

HYPOTHESIS

A hierarchical model for evolution of 23S ribosomal RNA

Konstantin Bokov¹ & Sergey V. Steinberg¹

The emergence of the ribosome constituted a pivotal step in the evolution of life. This event happened nearly four billion years ago, and any traces of early stages of ribosome evolution are generally thought to have completely eroded away. Surprisingly, a detailed analysis of the structure of the modern ribosome reveals a concerted and modular scheme of its early evolution.

The ribosome is an RNA–protein complex performing protein synthesis in all living cells¹. It is generally accepted that the ribosome originated from the so-called ‘RNA world’ when proteins did not exist and the primordial chemical reactions of life were catalysed by RNA^{2,3}. Although the contemporary ribosome contains several dozen proteins^{4–8}, the two major functions of the ribosome—the selection of the proper amino acid and the transpeptidation—are performed by RNA^{9–11}, whereas proteins have only an auxiliary role. Structurally, RNA forms the core of the ribosome, whereas proteins are mostly located at the periphery. Hence, the problem of the origin of the ribosome concerns the origin of ribosomal RNA. Because in all living organisms the core of the ribosome has a very similar structure, it must have formed before the split of the tree of life into three phylogenetic domains^{12,13}. Consequently, the comparison of the available nucleotide sequences of rRNA is not sufficient for the deduction of how the ribosome emerged. However, the ribosome tertiary structure could provide key clues about the details of this process.

Our initial observation was that, compared to other domains of the 23S rRNA secondary structure, A-minor interactions in domain V¹⁴ follow a very specific pattern. A-minor is a frequently found RNA arrangement consisting of a stack of unpaired nucleotides, predominantly adenosines, that pack with a double helix^{15,16}. In the A-minor interactions that domain V forms with other parts of 23S rRNA, the double helix almost exclusively belongs to domain V, whereas the adenosine stack usually belongs to the rest of the molecule (Fig. 1). This characteristic distinguishes domain V from other domains of 23S rRNA, in which the proportion of the adenosine stacks and the double helical regions that form A-minor interactions is reversed.

To explain this abnormality of domain V, we suggest that it reflects the order in which different parts were added to 23S rRNA as it evolved. In the A-minor motif, the conformational integrity of the adenosine stack depends on the presence of the double helix, whereas the helix can maintain a stable conformation without interaction with its counterpart. Presuming that the integrity of the ribosome structure has been maintained throughout its entire evolution, adenosine stacks should not have appeared in rRNA before the corresponding double helices. Because domain V contains the peptidyl-transferase centre (PTC)¹⁷, which performs the central function of the ribosome, we expect it to be among the most ancient elements of the ribosome structure. Thus, the abnormality of domain V can be explained by the formation of the A-minor interactions between double helices of a more ancient domain V and the adenosine stacks of more recently acquired parts of 23S rRNA.

The model

The ability of the A-minor motif to serve as an indicator of the relative age of its moieties can be used to determine the order in which different elements were added to the ribosome structure during its evolution. To demonstrate such ordered assembly, we developed a strategy of systematically dismantling the ribosome structure through elimination of those elements that could be considered as most recent acquisitions. As an element, we considered an individual double helix or a domain of stacked nucleotides that on addition to the ribosome structure would form a stable compact arrangement. We suggested a general principle that an element could not be a recent addition if its removal compromised the integrity of the remaining parts of the ribosome. The 5′ and 3′ ends of a removed element must be structurally close enough to each other to be considered a local insertion. This would guarantee that, after the fragment is removed, the remaining RNA chain maintains its structural integrity. Also, because the integrity of each strand of a double helix

