Conserved Nucleotides in the TAR RNA Stem of Human Immunodeficiency Virus Type 1 Are Critical for Tat Binding and *trans* Activation: Model for TAR RNA Tertiary Structure

ULRIKE DELLING,¹ LORNE S. REID,² RICHARD W. BARNETT,² MICHAEL Y.-X. MA,² SHANE CLIMIE,² MARTIN SUMNER-SMITH,² AND NAHUM SONENBERG^{1,3*}

Department of Biochemistry¹ and McGill Cancer Center,³ McGill University, Montreal, Quebec H3G 1Y6, and Allelix Biopharmaceuticals Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1,² Canada

Received 20 November 1991/Accepted 28 January 1992

Interaction between the human immunodeficiency virus type 1 (HIV-1) *trans*-activator Tat and its *cis*-acting responsive RNA element TAR is necessary for activation of HIV-1 gene expression. We investigated the hypothesis that the essential uridine residue at position 23 in the bulge of TAR RNA is involved in intramolecular hydrogen bonding to stabilize a unique RNA structure required for recognition by Tat. Nucleotide substitutions in the two base pairs of the TAR stem directly above the essential trinucleotide bulge that maintain base pairing but change sequence prevent complex formation with Tat *in vitro*. Corresponding mutations tested in a *trans*-activation assay strongly affect the biological activity of TAR *in vivo*, suggesting an important role for these nucleotides in the Tat-TAR interaction. On the basis of these data, a model is proposed which implicates uridine 23 in a stable tertiary interaction with the GC pair directly above the bulge. This interaction would cause widening of the major groove of the RNA, thereby exposing its hydrogen-bonding surfaces for possible interaction with Tat. The model also predicts a gap between uridine 23 and the first base pair in the stem above, which would require one or more unpaired nucleotides to close, but does not predict any other role for such nucleotides. In accordance with this prediction, synthetic propyl phosphate linkers of equivalent length to 1 or 2 nucleotides, were found to be fully acceptable substitutes in the bulge above uridine 23, demonstrating that neither the bases nor the ribose moieties at these positions are implicated in the recognition of TAR RNA by Tat.

Detailed study of human immunodeficiency virus type 1 (HIV-1) has revealed novel mechanisms of gene regulation. One of the essential viral regulatory proteins, the *trans*-activator Tat, strongly stimulates HIV-1 gene expression via a *cis*-acting RNA target sequence termed TAR (*trans*-activation-responsive element), which is located downstream of the transcription start site in the viral long terminal repeat (6, 22, 23, 31). Tat interacts with the TAR element as a nascent transcript (2), and the physical interaction between Tat and TAR RNA has been demonstrated *in vitro*, using RNA mobility shift and filter binding assays (5, 8, 9, 24). TAR RNA assumes a stable stem-loop structure which is critical for the interaction with Tat (1, 11, 20, 25, 28). Intact base pairing in the double-stranded RNA forming the stem of TAR is essential for recognition by Tat (8, 9, 24). In addition, a uridine residue at the 5'-most position (+23) of a 3-nucleotide bulge, which is absolutely conserved among all HIV isolates, is the only base which cannot be substituted by any of the other three natural bases (24, 35, 37). The formation of a nucleoprotein complex between Tat and TAR RNA *in vitro* correlates with Tat-mediated *trans* activation *in vivo*, inasmuch as mutations in TAR that prevent the binding of Tat also abolish *trans* activation (8, 24). These findings indicate that Tat-TAR RNA complex formation is essential but not sufficient for *trans* activation of HIV-1 gene expression by Tat. Mutations in the loop of TAR which strongly decrease or completely abolish Tat-mediated *trans* activation do not interfere with binding. It was therefore proposed that cellular TAR-binding factors stabilize the interaction between Tat

and TAR RNA *in vivo*, and indeed several of these factors have been identified in HeLa and CHO cells (12, 13, 18). This hypothesis is supported by recent reports that TAR-binding proteins stimulate Tat-mediated *trans* activation in an *in vitro* system (17, 38).

A highly basic domain (two lysine and six arginine residues between amino acids 49 and 57) in the carboxy half of Tat mediates the specific interaction with TAR RNA. Small peptides spanning this region bind TAR RNA as efficiently as the full-length protein when tested in an RNA mobility shift assay, indicating that this domain is sufficient for RNA recognition (3–5, 7, 24, 36, 37). *In vitro* studies with synthetic peptides and transient transfection assays using mutant Tat expression vectors have shown that the basic region can be effectively substituted by heterologous basic domains, by scrambled variants of the natural domain, and by homopolymers of arginines (3, 7, 33, 34). However, a minimum of 5 basic amino acids is necessary for *trans* activation *in vivo* (7).

