New RNA recognition features revealed in ancient ribosomal proteins

Ada Yonath[1,2] and Francois Franceschi[3]

A new, inherently flexible, predominantly α-helical RNA binding motif has been detected in two rather conserved ribosomal proteins. Comparative studies prompt the suggestion that these ribosomal proteins are the ancestors of this new helix-turn-helix fold, also found in homeodomains.

The intricate and essential process of the enzymatic translation of genetic information into proteins is performed by a universal ribonucleoprotein assembly—the ribosome. Structural studies aimed at unraveling the secrets of ribosomal function have taken two approaches: one focuses on isolated ribosomal components, the second aims at the elucidation of the structure of the entire ribosome.

For individual ribosomal proteins, it remains to be seen whether those that are structured in solution bear any resemblance to their *in situ* conformations. Nevertheless, a considerable effort is being invested in their study, as crystallization of isolated ribosomal proteins may indicate the existence of an independent fold which does not alter significantly upon removal from the ribosomal environment. For over two decades attempts at the crystallization of isolated proteins from the *Escherichia coli* ribosome resulted in considerable frustration, as the yield of useful crystals was extremely poor. This, together with the observations that some ribosomal proteins loose their *in situ* conformation upon isolation, led to the assumption that the conformations of almost all ribosomal proteins are dictated by their surrounding environment—the internal structure of the ribosome.

The introduction of three-dimensional NMR spectroscopy as a method for structure determination, the increasing sophistication in crystallographic procedures, the implementation of powerful genetic techniques and the gradual substitution of the *E. coli* ribosome—which used to be a favoured research subject—by more robust ribosomes (for example, those from thermophilic bacteria), has resulted in considerable success in the determination of the molecular structure of ribosomal proteins[1–6].

The quantum leap in the determination of structures of isolated ribosomal components reveals that the ribosomal proteins can be divided into three groups: first, those that loose their *in situ* conformation upon detachment from the ribosomal scaffold; second, those possessing intrinsic characteristic folds (which may yield crystals providing the proteins are not damaged during preparation, a condition better fulfilled when using thermophilic ribosomes); and third, those with flexible folds that can be stabilized in solution by other ribosomal components (such as neighbouring ribosomal proteins or short segments of rRNA with which they interact *in situ*)—a class that will most certainly be found to include the majority of the ribosomal proteins.

Attempts to determine the three-dimensional structure of intact ribosomal particles by crystallographic methods have shown promise far beyond initial expectations—crystals diffracting to 2.9 Å resolution[7] have been grown from native, chemically modified, mutated and complexed ribosomal particles, and initial phasing attempts indicate the feasibly of their structure solution[8].

A new fold

The structures of S15 (ref. 4) and the C-terminal fragment of L11, named L11-C76 (refs 5, 6), have been determined by heteronuclear three-dimensional NMR spectroscopy and are reported in this issue of *Nature Structural Biology* (the names of individual ribosomal proteins are composed of a letter—L or S, for the large or small ribosomal subunit—and a running number, according to the position of the protein on two-dimensional gels). Surprisingly, both structures show a very similar fold, composed predominantly of α-helices. This similar fold, which can be described as helix-turn-helix—a motif found in non-RNA binding domains in ribosomal proteins (for example, L12)—varies substantially from all RNA recognition motifs detected previously in ribosomal proteins[3,4]. Thus, among a dozen ribosomal proteins whose structures are known so far, many share a split-β-α-β fold, called the 'common' motif and abbreviated RRM (for RNA recognition motif). A few proteins do show different rRNA binding motifs, called 'unique' and 'multiple', in which the rRNA interacting regions are built mainly of loops. Even protein S8, which exhibits versatility in its two RNA contact sites—one interacting with rRNA and the other with tRNA[3]—does not contain a fold similar to those found in S15 and L11-C76. However, predominantly α-helical domains that interact directly with RNA have been detected in non-ribosomal proteins: serine and phenylalanine tRNA synthetase[9,10].

