Fitting peptides into the RNA world

Alan D Frankel

The structures of several peptide–RNA complexes have been reported in the past year, underscoring the diverse nature of RNA structure and protein interactions. In general, specific peptide conformations are stabilized by the surrounding RNA framework; this is strikingly similar to how peptides are stabilized upon interaction with proteins.

Addresses
Department of Biochemistry and Biophysics, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0448, USA; e-mail: frankel@cgl.ucsf.edu

Current Opinion in Structural Biology 2000, **10**:332–340

0959-440X/00/$ – see front matter

Abbreviations

BIV	bovine immunodeficiency virus
CD	circular dichroism
HTLV-I	human T-cell leukemia virus type I
PDB	Protein Data Bank
RRE	Rev response element
RxRE	Rex response element
SH	Src homology
TAR	*trans*-acting response element

Introduction

The wide structural diversity of protein–RNA interactions becomes more and more apparent each time the structure of a new complex is solved. This diversity should not be surprising given that proteins and RNAs individually fold into a wide range of architectures and, thereby, can generate many different types of recognition surfaces. This review focuses on a small subset of protein–RNA complexes in which the RNA-binding domain, or a portion of the domain, can be localized to a short, contiguous polypeptide region that recognizes its specific RNA site in the absence of an extensive protein scaffold. Even within this limited context, the high level of structural diversity is obvious. The review begins by more precisely defining what is meant by an RNA-binding peptide and draws analogies to protein–protein and peptide–protein interactions. The remainder of the review describes several new structures or models of peptide–RNA complexes published during the past year and discusses some of the features that help determine their high binding affinities and specificities.

RNA-binding proteins, domains and peptides

As for any macromolecular interaction, the specificity and strength of a protein–RNA interaction is determined by the positioning of chemical groups at the interface. A typical interface shows a combination of shape complementarity, hydrophobic, ionic and hydrogen bonding interactions, and sometimes water- or metal-mediated interactions. It is often observed that the protein, the RNA or both undergo substantial conformational changes upon binding or adopt stable folds only in the context of the complex. These characteristics are observed to varying degrees in all types of macromolecular complexes, but the dominant types of interactions from an energetic standpoint can differ substantially depending on the types of partners. For example, protein–protein interactions often are driven by hydrophobic contacts [1], whereas protein–nucleic acid interactions often contain a large electrostatic component [2,3]. One particularly dominant feature of peptide–RNA interactions is the stabilization of the peptide structure when bound in 'RNA binding pockets', as also observed when peptides are bound in protein binding pockets.

To define what is meant by an RNA-binding 'peptide', it is useful to examine the known structures of protein–RNA complexes. Each structure may be classified arbitrarily as a complex containing multiple proteins, a single protein, a protein domain, a peptide or an amino acid ligand (Table 1). According to this classification, an RNA-binding domain is defined as an independent folding unit, whereas an RNA-binding peptide is defined as a contiguous peptide chain of ≤30 amino acids that is unfolded or only weakly structured in the absence of RNA and yet binds RNA in a sequence-specific manner. Such definitions are, at best, operational as there exists, in reality, a continuum in which contiguous portions of proteins undergo disorder-to-order transitions upon RNA binding [4] and in which RNA-binding peptides can be stabilized prior to binding (see below). Furthermore, the detailed protein–RNA contacts at the various interfaces share many common features, irrespective of their classification in Table 1. Nonetheless, it can be instructive to ask whether some features of RNA recognition are largely confined to RNA-binding peptides and whether such features make sense from a biological perspective, just as it can be instructive to compare peptide–protein with protein–protein interactions.

There is a substantial amount known about how proteins specifically recognize short peptides. In some cases, the peptides are the biologically relevant ligands (as in complexes with MHC molecules, peptide transporters or peptide-binding proteins) and, in other cases, peptides have been found to mimic partially unfolded regions of intact proteins that mediate specific protein–protein interactions [5]. When bound to proteins, peptides most often adopt extended conformations and, occasionally, β-turn or α-helical conformations [5]. Proline-rich peptides recently have been shown to adopt left-handed helical conformations when complexed to EVH1 domains [6]. In some cases, the same peptide can adopt different conformations when bound to different proteins or placed in different protein contexts. For example, one peptide adopts an extended conformation when bound to the DnaK chaperone and an α helix when bound to GroEL [7], another

Table 1

Structures of RNA complexes*.

Complex	PDB entry code†	References
Multiprotein		
30S ribosomal subunit	1QD7, 1C59	[52,53]
50S ribosomal subunit	1C04	[54]
70S ribosomal complex	486D	[55]
U2B″–U2A′–U2 hairpin	1A9N	[56]
S15,S6,S18–rRNA site	1EKC	[57]
Protein		
tRNA synthetase–tRNA complexes		
Glutamine	1QTQ	[58]
Aspartic acid	1ASY	[59]
Serine	1SER	[60]
Lysine	–	[61]
Phenylalanine	–	[62]
Proline	–	[63]
Threonine	1QF6	[64]
Isoleucine	1QU2	[65]
EF-Tu–tRNA^{Phe}	1TTT	[66]
EF-Tu–tRNA^{Cys}	1B23	[67]
Met transformylase–fMet tRNA	2FMT	[68]
L11–23S rRNA site	1QA6	[69,70]
L25–5S rRNA site	1D6K, 1DFU	[71,72]
S15–rRNA site	1DK1	[73]
L30–autoregulatory RNA	1CK5	[74]
MS2 coat protein–operator	1ZDI, 1ZDK	[75,76]
MS2 coat protein–aptamer	5MSF, 6MSF, 7MSF	[77]
TRAP–ssRNA site	1C9S	[78]
Rho–poly(C)	2A8V	[79]
Ffh–SRP RNA	1DUL	[80]
Single domain		
RNP (RBD)		
U1A–U1 hairpin	1URN	[81]
U1A–3′ UTR loop	1AUD, 1DZ5	[82,83]
Sxl–polypyrimidine tract	1B7F	[84]
PABP–poly(A)	1CVJ	[85]
Nova KH–gly receptor pre-mRNA	–	[86]
dsRBD–dsRNA	1DI2	[87]
HIV-1 nucleocapsid–Ψ hairpin	1A1T	[88]
Peptide		
HIV-1 Rev–RRE	1ETF	[43]
HIV-1 Rev–aptamer	1ULL, 484D	[39,45••]
HTLV-I Rex–aptamer	–	[40••]
λ N–box B	1QFQ	[41](a)
P22 N–box B	1A4T	[42]
BIV Tat–TAR	1MNB, 1BIV	[29,30]
HIV-1 Tat–TAR	1ARJ	[23,25••,26]
Amino acid ligand		
Argininamide–HIV-2 TAR	1AKX	[89]
Arginine–aptamer	1KOC	[90]

*Complexes were solved by X-ray crystallography or NMR spectroscopy. Complexes classified as multiprotein contain at least two different protein subunits, whereas some of the complexes classified as protein are multimers of a single subunit (such as the MS2 coat protein and TRAP). †Protein Data Bank (PDB) entry numbers are provided where available, including an entry for a λ N–box B complex that is not yet published. (a) M Schaerpf *et al.*, personal communication. RBD, RNA-binding domain.

