If the loop fits...

Alan D. Frankel

The structure of the yeast L30 ribosomal protein bound to its autoregulatory RNA site has been determined by NMR spectroscopy. The intricate architecture of the RNA internal loop and the structure of the binding region of the protein both are stabilized in the complex, highlighting the importance of mutually-induced fit in RNA–protein interactions.

Each time the structure of a new RNA–protein complex is solved, a delightful new world of RNA shapes, twists, nooks, and crannies seems to be revealed. The structure of the yeast ribosomal L30 protein bound to its autoregulatory site described on page 1139 of this issue of *Nature Structural Biology* is no exception[1]. In that structure, an internal loop in the RNA brings together two helical stems in a highly distorted architecture, providing the docking site for the protein. One intriguing aspect of the L30–RNA complex is the degree to which it exhibits an induced fit mode of binding, a characteristic seen repeatedly in RNA–protein interactions.

The L30 ribosomal protein from Saccharomyces cerevisiae is among a large class of ribosomal proteins that regulate expression of their own genes by a variety of interesting mechanisms (reviewed in ref. 2). An elegant set of studies by Warner and colleagues3–5 showed that regulation of L30 expression occurs at the levels of pre-mRNA splicing and translation. A sequence near the AUG initiation codon pairs either with an adjacent intron sequence in the unspliced RNA or with a similar adjacent exon sequence in the mature RNA, forming binding sites for L30. Binding of L30 to the first site disrupts spliceosome assembly whereas binding to the second site inhibits translation. Phylogenetic comparisons, chemical probing experiments and in vitro selection identified a purine-rich internal loop as the binding site of L30[3,6,7] (Fig. 1a). The NMR structure described by Williamson and coworkers shows L30 bound to the internal purine-rich loop of the pre-mRNA autoregulatory site.

The asymmetric internal loop of the L30 binding site is comprised of two purines opposite five purines (Fig. 1*a*) and adopts a very unusual structure in which several of the purines are stacked in three layers (Fig. 1*b*). The bases on the two-purine side of the loop form G:G and A:A pairs with two of the five purines on the other side, and these two purine–purine pairs stack on a base triple platform created by an interaction between another A in the loop and the minor groove face of a Watson-Crick G:C pair. The two remaining purines in the loop are bound to pockets of the L30 protein. The entire internal loop unit stacks on two Watson-Crick helices, creating a very distinctive architecture with a highly distorted backbone and a sharp bend (Fig. 2*a*). The L30 protein displays an αβα fold and interacts with the RNA via an edge of the β sheet, a portion of one helix, and three loops. The RNA–protein interface includes hydrophobic and aromatic stacking interactions to the bases, as well as specific hydrogen bonds typical of many RNA–protein complexes. The major grooves of the adjacent A-form RNA helices are widened and accessible to the protein, as is also commonly observed.

The L30–RNA complex is the second structure of a ribosomal protein bound to RNA resolved at atomic detail. The cocrystal structure of the L11 protein bound to a ribosomal RNA site has recently been reported[8,9] and again illustrates the intricacies of RNA folding and the importance of induced fit. In that case, four double-helical segments are stacked and folded into a single compact domain (Fig. 2*b*), stabilized by several base-specific tertiary contacts. A striking feature of the RNA structure is the extensive stacking of bulged and looped bases in the interior core of the RNA. One key docking interaction involves intercalation of an A to form a Hoogsteen pair with a U. This interaction helps establish the tertiary fold of the RNA and is stabilized directly by hydrogen bonding to the L11 protein, which makes base-specific contacts to the RNA largely from two α-helices and two loops. Additional interactions to 2' OH and phosphate groups of the RNA from these and other regions of the protein are used to recognize the highly distorted RNA backbone structure and form a tight RNA–protein interface.

