Biochemistry

 Volume 40, Number 27 July 10, 2001

Current Topics

Current Topics in RNA−Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture†

Nicolas Leulliot and Gabriele Varani*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Received April 4, 2001; Revised Manuscript Received May 4, 2001

When RNA and protein bind each other, recognition occurs almost invariably by "induced fit" rather than by rigid "lock-and-key" docking (*1*, *2*). The protein, the RNA, and sometimes both undergo large conformational changes, and minor populated species are captured during complex formation, leading to large changes in the local as well as macroscopic properties of the interacting components. Although these conformational transitions have been observed in almost all RNA−protein complexes studied to date, no unifying picture has emerged concerning their implication in the affinity and specificity of the interaction or the functional role of conformational adaptability. Recent advances in the characterization of RNA−protein complexes and the application of concepts derived from protein folding and protein recognition studies are providing new insight into RNA−protein recognition.

Induced Fit and Conformational Capture in RNA−Protein Recognition

The concept of a folding funnel derived from protein folding studies provides useful insight into different mechanisms of intermolecular recognition (*3*). In its original context, the funneling concept discards the notion of a single folding pathway in favor of parallel processes where an ensemble of conformations coexist and go through a funnel composed of several transition state intermediates. In this description, the conformation of the hydrophobic core of a protein or that of interaction surfaces can be described in terms of the energy landscape of the conformations accessible to the protein itself (Figure 1) (*3*). The mechanism by which recognition occurs is determined by the properties of the bottom of the folding funnel. A smooth funnel with a single minimum would represent a stable and rigid interface. A rugged bottom with several energy minima separated by low-energy barriers describes instead the situation where a protein or RNA interface is best described in terms of multiple conformers. If the ligand recognizes a specific conformation of the receptor, the occupancy of that conformational state will progressively increase during binding, while the populations of the other conformers present in the ligand-free state are depleted. If the state selected by the ligand is a minor population in the ligand-free state of the receptor, this is equivalent to the "conformational capture" of a minor populated conformer by the ligand. This process would often

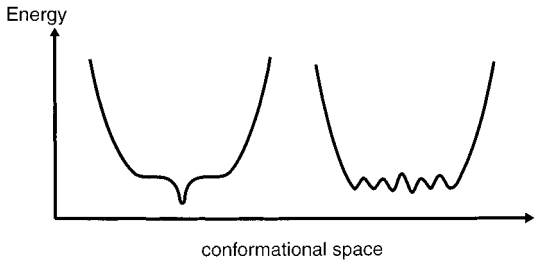

FIGURE 1: Schematic illustration of different energy landscapes for biological interfaces (*3*). A smooth funnel with a single minimum (left) would favor lock-and-key binding; a rugged bottom with low energy barriers separating multiple minima (right) would favor instead induced fit or conformational capture, should one of the low-energy states correspond to the bound conformation of the ligand.

† This work was supported by the Medical Research Council and by EMBO and a Marie Curie postdoctoral fellowship to N.L.

* To whom correspondence should be addressed: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K. Telephone: ++ 44-1223-402 417. Fax: ++ 44-1223-213 556. E-mail: gv1@mrc-lmb.cam.ac.uk.

10.1021/bi010680y CCC: $20.00
Published on Web 06/13/2001

be described as induced fit, but if the conformation present in the bound form pre-exists in the free form, even in a small fraction of the population, then binding is a special example of "tertiary structure capture". In this situation, the balance between binding interfaces characterized by broad specificity and highly selective interfaces depends to a significant extent on the energy landscape of the protein and RNA interfaces in the absence of the interacting partners (Figure 1).

RNA and proteins have fundamental differences with respect to the principles underlying the stability of their structural architecture (*4*). Of course, RNA has only four monomers of similar chemical nature, whereas proteins contain 20 amino acids of varied structure and different properties (hydrophobic, hydrophilic, polar, etc.). Another difference lies in the hierarchical nature of RNA folding; the energy involved in the formation of RNA secondary structure is substantially larger than the energy of RNA tertiary structure formation (*4*). With few exceptions (*5*), secondary structure is therefore formed before tertiary structure formation, which does not usually influence secondary structure formation. By contrast, formation of protein secondary and tertiary structures are generally intimately linked, making protein folding the result of a complex balance of the energies associated with secondary and tertiary structure formation.

The strongly hierarchical nature of RNA folding implies that RNA secondary structure (helices, bulges, loops, and junctions) is generally preformed regardless of whether any tertiary structure or protein is present. This observation has an immediate corollary: many RNA-binding proteins recognize single-stranded nucleotides presented in the context of a stable secondary structure. Induced fit does not imply the disruption of RNA secondary structure, which is rare, but the reorganization of local elements of secondary structure, the formation of a defined structure for disordered single-stranded regions, and the stabilization of a defined three-dimensional conformation of RNA.

Since RNA tertiary structure is not very stable, it can be easily molded in response to binding of proteins or even divalent metal ions. Several ribozymes, such as group I self-splicing introns, are inactive in the absence of protein cofactors that stabilize the active conformation of the enzyme through tertiary structure capture (*6*−*8*). RNA folding is characterized by rugged free energy landscapes and kinetic traps (*9*) which can be seen as intermediate states in RNA tertiary structure formation. The presence of tertiary elements that are not preorganized for docking can lead to the formation of alternative conformations that stabilize a nonactive domain. Sometimes, divalent metal ions suffice to induce tertiary structure reorganization in RNA. For example, the secondary structure of a three-way junction derived from group I self-splicing introns is notably different in solution and in the crystal structure, where the subdomain becomes part of its parent intron (*5*, *10*). The presence of Mg^{2+} rescues the folding of the domain, and the base-pairing pattern observed becomes similar to the crystal structure. Proteins and divalent metal ions can have the same effect. When S15 ribosomal protein binds RNA, there is a large conformational change in 16S ribosomal RNA upon binding not only of the protein but also of Mg^{2+} ions (*11*). The conformational equilibrium in the RNA is shifted by divalent ions toward the structure recognized by S15 protein.

Within proteins, poorly ordered loops between secondary structure elements are often remodeled upon RNA binding. Large-scale conformational changes are observed only when either molecule has several independent subunits, such as multiple protein domains or RNA helices anchored in multiple junctions. Since the relative orientation of these subunits is often weakly defined in the free molecule (*12*), it can be easily rearranged upon binding to the ligand.

Adaptive Binding in RNA−*Protein Recognition*

The last 10 years have unveiled many examples of RNA recognition by induced fit or tertiary capture, each of which contributes to illustrating how critical understanding this process is if we aim to understand how protein and RNA bind each other.