'intrinsically unstructured' peptide adopts very different conformations when bound to a GTPase domain or an autoinhibitory domain [8], and a third peptide folds into an α helix or β sheet when engineered into different regions of protein G [9,10]. The amino acids used to form specific interactions at peptide–protein interfaces vary greatly in sequence complexity and in the array of interactions used for recognition [5]. The recognition of proline-rich peptides by SH3 and WW domains seems particularly simple in that the domains just recognize the *N*-substituted nature of the proline sidechain [11]. The diversity of peptide structures observed in the context of proteins, their dependence on the surrounding protein scaffold for folding and the complexity of sidechain interactions used for recognition are quite similar to what is observed when peptides are bound to RNAs. The specific interactions that dominate peptide–RNA and peptide–protein interfaces may differ, but it is clear that both proteins and RNAs can provide appropriate environments to stabilize and evolve highly specific binding peptides.

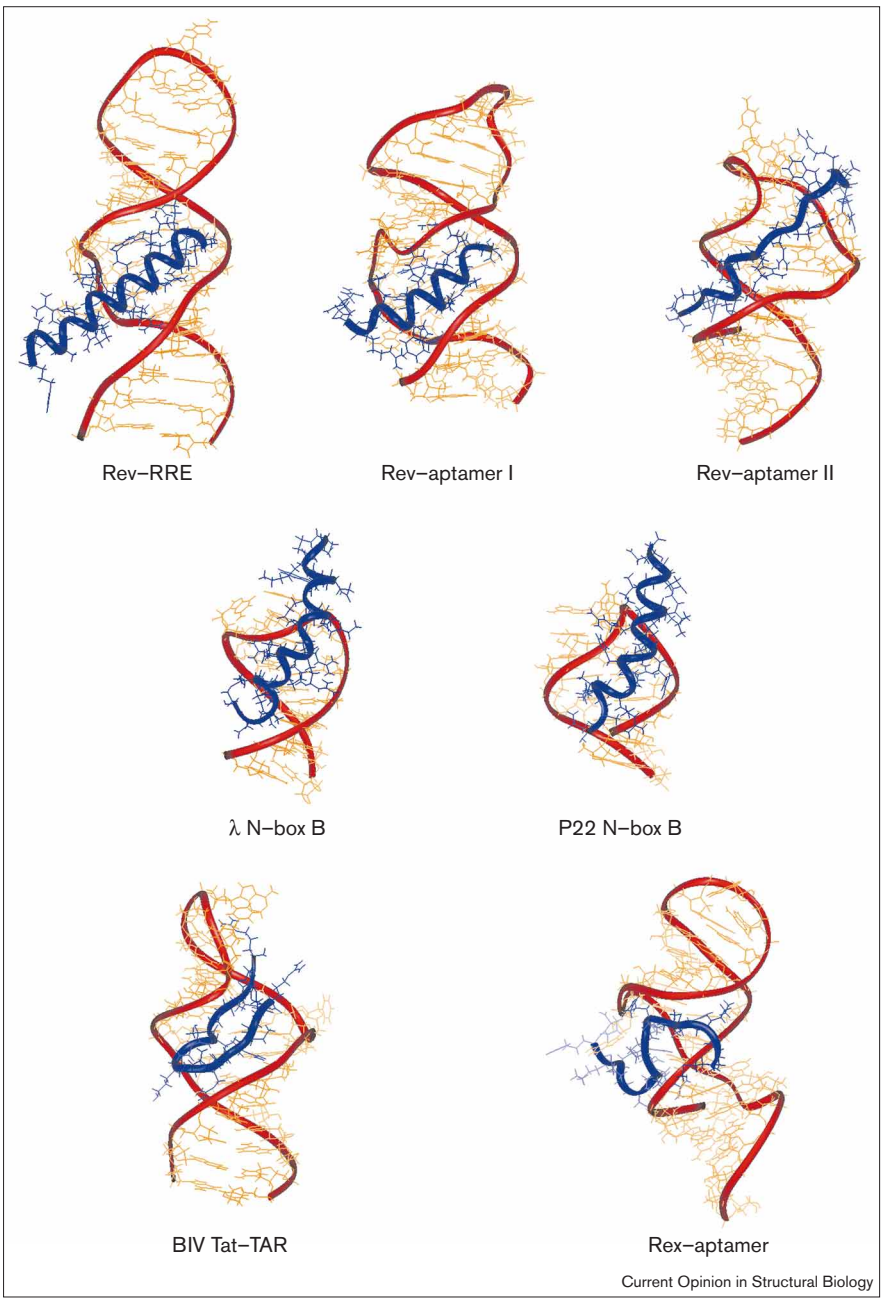
Arginine-rich RNA-binding peptides

Several recent reviews describe the structures of RNA–protein complexes, RNA-binding domains and RNAs, and define many of the important principles of recognition [3,12–16]. Two additional in-depth reviews focus on peptide–RNA interactions [17•,18•] and extensively summarize the structural information on peptide–RNA complexes available as of last year. The reader is referred to these reviews for additional background information.

Most RNA-binding peptides defined to date are highly basic and generally rich in arginines. Arginine-rich RNA-binding domains, sometimes referred to as arginine-rich motifs or ARMs, were initially found in bacteriophage antiterminator proteins, viral regulatory proteins, viral coat proteins and ribosomal proteins [19]. Studies with peptide fragments demonstrated that these regions alone, typically <20 amino acids in length, were able to bind their RNA sites with high affinity and specificity, and generally provided good models of the corresponding protein–RNA interaction (see [17•,18•]). Several plant virus coat proteins also contain short, basic RNA-binding domains that tend to be rich in lysines, rather than arginines [20]. It is perhaps not surprising that basic peptides can bind RNA with high affinity through electrostatic interactions, but what is more remarkable is their relatively high specificity. This apparently is possible because RNAs form distinct binding pockets that help 'mold' the peptide structure, with specific interactions sometimes involving just a few amino acid sidechains.

