VOLUME 30 NUMBER 5
ACCOUNTS of CHEMICAL RESEARCH®
MAY 1997

RNA−Protein Intermolecular Recognition

GABRIELE VARANI*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Received June 24, 1996

Background

RNA molecules play a central role in all the main functions of living molecules: storage of genetic information, propagation of the genetic material, and enzymatic activity. RNA molecules do not perform those functions alone, but in tight association with RNA-binding proteins. Thus, RNA−protein recognition is central to understanding a wide range of biological processes.

RNA−protein recognition cannot be undertstood without recognizing that the diversity and complexity of RNA structures are comparable to those of proteins. This structural diversity defines an enormous variety of sites and shapes for intermolecular recognition. In this Account, I will describe our current knowledge of the molecular basis of RNA−protein recognition. Since detailed thermodynamic and atomic-level structural information on RNA−protein complexes is very scarce, this Account will inevitably raise more questions than it will give answers. Thus, it will provide at least one good explanation for the ever increasing interest in this area of biochemistry.

Affinity and Specificity

Proteins bind RNA with the affinity required for complex formation at the concentrations of reagents present in living cells and for regulation of biological function. For example, transfer RNA (tRNA) synthetase enzymes bind tRNA with modest affinity (the bimolecular dissociation constant $K_d \approx 10^{-6}$ M): these enzymes would be ineffective if the substrate was not released after catalysis. In contrast, components of RNA−protein complexes that are permanently assembled (such as the human U1A protein) bind cognate RNAs much more tightly ($K_d < 10^{-9}$ M). The analysis of the few existing structures of RNA−protein complexes determined at atomic resolution reveals that affinity is not simply related to parameters such as protein charge or the size of intermolecular interface area. For example, tRNA−synthetase complexes have very large interface areas but bind RNA weakly, while the human U1A protein, with a small interface area, binds RNA very tightly.

Understanding the molecular determinants of the binding energy of an RNA−protein interaction is a very important but insufficient goal: specificity is equally important. Many RNA-binding proteins bind any RNA weakly, regardless of its sequence or structure. However, biological function requires discrimination of cognate RNAs (the correct, relevant targets) from noncognate RNAs, which are present in very large excess in the cell. The difference in binding energy between cognate and noncognate RNAs defines how specific an RNA−protein interaction is. Understanding this specificity adds a further, intriguing dimension to the characterization of these intermolecular recognition events.

RNA Recognition Differs from DNA Recognition

The many existing structures of DNA−protein complexes (>150) define an important paradigm in intermolecular recognition. Very often, DNA−protein recognition occurs by insertion of an α-helix into the major groove of double-stranded DNA.[1] A specific DNA sequence is then recognized through the formation of extensive hydrogen bonding and van der Waals interactions with the bases ("direct readout") and by recognition of sequence-dependent conformational features through electrostatic interactions with the negatively charged phosphodiester backbone

Gabriele Varani was born in Carrara, Italy, in 1959. He graduated from the University of Milano in physics and completed a Ph.D. in biophysics from the University of Milano under the supervision of Professor Giancarlo Baldini in 1987. After postdoctoral work with Professor Ignacio Tinoco at the University of California in Berkeley, Dr. Varani joined the MRC Laboratory of Molecular Biology in Cambridge in 1992 as a group leader and is now a senior member of the staff. Dr. Varani's interests are the determination of RNA structure by NMR spectroscopy and the study of RNA recognition by proteins and small molecular weight ligands.

* E-mail: gv1@mrc-lmb.cam.ac.uk. FAX: ++44-1223-213556. Phone: ++44-1223-402417.

(1) Steitz, T. A. *Q. Rev. Biophys.* **1990**, *23*, 205−280.

S0001-4842(96)00035-0 CCC: $14.00

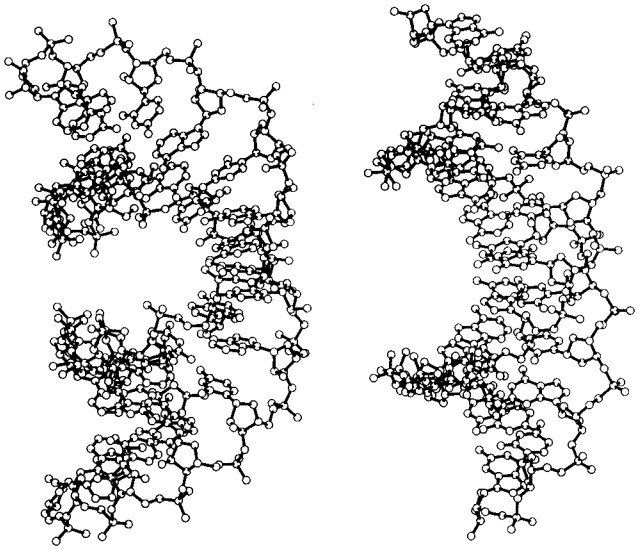

FIGURE 1. DNA and RNA double helices present different structures. The wide major groove of B-form DNA (right) sharply contrasts the narrow and deep major groove of A-form RNA (left).

("indirect readout"). This paradigm cannot be applied to RNA recognition. As shown in Figure 1, the major groove of double-helical RNA is too narrow to allow the insertion of a protein α-helix or β-strand. Some DNA-binding proteins[2−5] bind in the minor groove, and the RNA minor groove is well accessible (Figure 1). However, the chemical groups exposed in the minor groove of nucleic acid bases are not diverse enough among different nucleotides to allow effective discrimination.[6] Consequently, all known sequence-specific RNA-binding proteins recognize single-stranded regions and hairpin loops, where the functional groups on the bases become accessible (Figure 2). RNA double-helical regions are recognized only when structural distortions in the double helix generated by internal loops or bulges (Figure 2) allow access to the major groove.

