free energy than its relaxed equivalent, because of its torsional stress. A class of enzymes, known as topoisomerases because of their ability to alter the topological state of DNA, are able to relieve the torsional stress of supercoiled DNA, by a mechanism that must, of course, involve the transient breakage of one or both DNA strands (see Section 2.2.1; Chapter 5). The product of the action of such an enzyme is relaxed DNA; that is, the linking number of the DNA approaches Lk°. Treatment of closed-circular DNA with a topoisomerase in this way is equivalent to the closure of a linear or a nicked-circular molecule into a closed-circle by a ligation under the same conditions (20). The result of electrophoresis of this relaxed product on an agarose gel is shown in *Figure 2.11b*. The interesting thing about relaxed DNA is that it migrates as a series of bands rather than a single band (the nicked-circular band is still there, unchanged by the enzyme). This is because a topoisomerase catalyses equilibration between topoisomers (DNAs of different linking number). The energy difference between plasmid molecules of similar linking number around Lk° (relaxed DNA) is less than the thermal energy available at normal temperatures ($\sim kT$ per molecule or RT per mole, where k is the Boltzmann constant, R is the gas constant, and T is the temperature). So the equilibrated sample consists of a mixture of topoisomers, with a distribution of values of Lk, centred at Lk°. Adjacent bands on the gel correspond to isomers differing by one in their linking number and the most abundant consists of molecules with $\mathrm{Lk} = \mathrm{Lk_m}$. The range of Lk values in relaxed DNA represents

a Boltzmann distribution determined by thermal energy, and the form of the distribution is Normal or Gaussian (see Section 2.6.2).

For circular molecules of the same molecular weight, migration in an electric field through an agarose gel is essentially a measure of writhe. Although adjacent bands differ in linking number, it is any associated differences in writhe that cause their different mobilities on the gel. The so called 'native' negatively supercoiled plasmid DNA isolated from bacteria (*Figure 2.11a*) also consists of a mixture of topoisomers, but above a certain (negative or positive) value of writhe, the topoisomers run as an unresolved single band of high mobility.

Changes in the environmental conditions alter the helical repeat (h) of the DNA, and hence change the conformation of closed-circular DNA (see Section 2.5.3). Lk_m is the linking number of the most abundant topoisomer in the relaxed distribution, and under the conditions of relaxation will have the most energetically favourable conformation, that is, on average, close to a plane circle. The most abundant topoisomer should therefore have virtually the same mobility on a gel as the nicked-circular species, but this will only happen in the unlikely situation that electrophoresis is carried out under the same conditions as the relaxation reaction. In such circumstances, the Lk_m topoisomer will have the lowest mobility and the isomers with $\Delta Lk = \pm 1, \pm 2$, etc. will have steadily increasing mobilities (topoisomers of positive and negative ΔLk have positive and negative writhe, respectively, but each leads to a similar increase in the compactness of the DNA). In general, however, the nominally relaxed topoisomers formed in the topoisomerase reaction are induced to adopt new conformations by the changed conditions of gel electrophoresis (*Figure 2.11b*). Specifically, the change in twist caused by the altered conditions (see Section 2.5.3) forces the writhe to change from its initial average value of zero to a new value, and hence the most abundant topoisomer will run with higher mobility on the gel.

2.5.3 *The effect of solution conditions on supercoiling*

The average value of linking number obtained on relaxation depends on the solution conditions under which the relaxation reaction took place. There are two well-studied conditions that have an effect on the helical repeat (and hence on the twist) of double-stranded DNA: the availability of positively charged ions (counterions) to neutralize the negatively charged backbone of the DNA, and changes in temperature.

Repulsion between the negatively charged phosphates of the DNA molecule tends to unwind the DNA helix, so increasing the concentration

of counterions (decreasing the effective negative charge) reduces the effect and the helix becomes wound more tightly: the helical repeat decreases and the twist increases. Another way of thinking about this is that in a closed-circular molecule, a decrease in the helical repeat (h) corresponds to an increase in Lk° ($=N/h$; Equation (2.2)). Remember that Lk° is not a true linking number, so it can change. So, a plasmid DNA molecule relaxed by a topoisomerase under conditions of increasing concentrations of Na^+ ions, for example, equilibrates at increasing values of average linking number, as the DNA becomes more tightly wound; these samples all have no supercoiling (and no writhe) on average. A series of such samples run on an agarose gel with a low Na^+ concentration (where all the samples are now experiencing the same conditions, of course) will show the increasing average linking number as distributions of topoisomers with higher mobilities on the gel, corresponding to increasing levels of positive supercoiling, and hence positive writhe (*Figure 2.12*).

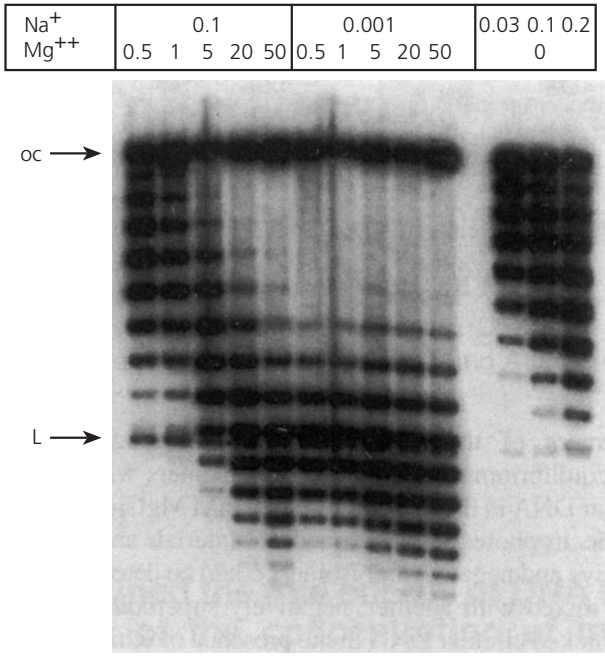


Figure 2.12 The effect of counterions on the relaxation of closed-circular DNA. An agarose gel illustrating the effect of relaxation of DNA in the presence of varying concentrations of Na^+ and Mg^{2+}. After relaxation, the samples were electrophoresed together under conditions of low counterion concentration. In general, increasing concentrations of counterions lead to relaxation at increasing values of Lk°, which is manifested on the gel as increasing positive supercoiling; oc: open-circular (nicked) DNA; L: linear DNA. See text for further details (reproduced from ref. 25).

The differences in average linking number (Lk°) of the different relaxed samples can be determined from the gel, and used to determine the magnitude of the effect. The effect does not have a smooth dependence on ionic strength; different ions (Na^+, K^+, Mg^{2+}) have different effects (24), and divalent ions such as Mg^{2+} bind to DNA much more tightly than monovalent ions, and so have an effect at much lower concentrations. Trivalent ions such as the polyamine, spermidine^{3+}, are even more potent (25). With mixtures of these ions, the situation can become complex; for example at fairly high concentrations of Mg^{2+} (50 mM), the addition of further sodium ions has little effect, whereas at low Mg^{2+} concentration (1–5 mM), the addition of sodium ions up to high concentrations actually decreases Lk°, as the ions compete for binding (*Figure 2.12*) (25, 26). For Na^+ ions alone, a change in concentration from 10 to 200 mM corresponds to a decrease in the helical repeat of about 0.04 bp/turn.

Changing the temperature also has an effect on the helical repeat of the DNA. Increasing temperature results in more thermal motion, and the gradual unwinding of the DNA helix. The effect of temperature on the linking number of relaxed DNA is shown in *Figure 2.19a* (cf. *Figure 2.12*). This effect has been studied over a wide range of temperature, between 0°C and 83°C, and is remarkably linear, with a change in the DNA winding angle per base pair of $-0.011°\ K^{-1}\ bp^{-1}$ (9, 27). This corresponds to a change in the helical repeat of DNA of, for example, 10.4 to 10.6 bp/turn between 25°C and 83°C.

To summarize, these environmental effects on the helical repeat are changing the Lk° of a closed-circular DNA. What does not change, of course, is the linking number, so this means that ΔLk ($= Lk - Lk°$) and consequently the writhe are also dependent on the conditions. Increasing the temperature or decreasing the counterion concentration reduces Lk°, so ΔLk is raised by the same treatment. Negatively supercoiled DNA will become absolutely less supercoiled (and writhed), and relaxed DNA will become positively supercoiled (writhed). Hence, in colloquial language, the level of supercoiling is a relative thing; whether a topoisomer of a given linking number is positively supercoiled, negatively supercoiled or relaxed depends on the conditions, in other words on the value of h and hence Lk°. Or, the term relaxed (no writhe; $\Delta Lk = 0$) corresponds to a different value of linking number under different conditions. Specifically, a DNA molecule that is relaxed by a topoisomerase under given conditions may no longer be relaxed when run on an agarose gel (*Figure 2.12*). Generally, electrophoresis buffers have lower counterion concentrations than the conditions of a topoisomerase reaction, so 'relaxed' DNA tends to be a little positively supercoiled on an agarose gel (*Figure 2.11b*).

In Sections 2.2.2 and 2.3 we suggested, using rubber tubing as an example, that the strained state of supercoiling is 'locked into' the system, but admitted that this was an oversimplification. While this is true for rubber tubing, it may not be true for DNA if the conditions change enough to make a significant difference to Lk°.

2.5.4 *The effect of intercalators*

An important example of a factor affecting the twist and helical repeat of a DNA is the presence of intercalating molecules. An intercalator contains a planar, usually polycyclic, aromatic structure that can insert itself between two base pairs of double-stranded DNA. This causes a local unwinding of the DNA helix, resulting overall in an increase in the helical repeat, that is, a decrease in the twist of the DNA. The classic example of an intercalating molecule is the dye, ethidium bromide (EtBr, *Figure 2.13*), which binds tightly to double-stranded DNA. Ethidium bromide is well known as a stain for DNA, since it exhibits a large enhancement of fluorescence on binding to DNA, but its additional interest from the point of view of DNA topology is that the binding of each molecule between adjacent base pairs causes a local unwinding of the helix of 26° (28, 29).

From Equation (2.6), a decrease in the twist of a closed-circular DNA molecule results in an increase in the writhe. The binding of ethidium bromide to initially negatively supercoiled DNA will reduce the twist and hence increase the writhe, so causing it to become more relaxed (*Figure 2.14*). Further addition of EtBr will lead to positive supercoiling of the DNA. Hence a DNA sample run on a series of gels containing increasing concentrations of ethidium bromide will exhibit a gradual shift towards more positive values of writhe, and the average mobility of the topoisomer distribution will change accordingly (*Figure 2.15*). For example, if the DNA sample is relaxed on the gel under certain conditions, increasing EtBr will cause an increasing positive writhe, and higher mobility, until the DNA migrates as an unresolved band. Alternatively, if the DNA is initially negatively supercoiled, and hence has negative writhe, the intercalation of EtBr into the double helix will cause a lowering of mobility and the appearance of resolved topoisomers (*Figure 2.15*). At a critical concentration of EtBr, the DNA will be relaxed on the gel; the average writhe of the original molecules being cancelled out by reduced twist caused by intercalation ($Wr = 0$, $Tw = Lk$). With a further increase in EtBr concentration, the DNA will exhibit increasing mobility due to positive writhe in an equivalent fashion to the initially relaxed sample described above. Thus intercalators such as ethidium bromide act in the same way as increasing temperature,

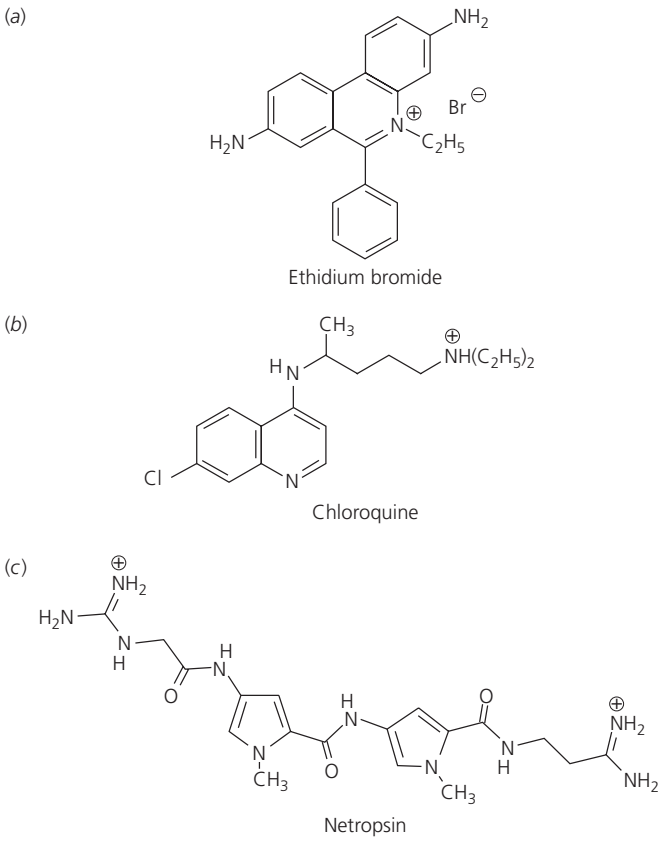


Figure 2.13 Structures of molecules whose binding alters the geometry of double-stranded DNA. (*a*) Ethidium bromide; (*b*) chloroquine; (*c*) netropsin.

or decreasing counterion concentration, in altering the level of supercoiling (but not the linking number) of closed-circular DNA. These changes may be described in terms of alterations in twist, leading to consequent changes in writhe, or in terms of changes in Lk° and hence ΔLk; these are essentially equivalent.

The concentration of ethidium bromide often used in gel electrophoresis of, for example, DNA restriction fragments (around 1 $\mu g\,ml^{-1}$) does not allow much resolution of topoisomers of plasmid DNA. Under these conditions DNA is essentially saturated with EtBr, and all closed-circular molecules have much reduced twist and high, unresolvable, positive writhe. Although nicked-circular DNA binds ethidium bromide in the same fashion, the resulting reduction in twist merely causes a swivelling of one strand around the other;

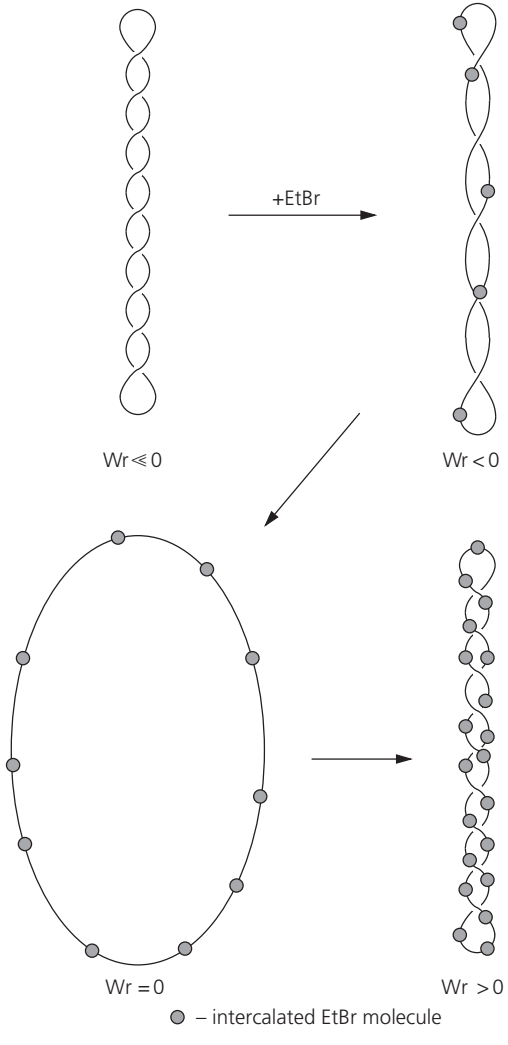


Figure 2.14 The effect of ethidium bromide (EtBr) on closed-circular DNA. Intercalation of ethidium bromide into an initially negatively supercoiled DNA (Wr ≪ 0) causes a reduction in twist, and a concomitant increase in writhe. The alteration in conformation is shown.

no writhing takes place and the molecule retains the most favourable open ring structure, hence its mobility on a gel does not change appreciably. For practical reasons, the alternative intercalating compound, chloroquine (*Figure 2.13*), which binds DNA less tightly, is often used in electrophoresis experiments, including that illustrated in *Figure 2.15* (30).

2.5.5 *Two-dimensional gels*

The effects of two different concentrations of ethidium bromide (or more usually chloroquine) on the mobility of topoisomers can be used in two

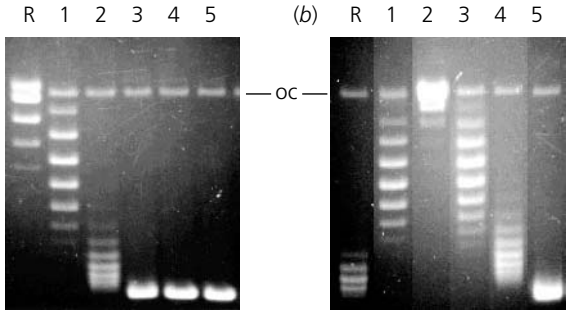


Figure 2.15 Effect of an intercalator on gel electrophoresis of closed-circular DNA. (*a*) An agarose gel of samples of pBR322 DNA of increasing negative supercoiling. R: relaxed DNA. 1–5: samples of increasing average negative specific linking difference, up to $\sigma = -0.07$. (*b*) The same samples, electrophoresed in the presence of an intercalator. The relaxed sample (R) now has positive writhe, and higher mobility. Negatively supercoiled samples (e.g. 3) now have reduced negative writhe and migrate more slowly. Although samples 3–5 appear identical in (*a*), they are resolved from each other in (*b*). Sample 1 has the same mobility in (*a*) and (*b*), but is negatively writhed in (*a*) and positively writhed in (*b*). oc: open-circular DNA.

dimensions on a single gel to enable a wider range of topoisomers to be resolved. The gel is first run in the normal way with a low (or zero) concentration of chloroquine, then the gel is soaked in buffer containing a higher concentration of the intercalator. The gel is then rotated through 90° and electrophoresed again. In the second dimension the topoisomers will have altered mobilities, as described for the one-dimensional gels above, and the topoisomers will be spread out in two dimensions. For example, consider the samples on the two gels in *Figure 2.15*. The sample R, which is an approximately relaxed set of topoisomers on the first gel, with no chloroquine, clustered around the position of the nicked-circular band, would run as such in the first dimension of a two-dimensional gel (*Figure 2.16*). In the second dimension, sample R would run as positively writhed topoisomers as in *Figure 2.15b*. The position of the bands, which now become spots, on the final two-dimensional gel can be estimated by plotting the mobilities from the two one-dimensional gels as co-ordinates on a two-dimensional plot (*Figure 2.16*). The same plotting process shows the behaviour of samples 1–3 from *Figure 2.15*. The topoisomers are thus resolved on an arc in two dimensions, as in the real example in *Figure 2.17* (31), with topoisomers that were increasingly positively supercoiled in the first dimension (spots 4, 3, 2, 1) on the right-hand limb of the arc, and those that were negatively supercoiled (spots 5–28) on the left-hand limb. In the second dimension, spot 15 is approximately relaxed, spots 14–1 are positively supercoiled, and spots 16–28 are negative. Two-dimensional gels are useful when we need to analyse

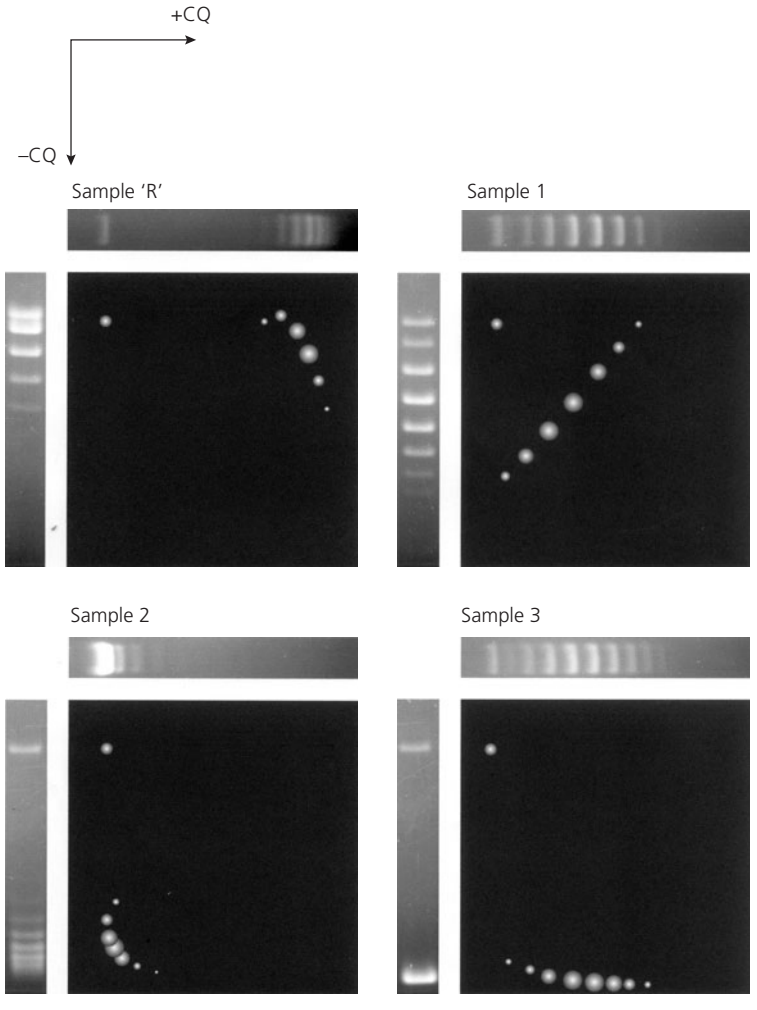


Figure 2.16 Two-dimensional agarose gels. The principle of two-dimensional agarose gels illustrated using the samples indicated from *Figure 2.15*. In each case, electrophoresis in the first dimension is as in *Figure 2.15*(*a*) (no intercalator), and the second is as in *Figure 2.15*(*b*) (+ intercalator); the corresponding tracks from *Figure 2.15* are shown at the left and above. The position of the spots expected on the two-dimensional gel are determined by plotting the positions from each dimension as co-ordinates on a two-dimensional plot. The spot at the top-left corresponds to open-circular DNA.

a wide range of topoisomers at one time, although the disadvantage is that only a small number of samples can be analysed on one gel. Two-dimensional gels have also been used to analyse supercoiling-dependent conformational changes of DNA (see Section 2.8.1).

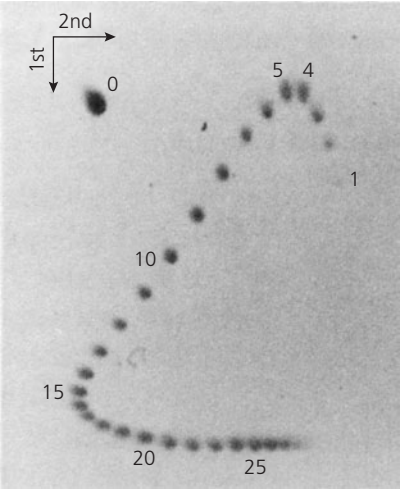


Figure 2.17 Example of a two-dimensional gel. A real two-dimensional agarose gel with a wide range of topoisomers electrophoresed under similar conditions to those in *Figures 2.15* and *2.16*. The gel is a 0.7% agarose gel run in TBE with 1.28 μM chloroquine in the second dimension (reproduced from ref. 31 with permission; copyright (1982) Cold Spring Harbor Laboratory Press).

2.5.6 *Relaxation in the presence of intercalators*

Another way to think of the effect of intercalators is to consider relaxation of DNA in the presence of ethidium bromide. Consider a topoisomer of 'relaxed' pBR322 DNA ($Lk = Lk_m = +415$, $Tw = +415$, $Wr \approx 0$); addition of ethidium bromide causes an untwisting of the DNA by, for example, eight turns ($Tw = +407$) (*Figure 2.18*). The molecule must hence adopt an equivalent positive writhe ($Wr = +8$). Treatment of the sample with a topoisomerase leads to relaxation of the DNA under the changed conditions (added intercalator, reduced twist), that is $Wr \rightarrow 0$, $Lk \rightarrow +407$, $Tw = +407$.[4] Subsequent removal of the ethidium bromide causes an increase in twist and the appearance of negative writhe; that is, the molecule becomes negatively supercoiled under the original conditions (e.g. $Lk = +407$, $Wr \approx -6$, $Tw \approx +413$). This procedure is commonly used for the preparation of negatively supercoiled DNA of any required average specific linking difference such as those in *Figure 2.15*; relaxation under conditions of increasing ethidium bromide concentration leads to increasing negative supercoiling after removal of the intercalator.

Some other classes of DNA binding molecule have the opposite effect on the DNA helix, that of increasing the twist of the DNA. The best-known example of such a molecule is netropsin (*Figure 2.13*; (32)), which binds to

[4] These numbers will not be strictly accurate, because intercalators have differential binding affinities for molecules of different levels of supercoiling (see Section 2.6.1).

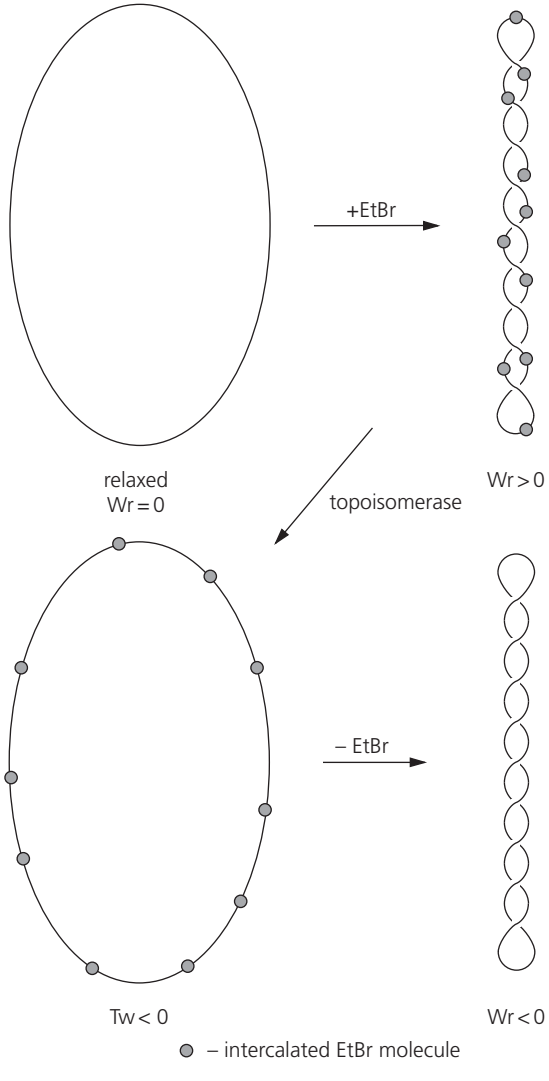


Figure 2.18 Relaxation in the presence of an intercalator. An initially relaxed DNA bound with EtBr is locally untwisted and becomes positively writhed (Wr > 0). Treatment with a topoisomerase removes the positive writhe, and subsequent removal of the EtBr converts local untwisting to negative writhe (i.e. the DNA becomes negatively supercoiled).

AT-rich DNA in the minor groove of the double helix, causing an increase in the winding of the helix of around 9° per molecule bound. Netropsin can consequently be used to produce the opposite effect to an intercalator in the experiments described above, although the relatively low efficiency of the process, and the restricted availability of the material, mean it has been much less widely used.

2.5.7 *The effect of protein binding*

The binding of certain proteins can also affect the geometry of supercoiled DNA. The best-known example of this process is the winding of DNA around the eukaryotic histone octamer to form the nucleosome. Each nucleosome involves the wrapping of about 146 bp of DNA in 1.8 left-handed turns around the histone core (33). This is a special case of the toroidal winding of DNA described in Section 2.4.3, and results in the stabilization of negative writhing in the structure of the complex (see Chapter 3, Section 3.5). Hence an initially relaxed DNA, when incorporated into nucleosomes, has negative writhe associated with the protein complex and must develop compensating positive writhe and/or ΔTw elsewhere in the molecule. Treatment with a topoisomerase will relax these compensatory distortions, with a corresponding negative linking number change. Subsequent removal of the bound protein yields a negatively supercoiled DNA. The average linking difference produced in a relaxed distribution of topoisomers in this sort of experiment is approximately -1 per nucleosome bound. This situation will be considered in more detail in Chapter 3. Other examples of proteins whose binding affects the supercoiling geometry of DNA are DNA gyrase (Chapter 5, Section 5.3.2) and RNA polymerase (see Chapter 6, Section 6.4.1). These should be distinguished from DNA-binding proteins that merely impart a planar bend to the DNA, although it may be that small amounts of twisting or writhing are involved in many DNA–protein interactions. The bending of DNA in response to protein binding has been analysed by Dickerson (34) (also see Chapter 1, Section 1.4.3).

2.6 **Thermodynamics of DNA supercoiling**

We have referred to the increase in free energy associated with the process of supercoiling several times in the preceding sections. The extent of supercoiling (ΔLk) is partitioned into elastic deformations of the axis of the DNA helix (writhe), and of the turns of the double helix itself (twist). The free energy associated with supercoiling is, in simple terms, the sum of the energies required for these bending and twisting deformations. More technically, the free energy depends on the torsional and bending rigidities of the DNA molecule, as well as the energetic consequence of bringing two segments of DNA close together in the interwound conformation. The next sections consider the experiments (Section 2.6) and theoretical approaches (Section 2.7) that have been used to analyse the thermodynamics of this process.

2.6.1 *Ethidium bromide titration*

We have seen that binding of ethidium bromide changes the conformation of closed-circular DNA, producing a reduction in twist and a corresponding increase in writhe. Since these changes have an energetic cost, the intrinsic binding affinity of ethidium bromide (EtBr) is modulated by the topological state of the DNA. In particular, EtBr will bind with higher affinity to a negatively supercoiled molecule ($\Delta Lk < 0$) than to an unconstrained molecule (e.g. nicked-circular), since binding results in a reduction in the energetically unfavourable (negative) writhing of the supercoiled molecule, as well as the reduction in twist that occurs in both closed-circular and nicked molecules. On the other hand, binding of ethidium bromide to relaxed or positively supercoiled DNA is less favourable than to nicked-circular, since this results in the introduction of (positive) writhing. This is the reason that the numbers used in the example in Section 2.5.6 are not strictly accurate: Relaxation of the DNA in the presence of ethidium bromide causes an alteration in the affinity for the intercalator; the twist after the relaxation will be a little less than that before. Ultimately, under saturating EtBr concentrations, closed-circular DNA binds less than half the amount of ethidium than does nicked or linear DNA, since high positive writhing disfavours further binding.

This differential binding of ethidium bromide to negatively supercoiled and nicked-circular DNAs was the basis of the first determinations of the free energy associated with the supercoiling process. In the 1970s, the binding isotherms were determined for ethidium bromide binding to both negatively supercoiled DNA and nicked-circular DNA formed from it by DNase I digestion. Bauer and Vinograd (21) measured the alteration of the buoyancy of DNA in caesium chloride gradients caused by titration of EtBr; Hsieh and Wang (35) followed the binding spectroscopically under similar solution conditions. At a critical stoichiometry of bound dye, all the writhing of the closed-circular molecule will be converted to untwisting caused by intercalation (see for example *Figure 2.14*), and the closed-circular and nicked molecules will have the same conformation and the same affinity for further binding of EtBr. However, below this critical value, while the closed-circular sample is still negatively supercoiled, the higher affinity of the ethidium for the closed-circular sample compared with the nicked sample can be used to determine the free energy associated with the initial ΔLk. The critical stoichiometry is directly related to the initial linking difference of the supercoiled molecule by the unwinding angle of ethidium bromide (26° per molecule bound). These experiments showed that the free energy of negative supercoiling in DNA molecules of several thousand base pairs in length had a quadratic dependence on the linking difference, that is, the free energy was approximately

BOX 2.3 **DEVIATIONS FROM QUADRATIC DEPENDENCE**

The derivation of the supercoiling free energy by ethidium bromide binding, as described, makes the implicit assumption that the bending and torsional rigidities of DNA are independent of supercoiling, and of the progressive occupancy of ethidium bromide intercalated into the molecule. A more recent and comprehensive investigation suggests that for certain DNA molecules, the behaviour is ideal in these respects, and the assumption that supercoiling free energy is quadratic up to high levels of supercoiling and EtBr binding appears to be valid (36). On the other hand, other DNAs, such as plasmid pBR322 behave in a non-ideal way, suggesting the influence of long-lived secondary structure changes at higher levels of negative supercoiling (37).

proportional to the square of the extent of supercoiling:

$$\Delta G_{sc} = K \cdot \Delta Lk^2 \tag{2.9}$$

Such a quadratic dependence of ΔG on the extent of deformation of the molecule is expected for a stretching or elastic process (cf. Hooke's Law), and thus suggests that the deformation of the DNA helix structure caused by the supercoiling of naturally occurring DNA is in the elastic range. Bauer and Vinograd did propose a small cubic term in Equation (2.9), which may be significant at higher levels of supercoiling, but not commonly considered (see Box 2.3).

2.6.2 *The Gaussian distribution of topoisomers*

The first use of ethidium bromide titration for the determination of the free energy of supercoiling pre-dated the resolution of topoisomers on agarose gels and the demonstration that the individual bands on a gel correspond to adjacent topoisomers (differing in Lk by ± 1), first shown by Keller and Wendel in 1974 (22, 23). The realization that, following ligation of the nicked-circular form (or topoisomerase-mediated relaxation), the relative concentrations of the topoisomers represent an equilibrium distribution suggested another method for the determination of the free energies.

In 1975 Depew and Wang (9), and Pulleyblank *et al.* (20) studied distributions of topoisomers of DNAs from 2200 to 9850 bp formed both by ligation of nicked-circular DNA and topoisomerase-mediated relaxation on agarose gels (*Figure 2.19a*), and determined the relative concentrations of the topoisomers from the intensities of the bands (*Figure 2.19b*). The analysis described here is based on that of Depew and Wang (9), although the nomenclature has been updated to that used throughout this book.

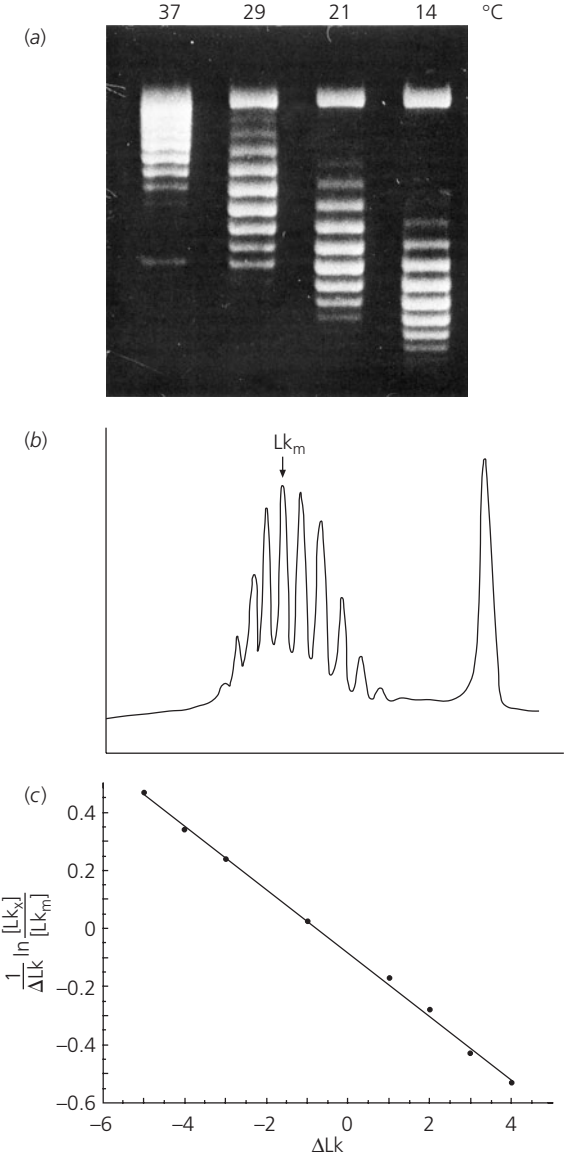


Figure 2.19 The Gaussian distribution of topoisomers. (*a*) Plasmid DNA relaxed at different temperatures. (*b*) Quantitation of the intensities of the bands corresponding to individual topoisomers in (*a*), 14°C (*c*). The data from (*b*) plotted according to Equation (2.14), to yield K and ω ((*a*) reproduced and (*b*), (*c*) redrawn from ref. 9 with permission).

For a relaxed distribution of topoisomers prepared by ligation of nicked-circular DNA, or relaxation with a topoisomerase, under given temperature and solution conditions (*Figure 2.19*), the most intense topoisomer band contains DNA of $Lk = Lk_m$. As we pointed out before, however, the formation of this most probable topoisomer usually requires

a small (energy requiring) twist and/or writhe displacement from the average conformation of the nicked-circular DNA. This hypothetically most stable conformation corresponds to the average linking number, Lk°, or the centre of the distribution. The small angular displacement required (ω) is hence $Lk_m - Lk^{\circ}(-0.5 \leq \omega \leq 0.5)$. Assuming that the free energy associated with this displacement is proportional to the square of the displacement (Hooke's Law), the free energy of supercoiling of the Lk_m isomer is:

$$G(Lk_m) = K\omega^2 \tag{2.10}$$

where K is a constant (the elastic constant) under the conditions of the experiment. Similarly, for the formation of another species in the distribution, Lk_x, with linking difference (ΔLk) here equal to the integral value $Lk_x - Lk_m$, the free energy is:

$$G(Lk_x) = K(\Delta Lk + \omega)^2 \tag{2.11}$$

The integral linking difference is in this case determined by counting from the most intense band in the distribution ($=Lk_m$). The sign of ΔLk may be determined from the dependence of the mobility of the DNA on different ligation temperatures (9) or chloroquine concentrations. The standard thermodynamic equation, $\Delta G = -RT \ln K'$ (the equilibrium constant is written here as K' to distinguish it from the elastic constant K in this derivation) relates the free energy to the concentrations of the equilibrated species, $[Lk_m]$ and $[Lk_x]$. Hence:

$$G(Lk_x) - G(Lk_m) = -RT \ln \frac{[Lk_x]}{[LK_m]} = K[(\Delta Lk + \omega)^2 - \omega^2] \tag{2.12}$$

which can be rearranged to yield

$$-\frac{RT}{\Delta Lk} \ln \frac{[Lk_x]}{[Lk_m]} = K(\Delta Lk + 2\omega) \tag{2.13}$$

or

$$\frac{1}{\Delta Lk} \ln \frac{[Lk_x]}{[Lk_m]} = -\frac{K}{RT}\Delta Lk - \frac{2K\omega}{RT} \tag{2.14}$$

The data obtained from densitometer scans of topoisomer distributions such as that illustrated in *Figure 2.19b* can be plotted according to

Equation (2.14), with:

$$y = \frac{1}{\Delta \mathrm{Lk}} \ln \frac{[\mathrm{Lk}_x]}{[\mathrm{Lk}_\mathrm{m}]}, \qquad x = \Delta \mathrm{Lk} \tag{2.15}$$

to yield a straight line (*Figure 2.19c*) with a gradient of $-K/RT$ and a y-intercept of $-2K\omega/RT$. The concentration distribution of relaxed topoisomers is thus a Gaussian (Normal) distribution centred around $\mathrm{Lk}_\mathrm{m} - \omega$ ($= \mathrm{Lk}^\circ$). This confirms that the supercoiling of DNA is an elastic process, at least for small ΔLk. The result is often extrapolated to higher levels of supercoiling; if ΔLk is large, ω becomes insignificant ($|\omega| \leq 0.5$), and therefore from Equation (2.12):

$$\Delta G_\mathrm{sc} = K \cdot \Delta \mathrm{Lk}^2 \tag{2.16}$$

In other words, at higher levels of supercoiling, the distinction between $\Delta \mathrm{Lk} = \mathrm{Lk} - \mathrm{Lk}_\mathrm{m}$ and $\Delta \mathrm{Lk} = \mathrm{Lk} - \mathrm{Lk}^\circ$ becomes insignificant.

2.6.3 *The effect of DNA circle size*

The energy required to introduce a given linking difference into a DNA circle should be inversely proportional to the size of the circle (see Section 2.3.2). As we showed earlier, the extent of supercoiling may be normalized to the size of the circle (Section 2.3.2), and the same applies to the free energy. Hence, dividing by N (number of base pairs) to give the free energy of supercoiling per base pair gives:

$$\frac{\Delta G_\mathrm{sc}}{N} = NK \left(\frac{\Delta \mathrm{Lk}}{N} \right)^2 \tag{2.17}$$

The constant NK was shown to be largely independent of the size of the DNA circle in the size range investigated. The value determined was approximately $1100RT$. The value of $\Delta \mathrm{Lk}/N$ is proportional to the specific linking difference; hence the free energy of supercoiling per base pair is proportional to the square of the specific linking difference, independent of circle size. Specifically, combining Equations (2.2) and (2.4) gives:

$$\frac{\Delta \mathrm{Lk}}{N} = \frac{\sigma}{h} \tag{2.18}$$

which we can substitute into Equation (2.17) to give:

$$\frac{\Delta G_{sc}}{N} = \frac{NK}{h^2}\sigma^2 \tag{2.19}$$

Since $NK = 1100RT$, and $h = 10.5$ bp/turn:

$$\Delta G_{sc} \approx 10NRT\sigma^2 \tag{2.20}$$

independent of circle size.

2.6.4 *Supercoiling free energy and solution conditions*

In the early determinations of supercoiling free energy, not too much attention was paid to the likely effect of solution conditions on the free energy of supercoiling, although a significant effect was later predicted in a theoretical simulation of supercoiling (see Section 2.7.4). One of the most comprehensive studies of supercoiling free energy was carried out more recently by Rybenkov *et al.* (25). Using essentially the same method as Depew and Wang (see Box 2.4), the authors studied the effect of a variety of DNA counterions (cations that will bind to the negatively charged phosphates and alter the properties of the DNA; see Section 2.5.3). The authors systematically varied the concentrations of Na^+, Mg^{2+}, and spermidine^{3+} ions and formed equilibrium distributions of topoisomers either using DNA ligase or a topoisomerase (in the absence of Mg^{2+}, when ligase will not work). The free energy of supercoiling was found to vary by more than 50%; NK is greater than $1500RT$ in very low concentrations of counterions (1 mM Na^+, <1 mM Mg^{2+}) and as low as $920RT$ in the presence of 3 mM spermidine, although $1100RT$ ($\Delta G_{sc} = 10NRT\sigma^2$; Equation (2.20)) is still a good estimate under *in vivo* conditions. The results correlate very closely with the ability of the ions to shield the charge on the DNA backbone, and allow a close approach of two DNA helices. In other words, DNA supercoiling is stabilized (reduced in free energy) by facilitating the close approach of the DNA helices in the interwound conformation—this makes good sense.

