Article No. jmbi.1999.2991 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 293, 255–270

JMB

Themes in RNA-Protein Recognition

David E. Draper

Department of Chemistry
Johns Hopkins University
Baltimore, MD 21218, USA

Atomic resolution structures are now available for more than 20 complexes of proteins with specific RNAs. This review examines two main themes that appear in this set of structures. A ''groove binder'' class of proteins places a protein structure (α-helix, 3_{10}-helix, β-ribbon, or irregular loop) in the groove of an RNA helix, recognizing both the specific sequence of bases and the shape or dimensions of the groove, which are sometimes distorted from the normal *A*-form. A second class of proteins uses β-sheet surfaces to create pockets that examine single-stranded RNA bases. Some of these proteins recognize completely unstructured RNA, and in others RNA secondary structure indirectly promotes binding by constraining bases in an appropriate orientation. Thermodynamic studies have shown that binding specificity is generally a function of several factors, including base-specific hydrogen bonds, non-polar contacts, and mutual accommodation of the protein and RNA-binding surfaces. The recognition strategies and structural frameworks used by RNA binding proteins are not exotically different from those employed by DNA-binding proteins, suggesting that the two kinds of nucleic acid-binding proteins have not evolved independently.

Keywords: Electrostatic interaction; mutual accommodation; OB fold; major groove; minor groove

Introduction

How proteins selectively bind specific sites on nucleic acids has been a challenging and interesting problem since the earliest days of molecular biology. The repressor hypothesis (Jacob & Monod, 1961) led to the demonstration that a protein is able to select a single DNA sequence out of an entire genome (Ptashne, 1967). This impressive feat of molecular recognition has been intensively studied ever since, and there are now a large number of atomic resolution structures of protein-DNA complexes which, with parallel thermodynamic studies, provide a picture of how sequence-specific DNA recognition is performed by a variety of repressors, transcription factors, and restriction enzymes. But perhaps the first protein-nucleic acid recognition problem to be defined was the enzymatic linking of an amino acid with its correct tRNA (Berg & Ofengand, 1958; Hoagland, 1960), a process whose specificity was seen as crucial for accurate gene expression. Protein recognition of specific RNA sites was also implicit in early studies of ribosome assembly (Traub & Nomura, 1969; Nomura, 1973). Since then, the participation of specific protein-RNA complexes in a large number of cellular processes has become evident.

For a variety of reasons, biochemical and physical characterization of protein-RNA complexes has lagged behind corresponding studies of protein-DNA complexes by more than a decade. In a comprehensive review of RNA-protein interactions written in 1994 (Draper, 1995), structures were available for three aminoacyl-tRNA synthetase-tRNA complexes and two complexes of proteins with RNA hairpins (MS2 coat protein and the U1A RNA binding domain). Since then, more than a dozen new peptide and protein-RNA complexes have been solved by NMR or crystallography, and many of these have been accompanied by binding studies with sequence variants. The present review is an update of the 1995 review in light of the new structural and thermodynamic information that has become available. The emphasis will be on two main categories that encompass most of the

Abbreviations used: BIV, bovine immunodeficiency virus; PRM, RNA recognition motif; RNP, ribonucleoprotein; OB fold, oligonucleotide/oligosaccharide binding.

E-mail address of the corresponding author: draper@jhunix.hcf.jhu.edu

0022-2836/99/420255–16 $30.00/0

structures: peptides and proteins that bind RNA helix grooves, and β-sheet proteins that recognize single-stranded bases in a sequence-specific manner. The two themes underscore the range of recognition strategies available to RNA-binding proteins, from purely sequence-dependent discrimination among single-stranded RNAs to recognition of highly structured RNA surfaces.

Peptides and proteins binding in RNA grooves

The *A*-form geometry of RNA helices has a deep and narrow major groove, which is not accessible to protein secondary structure in the same way as *B*-form DNA. However, there are two mitigating factors. First, RNA secondary structures rarely contain a more than half a turn of helix, and so have much of their major groove surface accessible from the ends. Second, mismatches and bulges confer a surprising plasticity on RNA helices and generate major grooves wide enough for protein secondary structure to penetrate. The wide and flat minor groove surface is, of course, easily accessible to protein. The structures of six different protein-RNA complexes feature a peptide or portion of a protein bound in the groove of an RNA helix, and are discussed here.

Arginine-rich sequences

Three groove-binding peptides belong to an arginine-rich class of protein sequences. (The term ''arginine-rich motif'' is still in use for these peptides, though their very different structures and lack of conserved arginine positions should disqualify them from being called a motif.) Despite the extensive use of arginine in each case, these peptides provide a broad picture of the strategies available to a protein recognizing the RNA major groove.

The Rev peptide adopts an α-helical conformation upon binding to an RNA hairpin (Tan *et al.*, 1993). Binding studies with a systematic series of alanine substitutions identified positions critical for RNA recognition (Tan *et al.*, 1993), and a second series of Arg → Lys variants identified positions where only positive charge was important for binding (Tan & Frankel, 1994). In the RNA, selection experiments had identified bases whose identity are important for recognition, including an unusual purine-purine mismatch (Bartel *et al.*, 1991). The structure of the complex was subsequently solved by NMR methods (Battiste *et al.*, 1996). Two purine-purine mismatches and a bulged base (U72) create an unusual S-shaped backbone in which G71 flips over and becomes parallel with G48, with which it pairs (Figure 1(a)). The net effect is a substantial widening of the major groove to accommodate the α-helix. The NMR structure of a complex between the same peptide and an RNA hairpin selected to have high binding affinity shows similar features (Ye *et al.*, 1996). In that RNA, an A·A pair isosterically substitutes for the G48·G71 pair, and a two base bulge places a U in position to make a Hoogsteen pair with an A at the equivalent position of C69. The extra U is sandwiched between Arg35 and Arg38; the additional non-polar contacts may account for its selection.

Schematics of protein-RNA contacts within the Rev-RRE complex are shown in Figure 1(a), and reveal a simple pattern of electrostatic, hydrogen bond, and non-polar contacts. The left-hand panel shows the appearance of amino acid side-chains on the surface of the α-helix. Strips of five arginine residues (upper left in Figure 1(a)) on one side of the helix and two arginine residues on the other side probably interact electrostatically with the RNA, on the basis that substitution with an alanine but not lysine residue is deleterious (Tan & Frankel, 1994). Four of these closely approach non-bridging phosphate oxygen atoms (<4 Å), but the remainder are not clearly associated with a specific phosphate (6-7 Å distance to nearest backbone charge). These are presumably sensing the electrostatic field of the RNA. Conversely, substitution of alanine for arginine residues at positions 41, 42, or 43 has negligible effect on RNA binding affinity (Tan *et al.*, 1993), even though each of these arginine residues closely approaches a phosphate oxygen atom (N—O distances of 3-4 Å). Some of the arginine side-chains may be poorly determined or disordered in the NMR experiments, contributing to uncertainties in their positions relative to the phosphate groups; other potential problems with interpreting the binding affinities of mutants in terms of electrostatic contributions are discussed below.

A group of four residues (three Arg and one Asn) lie between the two sets of electrostatically interacting amino acids and hydrogen bond to bases. A little more than two turns of α-helix spans the equivalent of seven base-pairs. This is a more extensive set of base-specific contacts than most DNA-binding proteins, which typically span three or four base-pairs with one subunit. Aliphatic carbon atoms from an arginine and an alanine residue make non-polar contacts with *endo* faces of backbone riboses, and a single tryptophan residue stacks against a bulged adenine base. All of the amino acid side-chains that hydrogen bond to bases, as well as Thr34 that hydrogen bonds to a phosphate, gave ten to 40-fold reductions in binding affinity when mutated to alanine (Tan *et al.*, 1993). Mutation of Arg43 or Trp45 did not affect RNA binding detectably, despite the non-polar contacts made by these residues. The themes of base-specific hydrogen bonds and non-polar contacts with sugars are repeated in other groove-binding peptides.

The N protein of phage λ prevents transcription termination of some phage operons. The N-terminal peptide of the protein is disordered, but folds into an α-helix upon binding the box B RNA hairpin (Tan & Frankel, 1995; Van Gilst *et al.*, 1997). An NMR-based structure of the complex shows that the peptide α-helix binds a GNRA-like tetraloop,

and from there accesses the major groove of the adjoining helix (Legault *et al.*, 1998). The α-helix extends about four base-pairs into the RNA helix, and ''caps'' the tetraloop by stacking a tryptophan residue on top of the terminal base (Figure 1(b)). The α-helix is severely bent to follow the contour of the RNA helix. Since the RNA surface recognized by the λ N peptide is similar to that recognized by the Rev peptide, it is not surprising that the pattern of contacts is qualitatively similar: two residues hydrogen bond to bases, these are flanked by electrostatic contacts, and several side-chains make non-polar contacts with the backbone. Particularly notable is Ala3, which is between the sugars of C4 and C5 and close to C5—C6 bonds of these two pyrimidines. The interaction may be modestly selective for pyrimidines, as inversion of the base-pairs to place G bases at positions 4 and 5 causes a sixfold reduction in anti-termination activity (Chattopadhyay *et al.*, 1995). A total of 19 amino acids have been substituted for Ala3 (Su *et al.*, 1997); only the serine-substituted peptide is comparable in activity to the parent sequence. A requirement for a small side-chain might be expected from the way Ala3 is surrounded by RNA. In contrast to the tryptophan-base stacking interaction in the Rev-RRE complex, mutation of Trp18 in the N peptide is quite deleterious (Su *et al.*, 1997). Other aromatic residues may substitute for a tryptophan residue, though with reduced binding affinity.