Tat−*TAR Interaction of HIV-1.* A first stunning example of the association of induced fit with RNA protein recognition was provided by the HIV-1 Tat−TAR system. Transcriptional elongation of the HIV-1 promoter is regulated through a mechanism dependent on the recognition of an RNA regulatory element, the transactivator response element (TAR) RNA, by the virally encoded transactivator protein Tat (*13*, *14*). This protein contains a region rich in basic amino acids, termed the basic domain, which mediates RNA recognition and is necessary, although not sufficient, for RNA binding. Studies with basic peptides derived from Tat protein (*15*, *16*) and, even more remarkably, earlier studies with a single arginine derivative (*17*), demonstrated that the bulge region of the TAR RNA binding site changes its conformation upon ligand binding.

The free and bound structures of HIV-1 TAR RNA represent two different ways of accommodating three bulged residues inside a RNA double helix (Figure 2). In the free RNA, U23 is looped out of the helix, and the two other bulged nucleobases assume continuous stacking interactions (*18*), inducing a kink in the RNA helix (*19*). In the bound RNA, a uracyl interacts instead with an arginine through hydrogen bonding and a cation$-\pi$ interaction (*15*). This conformational change relieves the helical twist and rise induced by the continuous stacking of the bulged residues and straightens the double helix. Mutational studies have shown that the most severe reduction in TAR RNA activity involves the two residues most critical to stabilization of the conformational change, leading to the conclusion that a specific ligand-induced RNA structure is critical for Tat binding. This is confirmed by the observation of "indirect readout" of the phosphate backbone charges defined by the tertiary structure of the RNA by the basic peptides (*20*, *21*). The structural consequences of the conformational change in TAR RNA are not just local. Studies of global properties of TAR have demonstrated that the bend of the RNA, a biologically important property, changes upon ligand binding (*22*). The functional consequences of these global conformational rearrangements in mediating binding of factors which interact with the loop region and mediate transactivation remain to be fully investigated.

A conformation much closer to the bound RNA is observed in free TAR in the presence of divalent metal ions under crystalline conditions (*23*). In the crystal, three calcium ions stabilize the trinucleotide bulge by providing electrostatic contacts that might drive the conformational transition

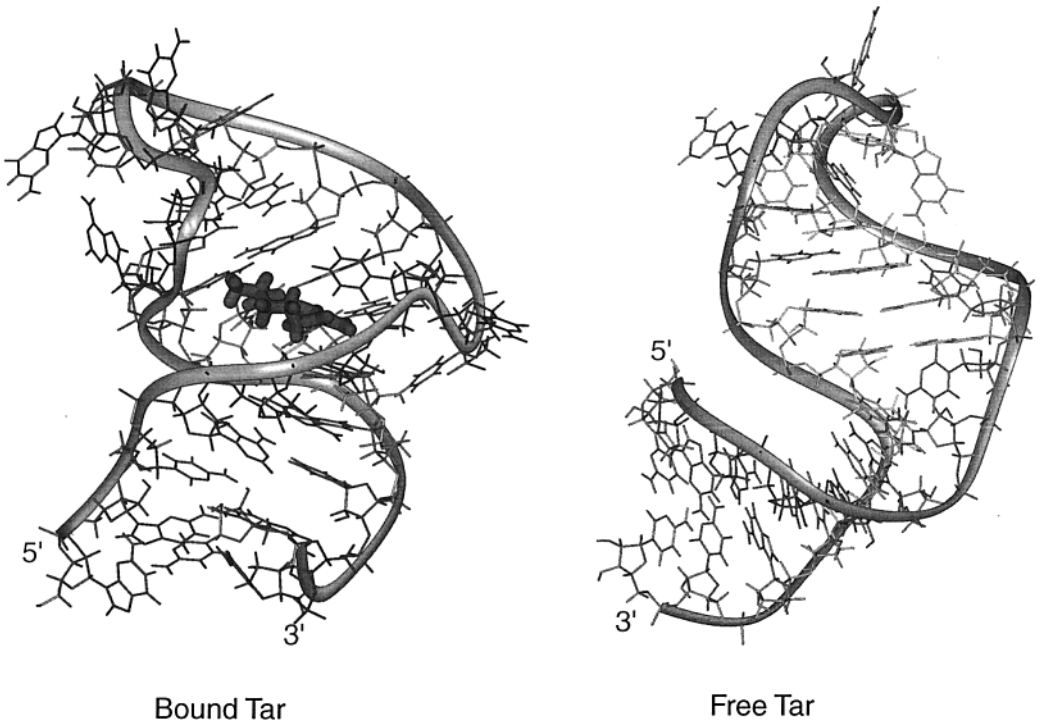

FIGURE 2: HIV-1 TAR RNA changes its conformation when peptides derived from Tat protein (*15*, *18*) or even a single-amino acid derivative argininеamide (*17*) binds to it.

like the highly basic region of the Tat peptide. In addition, crystal packing conditions might favor a straight double helix reminiscent of the bound TAR form rather than the kinked free TAR double helix. This observation and the NMR[1] studies of free TAR indicated that TAR RNA can adopt multiple conformations of comparable free energy separated by shallow free energy barriers in the absence of any ligand. Is it then possible that Tat peptides select one of these conformations upon binding by conformational capture, rather than inducing a completely new conformation of the RNA by induced fit?

Peptide−RNA Interactions: Basic Peptides and Aptamers. RNA aptamers selected by in vitro evolution methods for binding short basic peptides have provided interesting models of high-affinity RNA−peptide complexes. Small basic peptides have also been intensively studied as models for the RNA-binding basic domains of proteins such as Tat, Rev, and N from bacteriophage λ. The investigation of conformational rearrangements of structurally similar RNA binding sites was studied using RNA aptamers selected for binding closely related RNAs. While aptamers are selected on the basis of maximizing their affinity and specificity, natural RNA binding sites are selected under different evolutionary pressures. Thus, aptamers provide insight into "perfect" recognition, but biology seldom selects only for affinity, because it is difficult to regulate interactions that have very long off rates. Thus, RNA aptamer structures have explored a subspace of RNA recognition, but the unusual and novel binding modes discovered in these studies have revealed many new and important features of RNA recognition.

A first remarkable observation common to peptide−aptamer and basic peptide complexes with natural RNA substrates was that unstructured peptides acquire secondary structure upon RNA binding, and that the stabilization of peptide structure is critical for binding. A good example is provided by the HIV-1 Rev−RRE complex, a system controlling the export of the HIV-1 viral RNA from the nuclei of infected cells (*24*, *25*). Peptides derived from Rev protein and corresponding to the arginine-rich RNA binding domain of Rev, which are only partially helical in the free form, become stabilized in an α-helical conformation upon binding the RRE RNA (*26*, *27*). Even more remarkably, the structures of different RNA aptamers bound to the same Rev-derived peptide revealed different folds for the peptide itself (Figure 3), depending on the RNA with which it is interacting (*28*).