Eight NMR structures of peptide–RNA complexes have been solved to date (Table 1), although, as described below, the HIV-1 Tat peptide does not adopt a defined conformation when bound to TAR (*trans*-acting response element). In the other seven cases, the peptides are found to adopt a wide range of structures (Figure 1), including an α helix (HIV-1 Rev bound to the RRE [Rev response element] or to an aptamer), bent α helices (phage λ and P22 N peptides bound to box B RNAs), a β hairpin (BIV [bovine immunodeficiency virus] Tat bound to BIV TAR) and extended conformations (Rev and Rex peptides bound to aptamers). In these seven cases, the peptides are either unstructured or poorly structured in the absence of RNA and become

Figure 1



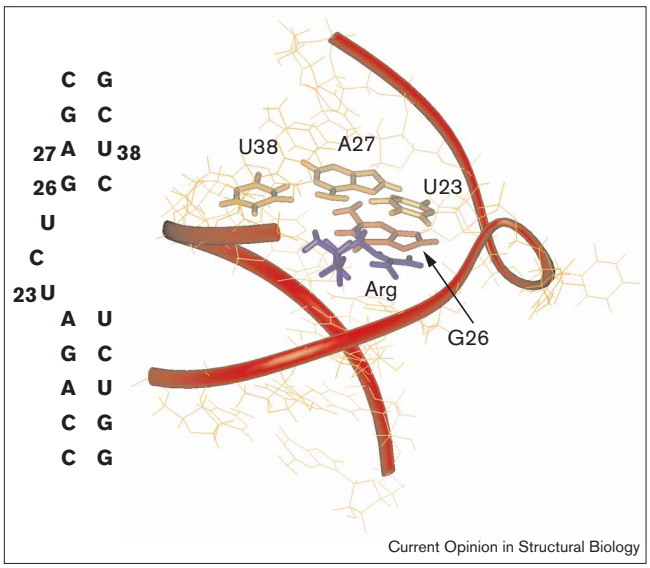
NMR structures of peptide–RNA complexes. Averaged minimized models of the HIV-1 Rev–RRE [43], λ N–box B [41] and BIV Tat–TAR [29] complexes and single calculated models of the HIV-1 Rev–aptamer I [39], HIV-1 Rev–aptamer II [45••], P22 N–box B [42] and HTLV-I Rex–aptamer [40••] complexes. Peptides are in blue and RNAs in red and orange.

fixed upon binding. Several papers on peptide–RNA complexes published in the past year add to the picture of structural diversity and are described, beginning with the most disordered peptides and ending with those that more closely resemble structured protein domains.

HIV-1 Tat

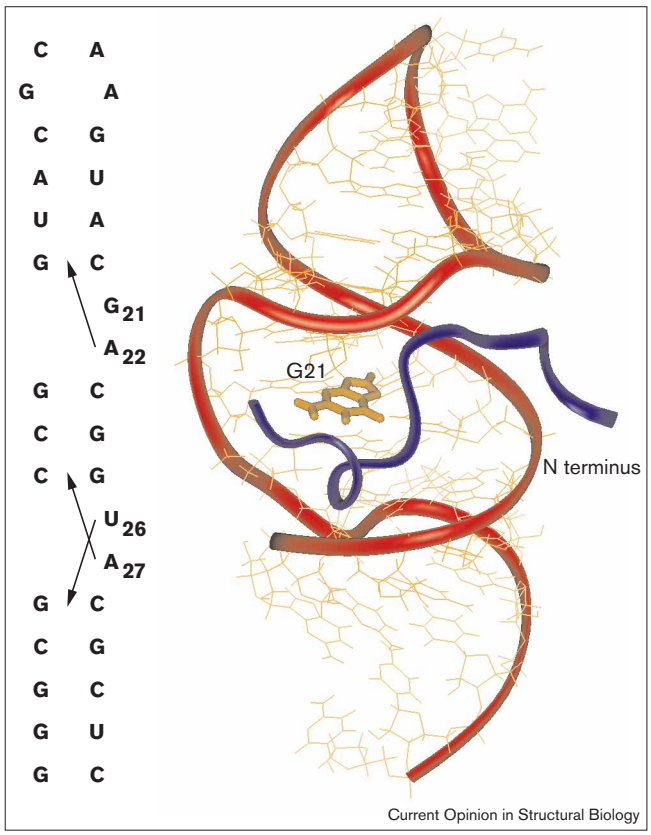
HIV-1 Tat is an essential viral transcription factor that enhances the processivity of RNA polymerase II and functions when bound to the TAR RNA hairpin located at the 5′ end of the viral mRNA transcripts (Figure 2). Tat interacts directly with a three-nucleotide bulge region in TAR and forms a ternary complex with cyclin T1, which extends recognition to the six-nucleotide loop [21]. It is not yet clear whether cyclin T1, Tat or both directly contact the TAR loop. The arginine-rich RNA-binding domain of Tat is just nine amino acids in length (RKKRRQRRR) and contains one key arginine, which forms specific hydrogen bonds with G26 and interacts with surrounding phosphates at the bulge [22,23]. An arginine-binding pocket in TAR is formed by the coaxial stacking of the two RNA helices and by the formation of a U23–A27:U38 base triple (Figure 2) induced by Tat peptide binding; a similar conformational change can be induced by the free amino acid arginine [23].

Figure 2



NMR model of an HIV-1 argininamide–TAR complex [23]. The arginine–G26 contact and U23–A27:U38 base triple are highlighted. The corresponding secondary structure of the bulge region is shown on the left. The six-nucleotide loop (not shown) is on top of the upper stem.

Figure 3



NMR structure of the HTLV-I Rex peptide–RNA aptamer complex [40••]. Base triples between bulge nucleotides and G:C base pairs in the surrounding stems are indicated by arrows in the corresponding secondary structure on the left.

Basic residues surrounding the arginine contribute important electrostatic interactions and raise binding affinity to the subnanomolar range (the free amino acid binds with millimolar affinity) [24]. Additional amino acids on the C-terminal side of the arginine-rich region and some additional arginines within the region enhance the kinetic stability of the complex *in vitro* [25••], although a single arginine in the region is sufficient for Tat function *in vivo* [22]. NMR and CD experiments show no evidence that Tat peptides adopt discrete structures upon TAR binding [24,25••,26] or in the context of the intact protein [27]; many types of amino acid substitutions, including changes to proline, are tolerated [28], further suggesting the absence of a defined structure. An analogous Tat–TAR complex in BIV displays a much more extensive peptide–RNA interface, with a variety of hydrogen bonding, electrostatic and hydrophobic contacts mediated by seven amino acids [29,30]. In this case, the peptide adopts a discrete β-hairpin conformation upon RNA binding and fits deeply within the major groove near a bulge (Figure 1). The HIV-1 Tat peptide–TAR complex seems to be an extreme case in which there are so few specific amino acid contacts and such a poorly defined structure that the peptide may be viewed as a simple ligand for the RNA. However, the cyclin T1 partner is likely to impart more structure to Tat in the context of the ternary complex.