As in the Rev-RRE complex, the roles of basic residues were not accurately deduced from mutagenesis experiments. Arg7, Arg8, and Arg11 could not be replaced by a lysine residue without reduction in the binding affinity, while positions 6, 10, and 14 could be switched between arginine and lysine but not alanine residues. It was therefore concluded that arginine residues 7, 8 , and 11 provide ''base or architecture-specific'' contacts and positions 6, 10 and 14 provide general electrostatic interactions (Su *et al.*, 1997). However, it is not obvious from the structure that positions 8 and 11 interact with the RNA other than electrostatically, and Lys14 is more distant from phosphate groups (6.4-7.0 Å) than most other basic residues in the peptide. Measurement of the salt dependence of the binding constant would be able to distinguish purely electrostatic effects of a mutation from other interactions the mutation might affect.

The bovine immunodeficiency virus (BIV) Tat protein contains an arginine-rich peptide that binds a hairpin (TAR) in the viral mRNA. NMR-based structures of nearly the same peptide and RNA hairpin sequences have been solved by two groups with essentially the same results (Puglisi *et al.*, 1995; Ye *et al.*, 1995). The helix contains two bulged bases, one of which (U12) is disordered and can be deleted without affecting peptide binding (Puglisi *et al.*, 1995), and the other of which (U10) sits in the major groove making a Hoogsteen base triple with A13 (Figure 1(c)). This kinks the helix and opens the major groove wide enough for the peptide to bind as an irregular β-hairpin. Both β-strands contact the RNA helix (Figure 1(c)). Base-specific hydrogen bonds are made from three arginine side-chains and two backbone positions. An isoleucine is packed against the U10 base triplet; this interaction is important to the formation of the complex but can be substituted by several other hydrophobic residues with only four to fivefold reduction in binding affinity (Chen & Frankel, 1995). Even U10 can be substituted by C or A with only twofold loss of affinity (Chen & Frankel, 1994). None of the three glycine residues can be mutated to alanine without substantial loss of binding activity: the α-carbon atoms of Gly71 and Gly74 are close to the A21 ribose and U21 base, respectively, where introduction of a β-carbon would cause steric clash, and Gly76 is important for the hairpin turn.

There is an interesting parallel between the BIV Tat-TAR complex and *met* repressor recognition of DNA (Somers & Phillips, 1992). The repressor binds DNA as a dimer, placing a two-stranded antiparallel β-ribbon in the major groove and using lysine and threonine residues from each strand to make base-specific contacts. Other parts of the protein make additional contacts with the DNA backbone. Thus both a DNA and an RNA binding protein use a β-ribbon to recognize a specific sequence from the nucleic acid major groove (see Figure 2).

Recognition of RNA helix grooves by proteins

The C-terminal domain of ribosomal protein L11 folds into three α-helices that are superimposable with α-helices of the homeodomain class of DNA-binding proteins (Xing *et al.*, 1997). The surface of α-helix 3 contacts RNA, just as homeodomain helix 3 lies in the major groove of DNA (Hinck *et al.*, 1997; Conn *et al.*, 1999). However, contacts are made with an RNA minor groove that has been distorted to present a relatively flat surface, in contrast to the major groove nucleic acid binding surfaces of homeodomains and the arginine-rich peptides discussed above. L11 is the only example of minor groove recognition among the RNA-protein complexes whose structures have been deduced by NMR or crystallography, though minor groove recognition of an acceptor stem G·U wobble pair has been demonstrated for alanyl-tRNA synthetase (Musier-Forsyth *et al.*, 1991; Musier-Forsyth & Schimmel, 1992).

A schematic of the recognized RNA helix surface is shown in Figure 1(d). Distortions of the RNA helix are maintained by tertiary interactions with other parts of the RNA domain. The most unusual of these is the formation of a Hoogsteen base-pair between U1060 and A1088, which intercalates into the helix major groove from a bulge loop elsewhere in the molecule. The entire U1060 nucleotide is flipped over so that the uridine base presents its major groove edge in the minor groove surface of the helix;

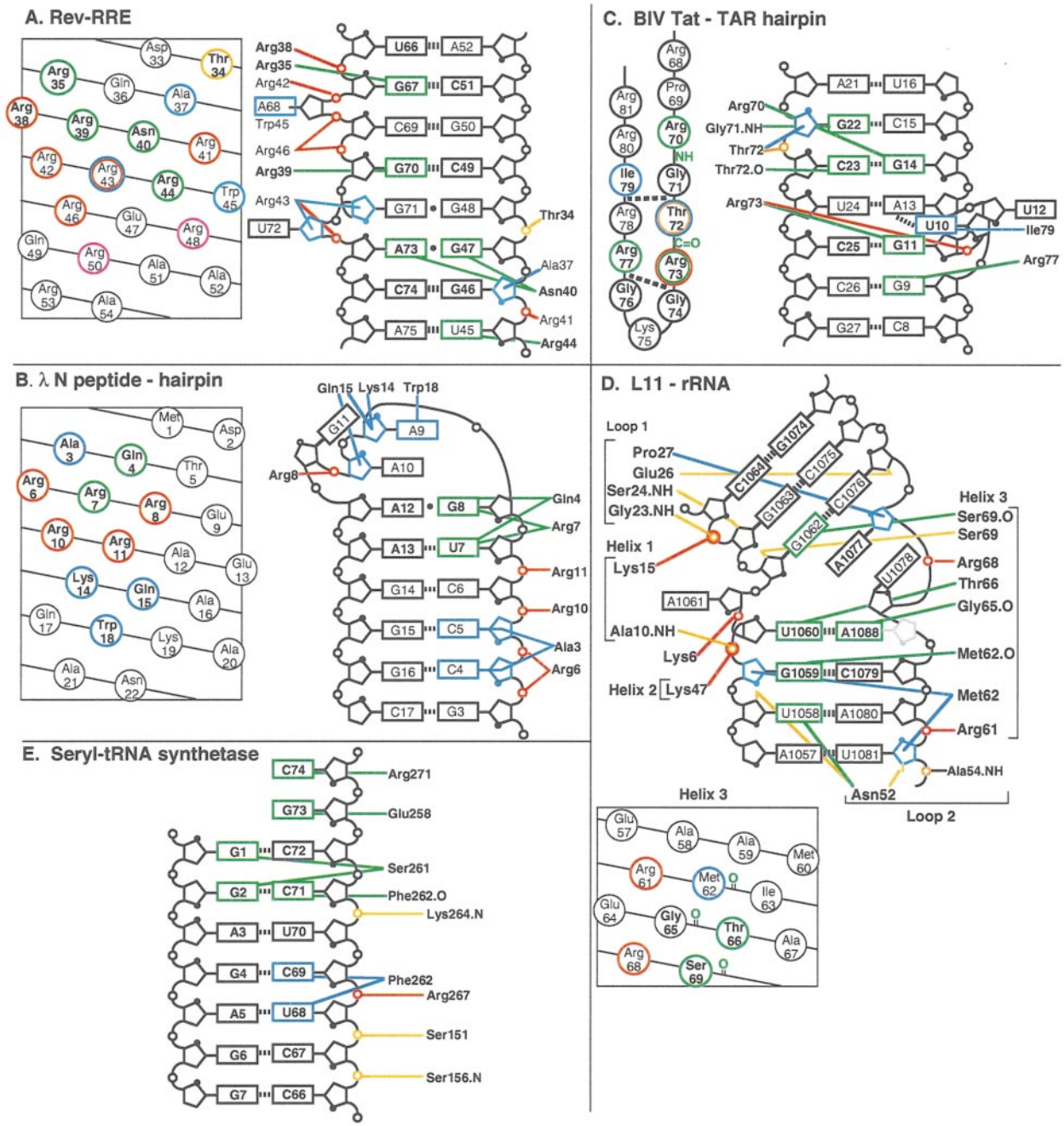

Figure 1. Contacts between peptides and major or minor grooves of RNA helices. The following color code applies to both protein and RNA schematics: red, positive charge is within 4 Å of a non-bridging phosphate oxygen atom; magenta, positive charge is 6-7 Å from a phosphate oxygen atom; green, hydrogen bond to base; orange, hydrogen bond to backbone; blue, non-polar contact. Contacts are made by the amino acid side-chain unless specified as amide (NH) or carbonyl (O). Dots on sugars indicate 4′O position. (a) Contacts between the Rev α-helix and an RNA major groove. The protein sequence is drawn on a two dimensional representation of the outside of an α-helix, as if a cylinder wrapped around the helix has been cut lengthwise and flattened (Crick, 1953). Amino acids in bold have at least a tenfold effect on binding affinity when mutated to alanine (Tan *et al.*, 1993). Bold nucleotides are invariant among RNAs selected from randomized sequences (Bartel *et al.*, 1991). The diagrammed interactions are based on the energy-minimized average structure from NMR studies (1left), which differs in several contacts from the diagram by Battiste *et al.* (1996). (b) Contacts between the λ phage N peptide and an RNA hairpin major groove (Legault *et al.*, 1998). Bold face residues in the α-helix reduce binding to less than ~20 % of control in a gel shift assay (Su *et al.*, 1997). Bases in bold face cannot be mutated without >tenfold reduction in activity (Chattopadhyay *et al.*, 1995) or binding affinity (Cilley & Williamson, 1997). The diagram is based on an averaged and minimized structure from NMR studies provided by P. Legault. (c) Contacts between BIV Tat protein and an RNA hairpin, based on PDB files 1mnb (Puglisi *et al.*, 1995) and 1biv (Ye *et al.*, 1995). The peptide is drawn in its β-ribbon conformation, with its two amide-carbonyl hydrogen bonds indicated as broken lines. Amino acid residues in bold face indicate <10 % activity of the peptide when substituted with alanine, in an *in vivo* assay (Chen & Frankel, 1995). Bases in bold face are pos-

the same S-shaped backbone conformation was observed for G71 in the Rev-RRE complex (Figure 1(a)). In both cases, a bulged base 3′ to the flipped base allows the backbone the extra length needed to rotate the nucleotide. On the other side of the helix, A1088 has displaced U1078, which then stacks with A1077 and causes an interruption in regular base stacking through the helix.

