Ying Zhao
Bethany L. Kormos
David L. Beveridge
Anne M. Baranger
Chemistry Department and Molecular Biophysics Program, Wesleyan University, Middletown, CT 06459

Molecular Dynamics Simulation Studies of a Protein–RNA Complex with a Selectively Modified Binding Interface

Received 7 September 2005; revised 22 October 2005; accepted 24 October 2005
Published online 8 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20408

Abstract: *The RNA recognition motif (RRM) is one of the most common RNA binding domains. We have investigated the contribution of three highly conserved aromatic amino acids to RNA binding by the N-terminal RRM of the U1A protein. Recently, we synthesized a modified base (A-4CPh) in which a phenyl group is tethered to adenine using a linker of 4 methylene groups. The substitution of this base for adenine in the target RNA selectively stabilizes the complex formed with a U1A protein in which one of the conserved aromatic amino acids is replaced with Ala (Phe56Ala). In this article, we report molecular dynamics (MD) simulations that probe the structural consequences of the substitution of A-4CPh for adenine in the wild type and Phe56Ala U1A–RNA complexes and in the free RNA. The simulations suggest that A-4CPh stabilizes the complex formed with Phe56Ala by adopting a folded conformation in which the tethered phenyl group fills the site occupied by the phenyl group of Phe56 in the wild-type complex. In contrast, an extended conformation of A-4CPh is predicted to be most stable in the complex formed with the wild-type protein. The calculations indicate A-4CPh is in an extended conformation in the free RNA. Therefore, preorganizing the structure of the phenyl-tethered base for binding may improve both the affinity and specificity of the RNA containing A-4CPh for the Phe56Ala U1A protein. Taken together, the previous experimental work and the calculations reported here suggest a general design strategy for altering RRM–RNA complex stability.* © 2005 Wiley Periodicals, Inc. Biopolymers 81: 256–269, 2006

This article was originally published online as an accepted preprint. The ''Published Online'' date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

Keywords: *RNA-binding protein; U1A; RNA recognition motif; molecular dynamics; design*

INTRODUCTION

The RNA recognition motif (RRM) is one of the most common RNA-binding domains in eukaryotes.[1] Proteins containing RRMs are involved in all steps of RNA processing, including pre-mRNA splicing, RNA editing, mRNA export, and RNA degradation. Therefore, an understanding of RNA recognition by RRMs is important for describing and controlling these fundamental biological processes. To investigate the recognition principles involved in RRM–RNA complexes, we recently synthesized an altered

Correspondence to: Anne M. Baranger; e-mail: abaranger@wesleyan.edu
Contract grant sponsor: NIH, National Center for Supercomputing Applications, Texas Advanced Computing Center; contract grant number: GM-56857 and GM-37909 (NIH).

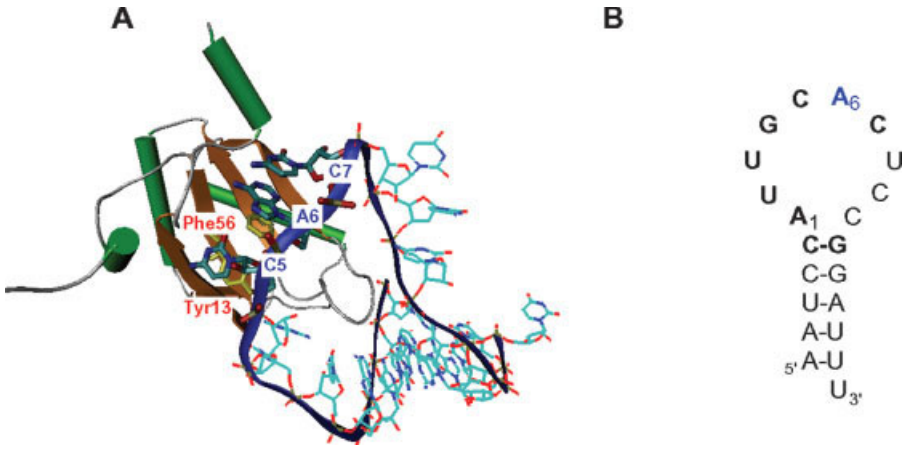

FIGURE 1 (A) Structure of the U1A-SL2 complex from the X-ray cocrystal structure.[3] Amino acids and bases involved in stacking are shown. (B) Stem loop 2 sequence used in the X-ray cocrystal and in the simulations reported in this article.

RRM–RNA interface by designing a modified RNA to bind selectively to a mutant RRM protein.[2] The complex formed between the modified RNA and the mutant RRM is more stable than the complexes formed by matching either with a wild-type component. Thus, the specificity of the complex is changed by concomitant protein mutation and RNA modification. In this article, we report molecular dynamics (MD) simulations that explore the structure and dynamics of the free and bound modified RNA and mutant RRMs, propose a molecular model for the altered specificity of the complex, and suggest a general design strategy for altering the target specificity of RRM proteins.

