Role of RNA structure in arginine recognition of TAR RNA

(transactivation response element/NMR spectroscopy/RNA structure/RNA–protein interactions)

JOSEPH D. PUGLISI*, LILY CHEN[†‡], ALAN D. FRANKEL[†‡], AND JAMES R. WILLIAMSON*[§]

*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; and [†]Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142

ABSTRACT   The human immunodeficiency virus Tat protein binds specifically to an RNA stem–loop structure (TAR) that contains two helical stem regions separated by a three-nucleotide bulge. A single arginine within the basic region of Tat mediates specific binding to TAR, and arginine as the free amino acid also binds specifically to TAR. We have previously proposed a model in which interaction of the arginine guanidinium group with guanosine-26 (G^{26}) and with a pair of phosphates is stabilized by formation of a base triple between U^{23} in the bulge and $A^{27} \cdot U^{38}$ in the upper helix. Here we show by NMR spectroscopy that formation of the base triple is critical for arginine binding to TAR. Mutants of TAR that cannot form the base triple or that remove the guanine contact do not bind arginine specifically. These mutants also showed reduced transactivation by Tat. A triple mutant designed to form an isomorphous base triple between C^{23} and $G^{27} \cdot C^{38}$ binds arginine and adopts the same conformation as wild-type TAR. These results demonstrate the importance of RNA structure for arginine binding and further demonstrate the direct correspondence between arginine and Tat binding.

The three-dimensional structures adopted by RNA molecules are crucial for their specific recognition by proteins (1–3). The human immunodeficiency virus (HIV) Tat protein binds specifically to an RNA stem–loop structure, TAR, located at the 5' end of the viral mRNA (4–6), and stimulates transcription (7). TAR contains three unpaired nucleotides (a bulge) that separate two helical stem regions. A domain of basic amino acids mediates RNA recognition by Tat, and peptides that correspond to this domain bind specifically to TAR (6, 8–10). A single arginine within the basic region of Tat is critical for specific recognition (11) and arginine as the free amino acid binds specifically to TAR (12). Certain positions in TAR—U^{23} in the bulge, $G^{26} \cdot C^{39}$, and $A^{27} \cdot U^{38}$—are required for transactivation (5, 13, 14), specific peptide binding (6, 9, 10, 15), and arginine binding (12, 16). This suggests similar recognition of TAR by arginine whether as the free amino acid or within the context of the peptide or protein.

We have recently shown that TAR changes conformation upon arginine binding, and we proposed a model in which formation of a base-triple interaction between U^{23} in the bulge and $A^{27} \cdot U^{38}$ in the upper stem of TAR stabilizes the interaction of arginine with G^{26} and phosphates in the major groove (17). However, direct evidence for the specific hydrogen bonds involved in the arginine–TAR complex was not obtained, since the exchangeable protons involved in the proposed base triple and arginine contacts did not give observable resonances. In order to demonstrate directly the formation of the base triple, a triple mutant of TAR was designed in which a C–G·C base triple was expected to form that would be isomorphic to the base triple in wild-type TAR. Comparison of the arginine complexes of the triple mutant

and wild-type TAR by NMR spectroscopy indicates that the two RNAs adopt the same conformation. This demonstrates the importance of the RNA structure, rather than only its particular base sequence, for arginine binding.

MATERIALS AND METHODS

NMR Spectroscopy. Milligram quantities of RNA oligonucleotides were synthesized in vitro by T7 RNA polymerase and purified by means of denaturing polyacrylamide gel electrophoresis (18). NMR experiments were performed in 50 mM NaCl/10 mM sodium phosphate, pH 7.5, 6.5, or 5.5 at 25°C; RNA strand concentrations were 1–1.5 mM. Samples were dialyzed against the final buffer. NMR experiments were performed at 500 MHz on a Varian VXR-500 Spectrometer. Data were processed on a Silicon Graphics Personal Iris using FELIX software (Hare Research, Bothell, WA). Exchangeable and nonexchangeable proton resonances were assigned by standard techniques (19, 20), including nuclear Overhauser effect spectroscopy (NOESY), double-quantum filtered correlated spectroscopy, and total correlation spectroscopy. All aromatic H8, H6, H5, and H2 resonances and sugar H1', H2', and H3' resonances were assigned. NOESY spectra were acquired with mixing times of 50, 100, and 400 ms by standard methods (20). Ribose sugar conformations were determined by using values of vicinal ribose $^1H^{-1}H$ coupling constants determined from double-quantum filtered correlated spectroscopy as described (21). Since the data for the triple mutant and wild-type TAR (17) were essentially identical, no explicit model building was performed. The nuclear Overhauser effects (NOEs) observed for the triple mutant are fully consistent with the model calculated in ref. 17.