L11 α-helix 3 makes four hydrogen bonds to bases, three of which involve backbone carbonyl groups (Figure 1(d)). Gly65 is essential to allow close approach of the adjacent α-helix carbonyl group to the RNA; its mutation to alanine reduces binding by about tenfold (Xing *et al.*, 1997). The BIV Tat peptide requires glycine residue at certain positions for a similar reason. Backbone contacts are made by basic residues on one side of the helix and by non-polar contacts of one methionine side-chain with ribose sugars. The Ser69 hydroxyl hydrogen bonds to a ribose 2′ hydroxyl, which is normally accessible only from the minor groove of a helix. α-Helix 3 is flanked at either end by loops that primarily interact with the RNA helix backbone, *via* hydrogen bonds to four 2′-OH and two phosphate groups and one non-polar contact between Pro27 and a ribose; the one base-specific contact is a hydrogen bond from Asn52 to U1058.

The effects of L11 binding to this site propagate through a 58 nt domain of tertiary structure: in melting experiments, no RNA unfolding can be detected until the protein dissociates from the RNA (Xing & Draper, 1995). L11 accomplishes this directly by interacting with a tertiary base-pair, U1060·A1088, that links two parts of the RNA, and indirectly by stabilizing the distorted helix structure that accommodates the intercalated A1088 and promotes other tertiary interactions.

Aminoacyl-tRNA synthetases frequently rely on ''discriminator'' bases in the acceptor helix and 3′-terminal sequence to distinguish different isoaccepting tRNA species; contacts must therefore be made in one of the acceptor stem grooves. The crystal structure of seryl-tRNA synthetase bound to its cognate tRNA shows how an irregular protein loop can extend deeply into the major groove of a helix to distinguish a base sequence (Figure 1(e)). An unusual base-specific contact is the packing of the ring of Phe262 against the 5,6 carbon atoms of two pyrimidines. In experiments with variant tRNA minihelices, the synthetase prefers pyrimidines at both these positions by factors of five- to 50-fold (Saks & Sampson, 1996); the N7-C8 bond of a purine would make a less favorable contact with Phe262. Favorable interaction between the edge of one aromatic ring and the face of another has been observed in proteins (Burley & Petsko, 1985).

A last example of RNA groove recognition is a retroviral nucleocapsid protein bound to an RNA hairpin packaging signal (De Guzman *et al.*, 1998). The protein contains two ''zinc knuckle'' domains that recognize the four-base loop. Nine residues of an N-terminal tail form a 3_{10}-helix that reaches into the adjacent major groove. Several basic residues hydrogen bond to phosphate groups, and two residues, an asparagine and an arginine, are within hydrogen bonding distance of bases. Use of a 3_{10}-helix is unprecedented among DNA-binding proteins.

Summary of RNA groove binding strategies

RNAs from many sources form extensive canonical secondary structures, yet it is rare to find even a full turn of helix uninterrupted by mismatches, bulges, or loops. The RNA-complexes discussed here, as well as other RNA structures that have been recently determined (Correll *et al.*, 1997, 1998), show that such non-canonical interruptions may preserve an approximately helical structure but with irregular dimensions considerably different from *A*-form. Proteins binding to structured RNAs are therefore confronted with a diverse array of major and minor groove surfaces that present considerable opportunity for site-specific recognition. The structures discussed here range from a nearly flat minor groove, to major grooves widened by bulges and mismatches, to *A*-form major grooves accessed from the ends. In turn, proteins have developed a number of strategies to create ''probes'' for grooves of different dimensions: α-helix, 3_{10}-helix, β-ribbon, and irregular loop have all been observed. Proteins have been able to further adjust the effective size of these probes by using glycine residues to allow close approach of the RNA to backbone amide and carbonyl groups.

To illustrate the diversity of the structures that have been discussed, Figure 2 shows ''end-on'' views of some of these protein-groove complexes. Two DNA-binding proteins that place either an α-helix or β-ribbon in the major groove are shown for contrast. The two RNA-binding peptides take advantage of the deeper RNA major groove to

itions at which mutations reduce apparent binding affinity >100-fold in a gel shift assay (Chen & Frankel, 1994). (d) Contacts between the C-terminal domain of ribosomal protein L11 (L11-C76) and its rRNA binding site, based on PDB file 1qa6 (Conn *et al.*, 1999). In the schematic of α-helix 3, amino acid residues in bold give >tenfold reduction in binding affinity when mutated to alanine (Xing *et al.*, 1997). Bases in bold are conserved in >97% of available large subunit rRNA sequences from bacteria, archaea, and eucarya (Gutell *et al.*, 1992; Conn *et al.*, 1999). The grey sugar for A1088 indicates that it is not contiguous with the RNA chains shown, but intercalates into the helix from a bulge loop elsewhere in the molecule. (e) Contacts between seryl-tRNA synthetase and the acceptor stem of $tRNA^{Ser}$ (Cusack *et al.*, 1996).

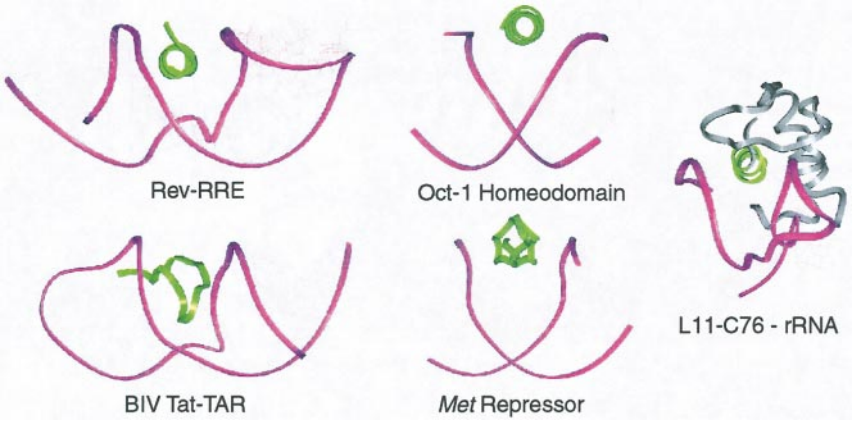

Figure 2. Proteins binding RNA and DNA grooves and discussed in the text. The two DNA-protein complexes show only α-helix 3 of Oct-1 (Klemm *et al.*, 1994) or the β-ribbon of *met* repressor (Somers & Phillips, 1992), which hydrogen bond to bases in the major groove. In the L11-C76-RNA complex, α-helix 3, which is structurally homologous to α-helix 3 shown in the Oct-1 complex, is colored.

bury more surface area against the RNA than possible for a protein binding *B*-form DNA. At the other extreme, L11 supplements its α-helix with loops at either end to create a large, nearly flat surface for binding.

The RNA-protein contacts in these complexes are dictated by the chemical nature of RNA grooves. Compensation of the RNA negative charge by basic residues and hydrogen bonding to the edges of bases are two obvious strategies which all of the proteins and peptides discussed here take advantage of; the same strategies are used by proteins binding specific duplex DNA sequences. More surprising is the major role of non-polar contacts with ribose and the 5,6 bond of pyrimidines in many complexes. Hydrophobic amino acids frequently contact 5-methyl groups of thymine bases in DNA, and the TATA binding protein makes extensive use of aromatic and hydrophobic residues in binding the minor groove of its DNA target sequence (Kim *et al.*, 1993), but the use of non-polar contacts seems much more common among RNA-binding proteins.

Sequence-specific RNA recognition by β-sheet proteins

Loops and bulges of folded RNAs may leave some nucleotides relatively unstructured, and some RNAs (such as mRNAs) contain regions of single-stranded or poorly structured nucleotides. A large fraction of the protein-RNA complexes of known structure target such regions. In contrast to the groove binders discussed above, these proteins tend to pull unstacked bases into binding pockets, composed partly or entirely of β-sheets, and ignore the RNA backbone. Several examples of this sequence-specific binding strategy are discussed in this section.

Ribonucleoprotein consensus proteins: recognition of two RNAs by Human U1A protein

One of the most widespread RNA binding domains is the so-called ribonucleoprotein (RNP) consensus sequence, sometimes also referred to as the RNA recognition motif (RRM). This class of proteins has an βαββαβ fold in which the first and third β-strands are the middle two strands of an antiparallel sheet and contain characteristic aromatic residues (Figure 3). There are now four atomic resolution structures of RNP proteins bound to cognate RNAs: human U1A domain bound to a 21 nt RNA hairpin from U1 snRNP (Oubridge *et al.*, 1994); the same protein bound to an mRNA internal loop (Allain *et al.*, 1996, 1997; Howe *et al.*, 1998); the related U2B″-U2A′ protein complex bound to a U2 snRNP hairpin; and two RNP domains of the sex-lethal gene bound to a single-stranded RNA sequence (Handa *et al.*, 1999). (PDB entries for these complexes are 1urn, 1a9n, 1aud, and 1b7f, respectively.) Since these proteins are homologous but recognize different sequences, the structures provide an opportunity to ask how sequence specificity is achieved.