A measure of this ability to shield the charge on DNA is given by the 'effective diameter' of the double helix. This is the diameter of an uncharged cylinder that will mimic the conformational properties of real, charged, DNA. The probability of two DNA molecules approaching closer than the effective diameter of the DNA is very low—effectively zero. The effective diameter is an important

BOX 2.4 **ALTERNATIVE NOMENCLATURE FOR THE FREE ENERGY OF SUPERCOILING**

For historical reasons (see Section 2.7.4), Vologodskii and co-workers have taken a statistical frequency distribution approach to the derivation of supercoiling free energy from topoisomer distributions (25), using the variance of the topoisomer distribution, although this is entirely equivalent to the treatment we quote in Section 2.6.2. The equation of a Normal distribution of topoisomers is:

$$P_i = Ae^{-(i-\delta)^2/(2\langle(\Delta\mathrm{Lk})^2\rangle)} \tag{2.21}$$

where P_i is the frequency of topoisomer i with linking difference $(i - \delta)$ (essentially its relative concentration in the distribution), $\langle(\Delta\mathrm{Lk})^2\rangle$ is the variance of the topoisomer distribution, and A is a constant. This means that $\delta = -\omega$ from our derivation in Section 2.6.2. So, for the most intense topoisomer ($i = 0$):

$$P_0 = Ae^{-\delta^2/(2\langle(\Delta\mathrm{Lk})^2\rangle)} \tag{2.22}$$

So, the free energy difference between topoisomer i and topoisomer 0 is:

$$\Delta G_i = -RT \ln \frac{P_i}{P_0} = \frac{RT}{2\langle(\Delta\mathrm{Lk})^2\rangle}[(i-\delta)^2 - \delta^2] \tag{2.23}$$

Comparing this with Equation (2.12), remembering that $\delta = -\omega$, we can see that the constant K for supercoiling free energy is:

$$K = \frac{RT}{2\langle(\Delta\mathrm{Lk})^2\rangle} \tag{2.24}$$

and

$$NK = \frac{NRT}{2\langle(\Delta\mathrm{Lk})^2\rangle} \tag{2.25}$$

Unfortunately for the sake of compatibility, Vologodskii and co-workers define a different K (let's call it K_1) as:

$$K_1 = \frac{N}{2\langle(\Delta\mathrm{Lk})^2\rangle} \tag{2.26}$$

So, when we say $NK = 1100RT$, they would say $K = 1100$; these are equivalent. Although this value is quoted as dimensionless, it actually has a dimension of length (N in Equation (2.26)), and should really be quoted in units of base pairs (bp). In the same way NK should really be quoted as $NK = 1100RT$ bp, although this never happens in the literature.

concept in the theory of polyelectrolytes (like DNA), and has been important in the computer simulation of DNA conformations (see Section 2.7.2).

2.6.5 *Supercoiling in small DNA circles*

In the early 1980s, studies of the ligation reactions of linear DNAs yielded interesting results in the case of small DNA fragments, below around 500 bp. Ligation of linear DNA fragments into circles becomes an increasingly slow process as the DNA size is reduced, reflecting the increasing difficulty of bending a smaller DNA fragment into a circle. Furthermore, below about 300 bp, the rate of ligation becomes a periodic function of the DNA length with a periodicity of around 10 bp (38, 39; see Section 2.7.1). This implies that the energy required to align the ends of the DNA strands by twisting or writhing, if the length of the DNA is not a multiple of the helical repeat, becomes a significant proportion of the activation energy of the ligation process in small circles. This makes sense because the required change in twist or writhe of up to 0.5 must be distributed over a shorter length of DNA. A corollary of this is that, below 300 bp, the energy required to change Lk by one unit becomes large, and only one or two topoisomers are visible in the relaxed distribution (e.g. *Figure 2.20*). In the case where two isomers are visible, each isomer has a significant level of positive or negative supercoiling. Hence, the specific linking difference must certainly be related to the centre of the topoisomer distribution, namely Lk°, rather than the most intense topoisomer in the distribution (Lk_m), since DNA of $Lk = Lk_m$ may be significantly supercoiled. The specific linking difference (σ) of a topoisomer of a small DNA circle of linking number Lk is hence:

$$\sigma = \frac{Lk - Lk^\circ}{Lk^\circ} \tag{2.27}$$

Shore and Baldwin (40) and Horowitz and Wang (41) determined the free energies of supercoiling in small circles (down to a size of 210 bp), from the ratios of relaxed topoisomers using modifications of the procedure outlined in Section 2.6.2. They showed that the constant NK (Equation (2.17)) increases gradually with decreasing DNA length (*Figure 2.21*). They interpreted this as being due to the increasing unfavourability of writhing relative to twisting of the DNA. A given twisting or untwisting of the helix should require the same energy per unit length of the DNA in any size of circle. The same is not true of writhing, however. Writhing is essentially a bending motion of the DNA helix, and most of the bending in the interwound conformation is in the end (apical) loops. As circles get smaller, a larger proportion

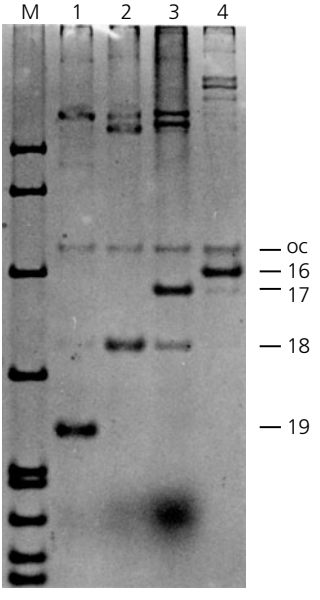


Figure 2.20 Topoisomers of small DNA circles. A polyacrylamide gel used to separate topoisomers of a small DNA circle. Tracks 1–4 contain a 196 bp DNA closed-circle, with increasing negative linking difference (the gel is run in the presence of ethidium bromide, so all the topoisomers have positive writhe; see *Figure 2.14*). Only one or two topoisomers are visible in each sample. The nicked-circular band (oc) is present in all samples. The actual values of Lk for the topoisomers are indicated.

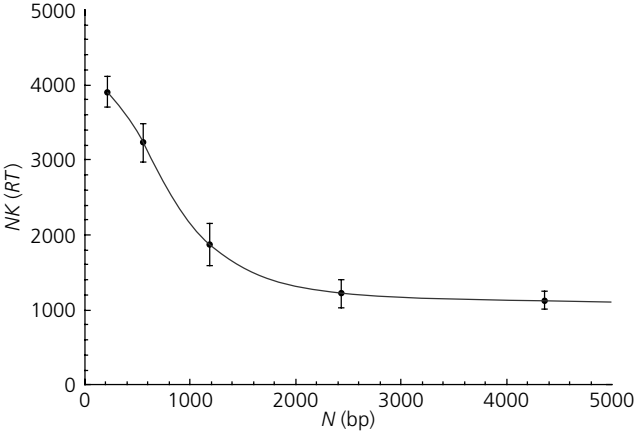


Figure 2.21 Length dependence of the constant *NK*. The value of *NK* (Equation (2.17)) is plotted against *N* for circles from 200 to 4500 bp. The bars show estimated errors (redrawn from ref. 41 with permission; copyright (1984) Elsevier).

of the circle is contained in the loops, so the relative bending cost is greater. It is thought, therefore, that supercoiling of small DNA circles is partitioned largely into twisting, rather than writhing. We can model this effect using rubber tubing (as always!). A very short length of tubing closed into a circle

will show, first, a much higher resistance to the introduction of a supercoil, and second, a higher proportion of twisting than was seen previously (Section 2.4.2). Horowitz and Wang (41) suggested that moderate linking difference in small DNA circles is partitioned almost exclusively into twisting of the DNA helix; indeed it had already been suggested from a theoretical simulation that a ΔLk of approximately 1.5–2.0 was required to cause a DNA of less than 300 bp to flip into a writhed conformation (42, 59). Consistent with this, analysis of the conformations of 178 bp circles by cryo-electron microscopy and gel electrophoresis (17) has suggested that one negative supercoil is partitioned into untwisting, with no significant change in the shape of the circle, but two supercoils cause the circle to flip into a figure-eight conformation, as long as a reasonable salt concentration is present so that the effective diameter is small enough to allow the close approach of the helices at the crossover of the figure-eight.

2.7 Theoretical approaches to DNA supercoiling

The experimental analysis of topoisomer distributions and the free energy of supercoiling does not directly address the question of the conformation of the molecules, or the relative distribution of twist and writhe, although some measurements have been made (see Section 2.4.3). A number of computer-based theoretical models of the conformations and energetics of supercoiled DNA have been made, which can complement experimental determinations since basic information about the energetics of deformation of the DNA helix can be used to simulate the actual conformation (including twist and writhe). In addition, the behaviour of DNA under a wide range of solution conditions can in principle be simulated. A computational approach works well for DNA, since it behaves as a fairly simple elastic polymer. We have already seen how the experimental data on free energy of supercoiling is consistent with an elastic deformation of the DNA structure (see Section 2.6). Calculation of the energetics of deformation of DNA depends on knowledge of three basic components: the energetic cost of bending or flexing the DNA (the bending rigidity), the cost of over- or undertwisting the DNA helix (the torsional rigidity) and the cost of bringing DNA double helices close together. These three components are described, respectively, by the bending and twisting force constants, and the effective diameter (see Section 2.6.4). All three of these parameters are known to a reasonable degree of precision from independent experiments.

2.7.1 *Bending and twisting rigidity and persistence length*

The energies associated with bending and twisting the DNA helix both show a quadratic dependence on the extent of these elastic deformations. The free energy for bending the helix is given by:

$$\Delta G_{\text{bend}} = \frac{1}{2} B \left(\frac{\theta^2}{L} \right) \tag{2.28}$$

where L is the axial length of helix which is bent through an angle θ radians and B is the bending force constant. Analogously, for elastic twisting:

$$\Delta G_{\text{twist}} = \frac{1}{2} C \left(\frac{\phi^2}{L} \right) \tag{2.29}$$

where a length L is over- or undertwisted through ϕ radians; C is the twisting force constant. Note the similarity in form to Equation (2.17), which, rearranged, is:

$$\Delta G_{\text{sc}} = NK \left(\frac{\Delta \text{Lk}^2}{N} \right) \tag{2.30}$$

This is not surprising, since these elastic twisting and bending energies underlie the free energy of supercoiling. B and C are analogous to the length-independent constant for supercoiling free energy, NK. The values of the force constants are usually quoted, for historical reasons, in units of erg cm: $B = 2.0 \times 10^{-19}$ erg cm and $C = 2.5 \times 10^{-19}$ erg cm. In more useful units for our purposes, they come out as $B = 350$ kJ mol^{-1} bp (83 kcal mol^{-1} bp; $140RT$ bp at 25°C) and $C = 440$ kJ mol^{-1} bp (105 kcal mol^{-1} bp; $180RT$ bp at 25°C), when the length of the bent or twisted DNA is measured in base pairs.

The bending force constant for DNA is a measure of the energy required to bend the helix, but this is directly related to the flexibility of the DNA due to thermal motion. This flexibility is normally described by a parameter called the persistence length, a, which is slightly hard to define, but describes the length scale over which the stiffness of the DNA helix is important. The easiest definition is that the persistence length describes the length over which there is some correlation between the initial and final direction of the helix. After a distance of one persistence length, due to thermal motion, the average deflection of the helix axis from its starting direction will be one radian, and after two persistence lengths the correlation between starting and ending directions

will be completely lost with an average deflection angle of about 90°. This value of $2 \times a$ is also known as the Kuhn statistical length. This is slightly confusing, but the value of the persistence length is directly related to B:

$$a = \frac{B}{kT} \tag{2.31}$$

where k is the Boltzmann constant; a has a value of 50 nm or 150 bp at 25°C. Another way of describing this is that the mean bend angle between adjacent base pairs of DNA due to thermal motion is around 7° at 25°C. A consequence of this is that DNA has different properties at different length scales: a short segment of DNA with length well below the persistence length acts as a relatively rigid rod, whereas lengths much longer than a behave as random coils. Of course, the local properties of DNA, such as bending and torsional rigidity do not depend on the actual length of the molecules analysed.

A number of methods have been used to measure the bending and twisting rigidities of DNA, including light scattering and hydrodynamic measurements of linear DNAs, fluorescence polarization anisotropy of intercalated ethidium bromide and direct measurements in electron micrographs. However, the most interesting from the perspective of this book involves the behaviour of small DNA fragments with cohesive ends when ligated into a circle (see Section 2.6.5). As a DNA fragment gets shorter, from several thousand to a few hundred base pairs, the probability of one end ligating to the other to form a circle increases, relative to the probability of forming a dimer by intermolecular ligation, as the ends are on average closer to each other, but as the length approaches about two persistence lengths, around 300 bp, the cyclization probability rapidly decreases, because the stiffness of the helix prevents it being easily bent into a circle (*Figure 2.22*). The exact dependence of the ligation probabilities on DNA length can be used to determine the bending force constant (and hence the persistence length) of the helix.

We have already seen that a DNA molecule can only be closed into a planar circle without deformation if its length is a multiple of the helical repeat (see Section 2.3.1), although the energy required for this deformation is very low if the DNA is several thousand base pairs in length. However, as the DNA length is reduced two things will happen; any deformation required to align the ends will become almost entirely a change of twist, as writhing is disfavoured in small DNA circles (see Section 2.6.5), and the required twist of up to 180° will have a greater and greater energetic cost, since the twist is contained in a shorter and shorter DNA length. Therefore, in short DNA fragments, this twisting requirement has an important effect on the cyclization probability,

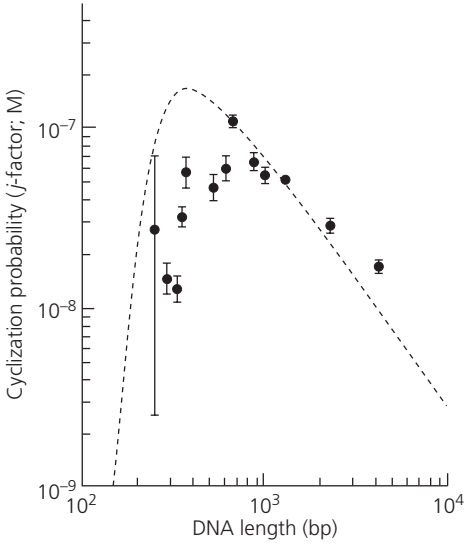


Figure 2.22 The cyclization probability of short DNA fragments. The dependence of cyclization probability (*j*-factor) on DNA length between 250 and 5000 bp. The cyclization probability drops at lengths below about 300 bp as the stiffness of DNA begins to impede bending into a circle. The large error bar for the first point represents the periodic variation in *Figure 2.23* (redrawn from ref. 39 with permission; copyright (1983) Elsevier).

which comes to have a periodic dependence on the DNA length with a period corresponding to the helical repeat (*Figure 2.23*). The amplitude of this period is a function of the torsional rigidity of the DNA provided that only changes in twist are involved, and C (the twisting force constant) can be determined from it.

The theory underlying cyclization probability was first worked out by Jacobsen and Stockmeyer (43). The cyclization probability was defined as the ratio of the equilibrium constants for the formation of the circular molecule by ligation of cohesive ends and the formation of a linear dimer by intermolecular ligation. This probability is called the *j*-factor, measured in units of concentration, and is equivalent to the effective concentration of a DNA end in the vicinity of the other end of the same molecule (in a conformation able to ligate). This method was first applied to DNA in the 1980s (38, 39) to yield the persistence length and the torsional rigidity, and these methods were later refined by Levene and Crothers (44) and Taylor and Hagerman (26), who showed that both parameters are essentially independent of the solution conditions over the range normally encountered with DNA. These methods have been reviewed by Crothers *et al.* (45).

2.7.2 *Excluded volume and effective diameter*

The physical behaviour and properties of a polymer like DNA are dependent on the thickness of the polymer chain, since there are restrictions on the path

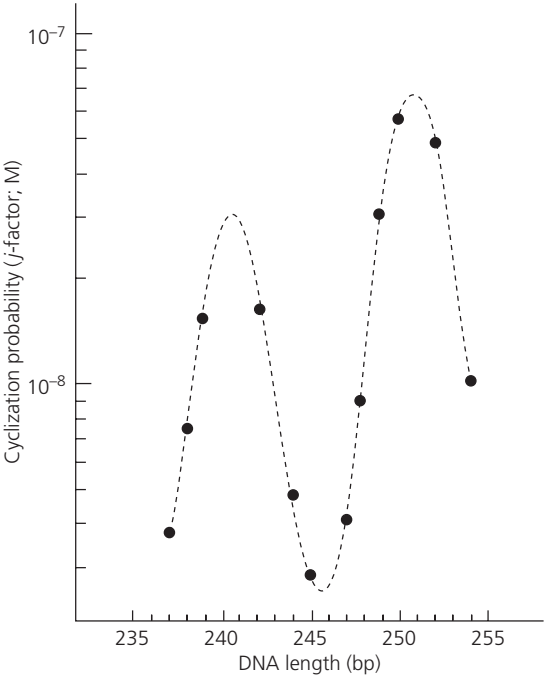


Figure 2.23 The effect of torsional rigidity on cyclization probability. The dependence of cyclization probability (*j*-factor) on DNA length at around 250 bp. The requirement to align the ends by twisting before ligation leads to a periodic dependence on length with a period corresponding to the DNA helical repeat (*h*) (redrawn from ref. 39 with permission; copyright (1983) Elsevier).

of the chain due to the fact that it clearly cannot intersect itself. The larger the diameter of the chain, the more likely such an intersection becomes. The volume occupied by the chain, and therefore unavailable to accommodate the path of another part of the chain, is known as the 'excluded volume'. In the case of charged polymers such as DNA, we have seen that this excluded volume is dependent on the effective charge on the DNA backbone, since charged DNA segments will in general repel one another at a distance larger than the physical diameter of the DNA (Section 2.6.4). This effect can be approximated by defining the effective diameter, *d*, such that no two DNA segment axes can approach closer than *d*. This sudden cut-off is not completely realistic, but modelling studies have shown that this simplification is normally justified (46). The effective diameter has been determined by hydrodynamic measurements (47), which measure the compactness of the DNA, since a larger excluded volume will cause a polymer to occupy a larger average volume overall. However, there is also a topologically interesting method for determining *d*. If a long DNA molecule is closed into a circle, then a small percentage of the molecules will form a knotted circle, instead of a simple closed loop (see Chapter 4, Section 4.2.2). The proportion of knotted molecules at equilibrium is related to the excluded volume of the DNA, since, in simple

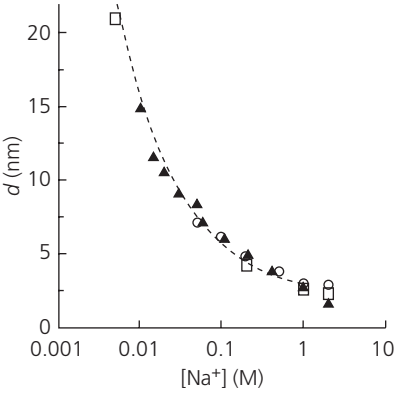


Figure 2.24 The effective diameter of DNA. The effective diameter as a function of Na^+ concentration. The data are from equilibrium sedimentation of linear DNA (squares (47)), DNA knotting probability (circles (48); triangles (49)) and a theoretical calculation (line (50)). The different methods show very good agreement (redrawn from ref. 46 with permission; copyright (1994) Annual Reviews www.annualreviews.org).

terms, the larger the diameter of the DNA, the less likely it is that it will adopt the more complex, knotted conformation. This method relies on a knowledge of the persistence length (bending rigidity), which also affects the knotting probability (see Section 2.7.4), but it has been used to determine *d* under many solution conditions (25, 48, 49). The results of an early theoretical calculation (50), hydrodynamic measurements (47) and the knotting probability method are all in remarkably close agreement (*Figure 2.24*). It is apparent from *Figure 2.24* that *d* is extremely sensitive to solution conditions, and that at high counterion concentrations, the value is close to the physical diameter of DNA (2 nm). Indeed, there is some evidence from knotting probabilities that the value of *d* can in some cases, notably high concentrations of divalent ions such as Mg^{2+}, be lower than 2 nm, corresponding to an attractive force between DNA molecules (48). This may explain the appearance of tight association between interwound segments of DNA in some electron micrographs (17, 19). Under the standard solution conditions of 0.2 M NaCl, *d* is about 5 nm.

2.7.3 *Simulations of DNA conformation and energetics*

With the basic parameters defining the energetics of DNA conformation determined, we can consider the methods that have been used to simulate the conformations of circular DNA, and to calculate their energies. The majority of these involve the modelling of DNA as a so-called wormlike chain and the use of a Monte Carlo simulation method to sample a series of possible conformations, calculate their energies and find either the conformation with the lowest energy, or a set of conformations representing the thermal fluctuations of the molecule (see Box 2.5). In principle, it would be possible to treat a DNA molecule as a set of individual base pairs, and consider the energy involved in

bending and twisting the DNA at every base-pair step. In practice, for DNAs of thousands of base pairs, this would be unrealistic in terms of computer time. Instead, DNA is normally modelled as a set of rigid segments with elastic bending joints between them (jointed rods). It has been shown that this simplification is acceptable as long as there are sufficient rods per persistence length of DNA. In general, five rods per persistence length, a, (10 per Kuhn length) is considered to be acceptable; each rod corresponds to around 30 bp. The elastic energy of bending is thus calculated as the sum of all the energies due to the

BOX 2.5 **MONTE CARLO SIMULATIONS**

Most computer simulations of DNA conformations are Monte Carlo simulations. In its most general sense, a Monte Carlo method simulates properties of a system using some method involving random numbers and probability statistics. In the more specific sense of molecular simulations, a model of a molecule's conformation is developed that can be used to calculate an energy, then random changes to the conformation are used to test the energies of a set of conformations, either to find the most stable one, or to generate a set of plausible conformations. The specific method used in most DNA studies, as well as in many others, was developed by Metropolis *et al.* in 1953 (52). From a starting conformation of calculated energy E_1, a random conformational change is induced in the model to give a new 'trial' conformation, and a new energy E_2 is calculated. The trial conformation is accepted if it has lower energy than the precursor ($E_2 < E_1$), but if the energy is higher, the conformation is accepted randomly, with a probability (P) that depends exponentially on the energy difference between the two conformations:

$$P = e^{(E_1 - E_2)/kT} \tag{2.32}$$

The probability thus reflects the likelihood that the conformation would occur by random thermal fluctuation. If the trial conformation is accepted, it is used as the basis for the generation of a new trial conformation. If it is rejected, the previous conformation is used for the next step. In this way, after many trials (typically hundreds of thousands) migrating through the space of possible conformations, the simulations converge on the conformations of a thermal distribution of the modelled molecules. It can be shown that subsequent trials effectively sample the thermal distribution of conformations at temperature T in Equation (2.32). The Metropolis Monte Carlo method relies on the assumption that the starting conformation is close enough to the true thermal distribution not to be separated from it by a very high energetic barrier with a low probability of being surmounted in successive trials. These methods were first applied to circular DNA conformation by Le Bret and Vologodskii *et al.* (51, 53).

bending angles between segments, where the elastic force constant for bending between straight segments is scaled from B to take account of the size of the segments. The application of the effective diameter to the simulated conformations is easy to understand; any conformations in which non-adjacent segments approach closer to each other than the value of d are rejected as effectively of infinitely high energy (see Box 2.5).

Consideration of the bending energy of the model and the excluded volume (in the guise of the effective diameter) is all that is required for the simulation of the conformations of nicked-circular DNA, where the average torsional energy of the DNA will be zero. While in some cases a model has been used which explicitly incorporates the torsional component of the elastic energy, the torsional energy is normally calculated indirectly by a trick. For a model of a closed circle made of rigid segments, it is possible to calculate the writhe of a given conformation trigonometrically (51). From the writhe, if one specifies the ΔLk of the closed-circular topoisomer one wishes to simulate, then the ΔTw can be determined directly from the standard twist and writhe relationship (Equation (2.6)). The twist energy of the conformation can then be calculated from the known torsional rigidity of DNA (Section 2.7.1), and the total elastic energy (bending plus twisting) can be used in the Monte Carlo simulation step.

2.7.4 *Applications of circular DNA simulations*

The applications of Monte Carlo simulations to the study of DNA conformation have been reviewed in detail by Vologodskii and Cozzarelli (46, 54). Many early simulations of closed-circular DNA concentrated on finding the minimum energy conformation of supercoiled DNA by analysis of models including the twisting and bending components. These calculations confirmed that a fairly regular plectonemic interwound superhelix was the lowest energy form (55), and were the first to suggest that in small DNA circles, ΔLk is initially manifested as twisting of the helix, and that flipping into a writhed figure-eight conformation should occur suddenly at a critical ΔLk of about 1.5–2.0 (42, 59). This latter effect has been confirmed experimentally from measurements of the free energy of supercoiling, and more directly by electron microscopy and gel electrophoresis (Section 2.6.5).

The Metropolis Monte Carlo simulation method allows one to generate sets of DNA conformations that represent the thermal distribution of conformations at a given temperature (see Box 2.5). This means that one can determine the distribution or variance of different parameters (see Box 2.4). For example, simulating an equilibrium set of conformations of nicked-circular DNA, using

the bending rigidity and the effective diameter, allows the calculation of variance of the writhe, $\langle(\mathrm{Wr})^2\rangle$. For a nicked DNA, the torsional rigidity is not relevant to the energy calculation, but thermal variations in twist will still occur. If we then imagine closing the nicked DNA by ligase to form a closed-circular molecule, then the sum of the variance in ΔTw and Wr will be equal to the variance in ΔLk:

$$\langle(\Delta\mathrm{Lk})^2\rangle = \langle(\Delta\mathrm{Tw})^2\rangle + \langle(\mathrm{Wr})^2\rangle \tag{2.33}$$

The variance in ΔLk is measurable from a topoisomer distribution generated under the same conditions (see Section 2.6.4; Box 2.4), so Equation (2.33) may be used to calculate the variance in the twist $\langle(\Delta\mathrm{Tw})^2\rangle$ of a nicked-circular DNA. This in turn leads to an estimate of the torsional rigidity of DNA, C, which is in good agreement with that measured by alternative methods (Section 2.7.1). This was the first use of the Metropolis Monte Carlo simulation for circular DNA (53).

Simulation of the conformations of a nicked-circular DNA is also the basis for the most versatile method for determining the effective diameter of DNA under a variety of conditions (see Section 2.7.2). As long as, in the steps of the Monte Carlo simulation, the DNA strands are allowed to pass through one another, then a proportion of the simulated conformations will consist of knotted-circular DNA. The proportion of knots is very sensitive to the chosen value of the effective diameter, d. Put simply, knotted-conformations are relatively more likely to be rejected than unknotted ones due to the close approach of segments, as d becomes larger. Tuning the value of d in a simulation to fit the proportion of knotted molecules formed in an experiment under particular solution conditions allows a value of d to be assigned to those conditions (25, 48, 49). As we have seen in Section 2.7.2, this method gives a very good match to independent determinations of the dependence of d on salt concentration. That these simulation methods of determining the effective diameter and the torsional rigidity of DNA give such a good match to previous measurements suggests that the simulations are producing a realistic description of DNA conformation.

Simulated sets of conformations of supercoiled DNA of particular values of σ (*Figure 2.25*) give very good qualitative agreement with the conformations seen in electron micrographs and inferred from studies using Int recombination (Section 2.4.3). Interwound molecules are almost exclusively present, although the conformations are rather loosely interwound at low specific linking difference or high effective diameter (low salt concentration). The partition of ΔLk into changes in twist and writhe in simulated molecules also

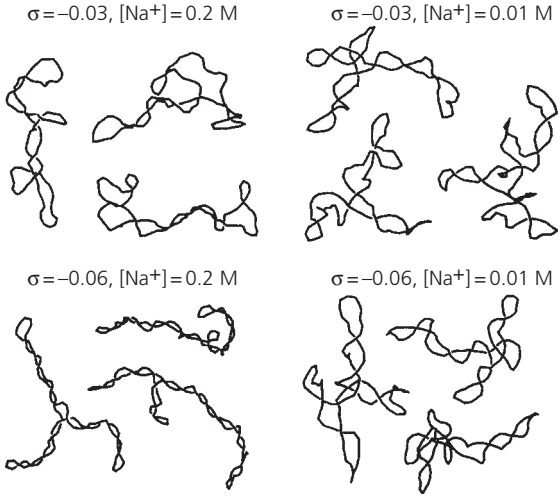


Figure 2.25 Simulations of supercoiled DNA conformations. Supercoiled DNA conformations simulated by a Monte Carlo method for a 3.5 kb circle with the indicated values of specific linking difference (σ), using values of the effective diameter (d) corresponding to the Na^+ concentrations shown (reproduced from ref. 46 with permission; copyright (1994) Annual Reviews www.annualreviews.org).

confirms experimental determinations; Wr $\approx$ 0.7 ΔLk for DNA circles more or less independent of DNA length (above 2.5 kb) or specific linking difference (46). The proportion of writhe increases with increasing salt (decreasing d) as the diameter of the interwound superhelix decreases; such an increase in writhe has also been observed by measurement of cryo-electron micrographs of supercoiled DNA (17, 19).

An interesting aspect of simulated DNA conformations is the preponderance of branching in the supercoiled DNA molecules (*Figure 2.25*). Branching ought to be disfavoured by the extra bending at the branch point and at the new apical loops, but is favoured by the increased configurational entropy of the system. Therefore, although the minimum energy conformations of supercoiled DNAs are unbranched, branches are very common in simulated thermal distributions of molecules above about 4 kb. The number of branches increases with DNA length, but decreases at higher values of specific linking difference and lower values of effective diameter, as both of these lead to tighter bending of the end loops, favouring conformations with fewer ends (56). Branches are also common in electron micrographs of supercoiled DNA (18).

The mean elastic free energy of simulated supercoiled molecules (i.e. the free energy of supercoiling) is available directly from the Monte-Carlo process. The values derived for standard conditions of 0.2 M NaCl (d = 5 nm) (57) correspond well to the experimental values from topoisomer distributions, but the simulations suggested that the free energy of supercoiling should be very dependent on the effective diameter (counterion concentration). These suggestions were later confirmed by experimental measurements (Section 2.6.4).

Monte Carlo simulations have also been used in studies of DNA knotting and catenation (reviewed in (54); see Chapter 4).

2.8 Biological effects of supercoiling free energy

The excess free energy associated with the negative supercoiling of DNA may be utilized in many cellular mechanisms. In general, processes that require an untwisting or a negative writhing of DNA, or which stabilize such deformations, are facilitated with negatively supercoiled as compared to relaxed DNA.

Examples of such processes include the replication and transcription of DNA, which require the unwinding of the DNA helix, the formation of nucleosomes and other protein complexes on DNA that stabilize the negative writhing of the helix, and the formation of altered DNA structures, such as Z-DNA and cruciforms. These and other effects of DNA supercoiling and its associated free energy, and their biological significance, will be considered in detail in Chapter 6. Here we will consider the stabilization of alternative structures, such as cruciforms and Z-DNA, the understanding of which relates to some of the methods discussed in this chapter.

2.8.1 *Stabilization of alternative structures by DNA supercoiling*

In the discussions of alternative DNA structures in Chapter 1 (Sections 1.2.4, 1.3.1, and 1.3.2), we said that certain structures such as Z-DNA, cruciforms, and H-DNA are stabilized by negative supercoiling. Now that we understand the conformation and energetics of supercoiled DNA, we can see how this works. The formation of all of these structures involves the local unwinding of DNA. This is perhaps most obvious in the formation of a cruciform, where the twist of the whole region encompassing the cruciform (*Figure 1.4*) is reduced to zero. The same is true for the single-stranded region of H-DNA. In the case of the formation of a region of Z-DNA, the right-handed twist of the B-DNA is converted to left-handed twist when the Z-DNA is formed, so the effect on the overall twist of the molecule is approximately twice that of a cruciform of the same length. So, although the formation of these structures is unfavourable in linear DNA, because of the denaturation of the loops in the cruciform, or the boundaries between B- and Z-DNA, in negatively supercoiled DNA, these reductions in twist are accompanied by a decrease in (negative) writhe. This reduction in the overall level of supercoiling of the DNA is energetically favourable, and above a threshold value of ΔLk, will

compensate for the energy of formation of the cruciform or Z-DNA. Another way to consider it is that the untwisting of the region containing the altered structure is essentially reducing the Lk° value of the plasmid, and so reducing the level of negative supercoiling. Since the formation of the alternative structure is accompanied by a change in writhe, then agarose gel electrophoresis, particularly on two-dimensional gels (Section 2.5.5), can be used to analyse the process.

Figure 2.26 illustrates a two-dimensional gel showing the behaviour of two plasmids of around 4400 bp, one of which contains a 32 bp insert that is capable of forming Z-DNA (31). The formation of the Z-DNA can be clearly seen in the discontinuity in the smooth pattern of topoisomers in the right-hand sample (between topoisomers 17′ and 18′). This discontinuity corresponds to a sudden drop in writhe in the first dimension of the gel, which was run under standard gel conditions (90 mM Tris-borate pH 8.0, 1 mM EDTA; TBE) with no intercalator. The Z-DNA is not formed in topoisomer 17′, but is stable in 18′, and so the midpoint of the B–Z transition would occur at a level of negative supercoiling between these two. The second dimension of the gel with added chloroquine serves to separate out the topoisomers, which would otherwise overlap at the point of formation of the Z-DNA. Note that in the second dimension, there is a smooth dependence of mobility on ΔLk. In particular, topoisomer 18′ has a slightly higher mobility than 17′ as expected. The presence of the intercalator has unwound the DNA, and topoisomer 15′ is

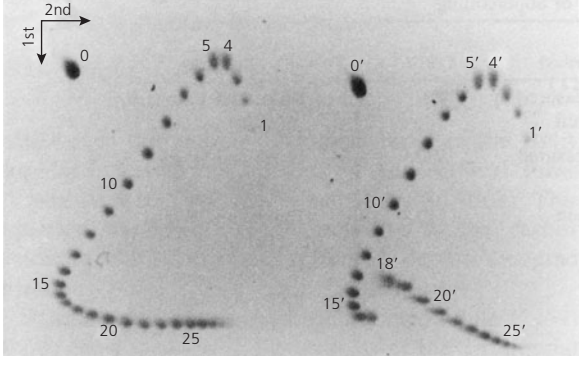


Figure 2.26 The effect of Z-DNA formation on the mobility of topoisomers in a two-dimensional gel. A two-dimensional agarose gel with two plasmid samples of the same size, one containing a Z-DNA forming region. The gel is a 0.7% agarose gel run in TBE with 1.28 μM chloroquine in the second dimension. The formation of Z-DNA is indicated by the discontinuity in the smooth progression of topoisomers (see text for details; reproduced from ref. 31 with permission; copyright (1982) Cold Spring Harbor Laboratory Press).

now approximately relaxed (Section 2.5.5); none of the negatively supercoiled topoisomers in the second dimension (16′–26′) is sufficiently supercoiled to stabilize the Z-DNA region.

The gel can be used to determine the precise level of supercoiling that stabilizes the formation of the Z-DNA. In the first dimension, Lk° lies between topoisomers 4 and 5, and the Z-DNA is formed between topoisomers 17′ and 18′; the B–Z transition occurs at a ΔLk of -13. The untwisting due to Z-DNA formation will be 32/10.5 turns for unwinding the B-DNA plus 32/12 turns for forming the Z-DNA (= 5.7 turns), so the new apparent value of ΔLk will be -7.3. The best value for the free energy of supercoiling under these conditions of gel electrophoresis (TBE) is $NK = 1450RT$ (25), giving $K = 0.33RT$ for a 4400 bp plasmid. So, the free energy for the formation of the Z-DNA, which is equal to the supercoiling free energy gain at equilibrium, is:

$$\Delta G_Z = 0.33RT(13^2 - 7.3^2) \tag{2.34}$$

which comes out as 95 kJ mol^{-1} (23 kcal mol^{-1}). These numbers are not identical to those in the original publication (31), since we have used the more recent values for the free energy of supercoiling. Two-dimensional gels have also been used in the same way to investigate cruciform formation (58).

2.9 Conclusions

This chapter contains the basics for an understanding of DNA supercoiling: the fundamental topological concepts of closed circularity and linking number, and the geometrical description in terms of twist and writhe. The practical aspects of the behaviour of supercoiled DNA during electrophoresis are important for the interpretation of many experiments where the topology of the DNA is a factor. The excess free energy associated with supercoiling and the conformation adopted by supercoiled molecules have an important influence on many DNA-associated processes *in vivo*. In the final chapter, we will consider the diverse effects of supercoiling in living cells.

2.10 Further Reading

Bloomfield, V.A., Crothers, D.M., and Tinoco, I. (2000). *Nucleic acids: structures, properties and functions*. University Science Books, Sausalito, CA.

Cozzarelli, N.R., Boles, T.C., and White, J.H. (1990). Primer on the topology and geometry of DNA supercoiling. In *DNA topology and its biological effects*. Cozzarelli, N.R. and Wang, J.C. (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 139–184.

Depew, R.E. and Wang, J.C. (1975). Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. USA* **72**, 4275–4279.

Sinden, R.R. (1994). *DNA structure and function*. Academic Press, London.

Vologodskii, A.V. and Cozzarelli, N.R. (1994). Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643.

Vologodskii, A.V. (1992). *Topology and physics of circular DNA*. CRC Press, Boca Raton, FL.

Vologodskii, A.V. (2001). Distribution of topological states in circular DNA. *Mol. Biol.* **35**, 285–297.

2.11 References

1. Watson, J.D. and Crick, F.H.C. (1953). General implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967.
2. Watson, J.D. and Crick, F.H.C. (1953). The structure of DNA. *Cold Spring Harbor Symp. Quant. Biol.* **18**, 123–131.
3. Delbrück, M. and Stent, G.S. (1957). On the mechanism of DNA replication. In *The chemical basis of heredity*. McElroy, W. and Glass, B. (eds.), The Johns Hopkins Press, Baltimore, MD, pp. 699–736.
4. Delbrück, M. (1954). On the replication of desoxyribonucleic acid (DNA). *Proc. Natl. Acad. Sci. USA* **40**, 783–788.
5. Dulbecco, R. and Vogt, M. (1963). Evidence for a ring structure of polyoma virus DNA. *Proc. Natl. Acad. Sci. USA* **50**, 236–243.
6. Weil, R. and Vinograd, J. (1963). The cyclic helix and cyclic coil forms of polyoma viral DNA. *Proc. Natl. Acad. Sci. USA* **50**, 730–738.
7. Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., and Laipis, P. (1965). Twisted circular form of polyoma viral DNA. *Proc. Natl. Acad. Sci. USA* **53**, 1104–1111.
8. Bauer, W.R. (1978). Structure and reactions of closed duplex DNA. *Annu. Rev. Biophys. Bioeng.* **7**, 287–313.
9. Depew, R.E. and Wang, J.C. (1975). Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. USA* **72**, 4275–4279.
10. White, J.H. (1969). Self-linking and the Gauss integral in higher dimensions. *Am. J. Math.* **91**, 693–728.
11. Fuller, F.B. (1971). The writhing number of a space curve. *Proc. Natl. Acad. Sci. USA* **68**, 815–819.
12. Fuller, F.B. (1978). Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl. Acad. Sci. USA* **75**, 3557–3561.

13. Crick, F.H.C. (1976). Linking numbers and nucleosomes. *Proc. Natl. Acad. Sci. USA* **73**, 2639–2643.
14. Bauer, W.R., Crick, F.H.C., and White, J.H. (1980). Supercoiled DNA. *Sci. Am.* **243 (July)**, 100–113.
15. Cozzarelli, N.R., Boles, T.C., and White, J.H. (1990). Primer on the topology and geometry of DNA supercoiling. In *DNA topology and its biological effects.* Cozzarelli, N.R. and Wang, J.C. (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 139–184.
16. Adams, C.C. (2001). *The knot book. An elementary introduction to the mathematical theory of knots.* W.H. Freeman, New York.
17. Bednar, J., Furrer, P., Stasiak, A., Dubochet, J., Egelman, E.H., and Bates, A.D. (1994). The twist, writhe and overall shape of supercoiled DNA change during counterion-induced transition from a loosely to a tightly interwound superhelix. Possible implications for DNA structure *in vivo. J. Mol. Biol.* **235**, 825–847.
18. Boles, T.C., White, J.H., and Cozzarelli, N.R. (1990). Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* **213**, 931–951.
19. Adrian, M., ten Heggeler-Bordier, B., Wahli, W., Stasiak, A.Z., Stasiak, A., and Dubochet, J. (1990). Direct visualisation of supercoiled DNA molecules in solution. *EMBO J.* **9**, 4551–4554.
20. Pulleyblank, D.E., Shure, M., Tang, D., Vinograd, J., and Vosberg, H.-P. (1975). Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: formation of a Boltzmann distribution of topological isomers. *Proc. Natl. Acad. Sci. USA* **72**, 4280–4284.
21. Bauer, W. and Vinograd, J. (1970). Interaction of closed circular DNA with intercalative dyes II. The free energy of superhelix formation in SV40 DNA. *J. Mol. Biol.* **47**, 419–435.
22. Keller, W. and Wendel, I. (1974). Stepwise relaxation of supercoiled SV40 DNA. *Cold Spring Harbor Symp. Quant. Biol.* **39**, 199–208.
23. Keller, W. (1975). Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **72**, 4876–4880.
24. Xu, Y.C. and Bremer, H. (1997). Winding of the DNA helix by divalent metal ions. *Nucleic Acids Res.* **25**, 4067–4071.
25. Rybenkov, V.V., Vologodskii, A.V., and Cozzarelli, N.R. (1997). The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling. *Nucleic Acids Res.* **25**, 1412–1418.
26. Taylor, W.H. and Hagerman, P.J. (1990). Application of the method of Phage T4 DNA ligase-catalysed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat. *J. Mol. Biol.* **212**, 363–376.
27. Duguet, M. (1993). The helical repeat of DNA at high temperature. *Nucleic Acids Res.* **21**, 463–468.

28. Wang, J.C. (1974). Degree of unwinding of DNA helix by ethidium. 1. Titration of twisted PM2 DNA-molecules in alkaline cesium-chloride density gradients. *J. Mol. Biol.* **89**, 783–801.
29. Pulleyblank, D.E. and Morgan, A.R. (1975). Sense of naturally occurring superhelices and unwinding angle of intercalated ethidium. *J. Mol. Biol.* **91**, 1–13.
30. Shure, M., Pulleyblank, D.E., and Vinograd, J.K. (1977). Problems of eukaryotic and prokaryotic DNA packaging and in vivo conformation posed by superhelix density heterogeneity. *Nucleic Acids Res.* **4**, 1183–1205.
31. Wang, J.C., Peck, L.J., and Becherer, K. (1982). DNA supercoiling and its effects on DNA structure and function. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 85–91.
32. Snounou, G. and Malcolm, A.D.B. (1983). Production of positively supercoiled DNA by netropsin. *J. Mol. Biol.* **167**, 211–216.
33. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
34. Dickerson, R.E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906–1926.
35. Hsieh, T.-S. and Wang, J.C. (1975). Thermodynamic properties of superhelical DNAs. *Biochemistry* **14**, 527–535.
36. Clendenning, J.B., Naimushin, A.N., Fujimoto, B.S., Stewart, J.M., and Schurr, J.M. (1994). Effect of ethidium binding and superhelix density on the supercoiling free energy and torsion and bending constants of p30δ DNA. *Biophys. Chem.* **52**, 191–218.
37. Naimushin, A.N., Clendenning, J.B., Kim, U.-S., Song, L., Fujimoto, B.S., Stewart, D.W., and Schurr, J.M. (1994). Effect of ethidium binding and superhelix density on the apparent supercoiling free energy and torsion constant of pBR322 DNA. *Biophys. Chem.* **52**, 219–226.
38. Shore, D., Langowski, J., and Baldwin, R.L. (1981). DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA* **78**, 4833–4837.
39. Shore, D. and Baldwin, R.L. (1983). Energetics of DNA twisting: 1. Relation between twist and cyclization probability. *J. Mol. Biol.* **170**, 957–981.
40. Shore, D. and Baldwin, R.L. (1983). Energetics of DNA twisting: 2. Topoisomer analysis. *J. Mol. Biol.* **170**, 983–1007.
41. Horowitz, D.S. and Wang, J.C. (1984). Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol.* **173**, 75–91.
42. Le Bret, M. (1979). Catastrophic variation of twist and writhing of circular DNAs with constraint? *Biopolymers* **18**, 1709–1725.
43. Jacobson, H. and Stockmayer, W.H. (1950). Intramolecular reaction in polycondensations I. The theory of linear systems. *J. Chem. Phys.* **18**, 1600–1606.
44. Levene, S.D. and Crothers, D.M. (1986). Topological distributions and the torsional rigidity of DNA. A Monte Carlo study of DNA circles. *J. Mol. Biol.* **189**, 73–83.