By examination instead of the Rev−response element (RRE) RNA and its conformation in the presence and absence of Rev-derived peptides, it is clear that the structures of the free and the bound RRE differ substantially (*29*, *30*). First, the major groove becomes open compared to that of regular A-form RNA. Second, the noncanonical G·G base pair adopts in the free RNA a symmetrical G(*anti*)·G(*syn*) conformation (*29*); after Rev has bound, the *syn* guanine flips to an *anti* conformation, resulting in a new base pair and the local reversal of the backbone chain (*30*). The accommodation of this conformational change is made possible by the conformational mobility allowed by a looped-out uracyl. Since this base can be replaced with a propyl linker with no effect on Rev binding, it acts as a spacer, providing the extra degrees of freedom to the phosphate backbone necessary to accommodate the peptide ligand. Since no contacts are made by the peptide with the G·G base pair, the role of these essential bases is to facilitate the interaction of the peptide with other RNA residues, probably through the widening of the major groove.

[1] Abbreviations: RRM, RNA recognition motif; NMR, nuclear magnetic resonance.

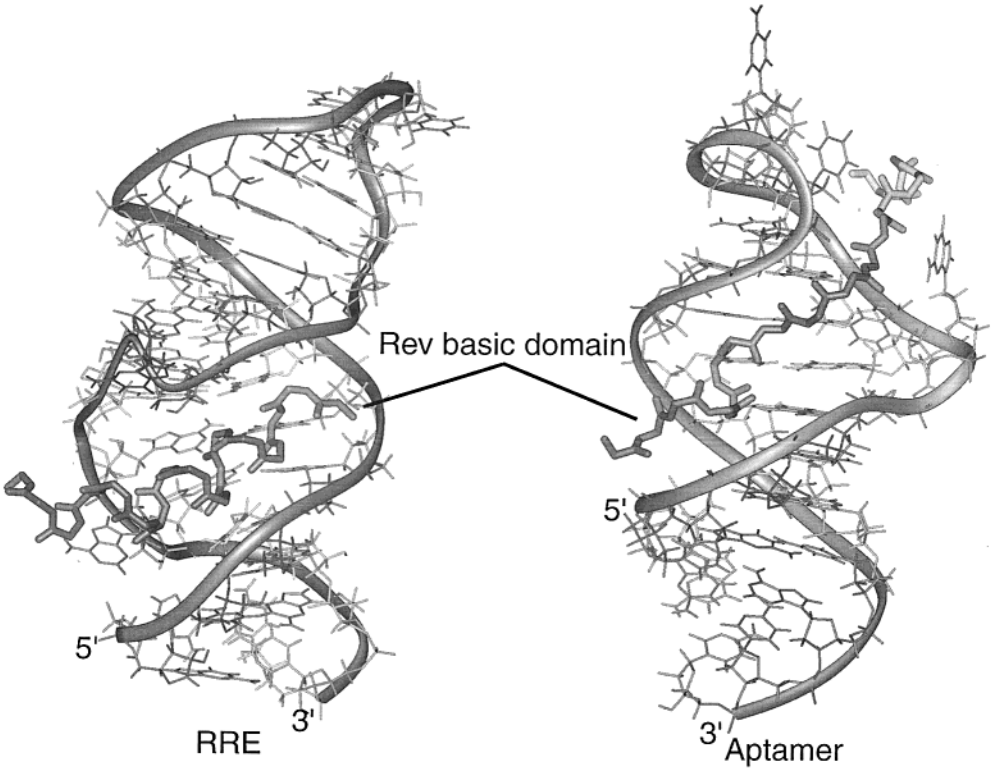

FIGURE 3: The same basic peptide can bind related RNAs in different conformations. A peptide corresponding to the basic domain of HIV-1 Rev protein binds to HIV-1 RRE in an α-helical conformation (left) (*27*) and a high-affinity aptamer sequence in an extended structure (right) (*28*).

In analogy with TAR, two crystal structures of the free RRE RNA showed structural variations in the conformation of the purine•purine base pairs (*31*, *32*). In contrast to the NMR structures, the G(*anti*)•G(*syn*) base pair is asymmetric in the crystal and mediated by a sodium ion. Two structural variants of the G•A base pair are present in one of these studies, one containing the NMR-observed G•A base pair and the other a novel water-mediated pair. This structural variability of the free RNA binding site is indicative of a metastable secondary structure and a rugged free energy landscape containing multiple conformations of comparable energy. This picture is consistent with many NMR observations, such as loss of intensity of NOE interactions and spectral broadening. In fact, the NMR data strongly suggest that the residues in the internal loop undergo conformational exchange between alternative conformations (*29*).

An additional excellent example of cofolding of RNA and peptides is provided by the bacteriophage λ N peptide−box B RNA complex (*33*), where even the full-length N protein is disordered in solution and remains disordered, with the exception of its RNA-binding domain in the presence of RNA (*34*). In each of these peptide−RNA structures, a disordered polypeptide gains secondary structure through intermolecular interactions with a specific RNA scaffold, to the same degree with which a protein folds into well-ordered secondary structure through tertiary interactions. It is even more striking that folding depends on the specific sequence of the RNA target, and is controlled by specific interactions between the peptide and RNA.

The N protein represents an extreme case of a totally unstructured protein, as opposed to unstructured loops or linkers inside a folded globular protein. The involvement of unfolded proteins in biological function, from transcriptional activation to membrane fusion, has recently been reviewed (*35*). Often these proteins participate in protein−protein interactions, with the binding sites becoming ordered upon complex formation. A particularly noteworthy example involves the KIX domain of the transcriptional activator CBP, which binds the kinase inducible domain (KID) of CREB and the transcriptional activation domain of c-Myb. Like N protein, KID is essentially unstructured in the free state but forms a pair of orthogonal α-helices upon binding to KIX, compensating for the unfavorable entropy through a very strong enthalpy contribution arising from the network of hydrogen bonds. In contrast, the transactivation domain of c-Myb is already folded into its α-helical binding structure in the free state, giving rise to a favorable entropic contribution, to the detriment of a weaker binding enthalpy, most probably because of the less optimal preformed binding surfaces.

