New features in RNA recognition: a Tat–TAR complex

The structure of a peptide from the BIV Tat protein bound to its cognate TAR RNA has been solved. This structure reveals a β-hairpin motif for the protein, bound at a widened bulge site in the major groove of the RNA.

Chemistry & Biology January 1996, **3**:17–19

The recognition of RNA by proteins is critical to many functions in cells, from the regulation of gene expression to the attachment of the correct amino acids during protein synthesis. The mechanics for recognition of RNA are much less well understood, however, than the recognition of duplex DNA by proteins. Many of the small number of RNA–protein complex structures available are tRNA–synthetase complexes. While these are valuable examples, it is clear that there is still much to learn about the way RNA can be recognized by proteins. A fascinating new example that has recently been reported is the basic region of the bovine immunodeficiency virus (BIV) Tat (*trans*-activating) protein bound to the TAR (*trans*-activation response element) RNA [1,2]. There are a number of novel features in this structure, and a few that are reminiscent of some previously seen protein–nucleic acid interactions. A particularly interesting feature is that the binding seems to be adaptive from both sides; probably neither the protein nor the RNA has the 'correct' structure before their interaction.

The importance of these structures goes beyond the implications for RNA–protein recognition. The Tat–TAR interaction is essential for the life cycle of all lentiviruses, a group of viruses that includes the major human pathogen HIV (human immunodeficiency virus). In all such viruses, a Tat protein regulates RNA production by binding to the TAR element within the 5' long terminal repeat of the mRNA of the virus. The presence of Tat stimulates transcriptional elongation of the mRNA by relieving a transcriptional block due to a stalled RNA polymerase. This action of Tat is apparently necessary for viral replication. The process governed by Tat is therefore a potential target for drugs, and, with the structural information now available, one can imagine rational drug design approaches aimed at blocking the Tat–TAR interaction in order to treat AIDS (acquired immune deficiency syndrome).

Previous structural studies

There have been a number of previous studies of the Tat–TAR systems from lentiviruses that lead to immunodeficiency diseases, including those of HIV, BIV and EIAV (equine infectious anemia virus). Although these are related viruses and all have a Tat–TAR system with similar function, there seem to be substantial differences in the way in which they work. Biochemical studies of HIV Tat have shown that one arginine residue is very important and essential for specific binding [3]. The HIV TAR RNA contains a hairpin loop with a three-base bulge in the stem. NMR studies have shown that the RNA changes conformation in response to peptide binding, and that an arginine amide induces a similar change, although a higher concentration is required than for the peptide [4]. A structural model was developed in which the guanidine portion of the arginine is bound at the three-base bulge, contacting two phosphates and the edge of a base. In addition, there was evidence for the formation of a base triple adjacent to the bulge site [4]. Recently a more detailed study was carried out using peptides [5], which supports the general site and mode of binding of the arginine, but disputes the formation of the base triple. Unfortunately in this study the peptide resonances were not assigned, and hence the conformation of the peptide was not addressed in any detail.

There have also been several NMR studies on Tat protein constructs. The data for the HIV [6] and EIAV [7] proteins suggested that these proteins are rather flexible, with some tendency to form α-helix, although this tendency is fairly weak in aqueous solutions. Thus studies of the protein alone have contributed little to understanding the mechanism by which recognition is achieved.

BIV Tat–TAR structures

A new but related system, from bovine immmunodeficiency virus, has now yielded more detailed information on the Tat–TAR complex. The structural work had also been preceded by biochemical studies [8], which showed that the RNA target for the Tat protein is again a bulge-containing hairpin, but with a different type of bulge from that seen in the TAR of HIV (Fig. 1). The better behavior of the complex of a 17-residue peptide from BIV Tat with a 28-nucleotide RNA has allowed detailed structural characterization by NMR of both components by two groups [1,2]. The two labs agree on the basic structure, though the report from Ye, Kumar and Patel provides considerably more detail due to more extensive assignments and identification of intermolecular contacts.

While previous studies had hinted at the formation of helices in the peptides that bind to the TAR RNA, the new structures show very clearly that the peptide forms a β-hairpin structure, which is inserted into the major groove of the RNA specifically at the site of the bulge. The major groove of normal A-form RNA is too narrow