45. Crothers, D.M., Drak, J., Kahn, J.D., and Levene, S.D. (1992). DNA bending, flexibility, and helical repeat by cyclization kinetics. *Meth. Enzymol.* **212**, 3–29.
46. Vologodskii, A.V. and Cozzarelli, N.R. (1994). Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643.
47. Brian, A., Frisch, H., and Lerman, L. (1981). Thermodynamics and equilibrium sedimentation analysis of the close approach of DNA molecules and a molecular ordering transition. *Biopolymers* **20**, 1305–1328.
48. Shaw, S.Y. and Wang, J.C. (1993). Knotted DNA rings: probability of formation and resolution of the two chiral trefoils. *Science* **260**, 533–536.
49. Rybenkov, V.V., Cozzarelli, N.R., and Vologodskii, A.V. (1993). Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. USA* **90**, 5307–5311.
50. Stigter, D. (1977). Interactions of highly charged colloidal cylinders with applications to double-stranded DNA. *Biopolymers* **16**, 1435–1438.
51. Le Bret, M. (1980). Monte Carlo computation of the supercoiling energy, the sedimentation constant and the radius of gyration of unknotted and knotted circular DNA. *Biopolymers* **19**, 619–637.
52. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.
53. Vologodskii, A.V., Anshelevich, V.V., Lukashin, A.V., and Frank-Kamenetskii, M.D. (1979). Statistical-mechanics of supercoils and the torsional stiffness of the DNA double helix. *Nature* **280**, 294–298.
54. Vologodskii, A.V. (2001). Distributions of topological states in circular DNA. *Mol. Biol. (Mosk.)* **35**, 285–297.
55. Hao, M.-H. and Olson, W.K. (1989). Global equilibrium configurations of supercoiled DNA. *Macromolecules* **22**, 3292–3303.
56. Vologodskii, A.V., Levene, S.D., Klenin, K.V., Frank-Kamenetskii, M., and Cozzarelli, N.R. (1992). Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.* **227**, 1224–1243.
57. Klenin, K.V., Vologodskii, A.V., Anshelevich, V.V., Dykhne, A.M., and Frank-Kamenetskii, M.D. (1991). Computer-simulation of DNA supercoiling. *J. Mol. Biol.* **217**, 413–419.
58. Vologodskaia, M.Y. and Vologodskii, A.V. (1999). Effect of magnesium on cruciform extrusion in supercoiled DNA. *J. Mol. Biol.* **289**, 851–859.
59. Le Bret, M. (1984). Twist and writhing in short circular DNAs according to first-order elasticity. *Biopolymers* **23**, 1835–1867.

This page intentionally left blank

3
DNA on surfaces

3.1 **Introduction**

The development of the ideas of linking number, twist and writhe (Chapter 2) has provided powerful concepts for understanding the geometry of supercoiled DNA. However, we have seen that the actual values of twist and writhe are difficult to determine, except in a few specialized cases (Chapter 2, Section 2.4.1). An important factor that can modulate DNA supercoiling is the wrapping of the DNA on a protein surface, where, in a closed-circular molecule, the writhing of the DNA around the complex and any changes in twist of the DNA helix contribute to the overall supercoiling of the molecule. The preeminent example of this is the wrapping of DNA around the histone octamer to form the nucleosome (*Figure 3.1*), which provides the primary mode of negative supercoiling in eukaryotic chromosomes (see Chapter 6, Section 6.2.2); other examples include DNA gyrase (see Chapter 5, Section 5.3.2) and *Escherichia coli* RNA polymerase (see Chapter 6, Section 6.4.1). In such cases, as we

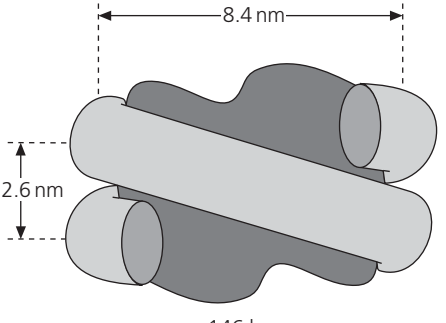


Figure 3.1 The nucleosome core particle. A schematic view of the path of DNA around the histone octamer in the nucleosome core particle.

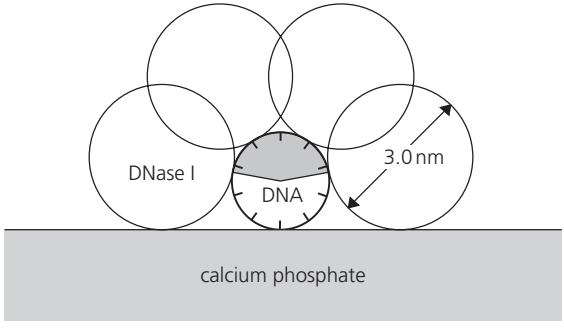


Figure 3.2 Measurement of the helical repeat of DNA by DNase I digestion. DNA (small circle) is shown bound to calcium phosphate (shaded area). The sites of possible cleavage of the DNA by DNase I (large circles) are indicated by the shaded segment (redrawn from ref. 2 with permission; copyright (1980) Nature Publishing Group).

will see, any change in twist of the DNA on protein binding is hard to disentangle from the influence of the curved surface on which the DNA is lying. This chapter describes an alternative topological description of closed-circular DNA (the surface-linking model), in terms of the winding of the DNA helix relative to the surface on which it lies, and the topology of that surface in space, which was specifically developed to address these issues (1). We will consider the application of this description to the nucleosome in some detail.

3.2 The helical repeat of DNA

The key to the discussion of the surface-linking model lies in the concept of the helical repeat of DNA, h, a quantity that we have already discussed in the first two chapters. One of the easiest ways to envisage the measurement of the helical repeat of linear DNA is to imagine the helix lying on a flat (plane) surface. h is then the average periodicity of the appearance of one of the DNA strands on the upper face of the helix, away from the surface beneath, measured in base pairs. In fact, h can be measured in exactly this way, by using DNA adsorbed to a perfectly flat calcium phosphate surface, and allowing the backbone to be cut by the nuclease DNase I, which cuts more effectively at phosphodiester bonds away from the surface (*Figure 3.2*; (2)). This is an application of the general method of DNA footprinting, in which material (e.g. protein or drugs) bound to DNA protects it from digestion by an enzymic or chemical agent at particular sites (3). The average periodicity of cutting of one of the DNA strands, measured by the length distribution of the products, is a direct measure of the helical repeat. As we saw in Chapter 2, Section 2.4.2, the average helical repeat is related to the twist of the DNA:

$$\mathrm{Tw} = \frac{N}{h} \tag{3.1}$$

This is logical since both the helical repeat and the twist are a measure of the winding of the DNA strands about the helix axis. However, most DNA, when supercoiled or lying on a protein surface such as the nucleosome (*Figure 3.1*), does not lie in a plane. Is it possible to measure the helical repeat of the DNA in such a situation? If the twist of the DNA was known, then one method would be to calculate the average helical repeat according to Equation (3.1). However, this puts things backwards: as we mentioned above, determination of the twist is not an easy matter. Twist is defined in terms of the instantaneous rotation of a point on one DNA strand around the tangent to the helix axis at that point, integrated over the whole length of the DNA (Chapter 2, Section 2.4.1, Box 2.1) (4). Since the effect of writhing the DNA is to continuously change the tangent to the helix axis, this complicates the calculation (this is in addition to the requirement that Tw + Wr = Lk, if the DNA is a closed circle (5)).

If the DNA does lie on a real surface such as a protein, we could measure the helical repeat using the DNase I method in *Figure 3.2*. This is good because we can make an actual physical measurement, but there is a problem. The measurement is being made relative to the surface on which the DNA lies, and so will depend on the shape of that surface in space, as well as on the 'real' helical repeat of the DNA. In more technical terms, by choosing a surface relative to which we will measure the periodicity of nuclease cleavage, such as a curved protein surface, we define a frame of reference for the measurement (see Box 3.1). The helical repeat measured by this method depends on the shape of the surface, in addition to any physical twisting or untwisting of the DNA helix.

The helical repeat of the DNA given by DNase I cleavage periodicity on a curved protein surface need not be the same as that derived from the twist of the DNA (i.e. N/Tw). If it differs from the standard 10.5 bp/turn, this may or may not reflect an actual over- or under-twisting of the DNA helix. The winding of DNA around, for example, the nucleosome can thus be described in principle using two frames of reference: by the twist of the DNA, which is measured relative to the local helix axis at any point (4), and by the helical repeat of the DNA relative to the surface on which it lies (i.e. in the surface frame; (1)).

We can define h_t for the twist-related, and h_s for the surface-related helical repeats. While h_s has the virtue that it is in principle an experimentally measurable quantity, h_t has the advantage that it better reflects the actual local twisting up or untwisting of the DNA helix in response to superhelical stress, or protein binding, or whatever. When DNA lies in a plane, $h_s = h_t$, but this need no longer be true if the helix axis is out of the plane. The value of h_t

BOX 3.1 **FRAMES OF REFERENCE**

The idea of different frames of reference is a potentially confusing one. The general sense of frame of reference that we are referring to is an environment or perspective (technically a set of co-ordinates) relative to which we can measure motion. One well-known example of two alternative frames is when we consider the air speed and ground speed of an aircraft (*Figure 3.3*). An observer on an aircraft measuring motion through the air (in the frame of reference of the air) might measure the speed and direction of the aircraft and represent it by the vector ***a***. On the other hand, an observer on the ground (ground frame of reference) might measure the aircraft's speed and direction as ***g***. In general these values need not be the same, but they are related by the vector representing the speed and direction of the wind over the ground (***w***). Air speed and ground speed are identical only when the wind speed is zero; normally the two frames of reference are moving relative to one another.

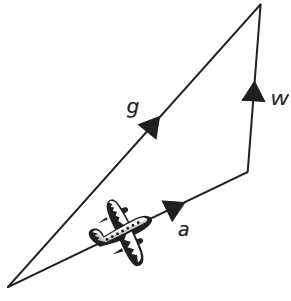


Figure 3.3 Air speed and ground speed. The relationship between the vectors representing the speed of an aircraft in two frames of reference, through the air (***a***) and over the ground (***g***), and the vector representing the wind speed (***w***).

Another example: it is clear to anyone standing on the surface of the Earth that the sun rotates around the Earth once every 24 hours. However, this apparent motion is only a quirk of the chosen frame of reference (the Earth's surface); it would equally apparent to someone in a spacecraft in interplanetary space that the sun stays still and the Earth rotates. Of course, neither perspective reflects the absolute truth, but the latter is a more useful frame of reference for a description of the mechanics of the solar system. In this case one frame of reference is rotating with respect to the other.

is a property only of the DNA conformation, whereas h_s is also a function of the surface on which the DNA lies.

As we will see, the binding of DNA to a protein surface, such as in the nucleosome, can certainly affect all the parameters describing DNA conformation, and the idea of the surface frame of reference can help us to analyse these changes.

3.3 The surface linking treatment

The surface linking treatment of DNA supercoiling, as its name implies, uses the surface frame of reference, since this has the virtue that in some cases, most notably in the nucleosome, the surface-related helical repeat can actually be measured, and the geometry of the surface on which the DNA lies can in principle be determined. The following discussion is based on the analysis of White and co-workers (1). This is the most comprehensive treatment, although the idea of the surface frame as that appropriate to DNase I measurements was developed earlier by Drew, Travers, and Klug (6, 7), and independently by Prunell and co-workers (8).

3.3.1 *DNA winding number* (Φ)

The winding number of a closed-circular DNA, Φ, describes the winding of the DNA helix around its axis in the surface frame of reference and is related to the physical measurement made by nuclease digestion of DNA on a surface. We can define this parameter by reference to *Figure 3.4*. The figure depicts a double-stranded DNA whose axis is represented by the curve *A*. One of the helical strands is represented by the curve *C*. The axis lies on a curved surface *M*. At any point along the axis (point *a*), a plane perpendicular to the axis will be crossed by *C* (at point *c*). $\boldsymbol{v}_{ac}$ is a vector in the plane along the line joining *a* and *c*. If we imagine travelling along the axis, then the vector $\boldsymbol{v}_{ac}$ will rotate around *A* as the helical strand *C* winds around it. To measure the rotation of the strand relative to the surface (in the surface frame), we need some reference marker of the curving of the surface as we move along the axis. The surface normal vector, $\boldsymbol{v}$, provides this reference. $\boldsymbol{v}$ is perpendicular to the surface at every point

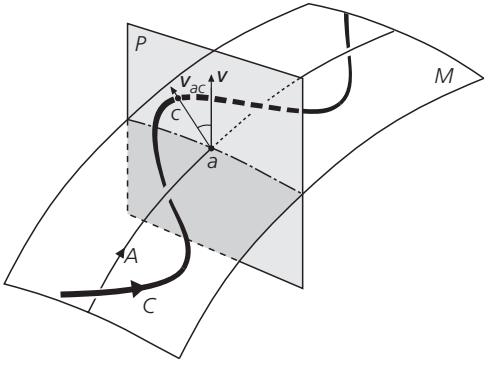


Figure 3.4 Definition of the winding number (Φ). The winding number (Φ) is given by the number of rotations of vector $\boldsymbol{v}_{ac}$ relative to vector $\boldsymbol{v}$ as point a traverses the DNA axis, *A*. See text for further details (redrawn from ref. 1 with permission; copyright (1988) AAAS).

(and is also in plane P). So, in closed-circular DNA (where the axis is a closed curve), we can count the number of times $\boldsymbol{\nu}_{ac}$ rotates past $\boldsymbol{\nu}$ as we travel round the circle once; this value is the winding number, Φ. A little thought should convince you that Φ must be an integer, since we have to end up back exactly where we started. Φ is positive if the direction of rotation is right-handed, as in *Figure 3.4* and in DNA. So, the average helical repeat relative to the surface, h_s, can now be defined in terms of the DNA winding number:

$$h_s = \frac{N}{\Phi} \tag{3.2}$$

3.3.2 *Surface linking number (SLk)*

If the closed-circular DNA is planar, and therefore the axis A lies in a plane (i.e. if M is a plane surface in *Figure 3.4*), the reference vector $\boldsymbol{\nu}$ will always point upwards. In this special case, the winding number will be equal to the linking number (Lk), the number of turns of the helix around the axis, (and also to the twist, since Wr = 0; *Figure 3.5a*). In general however, the DNA need not be planar, the helix axis A will not lie in a plane, and the reference vector $\boldsymbol{\nu}$ will not have a constant direction. In such cases, the linking number may not be equal to the winding number, but may also include a contribution

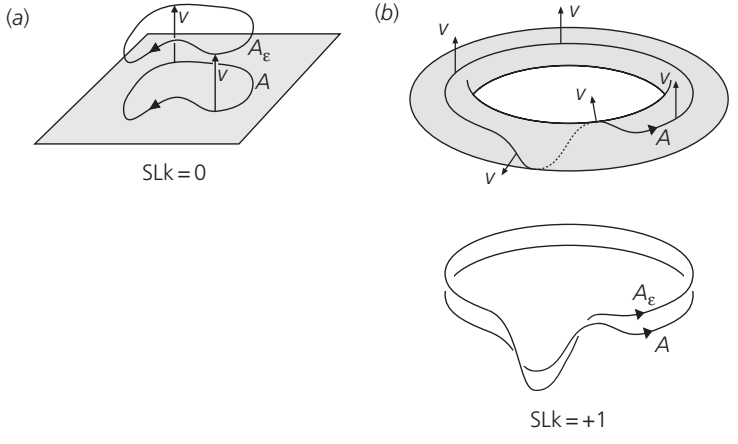


Figure 3.5 Definition of the surface linking number (SLk). The surface linking number (SLk) is the linking number of the DNA axis A and the curve A_ε, displaced along the surface normal, $\boldsymbol{\nu}$: (*a*) refers to DNA lying on a plane surface; (*b*) refers to DNA wrapped once round a torus. See text for further details ((*a*) redrawn from ref. 1 with permission; copyright (1988) AAAS).

from the change in the reference vector, that is, from the topology of the surface M. This contribution, called the surface linking number (SLk), is, in simplistic terms, the number of revolutions made by the vector $\boldsymbol{\nu}$ in space, as we travel along the DNA axis once. More rigorously, SLk is defined as follows: If a new curve A_ε is formed by joining all the points with a small displacement from A along the direction of the normal vector $\boldsymbol{\nu}$, SLk is the linking number of the two curves A and A_ε. For example, in *Figure 3.5a*, the DNA axis lies in a plane, the displaced curve A_ε lies entirely above the plane, and SLk $= 0$. In *Figure 3.5b*, the axis curve A lies on a toroidal surface, M, and makes one right-handed turn, passing through the hole in the torus. The surface normal, $\boldsymbol{\nu}$, makes one right-handed turn around A as the axis is traversed. The displaced curve A_ε and A have a linking number of $+1$ (compare with Chapter 2, Section 2.3.1; *Figure 2.6* to convince yourself of this if it is not clear), and hence SLk $= +1$.

The surface linking number is a function of the writhe of the DNA helix axis, since it is in part a property of the trajectory of the axis. The two terms are related by the following equation:

$$\text{SLk} = \text{STw} + \text{Wr} \tag{3.3}$$

where STw, the surface twist, is the twist of the displaced curve A_ε about the axis curve A. This equation thus has the same form as Lk $=$ Tw $+$ Wr (Chapter 2, Section 2.4.1). The surface twist, STw is also the correction between the measures of helical winding in the two frames of reference, the twist, Tw and the winding number, Φ:

$$\text{Tw} = \Phi + \text{STw} \tag{3.4}$$

The basic formulation of the surface linking treatment is thus that the linking number (Lk) of a closed-circular DNA may be partitioned into the DNA winding number (Φ), a measure of the rotation of the DNA helix relative to the surface on which it lies, and the surface linking number (SLk), which describes the topology of that surface in space; that is:

$$\text{Lk} = \text{SLk} + \Phi \tag{3.5}$$

For a rigorous proof of the relationship above, see White *et al.* (1). We can also incorporate the specific linking difference (σ) into this formulation. The linking difference of a closed-circular DNA is Lk $-$ Lk$^\circ$ (Chapter 2, Section 2.3.2), where Lk$^\circ$ $= \Phi^\circ = N/h^\circ$ (for a relaxed DNA whose axis

lies in a plane; *Figure 3.5a*). From Equation (3.5),

$$\Delta \mathrm{Lk} = \mathrm{Lk} - \mathrm{Lk}^\circ = \mathrm{SLk} + \Phi - \mathrm{Lk}^\circ \tag{3.6}$$

and therefore,

$$\sigma = \frac{\Delta \mathrm{Lk}}{\mathrm{Lk}^\circ} = \frac{\mathrm{SLk}}{\mathrm{Lk}^\circ} + \frac{\Phi - \mathrm{Lk}^\circ}{\mathrm{Lk}^\circ} \tag{3.7}$$

It is possible now to introduce the surface helical repeat $h_s (= N/\Phi)$, and $h^\circ (= N/\mathrm{Lk}^\circ)$:

$$\sigma = \frac{\mathrm{SLk}}{\mathrm{Lk}^\circ} + \frac{h^\circ}{h_s} - 1 \tag{3.8}$$

this equation may be solved for h_s:

$$h_s = \frac{h^\circ}{\sigma - (\mathrm{SLk}/\mathrm{Lk}^\circ) + 1} \tag{3.9}$$

In principle, this equation allows the calculation of the surface-related helical repeat for DNA of any specific linking difference, providing the surface linking number of the surface is known.

Since the surface linking number is a topological property of the path followed by the DNA on a surface, it follows that smooth deformation of that surface, that is stretching or bending, without breaking, will not change its value, as long as the DNA remains on the surface during the deformation. Since Lk is also a topological invariant, then the winding number and hence the average helical repeat also do not change on smooth deformation. In the following sections, potential applications of this approach will be considered.

As mentioned previously, in a number of situations DNA either does lie on a surface, for example that formed by a protein, or can be considered to lie on a virtual surface, and hence the surface linking treatment described above can be applied to a number of real problems.

3.4 Interwound supercoiled DNA

The most stable form adopted by negatively supercoiled DNA in solution is the interwound, or plectonemic form (see Chapter 2, Section 2.4.3). This

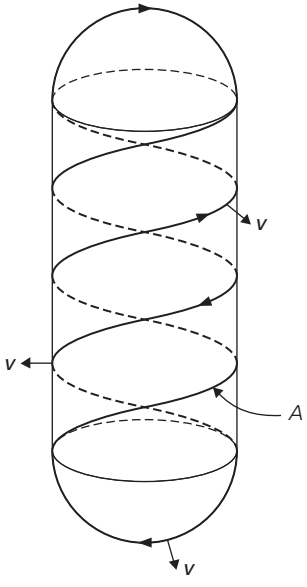


Figure 3.6 Representation of an interwound DNA helix. *A*, the axis of the DNA double helix, is wound on the surface of a cylinder with capped ends. ***v*** is the surface normal vector (see text) (redrawn from ref. 1 with permission; copyright (1988) AAAS).

conformation of DNA can be modelled as if the axis of the DNA helix lies on a cylinder with capped ends, as in *Figure 3.6*. The DNA axis *A* winds up and down the cylinder and crosses the hemispherical caps at the top and bottom; the actual dimensions of the cylinder are not too important. First, considering the value of the surface linking number for such a structure, the surface normal, ***v***, will always point out of the cylinder, and hence the displaced curve A_ε will always lie outside the cylinder. *A* and A_ε are not linked, and hence SLk = 0. This result means that from Equation (3.5):

$$\mathrm{Lk} = \Phi \tag{3.10}$$

and the surface helical repeat, h_s, from Equation (3.9), simplifies to:

$$h_s = \frac{h^\circ}{\sigma + 1} \tag{3.11}$$

This means, surprisingly, that the winding number is always equal to the linking number, and the apparent helical repeat of the DNA is a function simply of the initial helical repeat and the specific linking difference of the DNA. As an example, σ for naturally occurring plasmid DNA is typically −0.06 (Chapter 2, Section 2.3.2). If the initial helical repeat of DNA (h°) is 10.5 bp/turn, h_s comes out as 11.2 bp/turn at this level of supercoiling. It must

be emphasized again that this apparent change is a result of the changed frame of reference of the measurement (the surface of the virtual cylinder) and does not mean that the DNA is physically untwisted to the extent implied by this helical repeat. As we have seen, it is likely that the DNA is somewhat untwisted in a plectonemically wound supercoil (Chapter 2, Section 2.4.3). For interwound supercoiling, $\Delta Tw \approx 0.25\Delta Lk$, so $h_t \approx 10.65$ bp/turn for this specific linking difference. It is important not to become confused between h_s and h_t.

However, it is possible that the h_s value for plectonemic supercoiling of 11.2 bp/turn is important in real biological situations. A feature of many gene regulation and recombination systems is the formation of relatively small DNA loops (commonly up to a few hundred base pairs) by protein–protein interactions between DNA-binding proteins bound to different sites. Examples include loops in the Hin recombinase system and those formed by Lac and AraC repressors (see Chapter 6, Section 6.4.2). Measurements have been made of the helical periodicities in these loops, either by DNase I footprinting or from the periodic dependence of loop formation on the spacing between protein binding sites (see Chapter 6, Section 6.4.2). These measurements commonly give periodicities of 11.0–11.3 bp on negatively supercoiled DNA (9–12). The same periodicity of around 11 bp is seen for the probability of occurrence of certain dinucleotides such as AA/TT (AA on one strand, which is the same as TT on the other; see Section 3.3.4) in analyses of whole prokaryotic genomes, where the DNA is negatively supercoiled (13, 14). It is likely that these periodicities are revealing helical repeats in this 'plectonemic' frame of reference for negative supercoiling.

Interestingly, because of the fact, mentioned earlier, that the winding number is unchanged on smooth deformation, any change in the conformation of the DNA which results only in a smooth deformation of the virtual cylindrical surface leaves the winding number, and hence average h_s unchanged. Such a deformation would include that caused by changing conditions, such as temperature or ionic strength, or even the addition of intercalating drugs. Even though the effect of intercalators is to untwist the helix of the DNA, the effect would be to alter the winding of the DNA on the cylinder (through a change in writhe; remember, the cylinder is allowed to change shape) so as to leave the apparent helical repeat unchanged. In Equation (3.11), σ and h° change together so as to leave h_s unchanged. Thus, we can see then that the surface linking treatment is less powerful than the twist and writhe formulation for describing the conformations adopted by plectonemically supercoiled DNA, although the periodicities represented by h_s in this surface frame may reflect real periodicities in protein–DNA looping interactions, and across whole genomes.

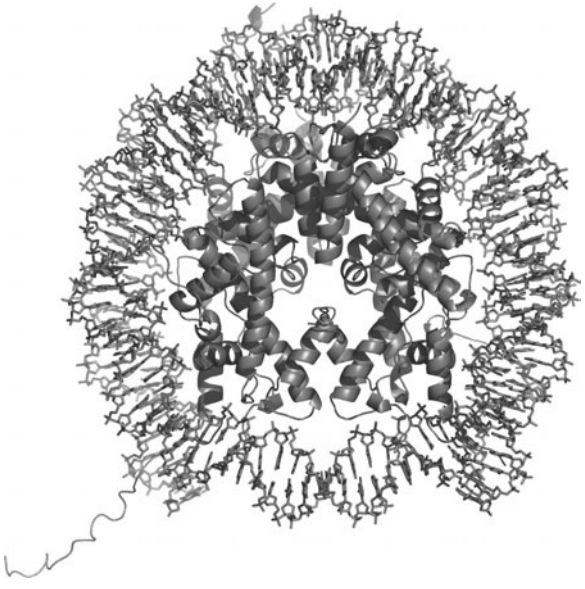


Figure 3.7 The structure of the nucleosome core particle. The image was generated using PyMOL (38) using crystallographic co-ordinates acquired from the PDB (39) and described in ref. 17.

3.5 **The nucleosome**

The most important application of the surface-linking model has been to the long-standing problem of the conformation of DNA when bound in the nucleosome (*Figure 3.7*). From the X-ray crystallographic studies of nucleosome core particles by Klug and co-workers (15, 16), it has been known for a long time that 146 bp of DNA wraps itself in about 1.8 turns around the histone octamer core, in a shallow left-handed superhelix. The dimensions of the DNA path are shown in *Figure 3.1.* More recently, X-ray crystallographic structures of the nucleosome at atomic resolution have become available (*Figure 3.7*) (17). We will consider the story of the nucleosome structure in a historical context and come to what we can learn from the high-resolution structures in Section 3.5.6.

3.5.1 *Linking difference and the 'linking number paradox'*

Since DNA is wrapped around the histone octamer in a superhelical path, it should come as no surprise that this writhing of the DNA can contribute to supercoiling if the nucleosomes are part of a closed-circular DNA. Indeed, if DNA is relaxed with a topoisomerase in the presence of a number of nucleosomes (see Chapter 2, Section 2.5.7), the DNA can be shown, when the histones are removed, to have a linking difference (ΔLk) corresponding

to approximately −1 per nucleosome present (the most precise value is -1.01 ± 0.08 (18)). This means that each nucleosome contributes a linking difference (ΔLk) of approximately −1. Although linking number is undefined for non-closed-circular DNA molecules, it is useful to consider such linking differences for segments of a closed molecule, which must sum to the total ΔLk of the whole molecule. Although the ΔLk attributable to one nucleosome is close to one, there is no requirement for a linking difference in this context to be an integer, and the ΔTw and ΔWr of the length of DNA being considered (relative to the twist and writhe of the equivalent linear unconstrained DNA) will still add up to the ΔLk (Equation (2.6)). The same argument also applies to Φ, SLk, and ΔLk for a given DNA segment; that is:

$$\Delta\text{Lk} = \Delta\text{SLk} + \Delta\Phi \tag{3.12}$$

for a segment of a closed-circular DNA, for example that bound in a nucleosome.[1]

The superhelical turns of DNA around the nucleosome core are of the toroidal type defined in Chapter 2, Section 2.4.3 (*Figure 2.9e*). We can see this by considering the representation in *Figure 3.8*. A closed-circular DNA incorporated into nucleosomes can be considered to lie wholly on the surface of a torus. The wrapping of DNA around the histones corresponds to turns around the torus, and the DNA linking the nucleosomes together lies along the line of the torus. From the definition of the surface linking number (Section 3.3.2), we can see that the total number of turns around the torus equals the SLk for the whole array of nucleosomes. This can be divided into a contribution from each nucleosome, which corresponds to ΔSLk in Equation (3.12) (refer to note 1). From the crystal structure (16), this value is −1.8.

The so-called linking number paradox refers to the discrepancy between the 1.8 turns of shallow superhelix of DNA around the histone core, which implies a contribution to linking number of approaching −2, and the measured linking contribution, which as we have seen is close to −1 per nucleosome (15, 19). Of course, although a linking difference can be defined as a characteristic of a single nucleosome structure, it is only possible to make a measurement of ΔLk in a closed-circular DNA, ideally one containing several

[1] Actually, this decomposition into values for individual segments of a circle, such as the regions bound in nucleosomes, is really only an approximation. Linking number, surface linking number, winding number, and writhe are really functions of the entire closed curve and/or surface, and the decomposition assumes that the DNA linking the nucleosomes makes no contribution to these parameters. While this will certainly *not* be true for a single nucleosome in a circular DNA, it may well be a pretty good approximation as an average for ensembles of DNA molecules each containing a number of nucleosomes. The approach outlined here relies on this assumption.

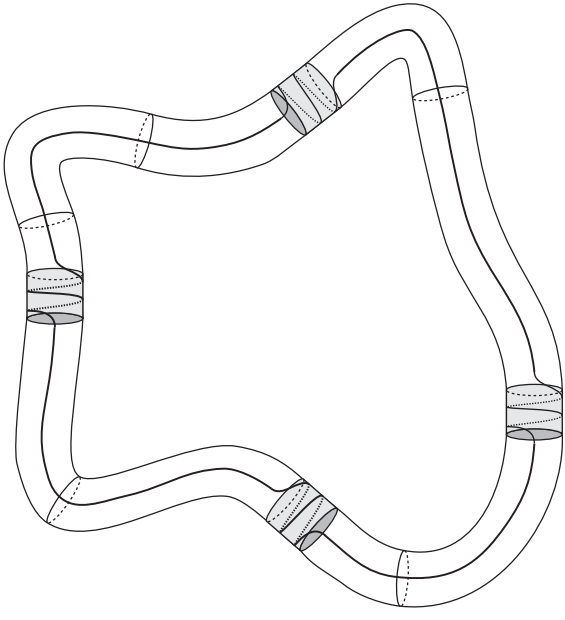


Figure 3.8 Toroidal supercoiling in nucleosome-wrapped DNA. The path of the DNA helix in an array of nucleosomes on closed-circular DNA can be assumed to lie on the surface of an imaginary torus (redrawn from ref. 1 with permission; copyright (1988) AAAS).

nucleosomes. This means that it is not known for sure whether the linking difference occurs within the structure of the nucleosome (*Figure 3.1*), or whether part of the effect reflects a superhelical ordering of the DNA which joins the nucleosomes together in the closed-circular array (*Figure 3.8*).

3.5.2 *Early resolution of the linking number paradox*

When the structure of the nucleosome was first elucidated in 1977 (15) and the nature of the linking number paradox discerned, it was soon worked out that an alteration in the helical repeat of the DNA bound in the nucleosome to 10.0 bp/turn from the value in free DNA would resolve the paradox (15, 19). This analysis was based on the twist and writhe parameters. The discrepancy between the linking difference on nucleosome formation (approx. -1) and the writhe associated with the structure (approx. -1.8)[2] was interpreted as being due to a twisting up of the DNA bound to the histone octamer. A straightforward calculation yielded the predicted value for the twist-related helical repeat (h_t) of 10.0 bp/turn.

[2] The number of superhelical turns is not exactly equal to the writhe contribution of a segment of DNA. For a shallow superhelix like the nucleosome, this is approximately true, however. See the telephone wire example (Chapter 2, Section 2.4.3) for an example where writhe decreases as the pitch of the superhelix increases.

So, one explanation for the linking number paradox is that there is a change of helical repeat of the DNA bound to the nucleosome, from 10.5 bp/turn in free DNA to 10.0 bp/turn in nucleosome-wrapped DNA. Alternatively, it may be that the DNA joining the nucleosomes in a closed-circular array contributes a linking difference of around +0.8 per nucleosome, to account for the difference between the contribution from the path of the DNA axis and the measured value.

3.5.3 *The helical repeat at the nucleosome surface*

At about the same time as the above calculations were made, the first accurate values of the periodicity of DNA on the nucleosome were made by DNase I digestion (19–21). Earlier experiments had yielded a value of roughly 10 bp/turn (22). However, in the more accurate determinations, the average value came out as 10.4 bp/turn, not dissimilar to that in free DNA, rather than the predicted 10.0 bp/turn (Section 3.5.2). However, this result and the prediction were reconciled by considering possible steric effects on DNase I digestion (19), owing to the large size of the DNase I molecule relative to the DNA helix (see *Figure 3.2*). It was reasoned that the DNase I molecule was prevented from cleaving directly opposite the surface by the steric effect of the adjacent turn of DNA helix.

As we have seen from considering the surface frame of reference (Section 3.2), the periodicity of the DNA helix relative to a surface on which it lies corresponds rigorously to h_s and is a function of the winding number, Φ. Subsequently, estimates of h_s have been made using methods that are not susceptible to the steric problem alluded to above. These data, and analysis using the winding number and surface linking number, made it possible to produce a convincing estimate of the topological and geometric parameters describing the path of DNA around the nucleosome.

3.5.4 *Measurements of h_s for the nucleosome*

When DNA is wound tightly around the nucleosome core, one of the consequences is the compression of the major and minor grooves (see Chapter 1, Section 1.2.2) on the inside and the widening of the grooves on the outside of the DNA curve. Certain sequences, notably runs of A–T base pairs, having an intrinsically narrow minor groove, should be most favourably positioned with their minor grooves on the inside of such a sharply curved segment of DNA. Analogously, runs of G–C base pairs should be more favourably aligned with minor grooves facing outwards (23). In 1985, Drew, Weeks,

and Travers (24) utilized this hypothesis in a study of the probabilities of the positioning of certain sequences in nucleosome-bound DNA. The periodicities of the appearance of short regions of A–T and G–C base pairs along many nucleosome-forming sequences, measured by DNase I footprinting of sequence-selective drugs, were averaged to yield a value of 10.17 ± 0.05 bp. Crucially, the periodicities of A–T and of G–C sequences were exactly out of phase, the A–T sequences appearing with minor grooves on the inside, as predicted. This value of 10.17 bp therefore reflects the periodicity of appearance of these sequences on the inside or outside of the DNA curve, relative to the surface of the histone core. In other words, this is a direct measure of h_s. Subsequently, sequencing of many nucleosome-containing sequences allowed the probability of occurrence of specific sequences to be plotted relative to their position in the nucleosome. The periodicity of positioning was most striking for AA/TT dinucleotide sequences (AA on one strand and therefore TT on the other) and for GC dinucleotides, and the average periodicity for these sequences was 10.22 bp (25). This same periodicity of 10.2 bp for the occurrence of AA/TT dinucleotides is apparent in more recent measurements from genome-wide sequencing (13).

More recently, an alternative DNA digestion technique, hydroxyl radical cleavage, has been used in footprinting experiments to determine the helical periodicity of DNA on the nucleosome (26, 27). Hydroxyl radical cleavage is not subject to the steric constraints associated with DNase I, and the extent of cleavage is modulated both by the protection of bound protein and by the width of the minor groove of the DNA helix (see above). The variation in such cleavage should thus provide an ideal measure of the periodicity of the DNA helix relative to a protein surface; that is h_s. Measurements of helical periodicities by hydroxyl radical cleavage in a single nucleosome (26) and in mixed sequence nucleosomes (27) have yielded an average value of 10.18 ± 0.05 bp/turn, a value in remarkably close agreement with that derived from sequence periodicities. Putting these values together, we can estimate the average value of h_s for nucleosomal DNA to be $\sim$10.2 bp/turn.

3.5.5 *Topology and geometry of DNA in the nucleosome*

With a knowledge of h_s ($\sim$10.2 bp/turn), and the length of the nucleosome wrap (146 bp), we can calculate the winding number Φ of nucleosome-wrapped DNA (Equation (3.2)):

$$\Phi = \frac{N}{h_s} = \frac{146}{10.2} = 14.3 \tag{3.13}$$

The winding number (equal to the twist) of 146 bp of linear DNA ($h_s = h_t = h^\circ = 10.5$ bp/turn) is:

$$\Phi = \frac{N}{h^\circ} = \frac{146}{10.5} = 13.9 \tag{3.14}$$

Therefore, the change in winding number on nucleosome formation is:

$$\Delta\Phi_{nuc} = +0.4 \tag{3.15}$$

The DNA is wrapped in 1.8 left-handed turns around the nucleosome core; hence ΔSLk, the change in surface linking number on nucleosome formation is -1.8 (SLk for planar, i.e. uncomplexed DNA, is zero, *Figure 3.5*). Hence, from Equation (3.12), the predicted linking difference attributable to the formation of one nucleosome can be calculated:

$$\Delta Lk_{nuc} = \Delta SLk + \Delta\Phi = -1.8 + 0.4 = -1.4 \tag{3.16}$$

The error in this value is probably around ± 0.1–0.2. Essentially the equivalent analysis has been presented by Travers and Klug (7) and by Wolffe and co-workers (26) although the terminology and the precise numbers vary. We have presented an averaged version of lower precision for simplicity. In the terminology used by Travers and Klug, the 'local' frame of reference is the equivalent of the surface frame; this analysis pre-dates the formal description of surface linking number, and should probably be recast in these terms.

When the calculated value for ΔLk on nucleosome formation (-1.4 ± 0.2) is compared to the experimentally determined value of -1.01 ± 0.08, we can see that this analysis does not resolve the linking number paradox. To account for this discrepancy, Hayes *et al.* (26) considered again that the DNA between successive nucleosomes in closed-circular DNA might be organized so as to provide a contribution to the ΔLk (Section 3.5.1). They suggested that more than 146 bp of DNA might be organized by the histone core, on the basis of their hydroxyl radical experiments. If this effect accounts for all of the discrepancy, then each linker DNA in a nucleosome array must contribute $+0.4$ to the linking difference, a value half that required in the original analysis (Section 3.5.2). Alternatively, it has been suggested by White *et al.* (28, 29) that small distortions in the shape of nucleosomes may lead to significant alterations in the surface linking number (SLk) attributable to the nucleosome structure. A similar point using different terminology has been made by Prunell and co-workers (8).

Although these findings did not completely resolve the linking number paradox, they did show that the intrinsic periodicity of DNA does change on binding to the nucleosome core. As we said previously, the winding number, Φ, is not the same as the twist of the DNA (Section 3.3.2); they are measured in different frames of references, and are related by the surface twist, STw (Equation (3.4)). The equation relating the change in these quantities on nucleosome formation is:

$$\Delta \mathrm{Tw} = \Delta\Phi + \Delta \mathrm{STw} \tag{3.17}$$

ΔSTw is a function of the geometry of the nucleosome (*Figure 3.7*) and is has been calculated from a derivation for a straight solenoidal helix to be -0.19 (5, 30). $\Delta\Phi = +0.4$ (Equation (3.15)); hence:

$$\Delta \mathrm{Tw} = 0.4 - 0.19 = +0.2 \tag{3.18}$$

In other words, the DNA is physically twisted up on nucleosome formation. In terms of the twist-related helical repeat, h_t (Section 3.2), this corresponds to a helical repeat of about 10.35 bp/turn in the frame of reference of the DNA axis. There are two possible physical explanations for this overtwisting of the helix. It has been suggested by thermodynamic calculations that bending of the DNA axis (such as in the nucleosome) can affect the twist of the DNA due to a physical deformation of the structure (31). That is, the act of bending may lead to a stable structure with a reduced helical repeat (increased twist). On the other hand, it may be that the interaction with the histone proteins of the nucleosome core may force the DNA into an overtwisted conformation (32). Hayes *et al.* prepared nucleosomes on DNA fragments containing rigid oligo(dA) · oligo(dT), flexible oligo$(\mathrm{d(AT)_n})$ and intrinsically curved segments (see Chapter 1, Section 1.4.1), which have distinctly different structural parameters, including helical repeat. They found that the different segments adopted a very uniform helical repeat, h_s, when bound in nucleosomes, as measured by hydroxyl radical cleavage. This implies that the interaction with the histone core, rather than any intrinsic property of the DNA may be the major determinant of the helical repeat on the nucleosome surface.

3.5.6 *Nucleosome structure at atomic resolution*

More recently, the quality of the structural information about the nucleosome core has been improved to atomic resolution by Richmond and

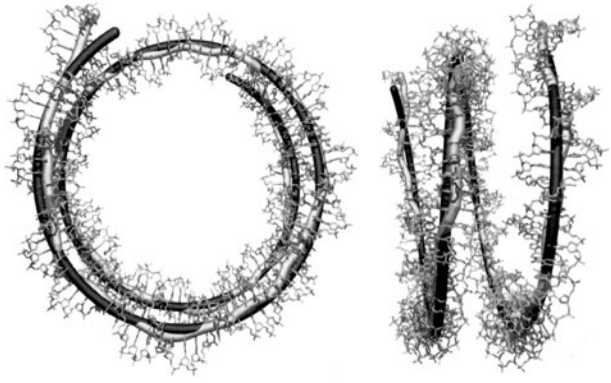


Figure 3.9 The path of the DNA superhelix in the nucleosome core particle. Two views of the structure of a symmetric 147 bp nucleosomal superhelix are shown, with the axis of the helix shown as a cylinder (light grey). The path of the ideal superhelix corresponding to this structure is shown in dark grey. The straightening of the helix at the terminus of the superhelix is particularly apparent at the top-left of the left-hand view (reproduced from ref. 35 with permission; copyright (2003) Nature Publishing Group).

co-workers (17). This has provided detailed information about the path of the DNA superhelix around the histone octamer (*Figure 3.7*), and has revealed a number of features of relevance to the discussion of the linking difference associated with nucleosome formation. Although a symmetrical 146 bp sequence was used in the structure determination, which should have led to a structure with the axis of symmetry passing between two base pairs, the structure turned out to be asymmetric, with a base pair on the symmetry axis, and one 72 bp half-sequence stretched in one short region to fit the dimensions of the other 73 bp half. Subsequently a symmetric structure containing 147 bp was produced (33, 34), the DNA conformation of which has been analysed in detail (35) (*Figure 3.9*). Since the individual base-pair steps in the sequence are visible in the structures, it is now possible to compute the twist (Tw) or h_t of the DNA by summing the individual twist values around the local DNA helix axis at each base-pair step (see Chapter 2, Section 2.4.1; Box 2.1). From the twist it is possible to compute Φ and h_s using the known geometry of the superhelix, the reverse of what we did in Section 3.5.5 using Equation (3.17). The averaged values from the whole superhelix, when taking into account the stretching of one half of a 146 bp sequence (which is probably a feature of the 'natural' 146 bp structures), are identical (within experimental error) to the values derived from sequence periodicities and hydroxyl radical footprinting (Sections 3.5.4 and 3.5.5): $h_s \approx 10.2$ bp/turn and $h_t \approx 10.35$ bp/turn (35). The way in which the DNA conformation adapts to produce the rather extreme bending of the DNA in the nucleosome superhelix is a complex function

of the DNA sequence and its interaction with the individual histone proteins, and the overtwisting of the DNA in the nucleosome is probably a consequence of this interaction.