The U1A RNP domain recognizes the same RNA sequence (AUUGCAC) either in the context of a hairpin loop (stem-loop II of U1 snRNP) or internal loop (U1A mRNA polyadenylation inhibition element, PIE) (Scherly *et al.*, 1989; van Gelder *et al.*, 1993). The identical seven nucleotide sequence was selected by U1A protein from pools of random sequence RNA (Tsai *et al.*, 1991), either in the context of a hairpin loop or an unstructured sequence. These results have suggested that U1A has strong sequence specificity and little regard for RNA structure. The two U1A protein-RNA structures in fact show that the protein hydrogen bonds extensively to bases of the AUUGCAC sequence. A schematic of protein-RNA contacts is shown in Figure 3. In the U1A-RNA hairpin crystal structure, a total of 14 hydrogen bonds between the protein and the seven bases have been identified; several more water-mediated base-protein hydrogen bonds were identified in the crystal structure but are not shown in Figure 3 (Oubridge *et al.*, 1994). The NMR structure of U1A with the PIE internal loop shows essentially the same conformation of the AUUCGAC sequence and pattern of hydrogen bonds with the protein (Allain *et al.*, 1997). Three of the bases (G9, C10, and A11) are stacked against amino acid side-chains, all three of which are conserved as aromatic residues in RNP

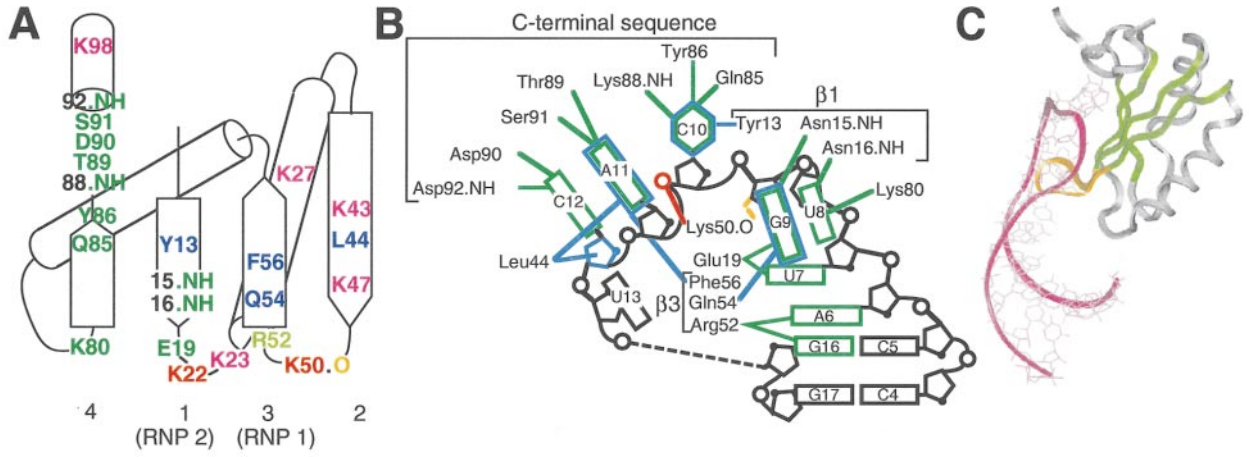

Figure 3. U1A protein complexed with an RNA hairpin. (a) Protein secondary structure with residues contacting RNA highlighted. Color-coding of the schematics are as described in the legend to Figure 1. Magenta indicates basic residues that do not contact RNA but for which substitution with glutamine is deleterious (Nagai *et al.*, 1990); these presumably interact with the RNA electrostatic field. The β-strands are numbered at the bottom of the figure. RNP1 and RNP2 refer to conserved sequences motifs first used to identify the RNP consensus proteins and corresponding to the middle two strands of the β-sheet (reviewed by Dreyfuss *et al.*, 1993). (b) RNA contacts with protein, based on the co-crystal structure (Oubridge *et al.*, 1994). A broken line substitutes for two bases that are disordered in the crystal. (c) structure of the U1A protein with bound RNA (PDB file 1urn). The β2-β3 loop (orange) inserts into the hairpin loop next to the closing base pair, and loop bases are held against the β-sheet surface (green) by residues at the protein C terminus.

consensus proteins. (One of the positions is Gln in U1A, but Phe or Tyr in most other RNP consensus proteins.) Since stacking is more favorable for purines than pyrimidines, the three aromatic protein residues might be expected to contribute a small sequence bias at the three interacting RNA bases. The fact that these positions are conserved as aromatic residues in RNP consensus proteins with many different sequence specificities suggest that their primary role is to orient bases on the protein surface, rather than select for particular sequences.

As suggested by the pattern of protein-RNA contacts in Figure 3, RNA bases are buried in the U1A structure and the phosphate-sugar backbone is relatively exposed to solvent. There are only two salt bridges between lysine side-chains and RNA phosphate groups (Lys22 hydrogen bonds to the phosphate of A2, which is not shown in Figure 3) and no other basic residues come within 5.0 Å of a phosphate oxygen atom. The same is true of the U1A-PIE complex. This observation is in surprising contrast to solution studies, which showed that (i) most phosphate groups within the hairpin loop and several in the stem are protected by U1A protein from reaction with ethylnitrosourea (Jessen *et al.*, 1991); (ii) hairpin binding affinity for U1A is very sensitive to salt concentration, weakening by a factor of 200 when NaCl concentration is raised from 250 to 500 mM (Hall & Stump, 1992); and (iii) besides the basic residues hydrogen bonding with bases or phosphate groups, mutation of six more basic residues to glutamine reduce the stability of the complex (Nagai *et al.*, 1990). The steep salt dependence of binding unambiguously indicates that the free energy of complex formation has a large electrostatic component. However, the mutagenesis and protection studies were quite misleading as to the extent of direct contacts of basic residues with phosphates, and prompted a model in which basic residues lined the RNA backbone and left bases exposed to solvent (Jessen *et al.*, 1991). It seems that strong electrostatic interactions in this RNA may not be associated with close distances between phosphate groups and basic residues.

The AUUGCAC sequence in the context of an unstructured RNA binds U1A poorly (Tsai *et al.*, 1991; Hall, 1994). Direct interactions of the protein with the stem of the RNA accounts for some of the additional binding free energy; thus Arg52 hydrogen bonds to the closing G·C base-pair in both the hairpin and internal loop RNA complexes, and additional backbone contacts are made by one or two residues in each complex. In the hairpin complex, the three bases of the loop 3′ to the recognized sequence are disordered. The equivalent positions in the U1A-PIE complex are occupied by other parts of the internal loop, and make several hydrogen bonds with the protein.

The RNA binding surface of U1A is formed mainly from the β-sheet surface and two other regions of the protein (Figure 3). The loop between β2 and β3 inserts into the middle of the RNA hairpin loop, bringing the N terminus of β3 into contact with the closing base-pair. Much of the recognized RNA sequence is then clamped in a groove formed from the β-sheet surface and the C-terminal sequence of the protein. The β2-β3 loop is somewhat disordered in the absence of RNA (Nagai *et al.*, 1990), and the C-terminal region probably rearranges upon RNA binding (Avis *et al.*, 1996). The RNA hairpin loop is also largely unstructured in the absence of protein (Oubridge

et al., 1994). Thus recognition should be viewed as a mutual adaptation of the structures of each macromolecule in which formation of both intra- and intermolecular contacts are important.

The mutual adaptation of the protein and RNA binding surfaces may account for the high cooperativity of forming the complex. Single base mutations within the seven nucleotide recognition sequence reduce the magnitude of the binding free energy by 1.4 to 7.9 kcal/mol (the larger numbers depend on a linear extrapolation of binding constants measured at low salt concentrations) (Hall, 1994). The smaller free energy changes might be expected for removal of one or two hydrogen bonding interactions (Fersht, 1987), but the very large effects associated with mutation of A6 → C (5.2 kcal/mol), G9 → A (7.6 kcal/mol), and A11 → G (7.9 kcal/mol) imply that these base substitutions prevent formation of large groups of contacts and not just the hydrogen bonds made directly to the mutated base.

Sequence discrimination by two similar RNP consensus proteins

U2B″, a homolog of U1A differing in only 22 out of 95 positions, binds to a hairpin from U2 snRNP that differs in only a few nucleotides from the U1 snRNP hairpin recognized by U1A. The general binding strategy and RNA contact surfaces in U1A and U2B″ are very similar (Price *et al.*, 1998). (U2B″ requires a second protein, U2A′, in order to bind RNA. U2A′ binds on the opposite face of U2B″ from the RNA and will not be discussed further here, though it may have some indirect effect on binding specificity (Rimmele & Belasco, 1998).) As swaps of relatively few nucleotides or amino acids are able to change the preferences of U1A and U2B″ proteins for the two RNA hairpins (Scherly *et al.*, 1990), it has been interesting to compare structures of the U1A and U2B″ complexes and ask if simple principles govern the differences in specificity.

The U2 snRNP hairpin contains the sequence AUUGCAG, which differs from the seven base recognition sequence of the U1 snRNP hairpin in only the last base. The six identical bases of each hairpin interact with their cognate protein in essentially the same way, making identical contacts with C-terminal and β1 residues. The seventh base is a C in the U1 hairpin, and fits tightly against the protein surface. In the U2 hairpin, a G base fits at the same position by adopting the *syn* conformation. The remaining three bases of the U1 hairpin loop remain disordered in the complex, and in fact can be replaced by ethylene glycol spacers without affecting the binding affinity (Williams & Hall, 1996). In the U2 hairpin, these three nucleotides have been replaced by four well-ordered nucleotides that make a number of contacts with the protein. Three of the protein residues holding this part of the RNA in place are identical in U1A; thus the extra nucleotide in the hairpin loop has enabled it to adopt a conformation which makes additional protein contacts.