RRMs bind to single-stranded RNA target sites in diverse structural contexts.[1] Three of the most highly conserved amino acids that contact RNA are aromatic, and these have been shown to stack with RNA bases when bound to RNA.[3–13] Thus, stacking interactions appear to be a general means for RRMs to gain affinity for RNA. We are studying the U1A protein as a well-characterized model system for RRMs. The N-terminal RRM of the U1A protein binds to stem loop 2 (SL2) of U1 snRNA, a component of the spliceosome.[14–17] U1A contains two of the three conserved aromatic amino acids found in RRMs. Tyr13 stacks with C5 in the complex, while Phe56 stacks with A6 (Figure 1).[3] Mutation of Phe56 to Ala resulted in a large loss in binding affinity.[18,19] We have designed a modified base to selectively improve the affinity of the Phe56Ala U1A protein for RNA.[2] A phenyl group was tethered to adenine using either 3 or 4 methylene groups, called A-3CPh and A-4CPh, respectively, so that the phenyl group could fold to stack against the adenine and thus compensate for the missing phenyl group in Phe56Ala U1A (Figure 2). We found using gel mobility shift assays that the substitution of A-4CPh for A6 in SL2 RNA (A6-4CPh) resulted in a 1.8 kcal/mol increase in the stability of the complex formed with Phe56Ala U1A, while the substitution of A-3CPh for A6 (A6-3CPh) resulted in only a 0.6 kcal/mol increase in the stability of this complex. The tethered phenyl group could stabilize the complex by preventing structural distortions of the complex that may have occurred as a result of the Phe56Ala mutation, by interacting with amino acids surrounding Phe56, and by solvation effects. The effect of the tethered phenyl group was specific for the complex formed with Phe56Ala U1A; the stability of the complex formed with the wild-type protein was decreased by 0.8 kcal/mol, while the stability of the complex formed with the Phe56Leu U1A protein was unchanged by the tethered phenyl group.

To investigate the effect of A-4CPh on the structure and dynamics of the free SL2 RNA and the complex, MD simulations of the free wild-type and Phe56Ala U1A proteins, the free wild-type and A6-4CPh SL2 RNAs, and the complexes were performed. In all of the simulations A-4CPh was considered in

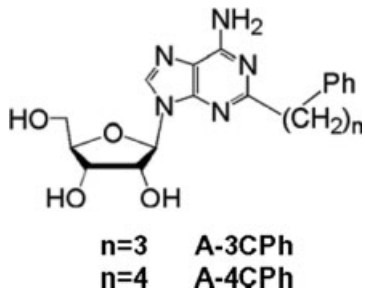

FIGURE 2 Modified base designed to compensate for the poor binding affinity of Phe56Ala U1A protein.[2]

Table I The MD Simulations Described in This Article

Name	Protein–RNA Complex
WT–WT complex	The complex formed between the wild-type U1A and the wild-type RNA
WT–folded A6-4CPh SL2 complex	The complex formed between the wild-type U1A and the A6-4CPh SL2 RNA with the folded tethered phenyl group
WT–extended A6-4CPh SL2 complex	The complex formed between the wild-type U1A and the A6-4CPh SL2 RNA with the extended tethered phenyl group
Phe56A1a–WT complex	The complex formed between the Phe56A1a U1A and the wild-type RNA
Phe56A1a–folded A6-4CPh SL2 complex	The complex formed between the Phe56A1a U1A and the A6-4CPh SL2 RNA with the folded tethered phenyl group
Phe56A1a–extended A6-4CPh SL2 complex	The complex formed between the Phe56A1a U1A and the A6-4CPh SL2 RNA with the extended tethered phenyl group
Free WT protein	Wild-type U1A protein
Free Phe56A1a protein	Phe56A1a U1A protein
Free WT RNA	Wild-type SL2 RNA
Free folded A6-4CPh SL2	A6-4CPh SL2 RNA with the folded tethered phenyl group
Free extended A6-4CPh SL2	A6-4CPh SL2 RNA with the extended tethered phenyl group

two conformations: one with the phenyl group in the folded conformation assumed in the design of A-4CPh and one with the phenyl group in an extended conformation. Thus, eleven simulations were performed (Table I). The results from these simulations suggest a solution structure for the complex formed between Phe56Ala U1A protein and A6-4CPh SL2 RNA, a likely cause of the destabilization of the wild-type complex compared to the stabilization of the complex formed with the Phe56Ala U1A protein, and reasons for the greater stability of the complex formed with the modified base containing 4 methylene groups (A-4CPh) than that containing 3 methylene groups (A-3CPh). Because the stacking interaction is highly conserved in RRM–RNA complexes, these calculations will inform future design strategies to develop a general approach for altering the affinity and specificity of RNA–RRM complexes.

METHODS

The structure of the extended A-4CPh free nucleotide was generated from adenosine using the program LEaP in AMBER 6.[29] The structure of the folded A-4CPh free nucleotide was generated by changing the dihedral angles of the butyl linker so that the tethered phenyl ring stacked parallel with the purine ring of adenosine. As the A-4CPh nucleoside is not included in the AMBER 6 library, parameters were developed for the modified nucleoside by following the protocol described for parameter development in the AMBER manual. Following the geometry optimization and the static potential calculations in Gaussian98 using Hartree–Fock (HF) theory[20–22] and the 6-31G(d) basis set,[23,24] the atomic charges of the A-4CPh free nucleotides were calculated by means of the restrained electrostatic potential (RESP) method.[25–27] The atom types and charges of tethered phenyl group and the connectivity of the molecule were added manually to the library file of A-4CPh free nucleotide. The r_{eq}, θ_{eq}, K_r, and K_θ values for the missing bond and angle parameters were taken to be those of similar cases already represented in the force field. Since the prescribed method for parameter development was followed quite rigorously, we expect the results reported herein to accurately portray the effects caused by the introduction of the modified A-4CPh nucleoside to the system (within the confines of AMBER and the Cornell et al. force field).