In addition to the requirements for a minimal basic charge, the positioning of the basic amino acids is also important. Substitution of two arginine residues at positions 52 and 53 of the basic domain strongly affects *trans* activation *in vivo* and complex formation *in vitro* (3, 4, 7, 17). The introduction of an arginine at either of these positions in a nonfunctional mutant Tat protein, containing a lysine stretch as the basic domain, restored *trans*-activation activity (4). On the basis of ethylation interference studies, it has been proposed that one arginine residue at position 52 or 53 mediates a specific interaction with TAR by forming hydrogen bonds with two phosphates, those between G21-A22 and A22-U23 (4, 11a). Such an interaction requires that the two phosphates be

* Corresponding author.

spaced more widely than in a typical RNA helix, but it does not directly predict why U23 (the uridine at position 23) itself should be essential. One possibility is that U23 is involved in a tertiary interaction required to stabilize a preferred structure.

In this study we identified specific nucleotides in the stem structure of TAR which are critical for complex formation with Tat and for Tat-mediated *trans* activation. On the basis of these results, a model for TAR tertiary structure is proposed which implicates the essential uridine residue (U23) in a stable hydrogen bond interaction with the essential GC pair at positions 26 and 39. We further show that the requirements for the second and third nucleotides in the bulge can be met by nonnucleotide spacers, demonstrating that neither the bases nor the ribose moieties of these nucleotides are functionally important for the recognition of TAR RNA by Tat.

MATERIALS AND METHODS

Modelling of TAR tertiary structure. Quanta Rev 2.1 and CHARMM Rev 21.2 were from Polygen Corp. CHARMM was updated every 20 cycles, and the smoothing of non-bonded and electronic interactions to zero was switched on at 11.0 Å (1.1 nm) and switched off at 18.0 Å (1.8 nm). The dielectric constant, without an explicit water shell, was set to a distance-dependent constant of 1.0. Explicit hydrogen bonds were turned off. The filtered, torsional angle search was based on a $\pm 25^\circ$ search window of the "back-side" angles around O3-P, P-O5, O5-C5, and C5-C4 between C39 and U40 and of the angles around O3-P, P-O5, O5-C5, C5-C4, and C1-N1 of U23. The initial starting position was canonical A-form of the lower stem plus U23 in a stacked position. The residues of the upper stem were locked into A-form helix for the research routine. A total of 100 trial structures were generated. Minimization was achieved with the Adopted Basis Set Newton-Raphson algorithm (ABNR). The bases in the triplex were initially set with a 10-kcal/mol harmonic constraint and target constraints between the triplex partners for 100 cycles of ABNR. This was followed by 500 cycles of unconstrained ABNR minimization. TIP3 water was then placed into the triplex region and allowed to minimize into a stable configuration.

In vitro transcription and RNA mobility shift assay. RNAs comprising nucleotides +18 to +44 of TAR were prepared from oligodeoxynucleotide templates by in vitro transcription using T7 RNA polymerase as described previously (19), except that the GTP concentration was raised to 0.5 mM. The transcripts were purified on an 8% polyacrylamide-8 M urea gel, eluted in a solution containing 0.5 M ammonium acetate, 1 mM EDTA, and 0.1% sodium dodecyl sulfate, and recovered by phenol-chloroform extraction and ethanol precipitation. Complex formation between Tat and TAR RNA was analyzed as described previously (24), except that the glycerol concentration in the incubation reaction was raised to 5%. Briefly, 32 P-labeled TAR RNA (2,000 cpm, 2.94×10^6 cpm/ μ g) was incubated with 100 ng (0.4 μ M) of purified Tat (see Fig. 2A, B, and D) or Tat peptide 49-62 (see Fig. 2C) at 30°C for 10 min. Complexes were analyzed on a 1.5-mm-thick 5% polyacrylamide gel (30:0.8, acrylamide-bisacrylamide) containing 5% glycerol. Electrophoresis was carried out in 0.5 $\times$ Tris-borate-EDTA (TBE) at a constant current of 30 mA for 2.5 h at 4°C. The interaction between Tat peptide 49-62 and synthetic TAR RNA (see Fig. 4) was performed similarly, except that 5,000 cpm of RNA was used in the assay.