Further differences exist between DNA−protein and RNA−protein recognition. All DNA double-helical structures are very similar (at least to to a first approximation). Thus, sequence-specific DNA recognition requires a very precise reading of the identity of individual nucleotides within the DNA double helix. The diversity of RNA structures (Figure 2) favors the recognition of unique shapes and charge distributions of different RNAs. Thus, despite the chemical similarities between RNA and DNA, the many important lessons of DNA−protein recognition are only partially applicable to RNA recognition.

$\alpha\beta$ Protein Domains

Many RNA-binding proteins contain modules of 60−90 amino acids that are responsible for RNA recognition and auxiliary domains that perform additional functions.[7−9]

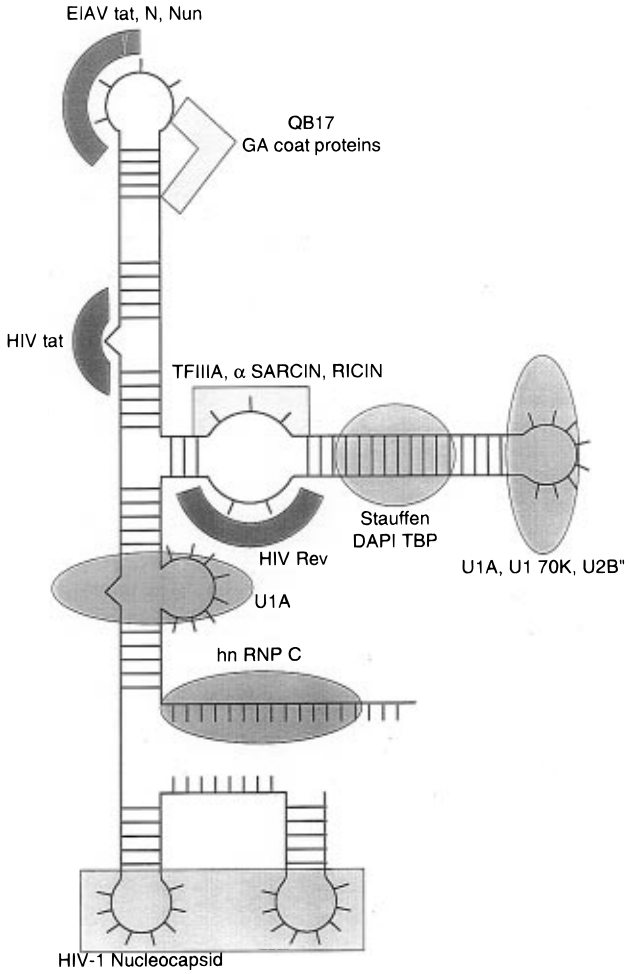

FIGURE 2. RNA-binding proteins do not target double-stranded RNA in a sequence-specific manner, but recognize instead single-stranded regions (hnRNP C) or sites of local distortions induced in double-helical regions by RNA hairpins (U1A, U1 70K, EIAV Tat, N, ...), bulges (HIV Tat), or internal loops (U1A, HIV Rev, TFIIIA, ...).

The three most common RNA-binding modules, ribonucleoprotein (RNP), K-homology (KH), and double-stranded RNA-binding (dsRBD) domains, have compact, globular structures (Figure 3) and constitute independent structural domains and RNA-binding units.[8]

The dsRBD domain is a general double-stranded RNA-binding module.[10−12] Isolated domains bind double-stranded RNA of any sequence with little or no specificity,[13] but multiple dsRBD domains may specifically recognize certain RNA structures.[13,14] KH domains appear to be non-sequence-specific single-stranded RNA-binding proteins. KH proteins have been associated with important biological functions; a single amino acid substitution

(2) Werner, M. H.; Huth, J. R.; Gronenborn, A. M.; Clore, G. M. *Cell* **1995**, *81*, 705−714.
(3) Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, *376*, 791−795.
(4) Kim, J. L.; Nikolov, D. B.; Burley, S. K. *Nature* **1993**, *365*, 520−527.
(5) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512−520.
(6) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804−808.
(7) Burd, C. G.; Dreyfuss, G. *Science* **1994**, *265*, 615−621.
(8) Nagai, K. *Curr. Opin. Struct. Biol.* **1996**, *6*, 53−61.
(9) Biamonti, G.; Riva, S. *FEBS Lett.* **1994**, *340*, 1−8.
(10) St Johnston, D.; Brown, N. H.; Gall, J. G.; Jantsch, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10979−10983.
(11) Bycroft, M.; Grünert, S.; Murzin, A. G.; Proctor, M.; St Johnston, D. *EMBO J.* **1995**, *14*, 3563−3571.
(12) Kharrat, A.; Macias, M. J.; Gibson, T. J.; Nilges, M.; Pastore, A. *EMBO J.* **1995**, *14*, 3572−3584.
(13) Clarke, P. A.; Mathews, M. B. *RNA* **1995**, *1*, 7−20.
(14) Ferrandon, D.; Elphick, L.; Nüsslein-Volhard, C.; St Johnston, D. *Cell* **1994**, *79*, 1221−1232.

