Molecular Dynamics and Binding Specificity Analysis of the Bovine Immunodeficiency Virus BIV Tat-TAR Complex

Carolina M. Reyes,* Riccardo Nifosì,† Alan D. Frankel,‡ and Peter A. Kollman*

Departments of *Pharmaceutical Chemistry and ‡Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143 USA; and †Scuola Normale Superiore and Istituto Nazionale di Fisica della Materia, Pisa 56126, Italy

ABSTRACT   We have performed molecular dynamics (MD) simulations, with particle-mesh Ewald, explicit waters, and counterions, and binding specificity analyses using combined molecular mechanics and continuum solvent (MM-PBSA) on the bovine immunodeficiency virus (BIV) Tat peptide-TAR RNA complex. The solution structure for the complex was solved independently by Patel and co-workers and Puglisi and co-workers. We investigated the differences in both structures and trajectories, particularly in the formation of the U-A-U base triple, the dynamic flexibility of the Tat peptide, and the interactions at the binding interface. We observed a decrease in RMSD in comparing the final average RNA structures and initial RNA structures of both trajectories, which suggests the convergence of the RNA structures to a MD equilibrated RNA structure. We also calculated the relative binding of different Tat peptide mutants to TAR RNA and found qualitative agreement with experimental studies.

INTRODUCTION

RNA-protein interactions are essential in regulating gene expression. Significant attention has focused on elucidating RNA-protein interactions in retroviruses, such as the human immunodeficiency virus (HIV). In retroviruses, RNA-protein interactions govern reverse transcription and the transcription and transport of the viral RNAs. One of the key elements in lentiviral transcription is the *trans*-activator of transcription (Tat) protein. Tat binds to the 5′ end of the newly transcribed *trans*-activation response element (TAR) RNA and enhances transcriptional elongation and yield (Rana and Jeang, 1999; Jones and Peterlin, 1994).

The Tat protein sequence can be roughly divided into five domains: the N-terminal, Cys-rich, core, basic, Gln-rich, and C-terminal regions. Among lentiviruses, only the basic and core regions are conserved (Rana and Jeang, 1999). The RNA-binding region has been narrowed down to a short sequence of 10–20 amino acids of the basic region (Frankel, 2000).

Structural studies of the RNA-binding region of Tat from HIV and equine infectious anemia virus (EIAV) suggest a highly flexible peptide with some tendency to form an α-helix (Mujeeb et al., 1996; Bayer et al., 1995; Aboul-ela and Varani, 1998; Sticht et al., 1993; Willbold et al., 1998; Tan and Frankel, 1995; Calnan et al., 1991). Structures of HIV TAR RNA with argininamide have also been determined (Puglisi et al., 1992b, 1993; Aboul-ela et al., 1995). Attempts to solve the structure of the HIV Tat peptide bound to TAR were hindered by the inability to assign spectra of the peptide protons (Aboul-ela and Varani, 1998;

Greenbaum, 1996; Puglisi et al., 1992a). Fortunately, the bovine immunodeficiency virus (BIV) Tat-TAR complex was more amenable to NMR studies. Puglisi and co-workers (Puglisi et al., 1995) and Patel and co-workers (Ye et al., 1995) independently solved the solution structure of the basic RNA-binding region of Tat, in complex with the TAR RNA.

Global features of both the Puglisi and Patel structures are similar, as illustrated in Fig. 1. The β hairpin peptide, represented as pink tubes in Fig. 1, is positioned in the major groove of the 28-nucleotide TAR RNA fragment. The widened major groove (9–17 Å) facilitates the tight fit of the peptide in the complex (Puglisi et al., 1995; Ye et al., 1995). For structure determination, both groups used a combination of ^{15}N- and ^{13}C-labeled samples with heteronuclear multidimensional NMR techniques. Puglisi and co-workers refined the entire complex simultaneously using 384 interproton distances (26 of which are intermolecular)(Puglisi et al., 1995). Patel and co-workers docked the peptide and RNA together after computing the monomer structures separately with a total of 1052 distance restraints (102 intermolecular) (Ye et al., 1995). Puglisi et al. deposited an averaged, energy-minimized structure, while Ye et al. reported a family of five structures in the Protein Data Bank (PDB) (Sussman et al., 1998).

One of the most notable differences between the Puglisi and Patel structures is the relative orientation of the upper and lower RNA helical stem regions. The Patel RNA structure shows a 40° bend (between C8-G27 and G11-C25 basepairs), whereas the Puglisi structure shows a relatively straight helix. However, in the absence of long-range NOE constraints, it is difficult to assign definite relative positions of the RNA helices. Therefore, the interhelical bend differences between the two structures can only be ascribed to their local conformational features because there are no specific interhelical constraints (Greenbaum, 1996).
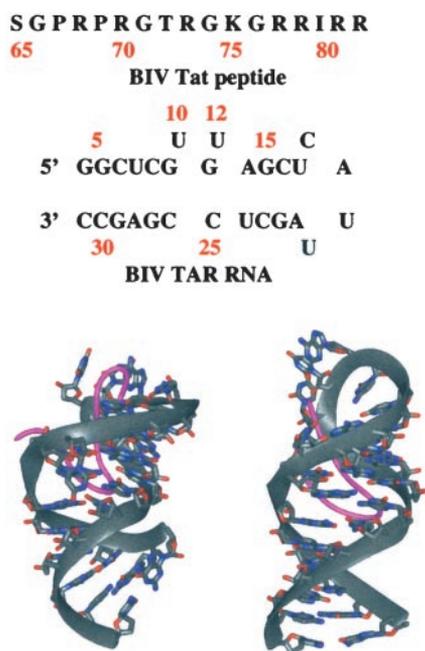
FIGURE 1 Sequence and initial structures of the Patel and Puglisi BIV Tat-TAR complexes. Peptide backbone is represented in magenta. The Puglisi Tat peptide does not contain residues 65–67.

In the RNA bulge region, U12 is looped out and solvent-exposed. This also contributes to the widening of the major groove and facilitates peptide binding (Puglisi et al., 1995; Ye et al., 1995). U12 is not well constrained by the NMR data and is looped out into the minor groove in a perpendicular orientation to the helix in the Patel structure, whereas it is pointing toward the solvent in the Puglisi structure. The U10 base of the RNA bulge is located in the major groove in both structures. Patel and co-workers established the existence of a co-planar U10-A13-U24 base triple with multiple NOE constraints, particularly between the imino proton of U10 to H8 of A13. Though their NMR data did not provide direct evidence, the Puglisi structure shows a remarkably similar base triple with H-bonds on the Hoogsteen edge of A13 and the Watson-Crick edge of U10.