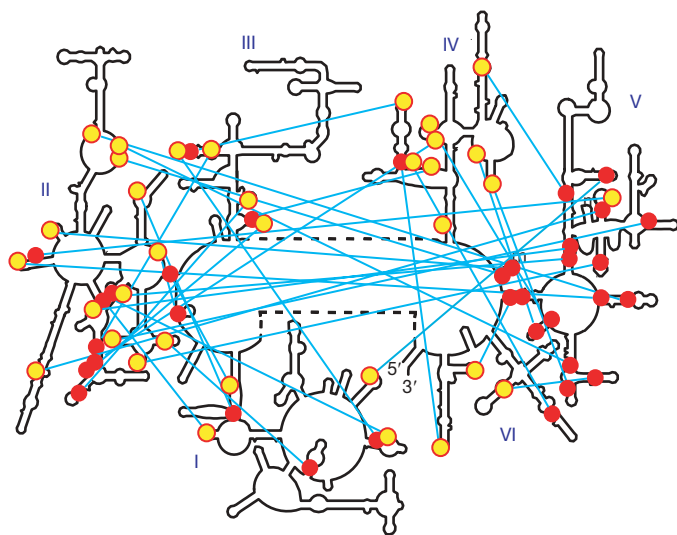


Figure 1 | Location of inter-domain A-minor interactions in the secondary structure of the *E. coli* 23S rRNA. The secondary-structure domains are marked by roman numerals. Each A-minor interaction is shown by a cyan line connecting the double helix (red circle) and the corresponding adenosine stack (yellow circle). Unlike other domains, domain V almost exclusively forms these interactions using double helices and not adenosine stacks.

¹Département de Biochimie, Université de Montréal, Montréal, H3C 3J7 (Québec), Canada.

depends on the presence of the other strand, a removed element must contain both strands of the same helix. Finally, if a removed element forms the A-minor motif with the remaining ribosome, it must contain the stack of unpaired nucleotides that form this interaction, and not the double helix. Because our analysis was focused on A-minor interactions, the exact location of the boundaries between different elements was not essential, as long as all adenosine stacks and all corresponding double helical regions remained intact. Additional requirements imposed on elements are discussed in Supplementary Data 1.

Analysis of the tertiary structure of the *Escherichia coli* 23S rRNA⁸ revealed 19 elements for which elimination does not compromise the integrity of the remaining part of the structure. These elements form level 1 in Fig. 2b; their location in the 23S rRNA secondary structure is shown in Fig. 2a and their complete description is given in Supplementary Data 1. The identified elements form a total of 13 A-minor interactions with regions located in the remaining part of the molecule (see Supplementary Data 2 and Supplementary Notes 1). In all of these interactions the adenosine stacks belong to the identified elements, whereas the double helices are located in the remaining part of 23S rRNA. Thus, the elements of layer 1 could be considered the final generation of acquired elements. We then identified a further 11 elements, the presence of which is essential for the integrity of only the elements of level 1. Accordingly, we describe these elements as constituting the penultimate generation of added elements (elements of level 2 in Fig. 2b). We repeated the same procedure ten more times and identified a total of 59 elements.

The position and the conformation of each identified element depend on the presence of only the elements of the preceding generations. In Fig. 2b, each dependency of element P on the presence of element Q is shown as arrow $Q \rightarrow P$. There are two types of dependencies, D1 and D2. A D1 dependency indicates that the removal of Q before P would split the whole molecule into two separate parts. A D2 dependency indicates that the removal of Q before P would compromise the conformation of P. In total, we identified 59 D1 dependencies and 56 D2 dependencies. Out of all D2 dependencies, 54 were based on the formation of A-minor interactions. The remaining 2 D2 dependencies corresponded to two non-local pseudoknots (discussed later).

The removal of the 12 generations of acquired elements eliminated 93% of the original 23S rRNA. The remaining part, located in domain V, is shown in Fig. 2a by the blue and red lines; its central loop forms the PTC. Recently, it was observed that this region consists of two consecutive parts having practically identical secondary and tertiary structures^{11,18} (blue and red parts in Fig. 2a; see also Supplementary Figs 1 and 2). The blue and red parts are arranged symmetrically to each other and form binding sites for the CCA-3'-termini of transfer RNA molecules in the P- and A-sites, respectively. Moreover, there is a very close correspondence between the positions of the nucleotides of both parts involved in the fixation of the equivalent elements of both tRNAs^{11,18–21}.