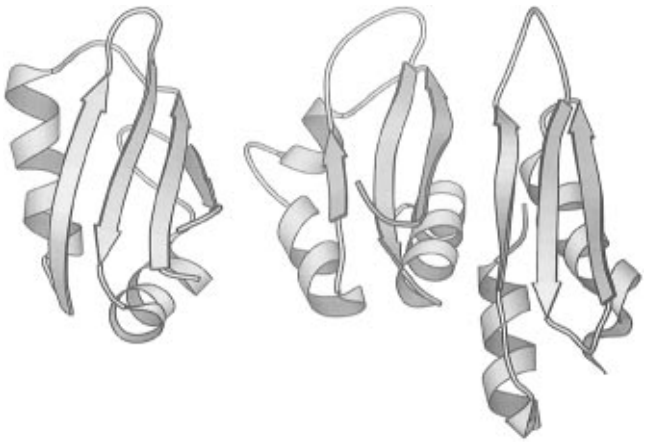

FIGURE 3. Three-dimensional structure of the three most common RNA-binding protein structures: RNP domain (left),[18,63] KH domain (center),[15] and dsRBD domain (right).[11,12]

that unfolds a human KH domain[15] leads to fragile-X syndrome,[16] the most common cause of inherited mental retardation. The RNP domain is one of the most common protein structures in higher organisms, comprising over 600 sequences. It is identified by two highly conserved amino acid sequences (RNP-1 and RNP-2)[17] located in the central strands of an antiparallel β-sheet.[18,19] RNP proteins specifically recognize both single-stranded and highly structured RNAs.

RNP,[18] KH,[15] and dsRBD[11,12] are $\alpha\beta$ proteins with an antiparallel β-sheet on one face of the protein packed by a hydrophobic core against an α-helical face (Figure 3). Although a number of all-helical RNA-binding proteins have been recently identified,[20−22] the $\alpha\beta$ structural theme is conserved in many RNA-binding proteins that do not share sequence homology with these three motifs, including ribosomal proteins[23] and other factors involved in protein synthesis.[23−26] In some cases (for example, ribosomal proteins L12 and L30 and RNP proteins, or dsRBD proteins and ribosomal protein S5) a similar arrangement of secondary structure elements indicates that these proteins originate from a common ancestor. In general, different $\alpha\beta$ RNA-binding proteins have different topology of the secondary structure elements: the RNP domain has a repeated $\beta\alpha\beta$ arrangement,[18,19] dsRBDs have $\alpha\beta\beta\beta\alpha$ topology,[11,12] and KH proteins have $\beta\alpha\alpha\beta\beta\alpha$ fold.[15] These structural differences and the low sequence homology indicate that the different domains represent distinct, convergent solutions to a common structure for RNA binding. These considerations strongly suggest that the $\alpha\beta$ structure represents a particularly favorable platform for RNA recognition.

Arginine-Rich Motif

A sequence of 10−15 amino acids rich in arginines and lysines mediates RNA recognition in the so-called "basic-domain" class of RNA-binding proteins.[27] This feature is often referred to as a domain, but does not constitute an independent structural domain or RNA recognition unit. Structural analysis at atomic resolution has so far been unsuccesful for this protein family, but extensive data exist on short peptide models of the arginine-rich region. Remarkably, peptides as short as 10−15 amino acids bind RNA with comparable affinity to the corresponding protein ($K_d \approx 10^{-9}$ M in many cases[28]), but fail to discriminate cognate binding sites from noncognate RNAs. For example, a basic peptide mimic of the human immunodeficiency virus (HIV-1) Rev protein binds its cognate RNA only 2-fold better than noncognate RNAs, while the full Rev protein binds its cognate RNA 1000-fold better than other substrates.[29]

Peptide models have provided very valuable insight into fundamental mechanisms of RNA recognition by basic-domain proteins, but have not addressed satisfactorily the origin of specificity. Arginine-rich peptides are unstructured (with a single exception[30]), but display a stable conformation upon RNA binding.[31−34] Each RNA structure presents a distinct array of hydrogen bond donors and acceptors and distribution of negative charges on the phosphates. Binding to either specific or nonspecific sites orders the unstructured peptides. Adaptation of the flexible peptide structures to the shape and charge distribution of different RNA targets would maximize favorable electrostatic interactions between the side chains of these very basic peptides and the negatively charged RNA backbone. Short peptides unconstrained by a protein scaffold may be unable to discriminate distinct RNA sites because they can adapt their conformation equally well to the shape of different RNAs.

Many basic-domain proteins recognize structural distortions in regular RNA double helices induced by internal loops or bulges (Figure 2). For example, HIV-1 Rev protein binds an internal loop containing non-Watson−Crick base pairs,[33−38] whereas HIV-1 Tat protein binds a tri-nucleotide