There also exist some subtle differences in the RNA-protein contacts between the two structures. The Patel peptide contains three additional residues at the N-terminus that are not present in the Puglisi structure; these additional residues contact the hairpin loop of the TAR RNA, but do not contribute to RNA-binding affinity (Chen and Frankel, 1995). Moreover, the interaction of Arg-73 differs in the two complexes, as well as the position of Ile-79 near the base triple.

Our theoretical studies can complement the structural studies of BIV Tat-TAR complexes to investigate the differences in their structures, dynamics, and energetics. Recent theoretical studies of RNA-protein complexes have highlighted the important driving forces in complex forma-tion, particularly for such highly flexible molecules that undergo significant conformational changes upon binding (Hermann and Westhof, 1999; Reyes and Kollman, 1999, 2000a, b; Tang and Nilsson, 1999). We have also explored the dynamics and energetics of the HIV-1 TAR RNA and its complex with argininamide (Nifosi et al., 2000). Both the previous HIV and current BIV theoretical studies investigate and attempt to understand, in general, the underlying forces of RNA-protein recognition. This is accomplished using molecular dynamics (MD) simulations and free energy analyses.

The conformational free energy is computed as the sum of molecular mechanics energy and solvation free energy as estimated by continuum methods (MM-PBSA)(Srinivasan et al., 1998a). The solvation free energy is approximated as the sum of the electrostatics, computed by the Poisson-Boltzmann (PB) method (Sharp and Honig, 1990), and nonpolar contributions, as a solvent-accessible surface area (SA) dependent term. The MM-PBSA method has recently been applied to nucleic acids (Srinivasan et al., 1998a, b; Cheatham et al., 1998), protein-ligand (Chong et al., 1999), and RNA-protein complexes, with relative success in reproducing both absolute and relative binding free energies (Reyes and Kollman, 2000b). Here, we use the MM-PBSA method to describe the conformational free energies of the Puglisi and Patel structures and explore the relative binding affinities of different mutants of the Tat peptide.

MATERIALS AND METHODS

Molecular dynamics

MD simulations were performed with the SANDER module of AMBER5 and the Cornell et al., force field (Cornell et al., 1996). One of the five NMR structures from Ye et al. was chosen arbitrarily (PDB: 1BIV) as the starting structure; the averaged, energy-minimized structure from Puglisi et al. (PDB: 1MNB) was obtained as the initial structure (Ye et al., 1995; Puglisi et al., 1995). Eighteen neutralizing Na^+ ions were added to both structures. The Ye et al. structure was solvated in a $73 \times 58 \times 50$ Å^3 box of 5670 TIP3P (Jorgensen et al., 1983) waters, and the Puglisi et al. structure was solvated in a $67 \times 56 \times 53$ Å^3 box of 5162 TIP3P water molecules.

During the initial equilibration stage of both structures, the peptide-RNA complex was minimized for 1000 steps, followed by 40-ps MD of the water molecules at 300 K to relax the waters around the macromolecules, and then another 1000 steps of minimization for all atoms. Both systems were heated to 298 K in gradual steps using the Berendsen algorithm for separate solute-solvent temperature coupling (Berendsen et al., 1981): 10 ps from 0 to 50 K, then 20 ps from 50 to 150 K, finally 30 ps from 150 to 298 K.

MD simulations were carried out at constant pressure (1 atm) and temperature (298 K), periodic boundary conditions, and particle-mesh Ewald (Darden et al., 1993) for a more accurate treatment of electrostatics. SHAKE was applied to all hydrogen bonds and a time step of 2 fs was utilized (Ryckaert et al., 1977). Both simulations ran for 1 ns after equilibration and gradual heating. Structural analyses of the trajectories were performed using the CARNAL package from AMBER5 and visual analyses and illustrations were done using MidasPlus (Ferrin et al., 1988).

Free energy analyses

The conformational free energy was approximated as the sum of the molecular mechanical energy, E(MM), and the solvation free energy, ΔG(solv) (Smith and Honig, 1994; Srinivasan et al., 1998b):

$$\langle \Delta G \rangle = \langle E(\text{MM}) \rangle + \langle \Delta G(\text{solv}) \rangle,$$

where

$$\langle \Delta G(\text{solv}) \rangle = \langle \Delta G(\text{PB}) + \Delta G(\text{np}) \rangle.$$

To calculate the energies, 50 snapshots at 10-ps intervals from the last 500-ps trajectories were selected, excluding the Na^+ ions and water molecules. The total molecular mechanical energies, E(MM), which include bond, angle, dihedral, E(BADH), van der Waals, E(vdW), and electrostatic, E(elec) terms, of these snapshots were computed and averaged over the simulation. Consistent with the MD simulations, the Cornell et al. force field was used in the energy analysis using the ANAL module in AMBER5.

The electrostatic component of the solvation free energy, G(PB), was computed with the nonlinear PB method as implemented in the Delphi II program (Nicholls et al., 1990). This was approximated as the reaction field energy of moving a solute from a low dielectric ($\epsilon = 1$) to a high dielectric continuum solvent ($\epsilon = 80$). Atomic charges were taken from the Cornell et al. force field and vdW radii from the PARSE parameter set (Massova and Kollman, 2000; Sitkoff et al., 1994). Salt concentration of 70 mM was specified for the calculations in accord with experimental binding measurements (Chen and Frankel, 1995). We defined a 0.5 Å grid spacing with lattice dimensions extending 20% beyond the solute dimensions and required 1000 linear iterations + 1000 nonlinear iterations for energy convergence.

The nonpolar solvation term was estimated as a solvent-accessible surface area-dependent term. G(np) = γ(SASA) + β, where γ = 0.00542 kcal/Å and β = 0.92 kcal/mol (Sitkoff et al., 1994; Srinivasan et al., 1998a; Cheatham et al., 1998). The solvent-accessible surface area was determined using Sanner's MSMS software (Sanner et al., 1996). The PB term and the nonpolar solvation term together make up the solvation free energy contribution. The average solvation term, computed from trajectory snapshots, was reported.