Several ribosomal proteins function as translational regulators through their ability to bind to their own polycistronic mRNA[11]. It has been suggested that the binding mode of these proteins to mRNA is similar to their interactions with rRNA: for example, L1 (ref. 12) and S8 (ref. 3). S15 is also an autoregulatory protein. However, in contrast to L1 and S8, the sequences of its rRNA and mRNA target sites do not resemble each other. Does this imply "specific recognition by S15 of a common RNA structure" as suggested by Härd and colleagues[4], or does it indicate that the flexibility of S15 is of a magnitude suitable for adjustment to two different RNA conformations? Since the structural information concerning such interactions is rather limited, it is difficult at this stage to draw firm conclusions. However it is noteworthy that L1 and S8, which interact with similar rRNA and mRNA sequences, posses stable structures even in isolation, whereas the structure of S15, which interacts with various RNA sequences, exhibits significant flexibility.

Comparative studies, carried out by Xing *et al.*[6], show that the RNA binding domain of ribosomal protein L11 can be aligned with the homeodomain, found in a group of eukaryotic DNA-binding proteins. Similarly, a fold resembling the N-terminal domain of S5 has been found in two proteins that bind DNA (HealII methyltrans-

ferase and DNAse II)[3]. In the absence of structural information about rRNA–protein complexes these comparisons lead to the suggestion that RNA and DNA binding proteins may use a similar strategy for recognition. However, the differences between the structures of helical DNA and RNA seriously challenge this suggestion. The major groove of DNA is able to accommodate an α-helix, whereas both the major and the minor grooves of RNA do not form a smooth surface for interactions with α-helices. Hence it was suggested that the key features in RNA binding are bulges and mismatches which can alter the contacting surface on the RNA segment[13,14], a situation which is dramatically different from the smooth contact area between α-helices and DNA.

Who influences whom?

Although the cores of S15 and L11-C76 are structurally stable, the proteins also contain unstructured loops, which move on a picosecond to nanosecond time-scale in the free proteins and yet acquire stable conformations upon binding rRNA[4,5], presumably imitating processes occurring during the course of ribosome assembly. Conformational re-adjustments have been predicted to be associated with the formation of the appropriate *in situ* micro-environment within the ribosome. It was widely assumed that the ribosomal components that undergo the major conformational changes were the rRNA molecules. The observation that flexible loops in ribosomal proteins become ordered upon binding to rRNA indicates that changes in the protein conformation may also be essential for the assembly of the ribosome.

Phenylalanine tRNA synthetase[10] is of interest in this regard not only because it posseses a recognition fold built solely of α-helices, but also because it provides an illuminating example of the stabilization of a flexible portion of a protein by interactions with RNA. In the unbound protein the N-terminal region stretches into the solvent and is extremely flexible. Its conformational freedom is so high that the entire region (85 residues) could not be identified in the 2.9 Å electron density map[15]. However, in the bound conformation this region appears as a long coiled coil built of antiparallel α-helices (Fig. 1), and is well resolved in the electron density map[9]. This α-helical arm forms many critical contacts with the tRNA molecule, in a fashion similar to that of the stable arms of the serine tRNA synthetase[9]. It remains to be seen whether the contacts made by the flexible loops of S15 and L11-C76 with rRNA resemble those which are created by the two tRNA synthetases.

Evolution and universality

As expected, ribosomes are a rich repository of structural motifs for RNA binding. These motifs may point to evolutionary relationships between ribosomal proteins and other groups of nucleic-acid binding proteins. Since S15 and L11 are rather conserved in all three phylogenetic domains—prokaryote, archaea and eukaryote—they should be considerably more ancient than the eukaryotic homeodomains and, therefore, may be evolutionary ancestors of the helix-turn-helix structural motif, although it cannot be excluded that the homeodomain fold arose independently. Furthermore, Xing and colleagues[6] conclude that those ribosomal proteins that bind directly to ribosomal RNA and are conserved in all organisms may have been among the earliest site-specific nucleic acid binding proteins. This suggestion is in accord with the classification of phenylalanine tRNA synthetase as the first member belonging to Class II synthetases, based on its binding mode[16].