There has been considerable controversy concerning the existence of the base triple in HIV-1 TAR (see [17•,25••]). The initial NMR data on the arginine–TAR complex [23] did not provide unambiguous evidence for base triple formation, as no imino proton corresponding to Hoogsteen-bonded U23 was observed. However, a mutant TAR in which the U–A:U triple was substituted with an isomorphic C^+–G:C triple formed the same structure as wild-type TAR when bound to arginine [31], additional TAR mutants showed a direct correlation between arginine-binding affinity and their ability to form the Hoogsteen interaction [32] and NMR experiments on a derivative of the closely related HIV-2 TAR showed imino protons expected of a C^+–G:C triple [33]. The base triple model was disputed by another NMR study, which suggested that, although U23 was positioned in the RNA major groove near the A:U base pair, the formation of the base triple was incompatible with the calculated NMR models [26]. A recent study on a kinetically stable HIV-1 Tat peptide–TAR complex appears to have resolved the controversy by finding the phantom imino proton attributed to U23 [25••]. NOEs (nuclear Overhauser enhancements) observed in this complex were consistent with a TAR structure containing a nearly ideal planar base triple.

Given the lack of a detailed structure of the Tat peptide bound to TAR, Rana and colleagues have been using

chemical cross-linking and cleavage methods to help further define the mode of peptide binding (summarized in [34•]). In one recent study, they incorporated Fe(II)–EDTA at a single position in the TAR bulge and observed that nearby hydroxyl radical cleavage sites were protected by peptide binding in a manner consistent with the extrusion of the modified bulge nucleotide [35]. Indeed, NMR experiments show at least two conformations in which bulge nucleotides are partially stacked between the two helical stems in the unbound RNA and become exposed upon peptide binding [25••]. In another study, a cross-link was observed between the Tat peptide and 4-thiol-uracil incorporated in the TAR loop and, although the peptide cross-link could not be localized, these and previous cross-linking data suggest that the C-terminal end of the arginine-rich domain may be located near the loop [34•].

As a step towards developing Tat–TAR inhibitors, several TAR-binding peptide analogs have been identified. Peptoids containing arginine sidechains were selected from combinatorial libraries and, according to NMR experiments, appear to bind TAR in a manner analogous to Tat peptides [36]. In contrast, a TAR-binding tripeptide (L-Lys-D-Lys-L-Asn) isolated from a combinatorial library containing a mixture of D- and L-amino acids did not produce TAR NMR spectra characteristic of the arginine-induced conformational change [37••]. A D-amino acid version of a Tat RNA-binding peptide binds TAR in a manner similar to the normal L-peptide [38]. This D-peptide, as well as the TAR-binding peptoids and tripeptides, shows the inhibition of Tat-mediated transcriptional activation in tissue culture.

Human T-cell leukemia virus type I Rex

The HIV-1 Tat–TAR interaction shows how a motif as simple as a base triple can help create a specific peptide-binding pocket in an RNA structure. A similar base triple is seen in the BIV Tat–TAR complex [30] and in the complex of an RNA aptamer bound to a Rev peptide [39] (see below). The importance of base triples was further underscored by the recently determined structure of a human T-cell leukemia virus type I (HTLV-I) Rex peptide bound to an RNA aptamer [40••]. Rex is an essential viral protein that transports partially spliced and unspliced viral mRNAs from the nucleus to the cytoplasm. Rex binds to the Rex response element (RxRE) RNA via its N-terminal arginine-rich domain and the NMR structure of a 15-mer N-terminal Rex peptide (MPKTRRRPRRSQRKR) bound to an RNA aptamer has been solved [40••]. A complex formed with the wild-type RxRE RNA gave poor NMR spectra and its structure could not be determined.

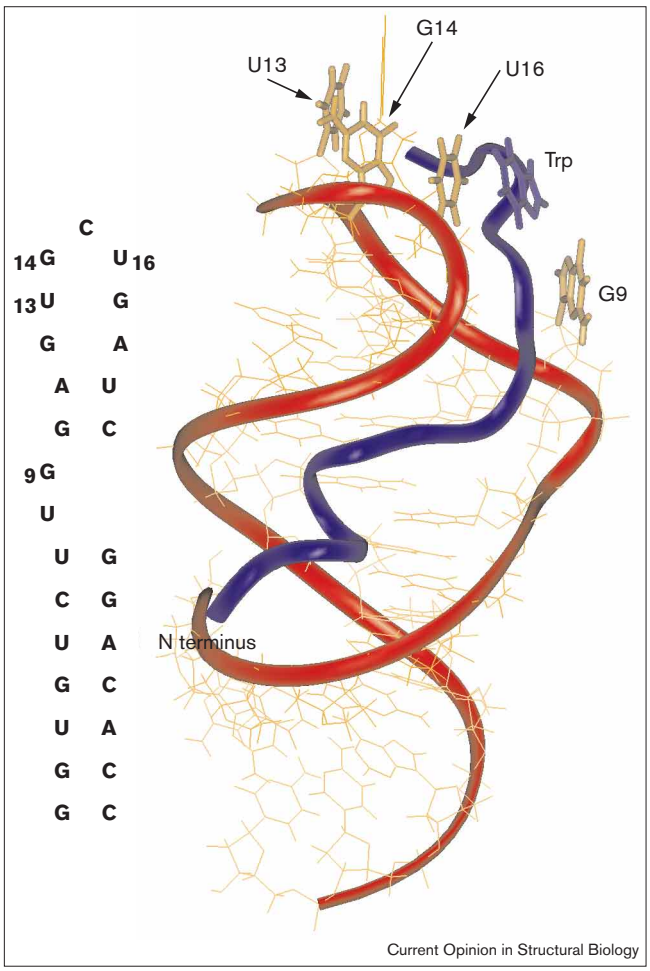
The aptamer Rex-binding site is composed of three stacked helices, in which the junctions are held together by three base triples (A27–G25:C6, U26–G5:C28 and A22–G9:C20) that, as in TAR, are formed by hydrogen bonding interactions between bulge nucleotides and adjacent base pairs (Figure 3). The stems at one junction are oriented at a 50° angle and are nearly parallel at the second junction, resulting in a roughly L-shaped RNA conformation (see the view in Figure 1). In addition to stabilizing the orientation of the stems, one of the base triples widens the major groove to allow access by the Rex peptide. Increased accessibility of RNA major grooves at nonhelical junctions is a general characteristic used by almost all RNA-binding proteins and peptides (see [3,12,13,15,17•,18•]). Another interesting facet of the RNA structure is the use of G21, located in a bulge, to stack over the peptide (Figure 3). Both the base and the sugar are used to help keep the peptide in place and this interaction alone accounts for about 10% of the total surface area of the interface. Hydrophobic interactions with looped-out or bulged bases also have been observed in the phage λ and P22 N peptide–box B RNA and BIV Tat–TAR complexes [29,30,41,42]. Hydrophobic interactions in which unpaired bases are buried in hydrophobic pockets of a protein are common in protein–RNA complexes (see [3,12,13,15]), but it seems quite remarkable that peptides without such pockets can utilize these types of interactions.

