Non-Watson–Crick base pairs in RNA–protein recognition

Thomas Hermann[1] and Eric Westhof[2]

The cellular functions of most RNA molecules involve protein binding, and non-Watson–Crick base pairs are hallmark sites for interactions with proteins. The determination of three-dimensional structures of RNA–peptide and RNA–protein complexes reveals the molecular basis of non-Watson–Crick base-pair recognition.

Addresses: [1]Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 557, New York, NY 10021, USA. [2]Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, F-67084 Strasbourg, France.

Correspondence: Thomas Hermann and Eric Westhof
E-mail: thermann@sbnmr1.ski.mskcc.org
westhof@ibmc.u-strasbg.fr

Chemistry & Biology December 1999, **6**:R335–R343

1074-5521/99/$ – see front matter

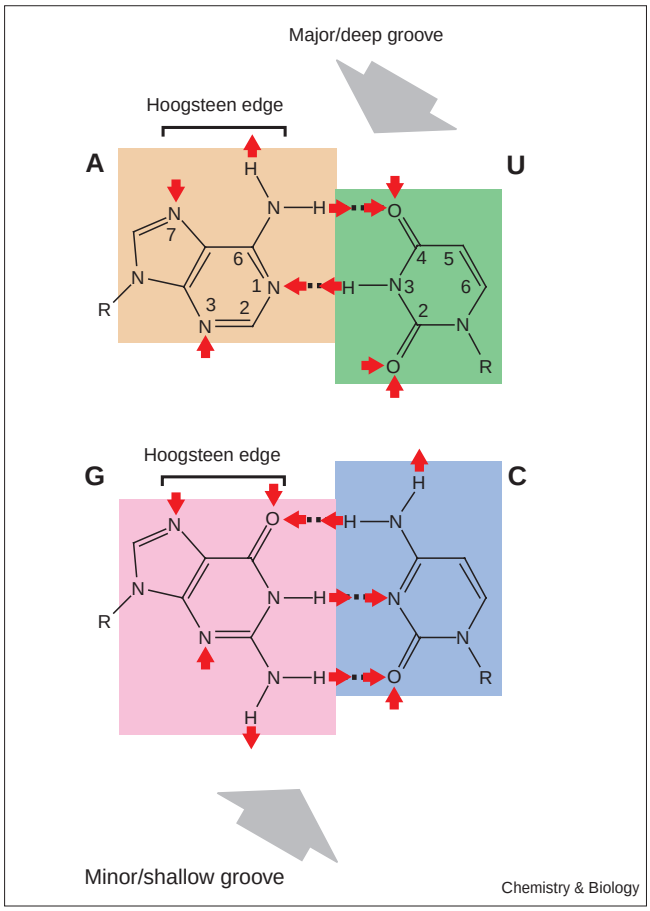
Introduction

RNA molecules are key players in all steps of gene expression. Beyond their function as messengers, RNAs are involved in regulatory processes and constitute essential parts of the cellular machinery responsible for mRNA splicing, transport and translation (reviewed in [1]). Several catalytic RNA molecules (ribozymes) have been discovered that function in the absence of proteins (reviewed in [2]). Some ribozymes are more catalytically efficient when associated with proteins, as has been found for the RNA component of eubacterial RNaseP (reviewed in [3]), whereas others are not catalytically active alone, but become active in presence of specific proteins (reviewed in [4]). Most cellular RNAs work in concert with protein partners, either in permanent complexes, such as the ribosome, or in transient associations, such as the mRNA splicing machinery. The assembly of functional RNA–protein complexes requires accurate recognition of the components. Here, we outline the important role of non-Watson–Crick base pairs in providing sites for the specific interaction of RNA folds with proteins. We will also discuss recent structures of protein–RNA complexes that do not rely on the involvement of base pairs for protein–RNA recognition. All base pairs observed in crystal and nuclear magnetic resonance (NMR) structures have been itemized and described recently [5].

The A-form RNA double helix is a poor target for specific interactions

RNA secondary structure is defined by contiguous canonical Watson–Crick pairs formed by hydrogen bonding between the complementary bases adenine–uracil (A–U) and guanine–cytosine (G=C; Figure 1). The definition of secondary structure in RNA includes the most common non-Watson–Crick pair: the wobble G•U pair. Stacking of canonical pairs gives rise to double-stranded helices; regular A-form helices are the basic building blocks of all RNA architectures known so far. In contrast with B-form DNA, which has a wide major groove and a narrow minor groove, A-form RNA has a narrow deep groove and a wide shallow groove (Figure 2). The discriminatory major-groove edges of the base pairs are buried in the inaccessible deep groove [6], whereas the shallow groove permits access to the rather uniform minor-groove side of canonical pairs. Moreover, the polar groups of the Watson–Crick face of the bases, potential sites for hydrogen bonding, are engaged in base-pair interactions. Regular A-form RNA helices therefore have little potential for specific recognition by proteins. Here, we emphasize the central role and importance of non-Watson–Crick pairs. In the DNA field, a mismatch (i.e. a base pair involving noncomplementary

Figure 1



Hydrogen bonding in canonical Watson–Crick base pairs A–U and G=C. Hydrogen-bond donor and acceptor sites are marked by red arrows. The Hoogsteen-pairing sites of the purines are indicated along with the orientation of the base-pair edges towards the major/deep and minor/shallow grooves in double-stranded helices. Hydrogen-bonding interactions also occur very frequently at the shallow-groove side of isolated bases or of base pairs. The color scheme (A, orange; U, green; G, pink; C, blue–purple) is used throughout in the following figures.

bases) is a potential locus for deficient or carcinogenic biological development. In the RNA world, however, non-Watson–Crick pairs are key determinants for proper native folding of the RNA and for RNA recognition by proteins or other ligands, such as ions or antibiotics. We will therefore avoid the term 'mismatch' when describing non-Watson–Crick base pairs in RNAs.