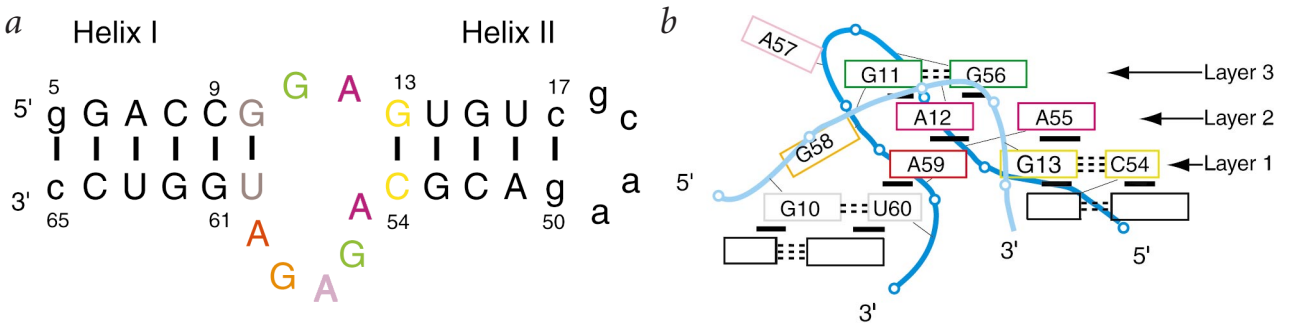

Fig. 1 ***a***, The RNA used in the NMR studies of the L30 pre-mRNA autoregulatory site complex. ***b***, Schematic drawing of the RNA internal loop regions (G10–G13 and C54–U60) derived from NMR. Rectangular boxes represent bases, blue lines with open circles represent the phosphate backbone, and thick black bars indicate aromatic stacking. The colors in (*a*) correspond to those in (*b*).

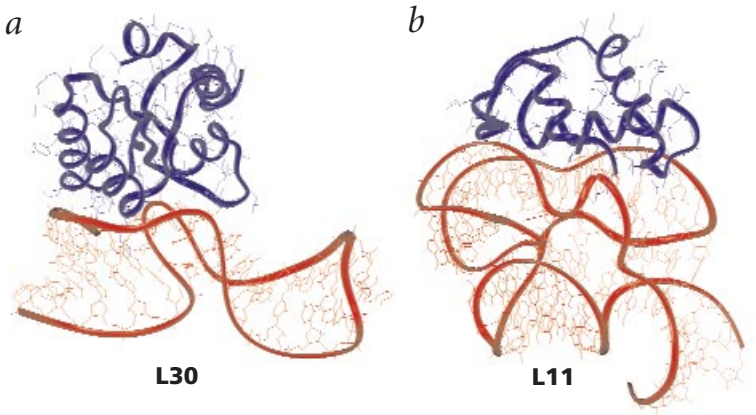

Fig. 2 ***a***, NMR structure of the *Saccharomyces cerevisiae* L30 ribosomal protein bound to its autoregulatory RNA site[1] and ***b***, crystal structure of the *Bacillus stearothermophilus* L11 ribosomal protein bound to its ribosomal RNA site[8]. The proteins are shown in blue and the RNAs in red.

Both the RNA and protein components of the L30 and L11 complexes undergo significant conformational changes or show substantial stabilization of poorly ordered parts of the structure upon binding. Thermal denaturation and NMR studies of the L30 and L11 RNA sites in the absence of protein suggest that the RNAs are relatively unstable and that many of the important RNA–RNA tertiary interactions observed in the bound complexes are sampled only sporadically in the free RNAs[1,10]. NMR and CD spectroscopy of the L30 protein in the absence of RNA indicate that, although the protein is folded globally, part of the α-helix and the three loops principally involved in RNA recognition are poorly ordered and dynamic[11]. These regions become ordered upon RNA binding, driven in part by the burial of exposed hydrophobic residues[1,11]. Similarly, NMR studies of the L11 protein show that part of an α-helix and a loop critical for recognition become ordered upon RNA binding[12,13]. Thus, both the L30 and L11 RNA complexes display characteristics that may be described as 'mutually-induced fit' or 'cofolding'. The L30 complex in particular involves costabilization of both an RNA loop and a protein loop.