Most of the Rex peptide (residues 5–13) adopts a defined conformation upon binding. The S-shaped peptide fold follows the contour of the RNA major groove (Figure 3) and appears to be defined almost entirely by its interactions at the peptide–RNA interface. A short stretch of 3_{10} helix (three residues near the C terminus) is the only apparent regular secondary structure. The dominant base-specific contacts that could be assigned unambiguously involve arginine–guanine hydrogen bonds and, in some cases, these are supported by peptide backbone interactions with adjacent cytosine bases. The arginine guanidinum sidechains may be further oriented by interactions with phosphate and 2′-OH groups. The charged nature of the arginine sidechain and its potential to donate five hydrogen bonds appears to provide many opportunities for arginine-rich peptides to recognize specific arrays of acceptors in an RNA site [17•,18•,22]. The Rex aptamer clearly has evolved a rather intricate RNA architecture in order to accommodate these interactions and to establish the fold of the peptide. It will be interesting to see whether the peptide folds in a similar manner in the context of its natural RxRE site.

HIV-1 Rev

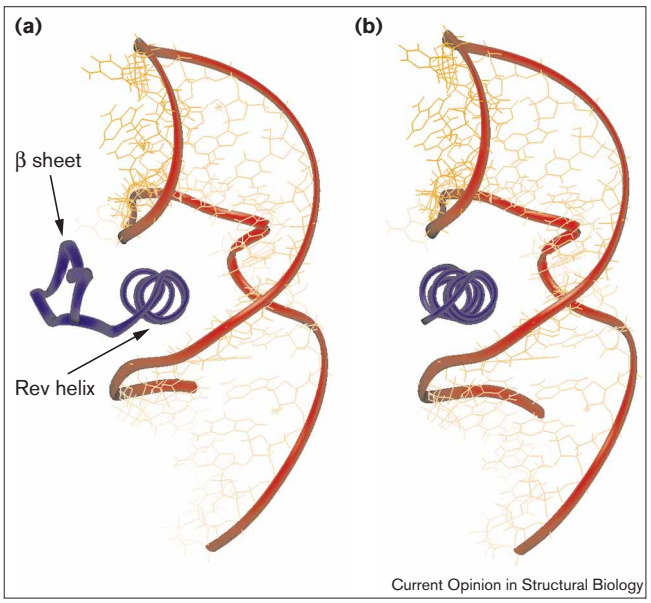
The HIV-1 Rev peptide provides an example in which a peptide can adopt two different folds when placed in the context of two different RNA sites. Like Rex, Rev is an essential viral protein that transports partially spliced and unspliced viral mRNAs from the nucleus to the cytoplasm. The Rev RNA-binding domain is 17 amino acids in length (TRQARRNRRRRWRERQR) and binds deeply within the major groove of its RRE site as an α helix [43] (Figure 1). In this case, the peptide is partially helical prior to binding and becomes stabilized in the

Figure 4



NMR model of the HIV-1 Rev peptide–RNA aptamer II complex [45••]. The corresponding RNA secondary structure is shown on the left.

Figure 5



Model of an HIV-1 RRE-binding zinc finger bound to RNA [47••] **(a)** compared with the Rev peptide–RRE complex **(b)**.

complex [44]. The NMR structure of a Rev peptide–RNA aptamer complex also showed an α helix buried deeply in the major groove [39] (Figure 1). Recently, the NMR structure of the same peptide complexed to a different aptamer was examined [45••] and, although the quality of the NMR spectra limited the structural analysis, the peptide was found to adopt an extended, rather than a helical, conformation (Figure 4).

The peptide-binding site on this second RNA aptamer encompasses a two-nucleotide bulge and a seven-nucleotide loop, with two intervening base pairs (Figure 4). The two bulge bases are involved in interactions grossly resembling those in the Rex–aptamer complex [40••]: U8 probably forms a base triple with the A11:U19 base pair, whereas G9 stacks over the peptide to help anchor it in the major groove (Figure 4). The loop contains a sheared G:A base pair that stacks on the Watson–Crick pairs below, three additional unpaired bases (U13, G14 and U16) that stack on the G:A pair and two looped-out pyrimidines that are not involved in peptide recognition. About half of the peptide backbone chain is well defined, but most sidechains are not. Nevertheless, it is clear that the peptide conformation is extended and that the indole ring of tryptophan is used to cap the stacked arrangement of bases in the loop (Figure 4) and to anchor the C terminus of the peptide. A related stacking arrangement has been observed in the phage λ N peptide–box B RNA complex, in which tryptophan extends the base stacking in a GNRA tetraloop [39]. An amide proton and an amide carbonyl adjacent to tryptophan in the peptide backbone form hydrogen bonds with U16 in the loop and several arginine sidechains point toward guanine bases in the major groove and toward the phosphate backbone, although no definitive hydrogen bonding patterns could be inferred. When the Rev peptide in its helical conformation binds to the RRE, it uses asparagine to hydrogen bond across a critical G:A base pair, but there is no evidence for an asparagine contact when the peptide is bound in its extended conformation to the aptamer site.

Despite the lack of structural detail, it seems clear that the Rev peptide can be forced into different conformations depending on the RNA scaffold, much as a peptide structure can be differentially molded within different protein contexts. It is expected that the *in vitro* selection of other peptide-binding RNA aptamers will reveal even more structural diversity, but it will be especially interesting if peptides are found to adapt their structures in different natural RNA contexts. In addition to selecting RNA aptamers, it also has been possible to select different peptide aptamers that bind the RRE in apparently nonhelical

conformations [46], further indicating how critical the RNA scaffold can be in molding peptide conformations.