The dual role of non-Watson–Crick pairs

Pairwise combinations of hydrogen-bonded coplanar bases other than Watson–Crick pairings give rise to noncanonical or mismatch pairs (reviewed in [5,7,8]). The G•U wobble base pair and G•A pairs are the most common non-Watson–Crick pairs in large RNA molecules such as ribosomal RNA (rRNA) [9,10]. Non-Watson–Crick interactions between nucleotides are also found in triples, in which a third base forms hydrogen bonds with a canonical pair.

Some types of non-Watson–Crick base pairs are incorporated into stacked RNA stems without disrupting the helical structure, but they do distort the regular A conformation [11]. Some non-Watson–Crick pairs particularly affect the lateral dimension of the deep groove, however, without affecting its characteristic depth (Figure 2). Unpaired nucleotides adjacent to noncanonical pairs increase the flexibility of the RNA backbone, thereby facilitating the widening of the deep groove. Non-Watson–Crick pairs and triples can distort the RNA deep groove to an extent that allows protein domains with ordered and regular secondary-structure elements, such as β turns (Figure 2) and α helices (Figure 3), to be accommodated.

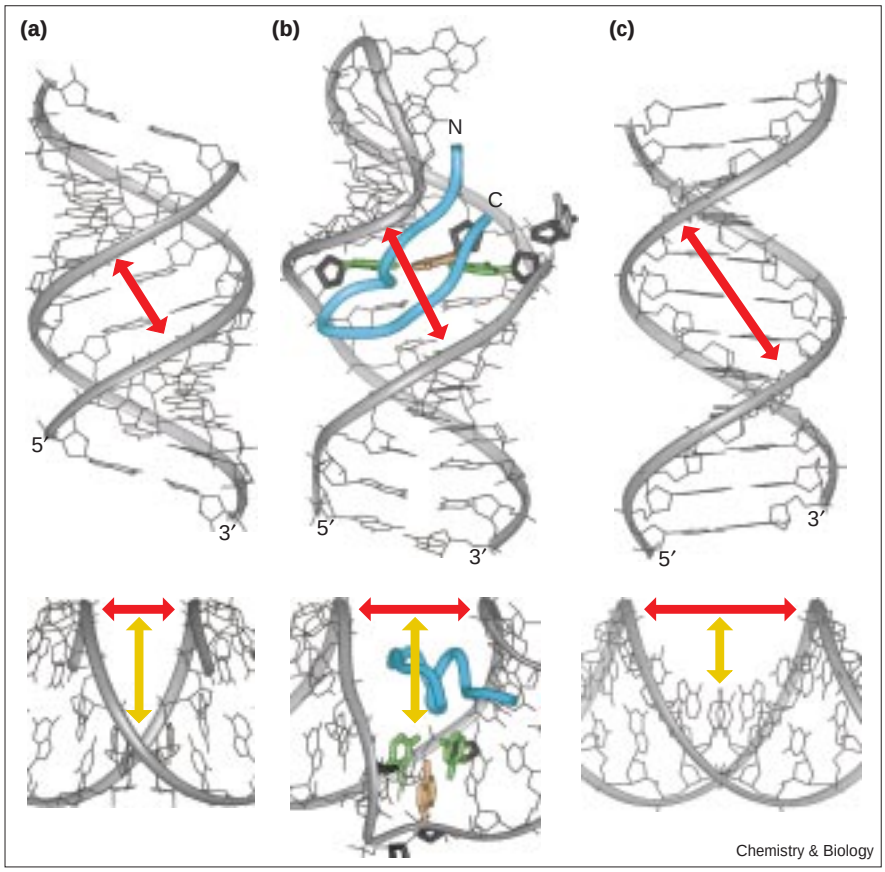
In addition to deforming the shape of the deep groove in RNA helices, non-Watson–Crick base pairs serve as specific recognition sites in hydrogen-bonding interactions with proteins. In noncanonical pairs, alternative sets of the polar donor and acceptor groups of the bases are available for intermolecular contacts because of the different hydrogen bonding patterns and arrangements of bases to those of canonical pairs.

Protein-binding sites formed by non-Watson–Crick pairs, triples and loops

The protein- and peptide-binding sites in three-dimensional structures of RNA complexes known so far suggest that single non-Watson–Crick base pairs are complemented by additional RNA motifs to form recognition surfaces for the substrates.

Tandem stacks of two noncanonical base pairs are found in the Rev peptide-binding sites of the HIV-1 Rev-response element (RRE) RNA [12] and a Rev-specific aptamer RNA (Figure 3) [13]. In the Rev complex of RRE and the aptamer, the peptide binds in an α-helical conformation to the RNA deep groove, which is widened by a *cis* Watson–Crick G•A pair (Figure 3d) followed by either a *trans* Watson–Crick G•G (RRE) or the isosteric A•A pair (aptamer) (Figure 3c,e). Adjacent bulged-out nucleotides facilitate the widening of the deep groove. In both complexes, the G•A pair is recognized by the same asparagine residue in the peptide forming intermolecular hydrogen bonds with the two purines simultaneously (Figure 3d). The symmetric G•G and A•A pairs in RRE and the aptamer, respectively, are not involved in direct contacts to the peptide. The opening of the peptide-binding pocket in the deep groove of RRE, however, strictly requires a homo purine pair isosteric to G•G. RRE variants obtained using *in vitro* selection showed high affinity for the Rev protein only when a G•G or A•A pair could be formed between the base positions 48 and 71 [14].