Transactivation Assay. HeLa cells were transfected with each reporter plasmid alone (10 ng; shown only for wild-type TAR), to determine the basal level of each mutant, or were cotransfected with each reporter and the Tat expression vector pSV2tat72 (5 ng). Total DNA concentrations were adjusted to 1 μg/ml. Chloramphenicol acetyltransferase (CAT) activities were quantitated as described (10) and the ratios of CAT activities in the presence or absence of Tat were used to calculate the fold transactivation. Assays were repeated four times with essentially the same results.

RESULTS

NMR of TAR Variants in the Absence of Arginine. The conformations of four TAR mutants (Fig. 1a; I–IV) with substitutions at positions known to affect Tat binding have

Abbreviations: HIV, human immunodeficiency virus; CAT, chloramphenicol acetyltransferase; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy.
[‡]Present address: Department of Biochemistry and Biophysics, and Gladstone Institute of Virology and Immunology, P.O. Box 419100, University of California, San Francisco, CA 94141.
[§]To whom reprint requests should be addressed.

Biochemistry: Puglisi *et al.*

Proc. Natl. Acad. Sci. USA 90 (1993) 3681

a

```
            G  G
          U      G
          C      A
           C₂₉ G₃₆              I
           G • C            ┌─────────┐
                            │ U₂₇ A₃₈ │
IV      ┌────────┐  A • U   └─────────┘
        │ G₂₇C₃₈ │←─┤A • U├──→          II
        │   +    │  ┌─────┐          ┌─────────┐
        │  C₂₃   │  │G • C│──────────→│ A₂₆ U₃₉ │
        └────────┘   26 39           └─────────┘
                 U  ┌─────┐             III
              C  │U₂₃│──────────────→┌─────────┐
                 └───┘A₂₂ U₄₀        │  C₂₃    │
                     G • C           └─────────┘
                     A • U
                     C • G
                     C • G
                     G • C
                  5'G • C3'
```
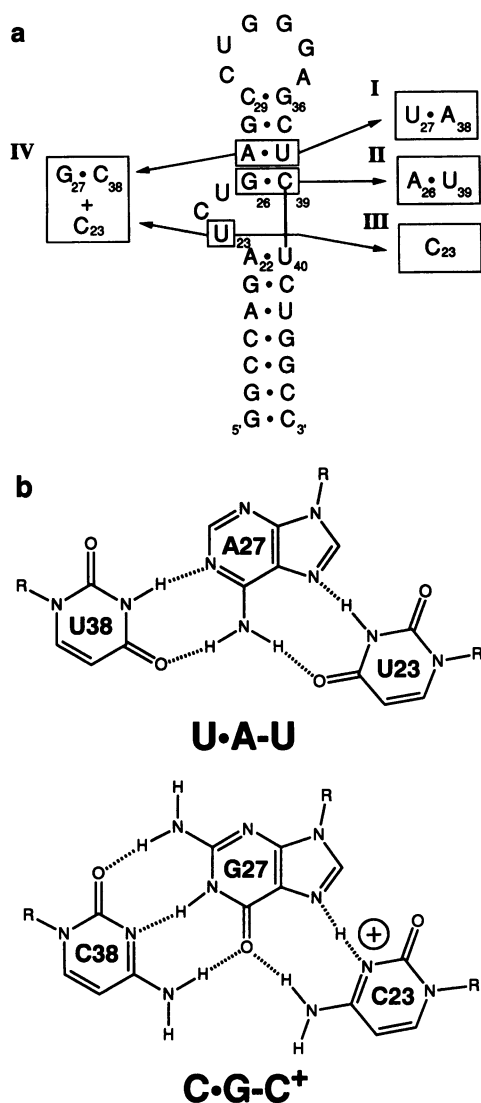
b