Conformational Transitions in RNA Binding Sites Significantly Affect the Binding Energetics. The peptide−RNA complexes described in the previous paragraphs demonstrate that RNA recognition requires not only a cluster of specific nucleotides providing an array of chemical groups to be recognized. Rather, it also requires the RNA to refold around an unstructured peptide, inducing conformational changes in the RNA and the peptide and locking both ligands into more rigid conformations [cofolding (*36*)]. Under these circumstances, the energetics of binding is determined not only by the interactions observed in the bound state but also by the energetic penalties associated with conformational changes and folding transitions observed in either component. Since these are sequence-dependent and probably large (*37*), the conformational properties of unbound peptide and RNAs are just as important in determining affinity and specificity

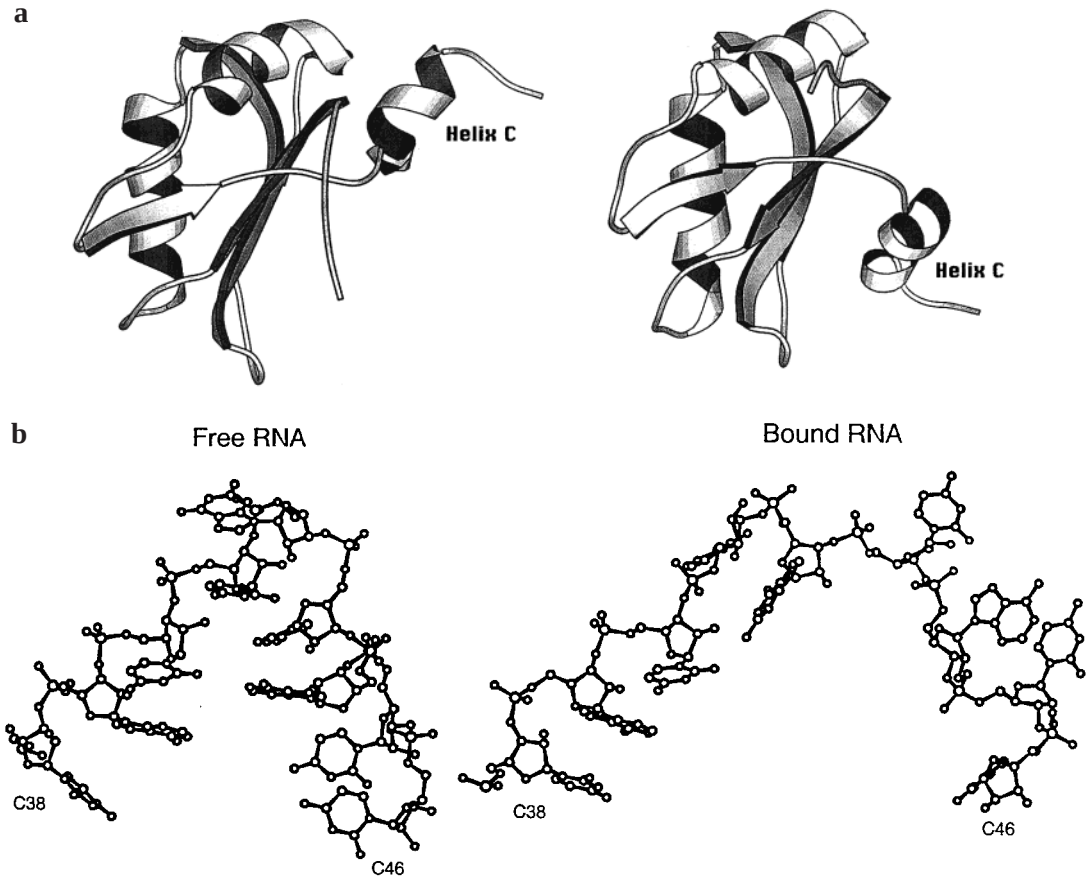

FIGURE 4: Conformations of U1A protein (*41*) (a) and its RNA substrate (*42*) (b) change significantly when they bind each other (*40*).

as the properties of the protein−RNA complex. As a consequence, it is difficult to interpret mutational data in RNA−peptide or RNA−protein complexes simply in terms of loss of individual intermolecular interactions. As discussed for the RRE−Rev system, residues critical for binding may not make any direct contact with the ligand, but simply define the correct conformation of the binding site. The residue might not even be directly involved in structure stabilization, but might favor the correct conformational change induced by the binding of the ligand or disfavor the incorrect conformation of the free state. For example, the looped-out uracyl in the Rev−RRE complex and two looped-out bases in TAR are critical for binding solely by allowing the conformational transitions to take place.

U1A and Other RRM Proteins. Peptide models have provided important insight into adaptive binding in protein−RNA recognition, but in the absence of a protein scaffold that is likely to restrain the conformation of the peptide, these have inherent flexibility and often bind with reduced specificity compared with the protein from which they are derived. These limitations could bias the conclusions of these studies. However, studies of bona fide, independently folded protein domains have provided a similar picture of the importance of induced fit in RNA recognition.

The RRM (RNA recognition motif, also known as the RNP domain or RBD) is the most common RNA binding module and one of the most common protein folds in eukaryotic genomes (*38*). The variety of RNA substrates and protein members of the superfamily provide a unique opportunity to study the basis of RNA recognition and specificity. The first structure of an RRM−RNA complex to be elucidated was the N-terminal fragment of human U1A protein bound to a stem−loop structure derived from U1 snRNA (*39*) and of the same protein domain bound to an internal loop regulatory element (*40*). An important element of recognition, relevant to this discussion, was the extent to which both the protein and RNA change structure upon formation of the complex. The protein undergoes a conformational change that repositions the C-terminal helix away from the RNA-binding surface (*41*), while the single-stranded RNA bases flip inside out and become involved in numerous interactions with the protein instead of other RNA bases (*42*) (Figure 4).

The major difference between the free and bound structures of U1A lies in helix C at the C-terminal end of the protein (*41, 42*). Residues immediately preceding this helix are involved in extensive interactions with three single-stranded nucleotides, but no direct contacts are made between helix C and the RNA. However, deletion of helix C abolishes specific RNA binding (*43*, *44*). The conformational rearrangement exposes the residues on the β-sheet for binding to the RNA. Intramolecular interactions involving helix C would in some respect act as a competitive ligand for the intermolecular RNA interactions and thus contribute to the specificity of the protein−RNA interaction, although it does not interact with RNA directly.

The conformational stability of loop 3, the region of the protein most critical for specificity, is also significantly altered upon RNA binding (*45*−*47*). In the free protein, this loop is flexible and may sample conformations that are captured by the RNA upon binding. In the final complex, the side chains of residues within loop 3 become rigid,

suggesting a correlation between restriction of motion and the regions of the intermolecular interface most important for specific binding (*45*). The loss of entropy with RNA binding must be compensated by a large number of favorable enthalpic contributions associated with specific RNA−protein contacts and possibly residual flexibility of residues from other regions of the protein.