While the crystal structures provide impressive confirmation of the earlier analyses of nucleosomal DNA, we are apparently not much closer to a solution to the mismatch between the calculated and measured linking differences of a nucleosome. However, the analyses up until now have been based on values averaged over the whole nucleosome, and it may be that variations in the structure of DNA along the path of the superhelix, which have become apparent from the high-resolution structures, may provide the answer. It was immediately clear from the first atomic resolution structure that the path of the superhelix is not smooth, and in particular the outmost 10 bp segments at each end of the superhelical wrap are approximately straight, and do not contribute to the number of superhelical turns (*Figure 3.9*). The precise value depends on how you make the measurement, but a value of 1.61 superhelical turns for the central 129 bp is pretty close (35). Recalculating on this basis, combining Equations (3.13)–(3.16), gives:

$$\Delta \mathrm{Lk}_{\mathrm{nuc}} = \Delta \mathrm{SLk} + \Delta\Phi = -1.61 + \left(\frac{129}{10.2} - \frac{129}{10.5}\right) = -1.25 \qquad (3.19)$$

Furthermore, it is likely that the value of the winding number or h_s is not uniform through the nucleosome structure. The value for the outer two turns is probably close to 10.5 bp/turn, and in addition, the central helical turn in the structure is also straight in the crystal structure and may have an h_s as high as 10.7 bp/turn (7). Sequence periodicities and footprinting data suggest that the remaining 118 bp (2×59 bp), comprising the two halves of the wrap that contribute to the superhelical turns, has a value of h_s of 10.0 bp/turn (7, 26, 27, 36). If that is the case, then the calculation becomes:

$$\Delta \mathrm{Lk}_{\mathrm{nuc}} = \Delta \mathrm{SLk} + \Delta\Phi = -1.61 + 2\left(\frac{59}{10.0} - \frac{59}{10.5}\right) = -1.05 \qquad (3.20)$$

This value is now within the error of the measured linking difference of a nucleosome (-1.01 ± 0.08). The likely validity of this calculation should not be overstated, since some of these data come from nucleosomes with rather artificial repeating sequences; it is included here merely to make the point that heterogeneity in the path and the winding of the nucleosome superhelix may help to explain the linking difference observed. It may be that the definitive version of the 25-year saga of the linking number paradox remains to be written, but application of the surface linking model to data from

measurements of the periodicity of the nucleosome DNA relative to the histone octamer surface, combined with high-resolution structural data, have come close to an explanation of the experimental linking difference measurement.

3.5.7 *Summary*

It may be that you found the previous sections on the nucleosome slightly hard going; there is no doubt that these different frames of reference and different measures of helical repeat have the potential to confuse. If so, here is the executive summary:

The most recent high-resolution information on the geometry of the nucleosome and measurements of the periodicity of DNA on the nucleosome surface, when analysed using the surface linking approach, have led to the following conclusions. The change in linking number attributable directly to the formation of the nucleosome core on a closed-circular DNA is calculated to be approximately −1.25. The experimentally determined value for an array of nucleosomes is −1.01 per nucleosome. The discrepancy may be accounted for either by superhelical ordering of the DNA linking the nucleosomes in the array, or, possibly, by distortions in the shape of the nucleosome. However, it is possible that variations in the bending and twisting of the DNA along the length of the nucleosome superhelix are sufficient to explain the remaining discrepancy. These results imply that the DNA is physically twisted up on nucleosome formation, by +0.2 turns over the whole DNA wrap; this corresponds to a helical repeat in the frame of the helix axis (given by N/Tw) of around 10.35 bp/turn, compared with 10.5 bp/turn for uncomplexed DNA. In natural 146 bp nucleosomes, the twisting up of the helix is probably a result of a complex interplay between the flexibility of a specific sequence and interactions with the histone proteins in the octamer core.

3.6 **Conclusions**

This chapter has described an alternative topological description of the supercoiling of DNA, which is complementary to the twist and writhe formulation described in Chapter 2. While it may not replace the twist and writhe approach for the analysis of the conformation of free DNA circles, it is very helpful for discussing the complexities of DNA helically wound on a protein surface, and may help us understand the periodicities measured for DNA loops. An alternative application, not discussed here, is to DNA catenanes

(see Chapter 4) toroidally wound about each other (37). A problem to be aware of when considering the surface linking approach is the confusion resulting from the seductive, but false, equation of the winding number, Φ, and the twist, Tw, of a closed-circular DNA, or reference to the 'helical repeat' of the DNA without being specific about the frame of reference used.

3.7 Further Reading

Cozzarelli, N.R., Boles, T.C., and White, J.H. (1990). Primer on the topology and geometry of DNA supercoiling. In *DNA topology and its biological effects*. Cozzarelli, N.R. and Wang, J.C. (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 139–184.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* **389**, 251–260.

Prunell, A. (1998). A topological approach to nucleosome structure and dynamics: the linking number paradox and other issues. *Biophys. J.* **74**, 2531–2544.

Richmond, T.J. and Davey, C.A. (2003). The structure of DNA in the nucleosome core. *Nature* **423**, 145–150.

White, J.H., Cozzarelli, N.R., and Bauer, W.R. (1988). *Science* **241**, 323.

White, J.H., Gallo, R.M., and Bauer, W.R. (1992). Closed-circular DNA as a probe for protein-induced structural changes. *Trends Biochem. Sci.* **17**, 7.

3.8 References

1. White, J.H., Cozzarelli, N.R., and Bauer, W.R. (1988). Helical repeat and linking number of surface wrapped DNA. *Science* **241**, 323–327.
2. Rhodes, D. and Klug, A. (1980). Helical periodicity of DNA determined by enzyme digestion. *Nature* **286**, 573–578.
3. Tullius, T.D. (1989). Physical studies of protein-DNA complexes by footprinting. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 213–237.
4. Fuller, F.B. (1978). Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl. Acad. Sci. USA* **75**, 3557–3561.
5. White, J.H. and Bauer, W.R. (1986). Calculation of the twist and the writhe for representative models of DNA. *J. Mol. Biol.* **189**, 329–341.
6. Drew, H.R. and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186**, 773–790.
7. Travers, A.A. and Klug, A. (1987). The bending of DNA in nucleosomes and its wider implications. *Phil. Trans. R. Soc. Lond.* **B 317**, 537–561.
8. Zivanovic, Y., Goulet, I., Revet, B., Le Bret, M., and Prunell, A. (1988). Chromatin reconstitution on small DNA rings II. DNA supercoiling on the nucleosome. *J. Mol. Biol.* **200**, 267–290.

9. Muller, J., Oehler, S., and MullerHill, B. (1996). Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* **257**, 21–29.
10. Haykinson, M.J. and Johnson, R.C. (1993). DNA looping and the helical repeat *in vitro* and *in vivo*: effect of HU protein and enhancer location on Hin invertasome assembly. *EMBO J.* **12**, 2503–2512.
11. Law, S.M., Bellomy, G.R., Schlax, P.J., and Record, M.T. (1993). *In vivo* thermodynamic analysis of repression with and without looping in *lac* constructs. Estimates of free and local *lac* repressor concentrations and of physical properties of a region of supercoiled plasmid DNA *in vivo*. *J. Mol. Biol.* **230**, 161–173.
12. Bellomy, G.R. and Record, M.T. (1990). Stable DNA loops *in vivo* and *in vitro*: roles in gene regulation at a distance and in biophysical characterisation of DNA. *Prog. Nucleic Acid Res. Mol. Biol.* **39**, 81–128.
13. Widom, J. (1996). Short-range order in two eukaryotic genomes: relation to chromosome structure. *J. Mol. Biol.* **259**, 579–588.
14. Herzel, H., Weiss, O., and Trifonov, E.N. (1999). 10–11 bp periodicities in complete genomes reflect protein structure and DNA folding. *Bioinformatics* **15**, 187–193.
15. Finch, J.T., Lutter, L.T., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* **269**, 29–36.
16. Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7Å resolution. *Nature* **311**, 532–537.
17. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* **389**, 251–260.
18. Simpson, R.T., Thoma, F., and Brubaker, J.M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell* **42**, 799–808.
19. Klug, A. and Lutter, L.C. (1981). The helical periodicity of DNA on the nucleosome. *Nucleic Acids Res.* **9**, 4267–4283.
20. Lutter, L.C. (1979). Precise location of DNase I cutting sites in the nucleosome core determined by high-resolution gel electrophoresis. *Nucleic Acids Res.* **6**, 41–56.
21. Prunell, A., Kornberg, R.D., Lutter, L., Klug, A., Levitt, M., and Crick, F.H.C. (1979). Periodicity of deoxyribonuclease I digestion of chromatin. *Science* **204**, 855–858.
22. Noll, M. (1974). Internal structure of the chromatin subunit. *Nucleic Acids Res.* **1**, 1573–1578.
23. Drew, H.R. and Travers, A.A. (1984). DNA structural variations in the *E. coli tyr T* promoter. *Cell* **37**, 491–502.
24. Drew, H.R., Weeks, J.R., and Travers, A.A. (1985). Negative supercoiling induces spontaneous unwinding of a bacterial promoter. *EMBO J.* **4**, 1025–1032.
25. Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675.

26. Hayes, J.J., Tullius, T.D., and Wolffe, A.P. (1990). The structure of DNA in a nucleosome. *Proc. Natl. Acad. Sci. USA* **87**, 7405–7409.
27. Hayes, J.J., Clark, D.J., and Wolffe, A.P. (1991). Histone contributions to the structure of DNA in the nucleosome. *Proc. Natl. Acad. Sci. USA* **88**, 6829–6833.
28. White, J.H., Gallo, R., and Bauer, W.R. (1989). Effect of nucleosome distortion on the linking deficiency in relaxed minichromosomes. *J. Mol. Biol.* **207**, 193–199.
29. White, J.H., Gallo, R., and Bauer, W.R. (1989). Dependence of the linking deficiency of supercoiled minichromosomes upon nucleosome distortion. *Nucleic Acids Res.* **17**, 5827–5835.
30. White, J.H. and Bauer, W.R. (1989). The helical repeat of nucleosome-wrapped DNA. *Cell* **56**, 9–11.
31. Levitt, M. (1978). How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl. Acad. Sci. USA* **75**, 640–644.
32. Hayes, J.J., Bashkin, J., Tullius, T.D., and Wolffe, A.P. (1991). The histone core exerts a dominant constraint on the structure of DNA in the nucleosome. *Biochemistry* **30**, 8434–8440.
33. Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *J. Mol. Biol.* **319**, 1097–1113.
34. Davey, C.A. and Richmond, T.J. (2002). DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl. Acad. Sci. USA* **99**, 11169–11174.
35. Richmond, T.J. and Davey, C.A. (2003). The structure of DNA in the nucleosome core. *Nature* **423**, 145–150.
36. Negri, R., Buttinelli, M., Panetta, G., De Arcangelis, V., Di Mauro, E., and Travers, A. (2001). Sequence dependence of translational positioning of core nucleosomes. *J. Mol. Biol.* **307**, 987–999.
37. Wasserman, S.A., White, J.H., and Cozzarelli, N.R. (1988). The helical repeat of double-stranded DNA varies as a function of catenation and supercoiling. *Nature* **334**, 448–450.
38. DeLano, W.L. (2002). The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org
39. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.

This page intentionally left blank

4
Knots and catenanes

4.1 Introduction

To most people the word 'knot' conjures up images of boy scouts and bowlines, but, in the context of this book, it refers to 'topological' knots in DNA. The distinction between the everyday sense of the word 'knot' (as applied, for example, to a piece of string) and a topological knot, is that the latter must be a closed curve. Like linking number (Chapter 2, Section 2.3.1), topological knotting is a concept that is only meaningful in a closed curve. Any knot tied in a piece of string becomes a topological knot when the two ends are joined together. 'Catenane' is a less-familiar term and refers to the topological interlinking of two or more rings, such as in the links of a chain.[1] In Chapter 2, rubber tubing and ribbon were used to model the behaviour of DNA, and with both these materials we can also make knots and catenanes. It may seem surprising that DNA can form these structures too, but, as this chapter shows, knots and catenanes do indeed occur in DNA and we discuss below their occurrence, properties, and biological relevance. Since knots and catenanes are topological in nature they can be investigated and understood using similar techniques and concepts to those used for supercoiling.

4.2 Knots

4.2.1 *Occurrence of knots*

James Wang and co-workers first observed knots in single-stranded DNA in 1976 (1). During studies of the single-stranded circular DNA of bacteriophage fd, they found that treatment of the DNA with the enzyme

[1] It's worth pointing out that ordinary closed-circular double-stranded DNA is a catenane of two single strands.

Escherichia coli topoisomerase I (then called ω protein, see Chapter 5), yielded a new species that sedimented faster than the untreated DNA. Examination of this new species by electron microscopy showed that it consisted of DNA rings containing at least three cross-over points. By contrast, the untreated single-stranded circular bacteriophage fd DNA showed very few cross-over points. This is reminiscent of the behaviour of polyoma DNA we described in Chapter 2, Section 2.2.2, although, of course, single-stranded DNA cannot be supercoiled, so that cannot be the explanation here. It was concluded that the fast-sedimenting species were in fact knotted single-stranded circles generated by the action of topoisomerase I. They further showed that the knotted species could be converted back to the unknotted form by the same enzyme; under high salt conditions topoisomerase I can convert circular bacteriophage fd DNA to a knotted form, whereas under low salt conditions the enzyme catalyses the reverse reaction. This can be explained by the effect of salt on the excluded volume of DNA, as described in Chapter 2, Section 2.7.2, that is, at high salt the intramolecular repulsion of DNA is greatly screened, which favours conversion of the DNA to the knotted form. Under low salt conditions, however, the intersegmental ionic repulsion is strong and therefore promotes conversion of knots into unknotted circles. Specific types of single-stranded knots can also be produced artificially using synthetic oligonucleotides and DNA ligase (2).

Knots in naturally occurring double-stranded DNA were first observed in bacteriophage P2 DNA in 1981, again by Wang and co-workers (3). Mature bacteriophage P2 DNA consists of double-stranded molecules 33 kilobases long with 19-nucleotide single-strand complementary extensions at each end. In solution these single-strand extensions can base pair to form circular and concatameric (joined end-to-end) species. The DNA found in tail-less P2 capsids (heads) is much less viscous than that found in the mature bacteriophage particle. This reduced viscosity is due to the presence of knotted DNA. Sedimentation experiments showed that head DNA contains a large fraction of material that sediments more rapidly than the DNA isolated from the mature bacteriophage. Electron microscopy showed that many of the molecules are highly condensed, consistent with them being in a highly knotted form. The structure of the knotted species is dependent upon hybridization of the 19-nucleotide extensions; disruption of the hydrogen bonding by heating converted the DNA to the linear form. Thus the DNA molecules are knotted, double-stranded DNA rings containing one nick (broken phosphodiester bond) in each strand. These nicks can be sealed by the action of DNA ligase. Sealing of one nick leads to the production of a singly nicked, knotted DNA circle which, under alkaline conditions,

dissociates to a single-stranded linear species and a highly knotted single-stranded ring. Sealing of both nicks generates fully closed-circular knotted circles that cannot be resolved into single-stranded components by alkali treatment.

Naturally occurring knots in double-stranded DNA have also been found in the tail-less capsids of bacteriophage P4 DNA (4). The origin of the knots in P2 and P4 head DNA is very likely to be random cyclization of the DNA in the confined space of the bacteriophage head, where the two ends of the DNA encounter each other and hybridize after a random walk through the tangle of DNA (5). (Imagine the knots that might occur when joining the ends of a long piece of string crammed into a restricted space.) Studies of knotting probability in P4 DNA have been carried out and show that knotting has a strong dependence on solution conditions, specifically counterion concentration; this is the basis for the determination of the effective diameter of DNA (as a descriptor of the excluded volume of the polymer) discussed in Chapter 2, Section 2.7.2 (6, 7). Coupling these experimental data with simulations of knotting probability, it was possible to calculate the effective diameter of the DNA double helix. For example, at approximately physiological salt concentrations (0.15 M NaCl), the effective diameter of double-stranded DNA was found to be ∼5 nm (50 Å), which is rather larger than the diameter of B-form DNA, ∼2 nm (20 Å) (see Chapter 1). The consequences of this are particularly relevant for plectonemically wound supercoiled DNA (Chapter 2) where there are potential close contacts between different portions of the DNA duplex (7, 8).

Double-stranded DNA knots can also be generated in the test tube using a variety of enzymes. We have seen that the enzyme *E. coli* topoisomerase I can knot single-stranded bacteriophage fd DNA. If double-stranded DNA circles bearing a single-strand nick are used as substrates for this enzyme, knotted DNA species can also be formed (9). These species can be distinguished from unknotted forms by agarose gel electrophoresis and electron microscopy (e.g. *Figure 4.1*). The knotting of intact double-stranded DNA circles can be achieved by the action of type II topoisomerases such as the enzyme from bacteriophage T4 (10). (For a fuller description of the reactions of topoisomerases, see Chapter 5.) Knotted double-stranded DNA can also be generated by recombination enzymes (see Chapter 6, Section 6.6).

So, in summary, knots can be found or created in both single- and double-stranded DNA molecules. The mathematical description and theory of knots is a vast subject that is largely outside the scope of this book. The section that follows just scratches the surface of this topic to give you enough information to adequately describe and understand the knots found in DNA.

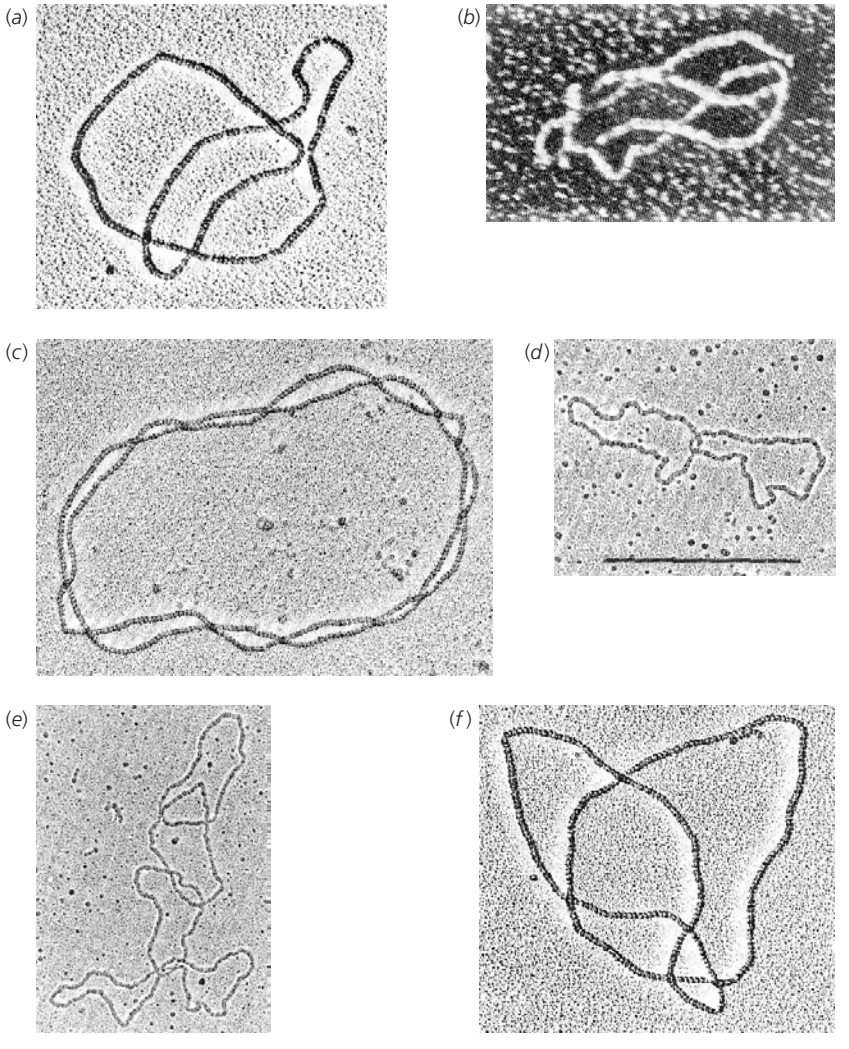


Figure 4.1 Electron micrographs of DNA knots and catenanes. Knotted and catenated DNAs were coated with RecA protein prior to visualization by electron microscopy. (*a*) Trefoil (three-noded knot) (11). (*b*) Five-noded twist knot (48). (*c*) 13-noded torus knot (49). (*d*) Singly linked catenane (32). (e) Catenane consisting of five circles (32). (*f*) Figure-eight (five-noded) catenane (11). (*a*), (*f*) reproduced with permission; copyright (1983) Nature Publishing Group. (*b*), (*c*), (*d*), (*e*) reproduced with permission; copyright (1987, 1985, 1980) Elsevier.

4.2.2 *Description of knots*

The simplest type of knot that can be made in closed-circular DNA, or any other closed curve, is a trefoil knot (or three-noded knot, see later),

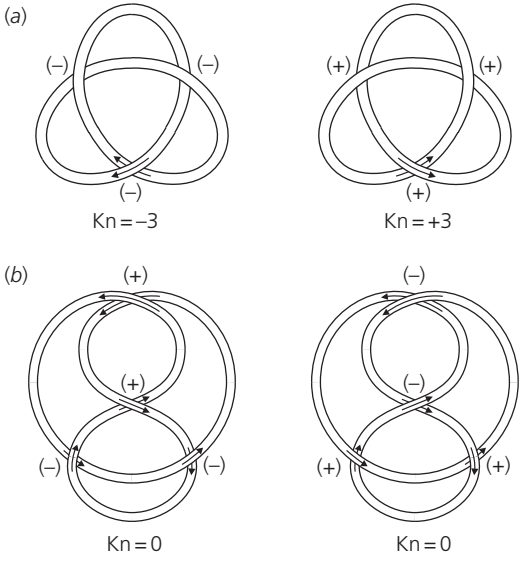


Figure 4.2 Illustration of knots. (*a*) Shows the two isomers of the trefoil (three-noded) knot. (*b*) Shows the four-noded (or figure-eight) knot and its superimposable mirror image.

so called because the structure when laid flat can be seen to have three lobes (*Figure 4.2a*). The easiest way to generate this is to take a length of string and pass one end around the other, forming a simple loose overhand knot, and then secure the ends, either by holding them firmly in one hand or by tying (e.g. a tight reef knot at the ends or sticking the ends together with glue or tape to make the string 'closed'). To reveal the trefoil structure you must lay the string flat on a surface to identify the three lobes and three crossings (nodes), which you can manipulate to produce something like the structures shown in *Figure 4.2a*. A little effort should convince you that it is impossible to make a representation showing less than three nodes (although you can easily introduce more). This representation thus shows the minimum number of nodes: it is the 'minimum plane projection' of the trefoil knot. There are in fact two distinguishable trefoil knots. If the original overhand knot was made in a right-handed (clockwise) sense then the resulting trefoil will be equivalent to that illustrated on the right in *Figure 4.2a*. Conversely, a left-handed (counterclockwise) knot will produce a trefoil like that on the left in *Figure 4.2a*. The illustrations in *Figure 4.2a* are mirror images of each other and are most easily distinguished by the signs of their 'nodes'. As described in Chapter 2, nodes are cross-over points of DNA helices. If a polarity (or directionality) is assigned to the DNA (the direction is completely arbitrary in the case of a knot), then the nodes can be given a $+1$ or -1 designation. As we describe in Chapter 2, Section 2.3.1, if arrows showing the

polarity are drawn at the site of the node (as shown in *Figure 4.2a*), the node sign is determined by the direction of movement required to align the arrows on the upper and lower curve with a rotation of the upper arrow of less than 180° (see Chapter 2, Section 2.3.1). A clockwise rotation defines a negative (−1) node; a counterclockwise rotation a positive (+1) node. (Alternatively, if rotation is only permitted in the clockwise direction, a rotation of $<180°$ defines a −1 node, while a rotation of $>180°$ defines a +1 node.)[2]

The discussion above demonstrates the usefulness of the node convention, which can be applied to supercoiling, knotting and, as described below, to catenation. In DNA the sign of a node can be determined by electron microscopy, particularly if the DNA is coated with protein, such as the *E. coli* RecA protein (11) (see *Figure 4.1*). The sum of the nodes in the minimal plane projection is defined as the intrinsic linkage of the knot, Kn, and is −3 or +3 for the two trefoils (12). Thus the Kn notation serves to distinguish the two forms of a three-noded knot. The four-noded, or figure-eight, knot (*Figure 4.2b*) has two negative and two positive nodes, with Kn = 0. In fact the mirror image of this knot is superimposable on itself and there is, rather surprisingly, only one four-noded knot. (To convince yourself of this, it is worth making a rubber tubing or string model of the knot on the left in *Figure 4.2b* and manipulating it so that it looks like the one on the right; this is quite fun to do and will help you in your understanding of nodes.) For knots with five or more nodes there are different knots with the same value of Kn; so, Kn does not describe such knots uniquely.

The classification and nomenclature of knots is a subject of some complexity and a number of systems have been devised to describe them (13). The most commonly used nomenclature is the Alexander–Briggs notation, n_i, where n is the number of nodes in the minimal plane projection, and i distinguishes knots with the same number of nodes. Tables illustrating knots with different numbers of nodes can be found in the classic book, *Knots and Links* by Dale Rolfsen (14) and more recently in *The Knot Book* by Colin Adams (15). For example, there is only one knot with 3 nodes and one with 4 (3_1 and 4_1), ignoring chirality (see later), but there are 49 listed that have 9 nodes (9_1–9_{49}). Note that the Alexander–Briggs notation does not designate the chirality (handedness) of the knot, but this can be indicated by using plus and minus signs. For example, the right-handed and left-handed trefoils can be written as $+3_1$ and -3_1.

[2] Note that the nodes in Chapter 2 referred to single strands of DNA crossing each other, whereas here we are considering self-crossing of a single knotted curve, which could equally apply to single- or double-stranded DNA knots.

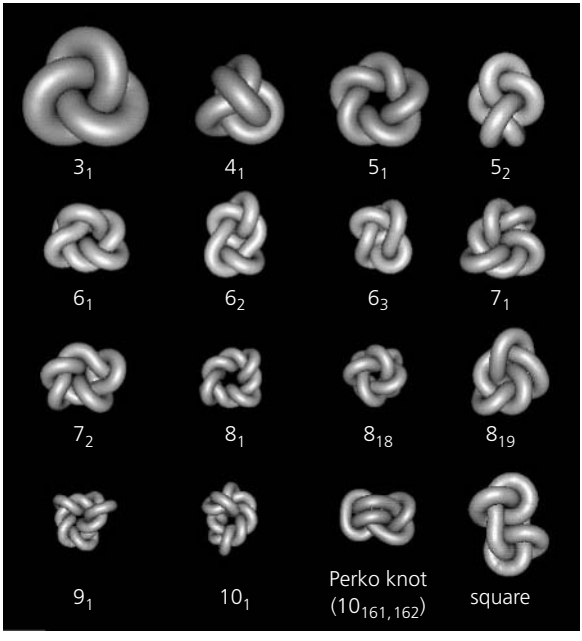


Figure 4.3 Ideal representation of knots. Ideal representations of various types of knots, labelled according to the Alexander–Briggs notation (reproduced from ref. 16 with permission; copyright (1996) Nature Publishing Group).

A development of this type of classification has been the definition of 'ideal' knots (16). For a knot of given topology assembled from a tube of uniform diameter, the ideal form is the geometrical configuration having the highest ratio of volume to surface area, that is, the shortest piece of tube that can be closed to form the knot. This is determined using a Metropolis Monte Carlo simulation, similar to that described for DNA conformation in Chapter 2, Section 2.7.3. In other words, one determines the shortest piece of tube of a given diameter that can be closed to form the knot in question. This ideal geometrical representation of knots is shown in *Figure 4.3*. As we can see, the more complex the knot the longer the piece of tubing (of a given diameter) required to form the knot. The length to diameter ratios (L/D) of the tubing in these ideal forms can be used to rank knots in terms of their complexity. Another measure of knot complexity for ideal knots is the average crossing number, which is the average number of crossings seen in different plane projections of the curve (for a given knot type, this is always greater than the Kn value). Average crossing number is found to have a linear relationship with knot L/D values for a number of simple knots (16).

Composite knots are knots composed of two or more independent knots (or factor knots) on the same string. One of the knots shown in *Figure 4.3*, the square knot, is in fact a composite knot, and is obtained by tying two trefoils

(3_1) of opposite handedness on the same string. As seen in *Figure 4.3*, such complex knots can also be represented and analysed using the ideal geometrical representation (17).

Knots can also be classified into a number of different generic types of which 'torus' knots and 'twist' knots are perhaps the best known. Torus knots are characterized by the all-interwound path of the DNA, which can be thought of as lying on the surface of an imaginary torus (cf. Chapter 2, Section 2.4.3). An example of such a knot is shown in *Figure 4.1c*. Torus knots can have only odd numbers of nodes. The other category, twist knots, consist of an interwound region with any number of nodes, and a two-noded interlocking region. An example of a twist knot is the four-noded knot shown in *Figure 4.2b*. Interestingly trefoils (three-noded knots) belong to both families.

4.3 **Catenanes**

4.3.1 *Occurrence of catenanes*

Although knots in DNA can be readily generated in the test tube, they are not particularly common in nature. Catenanes, on the other hand, are much more common and have been discovered in a number of diverse biological systems. Catenated DNA molecules were first observed in 1967 in the mitochondria of human cells (18, 19). Mitochondrial DNA is known to exist as closed-circular duplex molecules. Caesium chloride density gradient centrifugation of mitochondrial DNA in the presence of ethidium bromide produced three bands. The lowest band consisted largely of closed-circular monomer DNA and the top band of topologically unconstrained (nicked or linear) DNA. These two forms have different buoyant densities under these conditions due to the preferential binding of ethidium to topologically unconstrained DNA (see Chapter 2, Section 2.6.1). The middle band was shown by electron microscopy to contain catenated DNA molecules, in which one molecule was a closed-circular duplex and the other a nicked-circular duplex (18). Such species would be expected to have an intermediate buoyant density.

Catenated DNA has also been found in the mitochondria of trypanosomes. Here the mitochondrial genome consists of a compact network of DNA, known as kinetoplast DNA (kDNA), with a molecular mass of up to 4×10^{10} Da, consisting of thousands of DNA circles held together by catenation (20). These networks can be visualized by electron microscopy and contain two types of circular DNA molecules: maxicircles, 6–12 μm in length (20–38 kbp), and minicircles, 0.2–0.8 μm in length (0.7–2.5 kbp). It has been shown that kDNA can be decatenated by type II topoisomerases (see Chapter 5) but not

by type I topoisomerases (21), suggesting that the kDNA consists of covalently closed double-stranded circles (see Chapter 5, Section 5.2). For example, the *Crithidia fasciculata* mitochondrial DNA consists of ~5000 minicircles (which are relaxed) and ~25 maxicircles (21). Studies of the *C. fasciculata* minicircle network have shown that it behaves as a two-dimensional lattice, with each circle linked on average to three neighbours (22). This 'chain-mail' arrangement of kDNA provides a mechanism for packaging the DNA, although the biological significance of such networks is not entirely clear.

Catenanes in DNA were first made artificially by cyclizing phage λ DNA in the presence of circular phage 186 DNA (23). This generated a band on a caesium chloride gradient that was distinct from that derived from either of the phage DNAs alone. Positive identification of this band as catenated DNA was subsequently made by electron microscopy (e.g. *Figure 4.1*) (24).

Perhaps the most common occurrence of catenated DNA is as an intermediate during the replication of circular DNA molecules. This has been observed in a variety of systems including bacterial plasmids (25–27) and the animal virus SV40 (28) and may be ubiquitous. Catenane formation during replication is one of the potential topological consequences of replication (cf. supercoiling in Chapter 2, Section 2.2.1) and will be considered further in Chapter 6. DNA catenanes may also be generated by a variety of enzymes in the test tube; these include Int recombinases, resolvases, and DNA topoisomerases (29–32). As with unknotting, the DNA topoisomerases can also catalyse the decatenation (unlinking) of catenated DNA rings (see Chapter 5, Section 5.2) (10, 31, 32).

4.3.2 *Description of catenanes*

DNA catenanes can be described using the node convention discussed above for knots, although for interlocked DNA rings the polarity (i.e. the directionality of the DNA) is not necessarily arbitrary. For homologous DNA circles, or ones derived from a single ring by recombination, the DNA sequence provides a way to orientate the rings. For circles of different origins, the assignment is arbitrary. The simplest type of catenane is a pair of singly interlinked rings as shown in *Figure 4.4*. There are only two forms of such a catenane having either two positive or two negative nodes in their minimal plane projections. The easiest way to be convinced of this is to interlink two pieces of rubber tubing joined into circles with connectors. If you draw a direction arrow on each circle you can see that they can only be interlinked in two ways. Like knots, catenanes can be assigned an intrinsic linkage, Ca, equal to the sum of the intermolecular nodes (*Figure 4.4*; (12)).

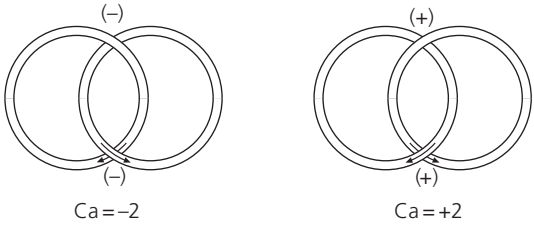


Figure 4.4 Illustration of catenanes. The diagram shows the two isomers of the singly linked catenane.

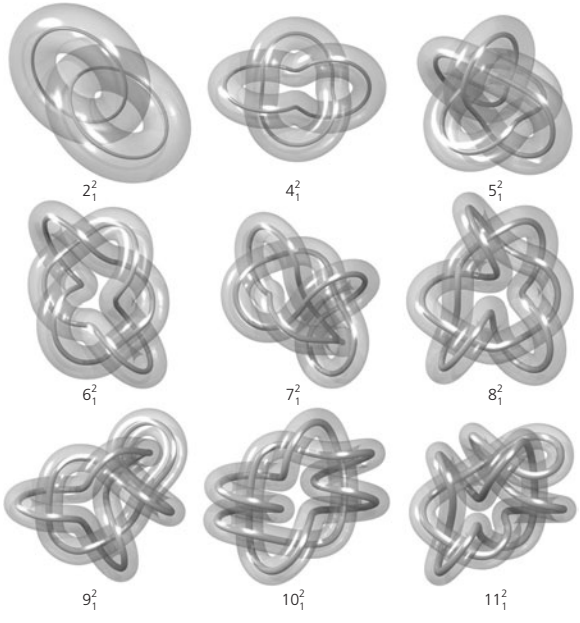


Figure 4.5 Ideal representation of catenanes. Ideal representations of various types of catenanes, labelled according to the Alexander–Briggs notation (redrawn from ref. 33).

The same complexities in the nomenclature and classification of knots (Section 4.2.2) also apply to catenanes. One method is based on the Alexander–Briggs notation for knots. In this notation, n_j^i, n is the number of intermolecular nodes, i is the number of component rings, and j distinguishes catenanes with the same number of nodes. Examples of catenanes and their nomenclature are given in Rolfsen's and Adams' books (14, 15), and some are shown in *Figure 4.5*. As with knots, the Alexander–Briggs notation does not designate the chirality of the catenane.

The ideal geometrical form for knots, described above, can also be applied to catenanes, and *Figure 4.5* shows a range of dimeric ($i = 2$) DNA catenanes represented in their ideal forms (33). The L/D ratio can be determined for these ideal forms and is found to have a linear relationship with average crossing number. For catenanes, average crossing number corresponds

to the average number of crossings seen when the axial trajectory of the ideal catenane is observed from all directions.

Catenanes can be divided into families, which include torus catenanes, where the component rings can be thought of as winding on a common torus. Torus catenanes must have an even number of nodes (cf. torus knots, with an odd number of nodes; Section 4.2.2). The singly linked catenanes shown in *Figure 4.4* are examples of torus catenanes. Many other types of catenanes are known (see *Figure 4.1*), including figure-eight catenanes (e.g. Chapter 6, *Figure 6.7*), complex multiply linked catenanes, and catenanes involving many DNA circles, such as those found in the kinetoplast DNA of trypanosomes.

4.4 Analysis of knots and catenanes

Knots and catenanes, like other topological isomers of DNA, are not necessarily easy to identify. In addition, the structural variety of knots and catenanes is very rich and a large number of possible isomers may have to be considered. The two principal methods that have been used to analyse knots and catenanes are gel electrophoresis and electron microscopy (34). In both cases it is usually desirable to nick the DNA prior to examination, so removing any supercoiling, which simplifies the analysis.

Agarose gel electrophoresis is a simple and rapid method for resolving knots and catenanes; complex mixtures can readily be separated into ladders (*Figure 4.6*). For molecules of the same size more complex forms migrate more rapidly in the gel; different forms can be identified by comparing their mobilities with markers. Experimentally it has been found, using high-resolution agarose electrophoresis, that there is a linear relationship between the speed of migration and the average crossing number of the ideal geometrical representation of knots (35), although this dependence tends to break down at high knot complexity (36). It has also been found that there is an approximately linear relationship between the electrophoretic migration of torus-type catenanes and their average crossing number (33). Simply put, there is a correlation between mobility on gels and complexity (e.g. number of nodes) for both knots and catenanes. Although the migration of DNA in agarose gels is incompletely understood, the linear relationships for knots and catenanes can be explained by the fact that the calculated sedimentation coefficients of knotted and catenated DNA molecules also show a linear correlation with average linking number (37), suggesting that electrophoretic mobility is directly related to sedimentation coefficient; for molecules with the same molecular weight a larger sedimentation coefficient means that they are more compact.

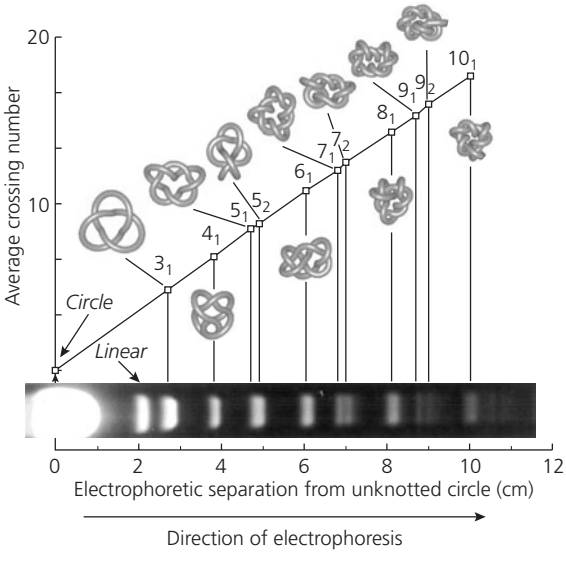


Figure 4.6 Gel electrophoresis of knotted DNA. A plot showing the linear relationship between average crossing number for various knots and their electrophoretic mobility. Knots are labelled according to the Alexander–Briggs notation (reproduced from ref. 37 with permission; copyright (1998) Elsevier).

As we have seen, coating DNA with a protein such as RecA greatly enhances the visualization of knots and catenanes by electron microscopy. The resultant thickening of the DNA by the bound protein permits more reliable identification of nodes at each crossing point (*Figure 4.1*). DNA catenanes have also been imaged using tapping mode atomic force microscopy (38). In this case naked DNA can be visualized directly without the need for protein coating.

4.5 Knots and catenanes as probes of DNA–protein interactions

As mentioned earlier, knots and catenanes in DNA can occur naturally as the products of a number of biological processes. They are also used as experimental tools to investigate the mechanisms of enzymes acting on DNA. The occurrence of knots and catenanes in DNA replication, transcription, and recombination will be discussed in Chapter 6. Here the use of knots and catenanes as probes for the analysis of these and other biological processes is briefly considered. Knots and catenanes can be generated and used as substrates in these reactions, or the knotted or catenated products from novel substrates can be used to determine mechanisms of DNA–protein transactions.

DNA replication is a fundamental process in biology and its topological consequences are considered in detail in Chapter 6. Artificial replication substrates that can generate knots have been constructed and the analysis of such knots can aid in the understanding of the replication process. For example, a plasmid has been constructed that contains two replication origins in a head-to-head orientation, that is, the replication forks will travel towards each other (39). Analysis of partially replicated molecules showed that a range of knotted species had been formed, which were analysed by electron microscopy. This analysis suggested that helical winding of replication bubbles can occur *in vivo* ((40); see Chapter 6).

The location of recognition sites by DNA-binding proteins is an important problem in molecular biology. In some systems it seems that proteins are able to locate their binding sites at rates that exceed those predicted by simple diffusion in solution (41). One mechanism that has been proposed to explain these observations is that the proteins 'track' (i.e. slide) along the DNA (42). In principle, tracking reduces the dimensionality of the search and could account for the high rates of binding. A number of tests have been devised to probe whether tracking occurs in DNA–protein interactions. One such test involves the engineering of dimeric catenanes in which each ring contains one protein-binding site for a protein that interacts with two sites on DNA. Such tests have been applied, for example, in the study of the mechanisms of restriction enzymes and transcriptional regulation.

Type II restriction enzymes are commonly used as tools in molecular biology. These enzymes are able to locate specific short, often symmetric, sequences on DNA and cleave both strands of the DNA at precise positions (43). Some type II restriction enzymes must interact with two copies of their recognition sites before DNA cleavage can occur. One such enzyme is *Sfi*I, which is a tetramer of identical subunits that interacts with two copies of its recognition sequence and cleaves both copies in a concerted reaction. In an experiment designed to determine whether the enzyme locates its second site by tracking from the first, the two recognition sequences were placed on separate rings of a singly linked dimeric catenane (*Figure 4.7a*). This was achieved by a recombination reaction using Tn*21* resolvase (see Chapter 6, Section 6.6.2.2; (44)). It was found that the catenated substrate was cleaved almost as readily as a single ring with two sites; this was in contrast to uncatenated rings with single sites that were cleaved only very slowly. This experiment rules out tracking as a mechanism for target-site location by *Sfi*I.

Type I restriction enzymes contrast with type II enzymes in a number of ways. These enzymes recognize asymmetric sequences on DNA and cleave non-specifically at sites a long distance from their recognition sites (45).