U•A-U

C•G-C⁺

FIG. 1. (*a*) Sequence and secondary structure of TAR RNA. The positions mutated in this study are boxed. The sequence changes in the four TAR variants (I–IV) are shown explicitly. Numbering refers to nucleotides in the HIV HXB-2 isolate relative to the cap site; G^{16}, G^{17}, C^{45}, and C^{46} are not part of HIV TAR and were added to increase the efficiency of *in vitro* transcription. (*b*) Isomorphic base triples between U^{23} and $A^{27}\cdot U^{38}$ in wild-type TAR and C^{23} and $G^{27}\cdot C^{38}$ in the triple mutant (IV).

been characterized in detail by proton NMR spectroscopy. Two mutants (17), U^{23} to C^{23} (**III**) and $A^{27}\cdot U^{38}$ to $U^{27}\cdot A^{38}$ (**I**), were expected to disrupt the base triple when bound to arginine. A third mutant, $G^{26}\cdot C^{39}$ to $A^{26}\cdot U^{39}$ (**II**), was expected to disrupt a direct arginine contact. A fourth mutant, containing a triple substitution (**IV**) in which U^{23} was changed to C^{23} and $A^{27}\cdot U^{38}$ was changed to $G^{27}\cdot C^{38}$, was designed to form an isomorphic base triple (Fig. 1*b*). In the absence of arginine, all four mutants were found to adopt conformations similar to that of wild-type TAR. The 5'-most nucleotide in the bulge (either U^{23} or C^{23}) stacks over A^{22} in the lower stem of TAR. The conformation of the other two bulge nucleotides is poorly defined by the NMR data. The upper and lower stems adopt A-form conformations, as previously described (17). Thus, these mutations have no effect on the unbound structure of TAR.

NMR of the Triple Mutant Plus Arginine. The triple mutant changes conformation upon binding of argininamide (a tight-binding arginine analog) in the same manner as previously observed for wild-type TAR (17). This triple mutant was

designed to form a $C^{23}\text{-}G^{27}\cdot C^{38}$ triple that is isomorphic with the wild-type $U^{23}\text{-}A^{27}\cdot U^{38}$ interaction (Fig. 1*b*). Changes in chemical shifts similar to those seen with wild-type TAR were observed upon addition of 6 mM argininamide to triple mutant TAR, suggesting a similar arginine binding site and similar conformational changes. The three bulge nucleotides in the bound conformation of the triple mutant are not stacked between the two helices, which are coaxially stacked. NOEs were observed between the C^{23}(H1') proton in the bulge and G^{26}(H8) and G^{26}(H3') (Fig. 2*a*). These NOEs position C^{23} in the major groove of the upper stem near G^{27}, consistent with formation of a base-triple interaction between $G^{27}\cdot C^{38}$ and C^{23}. The bound conformation of the triple mutant is essentially the same as that of wild-type TAR.

Argininamide interacts with the triple mutant in the same manner as with wild-type TAR. The same set of NOEs were observed between the argininamide δ proton and protons in the triple mutant or in wild-type TAR (Fig. 2 *a* and *b*); these NOEs position the arginine in the major groove so that the guanidinium group can donate a pair of hydrogen bonds to G^{26} and form additional interactions with the phosphates between G^{21} and A^{22} (*P22*) and A^{22} and C^{23} (*P23*). The arginine is positioned directly below the base moiety of C^{23} (or U^{23} in wild-type TAR) (Fig. 2*c*).

The bound conformation of the triple mutant is stabilized by lowering the pH to 5.5, consistent with protonation of C^{23} upon base-triple formation (23). Only very weak NOEs between the argininamide(δ) proton and TAR are observed above pH 7.0 (data not shown), and NOEs that indicate formation of the base triple are not observed. This suggests that the base triple and arginine binding in the triple mutant are unstable above pH 7.0.

The mutants of TAR that disrupt base-triple formation (U^{23} to C^{23} or $A^{27}\cdot U^{38}$ to $U^{27}\cdot A^{38}$) neither bind arginine nor undergo a conformational change (12, 16, 17). Similarly, mutation of $G^{26}\cdot C^{39}$, to which arginine forms bidentate hydrogen bonds with G^{26} (17), also disrupts specific arginine binding. On addition of 6 mM argininamide, small changes in structure were observed by NMR, but U^{23} remained stacked above A^{22} and no NOEs indicative of base-triple formation were observed (data not shown). Furthermore, no intermolecular NOEs were observed between argininamide and the $A^{26}\cdot U^{38}$ mutant.