When the U1A complex is examined instead from the RNA perspective, the unpaired nucleotides in the loop of the free RNA are also flexible in solution, but rigidify upon protein binding (*42*). Distinct changes between free and bound RNA are also observed in the pattern of base stacking interactions. In the free RNA, most of the single-stranded bases are oriented toward the inside of the loop, filling the cavity created by the sugar−phosphate backbone of the double-helical stem. The space in the cavity occupied by nucleobases in the free RNA is filled in the complex by protein residues from loop 3, while the RNA bases face the protein surface. Thus, the RNA undergoes at the same time a conformational change and a disorder−order transition upon binding, making it difficult to understand the sequence dependence of the binding energy.

Understanding the contribution of loop 3 and of the single-stranded nucleotides to RRM−RNA recognition is critical for understanding specificity in the RRM superfamily. Achieving this aim requires not only quantifying the energetics of intermolecular interactions and the conformational changes but also the disorder−order transition with its associated entropic costs. The observations described in the previous paragraphs indicate that affinity and specificity are defined by a delicate balance between entropic penalties and enthalpic gains and between rigidity and local conformational heterogeneity. On one hand, high specificity is obtained through rigid intermolecular interactions involving residues that are well-defined in the structure. On the other hand, the entropic cost for fixing certain degrees of freedom is alleviated by keeping residual flexibility in residues that are not key determinants for specificity. Thus, certain residues involved in intermolecular contacts can be mutated with a relatively small effect on binding, pointing toward a flexible interface capable of accommodating structural rearrangements. Flexibility of loop 3 could also allow different conformational states to be accessed during binding, providing a way of maximizing steric and ionic complementarity and functional group recognition. This optimization of interacting surfaces must occur not only in the final complex but also probably along the pathway from the free to the bound form of the reactants.

If these considerations were true, then induced fit would be more than a simple structural rearrangement of the side chains at the interface of the binding site, but would also involve an intricate balance of conformational rigidity and flexibility. Since the conformational changes occurring in the protein and RNA are large, any motion relevant to them will occur on a time scale slower than that which can be sampled by methods that detect motion on the nanosecond time scale, as is currently done by NMR and computational methods. In fact, the regions of U1A that are most important for RNA recognition move on a time scale of microseconds to milliseconds (*45*). Perhaps these motions within the protein, and similar motions in the RNA, lie on the pathway toward the conformation observed in the final complex and may provide favorable kinetic pathways for funnel complex formation. If this is the case, then mutations of certain nucleotides or amino acid side chains may affect binding not only by altering the conformation of the free components or disrupting key intermolecular interactions but also by altering the potential energy surface upon which the protein and RNA sample conformational space.

RNA-Induced Structural Reorganizations Contribute to the Specificity of Proteins Containing Multiple RRMs. While U1A represents a paradigm for RNA recognition by proteins utilizing a single RRM, several recent structures revealed how single-stranded RNA can be recognized by proteins containing two RRMs arranged in tandem (*48*−*51*). In each of these diverse structures, two structurally independent RRMs act synergistically to increase affinity and specificity for a specific single-stranded RNA sequence. Studies of RNA-free protein domains indicate that the domains are structurally independent in the absence of RNA (*12*, *52*). The binding surface is created by simultaneous interactions of the nucleic acid with both domains and the inter-RRM linker sequence whose conformation is stabilized by the RNA. In the poly(A) binding protein (Pab) complex (*49*), for example, the interface area defined by direct interactions between RRMs is only 550 Å^2, while the surface area buried by protein−RNA contacts is 2600 Å^2. Therefore, RNA−protein interactions are very likely important in defining the relative orientation of RRM1 and RRM2. In turn, this influences the ability of these proteins to bind different RNAs. The interdomain arrangement in the Pab−poly(A) complex (a flat extended surface) (*49*) is very different from those observed in other complexes involving recognition of pyrimidine residues (*48*, *51*) or of the complex of nucleolin with a stem−loop structure (*50*). The binding cleft between RRM1 and RRM2 is much narrower in the sex-lethal protein bound to a uridine-rich sequence than in the Pab protein bound to poly(A), allowing for discrimination based on the different size of purine and pyrimidine residues. Since the linker of the sex-lethal protein is not directly involved in RNA recognition (*48*), it might contribute to the affinity of the interaction only by providing a stable hinge when the two domains are correctly positioned. Thus, the relative conformation of the two RRMs, which is established through induced fit upon the interaction with RNA, is an important determinant of the specificity of proteins containing multiple RRMs.

tRNA Synthetases and the Control of Enzymatic Activity by RNA-Induced Conformational Changes. Perhaps the best examples of the importance of induced fit in the biology of RNA−protein complexes are provided by aminoacyl-tRNA synthetase (aaRS). These enzymes aminoacylate tRNAs before protein synthesis, a reaction whose accuracy (the correct amino acid must be attached to each tRNA) is essential for correct reading of the genetic code (*53*). If the tRNA is charged with the wrong amino acids, this will be incorporated in the protein sequence, resulting in miscoding (*54*, *55*). These enzymes often have two domains: an enzymatic domain that binds the acceptor end of the tRNA (*56*) and a second domain binding the anticodon region and therefore recognizing the distinct amino acid identity of each tRNA. Conformational rearrangements provide a mechanism for transmitting the information from the anticodon end of the tRNA to the acceptor end, which are more than 50 Å

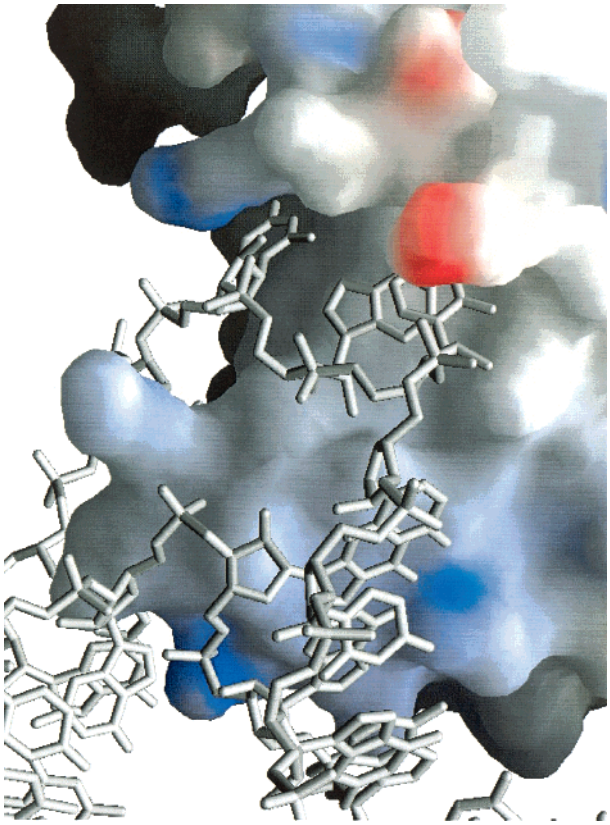

FIGURE 5: RNA adapts very unusual conformations when bound to proteins such as U1A (*39*, *40*). These conformations, often characterized by bases splayed out to the solvent, would not be stable unless the bases were deeply buried against the protein surface.