A particularly interesting residue is Arg52, which hydrogen bonds in different ways to bases in the two complexes. In the U1A-hairpin complex, the Arg52 guanidinium group interacts with the major groove edge of G16 in the base-pair closing the loop, and also with N1 of the neighboring A6 (Figure 3). The closing pair of the U2 hairpin is a U·U mismatch, rather than a C·G pair. Arg52 has rotated to hydrogen bond its guanidinium group to U17 phosphate and N1 of A6, and in doing so has placed its N^{ε} atom next to O2 of U7. Thus a small change of the Arg52 side chain conformation has been used to maintain extensive hydrogen bonding in both complexes. These kinds of adaptations make it difficult to deduce information about protein-RNA contacts from mutagenesis studies, as substitution of a base or amino acid side-chain may create a different, compensating set of interactions. In this example, the mutation Arg52 → Gln is very deleterious to U1A binding while Arg52 → Lys is not, from which it was erroneously concluded that position 52 only interacts with RNA electrostatically (Jessen *et al.*, 1991).

A set of mutagenesis experiments has shown that non-polar contacts are an important aspect of discrimination in this system. In the U1A-hairpin complex, C^{δ} of Leu44 is placed against C2 of A11 and C1′ of C12. Leu44 has been replaced by Val44 in U2B″. The smaller size of Val as compared to Leu, and the different positioning of the U2B″ *syn* G12 as compared to C12, allows additional contacts of the Val44 methyls with G12 C8; these contacts potentially stabilize the *syn* conformation of G12 and discriminate against C at this position. Rimmele & Belasco (1998) showed that substitution of Leu for Val in U2B″ increases its relative affinity for the U1 hairpin fivefold while decreasing its affinity for the U2 hairpin threefold. Thus very subtle differences in geometry and non-polar contacts can significantly alter binding specificity. U2B″ makes additional non-polar contacts between Leu46 and U13, and Thr48 and C16; these interactions could also be favoring the U2 hairpin (both of these residues change to Ser in U1A).

A crystal structure of the sex-lethal protein bound to a 17mer RNA, U_5GUUGU_7, appeared recently (Handa *et al.*, 1999). This protein contains two RNP consensus domains that combine to recognize a continuous stretch of nine nucleotides (UGU_7). The two protein domains are similar to the U1A and U2B″ proteins discussed above, and use approximately the same surfaces to contact RNA. Each of the nine recognized bases makes at least one hydrogen bond to the protein, and several aromatic and hydrophobic side-chains stack against bases as well. As the RNA is completely unstructured, recognition depends entirely on sequence.

In summary, extensive hydrogen bonding between RNP consensus proteins and bases of their cognate RNAs is probably responsible for

much of their sequence specificity. But this simple ''readout'' of a single-stranded RNA sequence is modulated in two ways. (i) Both the RNA and protein are partially unstructured as free macromolecules, so a lock-and-key type of docking between two complementary surfaces does not apply. Cooperative folding of the molecules upon binding results in single mutations having very large and unpredictable consequences. (ii) Non-polar residues also factor into the selectivity of a protein, as they can pack tightly around bases and promote discrimination based on size (pyrimidine *versus* purine) or ability to adopt unusual conformations (e.g. *syn* glycosidic bond). Both hydrogen bonds and non-polar contacts are easily enumerated from a crystal structure, but their consequences for binding free energy and discrimination cannot be inferred from structure alone; further thermodynamic studies are clearly needed to understand the basis for sequence discrimination in these systems.

RNA recognition by β-barrels

Many aminoacyl-tRNA synthetases directly sense the anticodon sequence of their cognate tRNAs. In aspartyl, asparaginyl, and lysyl-tRNA synthetases, this discrimination is performed by a β-barrel structure which belongs to the oligonucleotide/oligosaccharide binding (OB fold) class of proteins. This set of proteins is divided into several groups, most of which bind oligonucleotides or oligosaccarides (Murzin, 1993) and in some cases share similarity with RNP consensus proteins in the sequence of one β-strand (RNP1) (Landsman, 1992). The crystal structure of the aspartyl-specific enzyme with its cognate RNA shows that the three anticodon bases are pulled against the β-sheet surface, where a phenylalanine residue stacks against U35 and helps to orient the RNA on the protein (Cavarelli *et al*., 1993). This Phe-U35 feature is conserved in all members of this group of tRNAs and synthetases. As in RNP consensus proteins, there are two loops that extend from between two pairs of β-strands and help clamp the RNA against the β-sheet.

The recent structure of a rho protein-single stranded RNA complex shows another example of sequence-specific contacts by an OB fold protein (Bogden *et al*., 1999). The protein is a hexameric, RNA-dependent ATPase involved in transcriptional termination, and binds preferentially to poly(C). Two cytidine residues are placed against the β-barrel surface and enclosed by an overhanging loop, approximately as in aspartyl-tRNA synthetase. Hydrogen bonds are made to all three positions on each base, and a phenylalanine residue stacks against one of them.

TRAP recognition of repeating G/UAG

The *trp* RNA-binding attenuation protein (TRAP) of *Bacillus subtilis* recognizes a specific sequence in the leaders of the *trp* operon RNA and the *trp*G gene, where it affects formation of a transcription terminator secondary structure and/or translational efficiency (Babitzke, 1997). The protein-RNA binding affinity is modulated by tryptophan, so that expression of *trp* genes is linked to the cellular level of tryptophan. The protein is an unusual ring of 11 identical subunits containing only β-sheet secondary structure (Antson *et al*., 1995). TRAP correspondingly recognizes nine to 11 repeats of trinucleotide sequence, G/UAG, optimally separated by two nucleotides (Babitzke *et al*., 1995). Alanine scanning mutagenesis identified three basic residues, located on β-strands on the outside perimeter of the protein, that are essential for RNA recognition but do not affect tryptophan binding. At position 56 only lysine is functional, while residues 37 and 58 may be either lysine or arginine. Further studies with modified nucleotides have shown that the protein is relatively insensitive to the identity of the 5′ trinucleotide position, as elimination of the base (leaving an abasic site in each instance of 11 repeats) reduces binding only tenfold. In the second position, the exocyclic amine of adenine is essential, and the protein recognizes both O6 and N2 of guanine in the third position (Elliott *et al*., 1999). In addition, the only 2′-OH group required is in the third nucleotide of each repeat.

The crystal structure of a TRAP homolog from *Bacillus stearothermophilus* bound to $(GAGAU)_{11}$ RNA was recently solved (A.A. Antson *et al*., unpublished results). As in other β-sheet proteins recognizing single-stranded RNA, the RNA backbone is on the outside and the bases are placed against the β-sheet surface. Consistent with this, RNA binding is nearly independent of salt concentration (Baumann *et al*., 1996). The bases of the GAG repeats hydrogen bond to protein residues largely as expected from mutagenesis studies: adenine N6 and N1 hydrogen bond to Lys37; O6, N1, and N2 of guanine in the third position all hydrogen bond to either Lys56 or Arg58; and the 2′-OH group of the third ribose is hydrogen bonded to a phenylalanine residue. Additional non-polar contacts define pockets for the bases. An unanticipated hydrogen bond is formed between Asp39 and N2 of the first guanine base of the trinucleotide; neither alanine scanning mutagenesis or base substitutions had identified either of these positions as important. A preference for pyrimidines in the spacer nucleotides has been observed in *in vitro* selection experiments (Baumann *et al*., 1997), and was thought to reflect the weaker stacking propensity of these bases. In support of this, the co-crystal structure does not show any protein contact with spacer nucleotides. As expected from the way TRAP binds single stranded RNA, any

propensity for forming secondary structure decreases the apparent binding affinity of an RNA (Xirasagar *et al.*, 1998).

MS2 coat protein recognition of a hairpin: sequence and structure recognition

One of the best studied protein-RNA interactions is the complex between phage R17 coat protein and a hairpin from its RNA genome. This interaction, which causes translational repression of the phage replicase gene, was the first specific RNA-protein complex to be isolated and studied *in vitro* (Spahr *et al.*, 1969) and the first for which RNA sequence requirements for recognition were established in detail (Uhlenbeck *et al.*, 1990). The protein is extremely selective for the correct sequence, discriminating against other hairpins by more than six orders of magnitude in binding affinity (Carey & Uhlenbeck, 1983). The crystal structure of the homologous complex from MS2 phage has now been determined for several RNA and protein variants (Valegård *et al.*, 1994, 1997; van den Worm *et al.*, 1998). It is now possible to ask whether the thermodynamic selectivity of the interaction can be rationalized from the available structural information.

The general features of the recognition complex are simple; RNA contacts are presented in Figure 4. Protein binding is insensitive to the specific base-pairs present in the secondary structure; four of the five remaining unpaired bases are critical. The protein binds as a dimer in which two five-stranded β-sheets combine to form one large β-sheet surface. The RNA hairpin straddles the dimer interface and places A − 10 and A − 4 in symmetry-related pockets of the two monomers. The asymmetry of the RNA positions the two adenine bases differently, and different hydrogen bonds are made (note particularly the changed position of Ser47; Figure 4). Position −5 is stacked against a tyrosine residue on one side and A − 7 on the other. A − 7, which must be a purine for strong binding, does not make any direct contact with the protein. This recognition strategy is similar to that described for the other proteins in this section, in that a β-sheet surface is used to make sequence-specific contacts with RNA bases. In this case the contacted bases are highly constrained in their relative orientations by the structure of the hairpin loop; for instance, A − 4 and A − 10 are held in the correct spatial orientation for binding by the intervening two base-pairs. (The hairpin loop is disordered and the −10 bulge is intercalated in the free RNA (Borer *et al.*, 1995), but the recognized bases are still more restricted in relative position than any of the RNP consensus protein recognition sites discussed above.) To a larger degree than seen in the other β-sheet proteins discussed here, RNA structure is a determinant of recognition specificity.