Figure 2



Dimensions of the major/deep groove in nucleic acid duplexes shown from the side (top) and looking into the groove (bottom). (a) In regular A-form RNA helices, the major groove is deep and narrow. (b) Non-Watson–Crick base pairs, triples and adjacent loops distort the A-form geometry of RNA helices, leading to a expanded deep groove without reducing its characteristic depth. In the complex between BIV Tat peptide and TAR RNA [19,20], the peptide (cyan) binds in a β-turn conformation to the RNA deep groove widened by an U•A–U triple, in which adenine (orange) participates in noncanonical pairing with one of the uracils (green). Adjacent to the triple, an unpaired nucleotide (grey) facilitates the widening of the deep groove. (c) In B-DNA, the major groove is much wider but less deep than in double-stranded RNA.

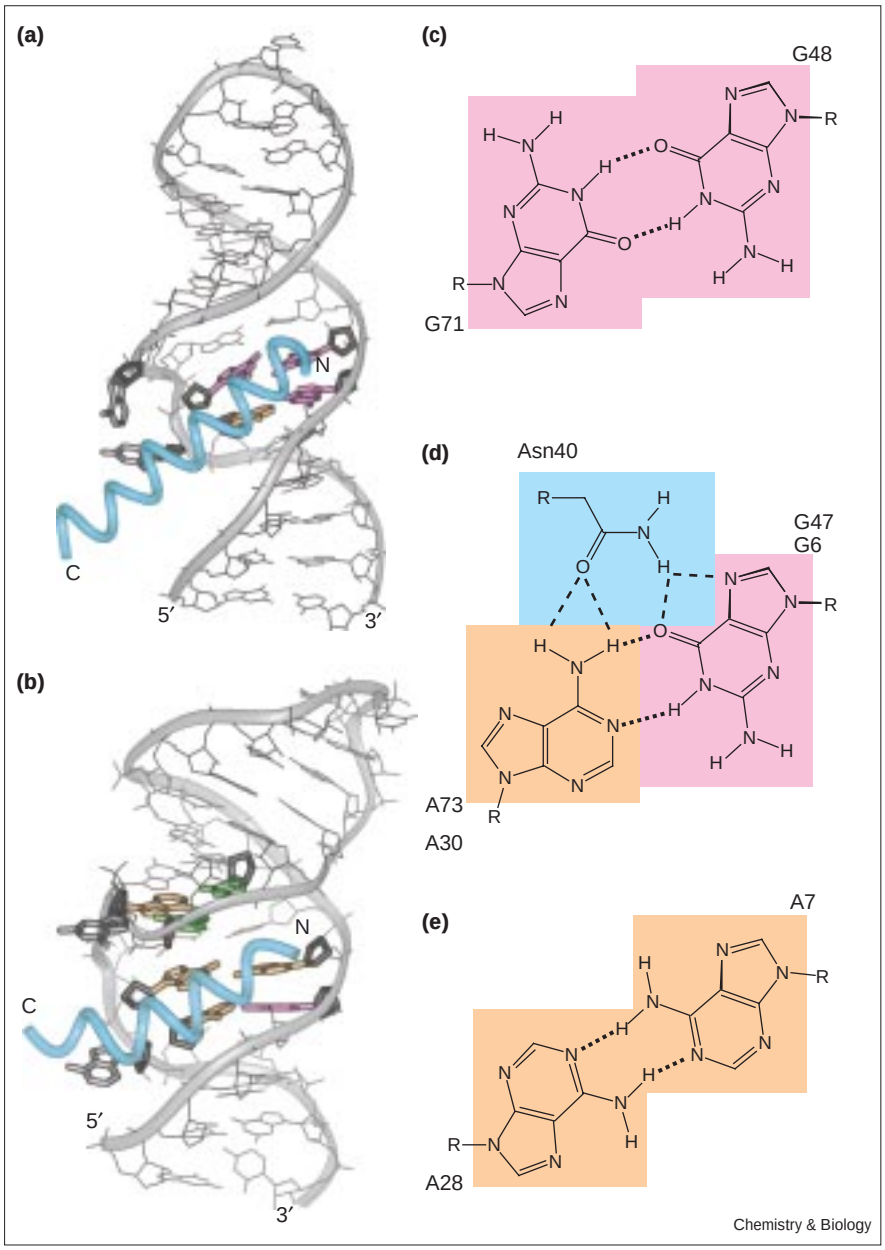
Interestingly, binding of Rev protein to RRE is inhibited by aminoglycoside antibiotics such as neomycin B, which specifically recognizes the G•G pair in RRE [15]. Footprinting experiments on RRE–neomycin complexes have revealed that the drug interacts with G47 and G48 in the G•A and G•G pairs [16]. Structural data available for other RNA–aminoglycoside complexes (reviewed in [17]) suggest that the aminoglycosides may form polar hydrogen bonds with their protonated amino groups to N7 and O6 accessible at the Hoogsteen edge of the guanines.

Similarly, a *cis* Watson–Crick G•A pair has been identified as a specific recognition site for a designed zinc finger protein motif [18]. It has been suggested that the G•A pair, embedded in a regular RNA duplex, is involved in two hydrogen bonds to a lysine sidechain that contacts the Hoogsteen edge of the guanine [18].

In the Rev-specific aptamer RNA, a U•A–U triple contributes to the peptide-binding site [13]. The arrangement of bases in the triple corresponds to a classical U•A–U trimer in which both Watson–Crick and Hoogsteen base-pairing sites of adenine are engaged simultaneously [7]. Stereochemically identical U•A–U triples open up the deep groove for substrate binding in BIV TAR RNA [19,20] in complex with a Tat peptide (Figure 2b) and a class II Rev-aptamer RNA [21] bound to a Rev peptide (Figure 4a). In the TAR complex, an isoleucine sidechain of the Tat peptide in the deep groove packs against the hydrophobic C5–C6 edge of the uracil base [19,20] that binds to the Hoogsteen face of adenine in the U•A–U triple.