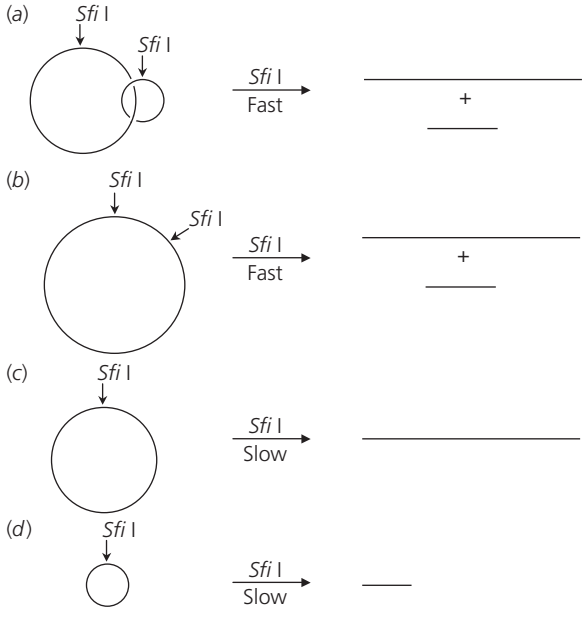


Figure 4.7 Target site location by the restriction enzyme *Sfi*I.
(*a*) A catenane in which each partner contains a single *Sfi*I restriction site. Cleavage of this substrate by the enzyme is almost as fast as the single circle with two *Sfi*I sites (*b*). Cleavage of the two uncatenated rings (c) and (*d*) is relatively slow (44).

It is thought that these enzymes track along the DNA in an ATP-dependent process to the site of cleavage, while remaining bound at their recognition site, thus generating a DNA loop. This idea has been tested for the enzyme *Eco*R124I by using dimeric catenanes (46). A plasmid containing one *Eco*R124I site is cleaved by the enzyme. When the same plasmid was converted into a catenane with the recognition site on the smaller ring, only the smaller ring was cleaved. It was concluded that, in this case, the enzyme must follow the DNA contour between the recognition and cleavage site, that is, the enzyme tracks along the DNA between the two sites.

Catenated substrates have also been used to probe the mechanism of transcriptional activation. Enhancers are DNA sequences that increase the rate of transcription of genes and can act over large distances. These elements occur in both eukaryotes and prokaryotes, and are thought to serve as protein binding sites. Models of their mechanism of action include those where the tracking or sliding of a protein along the intervening DNA mediates communication between the enhancer and the promoter. A test of such models involved the bacterial *glnA* enhancer and the nitrogen regulatory protein NtrC (47). A singly linked catenane was constructed, where one circle contained the enhancer, and the other a promoter. Transcription from the promoter was still stimulated despite the elements being on different DNA rings, although separation of the two rings by decatenation abolished stimulation. This experiment shows

that the enhancer and promoter need not be contiguous on the DNA, but do need to be in close proximity, favouring DNA looping (see Chapter 6) as a mechanism of transcriptional activation.

4.6 Conclusions

Knots and catenanes might appear to be esoteric structures and their descriptions can involve fairly complex mathematical concepts. Nevertheless, they are found in DNA in nature, and their origins may well reflect important features of the reactions of DNA-specific enzymes. Indeed, careful analyses of knots and catenanes generated by such enzymes have yielded important mechanistic information concerning these enzymes (see Chapter 6). Additionally, knots and catenanes have proved to be valuable synthetic tools for probing DNA–protein interactions.

4.7 Further Reading

Adams, C.C. (2001). *The knot book. An elementary introduction to the mathematical theory of knots.* W.H. Freeman, New York.

Dröge, P. and Cozzarelli, N.R. (1992). Topological structure of DNA knots and catenanes. *Methods Enzymol.* **212**, 120–130.

Postow, L., Peter, B.J., and Cozzarelli, N.R. (1999). Knot what we thought before: the twisted story of replication. *Bioessays* **21**, 805–808.

Rolfsen, D. (1976). *Knots and links.* Publish or Perish, Inc., Berkeley, CA.

Schvartzman, J.B. and Stasiak, A. (2004). A topological view of the replicon. *EMBO Rep.* **5**, 256–261.

Sossinsky, A. (2002). *Knots. Mathematics with a twist.* Harvard University Press, Cambridge, MA, USA.

Vologodskii, A.V. (2001). Distributions of topological states in circular DNA. *Mol. Biol. (Mosk)* **35**, 285–297.

4.8 References

1. Liu, L.F., Depew, R.E., and Wang, J.C. (1976). Knotted single-stranded DNA rings: a novel topological isomer of circular single-stranded DNA formed by treatment with *Escherichia coli* ω protein. *J. Mol. Biol.* **106**, 439–452.
2. Bucka, A. and Stasiak, A. (2002). Construction and electrophoretic migration of single-stranded DNA knots and catenanes. *Nucleic Acids Res.* **30**, e24.
3. Liu, L.F., Perkocha, L., Calendar, R., and Wang, J.C. (1981). Knotted DNA from bacteriophage capsids. *Proc. Natl. Acad. Sci. USA* **78**, 5498–5502.

4. Liu, L.F. and Davis, J.L. (1981). Novel topologically knotted DNA from bacteriophage P4 capsids: studies with topoisomerases. *Nucleic Acids Res.* **9**, 3979–3989.
5. Arsuaga, J., Vazquez, M., Trigueros, S., Sumners de, W., and Roca, J. (2002). Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proc. Natl. Acad. Sci. USA* **99**, 5373–5377.
6. Shaw, S.Y. and Wang, J.C. (1993). Knotted DNA rings: probability of formation and resolution of the two chiral trefoils. *Science* **260**, 533–536.
7. Rybenkov, V.V., Cozzarelli, N.R., and Vologodskii, A.V. (1993). Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. USA* **90**, 5307–5311.
8. Bednar, J., Furrer, P., Stasiak, A., Dubochet, J., Egelman, E.H., and Bates, A.D. (1994). The twist, writhe and overall shape of supercoiled DNA change during counterion-induced transition from a loosely to a tightly interwound superhelix. Possible implications for DNA structure *in vivo. J. Mol. Biol.* **235**, 825–847.
9. Brown, P.O. and Cozzarelli, N.R. (1981). Catenation and knotting of duplex DNA by type I topoisomerases: a mechanistic parallel with type II topoisomerases. *Proc. Natl. Acad. Sci. USA* **78**, 843–847.
10. Liu, L.F., Liu, C.-C., and Alberts, B.M. (1980). Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule *via* a reversible double-strand break. *Cell* **19**, 697–707.
11. Krasnow, M.A., Stasiak, A., Spengler, S.J., Dean, F., Koller, T., and Cozzarelli, N.R. (1983). Determination of the absolute handedness of knots and catenanes of DNA. *Nature* **304**, 559–560.
12. Cozzarelli, N.R., Krasnow, M.A., Gerrard, S.P., and White, J.H. (1984). A topological treatment of recombination and topoisomerases. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 383–400.
13. White, J.H., Millet, K.C., and Cozzarelli, N.R. (1987). Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination. *J. Mol. Biol.* **197**, 585–603.
14. Rolfsen, D. (1976). *Knots and links.* Publish or Perish, Inc., Berkeley, CA.
15. Adams, C.C. (2001). *The knot book. An elementary introduction to the mathematical theory of knots.* W.H. Freeman, New York.
16. Katritch, V., Bednar, J., Michoud, D., Scharein, R.G., Dubochet, J., and Stasiak, A. (1996). Geometry and physics of knots. *Nature* **384**, 142–145.
17. Katritch, V., Olson, W.K., Pieranski, P., Dubochet, J., and Stasiak, A. (1997). Properties of ideal composite knots. *Nature* **388**, 148–151.
18. Hudson, B. and Vinograd, J. (1967). Catenated circular DNA molecules in HeLa cell mitochondria. *Nature* **216**, 647–652.
19. Clayton, D.A. and Vinograd, J. (1967). Circular dimer and catenate forms of mitochondrial DNA in human leukaemic leucocytes. *Nature* **216**, 652–657.
20. Englund, P.T., Hajduk, S.L., and Marini, J.C. (1982). The molecular biology of trypanosomes. *Annu. Rev. Biochem.* **51**, 695–726.

21. Marini, J.C., Miller, K.G., and Englund, P.T. (1980). Decatenation of kinetoplast DNA by topoisomerases. *J. Biol. Chem.* **255**, 4976–4979.
22. Chen, J., Englund, P.T., and Cozzarelli, N.R. (1995). Changes in network topology during the replication of kinetoplast DNA. *EMBO J.* **14**, 6339–6347.
23. Wang, J.C. and Schwartz, H. (1967). Noncomplementarity in base sequences between the cohesive ends of coliphages 186 and λ and the formation of interlocked rings between the two DNAs. *Biopolymers* **5**, 953–966.
24. Martin, K. and Wang, J.C. (1970). Electron microscopic studies of topologically interlocked λb2b5c DNA rings. *Biopolymers* **9**, 503–505.
25. Kupersztoch, Y.M. and Helsinki, D.R. (1973). A catenated DNA molecule as an intermediate in the replication of the resistance transfer factor RK6 in *Escherichia coli. Biochem. Biophys. Res. Commun.* **54**, 1451–1459.
26. Novick, R.P., Smith, K., Sheehy, R.J., and Murphy, E. (1973). A catenated intermediate in plasmid replication. *Biochem. Biophys. Res. Commun.* **54**, 1460–1469.
27. Sakakibara, Y., Suzuki, K., and Tomizawa, J.-I. (1976). Formation of catenated molecules by replication of colicin E1 plasmid DNA in cell extracts. *J. Mol. Biol.* **108**, 569–582.
28. Jaenisch, R. and Levine, A.J. (1973). DNA replication of SV40-infected cells VII. Formation of SV40 catenated and circular dimers. *J. Mol. Biol.* **73**, 199–212.
29. Mizuuchi, K., Gellert, M., Weisberg, R.A., and Nash, H.A. (1980). Catenation and supercoiling in the products of bacteriophage λ integrative recombination *in vitro. J. Mol. Biol.* **141**, 485–494.
30. Reed, R.R. (1981). Transposon-mediated site-specific recombination: a defined *in vitro* system. *Cell* **25**, 713–719.
31. Tse, Y.-C. and Wang, J.C. (1980). *E. coli* and *M. luteus* DNA topoisomerase I can catalyse catenation or decatenation of double-stranded DNA rings. *Cell* **22**, 269–276.
32. Kreuzer, K.N. and Cozzarelli, N.R. (1980). Formation and resolution of DNA catenanes by DNA gyrase. *Cell* **20**, 245–254.
33. Laurie, B., Katritch, V., Sogo, J., Koller, T., Dubochet, J., and Stasiak, A. (1998). Geometry and physics of catenanes applied to the study of DNA replication. *Biophys. J.* **74**, 2815–2822.
34. Dröge, P. and Cozzarelli, N.R. (1992). Topological structure of DNA knots and catenanes. *Methods Enzymol.* **212**, 120–130.
35. Stasiak, A., Katritch, V., Bednar, J., Michoud, D., and Dubochet, J. (1996). Electrophoretic mobility of DNA knots. *Nature* **384**, 122.
36. Stark, W.M. and Boocock, M.R. (1994). The linkage change of a knotting reaction catalysed by Tn3 resolvase. *J. Mol. Biol.* **239**, 25–36.
37. Vologodskii, A.V., Crisona, N.J., Laurie, B., Pieranski, P., Katritch, V., Dubochet, J., and Stasiak, A. (1998). Sedimentation and electrophoretic migration of DNA knots and catenanes. *J. Mol. Biol.* **278**, 1–3.

38. Yamaguchi, H., Kubota, K., and Harada, A. (2000). Direct observation of DNA catenanes by atomic force microscopy. *Chem. Lett.* **29**, 384–385.
39. Sogo, J.M., Stasiak, A., Martinez-Robles, M.L., Krimer, D.B., Hernandez, P., and Schvartzman, J.B. (1999). Formation of knots in partially replicated DNA molecules. *J. Mol. Biol.* **286**, 637–643.
40. Schvartzman, J.B. and Stasiak, A. (2004). A topological view of the replicon. *EMBO Rep.* **5**, 256–261.
41. Berg, O.G., Winter, R.B., and von Hippel, P.H. (1982). How do genome-regulatory proteins locate their DNA target sites? *Trends Biochem. Sci.* **7**, 1513–1523.
42. Wang, J.C. and Giaever, G.N. (1988). Action at a distance along a DNA. *Science* **240**, 300–304.
43. Pingoud, A. and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727.
44. Szczelkun, M.D. and Halford, S.E. (1996). Recombination by resolvase to analyse DNA communications by the *Sfi*I restriction endonuclease. *EMBO J.* **15**, 1460–1469.
45. Murray, N.E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**, 412–434.
46. Szczelkun, M.D., Dillingham, M.S., Janscak, P., Firman, K., and Halford, S.E. (1996). Repercussions of DNA tracking by the type IC restriction endonuclease *Eco* R124I on linear, circular and catenated substrates. *EMBO J.* **15**, 6335–6347.
47. Wedel, A., Weiss, D.S., Popham, D., Dröge, P., and Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science* **248**, 486–490.
48. Shishido, K., Komiyama, N., and Ikawa, S. (1987). Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants. *J. Mol. Biol.* **195**, 215–218.
49. Spengler, S.J., Stasiak, A., and Cozzarelli, N.R. (1985). The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure. *Cell* **42**, 325–334.

5
DNA topoisomerases

5.1 Introduction

Virtually every reaction that occurs in biological systems is catalysed by an enzyme. The interconversions of the different topological forms of DNA are no exceptions. The enzymes that catalyse these processes are known as DNA topoisomerases and constitute a widespread and fascinating group.

The first DNA topoisomerase was discovered by James Wang in 1971, the so-called *ω* protein from *Escherichia coli* (1). This enzyme, now called DNA topoisomerase I, was found to reduce the number of negative supercoils in bacteriophage λ DNA, as measured by changes in sedimentation coefficient (see Chapter 2, Section 2.2.2). This finding was followed closely by the discovery of a 'nicking-closing' activity from nuclear extracts of mouse cells by James Champoux (2), which could remove supercoils from closed-circular polyoma virus DNA. Although it is also referred to as DNA topoisomerase I, this enzyme turns out to have a mechanism of action quite distinct from the prokaryotic enzyme found by Wang. These two enzymes are now classified in different groups: Type IA for the prokaryotic enzyme and type IB for the eukaryotic enzyme (see Section 5.2). An enzyme capable of introducing supercoils, DNA gyrase, was discovered a few years later by Martin Gellert and co-workers, during a search for host co-factors that support site-specific recombination by bacteriophage λ (3). They showed that the λ DNA substrate had to be supercoiled in order to support the recombination reaction. This substrate could be replaced by relaxed closed-circular DNA only if it was incubated with an *E. coli* cell fraction and ATP. An activity was purified from *E. coli* extracts that could supercoil DNA in the presence of ATP. The rather non-systematic but euphonic name, DNA gyrase, has stuck, although the enzyme is also referred to as DNA topoisomerase II, and its eukaryotic homologues are known by this name.

Every cell type so far examined contains DNA topoisomerases, and where a genetic test has been possible, at least one is essential for cell growth. Examples of the type of organisms in which topoisomerases have been studied include the bacteria *E. coli* and *Staphylococcus aureus*, yeasts, the model plant *Arabidopsis*, *Drosophila*, and man. In addition, several viruses are known to encode a topoisomerase, for example, bacteriophage T4 and the animal virus vaccinia. Examples of DNA topoisomerases are given in *Table 5.1*. Perhaps the most important aspect of topoisomerases for those interested in DNA topology is the mechanism of topoisomerase action: How do these enzymes achieve the complex interconversions of DNA supercoils, knots, and catenanes? The current understanding of how these enzymes work is summarized in Section 5.3.

Table 5.1 DNA topoisomerases

Enzyme	Type	Source	Subunit size (kDa) and composition	Remarks
Bacterial topoisomerase I (ω protein)	IA	Bacteria (e.g. *E. coli*)	97 Monomer	Cannot relax positive supercoils
Eukaryotic topoisomerase I	IB	Eukaryotes (e.g. human)	91 Monomer	Can relax both positive and negative supercoils
Vaccinia virus topoisomerase I	IB	Vaccinia virus	37 Monomer	ATP stimulates topoisomerase activity
Topoisomerase III[a]	IA	Bacteria (e.g. *E. coli*)	73 Monomer	Potent decatenating activity
Reverse gyrase	IA	Thermophilic Archaea (e.g. *Sulfolobus acidocaldarius*)	143 Monomer	Can introduce positive supercoils into DNA (ATP-dependent)
DNA gyrase	IIA	Bacteria (e.g. *E. coli*)	97 and 90 A_2B_2	Can introduce negative supercoils into DNA (ATP-dependent)
T4 topoisomerase	IIA	Bacteriophage T4	58, 51, and 18 2 copies of each subunit	Can relax, but not supercoil, DNA (ATP-dependent)
Eukaryotic topoisomerase II	IIA	Eukaryotes (e.g. human topoisomerase IIα)	174 Homodimer	Can relax, but not supercoil, DNA (ATP-dependent)
Topoisomerase IV[a]	IIA	Bacteria (e.g. *E. coli*)	84 and 70 C_2E_2	Can relax, but not supercoil, DNA, potent decatenase (ATP-dependent)
Topoisomerase VI	IIB	Archaea (e.g. *Sulfolobus shibatae*)	45 and 60 A_2B_2	Can relax, but not supercoil, DNA (ATP-dependent)

[a] Note that topoisomerases III and IV do not represent 'types' of topoisomerase mechanism (cf. type I and type II).

5.2 Reactions of topoisomerases

A reaction common to all topoisomerases is the ability to relax negatively supercoiled DNA; that is to convert it into a less-supercoiled form, or increase its linking number (Lk). It should be clear from the discussion in Chapter 2 that this conversion must involve the transient breaking and rejoining of DNA strands. A number of other reactions are also known and these are illustrated in *Figures 5.1 and 5.2*. During mechanistic studies it became clear that topoisomerases could be divided into two main types. Type I enzymes are able to carry out reactions involving the breaking of only one strand of the DNA (*Figure 5.1*), while type II enzymes can carry out those involving the breaking of both strands (*Figure 5.2*). Perhaps the best illustration of this distinction is the catenation and decatenation of double-stranded DNA circles. These interconversions can be achieved by type II topoisomerases but not by type I enzymes, unless one of the circles already contains a break in one strand (4). More recently topoisomerases have been further subdivided into type IA and B, and type IIA and B.

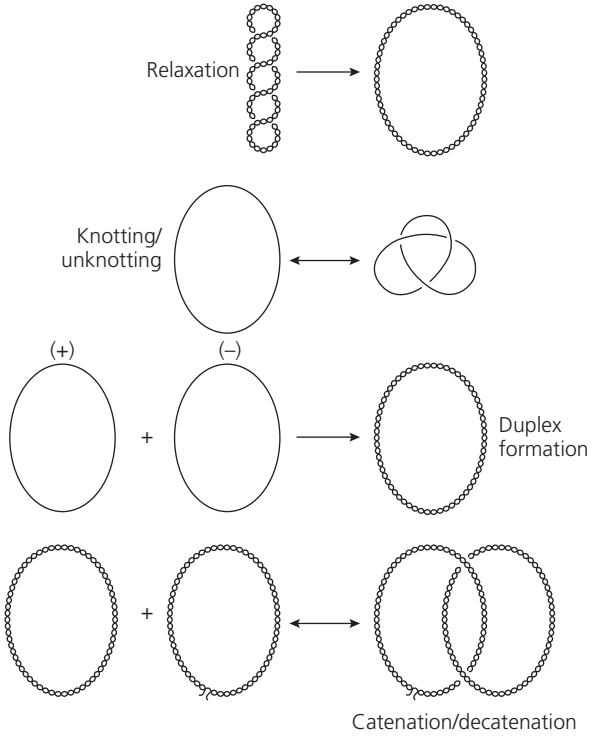


Figure 5.1 Reactions of type I topoisomerases (redrawn from ref. 48).

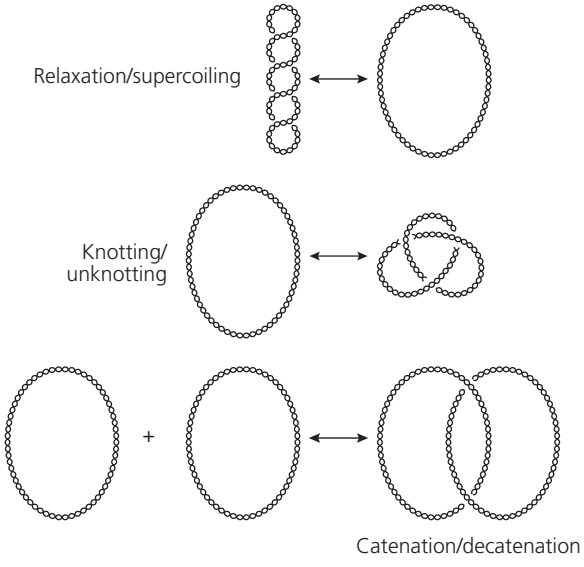


Figure 5.2 Reactions of type II topoisomerases (redrawn from ref. 48).

Not all topoisomerases can carry out the full range of reactions shown in *Figures 5.1 and 5.2*. For example, the type I topoisomerase from *E. coli* can relax only negatively supercoiled DNA, while the type I enzyme from calf thymus will relax both negatively and positively supercoiled DNA (1, 2). This turns out to be a consequence of the inability of *E. coli* topoisomerase I to bind to positively supercoiled DNA. Indeed, the *E. coli* enzyme can relax a positively supercoiled DNA molecule if it is constructed to contain a single-stranded loop (5), suggesting that it preferentially binds single-stranded DNA. A further example concerns the ability of topoisomerases to introduce supercoils into DNA. Only one type II enzyme, DNA gyrase, is able to do this. Gyrase is found in bacteria (and more recently plants (50)) and can introduce negative supercoils into DNA using the free energy from ATP hydrolysis (3). Thermophilic archaea and eubacteria, such as *Sulfolobus* and *Thermotogales* contain an enzyme called reverse gyrase that can introduce positive supercoils into DNA (6); reverse gyrase is actually a type I enzyme and is not a homologue of DNA gyrase (*Table 5.1*). Type II enzymes from archaea, bacteriophage, and most eukaryotes are unable to catalyse DNA supercoiling.

In addition to topoisomerase I and DNA gyrase (topoisomerase II), *E. coli* has two further topoisomerases (*Table 5.1*). Topoisomerase III is another type IA enzyme that is more active as a decatenating enzyme than as a DNA-relaxing enzyme (7). Topoisomerase IV is a type II enzyme with a high degree of sequence similarity to DNA gyrase, which can relax, but

not supercoil, DNA and again shows efficient decatenation activity (8, 9). Topoisomerase V, originally identified in *Methanopyrus kandleri* is the only known example of a type IB enzyme in bacteria. This organism possesses a number of other topoisomerases, including a novel heterodimeric reverse gyrase. Multiple topoisomerases are also found in other organisms, including humans (*Table 5.1* (10)). Mammalian cells possess two type II enzymes, topoisomerase IIα and topoisomerase IIβ; these isoforms differ in their patterns of expression with the α-isoform being expressed preferentially in proliferating cells. Archaea (and plants) possess a type II enzyme called topo VI (11), which appears to be distinct from the other type II enzymes, and its discovery has caused these enzymes to be subdivided into types IIA and IIB (*Table 5.1*). The enzyme from *Sulfolobus* consists of two subunits (A and B) one of which (A) has similarity to Spo11, a yeast protein involved in meiotic recombination (11). Sequence analysis of the *Arabidopsis thaliana* genome suggests that topoisomerase VI occurs in plants as well. Experimental evidence supports the idea that this enzyme is important in plant growth and development, possibly in the process of endoreduplication, in which the DNA in plant cells is duplicated without going through mitosis (12–14).

There are now a large number of enzymes that have been shown to perform reactions of the type shown in *Figures 5.1 and 5.2.* Some of these were not originally identified as topoisomerases but as enzymes involved in recombination reactions. Examples include resolvase proteins involved in the process of transposition and the Int protein involved in bacteriophage λ integration (see Chapter 6, Section 6.6). These proteins carry out reactions involving the breaking of DNA and the transfer of the broken end to another DNA molecule. As we will see, this is mechanistically very similar to a topoisomerase reaction except that topoisomerases rejoin the broken end to its original partner.

5.3 Structures and mechanisms of topoisomerases

If we consider the DNA relaxation reaction in the light of our discussions about linking number changes in Chapter 2 (see Sections 2.2.1, 2.3.1), it seems logical that topoisomerases should work by a swivel mechanism. This might involve breaking one (or both) strands of the DNA, allowing the free end (or ends) to rotate about the helix axis, and resealing the break. This would alter the linking number of the DNA as required by the relaxation reaction. However, if we consider the knotting/unknotting and catenation/decatenation

reactions (*Figures 5.1 and 5.2*), it should be clear that a swivel mechanism cannot account for the full range of topoisomerase actions. In fact a different type of mechanism, called strand passage, can account for the ability of topoisomerases to catalyse all these interconversions. In its most general form, strand passage involves the cleavage of one or both strands of the DNA by the enzyme and the passing of a single- or double-stranded segment of DNA through the break, which is then resealed. Topoisomerases stabilize the DNA break by forming a covalent bond between the enzyme (via a tyrosine hydroxyl group) and the phosphate at the break site. The strand passage event can involve segments of DNA from the same DNA molecule, in the case of relaxation/supercoiling and knotting/unknotting, or from separate DNA molecules, in the case of catenation/decatenation. The details of the strand-passage mechanism differ from one enzyme to another. The reactions of type I enzymes proceed through single-strand breaks in DNA and involve covalent attachment to either the 5′-phosphate (type IA) or the 3′-phosphate (type IB) at the break site. The reactions of type II enzymes proceed via double-strand breaks and all involve covalent attachment at the 5′-phosphate. Intramolecular topoisomerase reactions (e.g. relaxation/supercoiling) that occur via passage of a single strand of DNA through a transient single-strand break change the linking number in increments of 1; those that occur via passage of a double strand of DNA through a transient double-strand break change Lk in increments of 2 (15). This turns out to be an effective diagnostic test for distinguishing type I and type II topoisomerases.

Perhaps the most important advance in relation to understanding topoisomerase mechanism has been the application of x-ray crystallography to determine the structures of some of these enzymes. In only one case has the structure of an intact topoisomerase been solved, but there are also several examples of protein fragments whose structures have given us key insights into the likely mechanisms of action of the intact enzymes. Many of these fragments correspond to protein domains, and alignment of topoisomerases from the same group shows how these domains are conserved (*Figure 5.3*). This also demonstrates the evolutionary relationships among these enzymes.

Structures of some DNA topoisomerases are shown in *Figure 5.4*, and examples of the proposed mechanisms for certain enzymes are shown in *Figures 5.5–5.8.* Most of these structures have been solved in the absence of DNA and only models of the protein–DNA complexes are available in these cases. The only exception is human DNA topoisomerase I where structures of covalent and non-covalent complexes with DNA have been solved.

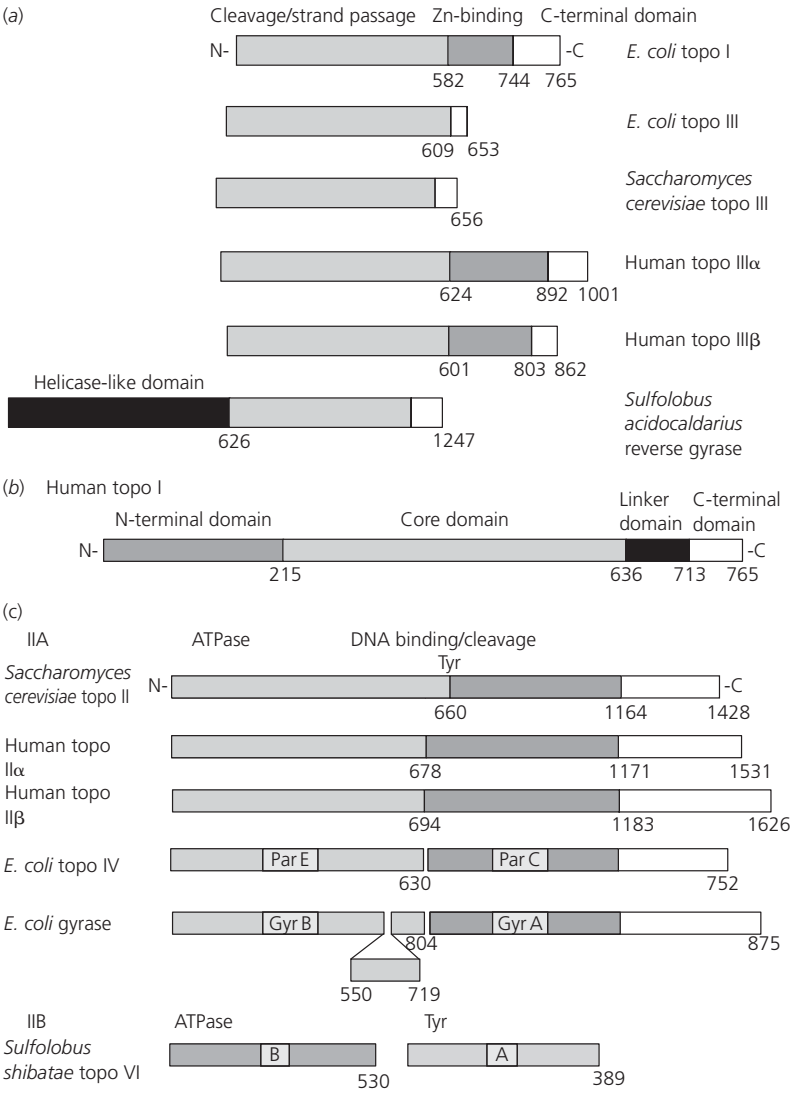


Figure 5.3 Domain alignments of type I and type II topoisomerases. (*a*) Domain alignments of type IA topoisomerases. (*b*) Domain alignments of type IB topoisomerases. (c) Domain alignments of type II topoisomerases. (Redrawn from ref. 49 with permission; copyright (2001) Annual Reviews www.annualreviews.org).

5.3.1 *Type I topoisomerases*

The first type I topoisomerase structure to be solved was that of a 67 kDa N-terminal fragment of *E. coli* DNA topoisomerase I (16), an example of a type

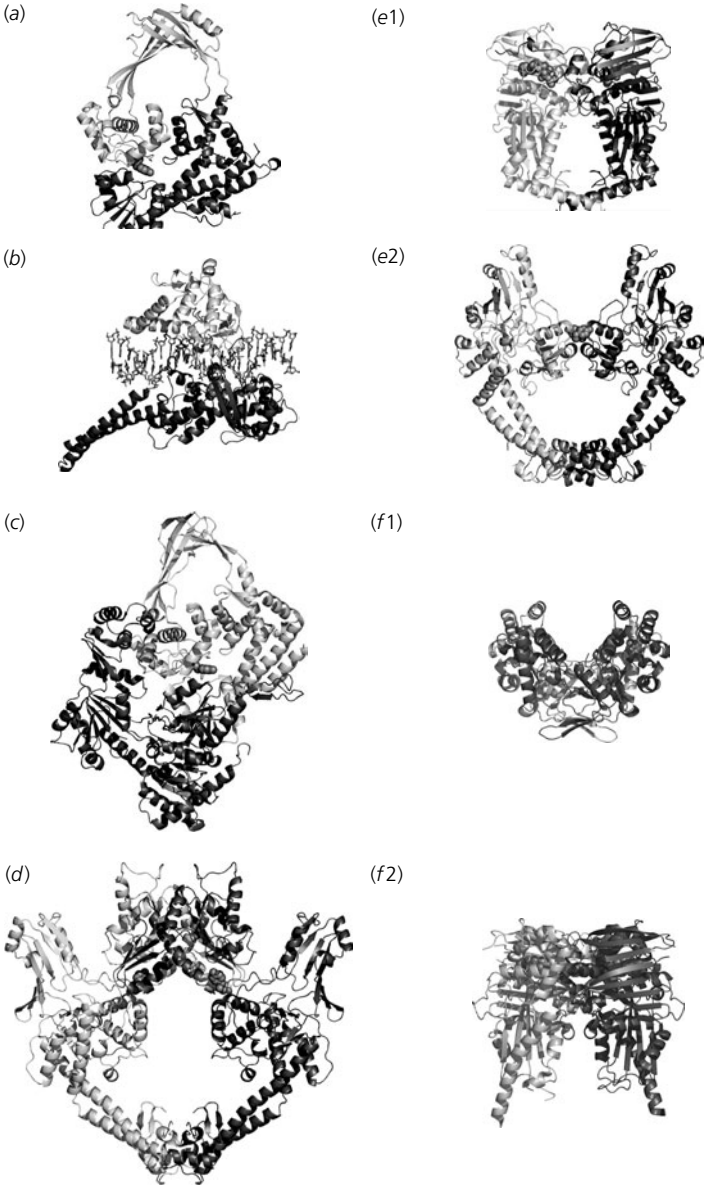


Figure 5.4 Structures of DNA topoisomerases. Crystal structures of selected topoisomerase fragments are shown in ribbon representation. (*a*) N-terminal 67 kDa fragment of *E. coli* topoisomerase I (16); (*b*) C-terminal 70 kDa fragment of human topoisomerase I complexed with DNA (49); (*c*) reverse gyrase (21); (*d*) 92 kDa fragment of yeast topoisomerase II (22); (*e1*) N-terminal 43 kDa fragment of the DNA gyrase B protein (26); (*e2*) N-terminal 59 kDa fragment of the DNA gyrase A protein (23); (*f1*) Topo VI A subunit (30); (*f2*) Topo VI B subunit (33). The images were generated using PyMol (51) using crystallographic co-ordinates acquired from the PDB (52).

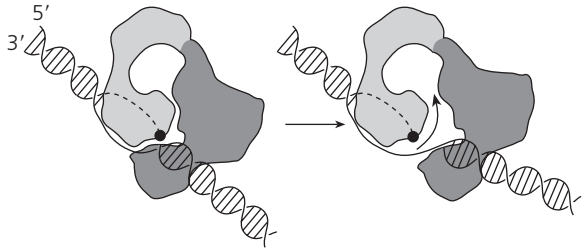


Figure 5.5 A proposed mechanism for *E. coli* topoisomerase I. The enzyme binds DNA and cleaves one strand forming a 5′-phosphotyrosine linkage (black circle). The complementary strand is passed through the gap and into the central cavity of the enzyme. Resealing of the nick and release of the passed strand changes the linking number by ±1, resulting in relaxation of the DNA. If the initial cleavage is opposite a nick or gap in a duplex circle, then passage of another duplex segment could lead to a catenation or knotting reaction.

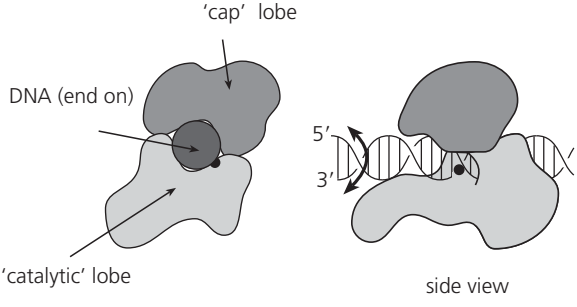


Figure 5.6 A proposed mechanism for human topoisomerase I. The enzyme binds duplex DNA and cleaves one strand forming a 3′-phosphotyrosine linkage (black circle). The free 5′-OH can then rotate (controlled rotation) before the break is resealed, resulting in DNA relaxation.

IA enzyme (*Table 5.1*). An important feature of the structure of this fragment (*Figure 5.4*) is the presence of a central cavity, large enough to accommodate a segment of double-stranded DNA. The active-site tyrosine is at the interface between two domains at the opening of this cavity, and this suggests a mechanism in which DNA (either single- or double-stranded) can enter the interior of the enzyme through the DNA break resulting from cleavage of a single-stranded segment (*Figure 5.5*). Cavities such as the one present in the *E. coli* DNA topoisomerase I turn out to be a common feature of topoisomerase structures and provide strong support for the concept of strand passage in topoisomerase mechanisms.

Although the idea of strand passage is now firmly established, it seems quite likely that one group of topoisomerases, the type IB enzymes, operate by a swivel mechanism, the 'logical' mechanism we mentioned earlier. One of

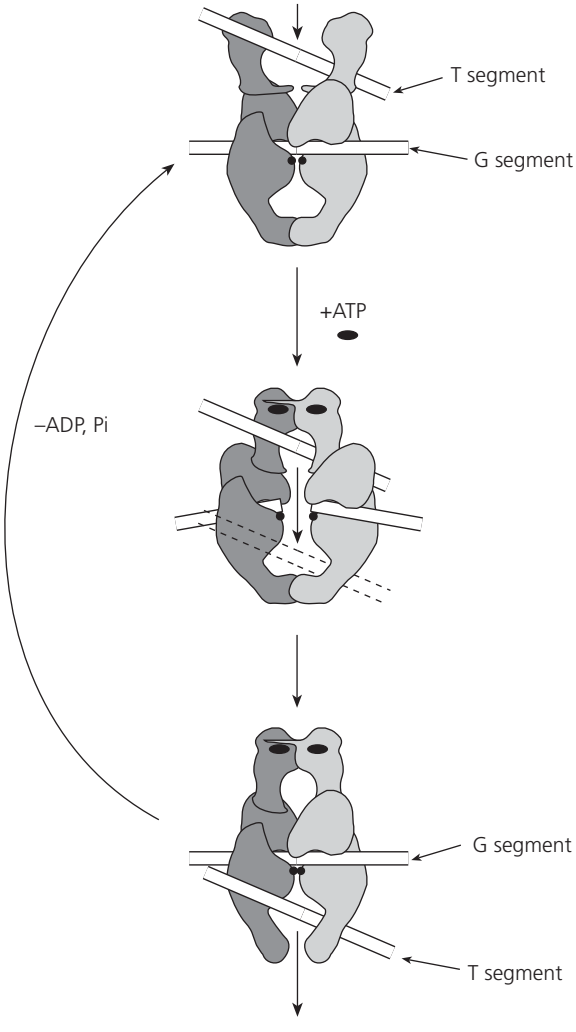


Figure 5.7 A proposed mechanism for yeast topoisomerase II. A segment of DNA (the G segment) is bound to the enzyme in the vicinity of the active-site tyrosines (black circles). Cleavage of the G segment (to yield 5′-phosphotyrosine linkages) allows passage of a second DNA segment (the T segment) through the G segment. The T segment is captured by ATP-dependent dimerization of the N-terminal (ATPase) domains of the enzyme, which form a 'clamp'. The T segment exits from the bottom of the enzyme as shown. ATP hydrolysis releases the clamp and allows the enzyme to reset.

the members of this group for which we have structural information is human topoisomerase I (*Figure 5.4*). The structure of a 70 kDa fragment of the human enzyme bound to DNA shows no tell-tale cavity appropriate to a strand-passage step; instead it surrounds the DNA double helix and, following cleavage and covalent attachment of the 3′-phosphate to the protein, the 5′-hydroxyl end may be able to swivel (17, 18). This has led to a mechanistic model for type IB enzymes termed 'controlled rotation' (*Figure 5.6*). Experimental evidence for this type of mechanism has also been obtained from work on another member of this group, vaccinia virus topoisomerase I (19).

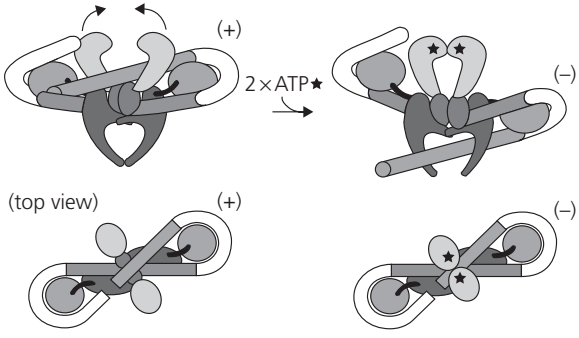


Figure 5.8 A proposed mechanism for *E. coli* DNA gyrase. Gyrase operates by a similar mechanism to topoisomerase II (*Figure 5.7*) except that the G segment is wrapped around the enzyme (as illustrated in the figure) so as to present the T segment to the DNA gate. This wrap strongly favours intramolecular strand passage and leads to the enzyme's ability to introduce negative supercoils into DNA. The asterix represents ATP. (Reproduced from ref. 53 with permission; copyright (2004) National Academy of Sciences, USA.)

Reverse gyrase, which is a type IA enzyme, may act via a mechanism that is distinct from those of the more conventional (DNA-relaxing) type I enzymes. The enzyme contains a helicase-like domain and a type IA topoisomerase domain in the same polypeptide (20). The crystal structure of the 120 kDa reverse gyrase from *Archaeoglobus fulgidus* has been solved with and without a bound ATP analogue (21). The enzyme has an N-terminal domain resembling helicase structures and a C-terminal domain equivalent to *E. coli* topoisomerase I (*Figure 5.4*). The structures suggest a strand-passage mechanism, similar to *E. coli* topoisomerase I (*Figure 5.5*), except that in the case of reverse gyrase it is adapted to give only a positively supercoiled product (21).

5.3.2 *Type II topoisomerases*

The crystal structure of a large (92 kDa) internal fragment of yeast DNA topoisomerase II has been solved (22). This structure again reveals a large cavity that could accommodate a double-stranded DNA (*Figure 5.4*). This structure is complemented by that of a 59 kDa N-terminal fragment of the *E. coli* DNA gyrase A protein, which is homologous to yeast topoisomerase II (23). These two structures show a high degree of similarity in their secondary and tertiary folds (*Figure 5.4*), but are significantly different in terms of their quaternary structures. It is thought that they represent two conformational states that occur in the reaction cycle of the type II topoisomerases. Mechanistic studies of yeast topoisomerase II have defined two segments of DNA that are bound to

the enzymes: a G (or gate) segment, which is cleaved, and a T (or transport) segment, which is passed through the break in the G segment (24). Coupled with the structural information on yeast topoisomerase II and additional structural information on DNA gyrase (see below), these mechanistic studies suggest a two-gate model for topoisomerase II (*Figure 5.7*), in which a T segment is captured by the enzyme and passes through two protein gates: one associated with the cleaved G segment, and one controlling the exit of the T segment from the protein. An ATP-operated clamp, comprising the N-terminal domain of the enzyme dimer, effects the initial capture of the T segment. The structure of this domain was first determined for the corresponding region of bacterial gyrase (see below) and has also been solved for yeast topoisomerase II (25). It is proposed that this domain dimerizes in the presence of ATP capturing the T segment so that it can be passed through the cleaved G segment (*Figure 5.7*).

