The EMBO Journal Vol. 20 No. 1 & 2 pp. 178–186, 2001

Structure-based design of a dimeric RNA–peptide complex

Donna M.Campisi, Valerie Calabro and Alan D.Frankel[1]

Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448, USA

[1]Corresponding author
e-mail: frankel@cgl.ucsf.edu

The arginine-rich RNA-binding domain of bovine immunodeficiency virus (BIV) Tat adopts a β-hairpin conformation upon binding to the major groove of BIV TAR. Based on its NMR structure, we modeled dimeric arrangements in which two adjacent TAR sites might be recognized with high affinity by a dimeric peptide. Some dimeric RNAs efficiently bound two unlinked BIV Tat peptides *in vitro*, but could not bind even one monomeric peptide *in vivo*, as monitored by transcriptional activation of human immunodeficiency virus long terminal repeat reporters. Results with additional reporters suggest that extending the RNA helix in the dimeric arrangements inhibits peptide binding by decreasing major groove accessibility. In contrast, a dimeric peptide efficiently bound an optimally arranged dimeric TAR *in vivo*, and bound with an affinity at least 10-fold higher than the monomeric peptide *in vitro*. Mutating specific nucleotides in each RNA 'half-site' or specific amino acids in each β-hairpin of the dimeric peptide substantially decreased binding affinity, providing evidence for the modeled dimer–dimer interaction. These studies provide a starting point for identifying dimeric RNA–protein interactions with even higher binding affinities and specificities.

Keywords: bovine immunodeficiency virus/RNA–protein interactions/TAR/Tat

Introduction

The large body of structural information on DNA–protein complexes has made it possible over the past few years to utilize structure-based approaches to design sequence-specific DNA-binding molecules that might be used to control the expression of specific genes. Two of the most successful examples involve the use of zinc finger proteins to target the DNA major groove and polyamides to target the DNA minor groove (for recent reviews see Dervan and Burli, 1999; Segal and Barbas, 2000). Most of the strategies have combined information from high-resolution structures with combinatorial experiments to help identify the tightest and most specific binders. In addition to targeting DNA, in some cases it also may be desirable to target specific RNA elements involved in transcriptional, post-transcriptional or other RNA-based processes. Structure-based design of RNA binders is currently more difficult than for DNA because fewer structures of RNA–protein complexes have been solved and because the greater structural diversity of RNA makes the 'rules' of recognition more complex (Draper, 1999; Hermann and Patel, 1999; Hermann and Westhof, 1999). Nevertheless, some promising approaches have been identified, with much of the effort to date focused on using aminoglycoside or peptide scaffolds to design specific RNA binders (for recent reviews see Walter *et al*., 1999; Frankel, 2000). We have been particularly interested in using arginine-rich peptides for this purpose.

Arginine-rich RNA-binding domains are found in a relatively large group of proteins and often can specifically recognize their RNA targets even as short (<20 amino acid), isolated peptides (Frankel, 2000). The structures of several RNA–peptide model systems have been solved by NMR and illustrate the wide variety of ways in which specific RNA sites can be recognized by small peptides or proteins. Arginine-rich peptides from human immunodeficiency virus (HIV) Rev, and bacteriophage λ and P22 N proteins recognize internal or terminal RNA loops using α-helical or bent helical conformations, a bovine immunodeficiency virus (BIV) Tat peptide recognizes an RNA bulge using a β-hairpin conformation and an HIV Tat peptide recognizes an RNA bulge using an apparently extended conformation (Calnan *et al*., 1991; Tan *et al*., 1993; Aboul-ela *et al*., 1995; Puglisi *et al*., 1995; Tan and Frankel, 1995; Ye *et al*., 1995, 1996; Battiste *et al*., 1996; Cai *et al*., 1998; Legault *et al*., 1998). The NMR structures of two RNA aptamer–peptide complexes (Jiang *et al*., 1999; Ye *et al*., 1999) and the identification of novel RNA-binding peptides from combinatorial libraries (Harada *et al*., 1996, 1997) demonstrate how RNAs can help 'mold' the conformations of the bound peptides and further highlight the structural versatility of the arginine-rich RNA-binding motif. A comparison between the BIV and HIV Tat–TAR complexes also shows how two RNA sites that are virtually superimposable in structure can be recognized by arginine-rich peptides using entirely different binding strategies; the BIV and HIV complexes use different peptide conformations and amino acid–RNA interactions for recognition and, in addition, the HIV interaction also requires a cellular protein, cyclin T1, for high-affinity binding (Smith *et al*., 1998, 2000; Wei *et al*., 1998; Bogerd *et al*., 2000).

Given the versatility of the arginine-rich motif, we have been interested in exploring whether such peptides might provide reasonable starting points for the design of novel sequence-specific RNA binders. In previous experiments, we identified arginine-rich peptides from combinatorial libraries that bind to the Rev response element (RRE) RNA with high affinities and specificities but appear to bind in non-helical conformations rather than the α-helical conformation used by Rev (Harada *et al*., 1996, 1997).

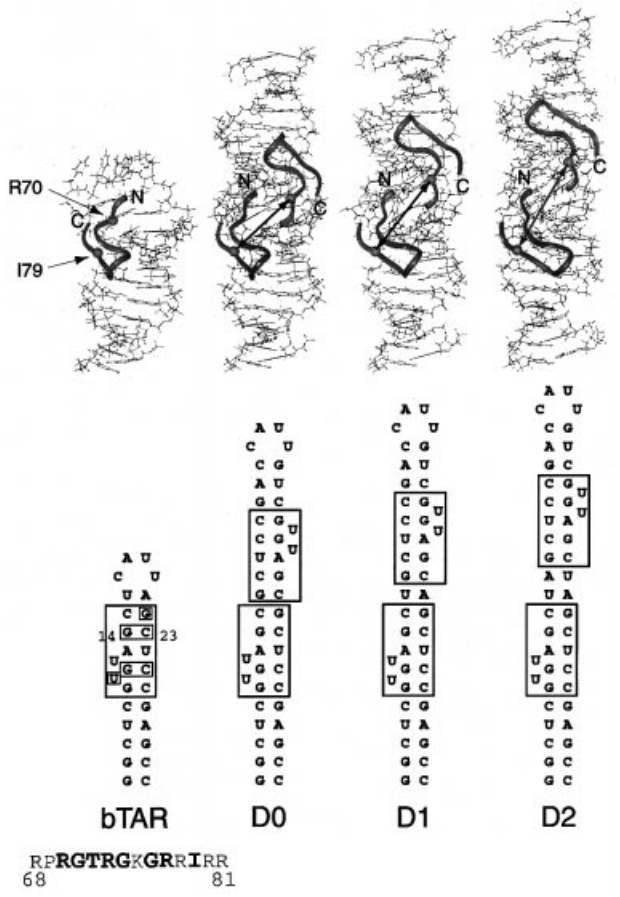

Fig. 1. Modeled structures of D0, D1 and D2 dimeric TAR RNAs with bound BIV Tat peptides, based on a single calculated NMR structure of the BIV Tat peptide–TAR complex (Puglisi *et al*., 1995) (shown on the left). The C_α atoms of Arg70 and Ile79, to be bridged by a linker in the dimeric peptides, are shown as balls. Corresponding RNA secondary structures are also shown, where the boxed regions correspond to the minimal BIV TAR-binding site and important nucleotides defined by mutagenesis (Chen *et al*., 1994), and the numbers indicate the critical G14:C23 base pair. The BIV Tat peptide sequence used in the NMR structure is shown on the bottom left, with important amino acids defined by mutagenesis (Chen *et al*., 1995) highlighted in bold.