Synthesis of linker-TAR RNA (ST61RCC) containing a propyl phosphate linker at positions 24 and 25. Oligoribonucleotides were prepared according to the procedure published previously (35). The modified phosphoramidite for synthesizing the sequence ST61RCC was prepared as follows: 1,3-propanediol (Aldrich) was first alkylated with 4,4'-dimethoxytrityl chloride in anhydrous pyridine (16). *O*-Dimethoxytrityl-1,3-propanediol was then phosphorylated with 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (Aldrich) in dry tetrahydrofuran (29). The product was purified by flash chromatography on silica gel. To synthesize the modified TAR analog ST61RCC, this modified phosphoramidite was dissolved in dry acetonitrile (0.2 to 0.3 M) and coupled into the oligonucleotide, using the standard synthesis cycle. The average coupling yield, as assayed by trityl measurement, was 95 to 97%.

In vitro mutagenesis and transfections. Mutations were introduced by use of an in vitro mutagenesis system (Amersham) based on the method of Sayers et al. (27). HeLa cells or HeLa cells stably expressing Tat were plated at 10^6 cells in 10-cm-diameter petri dishes one day before transfection. pU3RIII (2 μ g) (30) was transfected using the DEAE-dextran method (15). Cell extracts were prepared 40 to 48 h after transfection by freeze-thawing three times, heating 7 min at 60°C, and centrifuging in a microfuge for 5 min. Chloramphenicol acetyltransferase (CAT) assays were performed with 5 μ g of HeLa-Tat cell extract or 50 μ g of HeLa cell extract for 1 h at 37°C as described previously (14). The percentage of acetylated [14 C]chloramphenicol was determined by scintillation counting of the spots cut out from the thin-layer chromatographic plate. Assays giving under 1% or over 70% acetylation were repeated with more or less extract. Transfections were repeated at least three times. Fold *trans* activation was defined as the ratio of CAT activity (corrected for the amount of protein used in the assay) obtained in HeLa-Tat cells to that in the parental cells.

RESULTS

Complex formation between Tat and TAR RNA requires specific nucleotides above the pyrimidine bulge. To investigate possible base targets for an interaction with U23, we made nucleotide substitutions in TAR which conserve base pairing but change sequence in each of the three base pairs above the bulge (Fig. 1). In vitro-transcribed T7-derived TAR RNAs were tested for complex formation with purified Tat protein by RNA mobility shift assay. The substitution of CG or AU for GC26/39 completely abrogated Tat binding (Fig. 2A, lanes 3 and 4 [TAR 43] and lanes 5 and 6 [TAR 44]). Replacing AU27/38 by either GC or UA also abolished the binding of Tat (lanes 7 and 8 [TAR 45] and lanes 9 and 10 [TAR 46]). These results demonstrate that highly specific base pairs immediately above the bulge are essential for the interaction with Tat, as other purine-pyrimidine combinations led to the loss of Tat binding. In contrast, nucleotide substitutions in the third base pair above the bulge (GC28/37 to AU [lanes 11 and 12; TAR 55] or to CG [lanes 13 and 14; TAR 56]) had no effect on complex formation, suggesting that this base pair is not critical for Tat-TAR interaction. This finding is in accordance with a recent report showing that the substitution of CG for GC28/37 does not significantly affect Tat binding (8). Further evidence that sequence requirements at this position are rather flexible are provided by studying the interaction between Tat and HIV-2_{ROD} TAR. One of the two stem-loops of HIV-2_{ROD} TAR which is bound and *trans* activated by HIV-1 Tat (8, 10) is identical to HIV-1

FIG. 1. Sequence of TAR mutants. Nucleotide substitutions are indicated by boxes. RNAs were transcribed *in vitro* by T7 RNA polymerase from oligodeoxynucleotide templates as described in Materials and Methods.

TAR from positions 20 to 40 (bulge and loop region) except for an additional GC pair and a GU pair at positions 28 and 37.