The similarity between both parts is so high that it is logical to suggest that they originated by a duplication of the same RNA fragment. From this point of view, the evolution of 23S rRNA started with an initial fragment of about 110 nucleotides, which, probably, was able to bind the CCA-3'-terminus of what would later be tRNA. The duplication of this fragment allowed the resulting molecule to bind simultaneously two CCA-3'-termini. Within this arrangement, the two CCA-3'-termini associated with both parts are juxtaposed in space to allow for the transpeptidation reaction. Most probably, this dimer was already able to synthesize oligopeptides with random amino acid sequences, which would allow us to call it proto-ribosome. This view is supported by the fact that *in-vitro*-selected small RNA molecules resembling the PTC are able to perform transpeptidation²², thus demonstrating that this reaction does not require any other elements of the ribosome structure. All other elements of 23S rRNA were gradually added to the structure, one element at a time, in essentially the same way. Each element could appear only when all elements that were required for its proper positioning had already

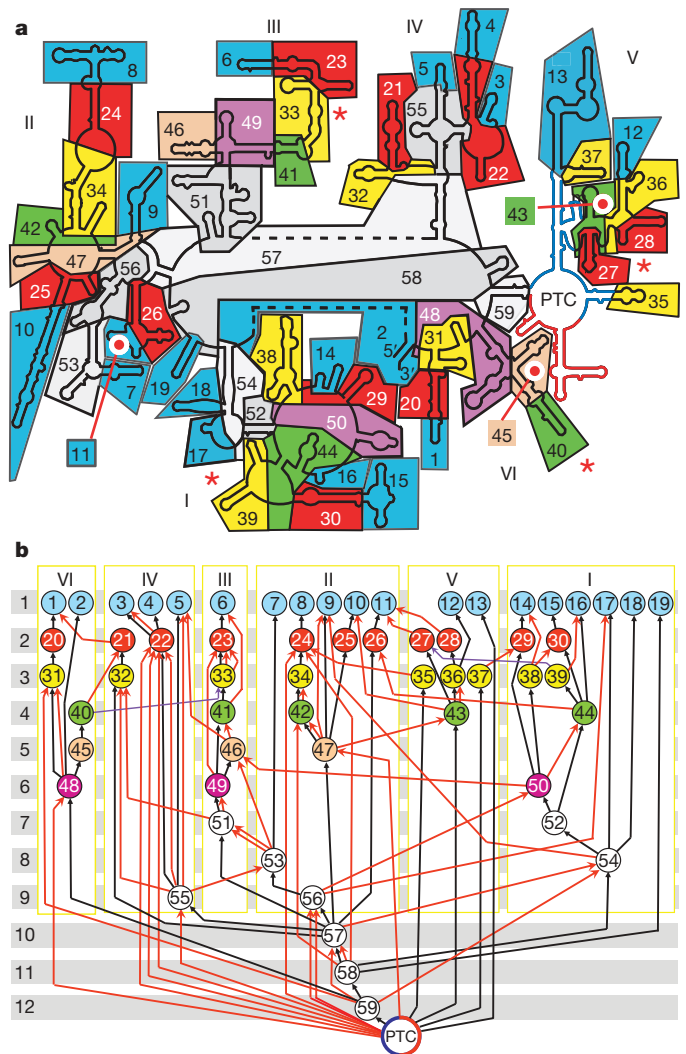


Figure 2 | The location of the identified elements in the *E. coli* 23S rRNA secondary structure (a) and the network of D1 and D2 dependencies between them (b). Each element has the same colour in a and b. The roman numerals indicate secondary-structure domains. PTC stands for the symmetrical arrangement in domain V containing the peptidyl-transferase centre (the proto-ribosome). a, The two halves of the proto-ribosome are blue and red. Red asterisks indicate the four elements that form two non-local pseudoknots 27–39 and 33–40. b, An arrow connecting two elements $Q \rightarrow P$ indicates that the position of P depends on the presence of Q. Black and coloured arrows represent D1 and D2 dependencies, respectively. Red arrows $Q \rightarrow P$ represent A-minor interactions formed by a double helix of element Q and a nucleotide stack of element P. Two violet arrows originate from the dissection of two non-local pseudoknots (see Supplementary Notes 1). The numbers of levels are shown on the left. The detailed description of all elements and of all D2 dependencies is given in Supplementary Data 1 and 2.

been placed. New elements were added as insertions containing all necessary details to dock with the surface of the evolving ribosome without disturbing already existing parts. The most common way for a new element to be fixed on the ribosome surface would be through the formation of an A-minor interaction with an already existing double helix.

Justification of the model