In another experiment, we used existing structural information to engineer the Rev α-helix into a zinc finger scaffold in order to pre-stabilize the helix prior to RNA binding, thereby generating a high-affinity RRE-binding zinc finger (McColl *et al*., 1999). Specific RRE-binding zinc fingers have also been identified using phage display (Friesen and Darby, 1998). To extend our structure-based design studies, we now report the construction of a dimeric RNA-binding peptide based on the NMR structure of the BIV Tat–TAR complex (Puglisi *et al*., 1995; Ye *et al*., 1995).

The BIV Tat peptide is unstructured in the absence of RNA but adopts a β-hairpin conformation upon binding in the major groove of BIV TAR adjacent to a bulge (see Figure 1). Biochemical and mutagenesis experiments identified eight amino acids within the 14 amino acid arginine-rich domain (three arginines, three glycines, one threonine and one isoleucine) and two G:C base pairs, an additional G nucleotide and one bulge nucleotide in BIV TAR that are essential for binding (Chen and Frankel, 1994, 1995) (see Figure 1). The NMR structure of the complex is well defined, and contacts between all the previously defined essential amino acids and bases have been observed (Puglisi *et al*., 1995; Ye *et al*., 1995). Here we have used computer modeling to generate a plausible binding arrangement in which two covalently linked β-hairpin peptides could be bound to two adjacent TAR sites, and present evidence for the formation of a high-affinity dimeric complex. The strategy of tethering multiple binding domains together has been used to generate DNA–protein interactions with high affinities and specificities using a variety of structural motifs (for examples see Percipalle *et al*., 1995; Robinson and Sauer, 1996; Kim *et al*., 1997; Jana *et al*., 1998; Kim and Pabo, 1998; Pomerantz *et al*., 1998; Kohler *et al*., 1999; Zondlo and Schepartz, 1999; Beerli *et al*., 2000), and we show that similar design principles can be applied to RNA–protein interactions. In addition, we show that accessibility of the RNA major groove *in vivo*, as monitored by peptide binding, correlates with previous *in vitro* studies showing how bulged nucleotides or non-Watson–Crick base pairs can widen the major groove of A-form RNA helices and thereby enhance accessibility.

Results

Design of a dimeric RNA–peptide complex

The BIV Tat–TAR interaction is highly specific and forms with subnanomolar affinity (Chen *et al*., 1994). Based on the NMR structure of the complex (Puglisi *et al*., 1995; Ye *et al*., 1995), it seemed plausible that a dimeric complex could be designed that, in principle, would interact with even higher affinity and specificity. As a first step, we wished to determine an appropriate orientation and spacing for two TAR sites that would place the termini of the two bound β-hairpin peptides in closest proximity for bridging with a short peptide linker, reasoning that the shortest possible linker would maximally increase the effective concentration of the second RNA-binding domain after the first is bound, thereby lowering the entropic cost of binding and maximally enhancing affinity (see Kim and Pabo, 1998). The minimal peptide-binding site in BIV TAR (see Figure 1) is referred to as a 'half-site' in the context of the dimeric arrangements, by analogy to DNA half-sites recognized by dimeric DNA-binding proteins (see Rastinejad *et al*., 1995). Three half-site orientations are possible: 'head-to-head', 'head-to-tail' and 'tail-to-tail'. Given the β-hairpin structure of the peptide and its binding orientation, it was apparent that a head-to-head dimeric arrangement would place the peptide termini in closest possible proximity, with the N-terminus of one peptide located near the C-terminus of the second (see Figure 1). Three spacings of the RNA half-sites seemed reasonable: D0, in which the two sites were placed directly adjacent to each other; D1, in which 1 bp separates the sites; and D2, in which 2 bp separate the sites (Figure 1). Residues 70–79 of the BIV Tat peptide contain all the determinants required for RNA-binding specificity (Chen *et al*., 1995; data not shown) and, in the three head-to-head models, the distances between the C_α of Arg70 and the C_α of Ile79, which would be bridged by a linker, are 15.2 Å for D0, 17.7 Å for D1 and 19.9 Å for D2. No steric clashes were seen between the two minimal peptide units (residues

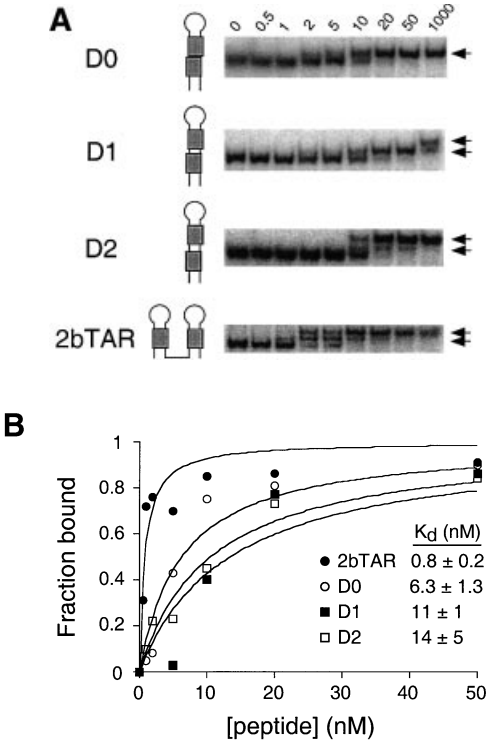

Fig. 2. Binding of the BIV $Tat_{(65–81)}$ peptide to dimeric BIV TAR RNAs *in vitro*. (**A**) Gel shift assays were performed with 0.02 nM RNA at the peptide concentrations indicated (nM). The 2bTAR RNA consists of two TAR hairpins in which the two lower stems have different sequences and the hairpins are separated by five uracils to help minimize the propensity to form alternative secondary structures. (**B**) Quantitation of the binding data derived from (A), fit to standard binding isotherms. The fraction of bound RNAs was estimated by measuring the disappearance of the unbound band, as described in Materials and methods.

