
27. Hydrophobic interaction rises above $k_B T$ such that the strength of the attractive force increases with distance.


30. Steric stabilization of hydrophobic colloids in aqueous solution occurs when a hydrophobic polymer is adsorbed at the colloid-water interface. This polymer prevents flocculation of two colloid particles in two ways. First, approach of the particles to a distance such that the strength of the attractive hydrophobic interaction rises above $k_B T$ (where $k_B$ is the Boltzmann constant and $T$ is temperature) is inhibited by polymer desorption. Second, for colloids, loss of conformational entropy due to the approach of two particles to one another creates a repulsive force that helps to oppose the attractive hydrophobic interaction. We refer to the second effect as "entropic repulsion," consistent with D. H. Everett (Basic Principles of Colloid Science (Royal Society of Chemistry, London, 1988), pp. 45–50).


34. Supported by the National Science Foundation (NSF) under the Engineering Research Center Initiative to the Massachusetts Institute of Technology Biotechnology Process Engineering Center (NSF cooperative agreement CDR-88-03041), Office of Naval Research, and the Defense Advanced Research Projects Agency (DARPA). The x-ray photoelectron spectrometer was provided by DARPA through the University Research Initiative and is housed in the Harvard University Materials Research Laboratory, an NSF-funded facility. Initial research in this project was carried out by C. Pale-Grudemange. We thank H. Biebuyck, J. Folkers, A. Jain, P. Lalbinius, and T. R. Lee for critical readings of the manuscript. K.L.P. was an NSF postdoctoral fellow, 1986 to 1989.

Arginine-Mediated RNA Recognition: The Arginine Fork

BARBARA J. CALNAN, BRUCE TIDOR, SARA BIANCALANA, DEREK HUDSON, ALAN D. FRANKEL*

Short peptides that contain the basic region of the HIV-1 Tat protein bind specifically to a bulged region in TAR RNA. A peptide that contained nine arginines ($R_9$) also bound specifically to TAR, and a mutant Tat protein that contained $R_0$ was fully active for transactivation. In contrast, a peptide that contained nine lysines ($K_9$) bound TAR poorly and the corresponding protein gave only marginal activity. By starting with the $K_0$ mutant and replacing lysine residues with arginines, a single arginine was identified that is required for specific binding and transactivation. Ethylation interference experiments suggest that this arginine contacts two adjacent phosphates at the RNA bulge. Model building suggests that the arginine-$\eta$ nitrogens and the $\epsilon$ nitrogen can form specific networks of hydrogen bonds with adjacent pairs of phosphates and that these arrangements are likely to occur near RNA loops and bulges and not within double-stranded A-form RNA. Thus, arginine side chains may be commonly used to recognize specific RNA structures.

RNA-protein interactions are important for many regulatory processes, but little is known about the details of sequence-specific recognition. From what is known, it appears that both RNA structure and nucleotide sequence function in recognition. The crystal structure of the glutaminyl tRNA synthetase-tRNA complex (1) has shown that specific contacts are made between amino acid side chains and bases in non-base paired regions of the RNA, while studies of the R17 coat protein (2) have suggested that the overall three-dimensional RNA conformation contributes substantially to recognition. Recently, an arginine-rich RNA-binding motif has been identified in the TAR RNA-binding proteins (3), including the human immunodeficiency virus (HIV) Tat protein. Peptides that contain this region of Tat bind specifically to an RNA stem-loop structure named TAR (4, 5), which is located in the HIV long terminal repeat, and RNA binding is essential for Tat-dependent transcriptional activation (5). The overall charge density of the Tat peptides is important for binding, however, the amino acid sequence requirements are flexible; the sequence can be scrambled and still bind specifically to TAR (5).

The basic RNA-binding region of Tat, RKKRRQRRR (residues 49 to 57), is nine amino acids long and contains a glutamine at position 54 that is not essential for binding or activity (5). Because it is known that a high positive charge density is important for RNA binding, we synthesized (6) two peptides, $R_9$, which contains a stretch of nine adjacent arginines (with a tyrosine at the NH$_2$-terminus and an alanine at the COOH-terminus), and $K_9$, which contains a stretch of nine lysines (and a surrounding tyrosine and alanine), and measured their binding to TAR RNA (7). The $R_9$ peptide bound to TAR RNA with the same affinity as the wild-type Tat peptide and with ten-fold higher affinity than $K_9$ (Fig. 1). The specificity of $R_9$ binding to TAR was identical to the wild-type peptide, whereas $K_9$ binding was nonspecific (7). Because RNA binding of Tat peptides correlates with Tat's function as a transcriptional activator (5), we asked whether $R_9$ or $K_9$ could function in the context of the intact protein. The nine- amino acid basic region of Tat was replaced by $R_9$ or $K_9$ in a Tat expression vector, and activation of HIV-1 transcription by the chimERIC Tat proteins was tested in transient transfection assays (8). The $R_9$-containing protein gave wild-type transactivation activity and was 100-fold more active.