Bacterial DNA gyrase differs from its eukaryotic counterpart in that it can actively introduce negative supercoils into DNA. Apart from the structure of an N-terminal fragment of the gyrase A protein mentioned above, the structure of an N-terminal fragment of the *E. coli* B protein is also known (*Figure 5.4*). The structure of this 43 kDa fragment was solved with a bound ATP analogue (5′-adenylyl-β,γ-imidodiphosphate) (26). As with the other type II topoisomerase fragments, this structure is dimeric and reveals a cavity large enough to accommodate DNA. Mechanistically, gyrase is thought to operate much like topoisomerase II, in that a captured T segment is passed though a cleaved G segment. Gyrase's unique ability to catalyse DNA supercoiling is thought to derive from the wrapping of a segment of DNA (which contains the G segment) around the enzyme in such a way as to present a T segment to the ATP-operated clamp (27) (*Figure 5.8*). The wrapping function resides in the C-terminal domain of the gyrase A protein, which shares little sequence similarity with the corresponding region of topoisomerase II. This ability to wrap DNA is reminiscent of other systems in which DNA is wrapped around proteins, such as the nucleosome (Chapter 3, Section 3.5) and RNA polymerase (Chapter 6, Section 6.4.1). The ATP-operated clamp is a common feature of all type II topoisomerases and appears to function to pass the T segment through the enzyme-stabilized break in the G segment (24). The structure of the N-terminal domain of GyrB indicates homology to a diverse range of other proteins, which have become known as the GHKL superfamily (28). These proteins include MutL (a mismatch repair protein), Hsp90 (a chaperone protein), histidine kinases, and the B subunit of topoisomerase VI (see below). The GHKL proteins share four conserved motifs involved in nucleotide binding and hydrolysis.

DNA supercoiling is an energetically unfavourable process and gyrase is somehow able to transform the chemical energy derived from the hydrolysis

of a phosphodiester bond in ATP into the torsional stress of supercoiling. Although it must be assumed that this process involves conformational changes in the protein, little is known of the detailed mechanism of this energy-coupling process. It is interesting to note that other type II topoisomerases hydrolyse ATP but are able only to relax, not supercoil, DNA (*Table 5.1*). Gyrase can also relax supercoiled DNA, but this reaction does not require ATP. The requirement for ATP hydrolysis for the reactions of topoisomerase II, most of which are apparently energetically favourable, has been something of a puzzle. Experiments have suggested that it is likely that this energy requirement enables topoisomerase II to catalyse reactions in which the products are a non-equilibrium distribution of topoisomers (29). For example, *E. coli* topoisomerase IV can relax supercoiled DNA to yield a distribution of relaxed topoisomers that is narrower than that obtained by relaxation by topoisomerase I under the same conditions (see Chapter 2, Section 2.6.2). This implies that type II enzymes can use the energy of ATP hydrolysis to shift the distribution of the relaxed topoisomer products away from equilibrium. In the case of gyrase this results in products that have negative supercoiling (σ; see Chapter 2, Section 2.2.2); in the case of topoisomerase II, the energy of ATP hydrolysis is serving to relax the DNA more than at equilibrium. A further illustration of this phenomenon occurs with catenated DNA. Rybenkov *et al.* (29) used a system that allows equilibration of catenanes with uncatenated DNA to show that topoisomerase IV, and other type II enzymes, are able to generate products with a much lower proportion of catenanes than those formed at equilibrium in the absence of enzyme, that is, the enzymes strongly favour the decatenation reaction.

The structure of a fragment of the A subunit of topoisomerase VI from *Methanococcus jannaschii* (*Figure 5.4*) suggests that the type IIB enzymes are distinct from the IIA enzymes (30). However, comparison of the structures of the type IA, IIA, and IIB enzymes has shown that there is structural similarity between these enzymes, specifically in a region containing a so-called Rossmann-like fold that is thought to co-ordinate metal ions (31). Sequence alignments have suggested that this homology extends to DNA primases and other enzymes and this structural feature has been termed the 'toprim' fold (32). In addition, the structure of a truncated form of the topoisomerase VIB subunit from *Sulfolobus shibatae* has been solved, both with and without an ATP analogue bound (33). Although the only region of topoisomerase VIB with a sequence apparently related to type IIA enzymes is a GHKL-type motif (see earlier), this structure shows impressive similarity to the N-terminal domain of GyrB (*Figure 5.4*). The structure reveals similarities with parts of the GyrB fragment involved in nucleotide hydrolysis and signalling of hydrolysis to other parts of the protein. This work reinforces the idea that type IIA

and IIB enzymes are evolutionarily related and are likely to share a similar mechanism of strand passage, and nucleotide binding and hydrolysis.

5.4 Topoisomerases as drug targets

In the opening section of this chapter we noted that topoisomerases are essential for cell growth. As a consequence, these enzymes are potential targets for cytotoxic drugs (34–38). For example, DNA gyrase and its close relative topoisomerase IV are targets for two groups of antibacterial compounds: the quinolones and the coumarins (*Table 5.2*). In the context of topoisomerase mechanisms it is interesting to consider how these drugs might work. The quinolones (e.g. nalidixic acid and ciprofloxacin) are thought to act by interrupting the DNA breakage–reunion step of the gyrase reaction (39). Indeed, under certain conditions, quinolones can lead to trapping of the gyrase–DNA intermediate in which the A subunits of the enzyme are covalently bound to the 5′-phosphates of DNA at the break site. The coumarin drugs (e.g. novobiocin and coumermycin A_1) act in an entirely different way by inhibiting the hydrolysis of ATP by gyrase and thus preventing DNA supercoiling. Although coumarins are structurally dissimilar to ATP, crystal structure information on complexes between gyrase and coumarins show that the drugs bind at a site that overlaps the ATP-binding site on the enzyme, and are therefore competitive inhibitors (40, 41).

The eukaryotic topoisomerases are also drug targets (*Table 5.2*). The antitumour drug camptothecin, whose derivatives are widely used in cancer chemotherapy, acts on eukaryotic topoisomerase I. The drug stabilizes the complex between the enzyme and DNA in a manner analogous to the action of quinolones on DNA gyrase. Crystal structure information on human topoisomerase I suggests a model in which the drug interacts with the bound

Table 5.2 Inhibitors of DNA topoisomerases

Inhibitor	Target enzyme	Therapeutic value
Quinolones (e.g. ciprofloxacin)	DNA gyrase and topoisomerase IV	Effective antibacterial agents
Coumarins (e.g. novobiocin)	DNA gyrase and topoisomerase IV	Antibiotics, but not widely used
Camptothecin (e.g. topotecan)	Human topoisomerase I	Anticancer drug
Amsacrine (mAMSA)	Human topoisomerase II	Anticancer drug
Epipodophyllotoxins (e.g. teniposide)	Human topoisomerase II	Anticancer drug

DNA by ring stacking and with the protein by hydrogen bonding (17). Several other antitumour drugs have been shown to be inhibitors of eukaryotic topoisomerase II. These include acridines (e.g. amsacrine), ellipticines (e.g. 2-methyl-9-hydroxy-ellipticinium acetate), and epipodophyllotoxins (e.g. teniposide). Again these drugs are thought to act by stabilizing the covalent complex between the enzyme and DNA, and in many cases this is thought to be mediated by intercalation of the drug into the DNA at the site of enzyme binding. Other topoisomerase II-targeted drugs act in a different way; for example, the bisdioxypiperazine ICRF-193 is thought to act by stabilizing the closed clamp form of the enzyme (42). A structure of the ATPase domain of yeast topoisomerase II complexed with the related compound ICRF-187 shows that the drug stabilizes a transient dimer interface between two ATPase domains (25).

As well as being the targets of antibacterial agents and anticancer drugs, topoisomerases can also be the targets of toxins; the bacterial toxins CcdB and microcin B17 both act on DNA gyrase (43). CcdB is a small protein (11.7 kDa) that is part of a toxin–antitoxin system, forming a complex with CcdA, for maintaining the F plasmid in *E. coli*. Loss of the F plasmid leads to loss of CcdA, which is relatively unstable, and release of CcdB from the complex, which then kills the host cell (44). Microcin B17 is a glycine-rich peptide (3.2 kDa) containing oxazole and thiazole rings, which is produced by some strains of *Enterobacteriaceae* to inhibit phylogenetically related species (45). Both CcdB and microcin B17 kill cells by stabilizing the complex between gyrase and DNA, in a manner reminiscent to that of quinolones, but using distinct mechanisms. Study of the mode of action of such proteinaceous inhibitors may yield new ideas for the design of novel inhibitors of gyrase.

5.5 **Biological role of topoisomerases**

As we will discuss in detail in Chapter 6, changes in DNA topology occur in many cellular processes and can have profound biological consequences. Therefore, it is not surprising to find that DNA topoisomerases are involved directly or indirectly in these processes. This fundamental requirement for DNA topoisomerases derives from the double-helical structure of DNA (Chapter 1; Chapter 2, Section 2.2.1), insofar as most processes involving DNA must involve the unwinding of the DNA helix to access the bases (at least temporarily). For example, the two strands of DNA must become completely unlinked during DNA replication; topoisomerases are thought to be involved in various stages of this process. DNA gyrase is important for initiation of

replication in prokaryotes; this requirement is likely to reflect the need for negative supercoiling prior to initiation. During the elongation steps of replication, the parental DNA is being continuously unwound and topoisomerases are required to prevent the accumulation of positive supercoils (see Chapter 2, Section 2.2.1). Both type I and type II topoisomerases can act to relieve the torsional stress generated during elongation. At the termination of DNA replication, the progeny DNA molecules are frequently catenated and both type I and type II topoisomerases can, in principle, unlink the intertwined molecules (see Chapter 6, Section 6.3.3).

Unlike replication, transcription could theoretically proceed without any topological problems. However, in bacteria, transcription can lead to the supercoiling of the DNA, when, for example, the DNA is anchored to fixed points or two genes on a circular plasmid are being transcribed in opposite directions (see Chapter 6, Section 6.5). Transcription can lead to positive supercoiling ahead of the transcription complex and negative supercoiling behind it. Experiments have suggested that DNA gyrase can relax the positive supercoils and topoisomerase I can relax the negative supercoils (46). For example, highly positively supercoiled plasmids have been isolated from *E. coli* treated with coumarin drugs (which inhibit DNA gyrase), and plasmids isolated from *E. coli* strains carrying mutations in *topA* (the gene encoding topoisomerase I) can exhibit high levels of negative supercoiling (see Chapter 6, Section 6.5).

An important function of topoisomerases within the cell is the maintenance of DNA supercoiling. In bacteria the level of intracellular supercoiling influences the rate of transcription of many genes. Given that RNA polymerase unwinds DNA on binding to promoters it is to be expected that negative supercoiling should stimulate transcription. In fact negative supercoiling can both increase and decrease gene expression (see Chapter 6, Section 6.4.1). The topoisomerase genes themselves are affected by supercoiling. Lowering negative supercoiling raises the expression of the gyrase genes and reduces the expression of *topA*. It is thought that this represents a homeostatic mechanism for the control of DNA supercoiling within the bacterial cell (47).

5.6 **Conclusions**

The existence of an essential class of enzymes whose function is to interconvert different topological isomers of DNA demonstrates the vital importance of DNA topology in cells. Apart from their relevance to DNA topology, topoisomerases are of great interest from the standpoint of their mechanistic enzymology and because of their potential as drug targets.

5.7 Further Reading

Bates, A.D. and Maxwell, A. (1997). DNA topology: topoisomerases keep it simple. *Curr. Biol.* **17**, R778–R781.

Biochim. Biophys. Acta. (1998) **1400**, Issues 1–3. Collection of articles on DNA topoiscmerases.

Bjornsti, M.-A. and Osheroff, N. (1999). DNA topoisomerase protocols. Volume I: DNA topology and enzymes. In *Methods in molecular biology*, vol. 94. Walker, J.M. (ed.), Humana Press, Towota, New Jersey.

Bjornsti, M.-A. and Osheroff, N. (1999). DNA topoisomerase protocols. Volume II: Enzymology and drugs. In *Methods in molecular biology*, vol. 95. Walker, J.M. (ed.), Humana Press, Towota, New Jersey.

Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369–413.

Drlica, K. and Malik, M. (2003). Fluoroquinolones: action and resistance. *Curr. Top. Med. Chem.* **3**, 249–282.

Gatto, B., Capranico, G., and Palumbo, M. (1999). Drugs acting on DNA topoisomerases: recent advances and future perspectives. *Curr. Pharm. Des.* **5**, 195–215.

Orphanides, G. and Maxwell, A. (1994). Topoisomerases: in one gate, out the other. *Curr. Biol.* **4**, 1006–1009.

Wang, J.C. (1998). Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Quart. Rev. Biophys.* **31**, 107–144.

5.8 References

1. Wang, J.C. (1971). Interaction between DNA and an *Escherichia coli* protein ω. *J. Mol. Biol.* **55**, 523–533.
2. Champoux, J.J. and Dulbecco, R. (1972). An activity from mammalian cells that untwists superhelical DNA—a possible swivel for DNA replication. *Proc. Natl. Acad. Sci. USA* **69**, 143–146.
3. Gellert, M., Mizuuchi, K., O'Dea, M.H., and Nash, H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc. Natl. Acad. Sci. USA* **73**, 3872–3876.
4. Liu, L.F., Liu, C.-C., and Alberts, B.M. (1980). Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. *Cell* **19**, 697–707.
5. Kirkegaard, K. and Wang, J.C. (1985). Bacterial DNA topoisomerase I can relax positively supercoiled DNA containing a single-stranded loop. *J. Mol. Biol.* **185**, 625–637.
6. Declais, A.C., de La Tour, C.B., and Duguet, M. (2001). Reverse gyrases from bacteria and archaea. *Methods Enzymol.* **334**, 146–162.

7. DiGate, R.J. and Marians, K.J. (1988). Identification of a potent decatenating enzyme from *Escherichia coli*. *J. Biol. Chem.* **263**, 13366–13373.
8. Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., and Suzuki, H. (1990). New topoisomerase essential for chromosome segregation in *E. coli*. *Cell* **63**, 393–404.
9. Peng, H. and Marians, K.J. (1993). *Escherichia coli* topoisomerase IV. *J. Biol. Chem.* **268**, 24481–24490.
10. Nitiss, J.L. (1998). Investigating the biological functions of DNA topoisomerases in eukaryotic cells. *Biochim. Biophys. Acta* **1400**, 63–81.
11. Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from *Archaea* with implications for meiotic recombination. *Nature* **386**, 414–417.
12. Yin, Y., Cheong, H., Friedrichsen, D., Zhao, Y., Hu, J., Mora-Garcia, S., and Chory, J. (2002). A crucial role for the putative *Arabidopsis* topoisomerase VI in plant growth and development. *Proc. Natl. Acad. Sci. USA* **99**, 10191–10196.
13. Sugimoto-Shirasu, K., Stacey, N.J., Corsar, J., Roberts, K., and McCann, M.C. (2002). DNA topoisomerase VI is essential for endoreduplication in *Arabidopsis*. *Curr. Biol.* **12**, 1782–1786.
14. Hartung, F., Angelis, K.J., Meister, A., Schubert, I., Melzer, M., and Puchta, H. (2002). An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr. Biol.* **12**, 1787–1791.
15. Fuller, F.B. (1978). Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl. Acad. Sci. USA* **75**, 3557–3561.
16. Lima, C.D., Wang, J.C., and Mondragón, A. (1994). Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* **367**, 138–146.
17. Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279**, 1504–1513.
18. Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G.J., and Champoux, J.J. (1998). A model for the mechanism of human topoisomerase I. *Science* **279**, 1534–1540.
19. Stivers, J.T., Harris, T.K., and Mildvan, A.S. (1997). *Vaccinia* DNA topoisomerase I: evidence supporting a free rotation mechanism for DNA supercoil relaxation. *Biochemistry* **36**, 5212–5222.
20. Confalonier, F., Elie, C., Nadal, M., Bouthier de le Tour, C., Forterre, P., and Duguet, M. (1993). Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc. Natl. Acad. Sci. USA* **90**, 4753–4757.
21. Rodriguez, A.C. and Stock, D. (2002). Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA. *EMBO J.* **21**, 418–426.
22. Berger, J.M., Gamblin, S.J., Harrison, S.C., and Wang, J.C. (1996). Structure at 2.7 Å resolution of a 92K yeast DNA topoisomerase II fragment. *Nature* **379**, 225–232.

23. Morais Cabral, J.H., Jackson, A.P., Smith, C.V., Shikotra, N., Maxwell, A., and Liddington, R.C. (1997). Structure of the DNA breakage–reunion domain of DNA gyrase. *Nature* **388**, 903–906.
24. Roca, J. and Wang, J.C. (1992). The capture of a DNA double helix by an ATP-dependent protein clamp: a key step in DNA transport by type II DNA topoisomerases. *Cell* **71**, 833–840.
25. Classen, S., Olland, S., and Berger, J.M. (2003). Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187. *Proc. Natl. Acad. Sci. USA* **100**, 10629–10634.
26. Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A., and Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* **351**, 624–629.
27. Heddle, J.G., Mitelheiser, S., Maxwell, A., and Thomson, N.H. (2004). Nucleotide binding to DNA gyrase causes loss of DNA wrap. *J. Mol. Biol.* **337**, 597–610.
28. Dutta, R. and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **25**, 24–28.
29. Rybenkov, V.V., Ullsperger, C., Vologodskii, A.V., and Cozzarelli, N.R. (1997). Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277**, 690–693.
30. Nichols, M.D., DeAngelis, K., Keck, J.L., and Berger, J.M. (1999). Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo 11. *EMBO J.* **18**, 6177–6188.
31. Berger, J.M., Fass, D., Wang, J.C., and Harrison, S.C. (1998). Structural similarities between topoisomerases that cleave one or both DNA strands. *Proc. Natl. Acad. Sci. USA* **95**, 7876–7881.
32. Aravind, L., Leipe, D.D., and Koonin, E.V. (1998). Toprim–-a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213.
33. Corbett, K.D. and Berger, J.M. (2003). Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *EMBO J.* **22**, 151–163.
34. Pommier, Y., Pourquier, P., Fan, Y., and Strumberg, D. (1998). Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochim. Biophys. Acta* **1400**, 83–105.
35. Hande, K.R. (1998). Clinical applications of anticancer drugs targeted to topoisomerase II. *Biochim. Biophys. Acta* **1400**, 173–184.
36. Froelich-Ammon, S.J. and Osheroff, N. (1995). Topoisomerase poisons: harnessing the dark side of enzyme mechanism. *J. Biol. Chem.* **270**, 21429–21432.
37. Drlica, K. and Malik, M. (2003). Fluoroquinolones: action and resistance. *Curr. Top. Med. Chem.* **3**, 249–282.
38. Maxwell, A. and Lawson, D.M. (2003). The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Curr. Top. Med. Chem.* **3**, 283–303.

39. Heddle, J.G., Barnard, F.M., Wentzell, L.M., and Maxwell, A. (2000). The interaction of drugs with DNA gyrase: a model for the molecular basis of quinolone action. *Nucleosides Nucleotides Nucleic Acids* **19**, 1249–1264.
40. Lewis, R.J., Singh, O.M.P., Smith, C.V., Skarynski, T., Maxwell, A., Wonacott, A.J., and Wigley, D.B. (1996). The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by x-ray crystallography. *EMBO J.* **15**, 1412–1420.
41. Tsai, F.T.F., Singh, O.M.P., Skarzynski, T., Wonacott, A.J., Weston, S., Tucker, A., Pauptit, R.A., Breeze, A.L., Poyser, J.P., O'Brien, R., Ladbury, J.E., and Wigley, D.B. (1997). The high-resolution crystal structure of a 24-kDa gyrase B fragment from *E. coli* complexed with one of the most potent coumarin inhibitors, clorobiocin. *Prot. Struct. Funct. Eng.* **28**, 41–52.
42. Andoh, T. and Ishida, R. (1998). Catalytic inhibitors of DNA topoisomerase II. *Biochim. Biophys. Acta* **1400**, 155–171.
43. Maxwell, A. (1997). DNA gyrase as a drug target. *Trends Microbiol.* **5**, 102–109.
44. Couturier, M., Bahassi, E.M., and Van Melderen, L. (1998). Bacterial death by DNA gyrase poisoning. *Trends Microbiol.* **6**, 269–275.
45. Heddle, J.G., Blance, S.J., Zamble, D.B., Hollfelder, F., Miller, D.A., Wentzell, L.M., Walsh, C.T., and Maxwell, A. (2001). The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. *J. Mol. Biol.* **307**, 1223–1234.
46. Liu, L.F. and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027.
47. Lilley, D.M.J., Chen, D., and Bowater, R.P. (1996). DNA supercoiling and transcription: topological coupling of promoters. *Quart. Rev. Biophys.* **29**, 203–225.
48. Maxwell, A. and Gellert, M. (1986). Mechanistic aspects of DNA topoisomerases. *Adv. Prot. Chem.* **38**, 69–107.
49. Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369–413.
50. Wall, M.K., Mitchenhall, L.A., and Maxwell, A. (2004). *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. USA* **101**, 7821–7826.
51. DeLano, W.L. (2002). The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org
52. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
53. Corbett, K.D., Shultzaberger, R.K., and Berger, J.M. (2004). The C-terminal domain of DNA gyrase A adopts a DNA-bending β-pinwheel fold. *Proc. Natl. Acad. Sci. USA* **101**, 7293–7298.

6

Biological consequences of DNA topology

6.1 Introduction: the ubiquity of topology

We hope that by the time you have reached this chapter you will have a good grasp of the concepts of DNA topology and will appreciate that it is an unavoidable issue when dealing with virtually any naturally occurring DNA molecule. The purpose of this chapter is to put DNA topology into the context of important biological processes: DNA replication, transcription, and recombination. You will see how for some processes (e.g. initiation of DNA replication), supercoiling is an essential prerequisite for the reaction to occur, whereas for others (e.g. transcription elongation), it is a consequence of the reaction that has to be controlled. In general, supercoiling is a 'good thing' in cells, and is maintained and controlled, whereas knotting and catenation are 'bad things' to be systematically removed. This distinction will be reflected in the discussions below. Organisms go to a lot of trouble to control the topology of their DNA molecules as indicated by the number of DNA topoisomerases that have been identified (Chapter 5). In some cases there is redundancy with more than one enzyme being capable of carrying out the same reaction.

Most of the following sections will centre around supercoiling as virtually all naturally occurring DNA molecules seem to be supercoiled. With the exception of some thermophilic species (see Section 6.1.1), all organisms appear to maintain their DNA molecules in a negatively supercoiled state. From Chapter 2 you will know that this favours DNA unwinding and is likely to facilitate processes where proteins require access to the DNA bases, such as replication

and transcription. Before talking about these processes in detail, we will first review the effects of negative supercoiling *in vivo*.

6.1.1 *General biological consequences of negative supercoiling*

DNA supercoiling is an attribute of almost all DNA *in vivo*. Plasmids, bacterial chromosomes, mitochondrial and chloroplast DNA, and many viral genomes can occur as closed-circular DNA, and as we explained in Chapter 2, linking number is an inherent property of such molecules. Eukaryotic chromosomes and other DNAs such as some bacterial chromosomes and yeast plasmids, although consisting of linear DNA, appear to be anchored to a nuclear matrix (scaffold) at a number of sites, and the domains between such attachment sites behave in a topological sense as closed-circular loops, since their ends are fixed relative to one another and rotation is restricted (1, 2). Such closed-circular molecules or closed domains are almost invariably negatively supercoiled, equivalent to an unwinding of the right-handed DNA helix, the only exceptions being in extremely thermophilic archaea and eubacteria, where plasmids have been isolated with positive supercoiling (3). As we saw in Chapter 5, these bacteria contain a 'reverse gyrase', which is able to introduce positive supercoiling into a relaxed DNA substrate. It seems clear that the presence of positively supercoiled DNA in such species is an adaptation to the extreme conditions in which they live; positively supercoiled DNA is likely to be more resistant to the unwinding and denaturation of DNA that would be expected at very high temperatures. It has also been suggested that positive supercoiling (which will tend to increase the twist of DNA) acts to compensate for the untwisting of the helix caused by high temperature (see Chapter 2, Section 2.5.3), so that the value of the helical repeat, which may be important for regulatory processes, is relatively unchanged (4).

An often-quoted consequence of the supercoiling of DNA is that it allows its compaction into a very small volume. Chromosomal DNA molecules are amazingly long, by comparison with the size of bacteria and eukaryotic nuclei; for example, the *Escherichia coli* chromosome is around 1.5 mm long, while the cell has a diameter of less than 1 μm. The problem is more acute in eukaryotes, where chromosomes can be several centimetres in length and must be contained within a nucleus also less than 1 μm across. However, it has been pointed out (5) that plectonemic supercoils, the conformation adopted by supercoiled free DNA, are very inefficiently compacted (around 2.5-fold). Toroidal winding (see Chapter 2, Section 2.4.3) is much more efficient, and in eukaryotes at least, it is the wrapping of DNA in nucleosomes and higher-order structures (see Section 6.2.2) that provides the major compacting effect.

More important, perhaps, is that DNA supercoiling has a direct influence on DNA-associated processes *in vivo*, for the most part involving the interaction of specific proteins with DNA. Many interactions between DNA and protein can be divided into a number of discrete steps that individually may be affected by DNA supercoiling. These include the binding of proteins to the DNA, the bringing together of two or more sites on the DNA ('synapsis') and, in the case of recombinases and topoisomerases, strand-transfer processes.

The binding of proteins to DNA is often supercoiling dependent. Negatively supercoiled DNA is in a high-energy conformation compared with unconstrained DNA (see Chapter 2, Section 2.6). Protein binding may relieve this excess energy. The unfavourable deformations associated with negative supercoiling are the untwisting of the DNA helix and the negative writhing of the helix axis. It follows that any binding process that requires or makes use of these distortions will be favoured by negative supercoiling. The most obvious examples are the binding of proteins that require the unwinding of the DNA helix (such as those involved in DNA replication and transcription). In addition, a number of DNA–protein complexes involve the wrapping of DNA around the protein (e.g. nucleosomes (Section 6.2.2), bacteriophage λ integration complex (Section 6.6.2.1)), and their binding is promoted by supercoiling, due to the stabilization of writhing.

The activities of a number of DNA-specific proteins involve synapsis (e.g. certain transcriptional activators and repressors (Section 6.4.2), and site-specific recombination proteins (Section 6.6.2)). Plectonemic, or interwound, supercoiling (Chapter 2, Section 2.4.3) provides a way of bringing two distant sites on DNA together (6, 7). 'Slithering' refers to the movement of the DNA around a fixed superhelical axis that occurs in plectonemically supercoiled DNA (referred to as a 'conveyor-belt' motion). One can imagine such motion bringing remote sites into apposition, thus facilitating synapsis (see Sections 6.6.2.1 and 6.6.2.2).

DNA breakage–reunion reactions are features of both recombinases and topoisomerases (see Chapter 5); in the former case the broken ends are rejoined to new partners, in the latter they are reunited with their original partner. Often, these reactions lead to changes in linking number of the DNA substrate. Clearly, if the reaction occurs on negatively supercoiled DNA, then processes leading to a positive linking difference (ΔLk), such as DNA relaxation by topoisomerases, will be favoured. The processes of binding, synapsis, and strand transfer are general phenomena found in many systems of DNA–protein interaction; Section 6.6.2 discusses these processes in detail in relation to site-specific recombination.

The excess free energy associated with supercoiling also influences the formation of Z-DNA, cruciforms, and DNA triplexes (see Chapter 1, Sections 1.2.4, 1.3.1, and 1.3.2; Chapter 2, Section 2.8.1). The formation of all these structures has the effect of dramatically reducing the twist of the DNA. Z-DNA formation converts a section of DNA from right-handed (+ve) to left-handed (−ve) twist, and the formation of a cruciform or intramolecular triplex corresponds to a complete untwisting of a length of DNA. In negatively supercoiled DNA, this twist change will be accompanied by an energetically favourable reduction in the (negative) writhe. This saving in energy compensates for the positive free energy associated with the formation of the more unstable structure and its junctions with normal B-DNA, and above a threshold specific linking difference the alternative structure will be stable (see Chapter 2, Section 2.8.1). It is thought that the average levels of supercoiling *in vivo* are too low to stabilize these structures, but we will discuss a situation in which they might form in Section 6.5.1.

Additionally, the helical structure of DNA and its closed-circular nature cause a number of complications in processes such as replication and transcription that require specific mechanisms for their resolution. The consequences of negative supercoiling and other DNA topological forms for DNA-associated processes will be considered in more detail in the following sections.

6.2 **Genome organization**

6.2.1 *Prokaryotes*

An important distinction between prokaryotes and eukaryotes is that the former have no membrane-bound nucleus that contains the genetic material. It is often thought that bacteria therefore have a less-organized chromosome structure than eukaryotes. However, it is clear that structural organization of bacterial DNA does exist and that DNA topology has an important role to play. The best-studied prokaryotic genome is that of *E. coli*, which consists of a 4639 kb closed-circular DNA molecule. The *E. coli* chromosome is negatively supercoiled (8) and is compacted about 1000-fold into a structure called the nucleoid within the bacterial cell (9). Nucleoids may be visualized by electron microscopy of thin sections of bacteria or fluorescence microscopy of living cells (*Figure 6.1*). They appear as irregular dispersed structures that occupy less than half of the intracellular volume, and are characterized by their apparent lack of ribosomes. Although the nucleoid is not surrounded by a membrane, as in the eukaryotic nucleus, it appears to have some similarities in overall organization. The average concentration of DNA in the

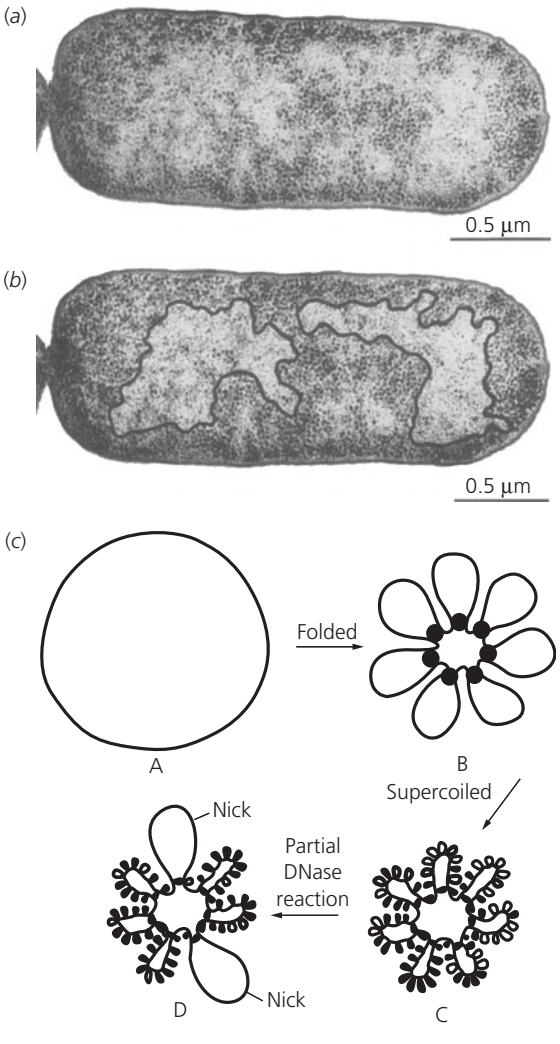


Figure 6.1 The nucleoid. (*a*) Electron micrograph of a thin section of an *E. coli* cell prepared by cryosubstitution (samples were frozen and placed in acetone at −90°C, where the water ice is replaced by the acetone) (9). Panel (*b*) shows the same cell with the boundaries of the nucleoid indicated. (*c*) Model of the bacterial chromosome: A—Completely unfolded DNA. B—DNA folded into domains by protein–DNA interactions. C—Supercoiling of the DNA and further condensing by short-range interactions. D—Consequence of nicking in domains, which relaxes supercoiling (reproduced from ref. 9 with permission; copyright (1999) ASM).

nucleoid is 20–50 mg ml^{-1}, which is comparable to that of interphase nuclei of eukaryotes, but considerably less than the 6–800 mg ml^{-1} concentration in bacteriophage heads.

A major difference between prokaryotic and eukaryotic DNA is that, while the writhing of DNA around histones accounts for virtually all the negative supercoiling in eukaryotic chromosomes (see Section 6.2.2), in prokaryotes, some of the negative supercoiling is not accounted for by protein binding. Although a plasmid isolated from an *E. coli* cell typically has $\sigma = \sim -0.06$, it has been estimated that approximately −0.025 (40%) of this is unconstrained,

that is, able to adopt the plectonemic conformation normally seen *in vitro* (see Section 6.2.1) (10); the remainder must consist of writhe and/or changes in twist stabilized by the binding of proteins of various kinds. It is assumed that the same proportions apply to the bacterial chromosome. This difference reflects the alternative methods for the introduction of negative supercoiling in the two types of cell. In eukaryotes, the formation of nucleosomes on initially relaxed DNA, followed by relaxation of the compensatory positive supercoiling by topoisomerases leads to negative supercoiling wholly constrained within nucleosomes (see Chapter 2, Section 2.5.5; Chapter 3, Section 3.5). In contrast, prokaryotes have an enzyme, DNA gyrase, that actively introduces negative supercoiling (see Chapter 5), as well as other topoisomerases capable of relaxing negative supercoils. Hence, the level of unconstrained supercoiling in bacteria is controlled by the relative activities of competing topoisomerases and the presence of DNA-binding proteins that stabilize negative supercoiling.

The unconstrained supercoiling in bacteria can be relaxed by the introduction of nicks into the DNA. Studies of the number of nicks required to completely relax the chromosomal DNA have suggested the existence of around 50 discrete topological domains (11); this proposal is also supported by electron microscopy (9). This gave rise to the idea that the DNA is organized as long supercoiled loops that are constrained from rotating at each end by interaction with nucleoid proteins. There is also evidence for the anchoring of bacterial DNA to the cytoplasmic membrane via membrane-associated proteins (9), and there is good evidence for the existence of transcription factories (see Section 6.5) in which RNA polymerase and other proteins form an essentially immobile complex through which the DNA translocates (12). There is thus a variety of interactions which may serve to organize the bacterial chromosome. A diagrammatic representation of the bacterial nucleoid is given in *Figure 6.1c*.

Although a search for analogues of the eukaryotic histone proteins in bacteria has led to the discovery of a number of proteins that may be involved in the compaction of the DNA, structures as well defined as the nucleosome have not been discovered. The proteins involved in genome organization in prokaryotes are exemplified by the *E. coli* proteins HU, H1 (or H-NS) and its paralogue StpA, and IHF (9, 13). HU is a major DNA-binding protein in bacteria that is known to constrain negative supercoils, preferentially bind bent DNA, and, like histones, to have little sequence specificity. However, there is insufficient HU in the cell for it to package the entire bacterial chromosome, and it may act to facilitate short-range structural transitions in conjunction with other regulatory proteins. H-NS has a number of similarities to HU in

that it can compact DNA and shows little sequence specificity. Also like HU it binds preferentially to bent DNA. IHF (integration host factor) was first identified as a sequence-specific DNA-binding protein that binds at the *att* site of *E. coli* to facilitate phage λ integrative recombination (see Section 6.6.2.1). Like HU, IHF facilitates DNA bending and has been shown to act as a positive regulator of a number of genes. It appears that these three DNA-binding proteins have overlapping functions and together participate in the organization of the bacterial nucleoid. In addition to their roles in bacterial chromosome structure, nucleoid proteins have also been implicated in the control of bacterial gene expression (14) (see Section 6.4).

Although uncommon and not widely studied, knotted DNAs are found in biological systems; these include plasmids in *E. coli* cells harbouring mutants in the gyrase genes (15) and DNA in bacteriophage P2 and P4 heads (16, 17) (see also Chapter 4). However, as mentioned before, knots are undesirable and tend to be removed by topoisomerases.

6.2.2 *Eukaryotes*

As we have mentioned, the genomic DNA in cells must be highly compacted in order to be contained in the space required. In eukaryotes, the first stage of this compaction is the winding of the DNA in the nucleosome, and the resulting histone-coated DNA is referred to as chromatin.

The nature of the wrapping of DNA in the nucleosome has been considered in detail in Chapter 3, Section 3.5. Briefly, about 146 bp of DNA are wrapped in 1.6 left-handed superhelical turns around a protein octamer consisting of histones: Highly basic, positively charged proteins that help to neutralize the negatively charged phosphate backbone of the DNA. The 'core' histones are H2A, H2B, H3, and H4; two copies of each are present in the nucleosome particle. This basic unit is repeated many times in eukaryotic chromatin; indeed it seems likely that almost all the DNA is constrained in this way. It follows from this that the negative supercoiling of the DNA is largely partitioned into writhing of the helix axis stabilized by interaction with the histones, and hence is not directly available for promoting the binding of ligands that unwind the DNA helix (see Section 6.1.1). It is likely that dissociation of DNA from the histone octamer has a part to play in the initiation of processes such as replication and transcription. The unconstrained supercoiling resulting from this dissociation may be harnessed to promote the untwisting of the DNA helix.

The structure of the nucleosome core particle with a 146 bp DNA molecule has been solved by X-ray crystallography at 2.8 Å, and more recently at

a resolution of 1.9 Å with a 147 bp DNA molecule (18, 19). It is now possible to see in atomic detail how the bound DNA is organized by the core histones (see Chapter 3, *Figure 3.7*.). The DNA is wrapped around the histone octamer in 1.6 turns as an irregular superhelix with negative writhe. Many contacts are made between the DNA phosphates and main-chain atoms of the histone proteins, allowing the interaction to be rigid but non-specific. The path of the DNA around the protein is not smooth but is distorted through bends at various positions. The DNA is actually overtwisted, not untwisted as might be expected for negatively supercoiled DNA; the helical repeat of the bound DNA is 10.35 bp/turn, in contrast to the value of ~10.5 bp/turn for B-DNA in solution (Chapter 1).

There are several increasing levels of compaction that have been identified in chromatin, in addition to the nucleosome itself. The first is the nucleosome filament, an extended array of nucleosomes along a length of DNA ('beads on a string'), which also contains histone H1, a protein associated with the entry and exit points of DNA from the histone octamer (this filament has a zigzag appearance in electron micrographs). In the 30 nm filament or solenoid, nucleosomes have been suggested to be wrapped into a compact left-handed helix of diameter 30 nm, although there is considerable evidence that this fibre is more irregular (*Figure 6.2*; (20–22)). Higher orders of compaction, for example, a helical winding of the 30 nm filament, also occur. The level of compaction exhibited by chromatin is dependent on the stage of the cell cycle; chromosomes are highly compacted at mitosis, but are more extended during interphase, consisting mostly of the 30 nm fibre. In addition, chromatin that is being actively transcribed has a less condensed structure than inactive chromatin, characterized by the loss of histone H1 and most higher-order structure, although many nucleosomes are still present.

It is now clear that chromatin structure is dynamic and that chromatin remodelling complexes and histone modifications can change the structure of nucleosomes such that the DNA becomes less tightly bound to the histone core (23, 24). Such remodelling means that the DNA is more readily accessed by other proteins for the processes of gene expression, replication, and repair, and also means that nucleosomes are not static and can move to alternative positions on DNA.

6.3 **Replication**

DNA replication is a fundamental cellular process that occurs by a semi-conservative process in which each parental strand of the DNA duplex is the

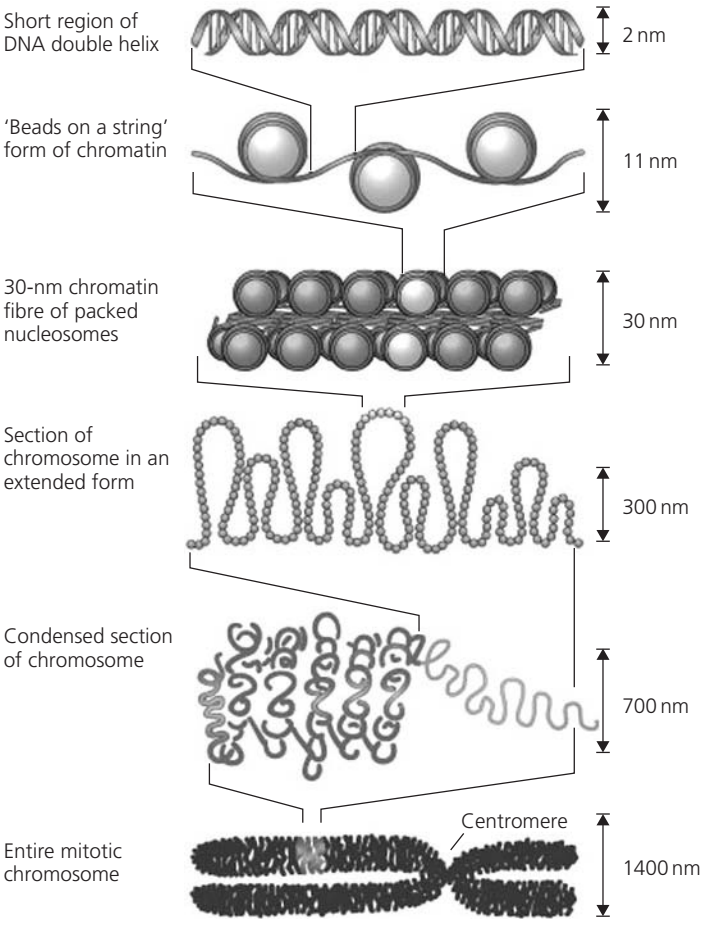


Figure 6.2 A model for chromatin condensation. Schematic illustration showing the steps involved in the folding of DNA into chromatin. DNA (top) is wound around nucleosomes, which can be packed into the 30 nm solenoid, which is further condensed in stages to finally form the metaphase chromosome (bottom) (reproduced from ref. 106 with permission; copyright (2003) Nature Publishing Group).

template for the synthesis of a new daughter strand. We have already seen the problem of unwinding the DNA ahead of replication (Chapter 2, Section 2.2.1). Since the realization that the *E. coli* chromosome is a single closed circle, this problem has been recognized as an explicitly topological one (25); completion of replication requires that the number of links (linking number) of the parental duplex be reduced to zero. Later it was proposed that this topological problem could be resolved by the action of DNA topoisomerases (26);

see Chapter 5). In eukaryotes, where the chromosome has been shown to be linear, topological problems will still occur due to the constraining of DNA into loops by association with proteins or other cellular components. DNA replication can be conveniently divided into three stages: initiation, elongation, and termination, and the topological aspects of all three are dealt with in the following sections. These issues are relevant to both prokaryotic and eukaryotic systems, but the most accessible work in this area has been carried out in bacteria (principally *E. coli*) particularly with plasmid or bacteriophage DNA substrates (27), where the analysis of replication products and intermediates is more tractable. Replication in eukaryotes is more complex (28) but the same topological effects will occur.

6.3.1 *Replication initiation*

In all systems, replication of DNA proceeds from specific sites known as origins of replication. In bacteria a single origin is present in the chromosome (*oriC* in *E. coli*), whereas in eukaryotes there are multiple replication origins. The binding of sequence-specific proteins at the origin precedes DNA synthesis. As the origin region must be unwound before the initiation of DNA synthesis it is clear that negative supercoiling is likely to favour the initiation process. In circular bacterial genomes DNA replication proceeds bidirectionally from the unique origin sequence, in so-called theta-type replication. In *E. coli*, DnaA binds to the origin and organizes it into a nucleoprotein complex. The complex of DnaA and *oriC* is more stable if *oriC* is present on a negatively supercoiled DNA (29). A similar requirement for negative supercoiling has been demonstrated for the bacteriophage λ O protein, which binds to *ori*, the replication origin (30). A further demonstration of the importance of supercoiling in initiation is provided by the yeast autonomously replicating sequences (ARS), which unwind readily when present in negatively supercoiled DNA; the ease of unwinding is correlated with their efficiency as replication origins *in vivo* (31).