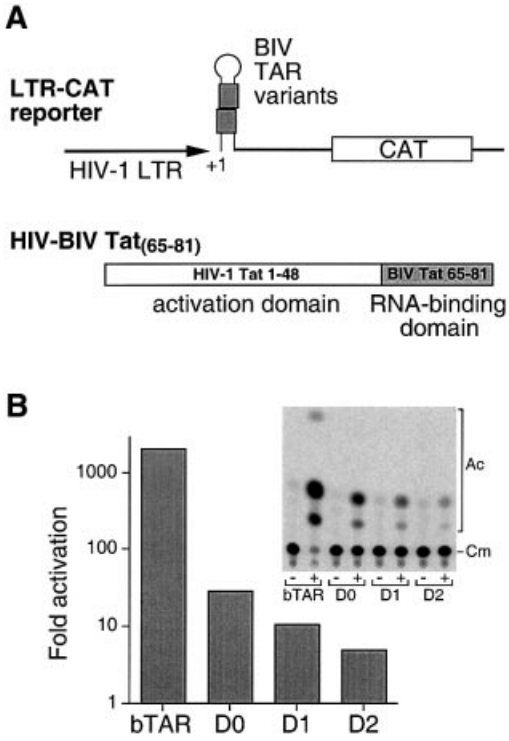

Fig. 3. Activation of HIV LTR–CAT reporters containing dimeric BIV TAR RNA sites by the HIV $Tat_{(1–48)}$–BIV $Tat_{(65–81)}$ fusion protein. (**A**) The HIV-1 LTR–CAT reporter was constructed with various RNAs in place of HIV-1 TAR located at the 5′ end of the transcript (+1). The reporter RNAs differ slightly from those shown in Figure 1 in that three additional base pairs were inserted into the lower stems to ensure stable hairpin formation *in vivo*. A schematic of the HIV–BIV Tat fusion protein containing the HIV-1 Tat activation domain fused to the BIV Tat RNA-binding domain is shown. (**B**) CAT assays with the HIV–BIV Tat fusion protein and various BIV TAR reporters. HeLa cells were co-transfected with 10 ng of the Tat expressor plasmid and 50 ng of each reporter plasmid, and CAT activity was measured after 44 h. The inset shows the raw CAT assay data, with unreacted chloramphenicol (Cm) and acetylated forms of chloramphenicol (Ac) indicated. CAT activity with the bTAR reporter is beyond the linear range of the assay and was repeated with an appropriate amount of extract for quantitation. Fold activation is the level of activity with Tat (+) normalized to the activity of each reporter plasmid alone (–).

70–79), but some side chain clashes near the N-termini seemed possible if the peptide ends were extended (see below). Thus, the D0 TAR arrangement seemed the best choice to accommodate two minimal BIV Tat peptides bridged by the shortest linker. However, we performed some initial experiments with all three dimeric RNAs (D0, D1 and D2) to allow for imprecisions in the NMR structure and dimer site modeling.

Characterization of peptide-binding sites on the dimeric RNAs

To test whether the binding sites would be functional in the dimeric RNA context, we first examined how well a monomeric BIV $Tat_{(65–81)}$ peptide bound the D0, D1 and D2 RNAs using gel shift assays *in vitro*. An RNA containing two tandem TAR hairpins designed to fold independently (2bTAR) was used as a control. As expected, two bound species were observed with 2bTAR, consistent with non-cooperative formation of one- and two-peptide complexes, but only a single peptide complex was observed with D0 (Figure 2A). Two peptides bound D1 only at high concentrations but readily bound D2 in which the two half-sites were spaced by 2 bp. Nearly identical results were obtained with a slightly shorter version of the peptide (residues 68–81; data not shown). The results are consistent with the modeling, suggesting that the N-termini of the two BIV $Tat_{(68–81)}$ peptides are in close proximity in the D0 and D1 complexes, with some potential for side chain clashes. All three dimeric RNAs bound the BIV Tat peptide with lower affinities (~7- to 17-fold) than did 2bTAR (Figure 2B), suggesting that some aspect of the RNA structures was not optimal for binding, as addressed below.

We next examined binding of the BIV $Tat_{(65–81)}$ peptide to the dimeric RNAs *in vivo* using an HIV long terminal repeat (LTR)–CAT reporter system in which the RNA-binding sites were engineered into the LTR in place of HIV TAR, and the BIV $Tat_{(65–81)}$ peptide was fused to the activation domain of HIV Tat (Figure 3A) (Chen *et al*., 1994). In this system, RNA binding results in transcriptional activation and CAT expression. All three dimeric reporters showed much lower activation than observed with a monomeric TAR hairpin, with activity increasingly reduced as additional base pairs were added between the two binding sites (Figure 3B). Thus, the dimeric reporters in general

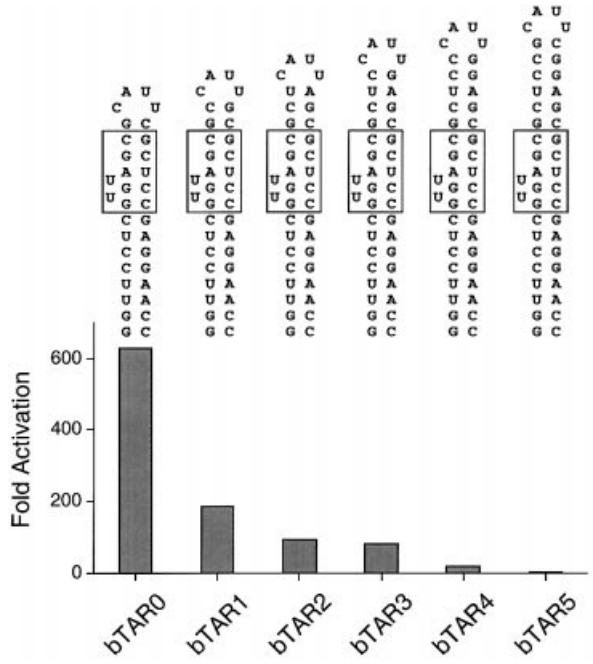

Fig. 4. Activation of HIV LTR–CAT reporters containing RNA sites with extended upper stems by the HIV $Tat_{(1–48)}$–BIV $Tat_{(65–81)}$ fusion protein. Activity was determined as in Figure 3B.

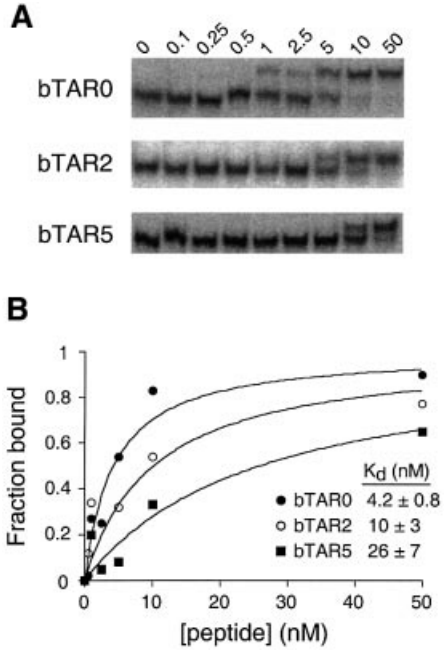

Fig. 5. Binding of the BIV $Tat_{(65–81)}$ peptide to BIV TAR RNAs with extended upper stems *in vitro*. (**A**) Gel shift assays were performed with 0.02 nM RNA at the peptide concentrations indicated (nM). (**B**) Quantitation of the binding data in (A), as described in Figure 2. Similar apparent dissociation constants were obtained at 4 and 25°C (data not shown). The RNAs used for the *in vitro* binding experiments were 3 bp shorter in the lower stem than those used in the *in vivo* experiments shown in Figure 4.

appear to create suboptimal binding sites for the monomeric BIV Tat peptide.

