Peptide models of the Tat-TAR protein-RNA interaction

A. D. FRANKEL

Protein Sci. 1992 1: 1539-1542

Protein Science (1992), *1*, 1539–1542. Cambridge University Press. Printed in the USA.

REVIEW

Peptide models of the Tat–TAR protein–RNA interaction

ALAN D. FRANKEL

Department of Biochemistry and Biophysics, and Gladstone Institute of Virology and Immunology, University of California at San Francisco, San Francisco, California 94143

(RECEIVED September 8, 1992; ACCEPTED September 14, 1992)

RNA–protein interactions occur in a wide variety of biological systems; however, in only a few cases have the determinants of sequence-specific recognition been identified. Because RNA molecules can fold into rather complex structures, it had been widely anticipated that features of RNA tertiary structure, in addition to base-specific interactions, would contribute to recognition. Indeed, recent structural studies have demonstrated that RNA tertiary structure plays a crucial role in specific protein binding. The cocrystal structures of two tRNA synthetase–tRNA complexes (Rould et al., 1989, 1991; Ruff et al., 1991) have shown that the shape of the synthetases is complementary to the shape of the tRNAs and that, upon binding, an extensive protein–RNA interface is formed in which many specific interactions occur. Some interactions require conformational changes in the RNA that expose functional groups on bases, whereas other interactions occur with groups already accessible in the RNA major or minor grooves. In addition to base-specific interactions, contacts are observed to phosphates and hydroxyl groups, particularly those located at nonhelical regions of the RNA backbone. A somewhat simpler interaction occurs between the human immunodeficiency virus (HIV) Tat protein and its RNA-binding site, TAR. In this case, it appears that only a single amino acid is responsible for specific binding, nevertheless, the principal determinants of recognition are similar to those of the synthetases.

Tat is a virally encoded protein that activates HIV transcription through binding to the TAR RNA hairpin located at the 5′ end of viral mRNAs (Frankel, 1992). TAR contains a six-nucleotide loop and a three-nucleotide bulge (Fig. 1) that are both necessary for Tat function. Detailed analyses of TAR mutants have shown that Tat binding is localized to the bulge region (Cordingley et al., 1990; Dingwall et al., 1990; Roy et al., 1990) whereas cellular proteins appear to bind to the loop (Marciniak et al., 1990; Sheline et al., 1991; Wu et al., 1991).

The Tat protein is 86 amino acids long (Fig. 1) and contains two prominent features: a cysteine-rich region thought to form a metal-binding domain (Frankel et al., 1988) and an arginine-rich region of nine amino acids. Proteolytic digestion studies suggested that these regions are contained within two separate structural domains (Frankel et al., 1988). The C-terminal domain alone (residues 49–86) binds specifically to TAR RNA and appears to bind with the same specificity as the intact protein (Weeks et al., 1990). Studies with synthetic peptides further localized the RNA-binding domain to the arginine-rich region (residues 49–57; RKKRRQRRR); peptides as short as nine amino acids bind TAR with specificity and affinity ($K_d \approx 6$ nM) similar to the intact protein (Weeks et al., 1990; Calnan et al., 1991a,b). Thus, the Tat peptide–TAR interaction provides a relatively simple system to study RNA–protein recognition.

The RNA-binding peptide from Tat has several unusual characteristics in addition to its small size. First, circular dichroism (CD) spectra indicate that even though it binds specifically to TAR, the peptide is unstructured on its own (Calnan et al., 1991a). CD experiments further suggest that the peptide conformation may become ordered upon RNA binding, perhaps analogous to increased α-helix formation of bZIP basic peptides upon specific DNA binding (O'Neil et al., 1990; Talanian et al., 1990; Weiss et al., 1990). Second, the precise amino acid sequence required for specific RNA binding is flexible. For example, the wild-type Tat sequence, RKKRRQRRR, can be reversed or scrambled or changed into a homopolymer of nine arginines and still retain specific RNA-

Reprint requests to: Alan D. Frankel, Department of Biochemistry and Biophysics, and Gladstone Institute of Virology and Immunology, University of California at San Francisco, P.O. Box 419100, San Francisco, California 94141.