For justification of the suggested evolutionary model, we analysed those features of the 23S rRNA tertiary structure for which the 59 elements could be consecutively removed without damaging the integrity of the remaining part. Our analysis shows that removal of these elements is possible if and only if the arrows representing D1 and D2 dependencies do not form cyclic structures—that is, cases

where a chain of several consecutive arrows arranged head-to-tail starts and finishes at the same element. A mathematically rigorous proof of this statement and the explanation of why the absence of cycles of dependence is essential for dismantling the 23S rRNA structure are given in Supplementary Notes 2.

The absence of cycles built of D1 dependencies reflects the hierarchical topology of the secondary structure of 23S rRNA in which the removal of remote elements of each domain would not compromise the integrity of the remaining RNA chain. Such topology could be disrupted by pseudoknots, several of which exist in 23S rRNA¹². However, most pseudoknots are arranged in the same region of the secondary structure and can be removed as a single element. Only two pseudoknots between the loops of elements 27–39 and 33–40 are not local (Fig. 2a). However, in both cases it was possible to split the two strands of the inter-loop double helix on the grounds that the conformational integrity of only one of the two loops requires the presence of the other loop (Fig. 2b and Supplementary Notes 1).

For the given 23S rRNA secondary structure, the absence of cycles involving D2 dependencies is a consequence of the particular orientation of many A-minor interactions. For example, dependency 41→6 (Fig. 2b) stands for the A-minor interaction between the double helix of element 41 and the adenosine stack of element 6. In the opposite situation, if the double helix occurred in element 6, while the adenosine stack was provided by element 41, four elements would have formed cycle 41→33→23→6→41. Similar cycles would have occurred in many other parts of 23S rRNA if the orientations of A-minor interactions were different. The existence of any such cycle would have arrested the procedure of dismantling the 23S rRNA structure before it reached PTC.

To demonstrate how unlikely the absence of cycles really is, we calculated the probability for the 23S rRNA structure to be cycle-free if the orientations of all A-minor interactions were chosen randomly. Our analysis presented in Supplementary Notes 2 shows that the total probability of a cycle-free arrangement in this case would be $P < 10^{-9}$. Such low probability excludes the possibility that the absence of cycles of dependence in 23S rRNA has occurred by chance. Instead, it strongly supports a hierarchical scenario for its evolution, according to which the integrity of each element of 23S rRNA depends only on the presence of more ancient elements of its structure. The absence of cycles in the 23S rRNA tertiary structure does not depend on the way we defined individual elements, but instead represents a fundamental property of this molecule.