Zinc fingers

In this review, an RNA-binding peptide has been defined as a contiguous peptide chain of ≤30 amino acids that is unfolded or only weakly structured in the absence of RNA and yet binds RNA in a sequence-specific manner. The arbitrary nature of this definition and the structural classification in Table 1 is highlighted by studies with RNA-binding zinc finger peptides. Zinc fingers are stable, compact domains composed of an α helix and a β sheet held together by a zinc ion; they are often used in nucleic acid recognition. A recent study engineered the Rev peptide, which is poorly structured in the absence of RNA, into a zinc finger framework in order to stabilize the helix prior to binding [47••]. It is known that the specific RNA-binding affinity of the Rev peptide is proportional to its preformed helical content [48] and, thus, it was predicted that the stabilized helix would recognize the RRE with high affinity. Modeling indicated that the β-sheet portion of the finger could be accommodated when bound to the RNA if the helix bound in the same manner as in the Rev peptide–RRE complex (Figure 5). The hybrid finger folded into a stable structure and specifically recognized the RRE with the expected high affinity [47••]. Thus, an independently folded RNA-binding domain was created from an unstable arginine-rich RNA-binding peptide.

Peptides composed of two zinc fingers that specifically recognize the RRE have been selected by phage display [49]. On the basis of the library design, these peptides probably use the α-helical portion of at least one of the fingers for binding. In another phage display experiment, a zinc finger was selected that specifically binds the G:A base pair of an RNA triplet when anchored by two surrounding zinc fingers bound to a DNA duplex [50•]. Again, according to the library design, amino acids from the helix are likely to be used in binding. NMR experiments indicated that the G:A pair in this context adopted a nonsheared pairing arrangement and modeling suggested several plausible recognition strategies for the base pair. Thus, it appears that zinc fingers in monomeric or multimeric forms will provide yet another layer of structural diversity for peptide–RNA recognition.

Conclusions

This past year has seen several new additions to the diverse world of peptide–RNA interactions. It is obvious that RNA structural frameworks provide excellent environments for peptide folding, perhaps reflecting early events in the 'RNA world'. Arginine-rich peptides seem ideally suited to adapting to different RNA structures, being rich in hydrogen bonding potential, while, at the same time, having an inherently high affinity for negatively charged nucleic acids. For a well-structured RNA, it appears to be possible to find highly specific RNA-binding peptides in relatively small collections of arginine-rich sequences. Perhaps this provides an evolutionary strategy for RNA viruses to rapidly co-evolve RNA-binding proteins together with their RNA targets and may explain why arginine-rich peptides are commonly found in viruses.

In an interesting recent study on protein structure, it was found that mutating a single amino acid in an N-terminal peptide from the Arc repressor caused the peptide to switch from a β-strand conformation to a right-handed helix, while still maintaining a stable protein fold [51]. It was suggested that flexible or late folding regions of protein located at their termini might be especially well suited to evolving new folds with only small changes in sequence. This reinforces the idea that peptide structures can readily adapt when given an appropriate framework for folding, such as a structured protein or RNA.

Acknowledgements

I thank Dinshaw Patel, Paul Roesch and Jennifer Doudna for providing coordinates and communicating results prior to publication, and Wendell Lim for helpful discussions. Supported by grants from the National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Jones S, Thornton MJ: **Principles of protein-protein interactions.** *Proc Natl Acad Sci USA* 1996, **93**:13-20.

2. Jones S, van Heyningen P, Berman HM, Thornton JM: **Protein–DNA interactions: a structural analysis.** *J Mol Biol* 1999, **287**:877-896.

3. Draper DE: **Themes in RNA-protein recognition.** *J Mol Biol* 1999, **293**:255-270.

4. Frankel AD: **If the loop fits...** *Nat Struct Biol* 1999, **6**:1081-1083.

5. Stanfield RL, Wilson IA: **Protein-peptide interactions.** *Curr Opin Struct Biol* 1995, **5**:103-113.

6. Prehoda KE, Lee DJ, Lim WA: **Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly.** *Cell* 1999, **97**:471-480.

7. Landry SJ, Jordan R, McMacken R, Gierasch LM: **Different conformations for the same polypeptide bound to chaperones DnaK and GroEL.** *Nature* 1992, **355**:455-457.

8. Kim AS, Kakalis LT, Abdul-Manan N, Liu GA, Rosen MK: **Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein.** *Nature* 2000, **404**:151-158.

9. Minor DL Jr, Kim PS: **Context-dependent secondary structure formation of a designed protein sequence.** *Nature* 1996, **380**:730-734.

10. Cregut D, Civera C, Macias MJ, Wallon G, Serrano L: **A tale of two secondary structure elements: when a beta-hairpin becomes an alpha-helix.** *J Mol Biol* 1999, **292**:389-401.

11. Nguyen JT, Turck CW, Cohen FE, Zuckermann RN, Lim WA: **Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors.** *Science* 1998, **282**:2088-2092.

12. Cusack S: **RNA-protein complexes.** *Curr Opin Struct Biol* 1999, **9**:66-73.

13. Steitz TA: **RNA recognition by proteins.** In *The RNA World*, edn 2. Edited by Gesteland RF, Cech TR, Atkins JF. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999:427-450.

14. Siomi H, Dreyfuss G: **RNA-binding proteins as regulators of gene expression.** *Curr Opin Genet Dev* 1997, **7**:345-353.

15. Hermann T, Westhof E: **Non-Watson-Crick base pairs in RNA–protein recognition.** *Chem Biol* 1999, **6**:R335-R343.

16. Hermann T, Patel JD: **Stitching together RNA tertiary architectures.** *J Mol Biol* 1999, **294**:829-849.

17. • Patel DJ: **Adaptive recognition in RNA complexes with peptides and protein modules.** *Curr Opin Struct Biol* 1999, **9**:75-87.
An extensive and well-written review that describes the structures of RNA complexes with peptides and small protein modules, emphasizing the critical importance of induced fit.

18. • Narayana N, Weiss MA: **RNA recognition by arginine-rich peptide motifs.** *Biopolymers* 1999, **48**:167-180.
Another in-depth and well-written review that focuses on RNA interactions with arginine-rich peptides and provides an excellent background on the biology of the interactions.

19. Lazinski D, Grzadzielska E, Das A: **Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif.** *Cell* 1989, **59**:207-218.

20. Ansel-McKinney P, Scott SW, Swanson M, Ge X, Gehrke L: **A plant viral coat protein RNA binding consensus sequence contains a crucial arginine.** *EMBO J* 1996, **15**:5077-5084.

21. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA: **A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA.** *Cell* 1998, **92**:451-462.

22. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD: **Arginine-mediated RNA recognition: the arginine fork.** *Science* 1991, **252**:1167-1171.

23. Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR: **Conformation of the TAR RNA-arginine complex by NMR spectroscopy.** *Science* 1992, **257**:76-80.