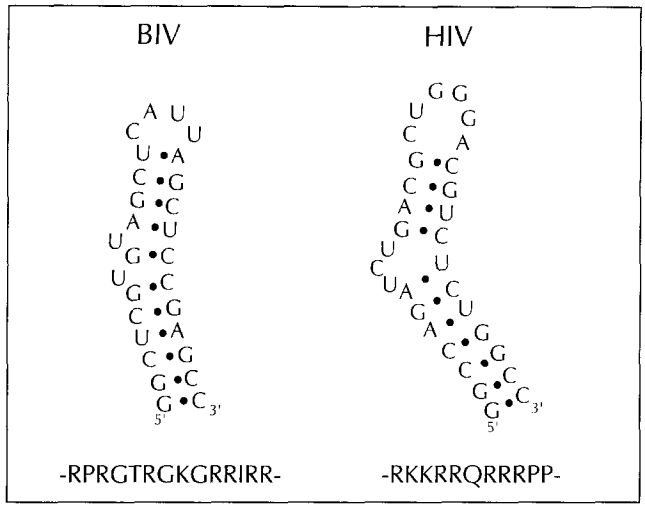

Fig. 1. The sequences of the BIV and HIV TAR RNA hairpin loops; the sequence of the basic region of the corresponding Tat proteins are given below.

to accommodate a peptide. The bulge bases widen the groove at the recognition site, however, and also fold back to form a base triple which probably helps to stabilize the complex. Thus this system provides a new example of adaptive recognition, in which both the peptide and the RNA seem to change their structures substantially upon binding.

Specific interactions

There are many specific interactions observed between the peptide and RNA [2], analogous to features seen in a variety of different protein–DNA complexes that have been solved over the past few years. The β-turn and sheet parts of the peptide are inserted deeply into the RNA groove, making extensive contacts with both the top and bottom surfaces of the groove (Fig. 2). As with HIV Tat, there are arginine residues that are crucial. The R73 sidechain is positioned near the G11•C25 base pair, and this residue hydrogen bonds both to the edge of G11 and to the phosphate of the 9–10 step, reminiscent of the arginine fork described in the HIV system [3], and also the interaction of zinc fingers with guanines [9]. The hydrophobic sidechain of I79 packs up against the C5–C6 edge of U10, part of the base triple U10•A13•U24 (not specifically identified by Puglisi *et al.* [1], although the U10 base is found in approximately the same position in the major groove in their structure). The shape of the sidechain fits well, and indeed substitution of this residue reduces the binding affinity substantially. The importance of hydrophobic packing of Ile has also been seen in the structure of SRY bound to DNA [10]. Other key hydrogen bonds are formed from the hydroxyl group of T72 to a phosphate, and from the backbone carbonyl of T72 to the amino group of a cytosine (only reported in [2]). The sidechains of R70 and R77 again make specific hydrogen bonds to guanines at positions 9 and 14. Replacement of these arginine residues with lysine drops the affinity dramatically. The structure also suggests why the three glycine residues in