(15) Musco, G.; Stier, G.; Joseph, C.; Castiglione Morelli, M. A.; Nilges, M.; Gibson, T. J.; Pastore, A. *Cell* **1996**, *85*, 237−245.
(16) Siomi, H.; Siomi, M. C.; Nussbaum, R. L.; Dreyfuss, G. *Cell* **1993**, *74*, 291−298.
(17) Query, C. C.; Bentley, R. C.; Keene, J. D. *Cell* **1989**, *57*, 89−101.
(18) Nagai, K.; Oubridge, C.; Jessen, T. H.; Li, J.; Evans, P. R. *Nature* **1990**, *348*, 515−520.
(19) Hoffman, D. W.; Query, C. C.; Golden, B. L.; White, S. W.; Keene, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *83*, 2495−2499.
(20) Markus, M. A.; Hinck, A. P.; Huang, S.; Draper, D. E.; Torchia, D. A. *Nat. Struct. Biol.* **1997**, *4*, 70−77.
(21) Xing, Y.; GuhaThakurta, D.; Draper, D. E. *Nat. Struct. Biol.* **1997**, *4*, 24−27.
(22) Berglund, H.; Rak, A.; Serganov, A.; Garber, M.; Härd, T. *Nat. Struct. Biol.* **1997**, *4*, 20−23.
(23) Liljas, A.; Garber, M. *Curr. Opin. Struct. Biol.* **1995**, *5*, 721−727.
(24) Biou, V.; Shu, F.; Ramakrishnan, V. *EMBO J.* **1995**, *14*, 4056−4064.
(25) Garcia, C.; Fortier, P.-L.; Blanquet, S.; Lallemand, J.-Y.; Dardel, F. *J. Mol. Biol.* **1995**, *254*, 247−259.
(26) Kang, C.-H.; Chan, R.; Berger, I.; Lockshin, C.; Green, L.; Gold, L.; Rich, A. *Science* **1995**, *268*, 1170−1173.
(27) Lazinski, D.; Grzadienska, E.; Das, A. *Cell* **1989**, *59*, 207−218.
(28) Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. *Science* **1990**, *249*, 1281−1285.
(29) Daly, T. J.; Doten, R. C.; Rusche, J. R.; Auer, M. *J. Mol. Biol.* **1995**, *253*, 243−258.
(30) Mujeeb, A.; Bishop, K.; Peterlin, B. M.; Turck, C.; Parslow, T. G.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8248−8252.
(31) Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. *Science* **1995**, *270*, 1200−1203.
(32) Ye, X.; Kumar, R. A.; Patel, D. J. *Chem. Biol.* **1995**, *2*, 827−840.
(33) Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. *Science* **1996**, *273*, 1547−1551.
(34) Ye, X.; Gorin, A.; Ellington, A. D.; Patel, D. J. *Nat. Struct. Biol.* **1996**, *3*, 1026−1033.
(35) Heaphy, S.; Finch, J. T.; Gait, M. J.; Karn, J.; Singh, M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 7366−7370.
(36) Iwai, S.; Pritchard, C.; Mann, D. A.; Karn, J.; Gait, M. J. *Nucleic Acids Res.* **1992**, *20*, 6465−6472.
(37) Battiste, J. L.; Tan, R.; Frankel, A. D.; Williamson, J. R. *Biochemistry* **1994**, *33*, 2741−2747.

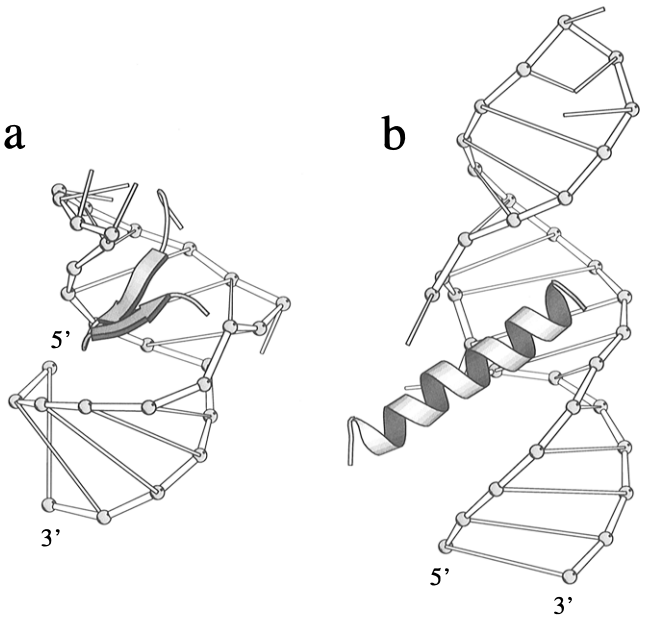

FIGURE 4. Structures of RNA−peptide complexes. The open major groove generated by an RNA bulge (left) or internal loop (right) is recognized by an antiparallel β-hairpin in the BIV Tat-TAR structure on the left[32] or an extended α-helix in the HIV Rev-RRE structure shown on the right.[33]

bulge.[39] Rev-derived peptides are superficially similar to DNA-binding proteins, since binding occurs in the RNA major groove and the peptide becomes α-helical upon binding[33,34] (Figure 4, right). However, considerations of the RNA structure highlight fundamental differences. The structural distortion generated by non-Watson−Crick base pairs opens up the major groove of the RNA to allow the "direct" recognition of the base functionality in the vicinity of the structural distortion (Figure 4). Furthermore, these intramolecular (as well as some intermolecular) interactions stabilize a unique RNA structure that may facilitate the indirect readout of the electrostatic potential of the phosphodiester backbone.

The complex between the bovine immunodeficienty virus (BIV) Tat-derived peptide and its cognate RNA[31,32] shows a remarkable similarity with DNA recognition by ribbon−helix−helix proteins.[40,41] In this complex, two unpaired nucleotides within an otherwise regular RNA helix open up the major groove to allow insertion of a β-ribbon peptide structure (Figure 4, left). The shape of two-stranded antiparallel β-ribbons closely matches that of double-stranded nucleic acids to easily fit DNA[4,5,42] and RNA[43] minor grooves, DNA major grooves,[40,41] and distorted RNA major grooves.[31,32] 2′-OH groups in the RNA minor groove are also regularly spaced to donate hydrogen bonds to backbone carbonyl groups from antiparallel β-sheets.[43] More examples of β-ribbon−RNA recognition will probably emerge in the future.

