Cell, Vol. 92, 149–151, January 23, 1998, Copyright ©1998 by Cell Press

Induced Folding in RNA–Protein Recognition: More than a Simple Molecular Handshake

Minireview

Alan D. Frankel and Colin A. Smith
Department of Biochemistry and Biophysics
University of California, San Francisco
513 Parnassus Avenue
San Francisco, California 94143-0448

When two macromolecules interact, one or both partners often undergo structural rearrangements to establish a complementary binding interface. Such induced fit interactions (sometimes called adaptive binding) are quite diverse, ranging from fine adjustments of a few atoms to large-scale folding or unfolding reactions to major domain rearrangements. In extreme cases, two entirely unfolded molecules can require the interaction to stabilize their structures, for example in forming protein–protein dimers. The energetic consequences of induced fit binding are similarly diverse; for example, locking a few side chains into particular conformations may have little or no energetic cost whereas remodeling an entire segment of protein secondary structure can substantially reduce the binding constant relative to a rigid molecular interaction. This simplistic view neglects changes in solvation at the binding interface, which can have major entropic effects. Induced fit appears to be a common theme in RNA–protein interactions, and two recent papers in *Molecular Cell* (Zheng and Gierasch, 1997; Mogridge et al., 1998) provide rather striking examples of how the RNA component of the complex can induce structure in a disordered or partially disordered protein. One tantalizing possibility from these papers is that the RNAs may not serve simply as molecular scaffolds for folding but also may influence protein function.

The λ *N*–boxB *Interaction*

In the first paper, Mogridge et al. (1998) identified three functional domains of the bacteriophage λ N protein, a transcriptional antiterminator, that individually interact with *E. coli* RNA polymerase (RNAP), the NusA protein, or *boxB* RNA of the *nut* (N utilization) site. These proteins, along with NusB, NusG, and S10, cooperatively assemble on the *nut* site to form a functional antitermination complex. NMR experiments by Mogridge et al. (1998) indicate that N is entirely disordered on its own (also shown by Van Gilst et al., 1997) and that upon binding to *boxB* RNA, only the amino-terminal RNA-binding domain becomes structured. It is presumed that interactions with RNAP and NusA similarly induce folding of the other domains. The RNA-binding domain of N is localized to a 22–amino acid arginine-rich region, and studies with model peptides indicate that the isolated domain binds *boxB* in an α-helical conformation (Tan and Frankel, 1995; Su et al., 1997a). The helical structure of the peptide is stabilized upon specific RNA binding, probably forming a bent α-helical structure (Su et al., 1997a), and a corresponding amount of helix is induced in intact N, as indicated by circular dichroism (CD) (Van Gilst et al., 1997). Induced α helix formation has been observed upon protein binding to specific DNA sites, most notably in the bZIP family of proteins in which the basic region becomes structured upon binding (Spolar and Record, 1994). On the other side of the N–*boxB* interaction, the structure of the RNA also becomes more ordered upon binding, including stabilization of a G:A base pair in the terminal loop (Su et al., 1997b; Mogridge et al., 1998). Thus, the interaction shows costabilization of the structure of both partners, analogous to that observed with an α-helical HIV-1 Rev peptide-RRE RNA complex (Tan and Frankel, 1994).

Induced folding is clearly an important aspect of the recognition process per se, but the papers by Mogridge et al. (1998), Van Gilst et al. (1997), and Su et al. (1997b) also suggest interesting possibilities for biological function. Previous models for antitermination have suggested that the N–*boxB* interaction is required simply to link the antitermination complex to the nascent transcript through an RNA-looping mechanism. It now seems possible that the RNA plays a more active role. Both Mogridge et al. (1998) and Van Gilst et al. (1997) have shown that N is entirely unfolded in vitro, and it is known that N is rather unstable in *E. coli* and is actively degraded by the lon protease. Thus, both papers speculate that the unfolded state of N may be monitored by the cell; proteases are proposed to limit the N concentration and thereby prevent *boxB*-independent antitermination that can occur at high protein concentrations. Other λ proteins involved in establishing the lysogenic state, including the cI and cII proteins, are under the control of cellular proteases, indicating the likely importance of such mechanisms in switching between lytic and lysogenic phage growth. RNA-induced folding of N may at least partially protect N from degradation, and Mogridge et al. (1998) suggest that partial proteolysis might leave *boxB* occupied by the helical RNA-binding domain, preventing other intact N molecules from binding and thereby blocking antitermination. Folding of other regions of N might be induced by the binding of protein ligands, such as RNAP or NusA, and also may modulate protease sensitivity, but whether RNA or other ligand-induced stabilization is actually used to regulate protein stability in vivo is not yet known. On the other side of the interaction, the N-induced folding of *boxB* RNA may effectively act as a biological "switch," providing an organized RNA platform for recognition by NusA (Su et al., 1997b; Mogridge et al., 1998).