Major periods in the 23S rRNA evolution

The scheme of dependencies presented in Fig. 2b can help us to elucidate some details of the evolution of the large ribosomal subunit after the emergence of the proto-ribosome. Our analysis shows that stabilization of the proto-ribosome tertiary structure was a major aspect of the 23S rRNA evolution in the post-proto-ribosome era. In Fig. 3 the structure of the proto-ribosome is shown without other parts of 23S rRNA (Fig. 3a) and with the gradually increasing number of added elements (Fig. 3b–e). The elements forming each structure are shown in Supplementary Fig. 3. The first 8 elements added to the proto-ribosome form a foundation that closely interacts with the bottom part of the proto-ribosome and effectively supports its conformation (Fig. 3b). Further addition of 12 elements makes this foundation wider and more massive (Fig. 3c). Finally, after the addition of a total of 50 elements, the proto-ribosome became surrounded by added elements on all sides except the side from which PTC must be reached by tRNAs (Fig. 3d). The added elements were arranged so they did not interfere with the release of the nascent peptide, leading to the formation of the exit channel (4 in Fig. 3f).

The emergence of the foundation provided new functional opportunities. In particular, it allowed the formation of the area of contact with the small ribosomal subunit (Fig. 3e), which was essential for the integration of this subunit into the ribosome. Another consequence of

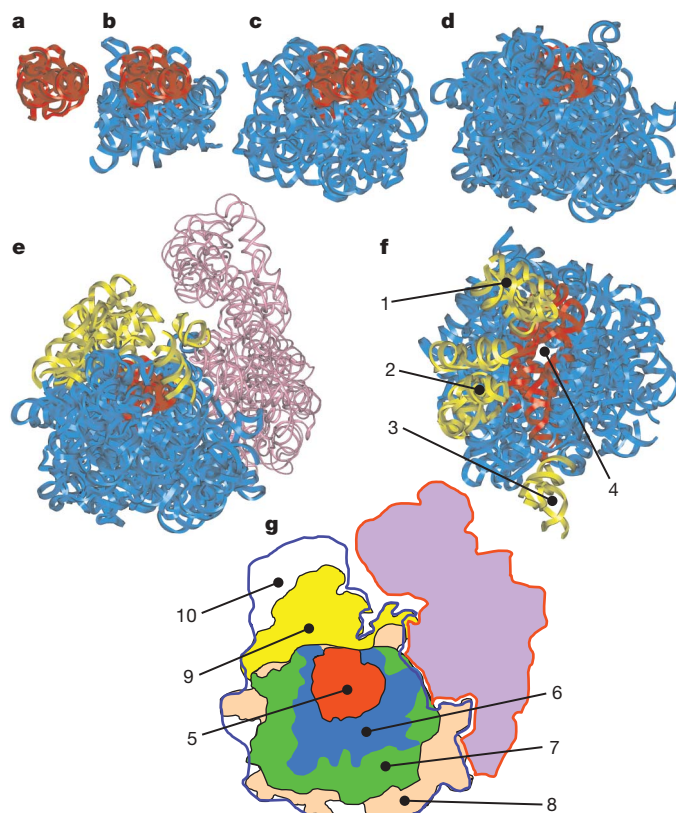


Figure 3 | The aggrandizement of the 23S rRNA structure during its evolution. **a–e**, the proto-ribosome with 0 (**a**), 8 (**b**), 20 (**c**), 50 (**d**) and all 59 (**e**) elements added. The proto-ribosome is red, elements forming the proto-ribosome foundation are blue, the protuberances are yellow, and 16S rRNA is purple. The complete list of the elements forming structures **a–e** is given in Supplementary Fig. 3. **f**, The top view of the 23S rRNA structure shown in **e**. **g**, The positions of the parts of 23S rRNA shown in **a–e** in the context of the whole ribosome. The structures of the 50S and 30S subunits are contoured by the blue and red line, respectively. 1–3 are the L7/L12, central and L1 protuberances, respectively; 4 is the exit channel; 5–9 are the structures shown in **a–e**, respectively; 10 is the part of 50S subunit that does not include 23S rRNA. This part is formed by ribosomal proteins and 5S rRNA.

the lateral expansion of the proto-ribosome foundation was that it allowed the formation of the three protuberances (yellow in Fig. 3e–g). In Fig. 2b, the elements forming these protuberances are positioned at the upper levels (see Supplementary Fig. 3f). Correspondingly, the particular functions associated with the protuberances—namely, the assistance in the selection of the proper aminoacyl-tRNA and the GTPase reaction^{23,24} (the L7/L12 protuberance) as well as in the release of the deacylated tRNA from the E-site²⁵ (the L1 protuberance)—should be relatively late acquisitions of the ribosome.