The initial structure for the MD simulations of the U1A–stem loop complexes and the free U1A proteins and stem loop 2 RNAs were derived from the structure of the U1A–stem loop 2 cocrystal.[3] As described previously,[28] six C-terminal amino acids were added and two mutations (H31Y and R36Q) were introduced to reproduce the experimental wild-type 2–102 system. The same modified U1A structure was used as the initial structure for the simulations of the free wild-type and Phe56Ala U1A proteins. A 25mer stem loop RNA was used in the previously reported biochemical studies.[2] However, the RNA in the crystal structure of the complex and in other MD studies was a 21mer RNA (Figure 1).[3,28] Previous studies have shown that the length of the RNA stem is not important for binding of

U1A.[43] Therefore, the 21mer RNA was used in the calculations described in this paper. Both *anti* and *syn* configurations of the modified A-4CPh nucleosides were considered, along with extended and folded geometries of the tethered phenyl group. A series of short simulations of the free RNA suggested that different structures and atomic charge sets of A-4CPh would not change the simulation results significantly. Therefore, in all of the simulations the A-4CPh is in an *anti* geometry with charges derived from A-4CPh in an *anti* geometry with the tethered phenyl group in an extended conformation. The A-4CPh modification was introduced to the A6 position of the free stem loop 2 RNA and the complexes in both the folded and extended geometries using LEaP. For the complexes formed between the U1A protein and A6-4CPh SL2 RNA, adding the tethered phenyl ring caused some steric clashes in the complexes. These clashes were removed by energy minimization in vacuum before the systems were set up for MD simulations. All molecular dynamics simulations were performed using the AMBER 6.0[29] suite of programs following the protocol described by Pitici et al.[28] There was a minimum distance of 9 Å between the solute atoms and each face of the water box. A certain number of Na^+ and Cl^- ions, which were calculated based on the water box volume, were subsequently added to the system to reach 250 m*M* salt concentration. The production run continued to 5 ns for the simulations of the free RNA and to 4 ns for the simulations of the complexes.

Average structures of the simulations, one-dimensional root mean square deviation (1D-RMSD) analyses, and sugar pseudorotation phase angles, were calculated using *ptraj* from AMBER 6.[29] Two-dimensional (2D)-RMSD analyses were performed using the 2D-RMSD function from the MD Toolchest, a packaged suite of tools developed by Ravishanker et al.[30] The intermolecular interactions in each U1A–RNA complex were calculated using the NUCPLOT program.[31] The cutoff distance of hydrogen-bonding interactions was 3.2 Å, and the cutoff distance of nonbonded interactions was 3.9 Å.

Free energies of binding were estimated using MM-PBSA in AMBER 8[32] according the method of Gohlke and Case,[33] using an additive free energy component analysis approach.[34–38] We chose this method because we were interested in a comparatively quick energetic comparison in support of the results observed from the simulations and did not feel the extensive computational time and resources required to perform more rigorous free energy approaches such as replica exchange or umbrella sampling were warranted. In addition, we have successfully used this method to reproduce observed binding affinities for other closely related U1A systems (manuscript in preparation), leading us to believe MM-PBSA can be a useful tool for gaining insight into the energetics of this system. Snapshots for analysis were collected every 20 ps over the final 1–3 ns of the molecular dynamics trajectories, as determined from convergence of RMSD plots. Electrostatic contributions to the solvation free energy were computed using the modified generalized Born (GB) method developed by Case and coworkers,[39,40] defined as model I in Ref. 41, and called by IGB = 2 in AMBER 8. The hydrophobic contribution to the solvation free energy was determined using solvent-accessible surface area (SASA) terms[41] computed with the program *molsurf*,[42] and entropic contributions were estimated using the normal mode analysis module *nmode* implemented in AMBER 8.

RESULTS

To explore the ability of the wild-type and Phe56Ala U1A proteins to accommodate folded and extended geometries of A-4CPh, eleven MD simulations were carried out (Table I). Six were simulations of the structures of the complexes formed between the wild-type and Phe56Ala U1A proteins and the wild-type SL2 RNA, SL2 RNA containing an *anti* extended tethered phenyl group at A6, and SL2 RNA containing the *anti* folded tethered phenyl group at A6. In addition, the structures of the free wild-type, extended and folded A6-4CPh SL2 RNAs and the structures of the free wild-type and Phe56Ala U1A proteins were simulated.

Comparison of the MD Simulations of the Free and Bound U1A Proteins

1D-RMSD Analyses. To evaluate the overall structural changes in the U1A protein that occurred during the simulations, 1D-RMSD analyses were performed. In all of the simulations, the protein equilibrated to a stable structure within 2–2.5 ns. The average RMSD values over the last 2 ns of the simulations are presented in Table II. For the simulations involving the wild-type protein, the protein in the complex formed with the folded A6-4CPh SL2 RNA is the most different from the initial structure with a RMSD value of 2.22 Å, while the protein in the complex with the extended A6-4CPh SL2 RNA is the most similar to the initial structure (RMSD of 1.18 Å). In contrast, for the simulations involving the Phe56Ala U1A protein, the protein in the complex with the folded A6-4CPh SL2 RNA is most similar to the initial structure (RMSD of 1.22 Å), while the protein in the complex with the extended A6-4CPh SL2 RNA is most different from the initial structure (RMSD of 1.72 Å). This data is consistent with our hypothesis that the folded phenyl group could fit into the cavity formed by the Phe to Ala mutation, which formed the basis for the original design of A-4CPh.[2]

Comparison of Simulated Structures. The structures of the wild-type (WT) proteins averaged over the last 2 ns of the simulations are superimposed in Figure 3A.

Table II Average RMSD Values of the Wild-type and Phe56A1a U1A Proteins Over the Last 2 ns of the Simulations Calculated Using Backbone Atoms[a]

Protein	RMSD
WT–folded A6-4CPh SL2 complex	2.22 (0.06)
WT–extended A6-4CPh SL2 complex	1.18 (0.12)
WT–WT complex	1.25 (0.11)
Free WT protein	1.63 (0.34)
Phe56A1a–folded A6-4CPh SL2 complex	1.22 (0.15)
Phe56A1a–extended A6-4CPh SL2 complex	1.72 (0.11)
Phe56A1a–WT complex	1.36 (0.11)
Free Phe56A1a complex	1.65 (0.20)

[a] Numbers inside the parentheses are the standard deviation of the RMSD.

