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 © 2000 Elsevier Science Ltd. All rights reserved.

Abbreviations
BIV bovine immunodeficiency virus
cD circular dichroism
HTLV-I human T-cell leukemia virus type I
PDB Protein Data Bank
RRE Rex response element
RxRE Rex responsive 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 peptide–protein interactions from an energetic standpoint 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 or 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]. Prolin-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

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 ≤50 amino acids that is unstructured 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 or whether such features make sense from a biological perspective, just as it can be instructive to compare peptide–protein with protein–protein interactions.

Abbreviations
BIV bovine immunodeficiency virus
cD circular dichroism
HTLV-I human T-cell leukemia virus type I
PDB Protein Data Bank
RRE Rex response element
RxRE Rex responsive 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 peptide–protein interactions from an energetic standpoint 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 or 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]. Prolin-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
where available, including an entry for a yet published. (a) M Schaerpf RNA-binding domain.

- Sheet when engineered into different regions of the 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 significant in that the domains just recognize the X-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 peptide–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 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 highly basic proteins, 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 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

<table>
<thead>
<tr>
<th>Complex</th>
<th>PDB entry code</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multisubunit</td>
<td>305</td>
<td>1DG7, 1C59 [52,53]</td>
</tr>
<tr>
<td>505</td>
<td>1C04 [54]</td>
<td></td>
</tr>
<tr>
<td>705</td>
<td>4BBM [55]</td>
<td></td>
</tr>
<tr>
<td>1260S</td>
<td>1A4N [56]</td>
<td></td>
</tr>
<tr>
<td>1S16, 5S16, 7B38–RNA site</td>
<td>1DG5, 1D4C [57]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>RNA-synthetase–RNA complexes</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>1A81 [59]</td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>18BR [60]</td>
<td></td>
</tr>
<tr>
<td>Lysine</td>
<td>– [61]</td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>– [62]</td>
<td></td>
</tr>
<tr>
<td>Proline</td>
<td>– [63]</td>
<td></td>
</tr>
<tr>
<td>Tryptophan</td>
<td>1Q66 [64]</td>
<td></td>
</tr>
<tr>
<td>Isoxazole</td>
<td>1QG2 [65]</td>
<td></td>
</tr>
<tr>
<td>EF-Tu–RNA/PM</td>
<td>1TT1 [66]</td>
<td></td>
</tr>
<tr>
<td>EF-Tu–RNA/PM</td>
<td>1BG3 [67]</td>
<td></td>
</tr>
<tr>
<td>Mot transformase–Met–RNA</td>
<td>2FM1 [68]</td>
<td></td>
</tr>
<tr>
<td>L12–2MS–RNA site</td>
<td>1QA6 [69,70]</td>
<td></td>
</tr>
<tr>
<td>L25–S5–RNA site</td>
<td>1DK1, 1UF4 [71,72]</td>
<td></td>
</tr>
<tr>
<td>S15–RNA site</td>
<td>1DK1 [73]</td>
<td></td>
</tr>
<tr>
<td>L30–autoregulatory RNA</td>
<td>1CK5 [74]</td>
<td></td>
</tr>
<tr>
<td>MS2 coat protein–operator</td>
<td>1ZDI, 1ZDK [75,76]</td>
<td></td>
</tr>
<tr>
<td>MS2 coat protein–aptamer</td>
<td>5M9SF, 6M6F, 7M6F [77]</td>
<td></td>
</tr>
<tr>
<td>TRAP–apupRNA</td>
<td>1CG6 [78]</td>
<td></td>
</tr>
<tr>
<td>Rho–poly(C)</td>
<td>2A8V [79]</td>
<td></td>
</tr>
<tr>
<td>Ffh–SRP RNA</td>
<td>1DUL [80]</td>
<td></td>
</tr>
<tr>
<td>Single domain</td>
<td>RNBP (BBB)</td>
<td>U1A–U1 hairpin</td>
</tr>
<tr>
<td>U1A–3′ UTR loop</td>
<td>1AUD, 1DZ5 [82,83]</td>
<td></td>
</tr>
<tr>
<td>Sol-glycopolymannurin</td>
<td>1BTF [84]</td>
<td></td>
</tr>
<tr>
<td>PRP3–pol(A)</td>
<td>1CVU [85]</td>
<td></td>
</tr>
<tr>
<td>Nova KH–gly receptor premRNA</td>
<td>– [86]</td>
<td></td>
</tr>
<tr>
<td>dsRBD–dsRNA</td>
<td>1DI2 [87]</td>
<td></td>
</tr>
<tr>
<td>HIV-1 nucleocapsid–poly(C)</td>
<td>1AT [88]</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td>HIV-1 Rev–RRE</td>
<td>1ETF [43]</td>
</tr>
<tr>
<td>HIV-1 Rev–aptamer</td>
<td>1ULL, 4D4 [39,40••]</td>
<td></td>
</tr>
<tr>
<td>HTLV-1 Rev–aptamer</td>
<td>– [40••]</td>
<td></td>
</tr>
<tr>
<td>λ-N–box B</td>
<td>1QFO [41,42]</td>
<td></td>
</tr>
<tr>
<td>P22 hi-box B</td>
<td>1AT [43]</td>
<td></td>
</tr>
<tr>
<td>BIV Tat–TAR</td>
<td>1MN1, 1B1V [29,30]</td>
<td></td>
</tr>
<tr>
<td>HIV-1 Tat–TAR</td>
<td>1ARU [23,25••,26]</td>
<td></td>
</tr>
<tr>
<td>Amino acid ligand</td>
<td>Arginine–meso-arginine</td>
<td>1AKK [89]</td>
</tr>
<tr>
<td>Arginine–aptamer</td>
<td>1KOC [90]</td>
<td></td>
</tr>
</tbody>
</table>