tRNA Recognition by Synthetase Enzymes

Aminoacyl tRNA synthetases (aaRS) enzymatically activate tRNA for protein synthesis by covalently joining an amino acid to the 3′-end of each of 20 different tRNAs.[44] The fidelity of translation of the genetic code depends on accurate aminoacylation. aaRS enzymes must recognize three-dimensional structural features common to all tRNAs, while discriminating different tRNAs by utilizing a limited set of identity elements (nucleotides or chemical modifications specific to each tRNA).[45,46] Striking features of tRNA−synthetase complexes are very large interface areas, ∼3000 Å^2 or ∼20% of free tRNA accessible surface.[47−50] By contrast, the nonspecific complex between elongation factor TU (EF-Tu) and tRNA^{Phe} has a smaller interface area and few direct contacts to the bases,[51] although EF-Tu binds tRNAs much more tightly ($K_d \approx 10^{-9}$ M) than RS enzymes ($K_d \approx 10^{-6}$ M). Differences in binding free energy are relatively small for aaRS enzymes binding different tRNAs: substrate discrimination occurs primarily at the level of the efficiency of aminoacylation. tRNA-induced protein conformational rearrangements appear to be a primary mechanism of substrate discrimination.[47,49] The structural complementarity expressed by very large interface areas allows RS enzymes to differentiate tRNAs from other cellular RNAs and may provide the free energy necessary for these conformational rearrangements.[49,52]

tRNA−synthetase complexes provide many examples of how different intermolecular interactions determine binding energy and contribute to intermolecular discrimination (i.e., specificity). Hydrogen-bonding interactions are clearly important. Direct and water-mediated hydrogen bonds between protein side chains and tRNA bases are formed in the typically wide minor groove of acceptor and anticodon stems of tRNA^{Gln}.[47,50] The RNA bases are also recognized by hydrogen bonding within the anticodon loop of tRNA^{Gln} [47,50] and tRNA^{Asp} [49] and in the unusually accessible major grooves at the end of the acceptor and anticodon stems of tRNA^{Asp}.[49,52] Electrostatic interactions are a second source of intermolecular contacts. Interactions with unique phosphodiester backbone geometries not only provide binding energy but also facilitate intermolecular discrimination. For example, a critical G·U base pair in tRNA^{Asp} is recognized only through the indirect readout of the backbone conformation.[49] Shape complementarity allows extensive van der Waals interactions, and the ability of tRNA to form unique structural features is critical for anticodon recognition in tRNA^{Gln} [47,50] and tRNA^{Asp}.[49] Conformational changes and

(38) Peterson, R. D.; Bartel, D. P.; Szostak, J. W.; Horwath, S. J.; Feigon, J. *Biochemistry* **1994**, *33*, 5357−5366.
(39) Gait, M. J.; Karn, J. *TIBS* **1993**, *18*, 255−259.
(40) Raumann, B. E.; Rould, M. A.; Pabo, C. O.; Sauer, R. T. *Nature* **1994**, *367*, 754−757.
(41) Somers, W. S.; Phillips, S. E. V. *Nature* **1992**, *359*, 387−393.
(42) Church, G. M.; Sussman, J. L.; Kim, S.-H. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1458−1462.
(43) Carter, C. W. J.; Kraut, J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 283−287.
(44) Cusack, S. *Nat. Struct. Biol.* **1995**, *2*, 824−831.
(45) McClain, W. H. *J. Mol. Biol.* **1993**, *234*, 257−280.
(46) McClain, W. H. In *tRNA: Structure, Biosynthesis and Function*; Söll, D., RajBhandary, U., Eds.; American Society for Microbiology: Washington, DC, 1995.
(47) Rould, M. A.; Perona, J. J.; Söll, D.; Steitz, T. A. *Science* **1989**, *246*, 1135−1142.
(48) Biou, V.; Yaremchuk, A.; Tukalo, M.; Cusack, S. *Science* **1994**, *263*, 1404−1410.
(49) Caverelli, J.; Rees, B.; Ruff, M.; Thierry, J.-C.; Moras, D. *Nature* **1993**, *362*, 181−184.
(50) Rould, M. A.; Perona, J. J.; Steitz, T. A. *Nature* **1991**, *352*, 213−218.
(51) Nilsen, P.; Kjeldgaard, M.; Thirup, S.; Polekhina, G.; Reshetnikova, L.; Clark, B. F. C.; Nyborg, J. *Science* **1995**, *270*, 1464−1472.
(52) Ruff, M.; Krishnaswamy, S.; Boeglin, M.; Poterszman, A.; Mitschler, A.; Podjarny, A.; Rees, B.; Thierry, J. C.; Moras, D. *Science* **1991**, *252*, 1682−1689.

ordering of flexible nucleotides occur in $tRNA^{Asp}$ upon synthetase binding, while two non-Watson−Crick base pairs stabilized by protein contacts extend the anticodon stem in the $aaRS^{Gln}$−$tRNA^{Gln}$ complex. Similarly, adaptation of the seryl-tRNA synthetase conformation to the unique shape of the extended $tRNA^{Ser}$ variable arm through ordering of a protein helical arm maximizes surface complementarity and facilitates discrimination.[48] A fourth category of interactions is intermolecular stacking between protein heteroaromatic side chains and RNA bases. Anticodon nucleotides are splayed-out against the surface of β-barrel domains in the $tRNA^{Asp\ 49}$ and $tRNA^{Gln\ 47}$ complexes allowing intermolecular stacking interactions to occur.

