Specific binding of arginine to TAR RNA

(RNA recognition/RNA structure/RNA-binding peptides/human immunodeficiency virus Tat protein)

JIANSHI TAO AND ALAN D. FRANKEL

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142

Communicated by Stephen C. Harrison, December 31, 1991 (received for review November 5, 1991)

ABSTRACT A single arginine residue within the basic region of the human immunodeficiency virus Tat protein mediates specific binding of Tat peptides to a three-nucleotide bulge in TAR RNA. It has been proposed that arginine recognizes TAR by forming a network of hydrogen bonds with two structurally distinct phosphates, an interaction termed the "arginine fork." Here it is shown that 1-arginine blocks the Tat peptide/TAR interaction, whereas L-lysine and analogs of arginine that remove specific hydrogen bond donors do not. Experiments using an L-arginine affinity column demonstrate that arginine and the Tat peptides bind to the same site in TAR. Modification of two phosphates located at the junction of the double-stranded stem and bulge and modification of two adenine N7 groups in base-paired regions of TAR interfere with specific arginine binding. The results emphasize the importance of RNA structure in RNA-protein recognition and provide methods to identify arginine-binding sites in RNAs.

The importance of RNA structure in sequence-specific RNA–protein recognition is becoming increasingly apparent. Biochemical studies of the R17 phage coat protein and its RNA-binding site were among the first to suggest a direct role of RNA structure in the interaction (1, 2). Subsequent structural studies of glutaminyl- and aspartyl-tRNA synthetase–tRNA complexes have highlighted its importance (3–5). RNA structure is also critical in the interaction of the human immunodeficiency virus (HIV) transcriptional activator, Tat, with TAR, an RNA stem–loop located at the 5′ end of viral mRNAs. Studies with Tat peptides have shown that the Tat/TAR interaction is mediated by a short (nine amino acid) region of basic amino acids (6–9). A single arginine residue in the peptide provides the only sequence-specific contact with the RNA (10). Chemical modification experiments identified two phosphates in TAR, located at the junction of the double-stranded stem and a three-nucleotide bulge, that are important in this interaction (10), suggesting that specificity may be derived largely from recognition of a defined backbone conformation of the RNA. Arginine may recognize this structure by forming a specific network of hydrogen bonds with the two phosphates, an interaction termed the "arginine fork" (10). Chemical modification experiments also suggest that N7 groups of two base-paired adenesines, one located above the bulge and one located below the bulge, are important in the interaction (6, 11).

Because only one arginine in the Tat peptides mediates specific recognition of TAR, it seemed plausible that the free amino acid arginine might also bind specifically to TAR. Here we show that 1-arginine does indeed bind specifically to TAR and that the arginine-binding site in TAR requires the same two phosphates and adenine N7 groups that are needed for Tat peptide binding. The results suggest that TAR RNA folds into a specific conformation containing a single arginine-binding site and emphasize the importance of RNA structure in an RNA–protein interaction.

MATERIALS AND METHODS

Gel-Shift Assays of RNA Binding. Peptide binding to TAR was measured by RNA gel-shift analysis using in vitro-transcribed 31-nucleotide TAR RNA as described (9, 10). Binding inhibition by 1-arginine, L-lysine, or L-arginine analogs was measured by adding several concentrations of each compound to reaction mixtures containing peptide and RNA under conditions that gave approximately 50–70% binding in the absence of inhibitor. Binding reactions were carried out for 30 min on ice. To determine Ki values, peptide-bound RNA and free RNA were quantitated with a β-scanner and the ratios of bound to free, relative to no competitor, were calculated. Ki is defined as the concentration of competitor required to decrease the fraction bound by 50%. A minimum of three sets of data were used to calculate each Ki.

RNA Binding to an L-Arginine-Agarose Column. A mixture of 10 μg each of in vitro-transcribed wild-type TAR (5P-labeled), TAR with a U23 → C substitution in the bulge (or an A27/U38 to U:A base pair substitution), and TAR with a deletion of the three-nucleotide bulge was bound to an L-arginine-agarose or an L-lysine-agarose column (1.0 ml; Sigma) equilibrated in 10 mM Tris·HCl, pH 7.5/70 mM NaCl/0.2 mM EDTA. RNAs were eluted at 4°C with a 70–500 mM NaCl gradient (150 ml), and fractions (1.0 ml) were collected and analyzed by 15% polyacrylamide gels. Wild-type and mutant TAR RNAs were transcribed and purified as described (9, 10).