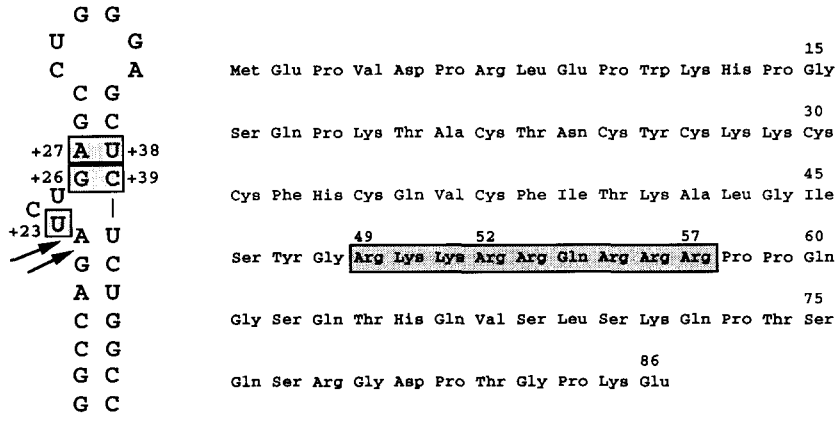

Fig. 1. Secondary structure of TAR RNA and sequence of the Tat protein. Nucleotides in TAR important for Tat recognition (boxed) were identified by mutagenesis and chemical modification experiments (Dingwell et al., 1990; Roy et al., 1990; Berkhout & Jeang, 1991; Sumner-Smith et al., 1991; Weeks & Crothers, 1991). Nucleotide numbering is relative to the start of HIV transcription (+1). Arrows indicate the positions of two phosphates whose modification interferes with binding (Calnan et al., 1991b). In the Tat protein, the nine-amino acid arginine-rich RNA-binding domain is highlighted.

binding activity in vitro and full transcriptional activation activity in vivo (Calnan et al., 1991a,b). The sequence flexibility suggests that the RNA-binding domain may be unstructured even in the context of the intact protein, and that upon binding TAR, each mutant domain can fold into a similar structure, perhaps determined by interactions with a defined RNA structure. Third, only a single arginine makes a sequence-specific contact with the RNA. A peptide containing one arginine within a set of lysines (KKKRKKKKK) binds specifically to TAR, whereas a homopolymer of nine lysines binds nonspecifically (Calnan et al., 1991b). Even as the free amino acid, arginine binds to TAR with specificity similar to the Tat peptides or Tat protein (Tao & Frankel, 1992). Given the unstructured nature of the RNA-binding domain and the relatively modest amino acid requirements for specific RNA binding, it appears that TAR structure provides the scaffold for Tat binding.

What features of TAR structure determine specific binding? Extensive mutagenesis and chemical modification experiments identified a set of nucleotides located near the bulge that are required for Tat protein, Tat peptide, and arginine binding (Fig. 1). The 5′-most uridine within the bulge (U23) and the two base pairs immediately above the bulge (G26–C39 and A27–U38) are essential for binding. Based on accessibility to chemical probes, it was proposed that the bulge serves to widen the RNA major groove of the adjacent duplex, allowing Tat to specifically recognize groups on base pairs above the bulge (Weeks & Crothers, 1991). Chemical modification experiments showed that two phosphates located at the junction of the bulge and the lower stem are also critical for binding (Calnan et al., 1991b). It was proposed that the guanidinium group of a single arginine might simultaneously donate hydrogen bonds to two adjacent phosphates (the "arginine fork"), thus allowing recognition of a specific structural feature of TAR. A preliminary NMR structure of the arginine–TAR complex (Puglisi et al., 1992; Fig. 2) shows that both the major groove and the RNA backbone contribute to specific arginine recognition. The guanidinium group of arginine appears to donate hydrogen bonds to the O6 and N7 groups of G26 in the major groove, and to the two phosphates. The complex appears to be stabilized by formation of a base triple between U23 in the bulge and the A27–U38 base pair above the bulge, which helps position the phosphates to interact with arginine. The existence of the base triple is supported by the observation that a C23–G27–C38 triple mutant forms a structure essentially identical to wild-type TAR (J.D. Puglisi, L. Chen, A.D. Frankel, & J.R. Williamson, submitted for publication). In the cocrystal structure of the Zif268 protein–DNA complex (Pavletich & Pabo, 1991), arginines make similar contacts to guanines in the major groove, but instead of using phosphates to position the guanidinium groups, aspartic or glutamic acid residues located elsewhere in the protein are used. Thus, RNA tertiary structure in TAR and protein tertiary structure in Zif268 can apparently play analogous roles in stabilizing arginine–guanine interactions.