These structures are quite similar to each other with the flexible regions, loop 3 and helix C, showing the most variation. As expected from the 1D-RMSD analyses, the MD structure of the wild-type U1A protein in the complex with folded A6-4CPh SL2 RNA differs most from the starting structure. The most significant change in structure is observed for helix C, whose helical axis rotated about 90° to direct the C-terminal end of helix C toward the β-sheet so that the C-terminal residues of helix C are within range to form van der Waals interactions with residues on loop 4. In addition, loop 6 moved away from the β-sheet, while both loops 3 and 5 moved away from the RNA. The structures of the Phe56Ala proteins averaged over the last 2 ns of the simulations superimposed better than the wild-type proteins (Figure 3B). Small structural variations were observed for loop 3 and helix C.

Comparison of the MD Simulations of the Free and Bound SL2 RNAs

1D- and 2D-RMSD Analyses. The all-atom RMSDs of SL2 RNAs were calculated using the initial structures of RNAs in each simulation as reference structures (Table III). In general, the free RNAs underwent more significant structural changes (average RMSD ~4.4 Å) during the simulations than the RNAs in the complexes (average RMSD ~2.3 Å). The stem regions of the free RNAs had lower RMSDs than the loop regions compared to the reference structures. In the initial structure, which was derived from the structure of the wild-type complex,[3] the bases on the loop region were spread out of the loop and exposed to form the interface for protein binding. During all of the free RNA simulations, many of these bases moved toward the interior of the loop and were no longer solvent exposed (Figure 4). This movement of bases toward the interior of the loop has been observed in previous MD simulations of the free SL2 RNA.[29,45,46] It is interesting that the interior of the loop is able to accommodate the large modified base A-4CPh.

The MD simulations of the free RNA were compared to each other using 2D-RMSD plots (Figure 5). A small structural transition was observed at approximately 4 ns in the free wild-type SL2 RNA simulation. During the free extended A-4CPh SL2 RNA simulation (Figure 5), the RNA structures varied around the starting structure for about 1.5 ns, changed to another structural form, and at 4 ns made another structural transition. These structural forms are significantly different from each other (RMSD ~5 Å). Dur-

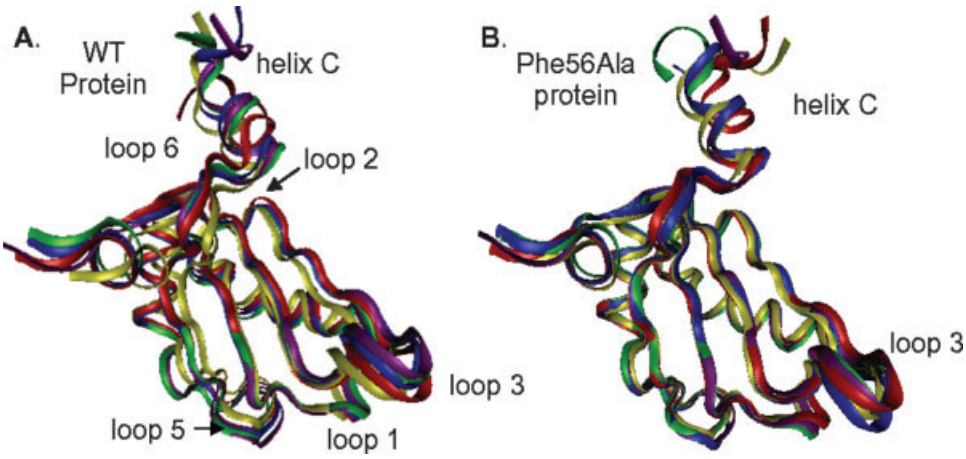

FIGURE 3 Superimposed protein structures averaged over the last 2 ns of the simulations of the WT (A) and Phe56Ala (B) U1A proteins. Purple: the initial structure of WT U1A; yellow: free protein; green: protein in complex with WT SL2 RNA; blue: protein in complex with the extended A6-4CPh SL2 RNA; red: protein in complex with the folded A6-4CPh SL2 RNA. The structures were oriented by superimposing the Cα atoms of residues 2–90.

Table III Average RMSD Values of RNA Over the Last 2 ns of the Simulations Calculated Using All Atoms[a]

RNA	Stem Loop	Stem	Loop
Phe56A1a–folded A6-4CPh SL2 complex	2.04 (0.23)	1.80 (0.26)	1.78 (0.31)
WT–folded A6-4CPh SL2 complex	2.52 (0.24)	2.01 (0.18)	2.11 (0.13)
Free folded A6-4CPh SL2	4.40 (0.47)	2.57 (0.23)	4.20 (0.20)
Phe56A1a–extended A6-4CPh SL2 complex	2.35 (0.39)	1.38 (0.25)	2.29 (0.24)
WT–extended A6-4CPh SL2 complex	2.92 (0.59)	2.26 (0.45)	2.55 (0.37)
Free extended A6-4CPh SL2	4.78 (0.34)	1.61 (0.31)	4.97 (0.18)
Phe56A1a–WT complex	1.71 (0.23)	1.55 (0.26)	1.23 (0.16)
WT–WT complex	2.25 (0.23)	1.93 (0.16)	1.69 (0.13)
Free SL2 RNA	4.15 (0.43)	1.81 (0.26)	4.31 (0.41)

[a]Numbers inside the parentheses are the standard deviation of the RMSD.

ing the free folded A6-4CPh SL2 RNA simulation, the structure oscillated around the starting structure for about 1.6 ns, changed to a new structural form, and then changed to a third structural form at approximately 3.2 ns. The three structural forms are also different from each other (RMSD 4–5 Å).