Computational mutagenesis

We applied the computational mutagenesis method to estimate the relative binding of different Tat peptides to the RNA on the Patel structure because its peptide sequence exactly matched the Tat peptide sequence measured experimentally. Two methods were used to calculate the relative binding of different peptide mutants. The first and simpler approach is the minimization method (Reyes and Kollman, 2000b). This method takes a representative snapshot, such as the final MD structure, and calculates its free energy after a series of minimizations with a distance-dependent dielectric. The minimization protocol entails first minimizing the residue to be mutated for 1000 steps, then minimizing only all hydrogen atoms for another 1000 steps, and finally minimizing all atoms for 1000 steps. Then, the coordinates of the same final MD structure are altered with MidasPlus to model the mutant, and its free energy is determined using the same minimization protocol as was applied to the wild-type complex. The relative binding free energy is then the binding free energy difference between the minimized wild-type and the minimized mutant structures.

The second protocol for calculating the relative binding generates mutant structures from the wild-type MD trajectory by altering the coordinates of the wild-type trajectory (Massova and Kollman, 1999). This method is best applied to mutations, which involve deleting atoms, such as mutation to an alanine or glycine. We obtained snapshots for energy calculations every 10 ps over the entire trajectory. The mutated snapshots are not minimized, so it is assumed that no local rearrangements occur with the mutation. With the trajectory mutation method, the average free energy difference between the wild-type snapshots and mutated snapshots is the relative binding free energy.

RESULTS AND DISCUSSION

Molecular dynamics

Structural analyses of the trajectories of the Puglisi and Patel structures reveal both differences and similarities. The root-mean-squared distance (RMSD) over the Puglisi trajectory fitted to the starting structure is plotted in Fig. 2. The
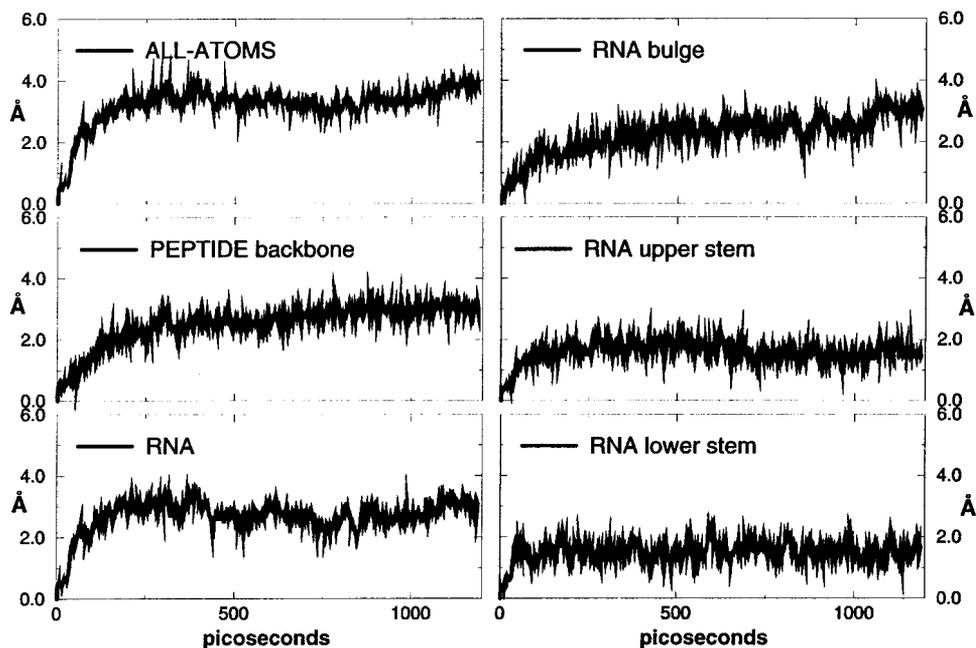

FIGURE 2 Root-mean-square deviations from the initial structure of the Puglisi trajectory fitted on all atoms, peptide backbone, RNA, RNA bulge, RNA upper stem, and RNA lower stem.

FIGURE 3 Average per residue/base fluctuations for the Patel Tat peptide (*upper left*), Puglisi Tat peptide (*lower left*), Patel TAR RNA (*upper right*), and Puglisi TAR RNA (*lower right*) computed over the entire trajectory.
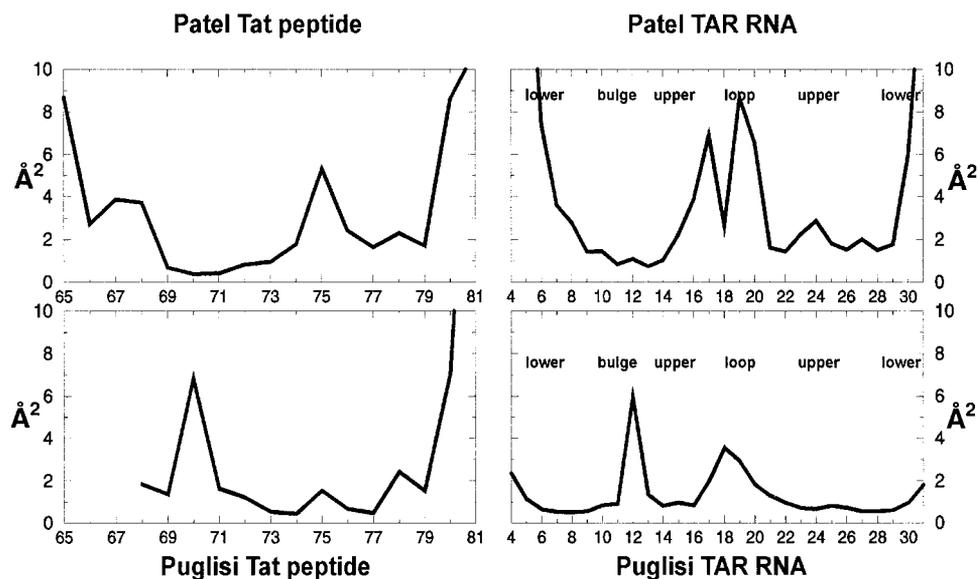
average RMSD of all atoms is 3.2 Å. The peptide backbone average RMSD is 2.5 Å, while the average RNA RMSD is 2.7 Å. Fitted to the RNA subregions, the average RMSDs are even lower, 2.3 Å for the bulge region, 1.6 Å for the upper stem, and 1.5 Å for the lower stem. Overall, the RMSD profile of the Puglisi trajectory shows a fairly stable simulation.