apart. This is illustrated in several examples, of which the recent crystal structure of arginyl-tRNA synthetase with its cognate $tRNA^{Arg}$ and in the presence or absence of the L-arginine provides a particularly compelling case (*57*). Comparison of the structure of free substrates, dimeric tRNA−aaRS complexes, and trimeric tRNA−aaRS−arginine complexes revealed distinctive local conformational changes upon binding. tRNA binding involves extensive structure adaptation for the recognition of the anticodon loop, which in turn triggers conformational changes in the active site of the enzyme, providing a more open crevice. A further tRNA-induced structural rearrangement of the ATP-binding site might be important for discrimination between different tRNAs. Arginine binding also induces an active conformation of the acceptor stem of the tRNA that is optimal for reactivity.

Conformational Capture, Induced Fit, and Residual Dynamics Modulate Affinity and Specificity of RNA−Protein Interactions

Intricate networks of intermolecular interactions between the protein and RNA functional groups form, leading to complexes of high affinity and selectivity, as ideally illustrated by the U1A structure (*39*, *40*) and by the structure of aptamers (*58*). It is instructive to observe what happens if the RNA or protein components of a complex are "removed", leaving only the counterpart visible (Figure 5**)**. When this representation of the structure is examined, it is difficult to imagine that the "naked" molecule would retain the same conformation in the absence of the binding partner. Since the nucleotide or amino acid sequence of an interface is not selected to form a stable structure, but for optimal binding, it is unlikely that a stable structure will be present in the absence of the cognate ligand. Either a new set of intramolecular interactions must form to compensate for the loss of intermolecular interactions, or the lowest-energy conformer will retain conformational flexibility, or both.

What are then the early steps in ligand recognition? An interesting possible mechanism invokes a combination of induced fit and rigid docking, with the target molecule being recognized first by rigid binding and then maximizing the interaction potential through induced fit, as suggested for a few protein−RNA complexes (*59*, *60*). The generality of this mechanism remains to be investigated.

Molecular dynamics studies of the U1A RNA−protein interface have provided quantitative insight into the dynamics of RNA−protein recognition (*61*). These simulations were aimed at studying the destabilization of the RNA−protein interface at high ionic strength. Cations compete with the protein for the RNA binding sites, and solvated ions subsequently destabilize residues at the interface. If microreversibility in molecular dynamics is assumed, then the simulation of the dissociation process would represent the reverse of the binding process and these studies will provide useful insight into the preorganization, solvation, and ion dependence of the interface. It has already been mentioned that metal ions can affect the structure of RNA and induce conformational transitions toward structures similar to those induced by proteins (*11*, *23*). It is widely accepted that water molecules and metal ions are an integral part of RNA structure (*62*), even more so in the exposed ligand-free start of binding interfaces. However, the energetic role of solvation and metal ions in controlling the conformation and dynamics of intermolecular interfaces is only beginning to be addressed in a quantitative fashion. A recent molecular dynamic simulation in explicit solvent of protein−protein interfaces has shed light on the dynamic nature of "misplaced" surface residues in the absence of the ligand. Key side chains frequently visit the rotamer conformation observed in the complex, pointing to a critical role of side chain−solvent interaction in recognition (*63*). Does this apply to RNA−protein interfaces as well?

Effect of Residual Conformational Flexibility on Specificity. Rigid docking provides a preorganized array of functional groups in well-defined spatial orientations that can be recognized by a ligand with high specificity. On the other hand, flexibility in the free reactants can maximize the affinity of an interaction by allowing optimal juxtaposition of interacting groups through conformational adaptation. In fact, complete interpenetration of the two binding surfaces, as often observed in RNA−protein complexes (Figure 6), is only possible if one or both of the binding partners undergo conformational rearrangements that are facilitated by conformational flexibility in the unbound state.

It appears that the flexibility of biological interfaces is controlled to subtly balance affinity and specificity. Fixing flexible residues in an interface requires a costly entropic contribution; too much flexibility in the unbound state is therefore detrimental to high-affinity binding. However, too little flexibility will stabilize a ground state that may not be recognized by the ligand. An estimate of the entropic costs

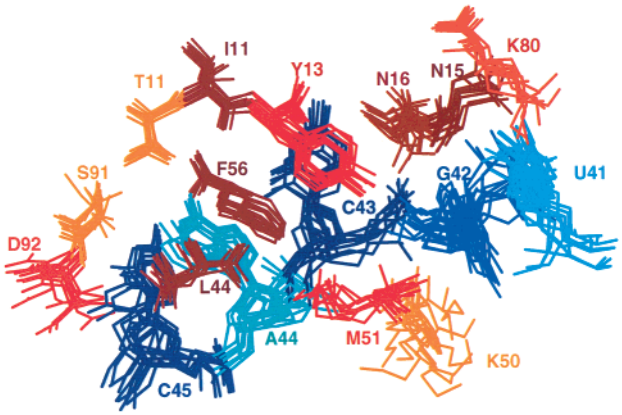

FIGURE 6: The U1A RNA−protein interface closely resembles a protein−protein interface, with considerable interpenetration and tight packing of RNA bases and amino acid side chains.

associated with induced fit interactions in protein−RNA recognition was provided by molecular dynamics studies that assessed the extent of residue motion in active and inactive variants of a peptide derived from a class I tRNA synthetase (*64*). Five residues were found to be essential for function, yet variants retaining these residues remained inactive. The simulation suggested that the five critical residues are fixed in the active form, but mobile in the inactive form. Inactive mutants (with high mobility for these residues) may pay a severe entropic cost upon binding, whereas the constraints imposed on these residues in the active mutants reduce the entropic cost of residue immobilization upon complex formation. It is very difficult to determine the precise entropic contribution to induced fit binding, especially because some interfacial residues retain conformational flexibility (*45*) and because the calculation of the entropic cost for immobilizing residues requires a thorough exploration of conformational space. Molecular dynamics simulations that only explore a relatively short time scale (nanoseconds) are not suited to addressing these calculations, nor are many NMR methods that also explore similar time scales. New computational and experimental techniques are needed to evaluate entropic effects in residue immobilization at RNA−protein interfaces, perhaps along the lines of recent studies of enzyme catalysis (*65*).