24. Tao J, Frankel AD: **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* 1993, **90**:1571-1575.

25. •• Long KS, Crothers DM: **Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA.** *Biochemistry* 1999, **38**:10059-10069.
A thorough NMR study of HIV-1 TAR in the absence and presence of a Tat peptide with which it forms a kinetically stable complex. TAR exists in major and minor conformers in its unbound state that differ in their base stacking arrangements at the bulge. Upon Tat peptide binding, TAR adopts a single conformation in which an imino proton resonance attributed to U23 of the base triple is observed. The kinetic stability of this complex prevents rapid exchange of the imino proton with solvent and depends on multiple arginines in the peptide. This paper provides the strongest evidence to date in support of the base triple model for TAR.

26. Aboul-ela F, Karn J, Varani G: **The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein.** *J Mol Biol* 1995, **253**:313-332.

27. Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P: **Structural studies of HIV-1 Tat protein.** *J Mol Biol* 1995, **247**:529-535.

28. Tan R, Frankel AD: **Structural variety of arginine-rich RNA-binding peptides.** *Proc Natl Acad Sci USA* 1995, **92**:5282-5286.

29. Puglisi JD, Chen L, Blanchard S, Frankel AD: **Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex.** *Science* 1995, **270**:1200-1203.

30. Ye X, Kumar RA, Patel DJ: **Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex.** *Chem Biol* 1995, **2**:827-840.

31. Puglisi JD, Chen L, Frankel AD, Williamson JR: **Role of RNA structure in arginine recognition of TAR RNA.** *Proc Natl Acad Sci USA* 1993, **90**:3680-3684.

32. Tao J, Chen L, Frankel AD: **Dissection of the proposed base triple in human immunodeficiency virus TAR RNA indicates the importance of the Hoogsteen interaction.** *Biochemistry* 1997, **36**:3491-3495.

33. Brodsky AS, Erlacher HA, Williamson JR: **NMR evidence for a base triple in the HIV-2 TAR C-G.C$^+$ mutant-argininamide complex.** *Nucleic Acids Res* 1998, **26**:1991-1995.

34. • Wang Z, Huq I, Rana TM: **Proximity of a Tat peptide to the HIV-1 TAR RNA loop region determined by site-specific photo-cross-linking.** *Bioconj Chem* 1999, **10**:512-519.
A paper that brings together the results of many chemical cross-linking and cleavage studies to arrive at a model for the Tat peptide bound to TAR.

35. Huq I, Tamilarasu N, Rana TM: **Visualizing tertiary folding of RNA and RNA-protein interactions by a tethered iron chelate: analysis of HIV-1 Tat-TAR complex.** *Nucleic Acids Res* 1999, **27**:1084-1093.

36. Hamy F, Felder ER, Heizmann G, Lazdins J, Aboul-ela F, Varani G, Karn J, Klimkait T: **An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication.** *Proc Natl Acad Sci USA* 1997, **94**:3548-3553.

37. •• Hwang S, Tamilarasu N, Ryan K, Huq I, Richter S, Still WC, Rana TM: **Inhibition of gene expression in human cells through small molecule-RNA interactions.** *Proc Natl Acad Sci USA* 1999, **96**:12997-13002.
This paper identifies a tripeptide from a combinatorial library composed of D- and L-amino acids that binds to HIV-1 TAR and inhibits transcriptional activation. This tripeptide appears to lock TAR into a different conformation from the Tat-bound conformation, providing an example of how an inhibitor might take advantage of RNA flexibility and stabilize a conformation that cannot be bound by its protein target.

38. Huq I, Ping Y-H, Tamilarasu N, Rana TM: **Controlling human immunodeficiency virus type 1 gene expression by unnatural peptides.** *Biochemistry* 1999, **38**:5172-5177.

39. Ye X, Gorin A, Ellington AD, Patel DJ: **Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex.** *Nat Struct Biol* 1996, **3**:1026-1033.

40. •• Jiang F, Gorin A, Hu W, Majumdar A, Baskerville S, Xu W, Ellington A, Patel DJ: **Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples.** *Structure* 1999, **7**:1461-1472.
This paper presents the structure of the human T-cell leukemia virus type I (HTLV-I) Rex peptide bound to an RNA aptamer. The RNA adopts a very interesting architecture that includes a series of three base triples and an unpaired base that stacks over the peptide to help stabilize its extended conformation. The structure adds considerable diversity to the repertoire of peptide–RNA complexes.

41. Legault P, Li J, Mogridge J, Kay LE, Greenblatt J: **NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif.** *Cell* 1998, **93**:289-299.

42. Cai Z, Gorin A, Frederick R, Ye X, Hu W, Majumdar A, Kettani A, Patel DJ: **Solution structure of P22 transcriptional antitermination N peptide-box B RNA complex.** *Nat Struct Biol* 1998, **5**:203-212.

43. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR: **Alpha helix major groove recognition in an HIV-1 Rev peptide-RRE RNA complex.** *Science* 1996, **273**:1547-1551.

44. Tan R, Frankel AD: **Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex.** *Biochemistry* 1994, **33**:14579-14585.

45. •• Ye X, Gorin A, Frederick R, Hu W, Majumdar A, Xu W, McLendon G, Ellington A, Patel DJ: **RNA architecture dictates the conformations of a bound peptide.** *Chem Biol* 1999, **6**:657-669.
A striking demonstration that a single peptide can adopt different conformations when bound to two different RNA aptamers. Instead of binding as an α helix (as previously observed [39,43]), the Rev peptide is anchored in the RNA loop via an interesting set of stacking interactions and the remainder of the RNA creates a binding pocket in the major groove that holds the rest of the peptide in an extended conformation.

46. Harada K, Martin SS, Tan R, Frankel AD: **Molding a peptide into an RNA site by in vivo peptide evolution.** *Proc Natl Acad Sci USA* 1997, **94**:11887-11892.

47. •• McColl DJ, Honchell CD, Frankel AD: **Structure-based design of an RNA-binding zinc finger.** *Proc Natl Acad Sci USA* 1999, **96**:9521-9526.
Using existing structural information, a specific RRE-binding zinc finger was created by placing the Rev α helix into the framework of a zinc finger from the Zif268 DNA-binding protein. This study demonstrates that a relatively unstructured RNA-binding peptide can be prestabilized prior to binding.

48. Tan R, Chen L, Buettner JA, Hudson D, Frankel AD: **RNA recognition by an isolated α helix.** *Cell* 1993, **73**:1031-1040.

49. Friesen WJ, Darby MK: **Specific RNA binding proteins constructed from zinc fingers.** *Nat Struct Biol* 1998, **5**:543-546.

50. Blancafort P, Steinberg SV, Paquin B, Klinck R, Scott JK, Cedergren R: **The recognition of a noncanonical RNA base pair by a zinc finger protein.** *Chem Biol* 1999, **6**:585-597.