Comparison of the Simulated Structures. The structural changes in the free RNA that were identified in the 2D-RMSD plots are shown in Figure 6. The large and global dynamical changes in the loop conformation are responsible for the large RMSD differences between the different SL2 RNAs. For the wild-type RNA, the loop region moved away from the RNA–protein interface. For the extended A6-4CPh SL2 RNA, the loop region also moved away from the starting position during the first 4-ns simulation. However, the loop region returned to the starting position during the last 1 ns of the simulation. For the folded A6-4CPh SL2 RNA, the loop region moved away from the starting structure in the first 3 ns. After moving toward the starting position from 3 to 4 ns, it moved away from the starting position in the last 1 ns. During the time course of the free folded A6-4CPh SL2 RNA simulation, the folded tethered phenyl ring opened up to an extended conformation (Figure 4). The final structures formed in the free folded and extended A6-4CPh SL2 RNA simulations are compared in Figures 4 and 7A and had an RMSD difference of 2.92 Å.

The MD structures of the RNA in the complexes with U1A were more stable and similar to each other than the MD structures of the free RNA. In the complexes with the wild-type U1A protein, the folded and extended A6-4CPh SL2 RNAs superimpose well with the wild-type RNA structure (Figure 7B). In the complexes with the Phe56Ala U1A protein, the general structures of the folded and extended A6-4CPh SL2 RNAs are also similar to the wild-type RNA structure (Figure 7C), but the region corresponding to the last three bases of the loop did not superimpose well. Since these three bases are not conserved and the number of bases can be varied without affecting the binding affinity,[47] structural variation in this region may not be important. In contrast to the simulations of the free RNA, the folded tethered phenyl ring remained in a folded conformation during the course of the simula-

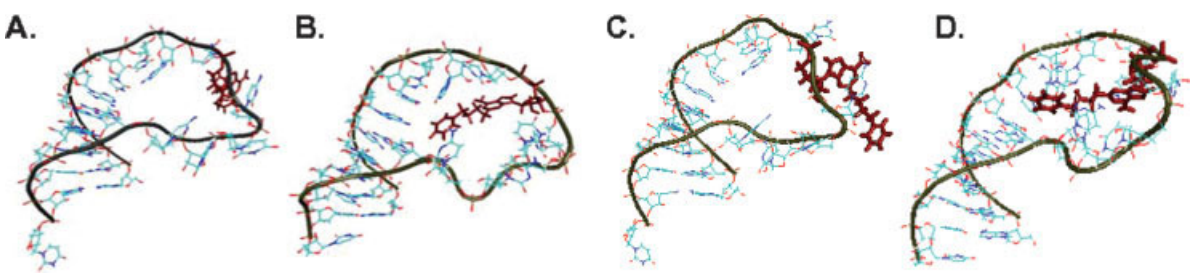

FIGURE 4 Comparison of the initial (A) and final (B) structures from the MD simulation of the free folded A6-4CPh SL2 RNA and initial (C) and final (D) structures from the MD simulation of the extended A6-4CPh SL2 RNA. The differences in the orientations of the loop bases are illustrated.

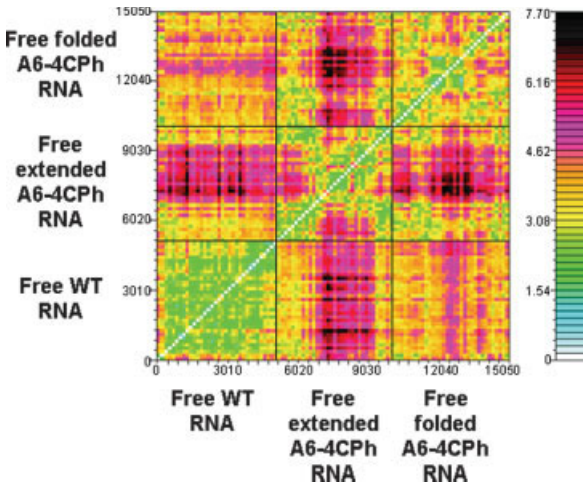

FIGURE 5 All atom 2D-RMSD plots of the free WT and A6-4CPh SL2 RNAs. The color scale ranges from 0 to 7 Å, with light blue associated with lower RMSD (more similar structures) and red or black associated with the higher RMSD (less similar structures).

tions of the complexes formed with both the WT and Phe56Ala U1A proteins.

Analysis of Sugar Pucker. The conformation of each ribose (the sugar pucker) in the AUUGCAC sequence in the loop of SL2 RNA was evaluated by creating plots of the pseudorotation angle every 5 ps vs. time (Figure 8).[47,48] For the free RNA, significant differences between the sugar puckers of nucleotides in the wild-type and modified SL2 RNAs were observed at nucleotides U2 through A6 (Figure 8B). The sugar puckers of U2 in both the free folded A6-4CPh and wild-type SL2 RNAs were approximately A form, while that of the free extended A6-4CPh SL2 RNA was approximately B form. The sugar pucker of U3 was more dynamic in the wild-type RNA than in the folded or extended A6-4CPh SL2 RNAs. For G4, the conformation of the sugar in the free wild-type SL2 RNA ranged from C2′-*exo* to C3′-*endo*, while the conformation of G4 in extended and folded A6-4CPh SL2 RNAs was C2′-*endo*. For C5, the sugar in the wild-type SL2 RNA was C2′-*endo*, while the sugar conformations of the folded and extended A6-4CPh SL2 RNAs were more dynamic than the wild-type RNA. The A6 sugar conformation was especially variable. For the free folded A6-4CPh SL2 RNA, the sugar pucker of A6 changed from a C3′-*endo* conformation to a C3′-*exo*/C4′-*endo* conformation at 1265 ps, then changed to a C2′-*exo*/C1′-*endo* conformation at 2110 ps. For free extended A6-4CPh SL2 RNA, the sugar pucker of A6 changed from a C2′-*exo* to a C3′-*exo* conformation at 1010 ps and stayed there for the remainder of the simulation. The sugar pucker of A6 in free WT RNA stayed in a C3′-*endo* conformation during the first 3 ns simulations, then changed to a C4′-*exo* conformation.