Transactivation. The *in vivo* transactivation by TAR mutants is consistent with base-triple formation and direct interaction of arginine and G^{26}. Mutation of U^{23} in the bulge (to C^{23}), of the $A^{27}\cdot U^{38}$ base pair (to $G^{27}\cdot C^{38}$), or of both positions simultaneously reduced transactivation to the same extent (by a factor of 3–4; Fig. 3). The similar effect of these mutations was expected for interacting positions. The triple mutant was not expected to function as well as wild-type because C^{23} would be unprotonated at physiological pH (see above). Thus, the $U^{23}\text{-}A^{27}\cdot U^{38}$ base triple is optimal for arginine binding (and Tat function) *in vivo*. Mutation of $G^{26}\cdot C^{39}$ reduced activity much more dramatically (Fig. 3), consistent with a direct interaction of arginine with G^{26}.

DISCUSSION

Our results show that the arginine interaction with TAR is stabilized by formation of a base-triple interaction. Mutations that disrupt the base triple destroy specific arginine binding to TAR, whereas the compensatory triple mutation restores binding. This suggests that arginine does not directly contact these bases. The base triple stabilizes arginine binding by positioning phosphates below the bulge to interact favorably with the arginine guanidinium group; the interaction with two well-positioned phosphates was predicted by the "arginine fork" model (11). It is possible that arginine binding is further
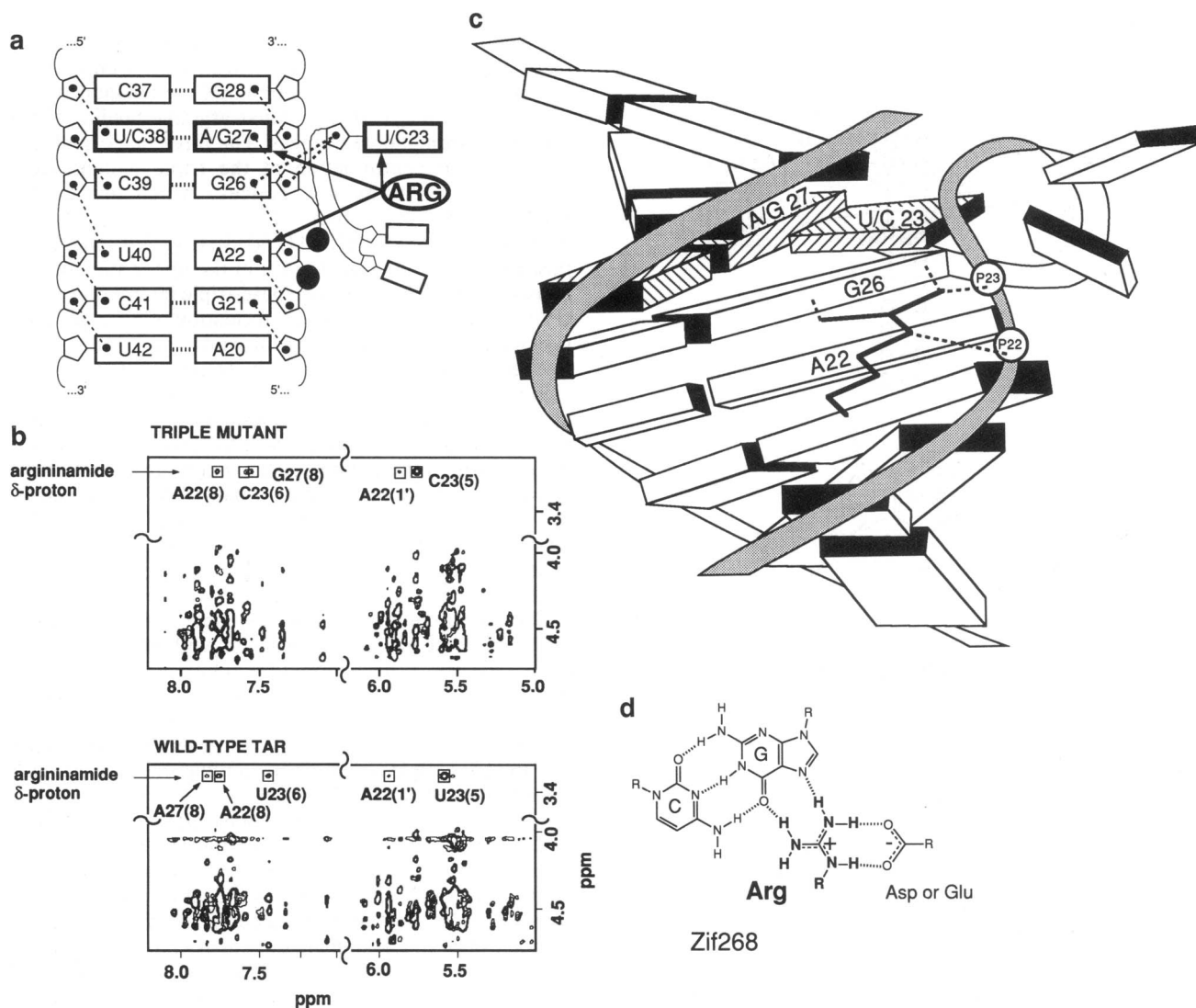
FIG. 2. (*a*) Summary of NMR data obtained on the bulge region of either wild-type TAR or the triple mutant in the presence of 6 mM argininamide. Bases are indicated by rectangles and ribose sugars by pentagons. Ribose, base H8/H6, and imino protons are represented by dots within pentagons or on the outside of bases, respectively. Large black circles represent phosphates between bases 21 and 22 and bases 22 and 23. Base-pair hydrogen bonding is shown by dashed lines between bases. Observed internucleotide NOEs are indicated by thin dashed lines; NOEs that indicate base-triple formation are shown by thicker dashed lines. The nucleotides that form the base triple are highlighted. Smaller riboses have a majority C2′-*endo* conformation, all other sugars adopt C3′-*endo* conformations. The argininamide (ARG) is represented by the oval; NOEs between the δ proton and positions in TAR are shown by arrows. (*b*) Portion of the NOESY spectra of triple mutant (*Upper*) and wild-type TAR (*Lower*) that shows NOEs between the argininamide δ proton and protons in the RNA. The mixing time for this experiment was 400 ms. (*c*) Schematic representation of the interaction of arginine with TAR. The arginine guanidinium group is black, and potential hydrogen bonds to G^{26} and two phosphates (P22 and P23) are indicated by dashed lines. U^{23} (or C^{23}) and $A^{27}·U^{38}$ ($G^{27}·C^{38}$), which form a base triple, are hatched. The two helical stems are coaxially stacked between A^{22} and G^{26}. Note the stacking of the guanidinium group below nucleotide 23 in the base triple. (*d*) Analogous contacts of arginine with guanine in the major groove are made in the Zif268 zinc-finger/DNA cocrystal structure (22) and in TAR.