Accessibility of the BIV TAR major groove

We wished to understand why the dimeric RNAs bound the BIV Tat peptide so poorly and suspected that the problem might be related to the addition of contiguous base pairs adjacent to the binding site. It is known that the deep and narrow major groove of an A-form RNA helix is relatively inaccessible for protein binding unless the helix is interrupted by discontinuities such as bulges or non-Watson–Crick pairs, or is located adjacent to a loop (Weeks and Crothers, 1993). Single nucleotide bulges, such as those found in BIV TAR, only marginally widen the major groove, and we presumed that close proximity of the binding site to the loop might be important to create a groove sufficiently wide for binding. To test this hypothesis, we constructed a series of reporters containing 1–5 additional base pairs in the upper stem of a monomeric TAR site, expected to decrease groove accessibility incrementally (Figure 4), and measured activation by the HIV–BIV $Tat_{(65–81)}$ fusion protein. The sequence of base pairs added corresponds to that in the D0 RNA to allow a rough comparison between the monomer and dimer structures. As shown in Figure 4, activation decreased as an increasing number of pairs was added, reaching a 200-fold decrease when five additional pairs were added. To examine whether peptide binding *in vitro* also correlated with the postulated differences in groove accessibility, we performed gel shift assays with bTAR0, bTAR2 and bTAR5, containing zero, two or five additional base pairs between the binding site and the loop, respectively (Figure 5A). We observed a 2- to 6-fold decrease in affinity as additional pairs were added (Figure 5B), further suggesting that the major groove becomes less accessible to peptide binding as the upper helix is lengthened. As observed with the dimeric RNAs described above, considerably larger decreases were observed by the *in vivo* activation assays than the gel shift assays, perhaps reflecting less hindrance to binding of the short peptide than the fusion protein, which contains the tethered Tat activation domain, or reflecting interference of binding of other cellular proteins such as cyclin T1 (Wei *et al*., 1998). The difference in magnitude is unlikely to reflect the different temperatures of the two assays (4°C *in vitro* versus 37°C *in vivo*) as binding assays performed at higher temperatures produced similar binding constants (data not shown).

To provide further evidence that the reduced activation observed *in vivo* probably results from decreased major groove accessibility, we designed a series of reporters with 'reopened' RNA structures and monitored activation by the HIV–BIV $Tat_{(65–81)}$ fusion protein. A previous *in vitro* study used diethylpyrocarbonate reactivity to monitor accessibility of RNA helices containing a systematic series of unpaired nucleotides (Weeks and Crothers, 1993). We engineered a similar set of bulges, asymmetric loops and symmetric loops into the upper stem of bTAR5, as well as two additional RNAs with bulges in the lower stem, to test the accessibility hypothesis (Figure 6). In general, bulges or asymmetric loops containing two or more unpaired nucleotides in the upper stem restored activity substantially, correlating well with the major groove accessibilities previously observed *in vitro* (Weeks and Crothers, 1993). A two-nucleotide bulge introduced into the lower stem had little effect, whether or not the upper stem contained a bulge (Figure 6). We attempted to engineer

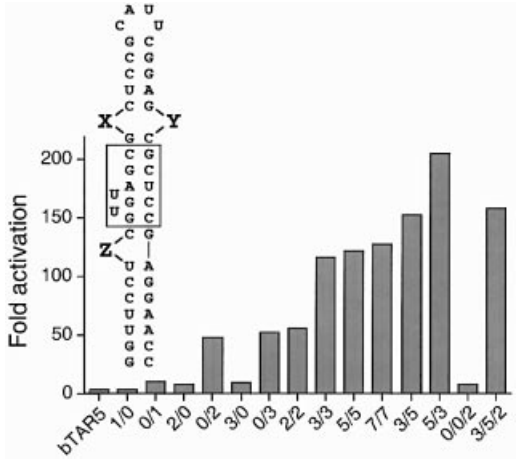

Fig. 6. Activation of HIV LTR–CAT reporters containing unpaired nucleotides in the extended BIV TAR stems by the HIV $Tat_{(1–48)}$–BIV $Tat_{(65–81)}$ fusion protein. The fully paired bTAR5 RNA contains 9 bp in the upper stem. Variable numbers of uridines were inserted at the 'X' and 'Y' positions, and adenines were inserted at the 'Z' position in two of the RNAs. RNAs are designated (X/Y/Z) by the number of nucleotides inserted at each position. Experiments were performed as described in Figure 3.

different nucleotides and additional types of bulges into the lower stem; however, the surrounding sequence led to many alternative base pairing possibilities and precluded a systematic analysis. The best configuration in the upper stem contained an asymmetric loop with three nucleotides on one strand and five on the other, and increased activation >75-fold compared with the fully paired bTAR5 (Figure 6). Thus, adding base pairs to the upper stem of BIV TAR substantially decreases BIV Tat peptide binding, whereas introducing unpaired nucleotides in configurations known to enhance major groove accessibility restores binding. Although we have not measured groove accessibility directly, and other possible differences in helical structure undoubtedly can affect peptide binding, the strong correlation between our *in vivo* binding results and the published *in vitro* results (Weeks and Crothers, 1993) suggests that inaccessibility of the major groove is at least one probable factor contributing to poor binding by the dimeric RNAs.

Design of BIV Tat peptide dimers

As described in the modeling section above, the ends of the two peptide β-hairpins are predicted to be in closest proximity in the D0 arrangement and thus D0 RNA appeared to be the best candidate for binding a dimeric peptide. Of the dimeric RNAs examined, D0 also showed the highest affinity for the monomeric BIV Tat peptide *in vitro*, was able to bind two N-terminally truncated peptides (residues 70–81; data not shown) simultaneously, and had the highest reporter activity and presumably most accessible major groove of the dimeric RNAs tested *in vivo*. To design an appropriate peptide dimer, we first considered the distance between the α carbons of Arg70 and Ile79 in the D0 complex (15.2 Å) and, assuming an extended peptide chain of 3.5 Å/residue, we reasoned that a six-amino-acid linker would be short enough to gain much of the entropic benefit of tethering yet still allow some degree of flexibility. We initially used six glycine residues [$BIV_{dimer(G6)}$; see Figure 7] to help maximize flexibility and minimize possible side chain steric clashes. The charge of a dimeric peptide containing two BIV Tat 70–79 units and an uncharged linker (see Figure 7) was only slightly greater than that of the BIV $Tat_{(65–81)}$ monomer used as a control (+10 versus +8) and thus was expected to enhance non-specific binding modestly. We also designed another dimer in which the sequence of the first BIV Tat RNA-binding domain was extended with the natural sequence found in BIV Tat (BIV_{dimer}; see Figure 7). This second dimer was used for *in vivo* experiments as well as *in vitro* experiments, as described below.