However, local structural changes did occur during the simulation. Fig. 3 shows the average fluctuations of the trajectory and highlights the flexible residues. Among the peptide residues, Arg-70, Arg-80, and Arg-81 become more extended and solvent-exposed during the simulation of the Puglisi structure. In the RNA, the single-stranded loop bases in the upper stem exhibit high fluctuations. The most flexible base is the solvent-exposed U12 in the bulge, which is freely rotating, as indicated by changes in its χ angle over time. Also, changes in the torsion angle of the phosphate backbone between U12 and A13 (A13's α and β) shift the orientation of U12 from perpendicular to parallel to the minor groove of the RNA, though it remains solvent-exposed. Puglisi et al. (1995) found U12 to be disordered in the NMR structure, as we observed in our MD simulation.

Another notable change in the bulge region is the formation of a definitive co-planar U · AU base triple, which was not directly specified by the NMR data of Puglisi et al. (1995), but was observed in the Patel structure (Ye et al., 1995). Fig. 4 highlights the difference between the initial and MD average Puglisi base triple. Initially, the U10, A13, and U24 bases do not lie on the same plane, but are within reasonable Watson-Crick and Hoogsteen hydrogen-bonding distances. During the simulation, changes in the phosphate-sugar backbone (α, β, γ) and χ torsions of U10, A13, and U24 contribute to the development of a more co-planar base triple.

The MD simulation of the Patel structure has a different RMSD profile, as shown in Fig. 5. The average all-atom RMSD is 3.0 Å, the average RMSD for the peptide backbone is 1.7 Å, and 3.1 Å for the RNA. Fitting to the RNA subregions, the average RMSDs for bulge, upper stem, and lower stem are 1.5, 1.9, and 2.0 Å, respectively. At ~450 ps there is a dramatic increase in RNA RMSD, dominated by changes in the RMSD of the lower stem region.

Fluctuation analyses of the RNA bases from the MD trajectory of the Patel structure in Fig. 3 show that the flexible bases are located in the upper-stem loop and the terminal bases of the lower stem. Further analyses of the RNA torsion angles over time (data not shown) reveal local structural changes in the lower stem and in the bulge occurring at about the same time as the shift in RNA RMSD.
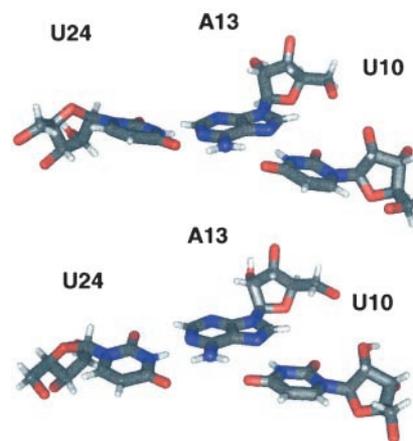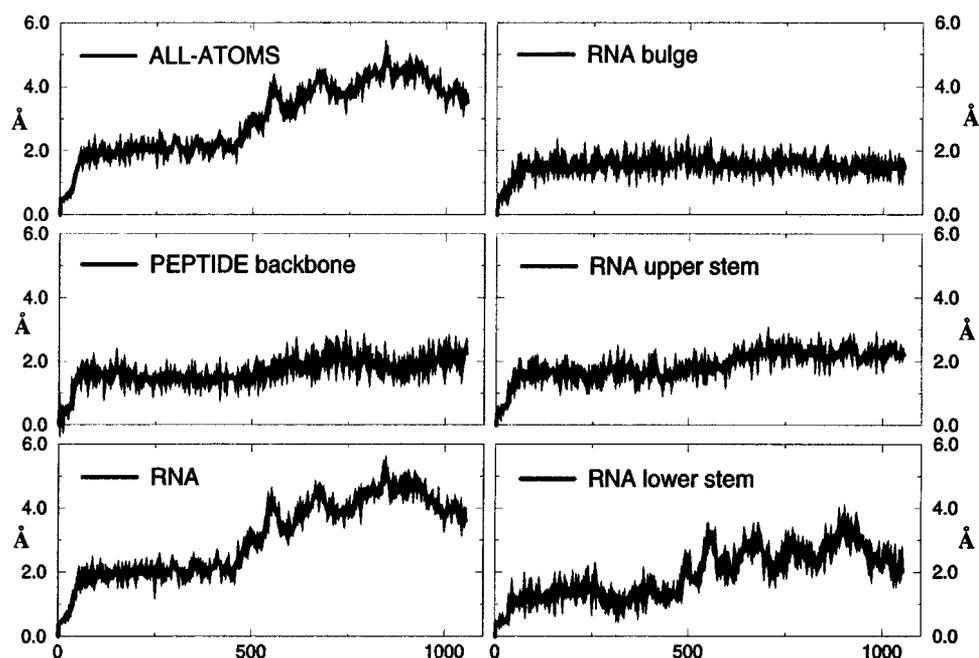

FIGURE 4 Initial U10-A13-U24 base triple in the Puglisi structure (*top*) and the more co-planar final 300-ps MD average base triple (*bottom*). Hydrogen bonds are drawn as dashed lines.

FIGURE 5 Root-mean-square deviations from the initial structure of the Patel trajectory fitted on all atoms, peptide backbone, RNA, RNA bulge, RNA upper stem, and RNA lower stem.

Changes in the backbone torsions (α, γ) of U10 and G27 around 400 ps produced a hinge-like motion, which straightened the interhelical angle bend in the RNA, and thus resulting in increased RMSDs in the lower stem and the whole RNA compared to the initial structure.

A more stringent test of how well the MD simulations represent the experimental structures was also performed by tracking the RNA-protein intermolecular NOE distances over the entire simulation. We compared our trajectory averaged intermolecular distances to the experimental NOE data found in Puglisi et al. (1995) and Ye et al. (1995). Of the 102 intermolecular NOEs from the Patel structure, seven NOEs, involving five amino acids listed in Table 1, have average distances >7 Å. The average distance of these seven NOEs is 8.4 Å. Assuming 7 Å as an upper distance constraint for weak NOEs or spin diffusion, the low number of NOE violations shows good agreement between the MD trajectory and the NMR data. Of the 26 total protein-RNA NOEs from the Puglisi structure, seven have average distances >7 Å (averaging 9.9 Å). Most of the violations result from the movement of Arg-70 away from the RNA during the simulation. Arg-70 in the Puglisi structure is fairly solvent-exposed in the major groove of the upper stem and swings gradually toward the solvent over the first 300 ps, remaining there for the rest of the simulation. In the Patel trajectory, however, Arg-70 stays in the major groove.