stabilized by stacking of the polarizable guanidinium group (24) below nucleotide 23 in the major groove (Fig. 2*c*).

The specific interaction of L-arginine with the guanosine binding site of the *Tetrahymena* group I intron (24, 25) shares similar features with the arginine–TAR interaction. The guanosine binding site in the intron consists of two stem regions separated by a single-nucleotide bulge, and arginine probably forms bidentate hydrogen bonds to a conserved G·C pair at the base of the bulge, as well as to phosphate groups (24). Thus, RNA structure, base- and phosphate-specific contacts, and stacking of arginine and nucleotides determine specific recognition of the intron by L-arginine. In contrast, it appears that the aliphatic part of the arginine side chain plays a role in binding to the intron (25) whereas the guanidinium group is sufficient for TAR binding (12, 16).

Bidentate hydrogen bonds between arginine and guanine (26) have been observed in a number of DNA–protein interactions. In our model for the arginine–TAR complex, arginine forms bidentate hydrogen bonds to O6 and N7 atoms of G^{26}, and mutation to A^{26} destroys specific arginine binding and the RNA conformational change. Arginine interactions with guanine in the Zif268 zinc-finger-domain/DNA cocrystal structure (22) are buttressed by interactions with negatively charged groups (either aspartic or glutamic residues) elsewhere in the protein (Fig. 2*d*). The arginine–guanine interaction in TAR is stabilized in an analogous manner, by phosphate groups in the RNA rather than by a protein side chain.