*To whom correspondence should be addressed.

17 December 1990; accepted 13 March 1991

B. J. Calnan, B. Tidor, A. D. Frankel, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142; S. Biancalana and D. Hudson, MillGen/Biosearch, Division of Millipore, 81 Digital Drive, Novato, CA 94949.
within the basic region, we changed each position individually (residues 49 to 57) from lysine to arginine and measured transactivation. The optimal location of the arginine was at position 52 or 53, with activity decreasing as the arginine was moved toward the ends of the basic region (Fig. 4). Thus, a single arginine surrounded by three to four basic amino acids on each side is sufficient for specific recognition of TAR RNA.

These and other results (4, 5, 11) clearly suggested that RNA structure is important in Tat-TAR recognition. It seemed plausible that the RNA backbone might be adopting a highly defined conformation and that a specific configuration of phosphates was being recognized by arginine. To identify phosphates involved in recognition, we performed ethylation interference experiments (12) with the R52, R53, K5, and wild-type peptides. Modification of two or more phosphates at the 5' end of the TAR RNA (between A22 and U23, and U23 and C24) interfered with specific binding of R52 (Fig. 5). An identical pattern was seen with the wild-type peptide (Fig. 5) and with R53 (10), while no interference was seen with K5 (Fig. 5). At higher R52 concentrations, where nonspecific binding occurred, no interference was observed (10). These results suggest that a single specific arginine simultaneously contacts two adjacent phosphates (even in the wild-type peptide, which contains multiple arginines), although we cannot rule out the possibility that ethylation alters the RNA structure and indirectly interferes with binding. Ethylation of some phosphates, particularly the phosphate between G26 and A27, seems to enhance binding (Fig. 5), possibly by stabilizing the RNA structure. The minimal interference observed with lysines or with nonspecific arginines suggests that these residues may make weaker contacts with the RNA or that alternative phosphates may be contacted when one is modified.

How does a single arginine recognize TAR RNA? Arginine contains two terminal amino (NH2) groups at the ε position and a secondary amine (NH) at the e position, each of which can donate hydrogen bonds to appropriate acceptor groups. The positions at which acceptor atoms would be located in order to form hydrogen bonds with ideal distances and geometries are shown in Fig. 6A. Clearly, an arginine side chain can form many possible hydrogen bonds with appropriately positioned acceptor groups on the RNA. These acceptors can include phosphate oxygens, the ribose 2' OH, and groups on the bases (for example, O-6 and N-7 on guanine or O-4 on uridine in the major groove, or N-3 on guanine or O-2 on uridine in the minor groove). In contrast, lysine, which contains a single terminal amino group, cannot form such an extensive network of hydrogen bonds; this amino group also has tetrahedral geometry rather than the planar geometry of the arginine amino groups. Our ethylation interference data indicate that a single arginine in Tat contacts two adjacent phosphates at the TAR bulge, suggesting that the phosphate backbone adopts a defined conformation that can be bridged by arginine in a fork-like arrangement. To determine a plausible con-
formation for this interaction, molecular modeling (13) was used to locate the most favorable positions of two phosphates with hydrogen (H)-bonds to arginine. The best arrangement (Fig. 6B) has a pair of H-bonds between a phosphate and two N$_\epsilon$'s, and another pair of H-bonds between the second phosphate and N$_\epsilon$ and N$_\eta$. Each phosphate is shared by a pair of nitrogens, with a distance between phosphates of 7.1 Å (center to center distance between phosphorus atoms). We define the arginine fork as an interaction between a single arginine and a pair of adjacent phosphates, which mediates specific recognition of RNA structure. Other arginine-phosphate arrangements are possible (for example, see legend to Fig. 6B), and arginine forks with additional H-bonds are possible (for example, with a specific base or a 2' OH).

To determine whether such phosphate arrangements are found in RNA structures, the modeled phosphate coordinates from Fig. 6B were superimposed on all phosphate pairs in tRNA crystal structures (14). The results indicate that double-stranded A-form RNA cannot readily accommodate this arrangement; the P-P distance in the model (7.1 Å) is longer than the P-P distance in A-form RNA (5.6 Å), and the phosphate oxygens in A-form RNA are not properly oriented to form H-bonds between a single arginine and a pair of adjacent phosphates. Reasonable H-bonding arrangements are much more likely to be found at discontinuous regions of RNA, for example, at junctions between double-stranded A-form RNA and a bulge or loop. The two critical phosphates in TAR are located precisely at the junction of the double-stranded stem and the 3-nucleotide bulge.