To detect possible low-affinity interactions between Tat and mutant TAR RNAs, we analyzed the binding of small synthetic peptides at high molar excess of peptide over RNA. Complex formation between Tat peptide 49-62 (amino acid sequence RKKRRQRRPPQNS) and wild-type TAR, TAR 43, or TAR 44 was examined at a 5-, 10- or 50-fold molar excess (Fig. 2B). At a fivefold molar excess of peptide, about 50% of wild-type TAR RNA is bound, while no complex formation could be detected with TAR 43 or TAR 44 (compare lane 2 with lanes 6 and 10). All of the peptide is bound to wild-type TAR RNA at 10- and 50-fold molar excess of Tat 49-62 (lanes 3 and 4). However, the mutant TAR RNAs do not bind the peptide even at high concentrations (lanes 7 and 8 and lanes 11 and 12), emphasizing that the GC pair at positions 26 and 39 is essential for Tat-TAR complex formation.

The mutational analysis of the two base pairs immediately above the bulge suggests their participation in intramolecular hydrogen bonding with additional bases. We therefore wished to determine whether one or both bases in each of these pairs are critical for the interaction with Tat. Uridine residues were substituted for C39 (TAR 52) and C37 (TAR 54), and adenosine 27 was replaced by guanosine (TAR 53)

(Fig. 2D). We observed two structures of TAR RNA for mutants TAR 52 (GU26/39) and TAR 53 (GU27/38). The slower-migrating RNA species is possibly a TAR RNA dimer. The faster-migrating monomer species does not complex with Tat (lanes 3 and 4 and lanes 5 and 6, respectively). The slower-migrating species of TAR 53 becomes diffuse upon addition of Tat. This was reproducible, but as there is no distinct shift, we attribute this to nonspecific interaction. In contrast to the previous mutants, binding was maintained when GU was substituted for GC at positions 28 and 37, the third base pair above the bulge (lanes 7 and 8 [TAR 54]). These results are in accordance with the data obtained from substituting both nucleotides of the base pairs and provide evidence that at least C39 and A27 are critical for complex formation between Tat and TAR RNA.

Mutation of nucleotides below the pyrimidine bulge does not interfere with the binding of Tat. Mutant TAR RNAs containing substitutions in the three base pairs directly below the bulge were also examined in the RNA mobility shift assay (Fig. 2C). Substitution of UA or GC for AU22/40 did not affect complex formation with Tat (lanes 3 and 4 [TAR 36] and lanes 5 and 6 [TAR 37]). Interaction of Tat with TAR RNA containing a substitution of CG for GC21/41 decreased binding by approximately fivefold (lanes 7 and 8 [TAR 38]), while the substitution of AU for GC21/41 (lanes 9 and 10 [TAR 39]) was comparable to that of wild-type TAR. Re-

FIG. 2. Effect of mutations in nucleotides above and below the pyrimidine bulge of TAR on complex formation with Tat or Tat peptide 49-62 (RKKRRQRRPPQNS). 32 P-labeled wild-type (wt) or mutant TAR RNAs were incubated in the presence or absence of 100 ng ($0.4 \mu\text{M}$) of protein (or Tat peptide 49-62 in panel B) and complex formation was analyzed on a nondenaturing, 5% polyacrylamide gel as described in Materials and Methods. (A) Effect of mutations in base pairs above the pyrimidine bulge of TAR on complex formation with Tat. (B) Complex formation between 5-, 10-, or 50-fold molar excess of Tat peptide 49-62 and wild-type (wt) or mutant TAR RNA. (C) Effect of mutations in base pairs below the pyrimidine bulge of TAR on complex formation with Tat. (D) Effect of substitution of a GU pair at positions 26 and 39 (TAR 52), 27 and 38 (TAR 53), or 28 and 37 (TAR 54).

placing AU20/42 with UA had no effect on binding (lanes 11 and 12 [TAR 47]), while substituting GC increased complex formation (lanes 13 and 14 [TAR 48]). This might be due to increased stabilization of the RNA secondary structure induced by the formation of a stretch of four GC pairs in the shortened TAR stem. Taken together, these results show that the primary sequence in the stem below the 3-nucleotide bulge is less critical for complex formation than the sequence above the bulge.