As expected, the sugar puckers of the loop nucleotides bound to U1A proteins were more similar to each other and more stable than those of the free RNA. Significant differences in sugar pucker were observed only for A6 and C7 (Figure 8C). In

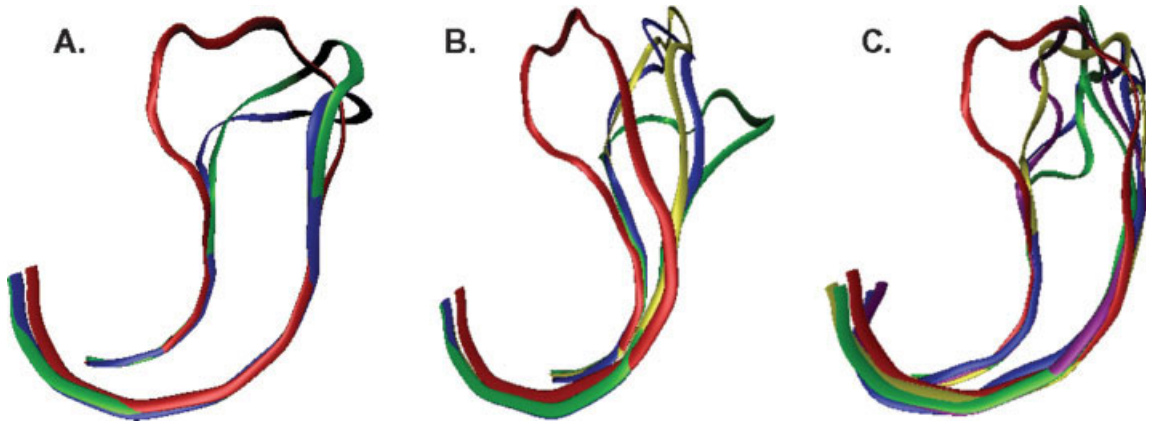

FIGURE 6 Superimposed structures of the free SL2 RNAs. The structures were oriented by superimposing the residues in the stem region. (A) Free WT RNA. Red: initial structure; green: the average structure of the first 4 ns; blue: the average structure between 4 and 5 ns. (B) Free extended A6-4CPh SL2 RNA. Red: initial structure; yellow: the average structure of the first 2 ns; green: the average structure between 2 and 4 ns; blue: the average structure of the last 1 ns. (C): Free folded A6-4CPh RNA. Red: initial structure; yellow: the average structure of the first 1.5 ns; green: the average structure between 1.5 and 3 ns; purple: the average structure between 3 and 4 ns; blue: the average structure of the last 1 ns.

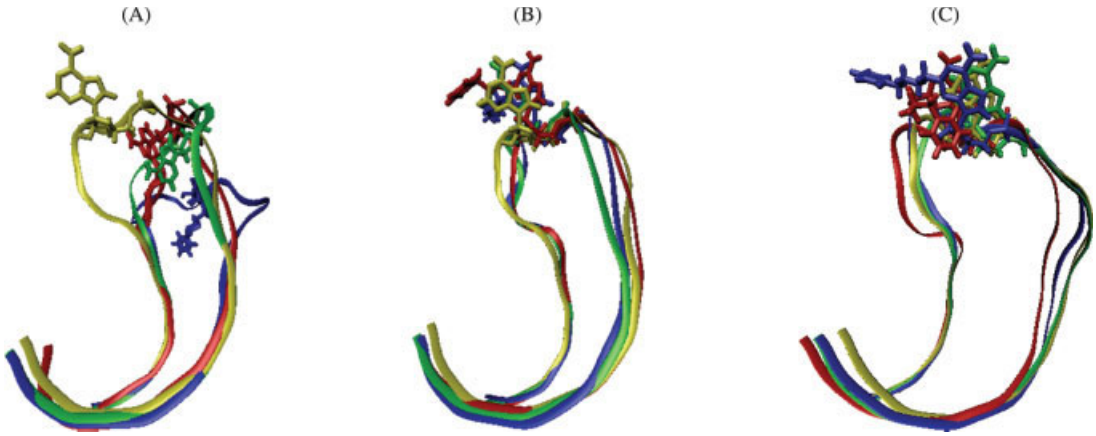

FIGURE 7 Superimposed structures from the MD simulations of the SL2 RNA (A) free, (B) bound to wild-type U1A, and (C) bound to Phe56Ala U1A averaged over the last 2 ns of the simulations. The structures were oriented by superimposing the residues in the stem region. The base orientations of the nucleotide at the A6 position of the loop are shown. Blue: extended A6-4CPh SL2 RNA; red: folded A6-4CPh SL2 RNA; green: wild-type SL2 RNA; yellow: wild-type SL2 RNA from the cocrystal structure.[3]

the complex between the wild-type U1A and the extended A6-4CPh SL2 RNA, the sugar pucker of the extended A-4CPh changed from A form to B form at 3435ps. In contrast, in the structures of the other complexes, A6 remained in the A form conformation. Similarly, in the wild-type complex, the sugar pucker of C7 changed to B form at 2020 ps. However, the sugar pucker of C7 in the Phe56Ala U1A–wild-type SL2 RNA complex remained mostly in the C3′-*endo* conformation, although some transitions toward the C2′-*endo* conformation were observed. The sugar pucker of C7 in the complexes formed between the extended and folded A6-4CPh SL2 RNAs and the wild-type and Phe56Ala U1A proteins quickly changed to a set of conformations that included the C2′-*endo* conformation. In general, the sugar pucker of the wild-type RNA was similar when free and bound. In contrast, the MD simulations suggest that a conformational change of the ribose is required at two nucleotides (G4 and A6) for the folded and three nucleotides (U2, G4, and A6) for the extended A6-4CPh SL2 RNAs upon binding U1A.