RNP Domain Paradigm of RNA Recognition

Hundreds of proteins from higher organisms recognize RNA substrates widely diverse in sequence and structure via RNP domains. Proteins which bind single-stranded RNA require multiple RNP domains,[53−55] while single RNP domains can recognize 5−10 single-stranded RNA bases with high affinity and specificity when the nucleotides are presented in a defined RNA structural context. For example, the human U1A protein recognizes seven single-stranded nucleotides in the context of a hairpin or internal loop (Figure 2) with very high affinity ($K_d \approx 10^{-11}$ M)[18,56,57] and specificity. When those seven nucleotides are presented in the absence of RNA secondary structure, the binding constant is reduced 100000-fold, i.e., almost to the level observed for binding to RNA of random sequence.[58,59]

Crystallographic[60] and NMR[61] structures of the complexes of the human U1A protein with two different RNA substrates have revealed important aspects of the molecular basis of RNP−RNA recognition. The N-terminal RNP domain of U1A binds two distinct RNA substrates that present the same single-stranded nucleotide sequence in a completely different structural context. In both the crystal structure of the U1A−hairpin loop[60] and in the NMR structure of the U1A−internal loop complex,[61] bases within the single-stranded loops are exposed to the surface of the β-sheet of U1A. Most hydrogen bond donors and acceptors from the single-stranded bases are recognized by an extensive network of interactions with protein residues, but an unusually large number of intermolecular contacts involve protein main chain donors and acceptors. In these complexes, all RNA bases are involved in intra- or intermolecular stacking interactions and the RNA phosphodiester backbone contacts a set of basic amino acids.

The comparison with the unbound RNA internal loop[62] and protein[63] structures (Figure 5a) reveals a complex recognition mechanism. A first set of intermolecular contacts involve rigid body docking between the protein and the well-ordered RNA double-helical regions and three single-stranded nucleotides. The remaining five single-stranded nucleotides are recognized instead by induced fit. The rigid interaction involves two loops within U1A that have unique sequences in different RNP proteins. Thus, recognition of a well-defined RNA structure allows U1A to binds its target and prevents other RNP proteins (lacking these unique amino acid sequences) from binding the same RNA. Complex formation reorients an α-helix at the end of the protein domain that is essential for binding.[59,64,65] At the same time, protein binding orders five single-stranded nucleotides against the β-sheet surface through intermolecular stacking interactions with three highly conserved aromatic amino acids. Intermolecular stacking interactions with exposed aromatic side chains are common to RNP proteins,[60,61] tRNA synthetases,[49,50] and viral coat proteins.[66] Although rare in DNA−protein complexes, intermolecular stacking interactions appear to be very common in RNA−protein recognition, perhaps because they provide large amounts (∼3 kcal/mol) of binding energy.[67]

Conformational Flexibility at RNA−Protein Interfaces

Structures of RNA−protein complexes have provided important insight into the mechanisms of intermolecular recognition, but have also raised intriguing questions concerning the molecular origin of binding energy and sequence discrimination. Surprisingly, binding energy is not related to interface area, presumably a measure of the van der Waals interaction energy and of the entropic contribution to binding energy from solvent and ion release. The very tight binding of the U1A protein ($K_d \approx 10^{-11}$ M) is remarkable, since the interface area is small and binding involves an entropically costly disorder−order transition in the single-stranded RNA loop.[61] If binding ordered exposed protein side chains at intermolecular interfaces to a high degree, changes in the distribution of molecular vibrations would contribute 15−25 kcal/mol (an amount comparable to the overall binding energy!) to increase the free energy of protein−protein and protein−DNA interactions.[68] However, the NMR relaxation properties of amino acid chains in DNA−protein and peptide−protein complexes indicate that motion is much less restricted at the intermolecular interface than in the highly

(53) Tacke, R.; Manley, J. L. *EMBO J.* **1995**, *14*, 3540−3551.
(54) Kanaar, R.; Lee, A. L.; Rudner, D. Z.; Wemmer, D. E.; Rio, D. C. *EMBO J.* **1995**, *14*, 4530−4539.
(55) Shamoo, Y.; Abdul-Manam, N.; Patten, A. M.; Crawford, J. K.; Pellegrini, M. C.; Williams, K. R. *Biochemistry* **1994**, *33*, 8272−8281.
(56) van Gelder, C. W. G.; Gunderson, S. I.; Jansen, E. J. R.; Boelens, W. C.; Polycarpou-Schwartz, M.; Mattaj, I. W.; van Venrooij, W. J. *EMBO J.* **1993**, *12*, 5191−5200.
(57) Hall, K. B.; Stump, W. T. *Nucleic Acids Res.* **1992**, *20*, 4283−4290.
(58) Tsai, D. E.; Harper, D. S.; Keene, J. D. *Nucleic Acids Res.* **1991**, *19*, 4931−4936.
(59) Hall, K. B. *Biochemistry* **1994**, *33*, 10076−10088.
(60) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C.-H.; Nagai, K. *Nature* **1994**, *372*, 432−438.
(61) Allain, F.-H. T.; Gubser, C. C.; Howe, P. W. A.; Nagai, K.; Neuhaus, D.; Varani, G. *Nature* **1996**, *380*, 646−650.
(62) Gubser, C. C. G.; Varani, G. *Biochemistry* **1996**, *35*, 2253−2267.
(63) Avis, J.; Allain, F. H.-T.; Howe, P. W. A.; Varani, G.; Neuhaus, D.; Nagai, K. *J. Mol. Biol.* **1996**, *257*, 398−411.
(64) Jessen, T. H.; Oubridge, C.; Teo, C. H.; Pritchard, C.; Nagai, K. *EMBO J.* **1991**, *10*, 3447−3456.
(65) Scherly, D.; Kambach, C.; Boelens, W.; van Venrooij, W. J.; Mattaj, I. W. *J. Mol. Biol.* **1991**, *219*, 577−584.
(66) Valegárd, K.; Murray, J. B.; Stockley, P. G.; Stonehouse, N. J.; Liljas, L. *Nature* **1994**, *371*, 623−626.
(67) LeCuyer, K. A.; Behlen, L. S.; Uhlenbeck, O. C. *EMBO J.* **1996**, *15*, 6847−6853.
(68) Spolar, R. S.; Record, M. T. J. *Science* **1994**, *263*, 777−784.