Dimeric RNA–peptide binding in vitro

In principle, forming a covalent dimer could enhance specific RNA-binding affinity by as much as the square of the binding constants (or the sum of the free energies) if the designed dimer was rigid and perfectly oriented. For DNA-binding proteins, covalent flexible linkage of binding domains has resulted in affinities generally enhanced by two or three orders of magnitude (for examples see Percipalle *et al*., 1995; Robinson *et al*., 1996; Kim *et al*., 1997; Jana *et al*., 1998; Kim and Pabo, 1998; Pomerantz *et al*., 1998; Kohler *et al*., 1999; Zondlo *et al*., 1999; Beerli *et al*., 2000). In our case, the designed $BIV_{dimer(G6)}$ and BIV_{dimer} peptides bound the D0 dimeric RNA *in vitro* with apparent K_d values of ~0.5 nM compared with 6 nM for the monomeric BIV $Tat_{(65–81)}$ peptide (Figure 7). In contrast, the dimeric peptides bound to the 2bTAR dimeric RNA composed of two independent hairpins with ~5 nM affinities, even weaker than the binding of the monomeric BIV $Tat_{(65–81)}$ peptide.

To test whether both half-sites are required for high-affinity binding to the dimeric RNA hairpin and to assess specificity, we generated mutant D0 RNAs in which an essential G14:C23 base pair (see Figure 1; Chen *et al*., 1994) was changed to C:G in one or both of the sites. Introducing this mutation into either single site reduced the binding affinity of both the $BIV_{dimer(G6)}$ and BIV_{dimer} peptides ~10-fold in each case, consistent with binding to a single site, whereas mutation of both sites abolished binding (the data in Figure 7 show K_d values >50 nM and data not shown indicate K_d values >1 μM), demonstrating that binding is highly specific. The binding affinity of the BIV $Tat_{(65–81)}$ peptide for each single site mutant decreased ~2-fold compared with D0, perhaps reflecting the entropic loss of removing one of two possible binding sites and/or communication between the half-sites upon peptide binding. The binding affinities of the dimeric peptides for the single site mutants remain 2- to 5-fold higher than for the monomeric peptide, perhaps reflecting some degree of non-specific binding by the dimeric peptides. The requirement for both half-sites to achieve high-affinity binding of the peptide dimers supports the proposed dimeric binding mode.

Dimeric RNA–peptide binding in vivo

Given the high affinity of the dimeric interaction observed *in vitro*, we next wished to compare activation of dimeric

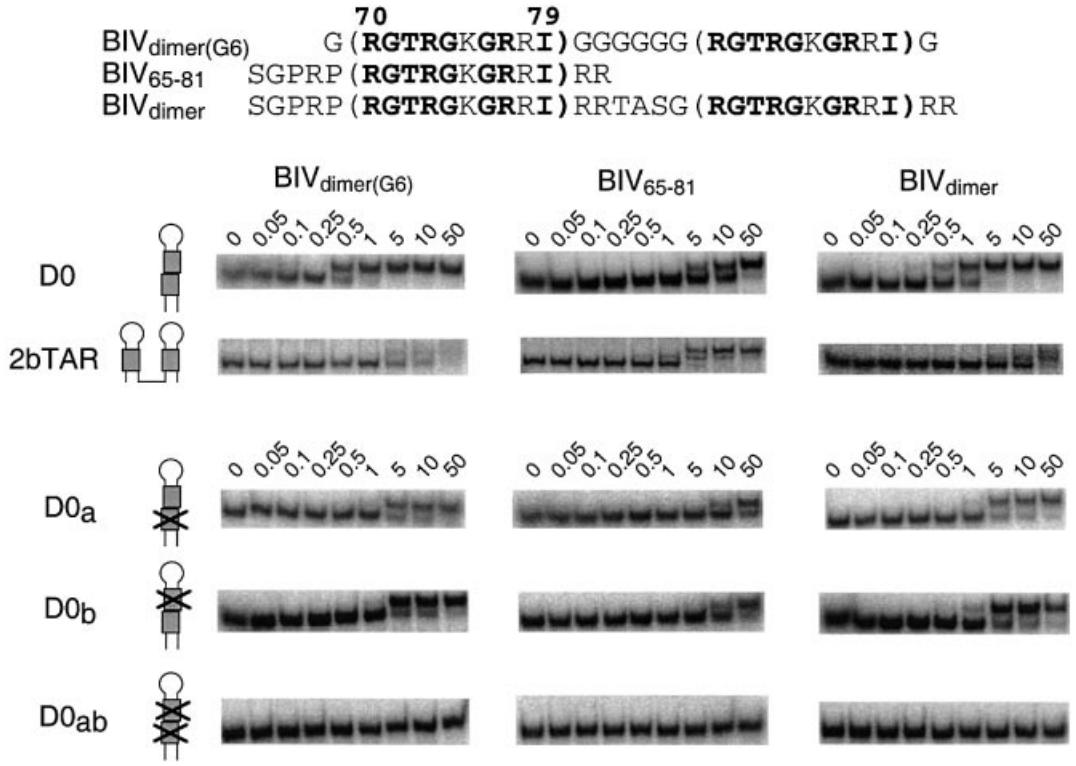

Fig. 7. Binding of the BIV $Tat_{(65–81)}$, $BIV_{dimer(G6)}$ and BIV_{dimer} peptides to dimeric TAR RNAs and RNA mutants *in vitro*. Gel shift assays were performed with 0.02 nM RNA at the peptide concentrations indicated (nM). $D0_a$ contains the G14:C23 to C:G mutation (see Figure 1; Chen *et al*., 1994) in the upper binding site, $D0_b$ contains the mutation in the lower site and $D0_{ab}$ contains the mutation in both sites.

and monomeric constructs *in vivo*. Initial experiments using the Tat activation domain fused to the $BIV_{dimer(G6)}$ peptide showed no activity on any of the reporter constructs tested (data not shown), including the highly activatible monomeric BIV TAR reporter. This result may possibly be explained by poor protein expression or stability. However, the second dimeric peptide, which utilizes a linker with the natural sequence of BIV Tat C-terminal to the RNA-binding domain (Figure 8), was ~20-fold more active on the D0 dimer RNA reporter than was the monomeric BIV $Tat_{(65–81)}$ peptide (Figure 8), suggesting a much higher affinity dimeric interaction. In contrast, both the monomer and dimer peptides were equally active on the 2bTAR reporter containing two independent hairpins. The overall levels of activation with the 2bTAR reporter were higher than with D0, probably due to the differences in major groove accessibility, as described above.

To examine further the specificity of the dimeric interaction and to test whether both units of the peptide dimer are needed for activity, we generated mutants in one or both halves of the peptide dimer. Introducing Arg73 to lysine mutations known to abolish specific RNA binding (Chen *et al*., 1995) into either the N- or C-terminal RNA-binding domain strongly reduced activity on the D0 dimeric RNA reporter, reducing activity to the weak level observed with the monomeric BIV $Tat_{(65–81)}$ fusion protein on the D0 reporter (Figure 8). A slight amount of residual activity was observed with the double mutant peptide, perhaps reflecting a small degree of non-specific binding. The requirement for both halves of the peptide dimer for high-affinity binding to the dimeric RNA *in vivo* is consistent with the modeled dimeric interaction.