Although TAR RNA provides an important structural framework for arginine or Tat recognition, its conformation is not static. CD difference spectra suggest that the extent of base stacking changes upon peptide, arginine, or even guanidine binding to TAR (Calnan et al., 1991a; Tan & Frankel, 1992). The precise nature of the conformational change can be seen by comparing the structures of unbound TAR and of the arginine–TAR complex, determined by NMR (Puglisi et al., 1992; Fig. 2). In unbound TAR, the three bulge nucleotides are at least partially stacked between two A-form RNA helices and there is no evidence of the base triple described above. Upon arginine or peptide binding, the bulge nucleotides become unstacked, U23 forms the base triple with A27–U38, and the helices coaxially stack to bring the two phosphates near G26, forming the arginine-binding pocket. The energetic cost of the conformational change is probably small and any unfavorable energy of unstacking the bulge nucleotides is likely to be offset by coaxial stacking of the helices, formation of the base triple, and hydrogen bonding to arginine. These interactions could

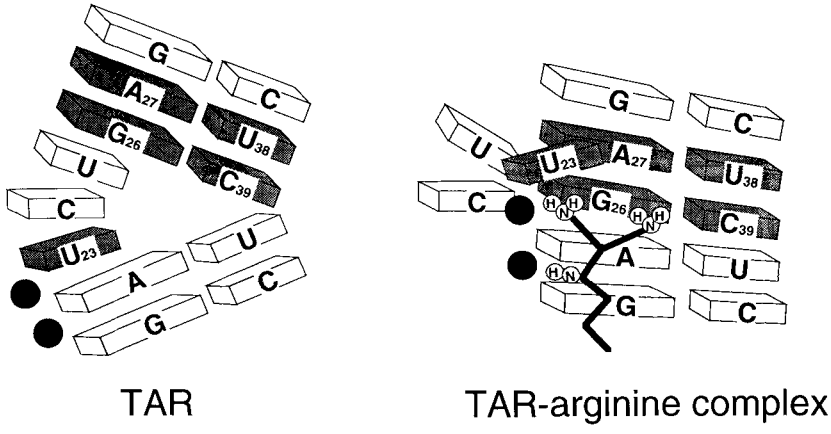

Fig. 2. Schematic representation of the conformations of TAR and the TAR–arginine complex, determined by NMR (Puglisi et al., 1992). Only the bulge region of TAR, where Tat and arginine binding is localized, is shown. In unbound TAR, the three bulge nucleotides are partially stacked within one strand of the duplex RNA stem. Arginine binding induces a conformational change in which U23 of the bulge appears to form a base triple with the A27–U38 base pair and the two other bulge nucleotides become unstacked and are apparently free in solution. Two phosphates important for recognition (black dots) appear to become positioned in the bound structure to accept hydrogen bonds from the arginine guanidinium group, and arginine appears to donate additional hydrogen bonds to the O6 and N7 groups of G26 in the major groove.

easily account for the 10–40-fold specificity observed for Tat and arginine binding (Weeks & Crothers, 1991; Tao & Frankel, 1992).

The change in TAR conformation upon binding of the rigid guanidinium group is reminiscent of induced fit in enzyme–substrate interactions, in this case with RNA playing the role of the enzyme. In the context of the peptide, however, the specificity of guanidinium binding may be influenced by the surrounding electrostatics and/or possible conformational changes of the peptide. Mutagenesis of basic residues N- and C-terminal to the sequence-specific arginine indicates that while each additional charge increases the overall RNA-binding affinity (Calnan et al., 1991a; Delling et al., 1991), basic amino acids at positions $i-2$ and $i+2$, relative to the arginine, are particularly important (Tao & Frankel, 1993). These charged residues appear to enhance the specificity of arginine binding to TAR (e.g., increase discrimination between TAR and TAR mutants), perhaps by neutralizing adjacent phosphates in the bulge that are brought close together in the bound structure. This is consistent with the observation that arginine binding is enhanced by chemically modifying (neutralizing) the bulge phosphates (Tao & Frankel, 1992). Unfavorable electrostatics between neighboring phosphates, as also observed in an RNA pseudoknot (Puglisi et al., 1990), may be common to complex folded RNA structures and may provide an important driving force for protein binding.