The BIV TAR RNA, the class II aptamer and the boxB RNA [22,23] (Figure 4b) provide examples of peptide-binding sites formed by combinations of base mismatches with adjacent loops. In the complexes of these RNAs and Tat, Rev and N peptide, respectively, the peptide-binding pocket opens up towards a loop that folds away from the groove to allow the substrate to enter. Sheared G•A pairs terminate the loops and participate in hydrogen bonding to arginine residues of the peptide in the class II Rev aptamer and the boxB RNA complexes (Figure 4). The arrangement of bases in sheared G•A pairs projects the Hoogsteen edge of the guanine towards the deep groove, making it readily available for contacts with amino acid sidechains (Figure 4c,d). Among mismatch pairs involved in protein recognition, sheared G•A pairs stand out, given

Figure 3



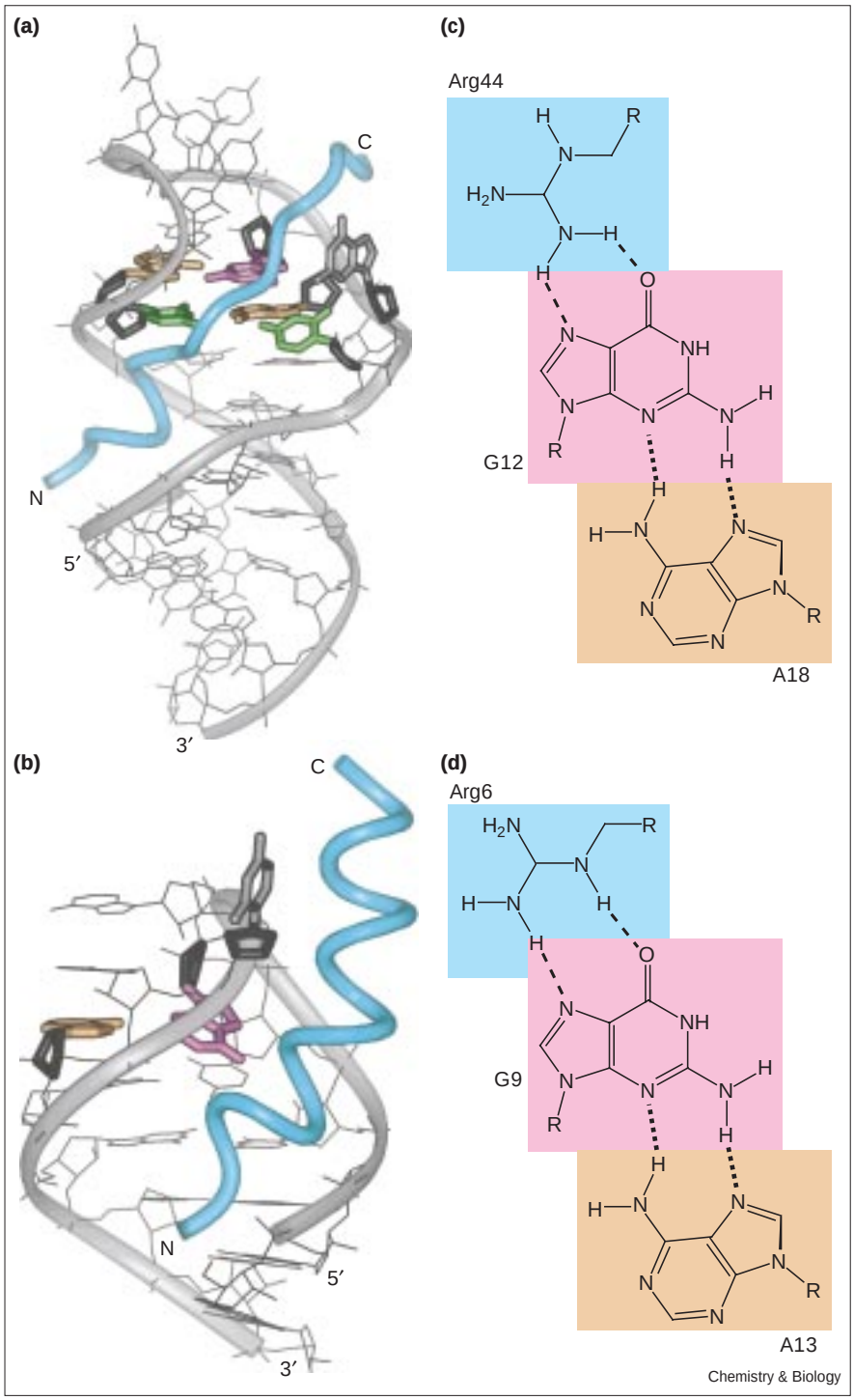
Peptide-binding sites involving stacks of tandem noncanonical base pairs in RNA duplexes. (a) In the complex between an HIV-1 Rev peptide and Rev-response element (RRE) RNA [12], the peptide binds as an α helix into the deep groove widened by a G•G pair (c) following on a *cis* Watson–Crick G•A pair (d). (b) The same peptide binds in the deep groove of an RNA aptamer [13] that contains an identical G•A pair and a symmetric A•A pair (e) isosteric with the corresponding G•G pair in the RRE RNA. A U•A–U triple identical to the triple in the BIV TAR RNA (see Figure 2b) participates in the peptide-binding site. (d) In both complexes, the sidechain of the same asparagine residue forms specific hydrogen bonds to groups at the deep-groove edge of the G•A pair. G47•A73 and G6•A30 correspond to the RRE RNA and the aptamer, respectively. Unpaired nucleotides in (a) and (b) are shown in grey.

their wide distribution in natural RNAs (reviewed in [24]). In addition to their ability to distort RNA duplexes and provide hydrogen-bond partners, sheared G•A pairs have a characteristic in-plane breathing motion that may facilitate the interaction with protein ligands [25].