Discussion

To our knowledge, these studies provide the first example of the design of a specific RNA-binding dimer. There are numerous examples in which enhanced DNA-binding affinity and/or extended specificity have been achieved by covalently tethering two or more DNA-binding modules (for examples see Percipalle *et al*., 1995; Robinson *et al*., 1996; Kim *et al*., 1997; Jana *et al*., 1998; Kim and Pabo, 1998; Pomerantz *et al*., 1998; Kohler *et al*., 1999; Zondlo *et al*., 1999; Beerli *et al*., 2000), in some cases by creating a covalent linkage between two subunits of a normally dimeric protein. In simplistic terms, the DNA recognition surface in these cases may be thought of as a linear DNA helix, with the flexible covalent linker between protein modules used to generate a high local concentration of the second domain upon binding of the first. Many dimeric DNA-binding proteins, however, achieve recognition of their dimeric sites not simply by a tethering mechanism but also by forming highly ordered dimerization interfaces that are structurally matched to the arrangements of the DNA half-sites (Rastinejad *et al*., 1995). This latter situation represents a better analogy to the designed BIV Tat–TAR dimer in which the orientation of half-sites appears critical; the high-affinity interaction occurs only with the D0 RNA dimer and not when two independently

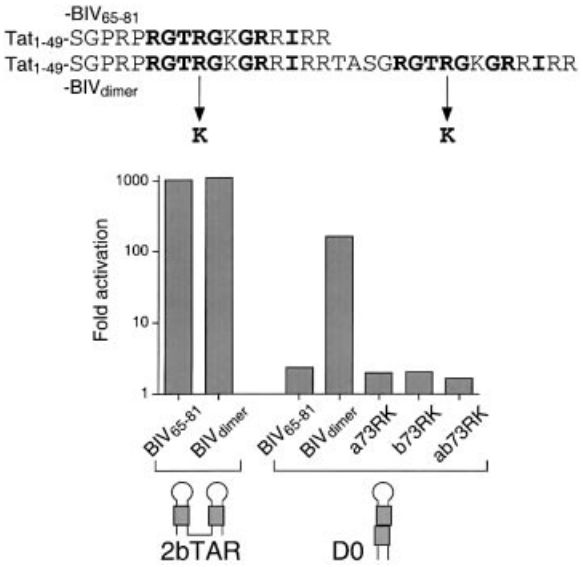

Fig. 8. Activation of HIV LTR–CAT reporters containing dimeric BIV TAR sites by Tat fusion proteins containing the activation domain of HIV Tat (residues 1–49) fused to monomeric or dimeric BIV Tat peptides. The sequences of the BIV Tat peptides and positions of Arg73 to lysine mutations are shown. Activation using the dimeric reporters indicated below was measured as described in Figure 3, except that 3 ng of each Tat expression plasmid were used in these transfections.

folded TAR hairpins are tethered. For some DNA-binding proteins, such as the nuclear receptors, the dimerization interface becomes ordered only after interacting with DNA, leading to cooperative binding to the two half sites (Rastinejad *et al*., 1995). A similar case of cooperative binding has been observed with a U1A RNP domain–RNA complex in which binding of one RNP domain to an internal loop in the 3′-untranslated region of its own mRNA stabilizes a C-terminal helix, which then interacts with an adjacent RNP subunit bound at an identical loop (Varani *et al*., 2000).

Because RNA-binding sites typically contain bulges, loops and non-Watson–Crick base pairs, and are not composed of simple helices, it can be difficult to predict the spatial orientations of two half-sites. To make the design of dimeric RNA–protein interactions even more challenging, alternative RNA structures may be formed when two identical sites are juxtaposed, and problems relating to the size and shape of the flanking helical regions may arise, such as accessibility of the major groove as described here. Nonetheless, it may be possible to optimize the positioning of two RNA half-sites, perhaps including intervening bulges, as well as the nature of the protein linkers using combinatorial selection experiments to identify high-affinity sequence-specific interactions (Laird-Offringa and Belasco, 1995; Harada *et al*., 1996, 1997; Friesen *et al*., 1998; Tan and Frankel, 1998). The structures of the MS2 coat protein bound to its operator RNA and a series of RNA aptamers show that a dimeric RNA-binding protein can recognize two ‘pseudo’ half-sites in which a bulge nucleotide in one ‘half-site’ and a loop nucleotide in the other ‘half-site’ are bound in similar ways, with subtle differences in the interactions with each site (Valegard *et al*., 1994; Rowsell *et al*., 1998). Given that dimeric complexes need not be symmetric, it seems plausible that new dimeric RNA–protein interactions with desired specificities may be designed by starting with one ‘anchored’ interaction to identify tethered peptides that interact specifically with an adjacent second site, and then identifying linkers that match the orientation of two peptide units more precisely to a particular configuration of the two binding sites, such as the D0 TAR configuration described here. In our initial BIV Tat–TAR experiments, we joined the domains with very simple linkers that almost certainly can be optimized better, potentially even introducing additional interactions to the RNA. Combinatorial experiments with Arc repressor and zinc fingers have been used both to optimize the linkers between two domains and to generate new DNA-binding specificities of the appended domains (Greisman and Pabo, 1997; Robinson and Sauer, 1998). The dimeric peptides described here provide a reasonable starting point for the further design of high-affinity RNA-binding peptides that bind therapeutically important RNA sites or that might be used to deliver proteins to specific RNAs.

In the case of our designed BIV Tat dimer, we presume that the peptide is disordered in the absence of BIV TAR, like the monomeric peptide, and folds into a double β-hairpin structure upon binding (see Figure 1), although we currently have no direct evidence for the structure. This putative coupled binding and folding is unlike most tethered DNA-binding interactions in which each domain is folded largely independently, with basic region leucine zipper and basic region helix–loop–helix proteins being notable exceptions (see Ellenberger, 1994). It may be possible to enhance RNA-binding affinity and/or specificity of the BIV complex further by pre-stabilizing the β-hairpin structure prior to binding, as was done with the α-helical RNA-binding domain of Rev (Tan *et al*., 1993; McColl *et al*., 1999), provided that a conformational change is not needed to fit the BIV peptide into its RNA site. Additional structural studies of the BIV dimeric complex, identification of other tight-binding peptides from arginine-rich combinatorial libraries in which the peptide conformation might be stabilized and studies of peptides engineered with non-natural amino acids or chemical linkers that stabilize the β-hairpin may shed further light on the coupled folding problem.