The authors placed an RNA triplet in the midst of a DNA helix and used phage display to isolate a zinc finger that recognized a G:A pair in the triplet when the peptide was anchored by two surrounding DNA-binding fingers. The authors suggest that fingers that are specific for other types of non-Watson–Crick RNA base pairs may be isolated by this method.

51. Cordes MH, Walsh NP, McKnight CJ, Sauer RT: **Evolution of a protein fold *in vitro*.** *Science* 1999, **284**:325-328.

52. Clemons WMJ, May JL, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V: **Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution.** *Nature* 1999, **400**:833-840.

53. Tocilj A, Schlunzen F, Janell D, Gluhmann M, Hansen HAS, Harms J, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A: **The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: pattern fittings and the identification of a functional site.** *Proc Natl Acad Sci USA* 1999, **96**:14252-14257.

54. Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA: **Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit.** *Nature* 1999, **400**:841-847.

55. Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF: **X-ray crystal structures of 70S ribosome functional complexes.** *Science* 1999, **285**:2095-2104.

56. Price SR, Evans PR, Nagai K: **Crystal structure of the spliceosomal U2B′-U2A′ protein complex bound to a fragment of U2 small nuclear RNA.** *Nature* 1998, **394**:645-650.

57. Agalarov SC, Prasad GS, Funke PM, Stoud CD, Williamson JR: **Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain.** *Science* 2000, **288**:107-112.

58. Rould MA, Perona JJ, Söll D, Steitz TA: **Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution.** *Science* 1989, **246**:1135-1142.

59. Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D: **Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp).** *Science* 1991, **252**:1682-1689.

60. Biou V, Yaremchuk A, Tukalo M, Cusack S: **The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser).** *Science* 1994, **263**:1404-1410.

61. Cusack S, Yaremchuk A, Tukalo M: **The crystal structures of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* tRNA(Lys) and a *T. thermophilus* tRNA(Lys) transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue.** *EMBO J* 1996, **15**:6321-6334.

62. Goldgur Y, Mosyak L, Reshetnikova L, Ankilova V, Lavrik O, Khodyreva S, Safro M: **The crystal structure of phenylalanyl-tRNA synthetase from Thermus thermophilus complexed with cognate $tRNA^{Phe}$.** *Structure* 1997, **5**:59-68.

63. Cusack S, Yaremchuk A, Krikliviy I, Tukalo M: **tRNA(Pro) anticodon recognition by *Thermus thermophilus* prolyl-tRNA synthetase.** *Structure* 1998, **6**:101-108.

64. Sankaranarayanan R, Dock-Bregeon A-C, Romby P, Caillet J, Springer M, Rees B, Ehresmann C, Ehresmann B, Moras D: **The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site.** *Cell* 1999, **97**:371-381.

65. Silvian LF, Wang J, Steitz TA: **Insights into editing from an Ile-tRNA synthetase structure with tRNAIle and mupirocin.** *Science* 1999, **285**:1074-1077.

66. Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J: **Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog.** *Science* 1995, **270**:1464-1472.

67. Nissen P, Thirup S, Kjeldgaard M, Nyborg J: **The crystal structure of Cys-tRNA(Cys)-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA.** *Structure* 1999, **7**:143-156.

68. Schmitt E, Panvert M, Blanquet S, Mechulam Y: **Crystal structure of methionyl-tRNA(fMet) transformylase complexed with the initiator formyl-methionyl-tRNA(fMet).** *EMBO J* 1998, **17**:6819-6826.

69. Conn GL, Draper DE, Lattman EE, Gittis AG: **Crystal structure of a conserved ribosomal protein-RNA complex.** *Science* 1999, **284**:1171-1174.

70. Wimberly BT, Guymon R, McCutcheon JP, White SW, Ramakrishnan V: **A detailed view of a ribosomal active site: the structure of the L11-RNA complex.** *Cell* 1999, **97**:491-502.

71. Stoldt M, Wohnert J, Ohlenschlager O, Gorlach M, Brown LR: **The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition.** *EMBO J* 1999, **18**:6508-6521.

72. Lu M, Steitz TA: **Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Å resolution.** *Proc Natl Acad Sci USA* 2000, **97**:2023-2028.

73. Nikulin A, Serganov A, Ennifar E, Tishchenko S, Nevskaya N, Shepard W, Portier C, Garber M, Ehresmann B, Ehresmann C *et al.*: **Crystal structure of the S15-rRNA complex.** *Nat Struct Biol* 2000, **7**:273-277.

74. Mao H, White SA, Williamson JR: **A novel loop–loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex.** *Nat Struct Biol* 1999, **6**:1139-1147.

75. Valegard K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L: **Crystal structure of an RNA bacteriophage coat protein-operator complex.** *Nature* 1994, **371**:623-626.

76. Grahn E, Stonehouse NJ, Murray JB, van den Worm S, Valegård K, Fridborg K, Stockley PG, Liljas L: **Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements.** *RNA* 1999, **5**:131-138.

77. Rowsell S, Stonehouse NJ, Convery MA, Adams CJ, Ellington AD, Hirao I, Peabody DS, Stockley PG, Phillips SEV: **Crystal structures of a series of RNA aptamers complexed to the same protein target.** *Nat Struct Biol* 1998, **5**:970-975.

78. Antson AA, Dodson EJ, Dodson G, Greaves RB, Chen X, Gollnick P: **Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA.** *Nature* 1999, **401**:235-242.

79. Bogden CE, Fass D, Bergman N, Nichols MD, Berger JM: **The structural basis for terminator recognition by the Rho transcription termination factor.** *Mol Cell* 1999, **3**:487-493.

80. Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA: **Crystal structure of the ribonucleoprotein core of the signal recognition particle.** *Science* 2000, **287**:1232-1239.

81. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K: **Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin.** *Nature* 1994, **372**:432-438.

82. Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, Varani G: **Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation.** *Nature* 1996, **380**:646-650.

83. Varani L, Gunderson SI, Mattaj IW, Kay LE, Neuhaus D, Varani G: **The NMR structure of the 38 kDa U1A protein–PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein.** *Nat Struct Biol* 2000, **7**:329-335.

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

85. Deo RC, Bonanno JB, Sonenberg N, Burley SK: **Recognition of polyadenylate RNA by the poly(A)-binding protein.** *Cell* 1999, **98**:835-845.

86. Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK: **Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the Fragile X syndrome.** *Cell* 2000, **100**:323-332.

87. Ryter JM, Schultz SC: **Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA.** *EMBO J* 1998, **17**:7505-7513.

88. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF: **Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element.** *Science* 1998, **279**:384-388.

89. Brodsky AS, Williamson JR: **Structure of HIV-2 TAR-argininamide complex.** *J Mol Biol* 1997, **267**:624-639.

90. Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M: **Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy.** *Science* 1996, **272**:1343-1347.