A combination of sheared G•A pairs with other noncanonical purine–purine base pairs participating in protein recognition sites is found in the loop E motif of 5S rRNA and the sarcin/ricin loop of 23S rRNA. NMR and X-ray structure analyses of the sarcin/ricin loop [26] and a 5S rRNA domain [27,28] have revealed the extensive base pairing in loop E, which comprises seven consecutively stacked noncanonical base pairs, including sheared G•A, *trans* Hoogsteen A•U and bifurcated G•G pairs. The non-Watson–Crick pairs in loop E are stabilized by bridging water molecules between bases and Mg^{2+} bound to the deep groove. Both the shallow and deep grooves are widened in the loop E motif, allowing protein domains access to the unique binding surface created by the hydrogen bond donors and acceptors of the base pair edges. Protein sidechains could specifically bind to the non-Watson–Crick pairs, displacing metal ions and water molecules by positively charged groups (see below). A three-dimensional structure for a loop E motif in

Figure 4

Peptide-binding sites involving sheared G•A pairs adjacent to a loop. In the complexes of (**a**) the HIV Rev peptide bound to a class II RNA aptamer [21] and (**b**) the N peptide bound to boxB RNA [23], an arginine residue forms hydrogen bonds to the Hoogsteen edge of a guanine in a sheared G•A pair. The arginine binds via a single amino group (**c**) in the aptamer complex and, in addition, via the secondary amino group (**d**) in the boxB complex. Bulged-out nucleotides are shown in grey stick representation.



complex with a protein is not available, but this RNA motif has been identified in a number of large RNAs [29] that are known to interact with proteins. Loop E of 5S rRNA is part of the binding site for ribosomal protein L25 in *Eubacteria* [30], L5 and transcription factor IIIA in *Eucarya* [31,32].

GNRA tetraloops as protein-recognition sites

The pentaloop in the boxB RNA adopts a GNRA tetraloop-like conformation (N is any nucleotide; R is a purine) by extrusion of one nucleotide [22,23] (Figure 4b) induced by binding of the N peptide [33]. GNRA tetraloops are very

frequent in large RNA folds because of both their conformational stability and their ability to participate in tertiary contacts with other RNA motifs [34–36]. A characteristic structural feature of GNRA loops is the terminating sheared G•A pair that involves the first and last residues of the tetraloop [37,38]. The sheared G•A contributes specific intermolecular hydrogen bonds to the N peptide in the boxB RNA complex [22,23] and, therefore, protein recognition of GNRA motifs is likely to be a general theme of RNA–protein interactions. Another example is the ribosomal sarcin/ricin loop capped by a GAGA tetraloop [26], which is an identity element for the recognition by the ribotoxic protein ricin [39].

Specific binding of proteins to GNRA motifs is expected to play a major role in large RNA–protein assemblies such as the ribosome [40,41]. Recent progress in the determination of three-dimensional structures of complete ribosomal subunits [42–44] will, therefore, greatly expand the repertoire of known GNRA loop–protein interactions.

Recognition of non-Watson–Crick pairs by displacing bridging water molecules or ions

In the crystal structure of the loop E motif of 5S rRNA, three non-Watson–Crick pairs at the center of a stack of seven noncanonical pairs are stabilized by water molecules and magnesium ions bridging the bases [27]. Proteins that bind to the loop E motif in different RNA folds (see above) might recognize the noncanonical pairs by replacing the bridging water molecules or ions with polar or charged amino acid sidechains.

The determination of the crystal structure of the spliceosomal U2B″–U2A′ protein complex bound to a fragment of U2 small nuclear RNA (snRNA) has revealed a similar situation [45]. The U2B″ protein binds to a 11-nucleotide loop that is closed by a noncanonical U•U pair joining the loop to a regular duplex (Figure 5a). The terminal amino group of a lysine sidechain is positioned at the deep-groove edge of the U•U pair in a proper orientation that allows hydrogen bonds to form with the O4 carbonyl atoms in both uracils (Figure 5c). At the position of the lysine amino group, a water molecule bridging the uracil bases is found in stereochemically identical U•U pairs in the three-dimensional structure of an RNA duplex containing an internal loop [46]. The interaction between the lysine sidechain of the U2B″ protein and the U•U pair directs the orientation of the adjacent RNA stem, providing a subtle mechanism for the discriminatory recognition of other spliceosomal proteins [45].

Recognition of noncanonical pairs in the shallow groove

The structural uniformity of the shallow-groove edges of Watson–Crick base pairs renders them poor targets for specific recognition. Noncanonical base pairs introduce asymmetries in the shallow groove of RNA duplexes that allow subtle structural discrimination in ligand binding.