Substitution of nucleotides above the pyrimidine bulge strongly affects Tat-mediated *trans* activation. We next examined the correlation between effects of mutations above the pyrimidine bulge on complex formation and Tat-mediated *trans* activation in vivo (Fig. 3). Nucleotide substitutions were introduced in vector pU3RIII, which contains the reporter gene for the bacterial CAT gene under the control of the HIV-1 long terminal repeat (30). The effect of the mutations on *trans* activation was determined by measuring CAT activity after transient transfection of pU3RIII into HeLa cells stably expressing Tat. Substitution of CG or AU for GC26/39 completely abolished *trans* activation (lanes 3 and 4 [TAR 43] and lanes 5 and 6 [TAR 44], respectively). The replacement of AU27/38 by UA (lanes 7 and 8 [TAR 45]) or GC (lanes 9 and 10 [TAR 46]) reduced *trans* activation by 10-fold. Furthermore, a TAR mutant substituting G for A26, and thereby introducing a GU pair at positions 26 and 39 (equivalent to TAR 52), was not *trans* activated by Tat (data not shown). In agreement with the in vitro data, all of the mutations in the two base pairs immediately above the bulge strongly affected *trans* activation. However, substitution of UA or GC for AU27/38 which completely prevented complex formation in vitro still retained 10% activity. Residual *trans*-activation activity of TAR mutants that are deficient in complex formation has previously been observed (8, 24) and might be due to stabilization of the Tat-TAR interaction by

cellular factors that bind to the loop sequence of TAR RNA (12, 13, 17).

A propyl phosphate linker can functionally substitute for nucleotides 24 and 25 in the bulge of TAR. We previously reported that TAR RNA containing only one bulged uridine residue does not form a stable complex with Tat, but that a minimum of two nucleotides is required. In accordance with these findings and the results presented here, we predicted that the nucleotides at positions 24 and 25 of the bulge function only to fill in the gap between U23 and G26 to

FIG. 3. Effect of mutations in base pairs above the pyrimidine bulge of TAR on Tat-mediated *trans* activation. HeLa cells (-) or HeLa cells stably expressing Tat (+) were transiently transfected with 2 μg of wild-type (wt) or mutant pU3RIII. Cell extracts were prepared and CAT activity was determined as described in Materials and Methods. *trans*-activation activity of the different mutants was calculated as an average of at least three independent transfection experiments.

FIG. 4. (A) Structure of linker-TAR RNA. (B) Complex formation between linker-TAR RNA substituting a propyl phosphate linker for nucleotides 24 and 25 in the bulge, and Tat peptide 49-62. RNAs were prepared as described in Materials and Methods. Wild-type (wt) or linker-TAR RNA was incubated with 1-, 10-, 100-, or 500-fold molar excess of Tat peptide 49-62.

stabilize the TAR structure. To determine whether nucleotides as such are dispensable at these positions, a TAR analog (ST61RCC) was made in which nucleosides C24 and U25 of the bulge were replaced by equivalent-length propyl groups $[-(\text{CH}_2)_3-]$, while retaining the intervening phosphates. Complex formation was analyzed using Tat peptide 49-62 in the RNA mobility shift assay (Fig. 4). The interaction between wild-type (lanes 1 to 5) and linker TAR RNA (lanes 6 to 10) was tested at peptide concentrations from 1- to 500-fold molar excess. The peptide bound wild-type and modified RNA with comparable affinities (compare lanes 2 to 5 and 7 to 10). Quantitative analyses determined similar dissociation constants (10^{-10} M) for the interaction between peptide 49-62 and the two RNAs (4a). The finding that a nonnucleotide linker can functionally substitute for residues 24 and 25 of TAR supports the notion that these nucleosides do not engage in base pairing or other hydrogen-bonding interactions. It also demonstrates that a ribose at these positions is not required for TAR recognition by Tat. Bulges of two rather than three nucleotides exist in some HIV-1 isolates and in HIV-2. We found that in TAR constructs containing only two bulged residues the second nucleoside could be replaced by a single propyl group (data not shown).

DISCUSSION

The results presented in this study show that nucleotide substitutions in the two base pairs above the pyrimidine bulge of TAR abolish complex formation with Tat. These results are in agreement with a recent report by Weeks and Crothers (37). The biological significance of the binding results is illustrated by the finding that the same mutations strongly decrease Tat-mediated *trans* activation. In contrast, base substitutions below the bulge had a weak effect on complex formation, suggesting that these nucleotides are not directly implicated in the binding of Tat.