Interference by Phosphate Ethylation or Adenine Carbethoxylation. TAR RNA was labeled with 32P at the 5′ end with T4 polynucleotide kinase, and RNA phosphates were ethylated with ethylnitrosourea (ENU) or adenine N7 groups were carbethoxylated with diethyl pyrocarbonate (DEPC), under denaturing conditions as described (10, 12). Modified RNAs (2 × 106 cpm) were loaded onto an L-arginine column and eluted with a salt gradient, as above. Fractions were collected and the modified RNAs were cleaved (10, 12) and analyzed on 20% polyacrylamide/8 M urea sequencing gels.

RESULTS

Inhibition of Tat Peptide Binding by Arginine. Previous studies had shown that short peptides containing the basic region of Tat bind specifically to TAR (6–9) and that a single arginine provides the only sequence-specific contact (10). To further examine the structural details of the interaction, we asked whether L-arginine, L-lysine, or L-arginine analogs would compete with Tat peptides for specific binding to TAR. Because only one arginine is involved in sequence-specific binding and because lysine is unable to substitute (10), it seemed reasonable that arginine would compete for binding whereas lysine or analogs of arginine that disrupt essential hydrogen bonds would not. To simplify the analysis, we measured inhibition of binding of R52, a peptide containing a

Abbreviation: DEPC, diethyl pyrocarbonate.
single arginine within a stretch of lysines (Tyr-Lys-Lys-Lys-Arg-Lys-Lys-Lys-Lys-Ala) that binds to TAR with the same affinity and specificity as the wild-type Tat 49-57 peptide (Tyr-Arg-Lys-Lys-Arg-Gln-Arg-Arg-Arg-Ala) (10). Using a gel shift assay, we found that binding of R52 was strongly inhibited by L-arginine but was only weakly inhibited by L-lysine (Fig. 1). Similar inhibition results were obtained with the wild-type peptide (data not shown). The inhibition constant, $K_i$, for L-arginine was 4 mM, whereas the $K_i$ for L-lysine was $>50$ mM (Table 1). This compares with Tat peptide and Tat protein binding constants of 6–12 nM (9, 13).

Arginine analogs were tested as inhibitors of binding (Table 1). The $K_i$ values for L-argininamide and agmatine, which have blocked or deleted COO$^-$ groups, were $\sim$1/4 of those for L-arginine, probably reflecting removal of an unfavorable electrostatic interaction with the RNA backbone. In contrast, $N^G$-monomethylarginine, which has one methylated $\eta$ N but retains the positive charge of the guanidinium group, has a significantly higher $K_i$ ($>50$ mM) than L-arginine. Methylation of the $N^G$-amino group eliminates one potential hydrogen bond donor that may be required for the specific interaction with TAR. No inhibition was detectable with L-citrulline, which removes one $\eta$ N and also eliminates the positive charge of the side chain. Thus, both favorable electrostatic interaction and an array of hydrogen bond donors seem to be required for recognition of TAR, consistent with the “arginine fork” model.

**Specific Binding of TAR to an L-Arginine Affinity Column.** To test whether arginine and the Tat peptides bind to the same site in TAR and to confirm the specificity for arginine, we asked whether TAR, or mutants of TAR, would bind to L-arginine or L-lysine affinity columns. A mixture of three RNAs—wild-type TAR, TAR with a U → C substitution in the bulge, and TAR with a deletion of the bulge—were bound to an L-arginine-agarose column and eluted with a salt gradient. Wild-type TAR was labeled with $^{32}$P to distinguish it from the U → C mutant; these 31-nucleotide RNAs could be resolved from the 28-nucleotide deletion mutant by gel electrophoresis. All three RNAs bound to the arginine column at the loading salt concentration (70 mM NaCl) and then eluted in the following order: bulge deletion ($\sim 120$ mM), U → C substitution ($\sim 140$ mM), and wild-type TAR ($\sim 165$ mM) (Fig. 2). Thus, the binding affinity of each RNA for the arginine column mimics the affinity and specificity of Tat peptide binding: wild-type TAR > bulge substitution > bulge deletion (9, 11). In contrast, all three RNAs eluted at the same salt concentration from an L-lysine-agarose column ($\sim 90$ mM NaCl; data not shown), consistent with nonspecific binding. These results demonstrate that arginine is sufficient for specific recognition of TAR, confirm the specificity for arginine over lysine, and suggest that arginine in the peptide and free L-arginine bind to the same site in TAR.