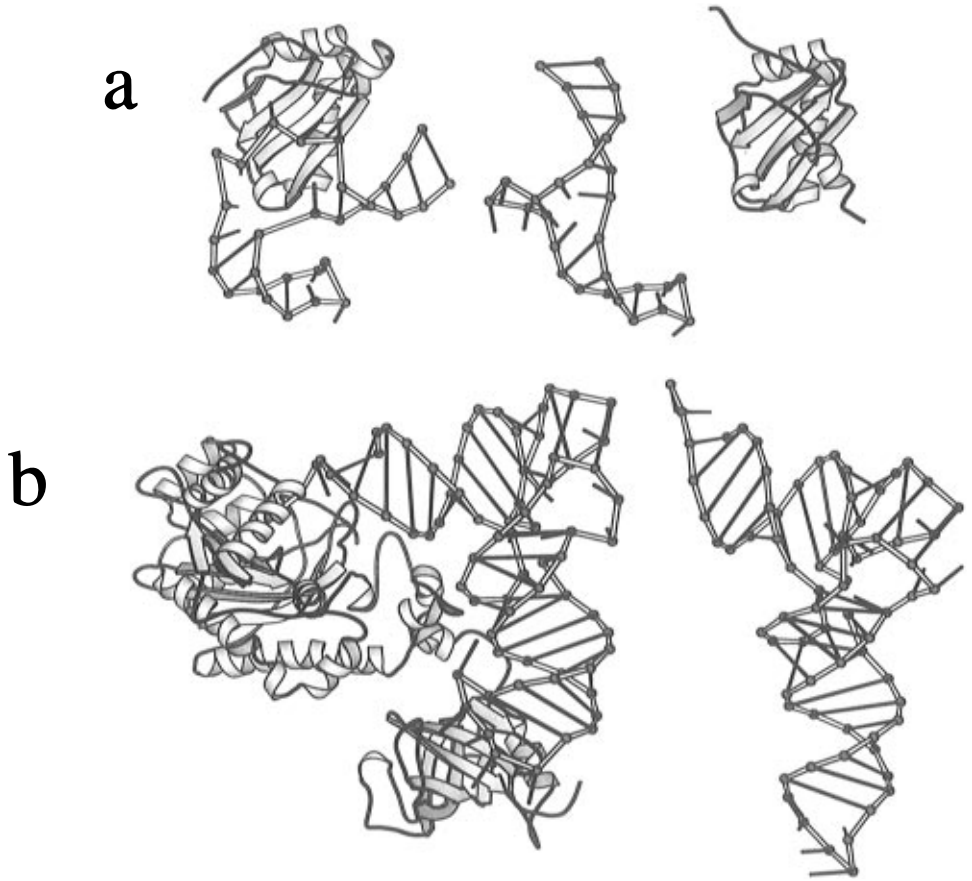

FIGURE 5. Examples of conformational adaptation in RNA−protein recognition. (a) NMR structures of the free U1A protein (right), the free RNA internal loop substrate (center), and the protein−RNA complex (left).[61] Protein binding induces a dramatic change in the RNA conformation, and requires a sharp reorientation of an α-helix at the carboxy end of the U1A domain. (b) Crystallographic structures of $tRNA^{Asp}$ free[80] (right) and in complex with aspartyl-tRNA synthetase (left).[49,52] In this orientation, the anticodon is at the bottom while the acceptor end is at the top left corner of the tRNA. Although the L-shaped tRNA structure is preserved, synthetase binding changes the relative orientation of the two major helical domains of $tRNA^{Asp}$.

ordered hydrophobic core.[69,70] Protein−nucleic acid interfaces may not be rigidly ordered. Rather, a fine balance between rigidity and flexibility may provide a compromise between complete specificity (at large entropic cost) and complete lack of selectivity.

Several observations derived from the U1A complexes suggest that protein−RNA interfaces are not rigid. Firstly, the arginine 52 side chain in U1A, involved in four hydrogen bonds in the crystalline state,[60] can be mutated to lysine (which cannot form those hydrogen bonds) without significant increase in the free energy of binding (<0.5 kcal/mol).[18,64] The solution structure is consistent with a less ordered conformation for arginine 52, where each hydrogen bond is only present part of the time,[61] perhaps rationalizing the remarkably small contribution to the binding energy. Secondly, a cytosine deeply buried at the intermolecular interface is recognized at every base functionality by the protein, yet can be mutated to guanine (which is much larger and cannot fit the space vacated by cytosine) with only a 10-fold increase in affinity.[57] NMR observations suggest once again that the intermolecular interface is flexible enough to accommodate either base through energetically inexpensive local conformational adjustments.

In contrast to the situation observed with the portion of the U1A−RNA interface determined by induced fit, it is more difficult to reorganize the intermolecular interface when the interaction occurs by rigid interlocking of preordered regions of the protein and RNA. For example, it is impossible for U1A to bind tightly if leucine 49 is mutated; this residue snugly fits a binding pocket at the junction between RNA helices and loops.[61,71] Arginine 52 nearby can only be mutated to lysine, which has similar size and positive charge,[18,64] but not to a much smaller glutamine. It has already been mentioned that disruption of the RNA secondary structure reduces binding 100000-fold.[58,59] Thus, the preformed RNA secondary structure may provide a structural counterpart to the rigid β-sheet of the protein in reducing the entropic costs of RNA folding[72] and in providing large amounts of binding energy through essential electrostatic interactions from the phosphodiester backbone.