It seems that topoisomerases may not be involved directly in replication initiation complexes but that the supercoiled state of the template DNA is a major determinant of the efficiency of initiation. In bacteria, DNA gyrase is required for the initiation of replication, but this requirement is thought to relate to its supercoiling activity, rather than to a specific role for gyrase at the origin. Indeed, in experiments using an *in vitro* system, in the absence of gyrase, both replication forks form at *oriC* but only one is allowed to proceed (32). It seems that this is the consequence of a topological block that can be relieved by gyrase, or another topoisomerase.

6.3.2 *Replication elongation*

As the replication fork proceeds along a DNA template, the unwinding of the strands causes positive supercoils to build up ahead of the fork (Chapter 2, Section 2.2.1). This is because the fork cannot rotate to follow the helix of the DNA, either through sheer size, or because it is tethered to organizing structures in the cell. In the case of *E. coli*, DNA gyrase is the obvious candidate topoisomerase to remove the positive supercoils (which would otherwise rapidly inhibit elongation) since it is efficient at relaxation of positive supercoils, whereas topoisomerase I will relax only negative supercoils (see Chapter 5, Section 5.3.1). However, it has been found that topoisomerase IV can also fulfil this function, at least in situations where gyrase is inhibited (33), suggesting that the two enzymes (gyrase and topo IV) may have overlapping functions. Additionally it has been found that in the absence of gyrase, *E. coli* topoisomerase III (see Chapter 5) is also capable of supporting nascent chain elongation during replication (34).

As the elongating replication fork nears the terminus of replication, the length of the unreplicated DNA becomes shorter and shorter, resulting in less room to accommodate the positive supercoils building up ahead of the fork and for the topoisomerases to act to remove them. A solution to this potential problem is the possibility that the positive supercoils may diffuse behind the fork, by rotation of the fork to create interwinding of the daughter duplexes (*Figure 6.3*) (35). These interwindings have been named 'precatenanes' as they will become proper catenanes if they are not removed before replication is complete (see Section 6.3.3). As precatenanes contain nicks, a type I topoisomerase such as topoisomerase III could support elongation by removing precatenanes behind the fork, before ligation of the daughter strands is complete, accounting for its ability to support this process (36). Evidence for the existence of precatenanes as replication intermediates has also been

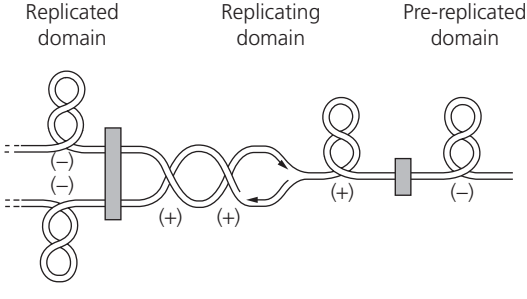


Figure 6.3 Model for the topology of a replicating chromosome. The replication fork (centre) is shown delimited by domain boundaries (grey boxes). Replication causes positive supercoiling ahead of the fork and, potentially, precatenanes behind it (reproduced from ref. 35 with permission; copyright (2001) National Academy of Sciences, USA).

obtained in eukaryotic systems (37). However, it is not easy to obtain definitive evidence for precatenanes so their origins and mechanisms of resolution remain somewhat speculative.

In the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, the elongation of DNA synthesis is much reduced by the inactivation of both topoisomerases I and II, but not by the inactivation of either (38), implying that in eukaryotes, either topoisomerase I or II can act to relieve the strain caused by elongation; this is to be expected since both enzymes can efficiently remove positive supercoils (see Chapter 5, *Table 5.1*). The type IA enzyme yeast topoisomerase III cannot support elongation in yeast.

6.3.3 *Termination of replication*

When two forks converge at the end of DNA synthesis during replication, the unlinking of the parental DNA strands may not coincide with their replication to two daughter DNA helices (*Figure 6.4*). The product may then be a pair of catenated DNA rings, or catenated topological domains in the case of linear replicons. Such structures have been documented as intermediates in replication in a number of systems, as discussed in Chapter 4, Section 4.3.1 (39).

In principle, catenated DNA molecules may be resolved by either type I or type II topoisomerases, but in the case of a type I, the reaction requires a nick,

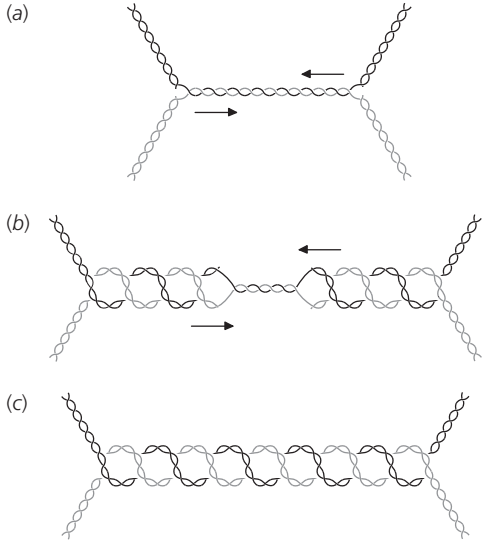


Figure 6.4 Formation of catenated DNA at the termination of replication. At the terminus of replication, converging replication forks (*a*), lead to the interwinding of daughter molecules and the formation of precatenanes (*b*). Upon completion of replication, the products are catenated DNA circles (*c*).

or single-stranded gap in at least one of the substrate DNAs (see Chapter 5). If the component rings of the product catenane are converted to a closed double-stranded form (*Figure 6.4c*), only a type II topoisomerase would be able to unlink the products. In *E. coli*, DNA gyrase has been shown to be capable of resolving catenanes formed by bacteriophage λ Int protein (see Section 6.6.2.1; (10)), and gyrase has been shown to be essential for proper segregation of the chromosome, which is consistent with a role in the decatenation of daughter chromosomes (40). However, more recent work has shown that DNA topoisomerase IV is much more effective at decatenating replicated DNA molecules and is most likely to be the enzyme required at the terminal stages of replication (41, 42). It has also been shown that *E. coli* topoisomerase III can decatenate replication products (34); in this case at least one of the products must contain a single-stranded break, which implies that any decatenation by this enzyme must occur before the daughter strands are completely ligated.

A further possibility at the end of replication is the formation of 'hemicatenanes' (43). In these molecules the parental strands of the two daughter chromosomes are still linked (*Figure 6.5*). Such molecules are electrophoretically distinct from conventional catenanes and can be resolved by the action of type I topoisomerases, since they contain single-stranded regions.

In eukaryotes, catenanes of a closed-circular plasmid in *Sacch. cerevisiae* have been shown to accumulate on inactivation of topoisomerase II, suggesting that this enzyme is necessary for the segregation of chromosomes at the termination of replication (44). In *Schizo. pombe* it has been shown that the enzyme is required during mitosis, consistent with a role in chromosome segregation (45). In studies of plasmid replication in *Xenopus* cell extracts it has been demonstrated that topoisomerase II is required for the unlinking of daughter chromosomes during DNA replication (37).

The picture that emerges is one of multiple roles for DNA topoisomerases during DNA replication and of a number of topological requirements and consequences of this process. Initiation of replication requires negatively supercoiled DNA, which in bacteria is the responsibility of DNA gyrase. The process of elongation generates positive supercoils ahead of the fork, which can be removed by both type I and type II topoisomerases. It is feasible for the positive supercoils to diffuse behind the fork and form precatenanes, which can, in principle, be resolved by both type I and type II enzymes. Converging forks at the termination of replication generate catenated products that require a type II topoisomerase; the resolution of such catenanes seems to be a (or perhaps the) principal function of topoisomerase II in eukaryotes and topoisomerase IV in bacteria.

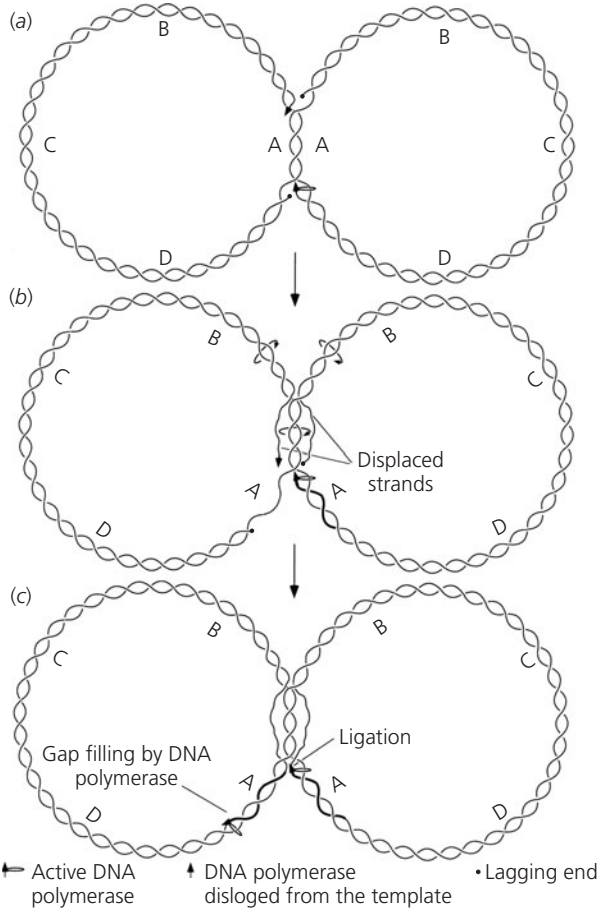


Figure 6.5 Termination of replication showing the formation of hemicatenanes. (*a*) Two replication forks approach each other at the termination of replication. (*b*) Displacement of leading and lagging strands from one fork. (*c*) Ligation of displaced strands causing formation of hemicatenane (reproduced from ref. 43).

6.4 **Control of gene expression**

Gene expression in its broadest sense encompasses all the factors that determine whether protein is produced from a given gene, or group of genes. DNA topology, specifically DNA supercoiling, is an important factor influencing gene expression at the level of transcription initiation; it is clear that the process of binding of RNA polymerase and associated transcription factors to DNA will have topological consequences. The initiation of transcription involves the binding of RNA polymerase to the promoter region of a gene, and its progression to continuous RNA synthesis through a number of discrete steps. This process has been intensively studied in prokaryotes and the ensuing description will focus primarily on results from bacterial systems (46). However,

significant progress has been made in eukaryotic systems, where the process of gene expression is necessarily more complicated, and, where appropriate, these will also be described.

The topological issues associated with transcription initiation can be divided into two aspects: The binding of RNA polymerase itself to DNA, and the binding of transcription factors that influence the initiation of transcription.

6.4.1 *RNA polymerase binding*

Transcription initiation is a complicated process that involves a number of phases, including promoter location by RNA polymerase, formation of the initiation complex, synthesis of the first phosphodiester bonds, and movement of RNA polymerase from the promoter region (elongation, see below) (47). This presents a number of opportunities for control, including control by DNA topology. The processes involved in RNA polymerase binding will mostly be discussed in terms of prokaryotes, as exemplified by *E. coli*, and it is to be expected that the general principles, if not the specifics, will be applicable to eukaryotic transcription, about which there is much less detailed information (48). A major advance in our understanding of the process of transcription initiation has come from structural information on RNA polymerases. These include phage T7 RNA polymerase, *E. coli* and *Thermus aquaticus* polymerases (including structures with DNA), and yeast RNA polymerase II (49–51). From these and other studies there is direct and indirect evidence of DNA distortion caused by interaction with RNA polymerase, which means that this interaction would affect, and be affected by, DNA topology. Prokaryotic RNA polymerase consists of four different protein subunits in the so-called core complex ($\alpha_2, \beta, \beta', \omega$); this complex is joined by the σ subunit in RNA polymerase holoenzyme, which is the form competent to initiate transcription.

The site on DNA to which RNA polymerase binds and where transcription initiation occurs is the promoter. Most promoter sequences in *E. coli* contain variations on the '−10' and '−35' consensus regions (the numbers refer to positions upstream of the transcription start point), which are recognized by the RNA polymerase holoenzyme (*Figure 6.6*) or more specifically by the σ subunit. Transcription is initiated when polymerase binds to the promoter sequences, locally unpairs the helix and forms the first phosphodiester bond. During this process a series of protein–DNA complexes are formed, which involve some degree of unwinding of the DNA (52). The initial RNA polymerase complex is known as the 'closed' complex, and it involves the wrapping of the DNA around the protein. The linking difference (ΔLk) associated with the closed complex is around −1.25 (53), that is, the wrapping is

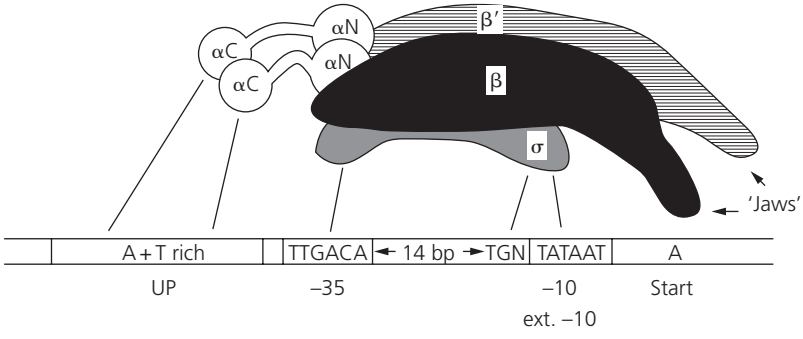


Figure 6.6 Interaction of RNA polymerase with DNA. Schematic representation of *E. coli* RNA polymerase interacting with a 'typical' promoter, showing consensus −10 and −35 sequences (reproduced from ref. 47 in modified form, with permission; copyright (1998) ASM).

left-handed. The closed complex is then converted to the 'open' complex; the ΔLk is now about −1.7, and the strands around the −10 region become completely unwound and separated over a length of ~14 bp (54). It is hypothesized that the negative writhe associated with the closed complex is converted to untwisting of the promoter in the open complex; in addition to the strand separation, the open complex must also contain some residual wrapping or untwisting. The open complex is competent to begin the synthesis of RNA; however, after the addition of a few nucleotides, the enzyme may release the transcript and begin synthesis again in a process known as abortive cycling. Transcription proper begins when the σ factor dissociates from the complex and RNA synthesis proceeds beyond the promoter region. The steps involved in transcription initiation are outlined in *Figure 6.7*.

We would expect that both the binding of RNA polymerase to form the closed complex, and the isomerization to the open complex will be promoted by increasing negative supercoiling. The closed complex stabilizes the writhing associated with negative supercoiling, and the underwinding of the DNA helix associated with negative supercoiling will promote the strand separation in the open complex. This suggests that transcription initiation is likely to be promoted by increasing levels of supercoiling, and indeed it was shown in the 1970s that, in general, levels of transcription are increased on negatively supercoiled templates. The response of individual promoters to supercoiling is more complex; of those that have been tested, some are stimulated by increasing negative supercoiling, some are unaffected, and others are inhibited (55). The precise effect is likely to be modulated by the sequence and local conformation of the −35 and −10 regions, as well as the spacing between them, that is, changes in supercoiling may be result in changes in the twist of the DNA in

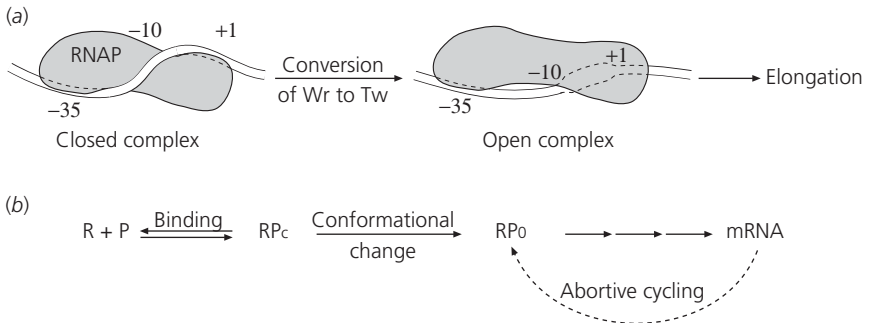


Figure 6.7 Transcription initiation at *E. coli* promoters. (*a*) The initial closed complex between RNA polymerase (RNAP) and the promoter involves the winding of DNA around the protein in a left-handed sense. The numbers indicate the approximate position of the −35, −10 regions and the transcription start point (+1). Isomerization to the open complex results in the conversion of writhe to twist, and the unpairing of a section of helix around the −10 region, prior to RNA synthesis. (*b*) The same process, shown symbolically. R, RNA polymerase; P, promoter; RP_C, closed complex; RP_O, open complex. The abortive cycling step, after the synthesis of a short oligoribonucleotide is also indicated (see text).

this region (see Chapter 2, Section 2.4). It is thought that subtle changes in twist in the −10 and −35 regions can have a marked effect on RNA polymerase interaction.

The variation of response to supercoiling in different promoters is exemplified by the genes encoding *E. coli* topoisomerase I and DNA gyrase. It is thought that the transcription of these genes is oppositely regulated in a homeostatic mechanism to maintain an appropriate level of supercoiling *in vivo*. Increased negative supercoiling increases the initiation of transcription from the topoisomerase I gene (*topA*) and reduces initiation from the *gyrA* and *gyrB* genes encoding the subunits of DNA gyrase, the latter contrary to the expected effect of supercoiling on DNA binding and unwinding by RNA polymerase (56). In the case of the gyrase genes, this effect has been shown to be dependent on the DNA sequence surrounding the site of initiation of RNA synthesis (57), and it has been proposed that this may be mediated by an effect on the stability of the abortive cycling complex; relaxation of the template may cause the destabilization of the complex and progression to elongation.

DNA supercoiling changes in response to environmental alterations, and the intracellular level of supercoiling may be used as a co-ordinate regulator of a number of genes by environmental conditions (58). For example, changes in anaerobic versus aerobic growth, changes in the nutrient conditions and osmolarity of the medium, and changes in growth temperature can influence the intracellular levels of supercoiling. Such environmental changes can be

signals for gene activation in virulent bacteria (56). One possibility is that the supercoiling effect is mediated through the intracellular [ATP]/[ADP] ratio and its effect on the activity of DNA gyrase. This mechanism has been implicated as a global regulator of gene expression in bacteria (59). Changes in DNA topology locally or globally will influence the ability of DNA to loop and for distant regions of DNA to contact each other. Support for this idea comes from experiments with the *proU* promoter of *Salmonella typhimurium*, from which transcription is induced upon increase in osmotic pressure (60). Experiments have shown that promoter activation is a consequence of increased DNA writhing with increased DNA supercoiling. Other changes induced by increased supercoiling will include local helix destabilization, which may stimulate transcription by promoting open complex formation by RNA polymerase. Finally it should be noted that at very high levels of negative supercoiling, beyond those normally present *in vivo*, the overall initiation of transcription is inhibited. A possible explanation for this is the formation of alternative DNA structures such as cruciforms or Z-DNA, which may not be efficient templates for RNA polymerase (see Section 6.1.1; Chapter 1, Section 1.2.4; Chapter 2, Section 2.8.1).

6.4.2 *Transcriptional regulatory proteins*

Another way in which DNA topology can influence the control of gene expression relates to the binding of regulatory proteins that affect (activate or repress) transcription by RNA polymerase. A great many such proteins are now known, in both prokaryotes and eukaryotes, and a considerable number have been subjected to high-resolution structure determination by X-ray crystallography or NMR (61, 62). The binding of such proteins to their DNA target sites can be influenced by DNA topology, and the formation of the protein–DNA complex can have topological consequences.

Transcriptional regulatory proteins can essentially operate in two ways: Either by contacting RNA polymerase directly or by altering the conformation of the promoter DNA (63). Direct contact with RNA polymerase is often through the C-terminal domain of the α subunit or through a region of the σ factor (*Figure 6.6*). Many transcriptional regulators act in this way and their interaction with polymerase can be influenced by DNA supercoiling. An example is NtrC of *E. coli*, which binds to a site (the enhancer) centred ~110 bp upstream of the *glaAp2* promoter. The enhancer is effective if placed up to 15 kb away from the promoter. Experiments have shown that DNA supercoiling is an essential requirement for enhancer action over

large distances and facilitates functional enhancer–promoter communication probably by bringing the enhancer and promoter into close proximity (64).

Alteration of DNA conformation by transcriptional regulatory proteins can occur in a number of ways, including the induction of sharps bends or kinks in the DNA, and the modulation of local twist. A good example of the latter is the MerR protein, which is encoded by the mercury resistance locus of the transposon Tn501 (65). The repressor form of MerR binds between the −10 and −35 regions of its target promoter and forms an inactive complex with RNA polymerase in which the DNA is bent towards the MerR protein. Binding of mercuric ions, Hg(II), triggers a conformational change in which the polymerase complex is converted to the open form. This activation involves unbending and untwisting of the operator DNA (*Figure 6.8*); in this case, addition of Hg(II) converts MerR from a repressor to an activator via local DNA conformation changes. Another system in which changes in local DNA twist play a role in protein–DNA interaction is the repressor of bacteriophage 434. Structural studies show that the DNA in the centre of the repressor–operator complex is overtwisted by ~30°, and the intrinsic twist of the operator appears to determine repressor affinity (66). Alteration of *in vivo* levels of supercoiling may modify DNA twist and therefore affect repressor affinity for the operator DNA.

A more common mechanism for DNA conformation-mediated activation/repression of gene expression is the introduction of sharp bends or kinks in the DNA. These can influence gene expression either by permitting

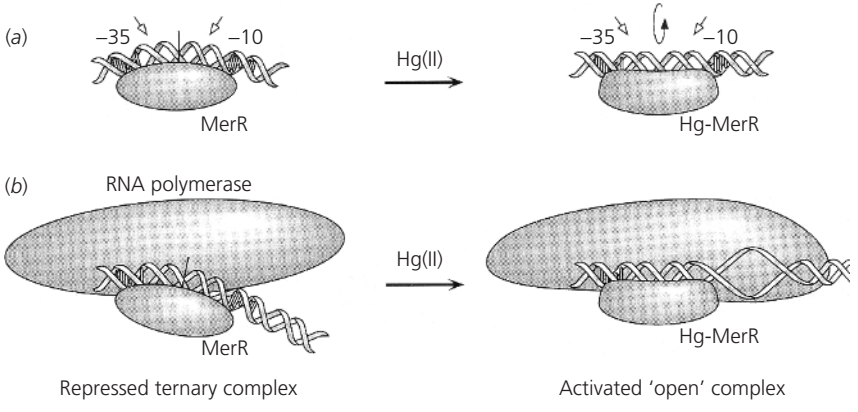


Figure 6.8 Model of MerR-induced distortions in DNA. Open arrows indicate the locations of kinks. The presence of Hg (II) ions leads to untwisting of the operator DNA (*a*), and in the presence of RNA polymerase (*b*), leads to activation of transcription (reproduced from ref. 65 with permission; copyright (1995) Nature Publishing Group).

upstream DNA sequences to interact with RNA polymerase or the induction of DNA conformational changes in the downstream promoter that influence initiation.

One of the most striking examples of a transcriptional regulatory protein that bends DNA is the eukaryotic TATA-box binding protein (TBP). Eukaryotes have three RNA polymerases (I, II, and III) that transcribe nuclear genes. Polymerase II genes (which include those encoding proteins) have promoter sites with a so-called TATA box (consensus sequence is TATATAAN) centred at about −25. TBP binds to the TATA box and is involved in the initiation of transcription (67). High-resolution structures of TBP bound to DNA show that it induces a distinct bend in the DNA (68, 69). The bend is the consequence of the introduction of two sharp kinks at either end of the consensus sequence; between the kinks the duplex is smoothly curved and partially unwound (*Figure 6.9*; Chapter 1, Section 1.4). The significance of the DNA distortion induced by TBP is thought to be to enable subsequent assembly of other protein factors and to bring together DNA elements on either side of the bend.

Another example of a protein that induces DNA bends is the prokaryotic transcriptional activator FIS (factor for inversion stimulation). FIS is a small dimeric protein that stimulates both DNA inversion and activates transcription in *E. coli*. DNA inversions are site-specific recombination events that lead to a reversal of the DNA sequence between the recombining sites. Studies with the *tyrT* promoter have suggested that the binding of three FIS dimers forms

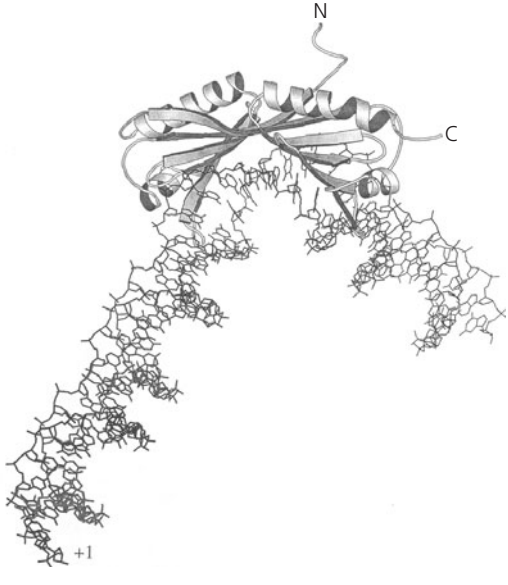


Figure 6.9 Interaction of TBP with DNA. The TBP2 protein of *Arabidopsis thaliana* complexed with DNA showing the induced DNA bend (the transcription start site is labelled +1) (reproduced from ref. 67 with permission; copyright (1996) Royal Society).

a nucleoprotein complex that stabilizes a short DNA loop, which facilitates the initiation of transcription (70). A loop in DNA stabilized by protein–protein interactions constitutes a discrete topological domain and, up to a point, can be treated in much the same way as a closed DNA circle (see Chapter 2). It is likely that the primary structural features of DNA that influence DNA curvature and flexibility (see Chapter 1, Section 1.4) will also influence the ability of DNA to form loops.

DNA looping is also a feature of the interaction of the *E. coli* AraC protein with DNA. The *ara* operon consists of three genes involved in the utilization of arabinose, transcribed from a single promoter. This promoter is controlled (both positively and negatively) by the co-operative binding of the AraC protein to two operator sites some 200 bp apart. Although the binding to one of the operators is tight, and to the other is weak, the two sites are normally occupied by the AraC protein (71). Moreover, insertion and deletion of small segments of DNA in the region between the two operators leads to cyclical changes in the repression of the *ara* promoter by AraC, with a period consistent with the helical repeat of DNA (72). These and other data are consistent with the co-operative binding of AraC at the two operators via 'looping' of the intervening DNA.[1] Other experiments have shown that loop formation by AraC can be promoted by DNA supercoiling. This is presumably due to the greater ease with which remote sites in plectonemically supercoiled DNA can be brought together, as in the NtrC example above.

DNA topology can influence transcription initiation in a variety of ways: affecting the interaction of RNA polymerase and transcription factors with DNA, and modulating the interaction between polymerase and transcription factors. It can act both globally, affecting the transcription of a large number of genes, or locally, influencing expression from a single promoter. We will now see how supercoiling plays a really fundamental role in the transcription process itself.

6.5 Transcription: the twin supercoiled domain model

The description of the organization of chromosomes in both eukaryotes and prokaryotes above (Section 6.2) seems to imply a reduction in the importance of supercoiling in the sense of free, torsionally strained DNA. Specifically we suggest that most negative supercoiling in eukaryotes is constrained by binding to histones in nucleosomes and their higher-order complexes, and

[1] Actually, the cyclical repression measured for AraC (a period of 11.1 bp) may reflect the surface-related helical repeat for a DNA loop wrapped on a virtual surface, rather than the intrinsic twist-related helical repeat of the DNA (see Chapter 3, Section 3.4).

that in prokaryotes (or *E. coli* at least), only ~40% of the supercoiling of isolated DNA is present *in vivo* as unconstrained plectonemic supercoiling. However, the situation *in vivo* is probably more complex than this. Several lines of evidence have suggested that supercoiling is a dynamic process that depends on more than the balance between opposing topoisomerase activities. For example, the inactivation of topoisomerase I in *E. coli* by genetic means leads to high negative supercoiling of plasmid pBR322, which is dependent on transcription from the strong promoter of the *tetA* (tetracycline resistance) gene on that plasmid (73, 74). In other work, inhibition of DNA gyrase *in vivo* by the drugs novobiocin and oxolinic acid was found to give rise to *positively* supercoiled pBR322 (75).

In 1987, Liu and Wang (76) speculated that these observations might reflect a property of the transcription process itself. As RNA polymerase transcribes a gene, the protein–RNA complex must follow the helical path of the DNA strands. The most obvious view of this has the DNA stationary and the polymerase rotating around the DNA axis. However, Liu and Wang suggested that the combination of the polymerase, the nascent RNA chain, and possibly even ribosomes translating the mRNA in a co-ordinated fashion, would form a complex so large that it would be unable to rotate around the DNA (*Figure 6.10*), and that instead, the DNA would rotate upon its axis. Such a rotation of the DNA around its axis, relative to the unpaired DNA region at the polymerase (*Figure 6.10b*), would tend to cause an increase in twist (positive supercoiling) ahead of the moving polymerase, and a reduction in twist (negative supercoiling) behind. This is exactly analogous to the accumulation of positive twist, and hence positive supercoiling, ahead of the replication fork, although in that case the strands are permanently separated, so there is no corresponding untwisting behind the fork (*Figure 6.10c*). Because these constitute two regions of supercoiling separated by the polymerase complex, this has been termed the 'twin supercoiled domain' model (77). The generality of these concepts suggests that this model will apply to both prokaryotes and eukaryotes. An alternative, but compatible, view envisages that the polymerase itself is anchored to a large and essentially immobile matrix within the cell, termed a 'transcription factory'. The DNA is then translocated through the stationary polymerase, rotating about its axis as it does so, with the results described above; the evidence for this view has been reviewed (12). These ideas concerning immobilized polymerases have been developed for transcription in eukaryotic nuclei, but could equally be applied to prokaryotic cells, and indeed for replication as well as transcription.

Since the DNA strands are not broken during transcription, an overall change in the linking number of a DNA is impossible by this mechanism

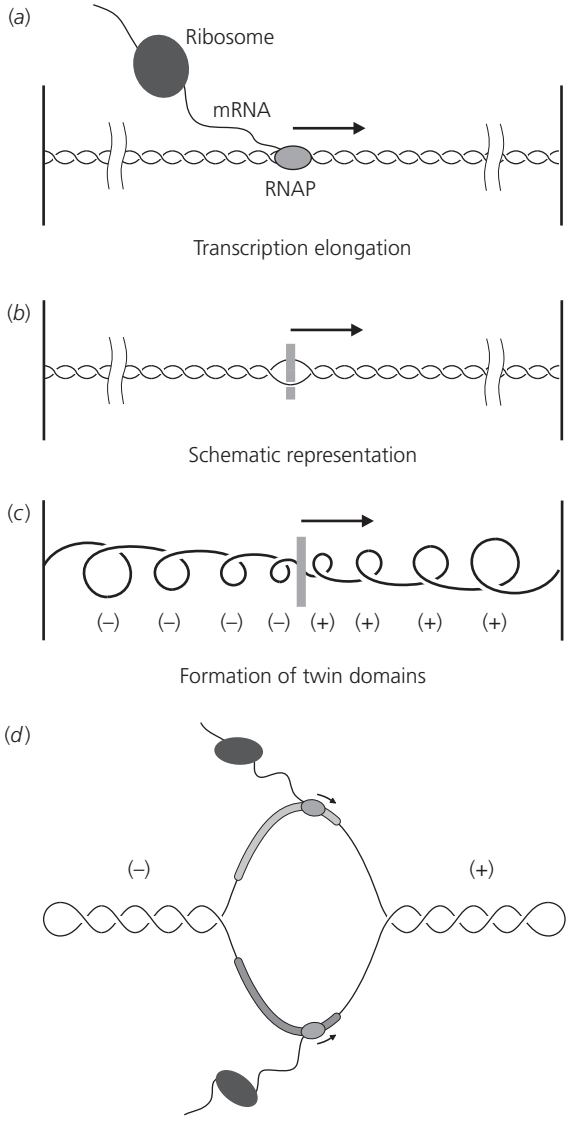


Figure 6.10 Formation of twin domains of supercoiling by transcription elongation. (*a*) RNA polymerase (RNAP) is shown transcribing a topologically isolated domain of DNA into mRNA, to which ribosomes become attached. This is shown schematically in (*b*) and leads to the formation of positive supercoiling ahead of the transcribing complex and negative supercoils behind (*c*). (*d*) Domains of supercoiling can also occur in plasmid DNA with two transcription complexes travelling in opposite directions (redrawn from refs 76, 77 with permission).

alone, the change in supercoiling (Tw + Wr) must be equal and opposite at each side of the polymerase. Of course, in a short linear molecule, such transient supercoiling would diffuse away, and with a single transcription unit on a circular plasmid, the positive and negative supercoiling could cancel out by diffusion around the circle. However, as we have seen, DNA may be organized into discrete domains that are anchored to a large matrix, and which may be

topologically independent. Such anchorage points would provide a block on the diffusion of transcription-generated supercoiling. Alternatively, transcription at a nearby gene proceeding in the opposite direction would also provide a blocking point (*Figure 6.10d*).

6.5.1 *The role of topoisomerases in transcription*

This alternative mechanism for the generation of supercoiling in DNA suggests a key role for DNA topoisomerases in cells. In prokaryotes, topoisomerase I can relax only negative supercoils, whereas gyrase can efficiently relax positive supercoils (Chapter 5), there is also a potential role for topoisomerase IV in the relaxation of both positive and negative supercoils (78). Hence it is likely that gyrase will act in front of and topoisomerase I behind the transcription complex to remove the transiently formed supercoils. If the activities of the two topoisomerases are significantly different, as is the case in the experiments quoted above where either *topA* (the gene for topo I) was deleted (73) or gyrase was inhibited by drugs (75), this would result in the build-up of either negative or positive supercoils in a closed-circular plasmid.

The likely operation of this mechanism for generating supercoils has been confirmed *in vivo* by experiments in *E. coli*, which demonstrate the requirement for transcription, translation, and signal sequence insertion and the importance of the relative orientation of transcribing genes (77). The same effects have been observed in yeast (79, 80); eukaryotic topoisomerases I and II can each relax both positive and negative supercoiling (see Chapter 5). The formation of twin-supercoiled domains has even been demonstrated *in vitro* using transcription from a single promoter on a closed-circular plasmid (81). These experiments showed that, in the presence of prokaryotic topoisomerase I, movement of RNA polymerase on the DNA template led to the rapid accumulation of unconstrained positive supercoils. Transient transcription-dependent increases in negative supercoiling have also been observed *in vivo* without inhibition of topoisomerases, suggesting that such effects occur in wild-type cells (82). The magnitude of these effects may be very great, particularly in the vicinity of actively transcribed genes, and it has been suggested that transiently high levels of supercoiling may have a role in regulatory processes. The idea, for example, that altered DNA conformations such as Z-DNA and cruciforms (see Chapter 1, Sections 1.2.4 and 1.3.1) may be important *in vivo* now seems more acceptable. Earlier estimations of the overall average levels of unconstrained supercoiling in cells were too low to be consistent with the formation of such structures (see Section 6.1.1; Chapter 2, Section 2.8.1). Support for the existence of Z-DNA in *E. coli* has come from the observation

that a recognition site is not a substrate for its specific methylase when the site is within a sequence with Z-DNA forming potential; the same site is a substrate when present in right-handed B-form DNA (83). Subsequently, similar experiments have suggested a role for transcription-induced supercoiling in the formation of Z-DNA *in vivo* (84).

6.6 **Recombination**

Recombination is the process of genetic exchange in DNA whereby chromosomes in both prokaryotes and eukaryotes are rearranged. It involves the breaking of phosphodiester bonds in DNA and the rejoining of the broken ends to new partners, as illustrated earlier in the discussion of Holliday junctions (Chapter 1, Section 1.3.1). Genetic recombination can be broadly divided into two types: general recombination and site-specific recombination. These are discussed separately below; the main points of relevance to this book are the topological aspects of site-specific recombination reactions.

6.6.1 *General recombination*

Generalized or homologous recombination is abundant in all species, and is the process responsible for generating genetic diversity; it is responsible for the crossing-over of chromosomes that occurs during meiosis in eukaryotes. It takes place between two DNA molecules that are 'homologous' (identical in sequence, or nearly so). The process requires breakage of the DNA backbones of the two partners, strand exchange and formation of product DNA molecules that are hybrids between the two parents. The intermediate formed in this process is the Holliday junction, which is illustrated in Chapter 1 (*Figure 1.5*). The key features of homologous recombination have been worked out in bacteria (principally *E. coli*) where the central role of the RecA protein has been worked out. RecA can catalyse DNA synapsis (pairing between different DNA molecules) by binding co-operatively to single-stranded DNA to form a nucleoprotein filament. Binding of a second double-stranded DNA molecule to this filament enables strand exchange between the two partners and the subsequent formation of heteroduplex DNA.

An important aspect of homologous recombination is that the branch point between the two parental DNA molecules can move (branch migrate), a process that is catalysed in *E. coli* by the RuvA and RuvB proteins, as shown in Chapter 1 (*Figure 1.6*). In eukaryotes an analogous process occurs and a homologue of RecA, Rad51, has been described in a number of

organisms including yeast, mice, and humans. As RecA and its homologues are known to unwind duplex DNA, negative supercoiling is expected to stimulate homologous recombination.

6.6.2 *Site-specific recombination*

In homologous recombination there is exchange between two DNA molecules that are very similar in DNA sequence. Site-specific recombination, by contrast, involves events in which recombination enzymes recognize short specific sequences on one or both of the recombining molecules; the molecules themselves may have very different sequences (85). Transpositional site-specific recombination involves the insertion of mobile genetic elements into essentially any DNA sequence. Here an enzyme called a transposase acts on specific DNA sequences at the ends of a transposon, a mobile genetic element, which is inserted, essentially randomly, into another site on DNA. There are a large number of such elements known, these include bacteriophage Mu, bacterial transposons (e.g. Tn*3*), retroviruses (including HIV, the AIDS virus), and retroviral-like retrotransposons such as the yeast Ty1 element. The recombination reaction carried out by the Tn*3* family of transposons is illustrated diagrammatically in *Figure 6.11*. This reaction occurs in two steps, the transpositional site-specific recombination step forming a co-integrate, containing two copies of the transposon, and the resolution reaction (catalysed by resolvase, see later) separating the donor and acceptor DNA molecules. Detailed discussion of transposable elements is outside the scope of this book, but the interested reader is directed to a recent book by Craig *et al.* for more detailed information (86). However, the resolution reaction, about which rather more is known, is discussed later (Section 6.6.2.2).

Rather than discuss the topological aspects of transpositional site-specific recombination, we will instead focus on conservative site-specific recombination (i.e. those where the target site is specified). There are many conservative site-specific recombination systems for which the proteins and the DNA partners have been well defined, such that the topological issues are well described. These reactions involve interaction between specific DNA sequences in both DNA partners. These sequences are aligned and recombination proteins bring about the exchange of DNA strands. The importance of DNA topology in these processes cannot be overemphasized, both in terms of facilitating the recombination reaction and in terms of understanding the recombination mechanism by analysis of the topology of the products. Two types of conservative site-specific recombination reactions have been described: the integrase-type and the resolvase-type (85). These two types are distinguished by the

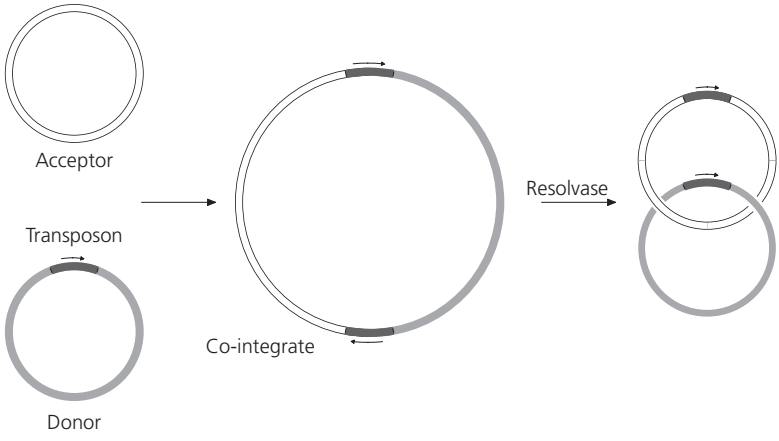


Figure 6.11 The recombination reaction carried out by the Tn*3* family of transposons. A donor circle carrying a transposon of the Tn*3* family forms a co-integrate with the acceptor DNA circle. The co-integrate carries two copies of the transposon. The action of resolvase catalyses a recombination reaction between *res* sites within the two transposons to yield the catenated product circles, each carrying a single copy of the transposon.

mechanisms of their recombination reactions. For example, integrase-type reactions involve a covalent link between 3′-phosphates in DNA and tyrosine residues in the protein, whereas resolvase-type reactions involve covalent links between 5′-phosphates and serine residues. The next sections describe examples of well-studied types of recombination reactions from each group, highlighting the aspects relevant to DNA topology.

6.6.2.1 *Integrase-type recombination*

The integrase family of recombinases comprises a large number of family members, which include bacteriophage λ Int (the founder member), Cre recombinase of bacteriophage P1, the FLP recombinase encoded by the yeast (*Sacch. cerevisiae*) 2 μm plasmid and the XerCD system of *E. coli* (87). We will exemplify the recombination reactions catalysed by these proteins using λ Int and XerCD. As illustrated in *Figure 6.12a*, the bacteriophage λ Int protein catalyses a recombination reaction between a site on the phage DNA (*att*P) and a site on the host DNA (*att*B), which leads to the integration of λ DNA into the *E. coli* chromosome. The reverse reaction, excision, occurs by a recombination event between *att*L and *att*R (*Figure 6.12a*) and shows different characteristics to the integration reaction. The λ integration reaction has been studied extensively *in vitro*, generally using artificial DNA substrates, such

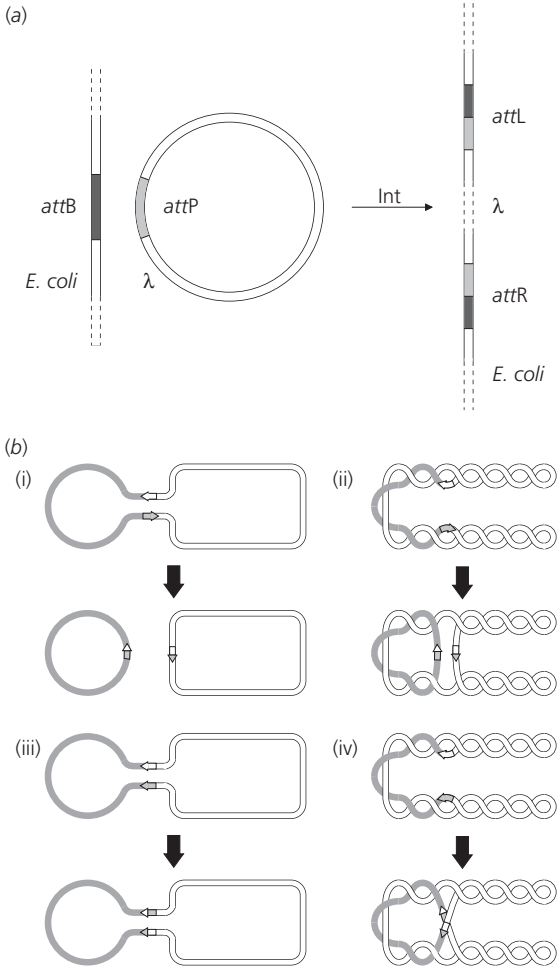


Figure 6.12 Site-specific recombination by the bacteriophage λ Int protein. (*a*) Bacteriophage λ DNA is integrated into the *E. coli* chromosome via a recombination event between the *att*P site on λ DNA and the *att*B site on *E. coli* DNA, mediated by Int and other proteins. The integrated phage (the prophage) is now flanked by two hybrid sites, *att*L and *att*R. (*b*) Recombination between *att* sites (shown as arrows) in direct repeat on the same circle generates two product circles, (i), which, if the substrate DNA was plectonemically supercoiled, will be catenated, (ii). When *att* sites are in inverted repeat (iii), the product is a single circle, which will be knotted if the substrate was plectonemically supercoiled (iv) ((*b*) redrawn from ref. 107).

as those shown in *Figure 6.12b*. DNA supercoiling has been shown to have a profound effect on the λ integration reaction; only negatively supercoiled closed-circular DNA molecules bearing the *att*P site are effective substrates for recombination. Supercoiling at *att*B is not necessary, and supercoiling is also not required for excision. The requirement for supercoiling of the phage DNA is thought to relate to the folding of *att*P into a higher-order structure in which the DNA is wrapped on itself in a negative writhe. In fact DNA gyrase, the bacterial supercoiling enzyme (see Chapter 5), was discovered during studies of λ integration as a host-encoded factor required for efficient integration (88). The higher-order structure formed during the λ integration

reaction has been termed the 'intasome', by analogy with the nucleosome (see Chapter 3, Section 3.5), and is made up of *att*P complexed with Int protein and IHF (integration host factor; see Section 6.2.1). This DNA–protein complex constitutes a binding site for *att*B, which probably binds to the complex as a naked piece of DNA.