The changes in binding free energy caused by many variants in both the RNA and protein have been studied in attempts to understand the selectivity of each of the base-protein interactions. A large number of modified bases have been substituted for the bulged A at −10, and can be at least qualitatively rationalized in terms of the crystal structure (Wu & Uhlenbeck, 1987). A − 10 makes three base-specific hydrogen bonds and lies in a non-polar pocket with Lys61 on one face and Val29 on the other (Figure 4). RNAs with pyrimidines at −10 do not bind detectably in most cases (>10^3 fold effect), perhaps because of loss of both stacking and hydrogen bonding contacts. Purine and several purines derivatives, including G, I, and deaza^{7}A, bind with only two- to fivefold reductions in affinity, consistent with their ability to both stack and either accept or donate a hydrogen bond from N1 or N6. m^{1}A reduces binding by >10^3, perhaps because of steric effects of the extra methyl group as well as the elimination of a hydrogen bond; similarly, m$^6_2$A binds more than 100-fold

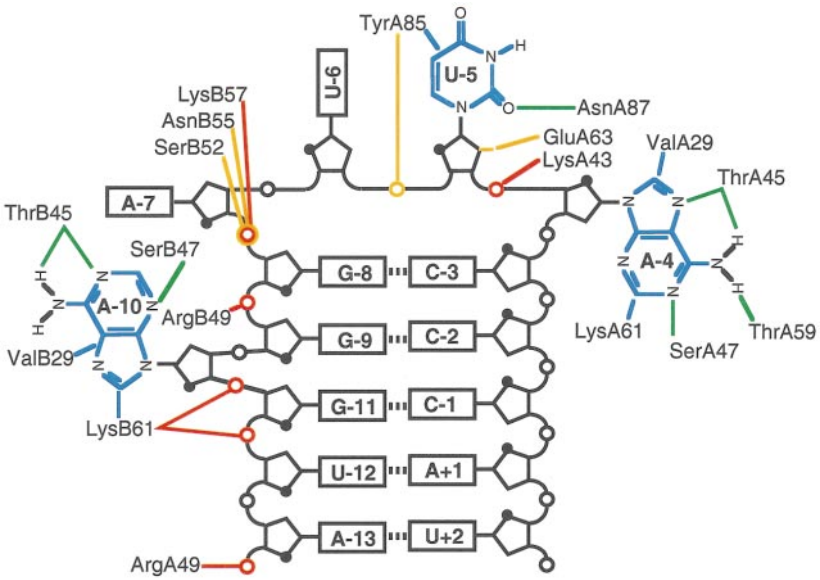

Figure 4. Protein-RNA contacts between MS2 coat protein and a phage RNA hairpin, adapted from Valegård *et al.* (1997). Nucleotide numbering is relative to the replicase initiation codon at A + 1. Color coding of interactions are as described in the legend to Figure 1. Amino acids are labeled A and B, to distinguish the two protein subunits forming the RNA binding site.

less well. Fewer substitutions have been made at the −4 position, where A binds ~10^3 more tightly than C, G, or U (Carey *et al.*, 1983). This degree of discrimination is surprising, as G would be able to form some of the same hydrogen bonds as A.

A crystal structure of the RNA hairpin complexed to a Thr45Ala variant has been solved (van den Worm *et al.*, 1998). The single mutation eliminates hydrogen bonds to both A − 4 and A − 10, yet binds only fivefold less strongly than the wild-type protein (Lago *et al.*, 1998). The crystal structure shows that both A − 4 and A − 10 have moved only slightly, by ~0.2 Å. This is enough to eliminate an additional hydrogen bond (Glu63 to U − 5 O2′). Elimination of this and the Thr45 hydrogen bonds is probably compensated in part by strengthening of other hydrogen bonds to the two nucleotides.

Position −5 is U in the wild-type RNA sequence, but a C substitution surprisingly enhances binding by nearly 100-fold; A or G are discriminated against by ten- or 100-fold, respectively (Carey *et al.*, 1983). Stacking of the base against Tyr85 provides some discrimination, as substitutions at this position are more deleterious for C binding than for U by as much as sevenfold (Johansson *et al.*, 1998). The hydrogen bond between Asn87 and the pyrimidine O2 (Figure 4) contributes virtually no favorable free energy to RNA binding, as its mutation to Ala has no effect on C − 5 or U − 5 RNAs. However, it does serve to discriminate against the larger purine rings, as Asn87Ala binds A − 5 and G − 5 three- to fivefold more tightly (Johansson *et al.*, 1998). These experiments do not point to any direct protein contacts with the base at −5 as a complete explanation for the discrimination at this site. However, a crystal structure of a C − 5 RNA with MS2 coat protein shows that a new hydrogen bond has formed from the C − 5 amino group to U − 6 phosphate, and has moved the −6 and −7 phosphate groups ~0.5 Å closer to the C − 5 base as compared to the wild-type U − 5 RNA. This shift may alter the strength of hydrogen bonds to the two non-bridging −7 phosphate oxygen atoms, and chemical modifications of these oxygen atoms have different effects on C − 5 and U − 5 RNAs. A last factor to be considered is that a C − 5 - phosphate hydrogen bond may help predispose the RNA loop (which is disordered in free RNA; Borer *et al.*, 1995) to the correct conformation for RNA binding.

It appears that coat binding specificity at the RNA −5 position, which ranges over four orders of magnitude with substitutions of the four standard bases, is the net result of several relatively small effects: stacking preferences of the bases for Tyr85; steric occlusion of purines by Asn87; conformational changes propagated through the protein by an intramolecular hydrogen bond in C − 5; and the effect of C − 5 on the unbound RNA conformation. None of these effects involves a direct hydrogen bond between the protein and −5 position, which underscores the difficulty of deducing a structural basis for specificity from hydrogen-bonding patterns in a single crystal structure.

Electrostatic interactions in RNA-protein complexes

All of the proteins discussed in this review derive some of their binding free energy from interactions of basic residues with the RNA electrostatic field, and it is relevant to ask (i) to what extent an RNA-binding protein depends on electrostatic interactions and (ii) how basic residues are arranged on the surface of a protein to enhance binding affinity. Three kinds of experiments have been used in answering these questions, and each approach has distinct limitations which need to be kept in mind.

To distinguish basic residues that interact solely with the electrostatic field from those in which the specific geometry of the side-chain is essential, variants with arginine, lysine, or a neutral residue (e.g. alanine or glutamine) at a position are frequently compared. If the position can tolerate lysine or arginine but not other side-chains, the interactions are presumed to depend only on positive charge and thus to be entirely electrostatic. A specific requirement for an arginine or lysine residue implies that the chemical nature of the side-chain is important, and other kinds of interactions are implied. This approach was first used to distinguish specific RNA contacts from general electrostatic interactions in an arginine-rich peptide from HIV Tat protein (Calnan *et al.*, 1991; Tao & Frankel, 1993), and has since been applied to a number of the complexes discussed here.

While the reasoning behind this approach seems sound, the correspondence between such mutagenesis data and actual structures is much poorer than anticipated. In several cases that have been mentioned, Arg52 of U1A protein and Lys56 and Arg58 of TRAP, mutagenesis of the residue in question fits the criterion of an amino acid that is important only for electrostatic interactions. Contrary to expectation, all of these residues make specific hydrogen bonds to RNA bases. In addition, those residues predicted to be engaged in electrostatic interactions do not necessarily make ''ionic contact'' with phosphate groups as sometimes assumed. The Rev-RRE complex is a good example: several arginine residues that approach phosphate groups as closely as 3.0 Å can be mutated to alanine with little effect on binding affinity, while others that place positive charges 6-7 Å distant from the nearest phosphate were identified as important. In an irregular RNA structure, calculations have shown that there may be large variations in electrostatic field strength which cannot easily be inferred by inspection of the structure (Sharp *et al.*, 1990). Thus the effect of a basic residue mutation on the RNA binding free energy is not necessarily proportional to its distance from the nearest phosphate; calculations are needed to

discern the electrostatic field expected in the vicinity of a basic residue.

Ethylnitrosourea footprinting identifies phosphate groups protected from alkylation when protein binds (Ehresmann *et al.*, 1987); ethylation interference locates phosphate groups at which ethylation weakens protein binding. Both experiments have been presumed to map phosphate groups that hydrogen bond to either basic or polar amino acids. Unfortunately, the correlation between ethylation protection or interference and phosphate-protein hydrogen bonds is poor. The most dramatic example discussed here is U1A protein, which protected many phosphate groups in the recognized RNA loop sequence and prompted a model with the phosphates against the protein and bases exposed to solvent (Jessen *et al.*, 1991), the opposite of the actual case. In the Rev-RRE complex, only three of five phosphate groups at which interference was observed are hydrogen bonded in the complex, and four other hydrogen bonded phosphate groups were missed (Kjems *et al.*, 1992).