*Composites were solved by X-ray crystallography or NMR spectroscopy. Complexes classified as multicomponent 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.

Fitting peptides into the RNA world Frankel 333
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].
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 b-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 exclusion 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-rich TAR peptide in an 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 Tat 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 dictate a specific peptide-binding pocket in an RNA structure. A similar base triple is seen in the HIV Tat–TAR complex [30] and in the complex of an RNA aptamer bound to a Rex 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 Rev response element (RRE) RNA via its N-terminal arginine-rich domain and the NMR structure of a 15-mer N-terminal Rex peptide (MKPRRPRRPRRQRKR) bound to an RNA aptamer has been solved [40••]. A complex formed with the wild-type RRE 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 S-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 feature 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- and 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 a helix (three residues near the C terminus) of the Rex peptide forms a helical groove 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 phosphates and 2'-OH groups. The cross-linking data strongly suggest that these interactions are tightly packed and that the overall 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 base triples (see [3,12,13,15]). The nature of these interactions is not well understood, but it is clear that interactions at the peptide–RNA interface. A short stretch of a helix (three residues near the C terminus) of the Rex peptide forms a helical groove 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 phosphates and 2'-OH groups. The cross-linking data strongly suggest that these interactions are tightly packed and that the overall 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 base triples (see [3,12,13,15]). The nature of these interactions is not well understood, but it is clear that interactions at the peptide–RNA interface. A short stretch of a helix (three residues near the C terminus) of the Rex peptide forms a helical groove 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 phosphates and 2'-OH groups. The cross-linking data strongly suggest that these interactions are tightly packed and that the overall 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 base triples (see [3,12,13,15]).

...
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 50 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 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 GA pair in this context adopted a nonsheared pairing. 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.

### 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

32. Tao J, Chen L, Frankel AD:
31. Puglisi JD, Chen L, Frankel AD, Williamson JR:

The structure of the human TAR adopts a single conformation in which an imino proton resonance for TAR.

This paper provides strongest evidence to date in support of the base triple model and depends on multiple arginines in the peptide. 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 peptides–RNA complexes.

19. Lazinski D, Grzadzielska E, Das A:
18. Narayana N, Weiss MA:
17. Patel DJ:
16. Hermann T, Patel JD:
15. Hermann T, Westhof E:

Structural variety of arginine-rich RNA-binding peptides. 

Using existing structural information, a specific RRE-binding zinc finger was adapted for arginine recognition of TAR RNA. 

Solution structure of a bound peptide. 

Specific RNA binding proteins constructed from zinc fingers. 

Fitting peptides into the RNA world Franks 339