**Ethylation Interference with the L-Arginine Column.** To identify possible phosphate contacts with L-arginine and to further compare the arginine-binding site to the Tat peptide-binding site in TAR, we performed an ethylation interference experiment with the L-arginine column. Phosphates in TAR RNA were ethylated (10), and the modified RNAs were bound to the column, eluted with a salt gradient, and analyzed. Modification of two phosphates located at the 5' end of the three-nucleotide bulge (between G21 and A22 and between A22 and U23) interfered with TAR binding to the

---

**Table 1.** Inhibition of Tat peptide binding to TAR by L-arginine, L-lysine, and L-arginine analogs

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>$K_i$, mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Arginine</td>
<td><img src="image1" alt="Structure" /></td>
<td>4</td>
</tr>
<tr>
<td>L-Lysine</td>
<td><img src="image2" alt="Structure" /></td>
<td>$&gt;50$</td>
</tr>
<tr>
<td>L-Argininamide</td>
<td><img src="image3" alt="Structure" /></td>
<td>1</td>
</tr>
<tr>
<td>Agmatine</td>
<td><img src="image4" alt="Structure" /></td>
<td>1</td>
</tr>
<tr>
<td>$N^G$-Monomethylarginine</td>
<td><img src="image5" alt="Structure" /></td>
<td>$&gt;50$</td>
</tr>
<tr>
<td>L-Citrulline</td>
<td><img src="image6" alt="Structure" /></td>
<td>No inhibition</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Specific inhibition of Tat peptide/TAR interaction by L-arginine. (A) R52 peptide (20 nM) was incubated with TAR RNA (2 nM) in the presence of the indicated concentrations of L-arginine (upper gel) or L-lysine (lower gel). The $K_i$ of R52 binding is 6 mM. Binding was determined by electrophoretic mobility shift analysis. (B) Inhibition curves with L-arginine (•) or L-lysine (○) were generated from the data in A. Peptide-bound RNA and free RNA were quantitated from three gels with a $\beta$ scanner and the ratios of bound to free, relative to no competitor, were calculated.
L-arginine column (Fig. 3). These are the same two phosphates required for Tat peptide binding to TAR (Fig. 3; ref. 10). [Note that these two phosphates were initially assigned incorrectly as the phosphates between A22 and U23 and between U23 and C24 (10); however, both pairs of phosphates are consistent with the computer modeling described previously (10).] Thus, the free amino acid arginine and the Tat peptide bind to the same site in TAR.

In addition to the striking interference observed at the two phosphates below the bulge, we observed some subtle enhancements at phosphates within the bulge. We noticed that there was a doublet of bands at each ethylated position (Fig. 3); the upper band is RNA that has retained the ethyl group after alkaline hydrolysis, whereas the lower band is RNA that has lost the labile ethyl group during hydrolysis (data not shown). We presume that deethylation occurs preferentially at one of the two phosphate oxygens (the pro-R or pro-S isomers) because their chemical reactivities, which depend on the precise geometry of each phosphate, are different. On the basis of this assumption, it appears that TAR binding to the arginine column is enhanced by ethylation of three particular phosphate oxygens in the bulge (Fig. 3), suggesting either that modification of these oxygens stabilizes the RNA structure recognized by arginine or that binding is enhanced due to electrostatic effects. Oxygen-specific enhancement is not seen in the context of the peptide (Fig. 3), suggesting that lysines surrounding the arginine may contact these oxygens and stabilize the arginine-binding RNA conformation or neutralize the charge. An "induced fit" model would be consistent with the changes in RNA conformation observed upon peptide binding (9, 14).