Our results also demonstrate that, despite its visible complexity, the structure of 23S rRNA follows a rather simple principle and could have evolved in a relatively short time on the evolutionary scale. Each new insertion emerged randomly and was accommodated only if it made the ribosome more stable and effective as a transpeptidase. At early stages of evolution, the ribosome existed exclusively as an RNA body. Later, when the ribosome functioning became sufficiently effective to produce proteins, the latter started playing an important part in the ribosome structure. We can argue that, among all structures shown in Fig. 3, the structure in Fig. 3b corresponds most closely to the moment when the RNA world changed for the protein-based world. This conclusion is based on the fact that although ribosomal proteins interact with the structure in Fig. 3b only marginally, they form extensive contacts with later structures (not shown). Whether indeed the structure in Fig. 3b corresponds to the end of the RNA world and thus

represents the most effective all-RNA ribosome, however, requires further experimental analysis.

1. Stillman, B. (ed.) *The Ribosome. Cold Spring Harbor Symposia on Quantative Biology* (Cold Spring Harbor Laboratory Press, 2001).
2. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–369 (1968).
3. Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).
4. Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
5. Ban, N. *et al.* The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
6. Harms, J. *et al.* High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–688 (2001).
7. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
8. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
9. Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
10. Noller, H. F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416–1419 (1992).
11. Nissen, P. *et al.* The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
12. Gutell, R. R., Larsen, N. & Woese, C. R. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**, 10–26 (1994).
13. Doudna, J. A. & Rath, V. L. Structure and function of the eukaryotic ribosome: the next frontier. *Cell* **109**, 153–156 (2002).
14. Cannone, J. J. *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**, 2 (2002).
15. Nissen, P. *et al.* RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
16. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. A universal mode of helix packing in RNA. *Nature Struct. Biol.* **8**, 339–343 (2001).
17. Polacek, N. & Mankin, A. S. The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit. Rev. Biochem. Mol. Biol.* **40**, 285–311 (2005).
18. Agmon, I., Bashan, A., Zarivach, R. & Yonath, A. Symmetry at the active site of the ribosome: structural and functional implications. *Biol. Chem.* **386**, 833–844 (2005).
19. Samaha, R. R., Green, R. & Noller, H. F. A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* **377**, 309–314 (1995).
20. Kim, D. F. & Green, R. Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol. Cell* **4**, 859–864 (1999).
21. Hansen, J. L., Schmeing, T. M., Moore, P. B. & Steitz, T. A. Structural insights into peptide bond formation. *Proc. Natl Acad. Sci. USA* **99**, 11670–11675 (2002).
22. Zhang, B. & Cech, T. R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390**, 96–100 (1997).
23. Savelsbergh, A. *et al.* Stimulation of the GTPase activity of translation elongation factor G by ribosomal protein L7/12. *J. Biol. Chem.* **275**, 890–894 (2000).
24. Kavran, J. M. & Steitz, T. A. Structure of the base of the L7/L12 stalk of the *Haloarcula marismortui* large ribosomal subunit: analysis of L11 movements. *J. Mol. Biol.* **371**, 1047–1059 (2007).
25. Nikulin, A. *et al.* Structure of the L1 protuberance in the ribosome. *Nature Struct. Biol.* **10**, 104–108 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Brakier-Gingras, A. Mankin, S. Michnick and I. Ponomarenko for advice and comments. This work was supported by a grant from NSERC.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.V.S. (serguei.chteinberg@umontreal.ca).