The Ffh-4.5S RNA Interaction

In the second paper, Zheng and Gierasch (1997) describe how binding of 4.5S RNA to the M domain of Ffh, a component of the signal recognition particle (SRP) in *E. coli*, stabilizes Ffh structure. Ffh (the *f*ifty-*f*our kDa *h*omolog of the mammalian SRP54 protein) is comprised of two rather distinct domains, the M (methionine-rich) domain, which binds the 4.5S RNA scaffold as well as the signal sequence, and an amino-terminal NG domain, which is a GTPase involved in regulating binding to the SRP receptor (FtsY in *E. coli*). The mammalian SRP counterpart contains additional proteins, with less well-defined functions, that decorate the 7S RNA scaffold. Zheng and Gierasch (1997) used protease mapping, CD,

and NMR experiments to show that the Ffh M domain by itself has characteristics of a molten globule, with high α-helical content but noncooperative melting behavior and high sensitivity to proteases. Upon specific RNA binding, the domain becomes stably folded, showing cooperative melting and resistance to proteases. Again, changes in RNA conformation are observed upon complex formation (Lentzen et al., 1996), providing yet another example of mutually induced fit in RNA–protein recognition.

As with the N protein, Ffh has multiple functional domains and binds protein ligands in addition to RNA. In this case, however, binding of signal peptides actually destabilizes the Ffh structure, causing enhanced protease susceptibility of the NG domain, either in the context of intact Ffh or as an isolated domain. Interestingly, binding of 4.5S RNA to the M domain protects the tethered NG domain from signal peptide-induced unfolding. Thus, RNA binding appears to stabilize indirectly a second domain of Ffh, and it is proposed that the M domain–RNA complex, and not the M domain alone, forms the proper docking surface for the NG domain when signal peptide is bound.

While the results of Zheng and Gierasch (1997) clearly demonstrate that 4.5S RNA protects Ffh from proteolytic digestion in vitro, it is unlikely that the unfolded state is a target for proteases in vivo because, unlike the N protein, Ffh is constituitively and stoichiometrically assembled into particles. In vivo it is known that expression of 4.5S RNA stabilizes the Ffh protein, but this may be a general feature of proteins that assemble into complexes and may not reflect an important regulatory step. Nevertheless, the RNA may still play an active role in SRP function, for example by communicating signal peptide binding to the translational apparatus, perhaps by sensing conformational unfolding of the NG domain. The results demonstrating signal peptide-induced unfolding of the NG domain were somewhat unexpected because earlier cross-linking experiments suggested that signal peptide binding was restricted to the M domain (at least for SRP54), and probably to a region of "methionine bristles" lining a helical hydrophobic binding pocket (see Zheng and Gierasch, 1997). It remains to be determined whether signal peptides bind to a specific site on the NG domain to induce unfolding or whether they act as relatively nonspecific hydrophobic denaturants, but it seems reasonable that signal peptides may contact surfaces on both the M and NG domains, thereby further contributing to interdomain communication.

The Induced Fit Continuum

The N and Ffh papers highlight the functional importance of induced fit interactions, which, as mentioned above, can be quite diverse. In DNA–protein interactions, there are many examples in which the DNA site remains relatively fixed in structure, often quite close to a B-form helix, and a protein is remodeled upon binding. The energetic consequences of protein remodeling have been compared in detail to "rigid body" associations, and thermodynamic "signatures" have been identified (Spolar and Record, 1994). In one extreme case of remodeling, the BamHI endonuclease undergoes large-scale domain rearrangements upon DNA binding, folding of disordered segments, restructuring of other segments, and even an unfolding of α helices to form extended arms that bind in the minor groove (Newman et al., 1995). Such large conformational changes appear to be important to ensure the fidelity of cleavage and to properly position active site residues. In another extreme case, the nuclear receptors are observed to bind as two independent monomers to adjacent DNA sites, using the DNA as a surface to mold the dimer interface (Rastinejad et al., 1995). Remarkably, the dimerization interface can be molded differently depending on the type of receptor bound and the spacing between adjacent sites. This type of induced fit allows for combinatorial usage of receptor proteins at different sites. There also are many cases in which the DNA becomes severely distorted upon binding, most notably resulting in bent DNA complexes or extrusion of bases from the DNA helix.

In the RNA–protein complexes studied to date, both partners have been observed to rearrange simultaneously or become stabilized. As mentioned above, the α-helical conformation of an HIV-1 Rev peptide is stabilized upon interaction with the RRE, together with formation of two purine–purine base pairs in the RNA. In a BIV Tat peptide-TAR complex, the peptide undergoes a transition from a completely unfolded state to a β hairpin conformation, together with formation of a base triple in the RNA (Sundquist, 1996). tRNA synthetases show a variety of conformational changes; in the phenylalanyl tRNA synthetase, for example, a disordered amino-terminal region forms a long coiled-coil helical domain upon binding (Goldgur et al., 1997). Ribosomal proteins also have been observed to undergo folding transitions upon RNA binding, and many are thought to be critically dependent on the ribosomal RNA scaffold to adopt a defined structure (Yonath and Franceschi, 1997), much like the case of SRP. Indeed, at least one ribosomal protein has characteristics of a molten globule (Zurdo et al., 1997). Given that RNA can help reorganize and mold protein structure, it would be interesting if RNAs also are found to catalyze folding reactions, analogous to the role that certain prodomains of proteases play in resolving trapped kinetic intermediates (Baker and Agard, 1994).