Do RNA Conformational Changes Contribute to Specificity?

The discussion of the previous paragraph suggests that even an extensive network of intermolecular interactions

(69) Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. *Biochemistry* **1996**, *35*, 361−368.
(70) Berglund, H.; Baumann, H.; Knapp, S.; Ladenstein, R.; Härd, T. *J. Am. Chem. Soc.* **1995**, *117*, 12883−12884.
(71) Laird-Offringa, I. A.; Belasco, J. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11859−11863.
(72) Nagai, K.; Oubridge, C.; Ito, N.; Avis, J.; Evans, P. *TIBS* **1995**, *20*, 235−240.

provide only limited specificity because of residual conformational flexibility. Characteristically large heat capacity changes ($\Delta C_p \ll 0$) observed in DNA−protein complexes have been convincingly attributed to local protein folding on DNA binding to create key parts of the intermolecular interface.[68] Thermodynamic data show that large heat capacity changes also occur upon U1A−RNA binding.[57,59] Formation of the U1A−RNA interface through induced fit depends on the identity of amino acid side chains exposed on the β-sheet surface and in neighboring loops, as well as on the identity of the single-stranded nucleotides. Both the driving force for the conformational change (the binding free energy) and the driven process (RNA folding) depend on RNA and protein sequences, providing a key step in intermolecular discrimination. In this interpretation, binding energy is dissipated to favor the conformational change to enhance discrimination. Since changes in sequence may critically affect the geometry of the interface far away from the site where the mutation occurs, nonadditive energetics of recognition are expected. In fact, mutations in U1A or in its RNA targets cannot be straightforwardly interpreted from the loss of intermolecular contacts observed in NMR and X-ray structures.[18,57,59,64,73]

Functional Implications

Direct evidence for the important functional role of protein-induced RNA conformational rearrangements is provided by group I self-splicing introns, a class of RNA enzymes that often require protein cofactors for catalysis in living organisms.[74−77] For example, a fungal tyrosyl-tRNA synthetase protein (Cyt-18) binds a group I intron and folds the preexisting RNA secondary structure into the catalytically active three-dimensional structure.[75,76,78] Thus, protein binding to group I introns induces higher order tertiary RNA structures required for catalytic function. Similarly, ribosomal proteins induce or stabilize compact tertiary structures which are largely absent in the naked ribosomal RNA.

Conclusions and Perspectives

Recent structures of RNA−protein complexes have revealed new principles of intermolecular RNA−protein recognition. These structures demonstrate different ways in which protein β-sheets provide large accesible surfaces for extensive interactions with RNA bases exposed in single-stranded regions. These β-sheet surfaces are so common in RNA-binding proteins to suggest a role as dominant as that of α-helices in DNA recognition. This dominance probably originates from the stereochemical complementarity with RNA structure due to the natural right-handedness and concavity of antiparallel β-sheets.[23,64,79] A balance of induced fit and shape selectivity through rigid fit is common to tRNA−synthetase complexes and to the recognition of highly structured RNAs by RNP proteins. Formation of RNA−protein complexes is clearly a highly dynamic process: RNA structure directs protein binding, which in turn modulates the RNA conformation to create a unique intermolecular interface. Atomic resolution structural information is only available for a handful of RNA−protein complexes, excluding ribosomal components, KH or dsRBD domains, and basic-domain proteins. The next several years will undoubtedly see exciting progress in understanding the thermodynamic and structural basis of RNA−protein recognition.

It is a pleasure to thank Drs. Kung-Yao Chang, Jacek Nowacelski, and Ignacio Tinoco, Jr. (University of California at Berkeley), and Mr. Charles Gubser and Luca Varani for critical reading of the manuscript, Mr. Charles Gubser and Luca Varani for preparation of illustrations, and Dr. Annalisa Pastore (EMBL, Heidelberg), Dinshaw Patel (Sloan-Kettering Cancer Center, New York), and Jamie Williamson (MIT, Boston) for the coordinates of structures determined in their laboratories.

AR960035X

(73) Policarpou-Schwartz, M.; Gunderson, S. I.; Kandels-Lewis, S.; Séraphin, B.; Mattaj, I. W. *RNA* **1996**, *2*, 11−23.
(74) Weeks, K. M.; Cech, T. R. *Cell* **1995**, *82*, 221−230.
(75) Mohr, G.; Caprara, M. G.; Guo, Q.; Lambowitz, A. M. *Nature* **1994**, *370*, 147−150.
(76) Mohr, G.; Zhang, A.; Gianelos, J. A.; Belfort, M.; A. M., L. *Cell* **1992**, *69*, 483−494.
(77) Weeks, K. M.; Cech, T. R. *Science* **1996**, *271*, 345−348.
(78) Caprara, M. G.; Lehnert, V.; Lambowitz, A. M.; Westhof, E. *Cell* **1996**, *87*, 1135−1145.
(79) Howe, P. W. A.; Nagai, K.; Neuhaus, D.; Varani, G. *EMBO J.* **1994**, *13*, 3873−3881.
(80) Moras, D.; Comarmond, M. B.; Fischer, J.; Weiss, R.; Thierry, J. C.; Ebel, J. P.; Giegé, R. *Nature* **1980**, *288*, 669−674.