Structural flexibility and induced fit seem to be common themes in RNA–protein interactions. It has been suggested that a subset of ribosomal proteins taken out of the ribosomal context is inherently unstable and requires the scaffolding of the RNA and/or other proteins to provide the correct folding environment[14]. The structures of many isolated ribosomal proteins have been determined (reviewed in refs 15 and 16) and some indeed show substantial flexibility or disorder. Like L30 and L11, the flexible parts tend to be localized to regions of expected interactions and often are found in loops. L9 is an interesting two domain protein thought to act as a 'molecular strut' in the ribosome, and amide proton exchange experiments indicate that basic amino acids and other likely RNA-binding residues located in loops of each domain (Fig. 3*a*) are among the least rigid in the molecule[17]. NMR experiments on the L25 protein indicate that it also contains a disordered loop that becomes structured upon binding to the E-loop of 5S ribosomal RNA[18]. The S7 protein contains a likely RNA-binding loop that adopts a well-ordered β-hairpin conformation in the crystal structure, but probably only as a result of strong crystal contacts[19,20] (Fig. 3*b*). The hairpin makes few interactions with the rest of the protein and is thought to be flexible in solution[19]. It has been suggested that a flexible loop in S5 might adopt a similar conformation[21]. From a very simplistic perspective, a β-hairpin may be viewed as an easy way to fold a protein loop, requiring just a few interactions to form across two extended strands provided that appropriate turn residues are present. Such folding can be induced by RNA binding, as seen with a 14-residue β-hairpin peptide from the bovine immunodeficiency virus Tat protein[22–24] (Fig. 3*c*). RNA-induced folding of L30 and L11 also involves stabilization of helices, as seen with RNA-binding peptides from the human immunodeficiency virus Rev and bacteriophage λ and P22 N proteins (reviewed in refs 25 and 26).

Induced folding or stabilization of unusual RNA structures, as observed in the L30 and L11 complexes, is characteristic of most, if not all, RNA–protein complexes studied to date. In some cases, such as the U1A protein–RNA complex-