In several *in situ* substructures in the ribosome, the level of conservation is so high that ribosomal proteins from one source can be fully exchanged by their homologues from other ribosomes, even when they belong to two different kingdoms or live under totally different conditions. It is not surprising, for example, that L11 from *E. coli* binds well to cores of *Thermus thermophilus* ribosomes lacking their own L11 protein, in view of the generally high homology of these ribosomes. However, the universality of the internal substructures L1 and a segment of the 23S rRNA was somewhat of a surprise. Despite the 'exotic' properties of ribosomes from the extreme halophile *Haloarcula marismortui* and the evolutionary distance between archaea and eubacteria, chimeric complexes of L1 were reconstituted between the halophilic components and the corresponding ones from *E. coli*[12].

Some facts about S15 and L11

S15 and L11 are two conserved ribosomal proteins that interact directly with conserved regions of rRNA[4–6]. However, their functional roles and their locations within the ribosome are dramatically different. S15 binds to the central domain of the rRNA and is believed to be essential for the integrity of the particle. In *in vitro* assembly experiments it was shown to be a primary rRNA binding protein. In contrast, L11 is located on the ribosomal surface and although it promotes GTPase activity and is suggested to regulate the rRNA conformation, cells lacking it are viable and their ribosomes—although significantly less efficient in protein biosynthesis—do not show major structural changes, as shown by their ability to yield crystals similar to those of native ribosomes[17,18]. In contrast to S15, which supports assembly of the ribosome, L11 forms part of the target site for binding the lethal antibiotics thiostrepton and micrococcin; interestingly, this target site is

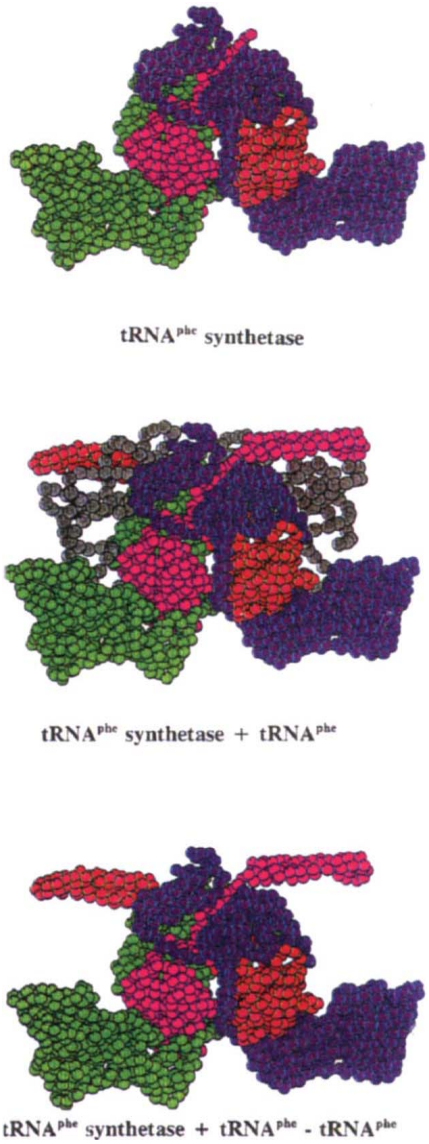

Fig. 1 *a*, Space-filling model of the structure of *T. thermophilus* phenylalanine synthetase[10]. The two heterodimers (violet/orange and green/pink) are related by two-fold symmetry axis which lies vertically in the plane of the figure. *b*, As in (*a*) but with two cognate tRNA molecules (grey). The two N-terminal domains, not resolved in the electron density map of the isolated protein become ordered and form long coiled coils of α-helices (orange and pink), embracing the tRNA molecules. *c*, A computer representation of the 'bound structure' of tRNA–Phe synthetase, but without the tRNA molecule, produced by computer subtraction of the tRNA molecule from (*b*).

conserved even in ribosomes which are not harmed by these antibiotics, such as those from *H. marismortui*.