Convergence of the global RNA structures

An interesting observation emerges from the comparison of the initial and the average MD structures from the last 300 ps of both simulations. Both Patel and Puglisi RNA structures appear to converge to an MD-equilibrated structure. Fig. 6 illustrates the converging similarity of both RNA structures, which are initially 5.0 Å apart in RNA RMSD and become 3.3 Å apart in their final 300 ps average MD structures. In the final average structures, the most apparent change is the more similar orientation of the two helical stems as the Patel RNA helix becomes less kinked. Convergence of the two structures is reasonable given that the same MD protocol and force field were applied to both structures. Previous work by Cheatham and Kollman on DNA showed convergence of A-DNA and B-DNA to an MD-equilibrated B-DNA-like structure (Cheatham and Kollman, 1996). However, a more rigorous test of convergence is to compare the RNA RMSD of Puglisi trajectory to the final Patel structure, and vice versa, as shown in Fig. 7. If the trajectories were converging, one would expect a decreasing trend in RMSDs as snapshots from one trajectory are compared to the final structure of the other. Linear regression analyses of both RNA RMSD plots show a slightly decreasing trend. By averaging the final 300 ps of both trajectories and comparing them, we can reduce the

TABLE 1 Average NOE distances >7 Å. Distances were averaged over the entire trajectories

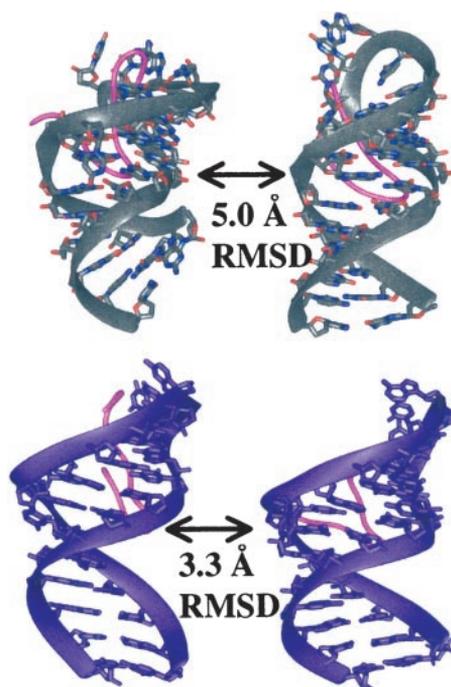| Patel Trajectory		Puglisi Trajectory	
Gly-66 αCH	C17 H5	Arg-70 δCH	G14 H1
Pro-67 αCH	A18 H1'	Arg-70 δCH	C23 NH
Pro-67 βCH	A18 H1'	Arg-70 αCH	C23 NH
Arg-68 δCH	U19 H1'	Arg-70 γCH	C23 NH
Arg-70 βCH	G14 H1	Arg-77 αCH	C26 NH
Ile-79 γCH	U24 NH3	Arg-77 γCH	C26 NH
Ile-79 δCH	U24 NH3	Arg-77 δCH	G9 H1

FIGURE 6 RNA RMSD comparison of the initial Patel (*left*) and Puglisi (*right*) structures (*top*) and the final 300-ps MD average structures (*bottom*). The PDB files of final 300-ps MD average structures are available in the supplementary materials.

noise among the snapshots and better observe that the global RNA structures of the Patel and Puglisi trajectories become more similar.

Differences in Tat peptide structures and Tat-TAR binding interface

Although the last 300-ps average RNA structures are more similar compared to their initial structures, the peptides and peptide-RNA contacts remain structurally different. A more detailed analysis in Fig. 7 shows increasing peptide RMSDs of the Puglisi trajectory to the final Patel peptide, and vice versa. The linear regression analysis shows an increasing trend, which implies that the peptide structures in both trajectories are not converging. The lack of convergence between the peptide structures is not surprising, given the highly flexible nature of the peptide in the Tat-TAR complex. This flexibility is indicated both by the broad envelope of NMR conformations in the reported structures and the high positional fluctuations of this region seen in our simulations (Fig. 3).

Moreover, the differences in the peptide-RNA contacts of the initial experimental structures are maintained in the MD. Analyses of intermolecular Tat-TAR H-bonding over both trajectories, listed in Table 2, reveal differences in high-occupancy ($> 50\%$) H-bonding. The guanidinium group of Arg-68 interacts mainly with the phosphate backbone of

U20 in the hairpin loop in the Puglisi trajectory, whereas it makes hydrogen bonds with the base of U20 in the Patel trajectory. Arg-70 makes consistent H-bonding with G14 (O6 and N7) in the Patel trajectory, in contrast to the lack of any Arg-70 interactions in the Puglisi trajectory, because it moved away from the RNA during the simulation. The backbone N of Thr-72 makes H-bonds with G22 N7 in the Patel trajectory, while the hydroxyl group of Thr-72 in the Puglisi trajectory interacts with G22 N7. The hydroxyl group of Thr-72 in the Patel trajectory interacts with the phosphate oxygen of C23. Arg-73, in both trajectories, makes contacts with N7 and O6 of G11, consistent with NMR data (Puglisi et al., 1995). As peptide contacts with the phosphate backbone are poorly resolved in the NMR structures, Arg-73 makes different high-occupancy phosphate contacts in the two trajectories: with phosphate of U10 in the Patel trajectory and with phosphate of G9 in the Puglisi trajectory. Arg-77 also differs in high-occupancy hydrogen bonding. It interacts with G9 in the Patel trajectory, but with the phosphate oxygens of C6 and U7 in the Puglisi trajectory. The flexible C-terminal arginines Arg-80 and Arg-81 contact the phosphates of G9, although there are no intermolecular NOEs to Arg-80 and Arg-81 determined by either Puglisi or Patel to support this. Therefore, although the final RNA average structures become more similar over the course of the simulations, the peptide structures and the peptide-RNA hydrogen-bonding patterns do not converge.