On the basis of the results presented here and elsewhere (3, 4, 7, 8, 24, 35, 36), it is likely that a tertiary RNA structure is required for the interaction between TAR and an arginine in the basic domain of Tat. As there is no evidence

yet for an interaction between an amino acid and the essential uridine residue at position 23, it is plausible that U23 is engaged in hydrogen bonding with other residues in TAR to engender a critical tertiary structure that is required for Tat binding. U differs from C, an unacceptable substitute, only at the N3 and C4 positions, implicating the hydrogen-bonding surfaces of U23 in a critical interaction. Indeed, we have shown previously that methylation of the N3 position of U23 makes TAR an unacceptable substrate for high-affinity TAR binding (35). Therefore, it is likely that U23 is interacting with another part of TAR through one or more hydrogen bonds. The potential sites with which U23 can form hydrogen bonds in the TAR structure are bases in the major or minor helical grooves and the phosphate oxygen atoms. However, only one hydrogen bond could be made between N3 of U23 and a phosphate oxygen on an opposite strand. Such an interaction has never been observed in any X-ray- or nuclear magnetic resonance-determined oligonucleotide structure. We therefore considered possible interactions of U23 with the GC and AU pairs immediately above the bulge. We modelled a stable hydrogen bond network between U23 and the conserved GC pair above the bulge (Fig. 5; our model does not take into consideration any possible effects of Tat interaction on the stability of TAR structure).

The interaction between U23 and GC26/39 was modelled into a canonical A-form RNA duplex through an exhaustive, randomized search of the torsional angle space of U23 and the backbone torsional angles between CU39/40 of the strand opposite the pyrimidine bulge (Fig. 5A). Any structures identified in the search which placed the hydrogen-bonding surface of U23 into close proximity with the hydrogen-bonding surface of the GC26/39 pair were energy minimized under the CHARMM force field. A single, stable triplexlike structure emerged (Fig. 5B). A bifurcated water molecule is proposed to fill and stabilize the third potential hydrogen-bonding site between U23 and G26. This structure can be visualized as a widening of the narrow major groove of the RNA duplex. U23 remains stacked over A22 in a somewhat distorted fashion and interacts simultaneously

FIG. 5. (A) Schematic model of the backbone atoms of the proposed three-dimensional structure of the TAR RNA stem-loop. The 5' strand (blue) is connected to the 3' strand (green) by a hexanucleotide loop (not modeled) at the location of the arrows. The core of the triplex structure is composed of the residues U23 (magenta), G26 (white), and C39 (yellow). The two remaining bulge residues (CU24/25 [orange]) are shown looped out into solution. (B) Detail of the proposed triplex-core hydrogen bond network. Hydrogen bonds are indicated by three arrowheads pointing from donor to acceptor. Atoms are colored as follows: oxygen in red, nitrogen in blue, phosphorus in green, and carbon and hydrogen in white.

with both bases of the GC base pair. This unusual configuration is in contrast to the traditional Hoogsteen pairing in which the third base interacts with only one partner of a base pair.

Our model differs from that of Weeks and Crothers (37) who have proposed only tertiary stacking interactions. In the model presented here, the hydrogen-bonding interactions between U23 and the GC26/39 pair would lock the stem-loop into a single structure leaving a gap of about 7 Å (0.7 nm) between the backbone phosphates to be filled by additional nucleotides of the bulge. For a 2-nucleotide bulge, the second nucleotide would effectively bridge this gap without distortion of the regular backbone structure. For a bulge of 3 or more nucleotides, some distortion of the backbone would occur as the nucleotides are thrust into solution rather than being stacked into the core. The model predicts that a unique, kinked stem structure is obtained with a bulge size of 2 and more bases. A single uridine residue in the bulge would not be able to form a triplex structure and achieve closure of the gap. This is in agreement with previous studies demonstrating that TAR RNA containing a bulge of 2 to 5 nucleotides, not a 1-nucleotide bulge, efficiently interacts with Tat (35).

We further show that a synthetic TAR RNA in which a propyl phosphate linker substitutes for nucleotides 24 and 25 in the bulge binds Tat peptide with an affinity similar to that of wild-type TAR RNA. This supports the notion proposed in our model that nucleotides 24 and 25 as such are not important but adopt a backbonelike configuration which stabilizes the kinked TAR tertiary structure. Previous phylogenetic comparison of the TAR region of different HIV strains (21) shows some sequence flexibility (either a uridine or cytosine residue) at these positions. Furthermore, extensive mutational analysis revealed that all bases can be substituted for position 24, and position 25 can be deleted without affecting complex formation or *trans*-activation activity of TAR (1, 8, 24, 35).