Lastly, the salt dependence of protein-RNA binding affinity measures the extent of electrostatic interactions. Since cations screen the electrostatic field of an RNA, increasing salt concentration should decrease the association constant of a protein that relies on electrostatic interactions (Record *et al.*, 1976). When oligopeptides of lysine or arginine bind long single-stranded or double-stranded RNA polymers, each additional basic residue makes the slope of a plot of $\log(K_a)$ *versus* $\log[M^+]$ more negative by a fixed increment which is related to the density of the ion atmosphere around the nucleic acid (Mascotti & Lohman, 1990, 1997). Unfortunately, a simple proportionality between the number of basic residues with electrostatic interactions and the slope of $\log(K_a)$ *versus* $\log[M^+]$ does not hold for RNA binding proteins. First, a proportionality factor can be estimated only for regular, infinitely long polymers. The relatively short RNAs studied in experiments with site-specific binding proteins should have very large ''end effects'' (Olmsted *et al.*, 1989) and their irregular structures require detailed calculations to estimate the extent of ion association. Second, the contributions of basic residues within a protein are not necessarily additive, as has been shown recently with single and double mutants in basic residues of L11 protein (D. GuhaThakurta & D.E.D., unpublished results). There has been an attempt to correlate the number of basic residue-phosphate contacts in the MS2 coat protein-RNA complex with the salt dependence of binding affinity (Carey & Uhlenbeck, 1983; LeCuyer *et al.*, 1996). The interpretation of the salt dependence cannot be correct for the reasons mentioned here. Much more sophisticated calculations, as have been done for some protein-DNA complexes (Misra *et al.*, 1994), are needed to analyze electrostatic interactions in this complex.

Though salt dependence data are difficult to interpret quantitatively, it is possible to identify basic residues making electrostatic contributions by measuring the salt dependence of proteins with neutral substitutions. In this way, six out of nine basic residues tested in L11 protein were found to contribute to salt dependence of RNA binding (D.G. & D.E.D., unpublished results). Five of these residues hydrogen bond to phosphate oxygen atoms, while the sixth is further away from the RNA backbone. This approach is particularly useful as it easily distinguishes mutations that weaken binding without affecting electrostatic interactions (i.e. the binding constant decreases but the slope of $\log(K_a)$ *versus* $\log[M^+]$ is unchanged) from those that affect only electrostatic interactions.

Mutual accommodation

In almost all of the cases discussed here, complex formation is accompanied by some degree of mutual accommodation between the RNA and protein binding surfaces, most commonly as a disorder → order transition upon binding, but sometimes as a conformational rearrangement. Thus NMR and CD studies of the separate and complexed components of the BIV TAT-Tar and Rev-RRE complexes show that both the peptide and RNA fold upon binding (Battiste *et al.*, 1994; Tan & Frankel, 1994; Chen & Frankel, 1995; Puglisi *et al.*, 1995); a 14 residue loop of the L11 RNA binding domain is completely unstructured in the absence of RNA, but becomes as ordered as the rest of the protein upon binding (Markus *et al.*, 1997); and the stacked, U-turn structure of the $tRNA^{Asp}$ anticodon loop is thoroughly disrupted upon binding to its cognate synthetase (Cavarelli *et al.*, 1993). In two cases, the box B hairpin bound by the λ N peptide and the L11-rRNA complex, the free RNA can be stabilized in the bound conformation by single base mutations, but this appears to be the exception rather than the rule. The unfavorable reduction in entropy that accompanies ordering of protein and RNA surfaces should reduce the stability of the complex, and stabilization of either protein or RNA folds can enhance binding affinity (Tan *et al.*, 1993; Xing & Draper, 1995).

One might ask why natural selection has not generated more rigid RNA and protein structures with higher affinities. The simplest possibility is that there may not be selective pressure to construct more elaborate protein or RNA structures with tighter binding affinity. Factors which transiently associate with an RNA should not bind too tightly, and proteins irreversibly associated with an RNA (as in ribosomes) are frequently held in place by cooperative interactions with other proteins. It is also possible that a small, flexible loop in a protein is a more cost-effective way to augment the RNA binding affinity of a protein than synthesis of a larger and more rigid structure. Rigid structures may be undesirable: some RNA target sites must be able to fold and unfold easily (e.g. in

mRNAs), and the ability of a protein to envelop RNA bases with flexible loops could be advantageous for specificity. Lastly, the propensity of RNAs for forming alternative folds (Gluick *et al.*, 1997; Pan & Woodson, 1998) may simply mean that it is difficult to evolve an RNA with unique secondary and tertiary structure.

Similarities between DNA and RNA-binding proteins

Since RNA adopts a much greater variety of non-canonical and tertiary structures than does DNA, one might have expected that the bestiary of RNA-binding proteins would be more exotic than the corresponding collection of DNA-binding proteins. But so far it appears that most of the protein frameworks for RNA recognition are not much different from duplex and single-strand-specific DNA binding proteins. Figure 2 makes this point clear for groove-binding proteins. The same is true for β-sheet proteins; for example, replication protein A holds single-stranded DNA bases against a β-sheet surface (Bochkarev *et al.*, 1997) in exactly the same way as RNP consensus proteins and other β-sheet RNA-binding proteins considered here. Of course RNA-binding proteins can potentially combine domains to recognize large and complex RNA shapes, as do aminoacyl-tRNA synthetases that require the anticodon loop and acceptor stem to be in the correct spatial orientation (Perret *et al.*, 1992). But at the binding domain level, RNA and DNA binding proteins do not appear fundamentally different.

Similarities between DNA and RNA binding strategies imply that the two sets of proteins may not have evolved independently. The strong structural homologies between L11 and homeodomain proteins (Xing *et al.*, 1997) or OB fold proteins binding both RNA and DNA (Bochkarev *et al.*, 1997; Bogden *et al.*, 1999) suggest (but do not prove) evolutionary relatedness. Thus the ribosome, which must include some of the first nucleic acid binding proteins to evolve, may have served as a reservoir of nucleic acid binding motifs during evolution.

Acknowledgments

I thank Dr Paul Gollnick for communicating results prior to publication, Debraj GuhaThakurta for help with illustrations, and Resi Gerstner for a critical reading of the manuscript. This work was supported by NIH grants R37 GM29048 and RO1 GM56968.

References

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature,* **380**, 646-650.

Allain, F. H.-T., Howe, P. W. A., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **16**, 5764-5772.

Antson, A. A., Otridge, J., Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Wilson, K. S., Smith, T. M., Yang, M., Kurecki, T. & Gollnick, P. (1995). The structure of *trp* RNA-binding attenuation protein. *Nature,* **374**, 693-700.

Avis, J. M., Allain, F. H.-T., Howe, P. W. A., Varani, G., Nagai, K. & Neuhaus, D. (1996). Solution Structure of the N-terminal RNP domain of U1A protein: the role off C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* **257**, 398-411.

Babitzke, P. (1997). Regulation of tryptophan biosynthesis: Trp-ing the TRAP or how *Bacillus subtilis* reinvented the wheel. *Mol. Microbiol.* **26**, 1-9.

Babitzke, P., Bear, D. G. & Yanofsky, C. (1995). TRAP, the trp RNA-binding attenuation protein of *Bacillus subtilis*, is a toroid-shaped molecule that binds transcripts containing GAG or UAG repeats separated by two nucleotides. *Proc. Natl Acad. Sci. USA,* **92**, 7916-7920.

Bartel, D. P., Zapp, M. L., Green, M. R. & Szostak, J. W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell,* **67**, 529-536.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1994). Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine-purine base pairs. *Biochemistry,* **33**, 2741-2747.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science,* **273**, 1547-1551.

Baumann, C., Otridge, J. & Gollnick, P. (1996). Kinetic and thermodynamic analysis of the interaction between TRAP (trp RNA-binding attenuation protein) of Bacillus subtilis and trp leader RNA. *J. Biol. Chem.* **271**, 12269-12274.

Baumann, C., Xirasagar, S. & Gollnick, P. (1997). The trp RNA-binding attenuation protein (TRAP) from *Bacillus subtilis* binds to unstacked trp leader RNA. *J. Biol. Chem.* **272**, 19863-19869.

Berg, P. & Ofengand, E. J. (1958). An enzymatic mechanism for linking amino acids to RNA. *Proc. Natl Acad. Sci. USA,* **44**, 78-86.

Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature,* **385**, 176-181.

Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. & Berger, J. M. (1999). The structural basis for terminator recognition by the Rho transcription termination factor. *Mol. Cell,* **3**, 487-493.

Borer, P. N., Wang, S., Rogeenbuck, M. W., Gott, J. M., Uhlenbeck, O. C. & Pelczer, I. (1995). Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry,* **34**, 6488-6503.

Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science,* **229**, 23-28.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science,* **252**, 1167-1171.

Carey, J. & Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry,* **22**, 2610-2615.

Carey, J., Lowary, P. T. & Uhlenbeck, O. C. (1983). Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site. *Biochemistry,* **22**, 4723-4730.

Cavarelli, J., Rees, B., Ruff, M., Thierry, J.-C. & Moras, D. (1993). Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature,* **362**, 181-184.

Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K. & Das, A. (1995). Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. *Proc. Natl Acad. Sci. USA,* **92**, 4061-4065.

Chen, L. & Frankel, A. D. (1994). An RNA-binding peptide from bovine immunodeficiency virus tat protein recognizes an unusual RNA structure. *Biochemistry,* **33**, 2708-2715.

Chen, L. & Frankel, A. D. (1995). A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl Acad. Sci. USA,* **92**, 5077-5081.

Cilley, C. D. & Williamson, J. R. (1997). Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA,* **3**, 57-67.

Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein - RNA complex. *Science,* **284**, 1171-1174.

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5 S rRNA domain. *Cell,* **91**, 705-711.

Correll, C. C., Munishkin, A., Chan, Y. L., Ren, Z., Wool, I. G. & Steitz, T. A. (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl Acad. Sci. USA,* **95**, 13436-13441.

Crick, F. H. C. (1953). The packing of α-helices: simple coiled-coils. *Acta Crystallog.* **6**, 689-692.

Cusack, S., Yaremchuk, A. & Tukalo, M. (1996). The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with $tRNA^{Ser}$ and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.* **15**, 2834-2842.

De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3-RNA recognition element. *Science,* **279**, 384-388.