In addition to its role in controlling the affinity of an interaction, does conformational dynamics contribute to specificity? NMR relaxation studies of the closely related estrogen receptor (ER) and glucocorticoid receptor (GR) DNA binding domains (DBD) were undertaken because these two domains assume very similar structures in the bound states, but in the free state, they differ significantly from each other and from the bound state (*66*). Most noticeably, the ER DBD is partially disordered in the free state whereas the GR DBD is well-structured; the ER undergoes a coupled folding reaction upon DNA binding. The analysis of ^{15}N relaxation revealed that the regions in GR that undergo conformational transitions upon binding coincide with those in conformational exchange in the free state for the ER. The very similar structures in the bound states and very different internal mobility in the free state point toward a major contribution of conformational dynamics to the specificity of the interaction. A triple mutant of GR, which recognizes the ER binding site, displays free state conformational dynamics reminiscent of the ER DBD. Could then dynamics be a signature of binding surfaces (*67*)?

A disadvantage of a binding surface having too much flexibility is that it can adapt itself to bind a broad range of ligands and therefore compromise specificity. However, nonspecific interactions can be biologically relevant for proteins or RNAs with multiple binding partners. In fact, the analysis of protein binding sites recognized by multiple protein partners highlighted the importance of conformational plasticity at interfaces, and suggested a fundamental role in multiple-ligand binding (*68*, *69*). Flexibility at a protein−RNA (or protein−protein) interface also facilitates coevolution of the two binding partners, thereby contributing to the generation of diversity. Mutations in one of the ligands can be accommodated by a structural rearrangement of the binding surface. This process can be re-created in vitro by inserting a disruptive mutation in one of the binding partners and selecting mutant ligands that restore the binding affinity. This technique was applied to the human growth hormone−receptor interface (*70*). The structures of the mutant complexes revealed that the binding was rescued by large conformational rearrangements in unaltered side chains, suggesting that conformational flexibility is a general feature enabling protein-binding surfaces to coevolve.

Better Experimental Techniques Are Needed To Characterize Conformationally Heterogeneous Systems

This article has emphasized that RNA−protein interfaces, before and after binding, sample a variety of conformations often interconverting between one another on a time scale of microseconds to milliseconds. Therefore, descriptions based on unique conformations are inadequate for describing the conformation of protein-binding sites within RNA and of certain regions of RNA-bound protein domains.

Crystallization conditions are optimized to obtain homogeneous crystals that diffract to the highest accuracy and generally yield a single conformer of the molecule. Only in some cases, several structurally independent molecules coexist inside a unit cell, or variation of the crystallization conditions leads to variation in the observed conformation. This opens the possibility of obtaining "snapshots" of the conformational space sampled by the molecule. Solution NMR spectroscopy provides a more direct way of probing the conformational flexibility of proteins and RNA in aqueous solution, and there are ways to measure the mobility or flexibility on a per residue basis. Proteins offer a wider range of techniques for studying the conformational flexibility of residues through the assessment of ^{15}N and ^{13}C relaxation. In contrast, NMR studies of RNA dynamics are limited so far to very few examples on model systems such as the UUCG tetraloop (*71*) or aptamers (*72*). This is a serious drawback, since we have argued that the flexibility of RNA structure is widely exploited by proteins that regulate its function.

Conformational dynamics critical to molecular function in protein recognition is likely to occur on a microsecond to millisecond time scale, or even slower, because it will involve several nucleotides or amino acid side chains; there is a need to develop methods for probing slow conformational dynamics. For example, the observation of slow motional processes in the region of free U1A protein that becomes involved in

interactions with RNA suggests that there may be motions in the free protein and RNA that lie on the pathway leading to the final conformation. If these motions exist, they could facilitate the conformational rearrangements necessary for binding by defining a favorable kinetic pathway for funneling the formation of intermediate complexes. NMR has the potential to define the time scale of the relevant motions and map the trajectories and potential energy landscapes on which these motions occur. The analysis of conformational dynamics and of the potential energy landscape of RNA and protein is key to understanding the binding mechanism and the entropy−enthalpy balance.

Conclusions and Future Prospects

Protein−RNA recognition often involves major conformational changes in both the protein and RNA. Yet we know little about the role of conformational dynamics in the initial encounter complex, or in subsequent binding events, nor do we understand how entropic losses and enthalpic gains are balanced against each other. Changes in dynamics and in the distribution of the reactants among different populated conformational states, as well as the intrinsic conformational properties of the reactants, are all key to understanding the energetic of recognition and the conformational pathways that are accessible to the system, upon which specificity depends. Yet we have few methods for studying these processes and fewer studies on structurally well established systems. The observation of induced fit or cofolding at RNA−protein interfaces has raised important questions about the mechanism of RNA−protein recognition, which cannot yet be satisfactorily addressed. Is RNA recognition a multistep process, with steric, ionic, and chemical features of the interacting partners being checked at different successive step? How are entropic losses balanced by enthalpic gains to achieve the required levels of affinity and specificity? Does binding involve the capture by the ligand of sparsely populated conformations of the free reactants? Are there motions in the free protein or RNA which lie on the pathway leading to the final conformation observed in the complex, thereby facilitating the conformational rearrangement and providing a kinetic pathway for funneling the system? What are the time scales of the relevant motions? What are the characteristics of the potential energy landscape on which these motions occur? Addressing these questions will provide several of the answers we seek in understanding how protein and RNA bind each other.