The occluded region of the rRNA synthetase–tRNA shows a similar interaction of arginine with the acceptor strand of tRNA (1). Arg$^{133}$ forms H-bonds with two adjacent phosphates and an additional H-bond with a ribose 2' OH. It is plausible that the arginine in TAR also interacts with a 2' OH, thus discriminating between RNA and DNA.

The recognition of TAR by Tat highlights fundamental differences between RNA recognition and DNA recognition. It
is clear from the structures of protein-DNA complexes that sequence-specific discrimination derives primarily from direct base-specific contacts, most commonly made in the DNA major groove (15). In most cases, DNA tertiary structure does not seem to be of major importance in recognition. That RNA recognition often seems to rely on RNA tertiary structure is emphasized by the finding that the Tat-TAR interaction uses only a single arginine side chain in the midst of major importance in recognition. That DNA tertiary structure does not seem to be RNA tertiary structure is emphasized by the of an apparently unstructured segment of the recognition site on both DNA and RNA. DNA recognition appears to be primarily backbone RNA-binding proteins, including heterogeneous nuclear RNA-binding proteins and nuclear proteins, contain clusters of methylated arginines, most commonly N<sup>2</sup>,N<sup>6</sup>-dimethylarginine (23). Because methylation would block H-bonding but would not alter the charge of the side chain, arginine methylation could provide a mechanism to regulate RNA binding between specific and nonspecific modes. While it is clear that RNA recognition will involve more than just arginine forks, it seems reasonable to suggest that arginine-mediated recognition of RNA structure may be an important part of many RNA-protein complexes.

REFERENCES AND NOTES

6. Synthesizes were performed with the use of Fmoc chemistry on a Milligen/Biosearch model 9600 peptide-synthesizer as described (5) except that a mixture of vinylglycol-polystyrene graft copolymer (PEG-PS) was used for the reactions (G. R. F. W. Hudson, unpublished data). All peptides were synthesized as their C-terminal amides with acetylated NH termini. Amino acid composition was synthesized as their COOH-terminal amides with acetylated NH termini. Amino acid composition was determined by hydrolysis at 110°C in HCl (6 M) that contained phenol (0.5%) and analysis on an LKB 4151 Alpha Plus analyzer. Peptides were pu-rified on a C<sub>4</sub> reversed-phased high-performance liquid chromatography (HPLC) column (Vydac) with an acetonitrile gradient (0.25% per min in 0.1% trifluoroacetic acid). Peptide absorption spectra were recorded and concentrations were determined by tyrosine absorbance at 278 nm [ε = 12400 M<sup>-1</sup> cm<sup>-1</sup>]. T. E. Creighton, Protein Structures and Molecular Principles (Freeman, New York, 1984), p. 17. Peptide masses were confirmed by fast atom bombardment mass spectrometry (University of California, Berkeley), and purity and concentrations were confirmed by native polyacrylamide (20%) gel electrophoresis in sodium acetate, pH 4.5 (30 M).
7. TAR RNAs were synthesized by T7 RNA polymerase in vitro with synthetic oligonucleotides (5) and subjected to electrophoresis on 10% polyacrylamide/4 M urea sequencing gels.
8. Local energy minima for the interaction of a single arginine side chain with two phosphate groups were determined with the use of the molecular modeling program CHARMM [B. R. Brooks et al., J. Comput. Chem. 4, 187 (1983)] with standard parameters (9) for the calculation of the potential energy function for the interaction of two arginines with two phosphates. The arginine side chain was fixed in the center of an 11 Å sphere and surrounded by 70 pairs of randomly placed dimethylimphosphate molecules. The potential energy function was modified so that each dimethylphosphoRUPEP was energy minimized with the protein, and the potential energy function was minimized with the protein in the presence of 11 Å dimethylphosphoRUPEP.
9. Local energy minima for the interaction of a single arginine side chain with two phosphate groups were determined with the use of the molecular modeling program CHARMM.
The signal recognition particle (SRP) directs signal sequence specific targeting of ribosomes to the rough endoplasmic reticulum. Displacement of the SRP from the signal sequence of a nascent polypeptide is a guanosine triphosphate (GTP)-dependent reaction mediated by the membrane-bound SRP receptor. A nonhydrolyzable GTP analog can replace GTP in the signal sequence displacement reaction, but the SRP then fails to dissociate from the membrane. Complexes of the SRP with its receptor containing the nonhydrolyzable analog are incompetent for subsequent rounds of protein translocation. Thus, vectorial targeting of ribosomes to the endoplasmic reticulum is controlled by a GTP hydrolysis cycle that regulates the affinity between the SRP, signal sequences, and the SRP receptor.