The DNA cleavage and rejoining reactions carried out by Int are accomplished in two steps. First, a tyrosine hydroxyl attacks the DNA at the scissile phosphate, nicking the DNA and forming a 3′-phosphotyrosine intermediate (analogous to type IB topoisomerases, see Chapter 5, Section 5.3.1). This covalent protein–DNA complex is then attacked by the 5′-hydroxyl from the other DNA molecule, which displaces the protein, forming the familiar Holliday junction. Repetition of this reaction for the other strand of each DNA partner will then generate the recombinant DNA molecules. It is the formation of the DNA–protein covalent bonds that conserves the energy of the broken bonds and allows the reaction to proceed without the requirement of a high-energy co-factor. Int itself is a ~40 kDa protein, which contains discrete domains; the crystal structure of the catalytic domain has given some insight into the mechanism of action of this protein (89).

DNA substrates that carry both the *att*P and *att*B sites (e.g. *Figure 6.12b*) generate knotted or catenated products, depending on the relative orientation of the *att*P and *att*B sites. Double-stranded circular DNA substrates containing *att*P and *att*B in the same orientation generate catenanes, while those containing *att* sites in inverted orientation generate knots (87). Analysis of the knotted and catenated products has suggested possible models for the bacteriophage λ integration reaction. For example, the complexity of the knots formed by recombination increases with increasing specific linking difference of the substrate. This is due to the trapping of writhes (or nodes) in interwound DNA (see Chapter 2, Section 2.4.3) between the interacting recombination sites. Non-supercoiled (nicked-circular) DNA molecules can, under certain conditions, constitute substrates for recombination and, at low frequency, generate both simple (unknotted) and knotted products. Increasing the distance between the *att* sites in supercoiled DNA substrates also increases the complexity of the knots produced. It has been concluded from these (and other) data that synapsis between *att* sites occurs essentially by random collision, which leads to the trapping of supercoils (writhes) between the two *att* sites, which will then be converted to knot or catenane nodes following recombination (*Figure 6.12b*). The knotted products of recombination between *att* sites on non-supercoiled DNA can be explained by formation of the intasome, which may lead to the trapping of a writhe in the DNA between the *att* sites and hence to knotted products.

The random collision mode of synapsis exhibited by Int recombination is in contrast to the behaviour of a related recombinase, Cre, from bacteriophage P1. Cre carries out recombination between two 34 bp sequences, known as *loxP*. In experiments using substrates with two *loxP* sites, analogous to those above, the predominant products are free (uncatenated) or unknotted circles (90), particularly *in vivo* (91). This suggests that synapsis can occur by an ordered process such as slithering (see Section 6.1.1) that excludes the trapping of any nodes between the interacting sites.

The bacteriophage λ integration reaction has been used as a probe of the conformation of supercoiled DNA *in vivo*. For example, it has been concluded that the writhing of DNA in *E. coli* is plectonemic rather than toroidal (see Chapter 2, Section 2.4.3). Toroidal supercoiling would not be expected to generate knotted and catenated products of the complexity found in the bacteriophage λ integration reaction (92). Furthermore, analysis of the complexity of the catenated products of Int-mediated recombination *in vivo*, has suggested that only 40% of the linking deficit of plasmid DNA in *E. coli* is in the form of interwound (plectonemic) supercoils (10). The remainder is thought to be constrained by protein binding (see Section 6.2.1).

The Xer site-specific recombination system from *E. coli* is involved in ensuring that circular plasmids and chromosomes are monomeric prior to segregation at cell division (93). This recombination system has a number of important similarities and differences compared to the λ Int system. Like Int it uses an active-site tyrosine to attack the scissile phosphates on DNA and forms a Holliday junction intermediate. Unlike Int it uses two recombination proteins, XerC and XerD, which are closely related in sequence, and the outcome of the recombination reaction is different depending on whether it occurs on plasmid or chromosomal recombination sites. It is thought that each of the Xer proteins catalyses one pair of strand exchanges potentially allowing more control over the production and resolution of the Holliday intermediate. Recombination on the chromosome occurs at a 32 bp sequence called *dif*, located at the replication terminus of *E. coli*. Recombination in plasmid ColE1 occurs at *cer* sites (∼210 bp) and requires the accessory proteins ArgR and PepA; recombination in plasmid pSC101 occurs at *psi* sites (∼190 bp) and requires only PepA (94). Plasmid recombination by Xer is preferentially intramolecular whereas chromosomal recombination can be intra- or intermolecular. Insight into the mode of interaction of the recombinases with DNA has come from crystal structure information on XerD (95).

Analysis of the topology of the products of Xer-mediated recombination reaction *in vitro* has given important insight into the mechanism of the recombination reaction. For example, it has been shown that the product

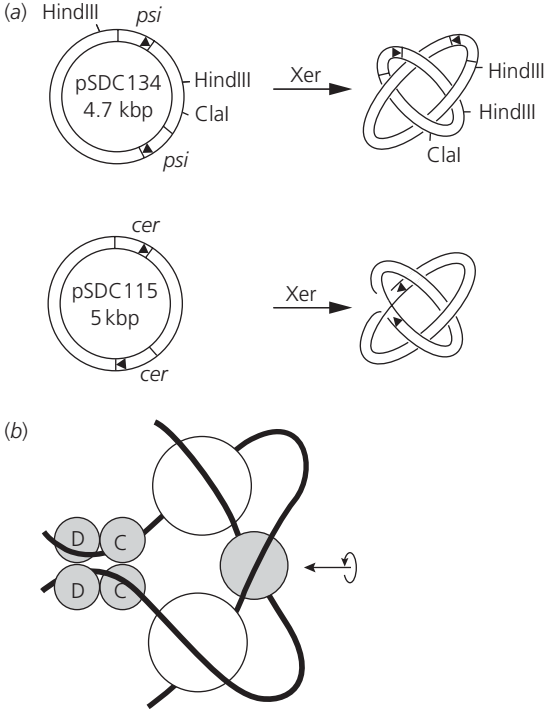


Figure 6.13 Recombination between directly-repeated sites by XerCD. (*a*) Products from recombination on plasmids with *psi* and *cer* sites. (*b*) Model for a synaptic complex generated by XerCD. The shaded circle is an ArgR hexamer; open circles represent PepA hexamers; C and D are XerC and XerD, respectively (reproduced from ref. 96).

of recombination between directly repeated *psi* sites is a right-handed antiparallel four-noded catenane; the product of the same reaction on *cer* sites has the same topology, except that only one pair of strands is exchanged (96) (*Figure 6.13*). These results suggest that the synaptic complex and strand exchange mechanism have a fixed topology and leads to a model for the synaptic complex showing how DNA wrapping around the proteins could trap three negative supercoils; this model is similar to that described for resolvase (see below).

6.6.2.2 *Resolvase-type recombination*

The resolvases constitute the second type of conservative site-specific recombinases. A well-studied system is the recombination reaction catalysed by the resolvase proteins from the transposons of the Tn*3* family, which include Tn*3* itself, $\gamma\delta$ (Tn*1000*) and Tn*21* (97, 98). Transposons are segments of DNA that can apparently 'jump' from one molecule to another. In the case of Tn*3* this process occurs in two stages. The first involves the fusion of the donor and recipient DNA molecules to yield a co-integrate bearing two copies of the

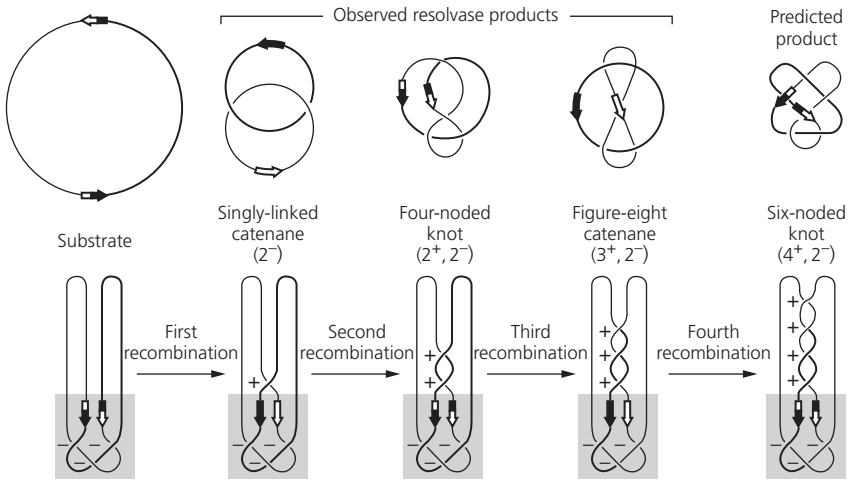


Figure 6.14 Site-specific recombination by resolvase. Scheme for the formation of multiple knotted and catenated products by Tn*3* resolvase. Successive rounds of recombination generate the products shown in the upper row. These products can be rationalized by proposing a synaptic process involving three nodes (lower row) (redrawn from ref. 100 with permission; copyright (1985) AAAS).

transposon, one at each junction of the donor and recipient DNA (*Figures 6.11* and *6.14*), known as transpositional site-specific recombination (see above). This co-integrate is then resolved into the product DNA molecules, each bearing a single copy of the transposon, by the action of the resolvase protein, which catalyses a site-specific recombination reaction between the *res* sites in each transposon. The product DNA molecules are catenated (*Figure 6.14*), and may subsequently be decatenated by the action of a topoisomerase (see Chapter 5).

The site-specific recombination reaction catalysed by resolvase has been studied extensively *in vitro* using DNA molecules carrying pairs of *res* sites. Such studies have shown that, unlike bacteriophage λ Int, resolvase will only efficiently recombine substrates bearing directly repeated *res* sites (*Figure 6.14*). Moreover, the major product of the resolvase recombination reaction is a singly linked catenane, regardless of the specific linking difference of the DNA substrate (*Figure 6.14*). The simplicity of this product implies a rigidly controlled mechanism of synapsis and not simply random collision as has been suggested for members of the integrase family (see above). Two resolvase-bound *res* sites must interact and form an organized complex, known as a 'synaptosome'. Topological analysis of recombination products has given crucial insight into the mechanism of the resolvase reaction. Early work showed that the resolvase product were catenated circles linked only once with two negative nodes

(Chapter 4, Section 4.3.2). Careful analysis of minor products of resolvase-mediated recombination has revealed that a range of knotted and catenated products is present at low frequency (99, 100) (*Figure 6.14*); these products can be explained by successive rounds of recombination. It is important to note that only the products with the node assignation given in *Figure 6.14* were detected, again suggesting a rigidly controlled synaptic process (e.g. there are eight possible six-noded knots, but only the one shown in *Figure 6.14* has been detected). It has been proposed that the resolvase-mediated synapsis of *res* sites involves the stabilization of three negative supercoils (i.e. three negative nodes) as shown in the lower panel of *Figure 6.14*. Indeed, measuring ΔLk for the synaptic complex involving resolvase and two *res* sites gives a value of -3 consistent with this proposal (101).

Significant insight into the mode of interaction of resolvases with DNA has come from the crystal structure of the $\gamma\delta$ resolvase complexed with a 34 bp sequence representing an artificially symmetrical version of the $\gamma\delta$ *res* site (102). This structure shows that the DNA is sharply bent by 60° towards the major groove and away from the protein. In the structure the active site serines are too far away from the scissile phosphates to be in an active configuration, possibly suggesting significant protein movement during the catalytic cycle. However, it must be borne in mind that the full synaptosome will also include sites I and II and a second resolvase-bound *res* site.

The requirement for directly repeated *res* sites in the resolvase recombination reaction has prompted much speculation as to the mechanism of bringing together the two sites. Random collision, suggested for the integrases, would seem to be inconsistent with the requirement for sites to be in directly repeated orientation, and with the observed simplicity of the recombination products. One model, known as tracking, involves the binding of the resolvase protein to one *res* site via one domain, and another domain of the protein sliding along the DNA until the other *res* site is encountered (see Chapter 4, Section 4.5). Several lines of experimentation now argue against this type of mechanism (reviewed in (6)). One test of this model involved a catenated substrate in which a DNA circle with two co-linear *res* sites was catenated with several circles lacking *res* sites (*Figure 6.15*) (103). If tracking was correct, the threading of DNA through resolvase would result in the non-*res* circles being excluded from one of the product circles (*Figure 6.15*). The result found was a random distribution of catenanes, arguing against a tracking model. An alternative explanation of the orientation requirement is that the *res* sites on supercoiled DNA find each other by slithering (Section 6.1.1) whereby the correct synaptic structure can only be assembled on negatively supercoiled DNA with the *res* sites in directly repeated orientation. Although slithering may be a significant

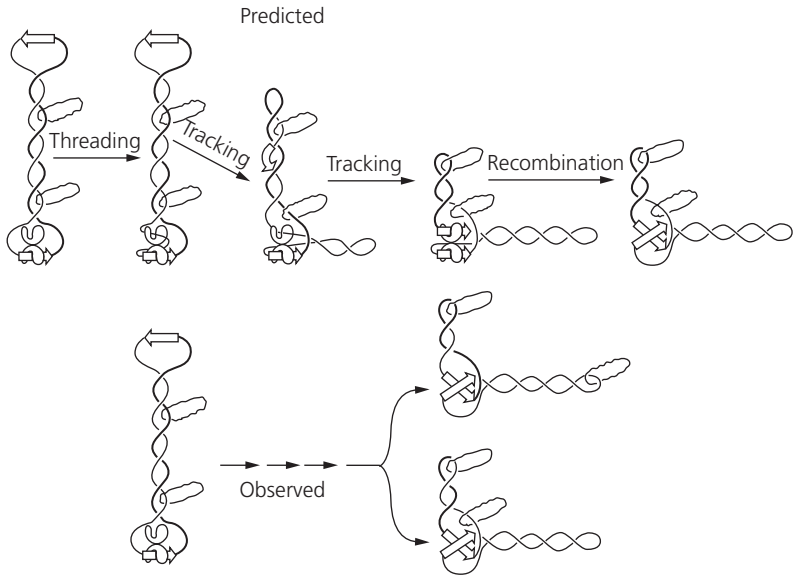


Figure 6.15 Experimental test for tracking. The substrate is a DNA circle, containing co-linear *res* sites, catenated with two reporter DNA circles. Tracking predicts exclusion of the rings from one of the product circles; the observed result shows that the reporter circles can reside on both products (reproduced from ref. 6 with permission; copyright (1987) Nature Publishing Group).

process in synapsis it is not absolutely required as recombination can occur between *res* sites on relaxed or linear DNA substrates, which are not expected to 'slither'. A model that seems to accommodate the current data on synaptosome assembly is called the 'topological filter' or 'two-step synapsis' model (104, 105). Here the two *res* sites can find each other using either random collision or slithering to form a mixture of complexes of varying topologies. Co-linear *res* sites readily form the synaptic complex, trapping 3 negative nodes, but other situations lead to topological complexities such that they are strongly disfavoured.

6.7 **Conclusions**

The examples given in this chapter demonstrate that DNA topology is of fundamental importance for a wide range of biological processes. Indeed, virtually every reaction involving DNA is influenced by DNA topology, or has topological effects. There is a tendency to think of DNA merely as a code

(G,A,T,C, etc.) that functions like 'ticker-tape' to be translated into RNA and ultimately protein. While this has some merit from a genetical perspective, it is important also to be aware of the higher-order structural features of DNA and how these have a profound effect on its function.

This book has progressed from the somewhat abstract concepts of linking number, twist, and writhe, through to their ultimate biological consequences. It is hoped that an appreciation of some of the finer points of DNA topology will provide greater insight into the biological functions of DNA. If you have read the book from cover to cover (well done if you have) you should better appreciate the topological aspects of DNA and their biological consequences. However, do not expect to have it all sorted. Some of the concepts are quite tricky and you will probably need to dip back into Chapter 2 from time to time to refresh yourself on the basics. And, do not forget, a piece of rubber tubing hanging on the back of your door will be an invaluable aid to getting to grips with future topological problems.

6.8 Further Reading

Baumberg, S. (ed.) (1999). *Prokaryotic gene expression.* Oxford University Press.

Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (2002). *Mobile DNA II.* ASM Press, Washington, DC.

Hansen, J.C. (2002). Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 361–392.

Hoppert, M. and Mayer, F. (1999). Principles of macromolecular organization and cell function in bacteria and archaea. *Cell Biochem. Biophys.* **31**, 247–284.

Horn, P.J. and Peterson, C.L. (2002). Molecular biology. Chromatin higher order folding—wrapping up transcription. *Science* **297**, 1824–1827.

Kahn, J.D. (2002). DNA topology: application to gene expression. *Curr. Org. Chem.* **6**, 815–826.

Lilley, D.M.J., Chen, D., and Bowater, R.P. (1996). DNA supercoiling and transcription: topological coupling of promoters. *Quart. Rev. Biophys.* **29**, 203–225.

Pettijohn, D.E. (1999). The nucleoid. In *Escherichia coli and Salmonella.* Neidhardt, F.C. (ed.), ASM Press, Washington, DC, Chapter 12.

Ptashne, M. (1992). *A genetic switch: phage λ and higher organisms.* Cell Press & Blackwell Science, Cambridge, MA, USA.

Schvartzman, J.B. and Stasiak, A. (2004). A topological view of the replicon. *EMBO Rep.* **5**, 256–261.

Wagner, R. (2000). *Transcription regulation in prokaryotes.* Oxford University Press.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3**, 430–440.

Wang, J.C. and Lynch, A.S. (1993). Transcription and DNA supercoiling. *Curr. Opin. Genet. Dev.* **3**, 764–768.

Zlatanova, J., Leuba, S.H., and van Holde, K. (1999). Chromatin structure revisited. *Crit. Rev. Eukaryot. Gene Expr.* **9**, 245–255.

6.9 References

1. Gasser, S.M. and Laemmli, U.K. (1987). A glimpse at chromosoma order. *Trends Genet.* **3**, 16–21.
2. Kramer, P.R. and Sinden, R.R. (1997). Measurement of unrestrained negative supercoiling and topological domain size in living human cells. *Biochemistry* **36**, 3151–3158.
3. Lopez-Garcia, P. and Forterre, P. (1997). DNA topology in hyperthermophilic archaea: reference states and their variation with growth phase, growth temperature, and temperature stresses. *Mol. Microbiol.* **23**, 1267–1279.
4. Forterre, P., Bergerat, A., and LopezGarcia, P. (1996). The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* **18**, 237–248.
5. Boles, T.C., White, J.H., and Cozzarelli, N.R. (1990). Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* **213**, 931–951.
6. Gellert, M. and Nash, H. (1987). Communication between segments of DNA during site-specific recombination. *Nature* **325**, 401–404.
7. Vologodskii, A. and Cozzarelli, N.R. (1996). Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophys. J.* **70**, 2548–2556.
8. Worcel, A. and Burgi, E. (1972). On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* **71**, 127–147.
9. Pettijohn, D.E. (1999). The nucleoid. In *Escherichia coli and Salmonella*. Neidhardt, F.C. (ed.), ASM Press, Washington, DC, Chapter 12.
10. Bliska, J.B. and Cozzarelli, N.R. (1987). Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. *J. Mol. Biol.* **194**, 205–218.
11. Sinden, R.R. and Pettijohn, D.E. (1981). Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc. Natl. Acad. Sci. USA* **78**, 224–228.
12. Cook, P.R. (1999). The organization of replication and transcription. *Science* **284**, 1790–1795.
13. Dorman, C.J., Hinton, J.C., and Free, A. (1999). Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria. *Trends Microbiol.* **7**, 124–128.
14. McLeod, S.M. and Johnson, R.C. (2001). Control of transcription by nucleoid proteins. *Curr. Opin. Microbiol.* **4**, 152–159.
15. Shishido, K., Komiyama, N., and Ikawa, S. (1987). Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants. *J. Mol. Biol.* **195**, 215–218.

16. Liu, L.F., Perkocha, L., Calendar, R., and Wang, J.C. (1981). Knotted DNA from bacteriophage capsids. *Proc. Natl. Acad. Sci. USA* **78**, 5498–5502.
17. Liu, L.F. and Davis, J.L. (1981). Novel topologically knotted DNA from bacteriophage P4 capsids: studies with topoisomerases. *Nucleic Acids Res.* **9**, 3979–3989.
18. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
19. Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.* **319**, 1097–1113.
20. Woodcock, C.L. and Horowitz, R.A. (1995). Chromatin organization reviewed. *Trends Cell Biol.* **5**, 272–277.
21. van Holde, K. and Zlatanova, J. (1995). Chromatin higher order structure: chasing a mirage? *J. Biol. Chem.* **270**, 8373–8376.
22. Widom, J. (1998). Structure, dynamics, and function of chromatin in vitro. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 285–327.
23. Wu, J. and Grunstein, M. (2000). 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* **25**, 619–623.
24. Kornberg, R.D. and Lorch, Y. (1999). Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* **9**, 148–151.
25. Cairns, J. (1963). The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.* **6**, 208–213.
26. Champoux, J.J. and Been, M.D. (1980). Topoisomerases and the swivel problem. In *Mechanistic studies of DNA replication and genetic recombination*: ICN-UCLA Symposia on Molecular and Cellular Biology. Alberts, B. (ed.). Academic Press, New York, pp. 809–815.
27. Benkovic, S.J., Valentine, A.M., and Salinas, F. (2001). Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181–208.
28. Bell, S.P. and Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374.
29. Fuller, R.S. and Kornberg, A. (1983). Purified *DnaA* protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. USA* **80**, 5817–5821.
30. Schnos, M., Zahn, K., Inman, R.B., and Blattner, F.R. (1988). Initiation protein induced helix destabilization at the lambda origin: a prepriming step in DNA replication. *Cell* **52**, 385–395.
31. Umek, R.M. and Kowalski, D. (1988). The ease of DNA unwinding as a determinant of initiation at yeast replication origins. *Cell* **52**, 559–567.
32. Smelkova, N. and Marians, K.J. (2001). Timely release of both replication forks from oriC requires modulation of origin topology. *J. Biol. Chem.* **276**, 39186–39191.
33. Khodursky, A.B., Peter, B.J., Schmid, M.B., DeRisi, J., Botstein, D., Brown, P.O., and Cozzarelli, N.R. (2000). Analysis of topoisomerase function in bacterial

replication fork movement: use of DNA microarrays. *Proc. Natl. Acad. Sci. USA* **97**, 9419–9424.
34. Hiasa, H. and Marians, K.J. (1994). Topoisomerase III, but not topoisomerase I, can support nascent chain elongation during theta-type DNA replication. *J. Biol. Chem.* **269**, 32655–32659.
35. Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D., and Cozzarelli, N.R. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc. Natl. Acad. Sci. USA* **98**, 8219–8226.
36. Nurse, P., Levine, C., Hassing, H., and Marians, K.J. (2003). Topoisomerase III can serve as the cellular decatenase in *Escherichia coli*. *J. Biol. Chem.* **278**, 8653–8660.
37. Lucas, I., Germe, T., Chevrier-Miller, M., and Hyrien, O. (2001). Topoisomerase II can unlink replicating DNA by precatenane removal. *EMBO J.* **20**, 6509–6519.
38. Kim, R.A. and Wang, J.C. (1989). Function of DNA topoisomerases as replication swivels in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **208**, 257–267.
39. Wasserman, S.A. and Cozzarelli, N.R. (1986). Biochemical topology: applications to DNA recombination and replication. *Science* **232**, 951–960.
40. Steck, T.R. and Drlica, K. (1984). Bacterial chromosome segregation: evidence for DNA gyrase involvement in decatenation. *Cell* **36**, 1081–1088.
41. Adams, D.E., Shekhtman, E.M., Zechiedrich, E.L., Schmid, M.B., and Cozzarelli, N.R. (1992). The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* **71**, 277–288.
42. Zechiedrich, E.L. and Cozzarelli, N.R. (1995). Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in *Escherichia coli*. *Genes Dev.* **9**, 2859–2869.
43. Laurie, B., Katritch, V., Sogo, J., Koller, T., Dubochet, J., and Stasiak, A. (1998). Geometry and physics of catenanes applied to the study of DNA replication. *Biophys. J.* **74**, 2815–2822.
44. DiNardo, S., Voelkel, K., and Sternglanz, R. (1984). DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proc. Natl. Acad. Sci. USA* **81**, 2616–2620.
45. Yanagida, M. and Wang, J.C. (1987). Yeast DNA topoisomerases and their structural genes. In *Nucleic Acids and Molecular Biology*, vol. 1. Eckstein, F. and Lilley, D.M.J. (eds.), Springer Verlag, Heidelberg, Berlin, pp. 196–209.
46. Lloyd, G., Landini, P., and Busby, S. (2001). Activation and repression of transcription initiation in bacteria. *Essays Biochem.* **37**, 17–31.
47. deHaseth, P.L., Zupancic, M.L., and Record, M.T., Jr. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.* **180**, 3019–3025.
48. Orphanides, G. and Reinberg, D. (2002). A unified theory of gene expression. *Cell* **108**, 439–451.

49. Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876.
50. Young, B.A., Gruber, T.M., and Gross, C.A. (2002). Views of transcription initiation. *Cell* **109**, 417–420.
51. Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassylyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **420**, 43–50.
52. Lilley, D.M.J., Chen, D., and Bowater, R.P. (1996). DNA supercoiling and transcription: topological coupling of promoters. *Quart. Rev. Biophys.* **29**, 203–225.
53. Amouyal, M. and Buc, H. (1987). Topological unwinding of strong and weak promoters by RNA polymerase: a comparison between the *lac* wild-type and the UV5 sites of *Escherichia coli*. *J. Mol. Biol.* **195**, 795–808.
54. Ebright, R.H. (1998). RNA polymerase–DNA interaction: structures of intermediate, open, and elongation complexes. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 11–20.
55. Steck, T.R., Franco, R.J., Wang, J.-Y., and Drlica, K. (1993). Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. *Mol. Microbiol.* **10**, 473–481.
56. Dorman, C.J. (2002). DNA topology and regulation of bacterial gene expression. In *SGM symposium 61: signals, switches, regulons and cascades: control of bacterial gene expression.* Hodgson, D.A. and Thomas, C.M. (eds.), pp. 41–56.
57. Menzel, R. and Gellert, M. (1994). The biochemistry and biology of DNA gyrase. *Adv. Pharmacol.* **29A**, 39–69.
58. Drlica, K., Wu, E.-D., Chen, R.-R., Wang, J.-Y., Zhao, X., Xu, C., Malik, M., Kayman, S., and Friedman, S.M. (1999). Prokaryotic DNA topology and gene expression. In *Prokaryotic Gene Expression.* Baumberg, S. (ed.), Oxford University Press, Oxford.
59. Hatfield, G.W. and Benham, C.J. (2002). DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.* **36**, 175–203.
60. Dorman, C.J. (1996). Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. *Trends Microbiol.* **4**, 214–216.
61. Luisi, B. (1995). DNA-protein interactions at high resolution. In *DNA–protein: structural interactions.* Lilley, D.M.J. (ed.), Oxford University Press, Oxford, pp. 1–48.
62. Burley, S.K. (1998). X-ray crystallographic studies of eukaryotic transcription factors. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 33–40.
63. Rhodius, V.A. and Busby, S.J. (1998). Positive activation of gene expression. *Curr. Opin. Microbiol.* **1**, 152–159.
64. Liu, Y., Bondarenko, V., Ninfa, A., and Studitsky, V.M. (2001). DNA supercoiling allows enhancer action over a large distance. *Proc. Natl. Acad. Sci. USA* **98**, 14883–14888.

65. Ansari, A., Bradner, J.E., and O'Halloran, T.V. (1995). DNA-bend modulation in a repressor-to-activator switching mechanism. *Nature* **374**, 371–375.
66. Koudelka, G.B. (1998). Recognition of DNA structure by 434 repressor. *Nucleic Acids Res.* **26**, 669–675.
67. Burley, S.K. (1996). X-ray crystallographic studies of eukaryotic transcription initiation factors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **351**, 483–489.
68. Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527.
69. Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.
70. Travers, A. and Muskhelishvili, G. (1998). DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J. Mol. Biol.* **279**, 1027–1043.
71. Martin, K., Huo, L., and Schleif, R.F. (1986). The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Natl. Acad. Sci. USA* **83**, 3654–3658.
72. Dunn, T.M., Hahn, S., Ogden, S., and Schleif, R.F. (1984). An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. USA* **81**, 5017–5020.
73. Pruss, G.J. (1985). DNA topoisomerase I mutants. Increased heterogenity in linking number and other replicon-dependent changes in DNA supercoiling. *J. Mol. Biol.* **185**, 51–63.
74. Pruss, G.J. and Drlica, K. (1986). Topoisomerase I mutants: the gene on pBR322 that encodes resistance to tetracycline affects plasmid DNA supercoiling. *Proc. Natl. Acad. Sci. USA* **83**, 8952–8956.
75. Lockshon, D. and Morris, D.R. (1983). Positively supercoiled plasmid DNA is produced by treatment of *Escherichia coli* with DNA gyrase inhibitors. *Nucleic Acids Res.* **11**, 2999–3017.
76. Liu, L.F. and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027.
77. Wu, H.-Y., Shyy, S., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53**, 433–440.
78. Zechiedrich, E.L., Khodursky, A.B., Bachellier, S., Schneider, R., Chen, D., Lilley, D.M.J., and Cozzarelli, N.R. (2000). Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli. J. Biol. Chem.* **275**, 8103–8113.
79. Brill, S.J. and Sternglanz, R. (1988). Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. *Cell* **54**, 403–411.
80. Giaever, G.N. and Wang, J.C. (1988). Supercoiling of intracellular DNA can occur in eukaryotic cells. *Cell* **55**, 849–856.

81. Tsao, Y.-P., Wu, H.-Y., and Liu, L.F. (1989). Transcription-driven supercoiling of DNA: direct biochemical evidence from *in vitro* studies. *Cell* **56**, 111–118.
82. Figueroa, N. and Bossi, L. (1988). Transcription induces gyration of the DNA template in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**, 9416–9420.
83. Jaworski, A., Hsieh, W.-T., Blaho, J.A., Larson, J.E., and Wells, R.D. (1987). Left-handed DNA *in vivo*. *Science* **238**, 773–777.
84. Rahmouni, A.R. and Wells, R.D. (1992). Direct evidence for the effect of transcription on local DNA supercoiling *in vivo*. *J. Mol. Biol.* **223**, 131–144.
85. Stark, W.M., Boocock, M.R., and Sherratt, D.J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439.
86. Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (2002). *Mobile DNA II*. ASM Press, Washington, DC.
87. Azaro, M.A. and Landy, A. (2002). λ integrase and the λ Int family. In *Mobile genetic elements II*. Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (eds.), ASM Press, Washington, DC, pp. 118–148.
88. Gellert, M., Mizuuchi, K., O'Dea, M.H., and Nash, H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc. Natl. Acad. Sci. USA* **73**, 3872–3876.
89. Kwon, H.J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**, 126–131.
90. Abremski, K., Hoess, R., and Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* **32**, 1301–1311.
91. Adams, D.E., Bliska, J.B., and Cozzarelli, N.R. (1992). Cre-lox recombination in *Escherichia coli* cells—mechanistic differences from the *in vitro* reaction. *J. Mol. Biol.* **226**, 661–673.
92. Spengler, S.J., Stasiak, A., and Cozzarelli, N.R. (1985). The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure. *Cell* **42**, 325–334.
93. Sherratt, D.J., Arciszewska, L.K., Blakely, G., Colloms, S., Grant, K., Leslie, N., and McCulloch, R. (1995). Site-specific recombination and circular chromosome segregation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **347**, 37–42.
94. Barre, F.-X. and Sherratt, D.J. (2002). Xer site-specific recombination: promoting chromosome segregation. In *Mobile genetic elements II*. Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (eds.), ASM Press, Washington, DC, pp. 149–161.
95. Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J., and Wigley, D.B. (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178–5187.
96. Colloms, S.D., Bath, J., and Sherratt, D.J. (1997). Topological selectivity in Xer site-specific recombination. *Cell* **88**, 855–864.

97. Oram, M., Szczelkun, M.D., and Halford, S.E. (1995). Recombination. Pieces of the site-specific recombination puzzle. *Curr. Biol.* **5**, 1106–1109.
98. Grindley, N.D.F. (2002). The movement of Tn*3*-like elements: transposition and cointegrate resolution. In *Mobile genetic elements II.* Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M., (eds.) ASM Press,Washington, DC, pp. 272–302.
99. Krasnow, M.A., Stasiak, A., Spengler, S.J., Dean, F., Koller, T., and Cozzarelli, N.R. (1983). Determination of the absolute handedness of knots and catenanes of DNA. *Nature* **304**, 559–560.
100. Wasserman, S.A., Dungan, J.M., and Cozzarelli, N.R. (1985). Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science* **229**, 171–174.
101. Benjamin, H.W. and Cozzarelli, N.R. (1988). Isolation and characterisation of the Tn*3* resolvase synaptic intermediate. *EMBO J.* **7**, 1897–1905.
102. Yang, W. and Steitz, T.A. (1995). Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell* **82**, 193–207.
103. Benjamin, H.W., Matzuk, M.M., Krasnow, M.A., and Cozzarelli, N.R. (1985). Recombination site selection by Tn*3* resolvase: topological tests of a tracking mechanism. *Cell* **40**, 147–158.
104. Craigie, R. and Mizuuchi, K. (1986). Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell* **45**, 793–800.
105. Boocock, M.R., Brown, J.L., and Sherratt, D.J. (1986). Structural and catalytic properties of specific complexes between Tn*3* resolvase and the recombination site res. *Biochem. Soc. Trans.* **14**, 214–216.
106. Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature* **421**, 448–453.
107. Pollock, T.J. and Nash, H.A. (1983). Knotting of DNA caused by a genetic rearrangement. Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. *J. Mol. Biol.* **170**, 1–18.

GLOSSARY

A-DNA	A conformation of right-handed double-stranded DNA characterized in part by a tilting of the bases with respect to the helix axis; found in DNA at low humidity.
Alexander–Briggs notation	A system for classification of knots and catenanes, based on the number of nodes.
average crossing number	The algebraic sum of crossings or nodes visible in a two-dimensional projection of a knot or catenane, averaged over all possible directions of projection.
B-DNA	A conformation of right-handed double-stranded DNA characterized in part by a perpendicular arrangement of the bases with respect to the helix axis; thought to be the form of DNA commonly found in nature.
bending force constant (*B*)	The constant relating the free energy of bending of DNA (or any elastic material) to the length of the bent segment and the angle of bend.
catenane	Interlinked circular DNA molecules, as in the links of a chain.
catenane number (Ca)	The algebraic sum of the intermolecular double-stranded nodes of the DNA molecules comprising a catenane.
closed-circular DNA	Double-helical DNA in which both strands are closed 5′ to 3′ into circles (i.e. circular double-stranded DNA with no free ends).
cruciform	A stem-loop structure in double-stranded DNA formed by intrastrand base-pairing of each strand. Cruciforms are stabilized by negative supercoiling.
curved DNA	DNA in which the preferred conformation has a curved axis.
directional writhe	The algebraic sum of the crossings or nodes visible in a two-dimensional projection of a closed curve (such as the axis of a closed-circular DNA). The writhe is the directional writhe averaged over all possible directions of projection.
effective diameter (*d*)	A parameter used to approximate the tendency of charged polymers like DNA to repel one another at a distance that may be greater than the physical diameter of the polymer. It is a reasonable approximation to model the polymer as an impenetrable cylinder of diameter d, where d will depend on the solution conditions.

excluded volume The volume occupied by a polymer of finite thickness, which is unavailable to another segment of the same polymer molecule. Related to effective diameter.

flexible DNA DNA that can be deformed, particularly by bending, more easily than average.

frame of reference An environment or perspective (technically a set of co-ordinates) relative to which one can measure motion.

H-DNA A triple-stranded conformation of DNA comprising a polypyrimidine strand lying in the major groove of a polypurine–polypyrimidine duplex. H-DNA is stabilized by negative supercoiling.

helical repeat (*h*) The number of base pairs per turn of the double-stranded DNA helix. This can be measured relative to the DNA helix axis (twist-related h; h_t) or relative to a surface on which the DNA lies (surface-related h; h_s). Also designated γ in some recent literature.

Holliday junction An intermediate in recombination formed by strand-exchange between two homologous duplexes. Structurally related to a cruciform.

Hoogsteen base pairs Base pairs involving alternative hydrogen bond donor and acceptor partners to the standard Watson–Crick base pairs.

ideal knots and catenanes The ideal form of a knot or catenane is the geometrical configuration having the highest ratio of volume to surface area, that is, the shortest piece of tube that can be closed to form the knot or catenane.

intercalator Planar small molecule that binds to DNA by stacking between base pairs, normally resulting in a local unwinding of the DNA helix. Examples include ethidium bromide and chloroquine.

interwound supercoiling See plectonemic supercoiling.

***j*-factor** Parameter describing the cyclization kinetics of molecules such as double-stranded DNA. The j-factor is the ratio of equilibrium constants for the formation of a circular molecule and a dimer linear molecule. The j-factor is the concentration of one end of the DNA in the vicinity of the other (in a conformation able to ligate to it).

knot A knotted length of linear DNA closed into a circle.

knot number (Kn) The algebraic sum of the minimum number of nodes in a knotted DNA circle.

linking difference (ΔLk) The difference between the linking number of a particular topoisomer of closed-circular DNA (or the average Lk of a DNA sample) and the average linking number of relaxed DNA under given conditions (i.e. $\Delta Lk = Lk - Lk^{\circ}$). Positive linking difference ($+\Delta Lk$) corresponds to positively supercoiled DNA and $-\Delta Lk$ to negatively supercoiled DNA.

linking number (Lk) The number of times the two strands of closed-circular DNA are linked. Strictly defined as half the algebraic sum of the interstrand nodes in a two-dimensional projection of the DNA. Linking number is distributed between the two geometric parameters twist (Tw) and writhe (Wr) (i.e. $Lk = Tw + Wr$). Lk is designated α in some early literature.

Lk° The mean linking number of relaxed DNA. The value need not be an integer, and is dependent on the solution conditions.

Monte Carlo simulation Generally, a simulation of a physical system that involves a random probabilistic component. For DNA, a computer simulation of DNA as an elastic jointed rod. Can be used to predict conformations and energies of supercoiled, catenated, and knotted DNA.

negatively supercoiled DNA Supercoiled DNA with a negative average linking difference.

nicked DNA See open-circular DNA.

node A crossover point of two DNA single- or double-strands in a two-dimensional projection. The sign of a node is determined by the directions assigned to the DNA strands.

open-circular DNA A double-strand DNA circle containing a broken phosphodiester bond in one strand. Also known as nicked DNA.

persistence length (a) Length describing the thermal flexibility of DNA (or any material). Lengths shorter than a persistence length behave as a fairly rigid rod, while those of many multiples of a persistence length behave as a random coil. The persistence length of DNA is approximately 150 bp.

plectonemic supercoiling A class of supercoiled DNA conformations in which the DNA axis follows a superhelical path around another segment of the same molecule. For negative supercoiling, the sense of the superhelix is right-handed. This is the conformation usually adopted by free supercoiled DNA. Also known as interwound supercoiling (cf. toroidal supercoiling).

positively supercoiled DNA Supercoiled DNA with a positive average linking difference.

recombination The rearrangement of a DNA molecule or molecules resulting from the breakage of the DNA duplex and its rejoining to another site.

relaxed DNA Closed-circular DNA formed without any constraint of the DNA helix, and with average $\Delta Lk = 0$; normally consists of a distribution of topoisomers.

specific linking difference (σ) The linking difference normalized to the length of the DNA, by dividing by Lk° (i.e. $\sigma = \Delta Lk/Lk^\circ$).

superhelix winding angle (γ) The pitch angle of a DNA superhelix. Also designated α in some recent literature.

supercoiled DNA Closed-circular DNA formed under a torsional stress. A vernacular term for DNA with a non-zero linking difference. Supercoiling can be manifested as a change in twist and/or a change in writhe.

superhelix DNA wound into a helix of a higher order than the normal DNA double helix; in other words, where the axis of the double helix itself follows a helical path; characteristic of supercoiled DNA. Can be plectonemic (interwound) or toroidal.

surface linking number (SLk) For a closed-circular DNA lying on a surface, SLk is the linking number of the helix axis and a line traced out by the surface normal.

surface twist (STw) The twist of a line traced out by the surface normal about the DNA axis, for DNA lying on a surface.

topoisomer A closed-circular DNA molecule with a particular value of linking number. Molecules differing only in linking number are topoisomers.

topoisomerase An enzyme that catalyses changes in the linking number of closed-circular DNA.

toroidal supercoiling A class of supercoiled DNA conformations in which the DNA axis follows a superhelical path around the axis of a torus. For negative supercoiling, the sense of the superhelix is left-handed. This conformation is not adopted by free DNA, but is often an appropriate model for DNA wound on a protein surface (cf. plectonemic supercoiling).

triple helix A three-stranded conformation of DNA comprising a single strand lying in the major groove of a DNA duplex.

twist (Tw) Simply stated, twist is the number of double-helical turns in a given length of DNA, measured relative to the DNA helix axis. In fact, twist cannot normally be determined by counting and is

	a more complex function of the path of a DNA strand about the helix axis.
twisting force constant (*C*)	The constant relating the free energy of twisting of DNA (or any elastic material) to the length of the twisted segment and the angle of twist.
type I topoisomerase	A topoisomerase whose reaction is characterized by linking number changes of ± 1.
type II topoisomerase	A topoisomerase whose reaction is characterized by linking number changes of ± 2.
winding number (Φ)	The number of double-helical turns in a closed-circular DNA measured relative to a surface on which the DNA helix axis lies.
writhe (Wr)	A parameter quantifying the coiling of the path of a DNA helix axis in space. Planar DNA, and DNA whose axis lies on a sphere without crossing, have $\mathrm{Wr} = 0$.
Z-DNA	A conformation of double-stranded DNA characterized by a left-handed helix. Can be formed in alternating purine–pyrimidine sequences. Z-DNA is stabilized by negative supercoiling.

This page intentionally left blank

INDEX