Adenine N7 Groups Involved in Arginine Recognition. Mutagenesis and chemical interference experiments have shown that the A27-U38 base pair is also important in the Tat peptide/TAR interaction (6, 11). To test whether free arginine can discriminate at this position, we prepared a mutant TAR RNA containing an A27-U38 to U:A base pair substitution and measured its binding to the L-arginine column. The A-U → U:A mutant eluted at the same salt concentration as the U → C bulge substitution mutant, between wild-type TAR and the bulge deletion mutant (data not shown; see Fig. 2). To further examine potential interactions with adenosines, we carbethoxylated the N7 groups by using DEPC (12) and performed a column interference experiment as above. Modification of A22 and A27 interfered with arginine binding (Fig. 4), exactly as seen with Tat peptide binding (6). This provides further evidence that free arginine occupies the same binding site in TAR as does arginine in the peptide.

DISCUSSION

The guanidinium group of the arginine side chain seems well suited for sequence-specific nucleic acid recognition. It is positively charged, providing a favorable electrostatic environment for interaction with nucleic acids, it can donate as many as five hydrogen bonds to appropriately positioned acceptor groups, and it has a rigid planar geometry that limits conformational entropy. In DNA recognition, arginine can form base-specific hydrogen bonds in the DNA major groove as well as electrostatic contacts with the backbone (15). In RNA recognition, arginine can form networks of hydrogen bonds with the sugar-phosphate backbone (3–5) and can interact directly with the bases (3–5, 16–18). In the Tat/TAR interaction, a single arginine appears to interact with two adjacent phosphates and may also contact groups on specific bases (for example, groups on U23 in the bulge or on A27 above the bulge; see below). The results presented here
suggest that both electrostatic and hydrogen bonding are essential for the arginine interaction with TAR and are consistent with the "arginine fork" model in which RNA recognition occurs, at least in part, through hydrogen bonds to two adjacent phosphates (10). The finding that methylation of a single amino group of arginine prevents binding is consistent with the proposal (10) that arginine methylation, commonly seen in RNA-binding proteins, may provide a mechanism to regulate RNA binding.

It is interesting that L-arginine can also bind specifically to a guanosine binding site in the Tetrahymena intron (16-18). Competition experiments with arginine analogs (16, 17) have suggested that the arginine-intron interaction, with a $K_d \approx 3$ mM, occurs through a combination of specific hydrogen bonds to a guanine base and hydrophobic interactions between the aliphatic part of the arginine side chain and surrounding groups in the RNA structure. The results from interference, mutagenesis, and competition experiments suggest that the arginine-binding site in TAR is different, probably involving only an array of hydrogen bond acceptors positioned specifically to accept hydrogen bonds from the guanidinium group (see below).

A recent model suggests that the Tat/TAR interaction may be mediated through specific interactions in the RNA major groove, facilitated by a widening of the major groove adjacent to the bulge (11). This is based on the observations that the N7 position of A27 (the second base pair above the bulge) is accessible to DEPC modification and that the A27-U38 base pair is essential for binding. Our results indicate that the same A/U base pair is important in the interaction with arginine. The most direct model to explain the strict requirement for U23 in the bulge (11, 19), the strong interference at the two phosphates, and the strong interference at the N7 of A27 would be a model in which the phosphates, U23, and A27 each present specific hydrogen bond acceptors within a single guanidinium-binding site. However, interference effects are seen at other positions (6, 11), and it is important to recognize that chemical modification and base substitution experiments cannot distinguish between direct disruption of specific contacts and indirect effects resulting from changes in RNA structure or steric hindrance. In fact, the interference observed at A22 probably reflects an indirect effect, since the identity of this base pair is not critical for Tat binding (11). Understanding the detailed set of specific interactions between arginine and TAR must await NMR and crystallographic analyses, but it is clear that the precise three-dimensional structure of TAR must be a major determinant of the specificity.

Other proteins are likely to use arginine side chains to recognize particular structural features of RNA (10). For example, several ribosomal proteins contain arginine-rich motifs similar to the RNA-binding region of Tat (20), suggesting that ribosomal RNAs will contain arginine-binding sites. The interference experiments presented here demonstrate methods for identifying such sites in RNAs. An RNA fragment of interest can be chemically modified (using any desired modification reagent), and positions that interfere with arginine binding can be identified by elution from an L-arginine column. Identification of these sites should facilitate mapping of protein–RNA interactions.

We thank S. Jaffrey, B. Calnan, and other members of the laboratory for helpful suggestions and P. Sharp, J. Williamson, and J. Puglisi for comments on the manuscript. This work was supported by a generous grant from the W. M. Keck Foundation, by the Lucille P. Markey Charitable Trust, and by National Institutes of Health Grant AI29135.