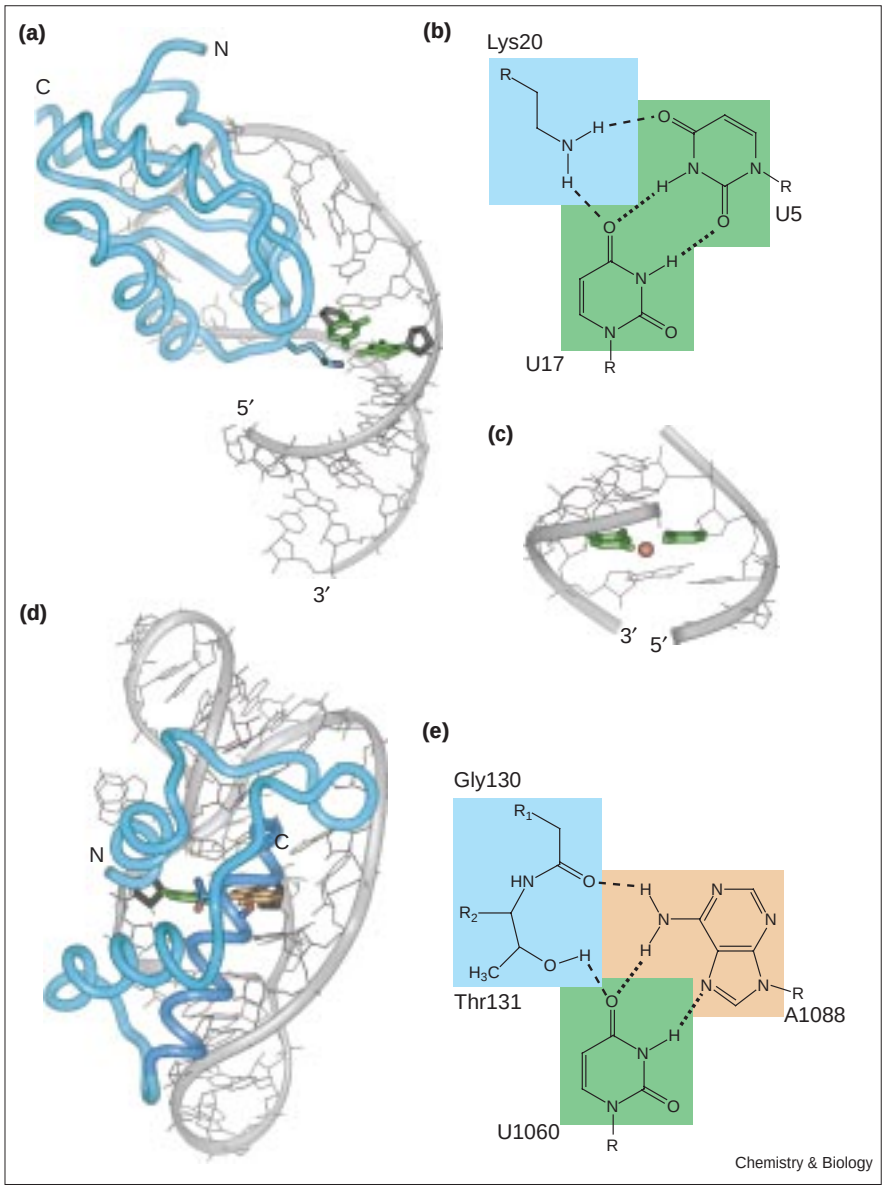
Specific recognition of a base pair in the shallow groove has been observed for G•U wobble base pairs. The alanine tRNA contains a single G•U pair in the acceptor stem that is a major determinant in specific aminoacylation by $tRNA^{Ala}$ synthetase [47]. Variant tRNAs in which the G•U pair is mutated or guanine is replaced by inosine, which lacks the 2-amino group of guanine, are not aminoacylated by the synthetase [48]. These findings indicate that $tRNA^{Ala}$ synthetase recognizes guanine in the shallow groove by its exocyclic 2-amino group, which is not involved in base pairing in G•U wobble pairs. Using an RNA microhelix derived from the acceptor stem of $tRNA^{Ala}$ as a binding substrate in phage-display selection, a 28-amino acid peptide has been obtained that binds to the shallow groove exclusively of RNA helices containing a G•U wobble pair [49].

The shallow-groove recognition of G•U wobble pairs is facilitated by the geometry of the base pair, in which the uracil is pushed into the deep groove, creating a depression on the shallow-groove surface. This site is, in many cases, occupied by a water molecule bridging the two bases of the wobble pair [50]. The specific binding of ligands to the shallow-groove edge of G•U pairs is therefore likely to involve displacement of a bridging water molecule (see above). The docking of a carbonyl oxygen atom into the shallow-groove depression and hydrogen bonding to the exocyclic 2-amino group of guanine have been suggested as the basis for specific recognition of G•U wobble pairs within helices by isoalloxazines [51].

tRNA–synthetase interactions provide yet another example of recognition in the shallow groove of a non-Watson–Crick pair. A sequence comparison analysis has shown that the first (32) and last (38) residues of the seven-membered tRNA anticodon co-vary so as to maintain characteristic bifurcated hydrogen-bonded pairs [52]. In the complex of $tRNA^{Gln}$ and its cognate synthetase, a contact between an asparagine sidechain and a uracil within a single-hydrogen-bonded U32•U38 pair has been discovered [53]. A hydrogen bond is formed between the amide group of the asparagine sidechain and the O2 carbonyl atom of the uracil, projecting into the shallow groove. Despite the recurrence of non-Watson–Crick pairs in tRNAs, examples of recognition of noncanonical pairs by tRNA-binding proteins are scarce. Clearly, most of them are necessary for maintaining the native architecture of tRNAs. Furthermore, one may speculate that the numerous modifications of nucleotides may be preferred as specific recognition elements of tRNA structures.

The recently determined three-dimensional structure of a 58-nucleotide RNA fragment of 23S rRNA in complex

Figure 5



Recognition of non-Watson–Crick pairs in larger RNA–protein complexes. (a) In the complex of the spliceosomal U2B″–U2A′ proteins bound to an snRNA hairpin fragment [45], a U•U base pair closing the RNA stem is recognized from the deep-groove side by a lysine sidechain of the U2B″ protein. (b) The lysine residue forms hydrogen bonds to O4 carbonyl atoms of both uracils in the U•U pair. (c) In place of the lysine sidechain, a water molecule (red sphere) has been found bridging the carbonyl groups in stereochemically identical U•U mismatches within an RNA duplex [46]. (d) In the complex between the ribosomal L11 protein and a fragment of the 23S rRNA [54], a noncanonical *cis* Hoogsteen A•U pair (e) is involved in protein contacts. The A•U mismatch forms hydrogen bonds to a threonine sidechain and the peptide backbone within an α helix (blue) that binds to a shallow groove face of the RNA fold.

with ribosomal L11 protein [54,55] shows the most extensive case of shallow-groove recognition in an RNA–protein complex yet. The L11 protein binds with a 15-residue α helix to a shallow-groove surface of the RNA fold (Figure 5b). Two consecutive amino acids within the α helix, namely Gly130 and Thr131, are involved in hydrogen bonds to a noncanonical A•U pair in the RNA (Figure 5e). The *trans*-Hoogsteen A•U pair is formed by a long-range tertiary interaction between an adenosine and a uridine, which ties together the RNA fold.