Requirement of GTP Hydrolysis for Dissociation of the Signal Recognition Particle from Its Receptor

TIMOTHY CONNOLLY,* PETER J. RAPEJKO, REID GILMORE†

The signal recognition particle (SRP) directs signal sequence specific targeting of ribosomes to the rough endoplasmic reticulum (RER)--specific signal sequences are cotranslationally recognized by SRPs and then delivered to the RER membrane via interaction between the SRP and the SRP receptor or docking protein (1–4). The SRP receptor–mediated displacement of the SRP from the signal sequence of the nascent polypeptide is a GTP-dependent reaction (5–7). One protein subunit from both the SRP receptor (SRAs) (7) and the SRP (SRP54) (8, 9) contains protein sequence motifs that are similar to those in GTP binding proteins (10). We examined the role of GTP hydrolysis in SRP receptor function by replacing GTP with the nonhydrolyzable analog β,γ-imidodiphospho-5′-triphosphate [Gpp(NH)p] during the targeting and insertion steps of a protein translocation reaction.

A truncated mRNA encoding the NH2-terminal 90 residues of the G protein of vesicular stomatitis virus was translated in vitro in the presence of 125I-labeled SRP to prepare complexes containing SRP, ribosomes, and a nascent polypeptide. After translation, ribonucleotides were removed by gel filtration chromatography, and the SRP-ribosome complexes were incubated in the absence or presence of Gpp(NH)p and microsomal membranes that were depleted of SRP (K-RM) (Fig. 1, A and B). The SRP-ribosome complexes were then separated from free SRPs by sedimentation on sucrose density gradients that were underlayered with a 2 M sucrose cushion. Under these conditions, membrane vesicles sediment at the interface between the sucrose layers. Addition of K-RM and GTP to the complexes increased the amount of unbound SRP recovered after centrifugation while the amount of SRP bound to ribosomes decreased (Fig. 1, A and C), indicating that the SRP enters a soluble pool. In closely, a broader range of phosphate orientations was accepted to accommodate other possible conformations indicated by the modeling (see legend to Fig. 6B) or conformational changes that might occur upon arginine binding. Three types of phosphate pairs were found: phosphates adjacent in the sequence (i, i+1), those one away from each other (i, i+2), and those distant in primary sequence but near each other in the tertiary structure. In no case did phosphate pairs in double-stranded RNA match template. The (i, i+1) pattern frequently appeared surrounding the first or last unpaired base in a bulge or loop, and the (i, i+2) pattern frequently appeared in a bulge or loop that bound a hydrated magnesium ion. In fact, water molecules coordinated to Mg2+ produce an array of hydrogen-bond donors similar to that of an arginine side chain (Fig. 6A) and bind to a phosphate pair in a manner similar to Fig. 6B.

1. Stolz, T., Quart, Rev. Biophys. 23, 205 (1990).

We thank D. Rio, P. Sharp, A. Sachs, J. Williamson, P. Kim, and A. Miranker for helpful discussions, and C. Palb for comments on the manuscript. Supported by the Lucile P. Markey Charitable Trust, by NIH grant AI29135 (A.D.F.), and by a grant from the Pittsburgh Supercomputing Center (B.T.).

7 March 1991; accepted 4 April 1991

Fig. 1. Recycling of SRP after GTP hydrolysis. A truncated mRNA transcript (7, 16) was incubated for 20 min in a wheat germ system containing 6.5 nM SRP (including 125I-labeled SRP) (3, 7, 19). SRP-ribosome complexes were separated from ribonucleotides (5) and incubated in 50 mM triethanolamine-acetate, pH 7.5, 150 mM potassium acetate, 2.5 mM magnesium acetate, and 1 mM diethiothreitol for 5 min at 25°C as follows. (A) No additions, (B) K-RM (5 equivalents), as defined (3), (C) K-RM (5 equivalents) and 100 μM GTP, and (D) K-RM (5 equivalents) and 100 μM Gpp(NH)p. The abbreviation K-RM refers to rough microsomal membranes depleted of SRPs by extraction with 0.5 M potassium acetate (3). Samples were applied to sucrose density gradients (10 to 30%) underlain with 0.5 ml of 2 M sucrose. The gradients contained 50 mM triethanolamine-acetate, pH 7.5, 150 mM potassium acetate, 5 mM magnesium acetate, and 1 mM diethiothreitol. Centrifugation, fractionation, and quantitation of gradients were as described (3, 7). The top and bottom of the gradient were in fractions 1 and 50, respectively. The interface between the sucrose layers was in fraction 45. The sedimentation position of 80S ribosomes (fractions 14 to 20) was determined from the ultraviolet-absorbance profile as recorded with a continuous flow cell. Free SRPs sedimented in fractions 1 to 5 in gradients lacking ribosomes. Similar results were obtained in three separate experiments.