Draper, D. E. (1995). Protein-RNA recognition. *Annu. Rev. Biochem.* **64**, 593-620.

Dreyfuss, G., Matunis, M. J., Piña-Roma, S. & Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289-321.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucl. Acids Res.* **15**, 9109-9128.

Elliott, M. B., Gottlieb, P. A. & Gollnick, P. (1999). Probing the TRAP-RNA interaction with nucleoside analogs. *RNA,* in the press.

Fersht, A. R. (1987). The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **12**, 301-304.

Gluick, T. C., Gerstner, R. G. & Draper, D. E. (1997). Effects of Mg^{2+}, K^+, and H^+ on an equilibrium between alternative conformations of an RNA pseudoknot. *J. Mol. Biol.* **270**, 451-463.

Gutell, R. R., Schnare, M. N. & Gray, M. W. (1992). A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. *Nucl. Acids Res.* **20**, 2095-2109.

Hall, K. B. (1994). Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry,* **33**, 10076-10088.

Hall, K. B. & Stump, W. T. (1992). Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucl. Acids Res.* **20**, 4283-4290.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. & Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the sex- lethal protein. *Nature,* **398**, 579-585.

Hinck, A. P., Markus, M. A., Huang, S., Gresiek, S., Kustonovich, I., Draper, D. E. & Torchia, D. A. (1997). The RNA binding domain of ribosomal protein L11: three-dimension structure of the RNA-bound form of the protein and its interaction with 23 S rRNA. *J. Mol. Biol.* **274**, 101-113.

Hoagland, M. B. (1960). The relationship of nucleic acid and protein synthesis as revealed by studies in cell-free systems. In *The Nucleic Acids* (Chargaff, E. & Davidson, J. N., eds), pp. 349-408, Academic Press, New York.

Howe, P. W., Allain, F. H., Varani, G. & Neuhaus, D. (1998). Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. Biomol. NMR,* **11**, 59-84.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318-356.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991). Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447-3456.

Johansson, H. E., Dertinger, D., LeCuyer, K. A., Behlen, L. S., Greef, C. H. & Uhlenbeck, O. C. (1998). A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl Acad. Sci. USA,* **95**, 9244-9249.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* **365**, 520-527.

Kjems, J., Calnan, B. J., Frankel, A. D. & Sharp, P. A. (1992). Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **11**, 1119-1129.

Klemm, J., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell,* **77**, 21-32.

Lago, H., Fonseca, S. A., Murray, J. B., Stonehouse, N. J. & Stockley, P. G. (1998). Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. *Nucl. Acids Res.* **26**, 1337-1344.

Landsman, D. (1992). RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucl. Acids Res.* **20**, 2861-2864.

LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. (1996). Mutagenesis of a stacking contact in the MS2 coat protein-RNA complex. *EMBO J.* **15**, 6847-6853.

Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. (1998). NMR structure of the bacteriophage

lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell,* **93**, 289-299.

Markus, M., Hinck, A., Huang, S., Draper, D. E. & Torchia, D. A. (1997). High resolution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding RNA. *Nature Struct. Biol.* **4**, 70-77.

Mascotti, D. P. & Lohman, T. M. (1990). Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. *Proc. Natl Acad. Sci. USA,* **87**, 3142-3146.

Mascotti, D. P. & Lohman, T. M. (1997). Thermodynamics of oligoarginines binding to RNA and DNA. *Biochemistry,* **36**, 7272-7279.

Misra, V. K., Hecht, J. L., Sharp, K. A., Friedman, R. A. & Honig, B. (1994). Salt effects on protein-DNA interactions. The λcI repressor and *Eco*RI endonuclease. *J. Mol. Biol.* **238**, 264-280.

Murzin, A. G. (1993). OB (oligonucleotide oligosaccharide binding) fold-common structural and functional solution for nonhomologous sequences. *EMBO J.* **12**, 861-867.

Musier-Forsyth, K. & Schimmel, P. (1992). Functional contacts of a transfer RNA synthetase with 2′-hydroxyl groups in the RNA minor groove. *Nature,* **357**, 513-515.

Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science,* **253**, 784-786.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature,* **348**, 515-520.

Nomura, M. (1973). Assembly of bacterial ribosomes. *Science,* **179**, 864-873.

Olmsted, M. C., Anderson, C. F. & Record, M. T. (1989). Monte Carlo description of oligoelectrolyte properties of DNA oligomers: range of the end effect and the approach of molecular and thermodynamic properties to the polyelectrolyte limits. *Proc. Natl Acad. Sci. USA,* **86**, 7766-7770.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature,* **372**, 432-438.

Pan, J. & Woodson, S. A. (1998). Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **280**, 597-609.

Perret, V., Florentz, C., Puglisi, J. D. & Giege, R. (1992). Effect of conformational features on the aminoacylation of tRNAs and consequences on the permutation of tRNA specificities. *J. Mol. Biol.* **226**, 323-333.

Price, S. R., Evans, P. R. & Nagai, K. (1998). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature,* **394**, 645-50.

Ptashne, M. (1967). Specific binding of the λ phage repressor to λ DNA. *Nature,* **214**, 232-234.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science,* **270**, 1200-1203.

Record, M. T., Lohman, T. M. & Haseth, P. d. (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **107**, 145-158.

Rimmele, M. E. & Belasco, J. G. (1998). Target discrimination by RNA-binding proteins: role of the ancillary protein U2A′ and a critical leucine residue in differentiating the RNA-binding specificity of spliceosomal proteins U1A and U2B″. *RNA,* **4**, 1386-1396.

Saks, M. E. & Sampson, J. R. (1996). Variant minihelix RNAs reveal sequence-specific recognition of the helical tRNA(Ser) acceptor stem by *E. coli* seryl-tRNA synthetase. *EMBO J.* **15**, 2843-2849.

Scherly, D., Boelens, W., Venrooij v., W. J., Dathan, N. A., Hamm, J. & Mattaj, I. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163-4170.

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″′. *Nature,* **345**, 502-506.

Sharp, K. A., Honig, B. & Harvey, S. C. (1990). Electrical potential of transfer RNAs: codon-anticodon recognition. *Biochemistry,* **29**, 340-346.

Somers, W. S. & Phillips, S. E. (1992). Crystal structure of the met repressor-operator complex at 2. 8 Å resolution reveals DNA recognition by β-strands. *Nature,* **359**, 387-393.

Spahr, P. F., Farber, M. & Gesteland, R. F. (1969). Binding site on R17 RNA for coat protein. *Nature,* **222**, 455-459.

Su, L., Radek, J. T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L. A. & Weiss, M. A. (1997). RNA recognition by a bent α-helix regulates transcriptional antitermination in phage lambda. *Biochemistry,* **36**, 12722-12732.

Tan, R. & Frankel, A. D. (1994). Costabilization of peptide and RNA structure in an HIV rev peptide-RRE complex. *Biochemistry,* **33**, 14579-14585.

Tan, R. & Frankel, A. D. (1995). Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl Acad. Sci. USA,* **92**, 5282-5286.

Tan, R., Chen, L., Buettner, J. A., Hudson, D. & Frankel, A. D. (1993). RNA recognition by an isolated α helix. *Cell,* **73**, 1031-1040.

Tao, J. & Frankel, A. D. (1993). Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins. *Proc. Natl Acad. Sci. USA,* **90**, 1571-1575.

Traub, P. & Nomura, M. (1969). Structure and function of *Escherichia coli* ribosomes VI. Mechanism of assembly of 30 S ribosomes studied *in vitro*. *J. Mol. Biol.* **40**, 391-413.

Tsai, D. E., Harper, D. S. & Keene, J. D. (1991). U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucl. Acids Res.* **19**, 4931-4936.

Uhlenbeck, O. C., Gott, J. M. & Witherell, G. W. (1990). The specific interaction between RNA phage coat proteins and RNA. *Prog. Nucl. Acid Res.* **40**, 185-220.

Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature,* **371**, 623-626.

Valegård, K., Murray, J. B., Stonehouse, N. J., Worm, S. v. d. , Stockley, P. G. & Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* **270**, 724-738.

van den Worm, S. H., Stonehouse, N. J., Valegard, K., Murray, J. B., Walton, C., Fridborg, K., Stockley, P. G. & Liljas, L. (1998). Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucl. Acids Res.* **26**, 1345-1351.

van Gelder, C. W., Gunderson, S. I., Jansen, E. J., Boelens, W. C., Polycarpou-Schwarz, M., Mattaj, I. W. & van Venrooij, W. J. (1993). A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.* **12**, 5191-5200.

Van Gilst, M. R., Rees, W. A., Das, A. & von Hippel, P. H. (1997). Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry,* **36**, 1514-1524.

Williams, D. J. & Hall, K. B. (1996). RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J. Mol. Biol.* **257**, 265-275.

Wu, H.-N. & Uhlenbeck, O. C. (1987). Role of a bulged A residue in a specific RNA-protein interaction. *Biochemistry,* **26**, 8221-8227.

Xing, Y. & Draper, D. E. (1995). Stabilization of ribosomal RNA tertiary structure by ribosomal protein L11. *J. Mol. Biol.* **246**, 319-331.

Xing, Y., GuhaThakurta, D. & Draper, D. E. (1997). The RNA binding domain of ribosomal protein L11 is structurally similar to homeodomains. *Nature Struct. Biol.* **4**, 24-27.

Xirasagar, S., Elliott, M. B., Bartolini, W., Gollnick, P. & Gottlieb, P. A. (1998). RNA structure inhibits the TRAP (trp RNA-binding attenuation protein)-RNA interaction. *J. Biol. Chem.* **273**, 27146-27153.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* **2**, 827-840.

Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nature Struct. Biol.* **3**, 1026-1033.