REFERENCES

1. Varani, G. (1997) *Acc. Chem. Res. 30*, 189−195.
2. Williamson, J. R. (2000) *Nat. Struct. Biol. 7*, 834−837.
3. Ma, B., Kumar, S., Tsai, C.-J., and Nussinov, R. (1999) *Protein Eng. 12*, 713−720.
4. Tinoco, I. J., and Bustamante, C. (1999) *J. Mol. Biol. 293*, 271−281.
5. Wu, M., and Tinoco, I. J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11555−11560.
6. Weeks, K. M., and Cech, T. R. (1995) *Cell 82*, 221−230.
7. Weeks, K. M., and Cech, T. R. (1996) *Science 271*, 345−348.
8. Caprara, M. G., Mohr, G., and Lambowitz, A. M. (1996) *J. Mol. Biol. 257*, 512−531.
9. Treiber, D. K., and Williamson, J. R. (1999) *Curr. Opin. Struct. Biol. 9*, 339−345.
10. Silverman, S. K., Zheng, M., Wu, M., Tinoco, I. J., and Cech, T. R. (1999) *RNA 5*, 1665−1674.
11. Batey, R. T., and Williamson, J. R. (1998) *RNA 4*, 984−997.
12. Crowder, S. M., Kanaar, R., Rio, D. C., and Alber, T. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 4892−4897.
13. Jones, K., and Peterlin, B. M. (1994) *Annu. Rev. Biochem. 63*, 717−743.
14. Gait, M. J., and Karn, J. (1993) *Trends Biochem. Sci. 18*, 255−259.
15. Aboul-ela, F., Karn, J., and Varani, G. (1995) *J. Mol. Biol. 253*, 313−332.
16. Long, K. S., and Crothers, D. M. (1995) *Biochemistry 34*, 8885−8895.
17. Puglisi, J. D., Tan, R., Canlan, B. J., Frankel, A. D., and Williamson, J. R. (1992) *Science 257*, 76−80.
18. Aboul-ela, F., Karn, J., and Varani, G. (1996) *Nucleic Acids Res. 24*, 3974−3981.
19. Riordan, F. A., Bhattacharya, A., McAteer, S., and Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305−310.
20. Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. G., Gait, M. J., and Karn, J. (1993) *J. Mol. Biol. 230*, 90−110.
21. Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharech, A. M., Singh, M., Karn, J., and Gait, M. J. (1994) *Nucleic Acids Res. 22*, 2592−2600.
22. Zacharias, M., and Hagerman, P. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6052−6056.
23. Ippolito, J. A., and Steitz, T. A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9819−9824.
24. Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W., and Lührmann, R. (1995) *Cell 82*, 475−483.
25. Stutz, F., Neville, M., and Rosbash, M. (1995) *Cell 82*, 495−506.
26. Tan, R., and Frankel, A. D. (1994) *Biochemistry 33*, 14579−14585.
27. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) *Science 273*, 1547−1551.
28. Ye, X., Gorin, A., Ellington, A. D., and Patel, D. J. (1996) *Nat. Struct. Biol. 3*, 1026−1033.
29. Peterson, R. D., Bartel, D. P., Szostak, J. W., Horwath, S. J., and Feigon, J. (1994) *Biochemistry 33*, 5357−5366.
30. Peterson, R. D., and Feigon, J. (1996) *J. Mol. Biol. 264*, 863−877.
31. Ippolito, J. A., and Steitz, T. A. (2000) *J. Mol. Biol. 295*, 711−717.
32. Hung, L.-W., Holbrook, E. L., and Holbrook, S. R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5107−5112.
33. Legault, P., Li, J., Mogridge, J., Greenblatt, J., and Kay, L. E. (1998) *Cell 93*, 289−299.
34. Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., and Greenblatt, J. (1998) *Mol. Cell 1*, 265−275.
35. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
36. Frankel, A. D., and Smith, C. A. (1998) *Cell 92*, 149−151.
37. Reyes, C. M., and Kollman, P. A. (2000) *J. Mol. Biol. 297*, 1145−1158.
38. Varani, G., and Nagai, K. (1998) *Annu. Rev. Biophys. Biomol. Struct. 27*, 407−445.
39. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994) *Nature 372*, 432−438.
40. Allain, F.-H. T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D., and Varani, G. (1996) *Nature 380*, 646−650.
41. Avis, J., Allain, F. H.-T., Howe, P. W. A., Varani, G., Neuhaus, D., and Nagai, K. (1996) *J. Mol. Biol. 257*, 398−411.
42. Gubser, C. C., and Varani, G. (1996) *Biochemistry 35*, 2253−2267.
43. Scherly, D., Kambach, C., Boelens, W., van Venrooij, W. J., and Mattaj, I. W. (1991) *J. Mol. Biol. 219*, 577−584.
44. Zeng, Q., and Hall, K. B. (1997) *RNA 3*, 303−314.
45. Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. E., and Varani, G. (1999) *J. Mol. Biol. 294*, 967−979.
46. Kranz, J. K., and Hall, K. B. (1998) *J. Mol. Biol. 275*, 465−481.

47. Kranz, J. K., and Hall, K. B. (1999) *J. Mol. Biol. 285*, 215−231.
48. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokohama, S. (1999) *Nature 398*, 579−585.
49. Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999) *Cell 98*, 835−845.
50. Allain, F. H.-T., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) *EMBO J. 19*, 6870−6881.
51. Wang, X., and Tanaka-Hall, T. (2001) *Nat. Struct. Biol. 8*, 141−146.
52. Allain, F. H.-T., Gilbert, D. E., Bouvet, P., and Feigon, J. (2000) *J. Mol. Biol. 303*, 227−241.
53. Arnez, J. G., and Moras, D. (1997) *Trends Biochem. Sci. 22*, 211−216.
54. McClain, W. H. (1993) *FASEB J. 7*, 72−78.
55. McClain, W. H. (1993) *J. Mol. Biol. 234*, 257−280.
56. Cusack, S. (1995) *Nat. Struct. Biol. 2*, 824−831.
57. Delagoutte, B., Moras, D., and Cavarelli, J. (2000) *EMBO J. 19*, 5599−5610.
58. Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A., and Nonin, S. (1997) *J. Mol. Biol. 272*, 645−664.
59. Allain, F. H.-T., Howe, P. W. A., Neuhaus, D., and Varani, G. (1997) *EMBO J. 16*, 5764−5774.
60. Stoldt, M., Wöhnert, J., Ohlenschläger, O., Görlach, M., and Brown, L. R. (1999) *EMBO J. 18*, 6508−6521.
61. Hermann, T., and Westhof, E. (1999) *Nat. Struct. Biol. 6*, 540−544.
62. Tinoco, I. J., and Kieft, J. S. (1997) *Nat. Struct. Biol. 4*, 509−512.
63. Kimura, S. R., Brower, R. C., Vajda, S., and Camacho, C. J. (2001) *Biophys. J. 80*, 635−642.
64. Ribas de Pouplana, L., Auld, D. S., Kim, S., and Schimmel, P. (1996) *Biochemistry 35*, 8095−8102.
65. Villa, J., Stajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., and Warshel, A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 11899−11904.
66. Wickstrom, A., Berglund, H., Hambraeus, C., van der Berg, S., and Härd, T. (1999) *J. Mol. Biol. 289*, 963−979.
67. Feher, V. A., and Cavanagh, J. (1999) *Nature 400*, 289−293.
68. Betts, M. J., and Sternberg, M. J. (1999) *Protein Eng. 12*, 271−283.
69. Sundberg, E. J., and Mariuzza, R. A. (2000) *Structure 8*, R137−R142.
70. Atwell, S., Ultsch, M., De Vos, A. M., and Wells, J. A. (1997) *Science 278*, 1125−1128.
71. Akke, M., Fiala, R., Jiang, F., Patel, D., and Palmer, A. G. I. (1997) *RNA 3*, 702−709.
72. Hoogstraten, C. G., Wank, J. R., and Pardi, A. (2000) *Biochemistry 39*, 9951−9958.

BI010680Y