The narrow major groove of RNA A-form helix is widened around the bulge in both our model and that of Weeks and Crothers (37). This opening creates a cleft of sufficient size for the basic domain of Tat to bind within. Since the interaction with Tat seems to be primarily electrostatic (3, 4, 7, 35), the resulting disposition of the phosphates on the edge of the groove may be particularly critical for maximal binding and specificity. The disposition of the two phosphates on either side of A22 is important for the hydrogen bonding network proposed in the "arginine fork" model (4, 11a).

Our model does not explain the preference for AU27/38, which is also essential for the interaction between Tat and TAR RNA *in vitro* and *in vivo*. The conformational search performed in developing the model did not identify any possible, stable tertiary interactions which this base pair may participate in. Possibly, one or both nucleotides of this AU pair are implicated in an interaction with an arginine of the basic domain and are therefore necessary for TAR recognition by Tat.

Interactions between RNA and proteins occur mostly in the large, accessible minor groove (32), although, for example, recognition of tRNA^{Asp} by aspartyl tRNA-synthetase involves contact sites in the major groove (26). Interestingly, these interactions implicate bases that are located at junctions between helical double-stranded structures and single-stranded regions of tRNA and therefore are more accessible to the protein. Thus, widening of the major groove of A-form RNA at junctions between double-stranded and single-stranded regions might be a common feature for protein recognition.

ACKNOWLEDGMENTS

We thank Paul Schimmel for critical comments on the manuscript, Craig Rosen for generously providing purified Tat protein, Raya Kuperman for synthesis of oligonucleotides, Wing Lin for binding analysis of the linker-TAR RNA, and Sophie Roy for helpful discussions.

This work was supported by grants (to N.S.) from the Health and Welfare Canada-National Health Research and Development Program and the Cancer Research Society (Montreal). U.D. is the recipient of a studentship from the Medical Research Council of Canada.