Analysis of RNA Conformation. The program AMIGOS was used to compare the conformations of the backbone in the wild-type and A6-4CPh SL2 RNAs.[49] This program simplifies the six backbone torsional angles of the RNA by the introduction of pseudobonds that connect the P and C4′ backbone atoms. The pseudotorsions η and θ are defined based on these pseudobonds as shown in Figure 9. The η and θ plots for nucleotides in the bound RNA were nearly identical (data not shown), suggesting that the introduction of the tethered phenyl group did not alter the backbone conformation of the bound RNA. In contrast, significant differences between the η and θ values for nucleotides in the free wild-type, extended A6-4CPh, and folded A6-4CPh SL2 RNAs were observed (Figure 9). Consistent with the analysis of sugar puckers, discussed above, the SL2 RNAs containing A6-4CPh are predicted to undergo greater conformational changes than the wild-type SL2 RNA upon binding U1A protein. Only the η and θ values of C5 and C7 of the wild-type SL2 RNA changed significantly upon binding protein. In contrast, the η and θ values of four nucleotides (G4–C7) of the folded A6-4CPh SL2 RNA and of seven nucleotides (A1–C7) of the extended A6-4CPh SL2 RNA were different in the free and bound forms. These conformational changes could contribute to the decreased binding affinity of the A6-4CPh SL2 RNA compared to the wild-type SL2 RNA for the wild-type U1A protein and would also be expected to reduce the affinity of Phe56Ala U1A for A6-4CPh SL2 RNA.

Analysis of the Interface Between U1A and SL2 RNA

Comparison of the Structures of the Complexes. The MD simulations of the complexes suggest that the interface of the complex formed between the wild-type U1A protein and the folded A6-4CPh SL2 RNA must undergo significant structural rearrangements to accommodate the phenyl tether (Figure 10B). Favorable edge–face interactions may stabilize the placement of the tethered phenyl ring of A-4CPh at

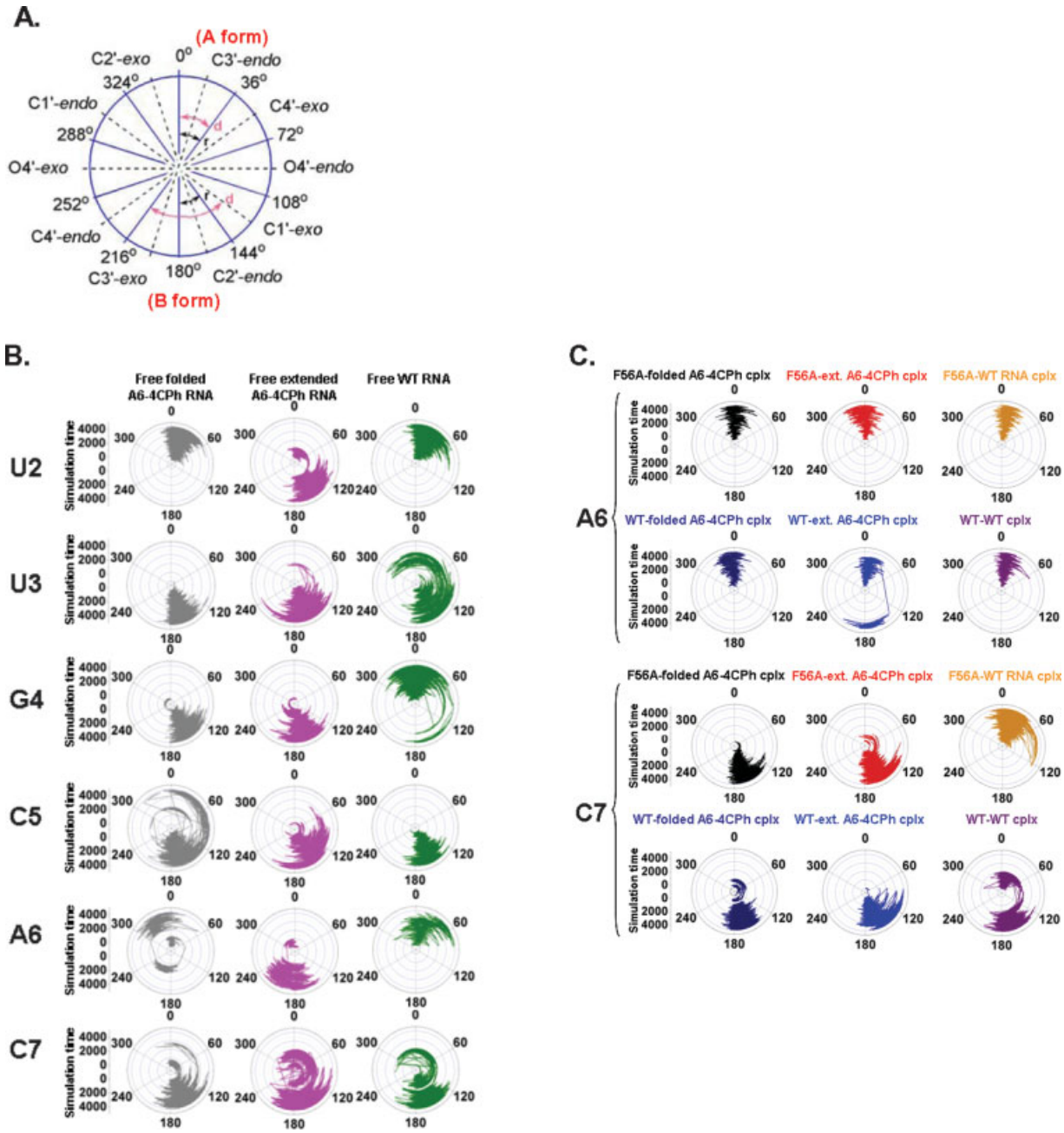

FIGURE 8 Sugar pucker diagram. (A) The sugar pseudorotation angle diagram. (B) The sugar pseudorotation angle diagram for U2–C7 of the free SL2 RNAs during the first 4 ns of the simulations. (C) The sugar pseudorotation angle diagram for A6 and C7 in the complexes with protein during the 4-ns simulations.

an angle to the adenine. Although the purine ring of A-4CPh is still stacked with C7, both rings have moved away from their original positions. The angle between the purine ring in its original position and in the new position is about 56°.