Relative binding free energies of peptide mutants

We also investigated the specificity determinants of Tat-TAR binding by comparing computational free energy analyses to experimental mutagenesis of the Tat peptide (Chen and Frankel, 1995). Table 3 summarizes our calculations for relative binding of various mutants applied to the final MD snapshot of the Patel structure because the Patel structure contains the exact same peptide sequence studied experimentally (Chen and Frankel, 1995). Fig. 8 illustrates the mutated residues in magenta. Overall, we obtained qualitative agreement with the experimental binding measurements.

Mutation of Arg-70 or Arg-73 to lysine was found to dramatically reduce binding to $<3\%$ of wild-type by in vitro binding and in vivo assays (Chen and Frankel, 1995). In our calculation, Arg-70Lys was unfavorable by 5.7 kcal/mol, while Arg-73Lys completely abolished binding ($\Delta\Delta G$ ~14.6 kcal/mol). Arg-70 in the wild-type structure interacts with O6 and N7 of guanine 14, which is disrupted in the mutant. The Arg-73 wild-type makes hydrogen bonds with both N7 and O6 of G11, and also forms charge interactions with the phosphates of U10; in the lysine mutant, the hydrogen bonds are eliminated and the distance to the phosphate oxygen is increased from 2.8 to 5.2 Å. It is interesting that Arg-73 in the BIV Tat-TAR structure has similar interactions to the single arginine residue bound to
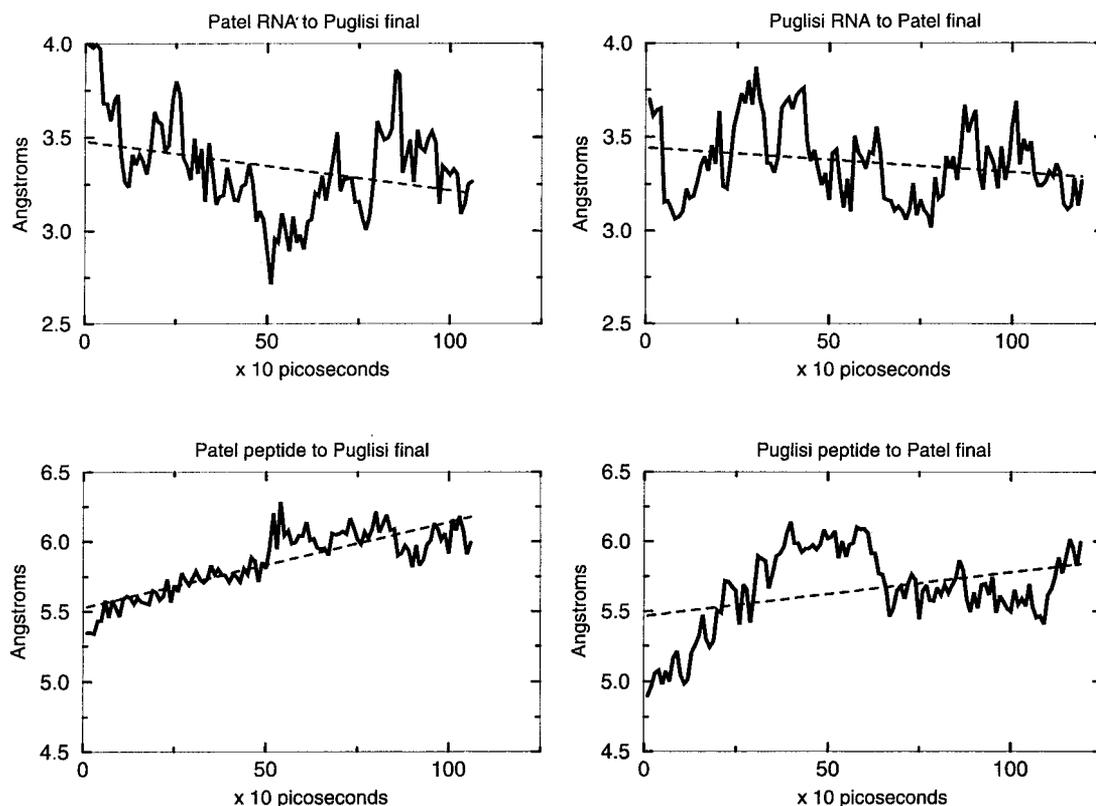
FIGURE 7 Plotting the peptide and RNA RMSDs of the Patel trajectory to the final snapshot of the Puglisi trajectory, and vice versa. The dotted lines represent the best-fit linear regression lines.

HIV-1 TAR RNA (Smith et al., 1998). Arg-73 interacts with G11, located below the base triple in the BIV Tat-TAR structure. Analogously, the single arginine residue interacts with G26 below the U-A-U base triple in the HIV-1 TAR-argininamide complex (Puglisi et al., 1992a). This single arginine of HIV Tat is largely responsible for recognition of HIV TAR (Tao and Frankel, 1992).

The Thr-72Ala mutation was computed both using minimization of the final snapshot and computing the mutation over the trajectory. The trajectory mutation method is only applicable to conservative changes because it operates by deleting atoms from the snapshots. In agreement with experiment, both methods calculate this mutation to be highly deleterious ($\Delta\Delta G \sim$ 7–8 kcal/mol). The backbone of Thr-72 forms hydrogen bonds with G22 and C23, and the hydroxyl OG1 with the phosphate oxygen of C23. In the mutation, only the OG1 interaction is disrupted. We also computed a binding free energy change for the Gly-74Ala mutant, which does not interact directly with the RNA, but facilitates the β-turn of the peptide. This mutation was also calculated to be highly unfavorable ($\Delta\Delta G \sim$ 8 kcal/mol) as found experimentally, presumably because it affects the stability of the β-turn.

A series of substitutions of Ile-79 have been examined experimentally (Chen and Frankel, 1995). In the Patel struc-ture, Ile-79 interacts with the hydrophobic edge of U10 and is located in the same plane as the U10-A13-U24 base triple. Over the course of the MD simulation, Ile-79 moves slightly below the plane of the base triple. This may explain why substitution with a larger tyrosine is tolerated at this position and costs only 0.6 kcal/mol experimentally (Chen and Frankel, 1995), although we calculate a large $\Delta\Delta G$ (4.7 kcal/mol). Tyrosine can potentially stack below U10 or perpendicular to it, and not necessarily on the same plane as the base triple. We observed that tyrosine formed a perpendicular stacking orientation relative to U10 in our calcula-tion of the Ile-79Tyr mutant. Although perpendicular aro-matic stacking interactions have been observed in protein structures (Burley and Petsko, 1985), they have not yet been observed in RNA complex structures. We also calculated the free energy changes for the mutations of Ile-79 to Ala and Val by minimization of the final MD snapshot and by mutating over the trajectory. The minimization method computed relatively larger unfavorable binding energies than observed experimentally (1.9 and 4.7 kcal/mol for the alanine and valine substitutions), but the trajectory mutation protocol yielded slightly more favorable binding energies, although with large standard deviations.