Conclusions

Canonical Watson–Crick pairs in RNA can be considered as the most basic unit for building three-dimensional frameworks. Together, these units form rather regular helices, interrupted at defined positions by unique interaction or recognition motifs that promote RNA–RNA or RNA–protein contacts. Because of their protean diversity, the simplest motifs generating irregularities and asymmetries suitable for specific interactions are noncanonical base pairs. The non-Watson–Crick pairs can occur in single occurrences within a helical stem or as stacks of tandem or more base pairs forming intricate and recurrent motifs, like that of the loop E motif [26–29]. The three-dimensional structures of RNA–peptide and RNA–protein complexes reveal non-Watson–Crick base pairs as key elements of RNA recognition. At the interface between the worlds of RNA and protein molecules, the exceptions from the

Watson–Crick geometry do indeed rule the geometry of base pairs involved in specific protein binding.

Two recent crystal structures of RNA–protein complexes are in stark contrast to the those described above: in both, the *Sex-lethal* protein complexed with the tra mRNA precursor [56] and the trp RNA-binding attenuation protein (TRAP) bound to its recognition RNA [57], the protein recognizes specifically a single-stranded RNA. The *Sex-lethal* protein binds to a characteristic U-rich polypyrimidine tract, and forces female-specific alternative splicing. The polypyrimidine tract has no base pairs and has an extended, but structured, conformation, with several intra-RNA hydrogen bonds between the hydroxyl groups and phosphate anionic oxygen atoms. Interestingly, all but one ribose ring adopts the nonhelical C2′-*endo* pucker, which is rare in structured RNAs. The RNA winds around a fairly long cleft within the protein, promoting numerous contacts between protein sidechains and the pyrimidine bases or the sugar-phosphate backbone. Mutations within the polypyrimidine tract would induce secondary-structure formation, preventing or weakening specific binding to the protein. In the striking structure of TRAP bound to RNA [57], the protein, an 11-mer activated by the co-factor L-tryptophan, folds into a circular belt of 80 Å diameter to which eleven GAG triplets are bound by specific interactions with the bases (e.g. a glutamate with the N1 and N2 nitrogens of a G or stacking of a lysine or a phenylalanine with a G). Although single-stranded, the GAG triplets adopt a helical-like conformation with the sugar-phosphate backbone, presenting hydration sites similar to those seen in free RNAs [50]. The nucleotide backbone is recognized solely via one H-bond involving a 2′-hydroxyl group and a NH mainchain group. The two complexes [56,57] illustrate the diversity of RNA recognition. As discussed above, the recognition of RNAs with defined secondary structure elements relies on the variety of non-Watson–Crick pairs. In contrast, in the *Sex-lethal* complex, the protein recognizes a nonhelical and convoluted RNA single-strand via base and backbone contacts, whereas, in the TRAP complex, the protein recognizes a helical-like complex using essentially the base polar groups. RNAs with positive selective pressure for avoiding secondary structure therefore use different recognition principles to govern their interactions with proteins than those used by RNAs under positive selective pressure for maintaining secondary structure.

Acknowledgements

We thank A. Gorin and D.J. Patel (Memorial Sloan-Kettering Cancer Center, New York) for communicating atomic coordinates of the class II RNA aptamer–Rev complex.

References

1. Gesteland, R.F., Cech, T.R. & Atkins, J.F. (1999). *The RNA World*, 2nd ed., Cold Spring Harbor Laboratory Press, New York.
2. Carola, C. & Eckstein, F. (1999). Nucleic acid enzymes. *Curr. Opin. Struct. Biol.* **3**, 274-283.
3. Frank, D.N. & Pace, N.R. (1998). Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153-180.
4. Lambowitz, A.M., Caprara, M.G., Zimmerly, S. & Perlman, P.S. (1999). Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In *The RNA World*, 2nd ed. (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds) pp.451-485, Cold Spring Harbor Laboratory Press, New York.
5. Leontis, N.B. & Westhof, E. (1999). Conserved geometrical base pairing patterns in RNA. *Quart. Rev. Biophys.* **32**, in press.
6. Weeks, K.M. & Crothers, D.M. (1993). Major groove accessibility of RNA. *Science* **261**, 1574-1577.
7. Saenger, W. (1984). *Principles of Nucleic Acid Structure.* Springer, New York.
8. Burkard, M.E., Turner, D.H. & Tinoco Jr., I. (1999). Structures of base pairs involving at least two hydrogen bonds. In *The RNA World*, 2nd ed. (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds) pp.675-680, Cold Spring Harbor Laboratory Press, New York.
9. Gautheret, D., Konings, D. & Gutell, R.R. (1994). A major family of motifs involving GoA mismatches in ribosomal RNA. *J. Mol. Biol.* **242**, 1-8.
10. Gautheret, D., Konings, D. & Gutell, R.R. (1995). G•U base pairing motifs in ribosomal RNA. *RNA* **1**, 807-814.
11. Holbrook, S.R., Cheong, C., Tinoco, I., Jr. & Kim, S.H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs. *Nature* **353**, 579-581.
12. Battiste, J.L., *et al.,* & Williamson, J.R. (1996). α-Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547-1551.
13. Ye, X., Gorin, A., Ellington, A.D. & Patel, D.J. (1996). Deep penetration of an a-helix into a widened RNA major groove in the HIV-1 Rev peptide-RNA aptamer complex. *Nat. Struct. Biol.* **3**, 1026-1033.
14. Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. & Ellington, A.D. (1993). Selection and design of high-affinity RNA ligands for HIV-1. *Rev. Gene* **137**, 19-24.
15. Werstuck, G., Zapp, M.L. & Green, M.R. (1996). A non-canonical base pair within the human immunodeficiency virus Rev-responsive element is involved in both Rev and small molecule recognition. *Chem. Biol.* **3**, 129-137.
16. Zapp, M.L., Stern, S. & Green, M.R. (1993). Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **74**, 969-978.
17. Hermann, T. & Westhof, E. (1998). Saccharide-RNA recognition. *Biopolymers (Nucleic Acid Sciences)* **48**, 155-165.
18. Blancafort, P., Steinberg, S.V., Paquin, B., Klinck, R. Scott, J.K. & Cedergren, R. (1999). The recognition of a noncanonical RNA base pair by a zinc finger protein. *Chem. Biol.* **6**, 585-597.
19. 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.
20. 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.
21. Ye, X., *et al.*, & Patel, D.J. (1999). RNA architecture dictates bound peptide conformations. *Chem. Biol.* **6**, 657-669.
22. Legault, P., Li, J., Mogridge, J., Kay, L.E. & Greenblat, J. (1998). NMR structure of the bacteriophage l N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289-299.
23. Cai, Z., *et al.,* & Patel, D.J. (1998). Solution structure of P22 transcriptional antitermination N peptide-box B RNA complex. *Nat. Struct. Biol.* **5**, 203-212.
24. Limmer, S. (1997). Mismatch base pairs in RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **57**, 1-39.
25. Hermann, T., Auffinger, A. & Westhof, E. (1998). Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* **27**, 153-165.
26. Szewczak, A.A., Moore, P.B., Chang, Y.L. & Wool, I.G. (1993). The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl Acad. Sci. USA* **90**, 9581-9585.
27. Correll, C.C., Freeborn, B., Moore, P.B. & Steitz, T.A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705-712.
28. Dallas, A. & Moore, P.B. (1997). The loop E-loop D region of *Escherichia coli* 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure* **5**, 1639-1653.
29. Leontis, N.B. & Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* **283**, 571-583.