The major features of the TAR bulge conformation (Fig. 2*c*) have been observed in other RNA structures. Base triples

Biochemistry: Puglisi *et al.*
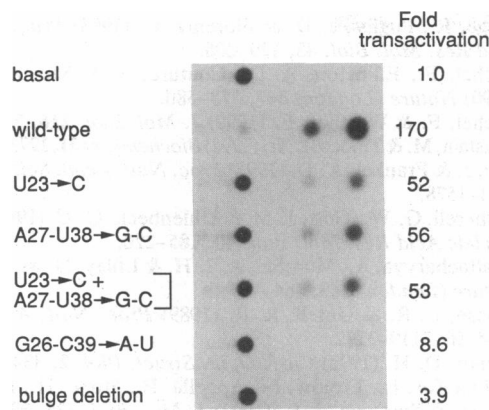
Proc. Natl. Acad. Sci. USA 90 (1993) 3683

FIG. 3. *In vivo* transactivation activity of TAR mutants. Tat transactivation of HIV long terminal repeat/CAT reporter constructs with wild-type or mutant TAR was measured. The far left TLC spot for each mutant represents unacetylated [^{14}C]chloramphenicol, whereas the spots on the right are acetylated [^{14}C]chloramphenicol. The level of transactivation was calculated as described in *Materials and Methods*. Note that mutation of the G^{26}·C^{39} base pair reduces activity almost as much as deleting the bulge entirely.

FIG. 4. Summary of RNA sequence requirements for TAR in transactivation function (*a*), Tat binding *in vitro* (*b*), Tat peptide binding (*c*), and arginine binding (*d*). Nucleotides that are critical for function are shown explicitly. Filled circles indicate positions whose mutation has no effect on TAR function; for stem positions, Watson–Crick base pairing was maintained on mutation. Untested positions are shown as open circles. The loop region of TAR is not required for Tat binding. Data are from references cited in the text.

are critical RNA structural elements (27), and formation of isomorphic base triples is suggested by phylogenetic covariations in tRNAs (28, 29) and group I introns (30, 31). Coaxial stacking of RNA helices appears to be a common theme in RNA structure and can direct adjacent single-stranded regions into the major groove (3' strand of unpaired nucleotides) or minor groove (5' strand of unpaired nucleotides) (27, 31, 32). Although bulge nucleotides in principle can form base triples in either groove, major-groove interaction is preferred in TAR, perhaps to facilitate arginine binding in the major groove.

Binding, mutagenesis, and chemical modification data strongly support the structural importance of the arginine–TAR complex in Tat–TAR recognition. In particular, the identical set of nucleotides—U^{23}, G^{26}·C^{39}, A^{27}·U^{38}—is important for Tat transactivation (5, 13, 14), Tat binding (4, 14), Tat peptide binding (6, 9, 10, 15), and arginine binding (12, 16, 17) (Fig. 4). These positions form the arginine binding site in TAR. The relatively weak discrimination of Tat peptides among TAR mutants (5- to 20-fold) corresponds to 1–2 kcal/mol of specific binding energy (9–11, 15). This small amount of energy could readily be provided by the arginine–guanine and arginine–phosphate hydrogen bonds and formation of the base triple. There is no thermodynamic need to invoke multiple base-specific protein contacts. Other basic amino acids in the RNA-binding domain of Tat may enhance arginine binding specificity through electrostatic contacts with the phosphodiester backbone, particularly near the bulge (33). Elucidation of the precise arrangement of non-base-specific contacts and the orientation of Tat protein on TAR awaits further structural analyses.

Bulges can be important sites of protein recognition (34), local bending (35), and RNA tertiary structure (36). Bulged nucleotides can form tertiary interactions, stack between helices, or protrude into solution (37, 38). All of these features are observed in the very simple TAR bulge structure. The architecture of the TAR bulge (Fig. 2c) presents a particular helical base pair and phosphates for specific recognition by an amino acid side chain. In other cases, additional interactions may occur. For example, nucleotides in a larger bulge region might form additional base triples, or the arginine interaction, which stabilizes the conformation of the TAR bulge, might be replaced by other RNA–RNA contacts. The structural characteristics illustrated by the TAR bulge
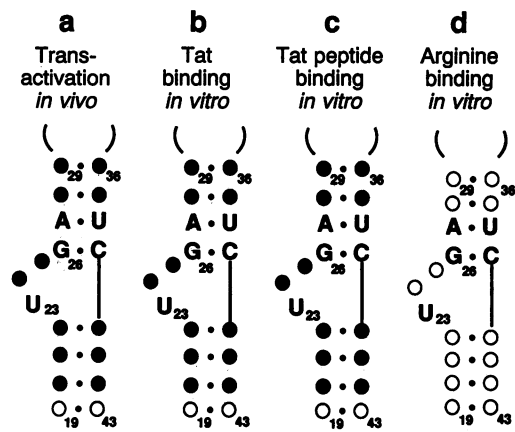
motif may help in prediction of analogous interactions in other systems.

1. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246**, 1135–1142.
2. Rould, M. A., Perona, J. J. & Steitz, T. A. (1991) *Nature (London)* **352**, 213–218.
3. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991) *Science* **252**, 1682–1689.
4. Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M. & Skinner, M. A. (1990) *EMBO J.* **9**, 4145–4153.
5. Roy, S., Delling, U., Chen, C.-H., Rosen, C. A. & Sonenberg, N. (1990) *Genes Dev.* **4**, 1365–1373.
6. 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) *Proc. Natl. Acad. Sci. USA* **87**, 8985–8989.
7. Frankel, A. D. (1992) *Curr. Opin. Genet. Dev.* **2**, 293–298.
8. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. (1990) *Science* **249**, 1281–1285.
9. Weeks, K. M. & Crothers, D. M. (1991) *Cell* **66**, 577–588.
10. Calnan, B. J., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) *Genes Dev.* **5**, 201–210.
11. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) *Science* **252**, 1167–1171.
12. Tao, J. & Frankel, A. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2723–2726.
13. Berkhout, B. & Jeang, K.-T. (1991) *Nucleic Acids Res.* **19**, 6169–6176.
14. Delling, U., Reid, L. S., Barnett, R. W., Ma, M. Y.-X., Climie, S., Sumner-Smith, M. & Sonenberg, N. (1992) *J. Virol.* **66**, 3018–3025.
15. Weeks, K. M. & Crothers, D. M. (1992) *Biochemistry* **31**, 10281–10287.
16. Tan, R. & Frankel, A. D. (1992) *Biochemistry* **31**, 10288–10294.
17. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) *Science* **257**, 76–80.

18. Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991) *BioTechniques* 11, 764–769.
19. Puglisi, J. D., Wyatt, J. R. & Tinoco, I., Jr. (1990) *J. Mol. Biol.* 214, 437–453.
20. Varani, G. & Tinoco, I., Jr. (1991) *Q. Rev. Biophys.* 24, 479–532.
21. Davis, P. W., Adamiak, R. W. & Tinoco, I., Jr. (1990) *Biopolymers* 29, 109–122.
22. Pavletich, N. P. & Pabo, C. O. (1991) *Science* 252, 809–817.
23. Sklenár, V. & Feigon, J. (1990) *Nature (London)* 345, 836–838.
24. Yarus, M. & Majerfeld, I. (1992) *J. Mol. Biol.* 225, 945–949.
25. Yarus, M. (1988) *Science* 240, 1751–1758.
26. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* 73, 804–808.
27. Chastain, M. & Tinoco, I., Jr. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* 41, 131–137.
28. Levitt, M. (1969) *Nature (London)* 224, 759–763.
29. Giegé, R., Puglisi, J. D. & Florentz, C. (1993) *Prog. Nucleic Acid Res. Mol. Biol.* 45, 129–206.
30. Michel, F., Ellington, A. D., Couture, S. & Szostak, J. W. (1990) *Nature (London)* 347, 578–580.
31. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* 216, 585–610.
32. Chastain, M. & Tinoco, I., Jr. (1992) *Biochemistry* 31, 12733–12741.
33. Tao, J. & Frankel, A. D. (1993) *Proc. Natl. Acad. Sci. USA* 90, 1571–1578.
34. Witherell, G. W., Gott, J. M. & Uhlenbeck, O. C. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* 40, 185–220.
35. Bhattacharyya, A., Murchie, A. I. H. & Lilley, D. M. J. (1990) *Nature (London)* 343, 484–487.
36. Woese, C. R. & Gutell, R. R. (1989) *Proc. Natl. Acad. Sci. USA* 86, 3119–3122.
37. Turner, D. H. (1992) *Curr. Opin. Struct. Biol.* 2, 334–337.
38. Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D. & Sussman, J. L. (1992) *J. Mol. Biol.* 225, 397–431.