Overall, we obtained qualitative agreement with experi-mental binding of different Tat peptides. In our calculations,

TABLE 2   Hydrogen bonding pairs that are >50% occupied computed in the entire Patel and Puglisi trajectories

Peptide Residue	Patel Trajectory	% Occupied	Puglisi Trajectory	% Occupied
Arg-68	Arg-68 NH1-U19 O2	52.2	Arg-68 NE-U19 O3′	89.0
	Arg-68 NH1-U20 O2	52.6	Arg-68 NE-U20 O2P	99.9
			Arg-68 NH2-U19 O3′	91.4
			Arg-68 NH2-U20 O2P	88.2
Pro-69	Pro-69 O-A21 O6	99.7		
Arg-70	Arg-70 NE-G14 N7	93.7		
	Arg-70 NE-G14 O6	100.0		
	Arg-70 NH2-G14 N7	100.0		
Gly-71	Gly-71 N-G22 N7	99.9		
Thr-72	Thr-72 N-G22 N7	99.7	Thr-72 OG1-G22 N7	68.4
	Thr-72 OG1-C23 O2P	99.9	Thr-72 OG1-C23 N4	69.6
	Thr-72 O-C23 N4	100.0		
Arg-73	Arg-73 NE-G11 N7	59.0	Arg-73 NE-G9 O1P	92.7
	Arg-73 NE-G11 O6	100.0	Arg-73 NH1-G11 O6	83.9
	Arg-73 NH1-U10 O2P	61.5	Arg-73 NH1-U10 O4′	84.9
	Arg-73 NH2-G11 N7	99.3	Arg-73 NH1-G11 N7	87.1
	Arg-73 NH2-G11 O6	59.1	Arg-73 NH2-G9 O1P	99.2
	Arg-73 NH2-U10 O2P	75.0	Arg-73 NH2-G9 O5′	83.6
			Arg-73 NH2-G10 O5′	84.7
			Arg-73 NH2-G10 O4′	78.1
			Arg-73 O-C25 N4	99.9
Arg-77	Arg-77 NE-G9 N7	100.0	Arg-77 NE-U7 O2P	95.1
	Arg-77 NE-G9 O6	59.7	Arg-77 NH2-U7 O2P	98.7
	Arg-77 NH2-G9 N7	51.0	Arg-77 NH2-C6 O1P	92.9
	Arg-77 NH2-G9 O6	100.0	Arg-77 NH2-C6 O5′	52.3
Arg-78			Arg-78 N-C8 O1P	98.2
Arg-80			Arg-80 N-G9 O1P	66.6
			Arg-80 N-G9 O2P	53.4
Arg-81	Arg-81 NH1-G9 O1P	85.5		
	Arg-81 NH1-G9 O2P	87.7		
	Arg-81 NH2-G9 O1P	74.9		
	Arg-81 NH2-G9 O2P	94.1		

mutations that disrupt interactions with the phosphate oxygens (Arg-70Lys, Thr-72Ala, and Arg-73Lys) have a tendency to be more unfavorable than experimental measurements. This deviation highlights the importance of interactions with the phosphate backbone, but it also highlights the inability of the MM-PBSA, at least as implemented with simple minimization or trajectory scanning, to

quantitatively reproduce the mutation free energies. We are currently investigating whether further improvement of the continuum solvent parameters (dielectric constant, continuum atomic radii) will improve the agreement between calculated and experimental free energies (J. Wang et al., work in progress).

CONCLUSIONS

We have presented a theoretical investigation of BIV Tat peptide-TAR RNA binding using molecular dynamics and free energy analysis with continuum solvent. We performed our simulations on two NMR structures determined by Puglisi and co-workers and Patel and co-workers. We found that last 300-ps trajectory averaged RNA structures are more similar than the initial experimental structures. The Patel RNA structure relaxed its interhelical angle bend over the course of the simulation. We observed these structural changes with our MD protocol, using explicit water, periodic boundary conditions, and full treatment of electrostatics with particle-mesh Ewald. Our MD simulations possibly corrected some inaccuracies from the initial experimental

TABLE 3   Computational mutagenesis on Patel structure by minimization and trajectory scanning methods. The standard deviations are given in parentheses

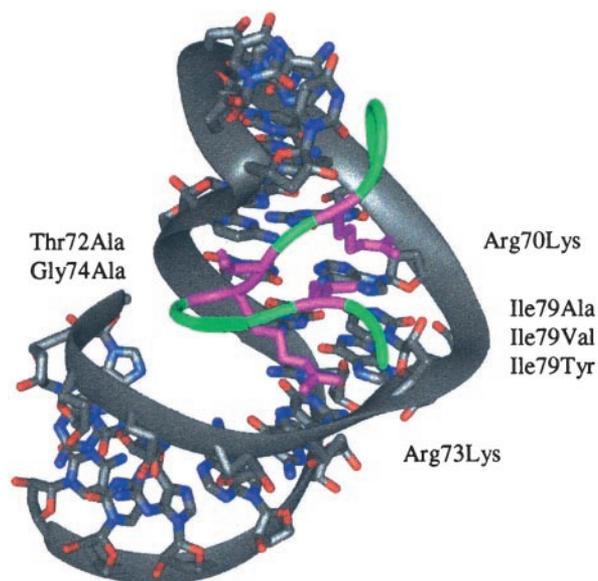| Mutant	Relative Binding Free Energies, $\Delta\Delta G$ (kcal/mol)		
	Minimization	Trajectory Scanning	Experimental*
Arg-70Lys	5.7	—	>2.7
Thr-72Ala	7.9	6.8 (1.4)	>2.7
Arg-73Lys	14.8	—	>2.7
Gly-74Ala	8.2	—	>2.7
Ile-79Ala	2.1	−0.64 (1.1)	1.5
Ile-79Val	4.9	−0.91 (3.0)	0.9
Ile-79Tyr	4.8	—	0.6

*Chen and Frankel, 1995.