30. Huber, P.W. & Wool, I.G. (1984). Nuclease protection analysis of ribonucleoprotein complexes: use of the cytotoxic a-sarcin to determine the binding sites of Escherichia coli ribosomal proteins L5, L18, and L25 on 5S rRNA. *Proc. Natl Acad Sci. USA* 81, 322-326.
31. Allison, L.A., Romaniuk, P.J. & Bakken, A.H. (1991). RNA-protein interactions of stored 5S RNA with TFIIIA and ribosomal protein L5 during Xenopus oogenesis. *Dev. Biol.* 144, 129-144.
32. Romaniuk, P.J. (1989). The role of highly conserved single-stranded nucleotides of Xenopus 5S RNA in the binding of transcription factor IIIA. *Biochemistry* 28, 1388-1395.
33. Su, L., *et al.,* & Weiss, M.A. (1997). An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. *Genes Dev.* 11, 2214-2226.
34. Woese, C.R., Winker, S. & Gutell, R.R. (1990). Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Natl Acad. Sci. USA* 87, 8467-8471.
35. Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I introns based on comparative sequence analysis. *J. Mol. Biol.* 216, 585-610.
36. Michel, F. & Westhof, E. (1996). Visualizing the logic behind RNA self-assembly. *Science* 273, 1676-1677.
37. Westhof, E., Romby, P., Romaniuk, P.J., Ehresmann, C. & Ehresmann, B. (1989). Computer modeling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S rRNAs. *J. Mol. Biol.* 207, 417-431.
38. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 253, 191-194.
39. Glück, A., Endo, Y. & Wool, I.G. (1992). Ribosomal RNA identity elements for ricin A-chain recognition and catalysis. *J. Mol. Biol.* 226, 411-424.
40. Brimacombe, R. (1995). The structure of ribosomal RNA: a three-dimensional jigsaw puzzle. *Eur. J. Biochem.* 230, 365-383.
41. Moore, P.B. (1998). The three-dimensional structure of the ribosome and its components. *Annu. Rev. Biophys. Biomol. Struct.* 27, 35-58.
42. Clemons, W.M., May, J.L.C., Wimberly, B.T., Mccutcheon, J.P., Capel, M.S. & Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* 400, 833-840.
43. Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. & Steitz, T.A. (1999). Placement of protein and RNA structure into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* 400, 841-847.
44. Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N. & Noller, H.F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science* 285, 2095-2104
45. 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, 645-650.
46. Baeyens, K.J., DeBondt, H.L. & Holbrook, S.R. (1995). Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. *Nat. Struct. Biol.* 2, 56-62.
47. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J. Green, R. & Schimmel, P. (1991). Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic group in the minor groove. *Science* 253, 784-786.
48. Beuning, P.J., Yang, F., Schimmel, P. & Musier-Forsyth, K. (1997). Specific atomic groups and RNA helix geometry in acceptor stem recognition by a tRNA synthetase. *Proc. Natl Acad. Sci. USA* 94, 10150-10154.
49. Frugier, M. & Schimmel, P. (1997). Subtle atomic group discrimination in the RNA minor groove. *Proc. Natl Acad. Sci. USA* 94, 11291-11294.
50. Westhof, E. (1988). Water: an integral part of nucleic acid structure. *Annu. Rev. Biophys. Biophys. Chem.* 17, 125-144.
51. Burgstaller, P., Hermann, T., Huber, C., Westhof, E. & Famulok, M. (1997). Isoalloxazine derivatives promote photocleavage of natural RNAs at GoU base pairs embedded within helices. *Nucleic Acids Res.* 25, 4018-44027.
52. Auffinger, P. & Westhof, E. (1999) Singly and bifurcated hydrogen-bonded base pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.* 292, 467-483.
53. Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S. & Steitz, T.A. (1998). How glutaminyl-tRNA synthetase selects glutamine. *Structure* 6, 439-449.
54. 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.
55. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* 97, 491-502.
56. 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.
57. Antson, A.A., Dodson, E.J., Dodson, G., Greaves, R.B., Chen, X. & Gollnick. P. (1999). Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* 401, 235-242.