S15 is a one-domain protein, whereas L11 forms two domains which vary considerably in their conformational stabilities. Apart for the structural heterogeneity of one of its loops, S15 displays a structured conformation in isolation, whereas under the same conditions, only the C-terminal domain of L11 exhibits a stable conformation—the N-terminal domain is completely unstructured. In a series of elegant experiments, the RNA binding properties of L11 were shown to be confined to the C-terminal domain[19]. This domain is a part of the GTPase centre. Its rRNA counterpart, a stretch of ~60 nucleotides, has been investigated thoroughly by Draper and colleagues[6], who show that one third of the nucleic acids are almost invariant among ribosomes from the three kingdoms. However, in contrast to the expectation that such a highly conserved region is intimately involved in intermolecular contacts, extensive mutagenesis indicates that only a few nucleotides are in close contact with L11. Thus, the recognition between L11 and the corresponding RNA is fold, rather than nucleotide dependent.

Confinement of the RNA binding site to the C-terminal domain suggests that the N-terminal domain is exposed to the solution. This assignment explains previous observations that the cysteine located in the C-terminal portion of L11 (for example, in L11 from *H. marismortui*) has reduced solvent accessibility[20] whereas the N-terminal cysteine in *Bacillus stearothermophilus* is on the ribosomal surface, exposed to the solution. This exposed cysteine is of interest due to its high level of conservation[21]. It is found at the same position (residue 39) in many organisms, including *E. coli* and *B. stearothermophilus*. But despite its conservation, this cysteine seems to have no role in ribosomal function or integrity, as it can be removed or modified without effecting the overall structure of the ribosome[17,18]. Indeed, despite the striking homology of L11 from *E. coli* and *T. thermophilus*, the latter does not contain any cysteine[21].

Early studies on isolated L11 showed that its core can undergo proteinase cleavage even when the protein was extracted from the ribosome under non-denaturing conditions[22]. These findings were substantiated in experiments which showed that the fully exposed cysteine in L11 from *B. stearothermophilus* becomes buried on isolation of the protein, and the chemical reactivity of the -SH group was regained only when the protein was denatured[17,18]. The ability of isolated L11 to regain its native fold upon reconstitution into the ribosomal core particle was further demonstrated in an experiment which showed its incorporation even when large chemical moieties, of a molecular weight approaching a third of its own (6,200 and 15,500 M_r respectively), were bound to its -SH group[23]. Thus, L11 is an appropriate example for a ribosomal protein which undergoes significant, albeit reversible, conformational changes upon isolation from the ribosome.

Open questions—future prospects

Recent years have been an exciting period in ribosome research. Major conceptual advances have been made: for example, ribosomal functions are no longer attributed solely to the ribosomal proteins and the ribosomal RNA—which was previously considered a passive scaffold—has been shown to take an active part in ribosomal function. Furthermore, the studies reported in this issue[4–6] indicate that the rRNA may also influence the process of ribosomal assembly. During the last years spectacular progress has also been made in the structure determination of isolated ribosomal components. Nevertheless, despite these recent achievements, many crucial and intriguing questions remain unresolved.

For example, is the diversity of the RNA recognition motifs designed for selectivity in binding? Why do some binding regions seem to be stable in isolation whereas others show structural heterogeneity—until they are confronted with their mates? In cases of clear protein flexibility, does the RNA possess a fixed structure, or is the RNA also flexible and its final conformation formed only upon protein binding? Does the conservation of ribosomal components reflect their essential role for ribosome integrity or function? Why are both components of the GTPase centre so highly conserved, despite the finding that most of the bases in the rRNA segment are not essential for recognition, and the main functional activity of the ribosome is preserved even when one of the protein components of this centre, L11, is totally removed? Do the structures determined for isolated proteins resemble their *in vivo* conformations? This last question is of a high operational importance, since the answer may dictate the direction of the research in many laboratories. However, because it seems that this issue will not be resolved until the structure of the ribosome is determined, two approaches are being taken: studies on entire ribosomal particles as well as those on isolated components.