FIGURE 8   The mutated residues are shown in magenta.

structural modeling of both structures, which were done in the absence of explicit water and with a less accurate force field. The flexible peptides, however, and RNA-peptide contacts do not similarly converge.

The U10-A13-U24 base triple, which was not directly observed by Puglisi et al., became more co-planar over the course of the MD simulation. MD simulations of the HIV TAR RNA bound to argininamide also observed the formation of an analogous U-A-U triple (Nifosi et al., 2000), which supports the hypothesis that the base triple is a common characteristic to both the HIV and BIV TAR RNAs in their bound conformation. We also computed the relative binding of different Tat peptide mutants, achieving qualitative agreement with experimental results and emphasizing the significance of interactions with the phosphate backbone. It is particularly interesting that mutation of Arg-73, which we computed to be the most deleterious, is analogous to the specificity determining arginine in the HIV Tat-TAR recognition.

REFERENCES

Aboul-ela, F., J. Karn, and G. Varani. 1995. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253:313–332.

Aboul-ela, F., and G. Varani. 1998. Recognition of HIV-1 TAR RNA by Tat protein and Tat-derived peptides. *Theochem. J. Mol. Struct.* 423: 29–39.

Bayer, P., M. Kraft, A. Ejchart, M. Westendorp, R. Frank, and P. Rosch. 1995. Structural studies of HIV-1 Tat protein. *J. Mol. Biol.* 247: 529–535.

Berendsen, H. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, The Netherlands. 331–342.

Burley, S. K., and G. A. Petsko. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science.* 229:23–28.

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.

Cheatham, T. E., and P. A. Kollman. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259:434–444.

Cheatham, T. E., J. Srinivasan, D. A. Case, and P. A. Kollman. 1998. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.* 16:265–280.

Chen, L., and A. D. Frankel. 1995. A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl. Acad. Sci. U.S.A.* 92:5077–5081.

Chong, L. T., Y. Duan, L. Wang, I. Massova, and P. A. Kollman. 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U.S.A.* 96: 14330–14335.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179, 1995.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N · log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089–10092.

Ferrin, T. E., C. C. Huang, L. E. Jarvis, and R. Langridge. 1988. The MIDAS display system. *J. Mol. Graphics.* 6:13–27.

Frankel, A. D. 2000. Fitting peptides into the RNA world. *Curr. Opin. Struct. Biol.* 10:332–340.

Greenbaum, N. L. 1996. How Tat targets TAR: structure of the BIV peptide RNA complex. *Structure.* 4:5–9.

Hermann, T., and E. Westhof. 1999. Simulations of the dynamics at an RNA-protein interface. *Nat. Struct. Biol.* 6:540–544.

Jones, K. A., and B. M. Peterlin. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* 63:717–743.

Jorgensen, W. L., J. Chandreskhar, J. D. Madura, R. W. Imprey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Massova, I., and P. A. Kollman. 1999. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121:8133–8143.

Massova, I., and P. A. Kollman. 2000. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design.* 18:1–23.

Mujeeb, A., T. G. Parslow, Y. C. Yuan, and T. L. James. 1996. Aqueous solution structure of a hybrid lentiviral Tat peptide and a model of its interaction with HIV-1 TAR RNA. *J. Biomol. Struct. Dyn.* 13:649–660.

Nicholls, A., K. A. Sharp, and B. Honig. 1990. DelPhi. Department of Biochemistry and Molecular Biophysics, Columbia University, New York.

Nifosi, R., C. Reyes, and P. A. Kollman. 2000. Molecular dynamics studies of the HIV-1. TAR and its complex with argininamide. *Nucleic Acids Res.* 28:4944–4955.

Puglisi, J. D., L. Chen, S. Blanchard, and A. D. Frankel. 1995. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science.* 270:1200–1203.

Puglisi, J. D., L. Chen, A. D. Frankel, and J. R. Williamson. 1993. Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. U.S.A.* 90:3680–3684.

Puglisi, J. D., R. Y. Tan, B. J. Calnan, A. D. Frankel, and J. R. Williamson. 1992b. Conformation of the TAR RNA-Arginine Complex By NMR Spectroscopy. *Science.* 257:76–80.

Rana, T. M., and K. T. Jeang. 1999. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* 365:175–185.

Reyes, C. M., and P. A. Kollman. 1999. Molecular dynamics studies of U1A-RNA complexes. *RNA-a Publication of the RNA Society.* 5:235–244.

Reyes, C., and P. A. Kollman. 2000a. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* 297:1145–1158.

Reyes, C. M., and P. A. Kollman. 2000b. Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* 295: 1–6.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* 23: 327–341.

Sanner, M. F., A. J. Olson, and J. C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 38:305–320.

Sharp, K. A., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Smith, C. A., S. Crotty, Y. Harada, and A. D. Frankel. 1998. Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry.* 37:10808–10814.

Smith, K. C., and B. Honig. 1994. Evaluation of the conformational free energies of loops in proteins. *Proteins: Struct., Funct., Genet.* 18: 119–132.

Srinivasan, J., T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case. 1998a. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.

Srinivasan, J., J. Miller, P. A. Kollman, and D. A. Case. 1998b. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16:671–675, 677–682.

Sticht, H., D. Willbold, P. Bayer, A. Ejchart, F. Herrmann, R. Rosinarbesfeld, A. Gazit, A. Yaniv, R. Frank, and P. Rosch. 1993. Equine infectious anemia virus Tat is a predominantly helical protein. *Eur. J. Biochem.* 218:973–976.

Sussman, J. L., D. W. Lin, J. S. Jiang, N. O. Manning, J. Prilusky, O. Ritter, and E. E. Abola. 1998. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. D.* 54:1078–1084.

Tan, R. Y., and A. D. Frankel. 1995. Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. U.S.A.* 92:5282–5286.

Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305.

Tao, J., and A. D. Frankel. 1992. Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. U.S.A.* 89:2723–2726.

Willbold, D., A. U. Metzger, H. Sticht, K. C. Gallert, R. Voit, N. Dank, P. Bayer, G. Krauss, R. S. Goody, and P. Rosch. 1998. Equine infectious anemia virus transactivator is a homeodomain-type protein. *J. Mol. Biol.* 277:749–755.

Ye, X. M., R. A. Kumar, and D. J. Patel. 1995. Molecular recognition in the bovine immunodeficiency virus Tat peptide TAR RNA complex. *Chemistry and Biology.* 2:827–840.