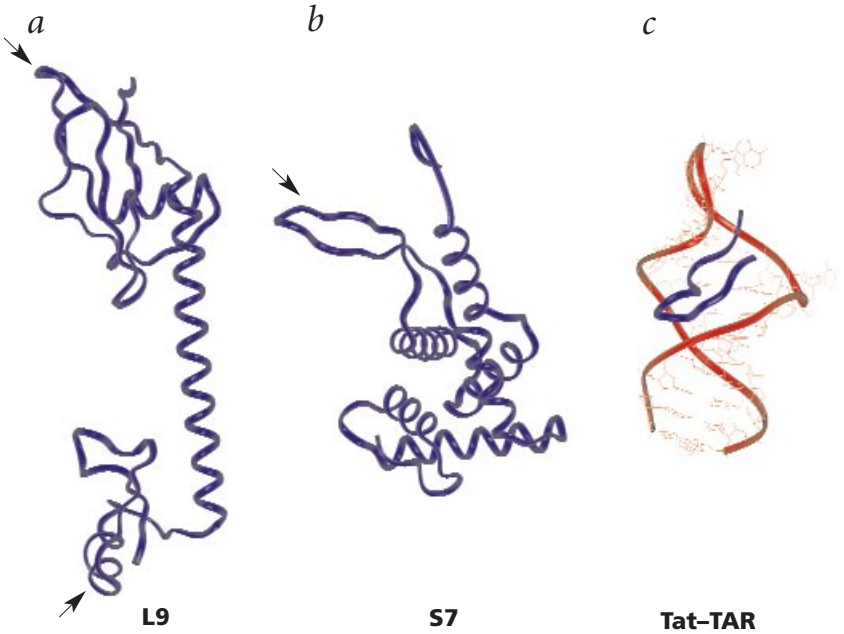

Fig. 3 ***a***, Crystal structure of the *Bacillus stearothermophilus* L9 ribosomal protein[34], ***b***, crystal structure of the *Thermus thermophilus* S7 ribosomal protein[19], and ***c***, NMR structure of the bovine immunodeficiency virus Tat RNA-binding peptide bound to TAR RNA[23]. The proteins or peptide are shown in blue and TAR RNA in red. Arrows indicate flexible regions in the L9 structure and the β hairpin in the S7 structure. The Tat peptide is unstructured in the absence of RNA and forms a β hairpin upon binding[22–24].

es[27,28], bases from unstructured loops are bound within protein pockets and use few RNA–RNA interactions to stabilize the RNA structure. In other cases, such as the L30 and L11 complexes, formation of the RNA architecture involves many stabilizing RNA–RNA interactions but still uses protein interactions to help stabilize the folded state. Placing an RNA in the context of surrounding charged and hydrophobic amino acids from a protein clearly creates a very different environment for RNA folding, and thus induced fit of the RNA is likely to be the rule in RNA–protein complexes.

Mutually-induced fit or cofolding may be particularly relevant to the assembly of large and dynamic ribonucleoprotein complexes, such as the ribosome, or to allow RNA-binding proteins to adapt to multiple binding sites, including autoregulatory sites in mRNAs. As large complexes assemble, or when conformational states change (during translation, for example), proteins must be sufficiently flexible to bind in the context of the changing environment. Coupling folding to binding provides one mechanism to adapt to these changes. As with L30, many ribosomal proteins bind to autoregulatory sites in addition to their ribosomal site and thus they must have the ability to recognize more than one site. In the simplest cases this is achieved through molecular mimicry in which the same RNA structure is used in both settings; however, this may not always be the case (reviewed in ref. 2). The S15 protein, for example, recognizes a three-way junction in ribosomal RNA and a pseudoknot in its mRNA. Because the structures of S15 bound to these two RNA sites are not yet known it is not clear whether the binding sites and protein interactions are similar or different. However, the structure of the S15 protein shows a flexible loop at the ends of two helices suggestive of a potentially adaptable RNA-binding region[29,30]. In the case of L30 there is no obvious ribosomal RNA site analogous to the internal loop of the regulatory site. Interestingly, an L30 mutant has been identified that eliminates binding to the regulatory site but still allows incorporation of L30 into ribosomes, suggesting that recognition of the two RNA sites may differ[31]. However the mutation is not located at the RNA-binding interface and it is possible that binding to the isolated regulatory site is sensitive to perturbation of the protein structure whereas binding to ribosomal RNA, which may rely on interactions with other proteins, is not. Further studies are needed to determine whether L30 complexed to the regulatory site mimics the ribosomal RNA interaction.

There are many cases in which folding accompanies complex formation in other macromolecular systems, including protein–protein interactions and protein–DNA interactions[32], and these can involve folding of just one partner or cofolding of both partners. In the case of the nuclear receptors, for example, folding of the retinoic acid receptor dimerization domain is induced upon DNA binding, generating different heterodimer interfaces depending on the DNA site recognized, thus allowing the protein to adapt to multiple binding partners[33]. Mutually-induced fit or cofolding in RNA–protein complexes may be considered analogous to the assembly of protein oligomers in which protein interfaces must mutually adapt, often via burial of exposed hydrophobic surfaces. In the case of RNA–protein complexes, the interaction happens to involve two different types of macromolecules. The structures of the L30–RNA and L11–RNA complexes serve to further emphasize the fascinating diversity of RNA architecture and the similarities to protein structure.