Perhaps the most striking and unexpected achievement has been reaching atomic resolution for intact ribosomal particles, as well as for some of the isolated components. This indicates a very high degree of order within the ribosome, which is still considered to be an extremely complicated ribonucleoprotein complex that is notoriously flexible and unstable. The two structural approaches will no doubt have to be united to achieve the determination of the complete molecular structure of the ribosome. Map interpretation can benefit significantly from the knowledge of the accurate atomic structures of individual ribosomal components, since they can serve as 'flags' and 'markers' which may be essential to gain a complete picture of this macromolecular complex.

[1]*Department of Structural Biology, Weizmann Institute, Rehovot 76100, Israel*
csyoanth@weizmann.weizmann.ac.il
[2]*Max-Planck Unit for Structural Molecular Biology, 22603, Hamburg, Germany*
yonath@mpgars.desy.de
[3]*Max-Planck Institute for Molecular Genetics, 14195 Berlin, Germany*
franceschi@mpimg-berlin-dahlem.mpg.de

1. Liljas, A. & Garber, M. *Curr. Op. Struct. Biol.* **5**, 721–727 (1995).
2. Nikonov, S. *et al. EMBO J.* **15**, 1350–1359 (1996).
3. Davies, C. Ramakrishnan, V. & White, S.W. *Structure* **4**, 1093–1104 (1996).
4. Berglund, H. *et al. Nature Struct. Biol.* **4**, 20–23 (1996).
5. Markus, M.A. *et al. Nature Struct. Biol.* **4**, 70–76 (1996).
6. Xing, Y., Takurta, D.G. & Draper, D.E. *Nature Struct. Biol.* **4**, 24–27 (1996).
7. von Bohlen, K. *et al. J. Mol. Biol.* **222**, 11–15 (1991).
8. Schluenzen, F. *et al. J. Biochem. Cell Biol.* **73**, 739–749 (1995).
9. Cusack, S. *et al. Nature* **347**, 249–255 (1990).
10. Goldgur, Y. *et al. Structure* **in the press.**
11. Nomura, M., Gourse, R. & Baughman, G. *Ann. Rev. Biochem.* **53**, 75–117 (1984).
12. Evers U., Franceschi, F., Boeddeker N., & Yonath, A. *Bioph. Chem.* **50**, 3–9 (1994).
13. Ye, X., Gorin, A., Ellington, A.D & Patel, D.J. *Nature Struct. Biol.* **3**, 1026–1033 (1996).
14. Batiste, T.L. *Science* **273**, 1547–1551 (1996).
15. Mosyak, L. *et al. Nature Struct. Biol.* **2**, 537–547 (1995).
16. Cusak, S. *Biochimi* **75**, 1077–1081 (1993).
17. Weinstein, S., Jahn, W., Wittmann, H.G. & Yonath, A. *J. Biol. Chem.* **264**, 19138–19142 (1989).
18. Bartels, H. *et al. J. Peptide Sciences* **37**, 411–419 (1995).
19. Xing, Y. & Draper, D.E. *Biochemistry* **35**, 1581–88 (1996).
20. Sagi, I. *et al. Biophys. Chem.* **55**, 31–41 (1955).
21. Heinrich, T., Erdmann, V.A. & Hartmann, R.K. *Gene*, **136**, 373–374 (1993).
22. Choli, T. *Biochem. Intern.* **19**, 1323–1338 (1989).
23. Thygesen, J., Weinstein, S., Franceschi., F. & Yonath, A. *Structure* **4**, 513–518 (1996).