One unexpected observation in our studies was the apparently poor major groove accessibility of the dimeric RNAs, particularly when assayed *in vivo*. When the two half-sites were placed in adjacent configurations, contiguous helices of 6, 7 or 8 bp were formed between the closest bulged uridines in each binding site (see Figure 1). Previous *in vitro* studies suggested that uninterrupted helices of perhaps as many as 8 bp could be accessed via the effective ‘ends’ created by adjacent bulges or loops (Weeks and Crothers, 1993), but it appears that the single bulge nucleotides in the BIV TAR configuration alone do not generate sufficient accessibility for peptide binding. The observation that the dimeric RNAs were relatively inactive provided an opportunity to use BIV Tat peptide binding as a probe for major groove accessibility *in vivo*. By engineering bulges and loops into an extended BIV TAR helix (Figure 4), we found an excellent correlation between RNA structures known to generate major groove accessibility *in vitro* and those that promote BIV Tat

peptide binding *in vivo*. It is interesting that asymmetric bulges and loops that contain extra nucleotides on the 5′ side of the peptide-binding site appear to create much more accessible grooves than when extra nucleotides are located 3′ to the site (Figure 6). BIV TAR contains two single nucleotide bulges that may already make the site partially accessible even when embedded in a long duplex. It seems reasonable that bulges introduced into the opposite strand would enhance access to the site further, with nucleotides on the 5′ side that make critical contacts to Gly71 and Thr72 of the β-hairpin peptide expected to become the most accessible (Weeks and Crothers, 1993). Thus, the BIV interaction provides a glimpse into RNA major groove accessibility *in vivo*, at least in one particular structural context.

Materials and methods

Modeling

Models for possible configurations of dimeric BIV TAR-binding sites were constructed using the Insight II graphics program (Biosym) on a Silicon Graphics workstation. Coordinates of a single BIV Tat–TAR NMR model (structure number 6; Puglisi *et al*., 1995) were used to explore various orientations of the two 'half-sites' in a dimer. This particular NMR model was chosen because it contained a well-formed U–A:U base triple, in agreement with observations by Patel and colleagues (Ye *et al*., 1995), and because it displayed most of the important RNA–peptide contacts. To model the dimeric RNAs with binding sites in three different orientations ('head-to-head', 'head-to-tail' and 'tail-to-tail'), the four-nucleotide BIV TAR hairpin loop was removed and the two RNA sites were positioned by visually overlapping the two end base pairs of each site. To model the 'head-to-head' dimeric RNAs more accurately, the orientation of 'half-sites' was approximated by generating 6 bp of an idealized A-form RNA helix and using this helix as a 'splint' to minimize the r.m.s. values of four phosphate groups (5′ to G14, C15, G22 and C23) at the ends of each of the TAR sites. To test whether any of the modeled RNA arrangements were likely to adopt alternative structures, all designs were run through the MFOLD secondary structure prediction algorithm (Zuker, 1989), and no reasonable alternative folds were found in any of the cases.

Peptide synthesis, purification and analysis

Peptides were synthesized on an Applied Biosystems model 432A peptide synthesizer using standard Fmoc chemistry and resins. The N-terminus of each peptide was acetylated using acetic anhydride, and the use of amidated resin resulted in amidated C-termini after cleavage. Peptides were purified by C_4 reverse-phase HPLC (Vydac) using an acetonitrile gradient of 0.2%/min in 0.1% trifluoroacetic acid. The molecular weights of the peptides were obtained by laser desorption mass spectrometry, and peptide concentrations were determined by quantitative amino acid analysis (University of Michigan Protein and Carbohydrate Structure Facility). Peptide purity and concentrations were confirmed using native PAGE with Coomassie Blue staining. The sequence of the BIV $Tat_{(65–81)}$ peptide is ac-SGPRPRGTRGKGRRIRR-am, the sequence of the $BIV_{dimer(G6)}$ peptide is ac-GRGTRGKGRRIGGGGGGRGTRGKGRRIG-am, and the sequence of the BIV_{dimer} peptide is ac-SGPRPRGTRGKGRRIRRTASGRGTRGKGRRIRR-am.

RNA-binding gel shift assays

Internally labeled RNAs were prepared by *in vitro* transcription using T7 polymerase and [α-^{32}P]CTP. RNA concentrations were determined by measuring incorporation of the isotope and calculating specific activities. RNAs were annealed by heating to 90°C and slow cooling to room temperature in renaturation buffer (10 mM Tris pH 7.5, 100 mM NaCl) at RNA concentrations of 5–20 nM. The binding reactions shown in the figures were carried out in 10 mM HEPES–KOH pH 7.5 buffer containing 100 mM KCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol, 10% glycerol, and 50 μg of tRNA as a competitor. Additional binding reactions (not shown) were carried out in 10 mM Tris–HCl pH 7.5, 70 mM NaCl, 0.2 mM EDTA and 25 μg/ml yeast tRNA as a competitor, and similar results were obtained. Reactions were incubated on ice for 10 min, loaded onto pre-run 20% native polyacrylamide gels (in 0.5× TBE buffer), and electrophoresed at 200 V at 4°C. For quantitation, dried gels were autoradiographed and scanned or exposed to a phosphorimaging plate and scanned with a Molecular Dynamics PhosphorImager, and free and bound RNA bands were quantified using ImageQuant software. To calculate apparent binding constants, the fraction of bound RNA was estimated by measuring the disappearance of the unbound band. This was necessary to normalize for differences among the different RNAs in the number of potential binding sites and variable numbers of complexes. For example, while 2bTAR and D2 are able to bind two monomeric BIV Tat peptides, D0 can bind only one peptide even though it contains two potential binding sites (see Results), and bTAR contains just a single binding site. Binding curves were fit to the data using Kaleidagraph software (Synergy Software, Reading, PA) and standard binding equations. For RNAs that showed singly and doubly bound species, in some cases we also calculated apparent binding constants by directly quantitating the bound bands and fitting the data to single binding isotherms, approximating the fraction of bound sites as the sum of the singly bound species and one half of the doubly bound species. Apparent binding constants estimated by both methods were in reasonable agreement, but because fitting errors were generally higher using the second method and because direct comparisons between RNAs with different numbers of binding sites are difficult, binding constant data are presented from calculations based on the disappearance of unbound RNAs.

Plasmid construction and CAT assays

HIV–BIV Tat hybrids were constructed by cloning synthetic oligonucleotide cassettes encoding each peptide directly after the codon for residue 49 of HIV-1 Tat in the pSV2tat72 expression vector (Tan *et al*., 1998), thereby replacing the C-terminal residues in HIV-1 Tat, including its own RNA-binding domain, with heterologous RNA-binding peptides. BIV TAR-related reporters were constructed by cloning oligonucleotide cassettes into an HIV-1 LTR–CAT reporter plasmid as described (Tao and Frankel, 1993). Reporter and expressor plasmids were transiently co-transfected into HeLa cells using lipofectin and Optimem media (Gibco-BRL). Total DNA amounts were adjusted to 1 μg per transfection using pUC19 plasmid DNA. Cells were harvested 44 h after transfection, and CAT activities were assayed as described (Tan *et al*., 1993) and quantified using a Molecular Dynamics PhosphorImager.