Alan D. Frankel is in the Department of Biochemistry and Biophysics, University of California San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0448, USA. email: frankel@cgl.ucsf.edu

1. Mao, H., White, S.A. & Williamson, J.R. *Nature Struct. Biol* **6**, 1139–1147 (1999).
2. Springer, M., Portier, C. & Grunberg-Manago, M. In *RNA structure and function* (eds. Simons, R.W. & Grunberg-Manago, M.) 377–413 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1998).
3. Eng, F.J. & Warner, J.R. *Cell* **65**, 797–804 (1991).
4. Vilardell, J. & Warner, J.R. *Genes Dev.* **8**, 211–220 (1994).
5. Li, B., Vilardell, J. & Warner, J.R. *Proc. Natl. Acad. Sci. USA* **93**, 1596–1600 (1996).
6. Li, H., Dalal, S., Kohler, J., Vilardell, J. & White, S.A. *J. Mol. Biol.* **250**, 447–459 (1995).
7. Li, H. & White, S.A. *RNA* **3**, 245–254 (1997).
8. Conn, G.L., Draper, D.E., Lattman, E.E. & Gittis, A.G. *Science* **284**, 1171–1174 (1999).
9. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. *Cell* **97**, 491–502 (1999).
10. Xing, Y. & Draper, D.E. *J. Mol. Biol.* **249**, 319–331 (1995).
11. Mao, H. & Williamson, J.R. *J. Mol. Biol.* **292**, 345–359 (1999).
12. Markus, M.A., Hinck, A.P., Huang, S., Draper, D.E. & Torchia, D.A. *Nature Struct. Biol.* **4**, 70–77 (1997).
13. Hinck, A.P. *et al. J. Mol. Biol.* **274**, 101–113 (1997).
14. Yonath, A. & Franceschi, F. *Nature Struct. Biol.* **4**, 3–5 (1997).
15. Draper, D.E. & Reynaldo, L.P. *Nucl. Acids Res.* **27**, 381–388 (1999).
16. Ramakrishnan, V. & White, S.W. *Trends Biol. Sci.* **23**, 208–212 (1998).
17. Lillemoen, J., Cameron, C.S. & Hoffman, D.W. *J. Mol. Biol.* **268**, 482–493 m(1997).
18. Stoldt, M., Wohnert, J., Gorlach, M. & Brown, L.R. *EMBO J.* **17**, 6377–6384 (1998).
19. Wimberly, B.T., White, S.W. & Ramakrishnan, V. *Structure* **5**, 1187–1198 (1997).
20. Hosaka, H. *et al. Structure* **5**, 1199–1208 (1997).
21. Davies, C. *et al. J. Mol. Biol.* **279**, 873–888 (1998).
22. Chen, L. & Frankel, A.D. *Proc. Natl. Acad. Sci. USA* **92**, 5077–5081 (1995).
23. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. *Science* **270**, 1200–1203 (1995).
24. Ye, X., Kumar, R.A. & Patel, D.J. *Chem. Biol.* **2**, 827–840 (1995).
25. Patel, D.J. *Curr. Opin. Struct. Biol.* **9**, 75–87 (1999).
26. Narayana, N. & Weiss, M.A. *Biopolymers* **48**, 167–180 (1999).
27. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & Nagai, K. *Nature* **372**, 432–438 (1994).
28. Allain, F.H.-T., Howe, P.W.A., Neuhaus, D. & Varani, G. *EMBO J.* **16**, 5764–5772 (1997).
29. Berglund, H., Rak, A., Serganov, A., Garber, M. & Hard, T. *Nature Struc. Biol.* **4**, 20–23 (1997).
30. Clemons, W.M., Jr., Davies, C., White, S.W. & Ramakrishnan, V. *Structure* **6**,429–438 (1998).
31. Vilardell, J. & Warner, J.R. *Mol. Cell. Biol.* **17**, 1959–1965 (1997).
32. Spolar, R.S. & Record, M.T., Jr. *Science* **263**, 777–784 (1994).
33. Rastinejad, F., Perlmann, T., Evans, R.M. & Sigler, P.B. *Nature* **375**, 203–211(1995).
34. Hoffman, D.W. *et al. EMBO J.* **13**, 205–212 (1994).