The interaction of Tat and TAR highlights the importance of RNA structure in RNA–protein recognition and demonstrates that short peptides, and even amino acids, are capable of sequence-specific RNA binding. Arginine has also been shown to bind to the *Tetrahymena* ribozyme (Yarus, 1988) but in contrast to Tat, appears to require more than just the guanidinium group for specific recognition. In DNA–protein recognition, hydrogen bonds between amino acid side chains and specific arrays of groups on the bases are often essential determinants of specificity. In the TAR–arginine interaction, additional hydrogen bonds to the RNA backbone appear to be required. "Nonspecific" electrostatic interactions between amino acids in the Tat peptide and RNA phosphates help to increase the overall binding affinity and may also modulate arginine-binding specificity. Conformational changes occur in both RNA and peptide upon binding; peptide electrostatics may help to stabilize the bound RNA conformation, and the RNA backbone may help to "shape" the peptide structure. Thus, as seen in more complex systems, RNA structure is a critical feature of Tat–TAR recognition. Because many RNAs can fold into unique tertiary structures, one may anticipate that other proteins will use short peptide motifs to recognize RNA structures and that other interesting peptide models of RNA–protein recognition will emerge.

References

Berkhout, B. & Jeang, K.-T. (1991). Detailed mutational analysis of TAR RNA: Critical spacing between the bulge and loop recognition domains. *Nucleic Acids Res. 19*, 6169–6176.

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

Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A.D. (1991b). Arginine-mediated RNA recognition: The arginine fork. *Science 252*, 1167–1171.

Cordingley, M.G., LaFemina, R.L., Callahan, P.L., Condra, J.H., Sardana, V.V., Graham, D.J., Nguyen, T.M., LeGrow, K., Gotlib, L., Schlabach, A.J., & Colonno, R.J. (1990). Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. USA 87*, 8985–8989.

Delling, U., Roy, S., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C.A., & Sonenberg, N. (1991). The number of positively charged amino acids in the basic domain of Tat is critical for trans-activation and complex formation with TAR RNA. *Proc. Natl. Acad. Sci. USA 88*, 6234–6238.

Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., & Skinner, M.A. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J. 9*, 4145–4153.

Frankel, A.D. (1992). Activation of HIV transcription of Tat. *Curr. Opin. Genet. Dev. 2*, 293–298.

Frankel, A.D., Bredt, D.S., & Pabo C.O. (1988). Tat protein from hu-

man immunodeficiency virus forms a metal-linked dimer. *Science 240*, 70–73.

Marciniak, R.A., Garcia-Blanco, M., & Sharp, P.A. (1990). Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation-response (TAR) element of human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA 87*, 3624–3628.

O'Neil, K.T., Hoess, R.H., & DeGrado, W.F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. *Science 249*, 774–778.

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.

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.

Puglisi, J.D., Wyatt, J.R., & Tinoco, I., Jr. (1990). Conformation of an RNA pseudoknot. *J. Mol. Biol. 214*, 437–453.

Rould, M.A., Perona, J.J., Soll, D., & Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science 246*, 1135–1142.

Rould, M.A., Perona, J.J., & Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature 352*, 213–218.

Roy, S., Delling, U., Chen, C.H., Rosen, C.A., & Sonenberg, N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev. 4*, 1365–1373.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science 252*, 1682–1689.

Sheline, C.T., Milocco, L.H., & Jones, K.A. (1991). Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes Dev. 5*, 2508–2520.

Sumner-Smith, M., Roy, S., Barnett, R., Reid, L.S., Kuperman, R., Delling, U., & Sonenberg, N. (1991). Critical chemical features in trans-acting-responsive RNA are required for interaction with human immunodeficiency virus type 1 Tat protein. *J. Virol. 65*, 5196–5202.

Talanian, R.V., McKnight, C.J., & Kim, P.S. (1990). Sequence-specific DNA binding by a short peptide dimer. *Science 249*, 769–771.

Tan, R. & Frankel, A.D. (1992). Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine. *Biochemistry 31*, 10288–10294.

Tao, J. & Frankel, A.D. (1992). Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. USA 89*, 2723–2726.

Tao, J. & Frankel, A.D. (1993). Electrostatic interactions modulate the RNA-binding and transactivation specificities of the HIV and SIV Tat proteins. *Proc. Natl. Acad. Sci. USA*, in press.

Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A., & Crothers, D.M. (1990). Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science 249*, 1281–1285.

Weeks, K.M. & Crothers, D.M. (1991). RNA recognition by Tat-derived peptides: Interaction in the major groove? *Cell 66*, 577–588.

Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrison, S.C., & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature 347*, 575–578.

Wu, F., Garcia, J., Sigman, D., & Gaynor, R. (1991). Tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185. *Genes Dev. 5*, 2128–2140.

Yarus, M. (1988). A specific amino acid binding site composed of RNA. *Science 240*, 1751–1758.