Acknowledgements

We thank Pat O'Farrell, Colin Smith and members of the Frankel laboratory for helpful suggestions, and Alan Cheng, Steve Landt, Rob Nakamura, Carl Pabo, Hadas Peled-Zehavi and Ralph Peteranderl for comments on the manuscript. This work was supported by an NIH postdoctoral fellowship (D.M.C.) and by grants from the National Institutes of Health.

References

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

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

Beerli,R.R., Dreier,B. and Barbas,C.F.,III (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl Acad. Sci. USA*, **97**, 1495–1500.

Bogerd,H.P., Wiegand,H.L., Bieniasz,P.D. and Cullen,B.R. (2000) Functional differences between human and bovine immunodeficiency virus Tat transcription factors. *J. Virol.*, **74**, 4666–4671.

Cai,Z., Gorin,A., Frederick,R., Ye,X., Hu,W., Majumdar,A., Kettani,A. and Patel,D.J. (1998) Solution structure of P22 transcriptional antitermination N peptide–box B RNA complex. *Nature Struct. Biol.*, **5**, 203–212.

Calnan,B.J., Biancalana,S., Hudson,D. and Frankel,A.D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA–protein recognition. *Genes Dev.*, **5**, 201–210.

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

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

Dervan,P.B. and Burli,R.W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.*, **3**, 688–693.

Draper,D.E. (1999) Themes in RNA–protein recognition. *J. Mol. Biol.*, **293**, 255–270.

Ellenberger,T. (1994) Getting a grip on DNA recognition: structures of the basic region leucine zipper and the basic region helix–loop–helix DNA-binding proteins. *Curr. Opin. Struct. Biol.*, **4**, 12–21.

Frankel,A.D. (2000) Fitting peptides into the RNA world. *Curr. Opin. Struct. Biol.*, **10**, 332–340.

Friesen,W.J. and Darby,M.K. (1998) Specific RNA binding proteins constructed from zinc fingers. *Nature Struct. Biol.*, **5**, 543–546.

Greisman,H.A. and Pabo,C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA sites. *Science*, **275**, 657–661.

Harada,K., Martin,S.S. and Frankel,A.D. (1996) Selection of RNA-binding peptides *in vivo*. *Nature*, **380**, 175–179.

Harada,K., Martin,S.S., Tan,R. and Frankel,A.D. (1997) Molding a peptide into an RNA site by *in vivo* peptide evolution. *Proc. Natl Acad. Sci. USA*, **94**, 11887–11892.

Hermann,T. and Patel,J.D. (1999) Stitching together RNA tertiary architectures. *J. Mol. Biol.*, **294**, 829–849.

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

Jana,R., Hazbun,T.R., Fields,J.D. and Mossing,M.C. (1998) Single-chain λ Cro repressors confirm high intrinsic dimer–DNA affinity. *Biochemistry*, **37**, 6446–6455.

Jiang,F., Gorin,A., Hu,W., Majumdar,A., Baskerville,S., Xu,W., Ellington,A. and Patel,D.J. (1999) Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples. *Structure Fold Des.*, **7**, 1461–1472.

Kim,J.-S. and Pabo,C.O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl Acad. Sci. USA*, **95**, 2812–2817.

Kim,J.S., Kim,J., Cepek,K.L., Sharp,P.A. and Pabo,C.O. (1997) Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression. *Proc. Natl Acad. Sci. USA*, **94**, 3616–3620.

Kohler,J.J., Metallo,S.J., Schneider,T.L. and Schepartz,A. (1999) DNA specificity enhanced by sequential binding of protein monomers. *Proc. Natl Acad. Sci. USA*, **96**, 11735–11739.

Laird-Offringa,I.A. and Belasco,J.G. (1995) Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc. Natl Acad. Sci. USA*, **92**, 11859–11863.

Legault,P., Li,J., Mogridge,J., Kay,L.E. and Greenblatt,J. (1998) NMR structure of the bacteriophage λ N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell*, **93**, 289–299.

McColl,D.J., Honchell,C.D. and Frankel,A.D. (1999) Structure-based design of an RNA-binding zinc finger. *Proc. Natl Acad. Sci. USA*, **96**, 9521–9526.

Percipalle,P., Simoncsits,A., Zakhariev,S., Guarnaccia,C., Sanchez,R. and Pongor,S. (1995) Rationally designed helix–turn–helix proteins and their conformational changes upon DNA binding. *EMBO J.*, **14**, 3200–3205.

Pomerantz,J.L., Wolfe,S.A. and Pabo,C.O. (1998) Structure-based design of a dimeric zinc finger protein. *Biochemistry*, **37**, 965–970.

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

Rastinejad,F., Perlmann,T., Evans,R.M. and Sigler,P.B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature*, **375**, 203–211.

Robinson,C.R. and Sauer,R.T. (1996) Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA. *Biochemistry*, **35**, 109–116.

Robinson,C.R. and Sauer,R.T. (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc. Natl Acad. Sci. USA*, **95**, 5929–5934.

Rowsell,S., Stonehouse,N.J., Convery,M.A., Adams,C.J., Ellington,A.D., Hirao,I., Peabody,D.S., Stockley,P.G. and Phillips,S.E.V. (1998) Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nature Struct. Biol.*, **5**, 970–975.

Segal,D.J. and Barbas,C.F.,III (2000) Design of novel sequence-specific DNA-binding proteins. *Curr. Opin. Chem. Biol.*, **4**, 34–39.

Smith,C.A., Crotty,S., Harada,Y. and Frankel,A.D. (1998) Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry*, **37**, 10808–10814.

Smith,C.A., Calabro,V. and Frankel,A.D. (2000) An RNA-binding chameleon. *Mol. Cell*, **6**, 1067–1076.

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

Tan,R. and Frankel,A.D. (1998) A novel glutamine–RNA interaction identified by screening libraries in mammalian cells. *Proc. Natl Acad. Sci. USA*, **95**, 4247–4252.

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

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

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

Varani,L., Gunderson,S.I., Mattaj,I.W., Kay,L.E., Neuhaus,D. and Varani,G. (2000) 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. *Nature Struct. Biol.*, **7**, 329–335.

Walter,F., Vicens,Q. and Westhof,E. (1999) Aminoglycoside–RNA interactions. *Curr. Opin. Chem. Biol.*, **3**, 694–704.

Weeks,K.M. and Crothers,D.M. (1993) Major groove accessibility of RNA. *Science*, **261**, 1574–1577.

Wei,P., Garber,M.E., Fang,S.M., Fischer,W.H. and Jones,K.A. (1998) 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*, **92**, 451–62.

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

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

Ye,X., Gorin,A., Frederick,R., Hu,W., Majumdar,A., Xu,W., McLendon,G., Ellington,A. and Patel,D.J. (1999) RNA architecture dictates the conformations of a bound peptide. *Chem. Biol.*, **6**, 657–669.

Zondlo,N.J. and Schepartz,A. (1999) Highly specific DNA recognition by a designed miniature protein. *J. Am. Chem. Soc.*, **121**, 6938–6939.

Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.

Received July 7, 2000; revised November 7, 2000;
accepted November 13, 2000