REFERENCES

- Berkhout, B., and K.-T. Jeang. 1989. *trans* Activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the *trans*-acting-responsive hairpin: a quantitative analysis. *J. Virol.* **63**:5501-5504.
- Berkhout, B., R. H. Silverman, and K.-T. Jeang. 1989. Tat *trans*-activates the human immunodeficiency virus through a nascent RNA target. *Cell* **59**:273-282.
- Calnan, B. J., S. Biancalana, D. Hudson, and A. D. Frankel. 1991. 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., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel. 1991. Arginine-mediated RNA recognition: the arginine fork. *Science* **252**:1167-1171.
- Climie, S. Unpublished data.
- Cordingley, M. G., R. L. LaFemina, P. L. Callahan, J. H. Condra, V. V. Sardana, D. J. Graham, T. M. Nguyen, K. LeGrow, L. Gotlib, A. J. Schlabach, and R. J. Colonna. 1990. Sequence-specific interaction of Tat protein and Tat peptides with the *trans*-activation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. USA* **87**:8985-8989.
- Cullen, B. R. 1990. The HIV-1 Tat protein: an RNA sequence-specific processivity factor? *Cell* **63**:655-657.
- Delling, U., S. Roy, M. Sumner-Smith, R. Barnett, L. Reid, C. A. Rosen, and N. Sonenberg. 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., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, and M. A. Skinner. 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.
- Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, M. Skinner, and R. Valerio. 1989. Human immunodeficiency virus 1 tat protein binds *trans*-activation-responsive region (TAR) RNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **86**:6925-6929.
- Emerman, M., M. Guyader, L. Montagnier, D. Baltimore, and M. A. Muesing. 1987. The specificity of the human immunodeficiency virus type 2 *trans*-activator is different from that of human immunodeficiency virus type 1. *EMBO J.* **6**:3755-3760.
- Feng, S., and E. C. Holland. 1988. HIV-1 tat *trans*-activation requires the loop sequence within tat. *Nature (London)* **334**:165-167.
- Frankel, A., and J. Tao. 1992. The arginine fork. *Science* **255**:665. (Author's correction.)
- Gatignol, A., A. Kumar, A. Rabson, and K.-T. Jeang. 1989. Identification of cellular proteins that bind to the human immunodeficiency virus type 1 *trans*-activation-responsive TAR element RNA. *Proc. Natl. Acad. Sci. USA* **86**:7828-7832.
- Gaynor, R., E. Soultanakis, M. Kuwabara, J. Garcia, and D. S. Sigman. 1989. Specific binding of a HeLa cell nuclear protein to RNA sequences in the human immunodeficiency virus transactivating region. *Proc. Natl. Acad. Sci. USA* **86**:4858-4862.
- Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
- Grosschedl, R., and D. Baltimore. 1985. Cell type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* **41**:885-897.
- Jones, R. A. 1984. Preparation of protected deoxyribonucleosides. In M. J. Gait (ed.), *Oligonucleotide synthesis: a practical approach*. IRL Press, Oxford.
- Marciniak, R. A., B. J. Calnan, A. D. Frankel, and P. A. Sharp. 1990. HIV-1 Tat protein *trans*-activates transcription *in vitro*. *Cell* **63**:791-802.
- Marciniak, R. A., M. A. Garcia-Blanco, and P. A. Sharp. 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.
- Meerovitch, K., J. Pelletier, and N. Sonenberg. 1989. A cellular protein that binds to the 5'-noncoding region of poliovirus RNA: implications for internal translation initiation. *Genes Dev.* **3**:1026-1034.
- Muesing, M. A., D. H. Smith, and D. J. Capon. 1987. Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. *Cell* **48**:691-701.
- Myers, G. 1990. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
- Pavliakis, G. N., and B. K. Felber. 1990. Regulation of expression of human immunodeficiency virus. *The New Biol.* **2**:20-31.
- Rosen, C. A. 1991. Regulation of HIV gene expression by RNA-protein interactions. *Trends Genet.* **7**:9-14.
- Roy, S., U. Delling, C.-H. Chen, C. A. Rosen, and N. Sonenberg. 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated *trans*-activation. *Genes Dev.* **4**:1365-1373.
- Roy, S., N. T. Parkin, C. Rosen, J. Itovitch, and N. Sonenberg. 1990. Structural requirements for *trans*-activation of human immunodeficiency virus type 1 long terminal repeat-directed gene expression by *tat*: importance of base pairing, loop sequence, and bulges in the *tat*-responsive sequence. *J. Virol.* **64**:1402-1406.
- Ruff, M., S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J. C. Thierry, and D. Moras. 1991. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA^{ASP}. *Science* **252**:1682-1689.
- Sayers, J. R., W. Schmidt, and F. Eckstein. 1988. 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res.* **16**:791-802.
- Selby, M. J., E. S. Bain, P. A. Luciw, and B. M. Peterlin. 1989. Structure, sequence, and position of the stem-loop in *tar* determine transcriptional elongation by *tat* through the HIV-1 long terminal repeat. *Genes Dev.* **3**:547-558.
- Sinha, N. D., J. Biernat, J. McManus, and H. Koster. 1984. Polymer support oligonucleotide synthesis XVIII: use of β -cyanoethyl-N,N-dialkylamino-/N-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.* **12**:4539-4557.
- Sodroski, J., C. Rosen, F. Wong-Staal, S. Z. Salahuddin, M. Popovic, S. Arya, R. C. Gallo, and W. A. Haseltine. 1985. *trans*-Acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. *Science* **227**:171-173.
- Steffy, K., and F. Wong-Staal. 1991. Genetic regulation of human immunodeficiency virus. *Microbiol. Rev.* **55**:193-205.
- Steitz, T. A. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence specific binding. *Q. Rev. Biophys.* **23**:205-280.
- Subramanian, T., R. Govindarajan, and G. Chinnadurai. 1991. Heterologous basic domain substitutions in the HIV-1 Tat protein reveal an arginine-rich motif required for *trans*-activation. *EMBO J.* **10**:2311-2318.
- Subramanian, T., M. Kuppuswamy, L. Venkatesh, A. Srinivasan, and G. Chinnadurai. 1990. Functional substitution of the basic domain of the HIV-1 *trans*-activator, Tat, with the basic domain of the functionally heterologous Rev. *Virology* **176**:178-183.
- Sumner-Smith, M., S. Roy, R. Barnett, L. S. Reid, R. Kuper-

- man, U. Delling, and N. Sonenberg. 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.
36. Weeks, K. M., C. Ampe, S. C. Schultz, T. A. Steitz, and D. M. Crothers. 1990. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **249**:1281–1285.
37. Weeks, K. M., and D. M. Crothers. 1991. RNA Recognition by Tat-derived peptides: interaction in the major groove? *Cell* **66**:577–588.
38. Wu, F., J. Garcia, D. Sigman, and R. Gaynor. 1991. Tat regulates binding of the human immunodeficiency virus *trans*-activating region RNA loop-binding protein TRP-185. *Genes Dev.* **5**:2128–2140.