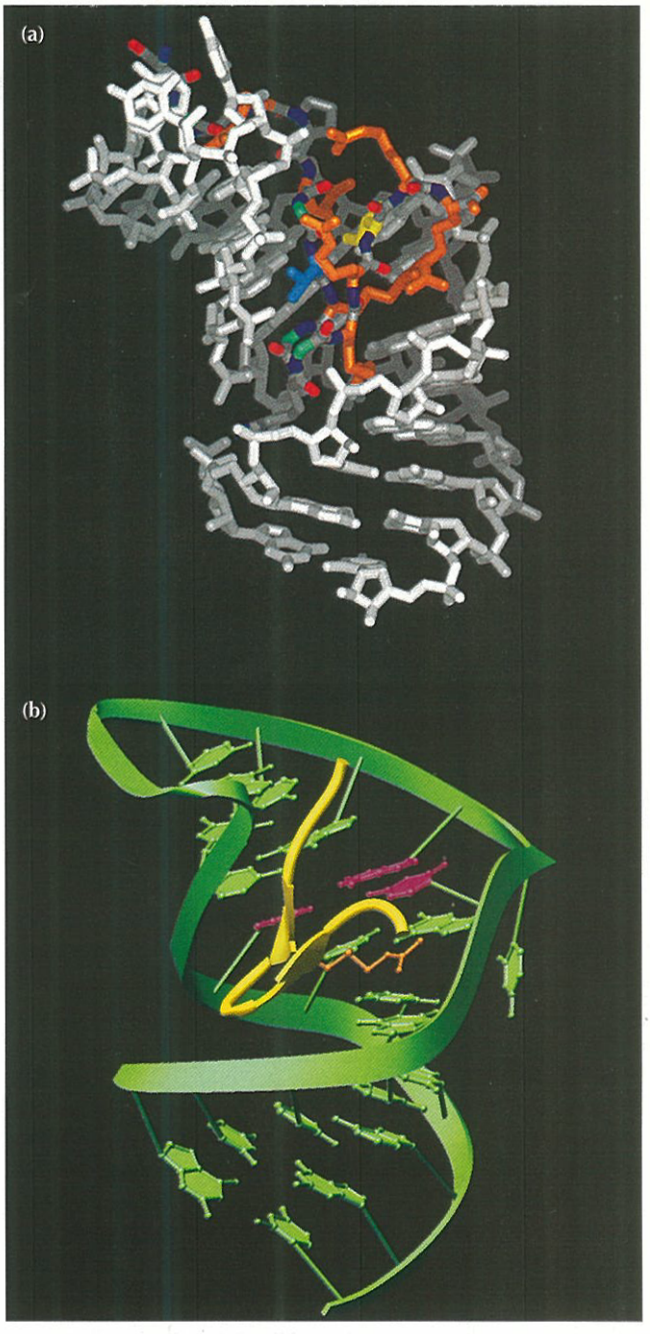

Fig. 2. The structure of the BIV Tat–TAR complex, derived from [2]. **(a)** A stick representation. The RNA is shown in gray, with the peptide atoms shown in color. Residues that are important for sequence-specific binding and folding are color-coded: the Arg sidechains are orange, Gly α-carbons green, Thr sidechain light blue, and Ile sidechain yellow. The backbone atoms are colored according to atom type: carbonyl carbon gray, carbonyl oxygen red, and amide nitrogen dark blue. The amino terminus of the peptide comes from the top/back of the RNA hairpin, extending down the groove. The β-hairpin can be identified by the two Gly residues, followed by the important Arg-Arg-Ile-Arg-Arg sequence at the carboxy terminus of the peptide. This figure was made from coordinates provided by the authors of [2]; the author is grateful for access to them. **(b)** A ribbon representation of the complex. The Tat peptide folded in a β-hairpin is shown in yellow with arrows indicating the β-sheet segment of the polypeptide chain. The side chain of R73 is shown in orange. The RNA is shown in green with the bases of residues U10, A13 and U24, which form the base triple, highlighted in magenta. Reprinted with permission from [2].

the peptide are important. G71 and G74 point into the RNA groove in places where the presence of a sidechain could not be accommodated without substantial steric clash, and G74 and G76 have positive ϕ values which would be energetically unfavorable for other residues. There are other contacts between the peptide and RNA, although those cited above are probably the major determinants for sequence-specific binding. The presence of the RNA bulge is clearly important, allowing the RNA to alter its conformation locally to adapt to the peptide. The amino terminus of the peptide extends over the loop in an irregular fashion, but these contacts probably also contribute to the binding.

Some of the features seen in these Tat–TAR systems are likely to occur in other systems as well. Another peptide–RNA system from HIV, the interaction of Rev with the RRE (Rev responsive element) mRNA, is also being studied by NMR. The RNA hairpin loop contains an internal loop, primarily of purines, which is structured despite the lack of normal base pairs [11]. It has been shown that the mismatches lead to a widened groove, allowing for greater penetration of the peptide, and that the RNA changes conformation upon interaction with the peptide [12]. The assignments of the peptide and RNA bound to each other are not yet complete, but one can anticipate that a detailed structure of the Rev–RRE complex should also appear in the not too distant future.

These first glimpses into the interactions of unstructured regulatory Tat proteins with RNA targets indicate a new level of complexity — adaptive binding by both protein and RNA. There will surely be many more new variations on the themes observed to date for protein–RNA interaction. Solution NMR is clearly establishing itself as a powerful tool for characterization of these structures; in the future crystallography may also contribute to an understanding of the details of the molecular recognition process in these systems.

References

1. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat–TAR peptide RNA complex. *Science* **270**, 1200–1203.
2. Ye, X., Kumar, R.A. & Patel, D.J. (1995). Molecular recognition in bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chemistry & Biology* **2**, 827–840.
3. Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A.D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science* **252**, 1167–1171.
4. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. & Williamson, J.R. (1992). Conformation of the TAR RNA–arginine complex by NMR spectroscopy. *Science* **257**, 76–80.
5. Aboul-ela, F., Karn, J. & Varani, G. (1995). The structure of human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313–332.
6. Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R. & Rösch, P. (1995). Structural studies of HIV-1 Tat protein. *J Mol. Biol.* **247**, 529–535.
7. Willbold, D., *et al.*, & Rösch, P. (1993). Sequence-specific resonance assignments of the ^{1}H-NMR spectra of a synthetic, biologically active EIAV Tat protein. *Biochemistry* **32**, 8439–8445.
8. Chen, L. & Frankel, A.D. (1994). An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure. *Biochemistry* **33**, 2708–2715.
9. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* **252**, 809–817.
10. Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY–DNA complex. *Cell* **81**, 705–714.
11. Peterson, R.D., Bartel, D.P., Szostak, J.W., Horvath, S.J. & Feigon, J. (1994). ^{1}H-NMR studies of the high-affinity Rev binding site of the Rev responsive element of HIV-1 mRNA: base pairing in the core binding element. *Biochemistry* **33**, 5357–5366.
12. Battiste, J.L., Tan, R., Frankel, A.D. & Williamson J.R. (1994). Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine–purine base pairs. *Biochemistry* **33**, 2741–2747.

David E Wemmer, Department of Chemistry, University of California, Berkeley, CA 94720-1460, USA.