The wide diversity of RNA rearrangements in RNA–protein interactions, for example the ordering of loop nucleotides within binding pockets of the U1A protein or tRNA synthetases (Allain et al., 1996; see Goldgur et al., 1997), reflects the wide diversity of RNA structure itself. Because alternative RNA structures often can be energetically as stable as, or even more stable than, the folded state, and because misfolded structures may be kinetically trapped, proteins can play crucial roles in RNA folding. Protein-assisted RNA folding, in which structures are stabilized through specific protein binding, and chaperone-mediated RNA folding, in which nonspecific binding proteins are used to resolve misfolded species, recently have been reviewed (Herschlag, 1995; Weeks, 1997). Both the N–*boxB* and Ffh–4.5S RNA interactions provide examples of protein-assisted RNA folding, and in each case the reorganization forms a new recognition surface, for NusA or the NG domain, respectively. Recognition of a preformed protein–nucleic acid surface has been seen in ternary complexes between TBP, TATA-box DNA, and TFIIB or TFIIA in which

the bent DNA–protein surface formed by TATA-TBP is subsequently recognized by TFIIB or TFIIA (Nikolov and Burley, 1997).

The Benefit of an Unfolded State: Flexibility

Finally, we consider a few possible advantages of using unfolded or partially folded proteins to recognize RNA, though they are by no means restricted to RNA–protein interactions. As described for N, maintaining proteins in an unfolded state provides the opportunity to monitor the bound (or functional) status of a protein, using proteases to regulate protein concentration. Recognition of a disordered protein may allow interactions with multiple partners, each appropriately molding a binding surface to its own needs, as described for nuclear receptor dimerization and as is thought to occur with disordered activation domains of transcription factors. Induced folding or the ability to remodel surfaces can facilitate the ordered addition of components or allow signaling of binding events to other partners; for example, both the N–*boxB* and Ffh–4.5S RNA complexes provide new recognition surfaces for other proteins or domains. This feature is expected to be especially important for large ribonucleoprotein complexes such as the ribosome or spliceosome, in which many components must be accurately assembled or ordered events must take place. The ability to remodel also might facilitate ordered exchanges with multiple partners, as might occur in pre-mRNA splicing, and the disordered state itself might provide important nonspecific RNA contacts at appropriate times. Some interactions are topologically impossible using two rigid surfaces, and full specificity may be achieved only through induced fit mechanisms. For example, an RNA site cannot be surrounded using a preformed protein structure; in DNA–protein complexes, encircling often is achieved using flexible protein arms. From an evolutionary perspective, many sequences are expected to form disordered domains and some will contain adaptable interaction surfaces. During the transition from an RNA world to a protein-based world, flexible or even unfolded structures might well have provided adequate affinities and specificities to evolve the necessary functions.

Selected Reading

Allain, F.H., Gubser, C.C., Howe, P.W., Nagai, K., Neuhaus, D., and Varani, G. (1996). Nature *380*, 646–650.

Baker, D., and Agard, D.A. (1994). Biochemistry *33*, 7505–7509.

Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997). Structure *5*, 59–68.

Herschlag, D. (1995). J. Biol. Chem. *270*, 20871–20874.

Lentzen, G., Moine, H., Ehresmann, C., Ehresmann, B., and Wintermeyer, W. (1996). RNA *2*, 244–253.

Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E., and Greenblatt, J. (1998). Mol. Cell *1*, 265–275.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1995). Science *269*, 656–663.

Nikolov, D.B., and Burley, S.K. (1997). Proc. Natl. Acad. Sci. USA *94*, 15–22.

Rastinejad, F., Perlmann, T., Evans, R.M., and Sigler, P.B. (1995). Nature *375*, 203–211.

Spolar, R.S., and Record, M.T., Jr. (1994). Science *263*, 777–784.

Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A., and Weiss, M.A. (1997a). Biochemistry *36*, 12722–12732.

Su, L., Radek, J.T., Labeots, L.A., Hallenga, K., Hermanto, P., Chen, H., Nakagawa, S., Zhao, M., Kates, S., and Weiss, M.A. (1997b). Genes Dev. *11*, 2214–2226.

Sundquist, W.I. (1996). Nat. Struc. Biol. *3*, 8–11.

Tan, R., and Frankel, A.D. (1994). Biochemistry *33*, 14579–14585.

Tan, R., and Frankel, A.D. (1995). Proc. Natl. Acad. Sci. USA *92*, 5282–5286.

Van Gilst, M.R., Rees, W.A., Das, A., and von Hippel, P.H. (1997). Biochemistry *36*, 1514–1524.

Weeks, K.M. (1997). Curr. Opin. Struc. Biol. *7*, 336–342.

Yonath, A., and Franceschi, F. (1997). Nat. Struct. Biol. *4*, 3–5.

Zheng, N., and Gierasch, L.M. (1997). Mol. Cell, *1*, 79–87.

Zurdo, J., Sanz, J.M., Gonzalez, C., Rico, M., and Ballesta, J.P.G. (1997). Biochemistry *36*, 9625–9635.