Fewer structural modifications are required to accommodate the phenyl tether in the complex formed between the Phe56Ala U1A protein and the folded A6-4CPh SL2 RNA (Figure 10D). The tethered phenyl ring of A-4CPh changes its position from overlapping with the purine ring to off-center stacking. However, the angle between the tethered phenyl ring and the purine ring is not changed. The tethered phenyl group moves away from Phe56Ala residue to a position close to Tyr13 and forms a T-shaped stacking interaction with Tyr13. The stacking interaction

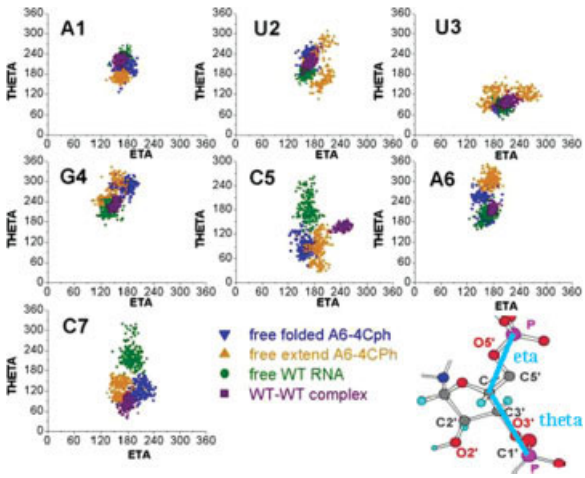

FIGURE 9 AMIGOS η and θ plots of the conserved AUUGCAC nucleotides in the loop region of the SL2 RNA for snapshots every 5 ps of the last 1 ns of the simulations.

between Tyr13 and C5 is not disturbed. The distance between Tyr13 and the phenyl ring of A-4CPh is about 5 Å. There are also nonbonding interactions between Phe56Ala and the butyl linker. All of these interactions may contribute to the stability of the Phe56Ala–folded A6-4CPh SL2 RNA complex.

In the average structure from the MD simulation of the wild-type protein–extended A6-4CPh SL2 RNA complex (Figure 10C), the conserved stacking interactions among Phe56, the purine ring of A-4CPh, and C7 are maintained. The tethered phenyl ring is pointing out of the protein–RNA interface, toward the loop 6 region between the β4 strand and the helix C. The extended A6-4CPh SL2 RNA adopts a similar structure in the complex with the Phe56Ala protein as in the complex with the wild-type protein (Figure 10E).

Analysis of Intramolecular Contacts. In general, the tethered phenyl group in A6-4CPh causes modifications throughout the interface between the SL2 RNA and the U1A protein in the structures predicted by these simulations. Significant changes in stacking interactions are observed in the wild-type protein–folded A6-4CPh SL2 RNA complex. As discussed earlier, the folded A-4CPh at the A6 position disrupts the four-step-stacking interaction among Phe56–A6–C7–Asp92 to form new edge-to-face interactions among Phe56, the purine ring of A-4CPh, and the tethered phenyl group of A-4CPh. Changes in other nonbonded contacts are found between the highly flexible loop 3 region and SL2 RNA. The complex formed between the wild-type protein and folded A6-4CPh SL2 RNA contains a large number of nonbonded contacts because the tethered phenyl group is folded into a tight binding interface.

The analysis of hydrogen bonding in the complexes demonstrates the ability of the complex to adjust to the Phe56Ala mutation and the tethered phenyl group. For example, in the complex formed between the wild-type protein and the wild-type or modified RNAs, N2 of G4 forms a hydrogen bond with the side chain of Glu19. However in the complex formed between the Phe56Ala U1A protein and the modified RNA, N2 of G4 forms a hydrogen bond with the main-chain carbonyl of Leu49. Overall, more hydrogen bonds are predicted to be present in the complex of the Phe56Ala U1A protein with the folded A6-4CPh SL2 RNA than with the extended A6-4CPh SL2 RNA. In contrast, the simulations suggest that there are fewer hydrogen bonds formed in the complex of the wild-type protein with the folded A6-4CPh SL2 RNA than with the extended A6-4CPh SL2 RNA. Several base functional groups that form hydrogen bonds in other complexes do not participate in any hydrogen bonds in the complex of the wild-type protein with the folded A6-4CPh SL2 RNA. For example, N1 of A6 and N4 of C7 do not form hydrogen bonds with any protein functional groups in this complex, while N1 of A6 typically forms a hydrogen bond with the side chain of Ser91 and N4 of C7 forms hydrogen bonds with the carbonyl of Asp90 or the side chain of Thr89. Thus, the analysis of the predicted hydrogen-bonding pattern in these complexes is consistent with A6-4CPh SL2 RNA being in an extended conformation in the complex formed with the wild-type protein, but in a folded conformation in the complex formed with the Phe56Ala U1A protein. Free energy component analysis[35–37] suggests that for the wild-type protein, the complex formed with A6-4CPh SL2 RNA in an extended conformation is 5 kcal/mol more stable than that formed with A6-4CPh SL2 RNA in a folded conformation. In contrast, for the Phe56Ala U1A protein, the complex formed with the folded A6-4CPh SL2 RNA is approximately 1 kcal/mol more stable than the complex formed with the extended A6-4CPh SL2 RNA.

DISCUSSION

The simulations provide a model for the selective stabilization of the complex formed with the Phe56Ala U1A protein by the substitution of A-4CPh for A6 in SL2 RNA. The complex is predicted to be stabilized by a folded conformation of A6-4CPh that maintains

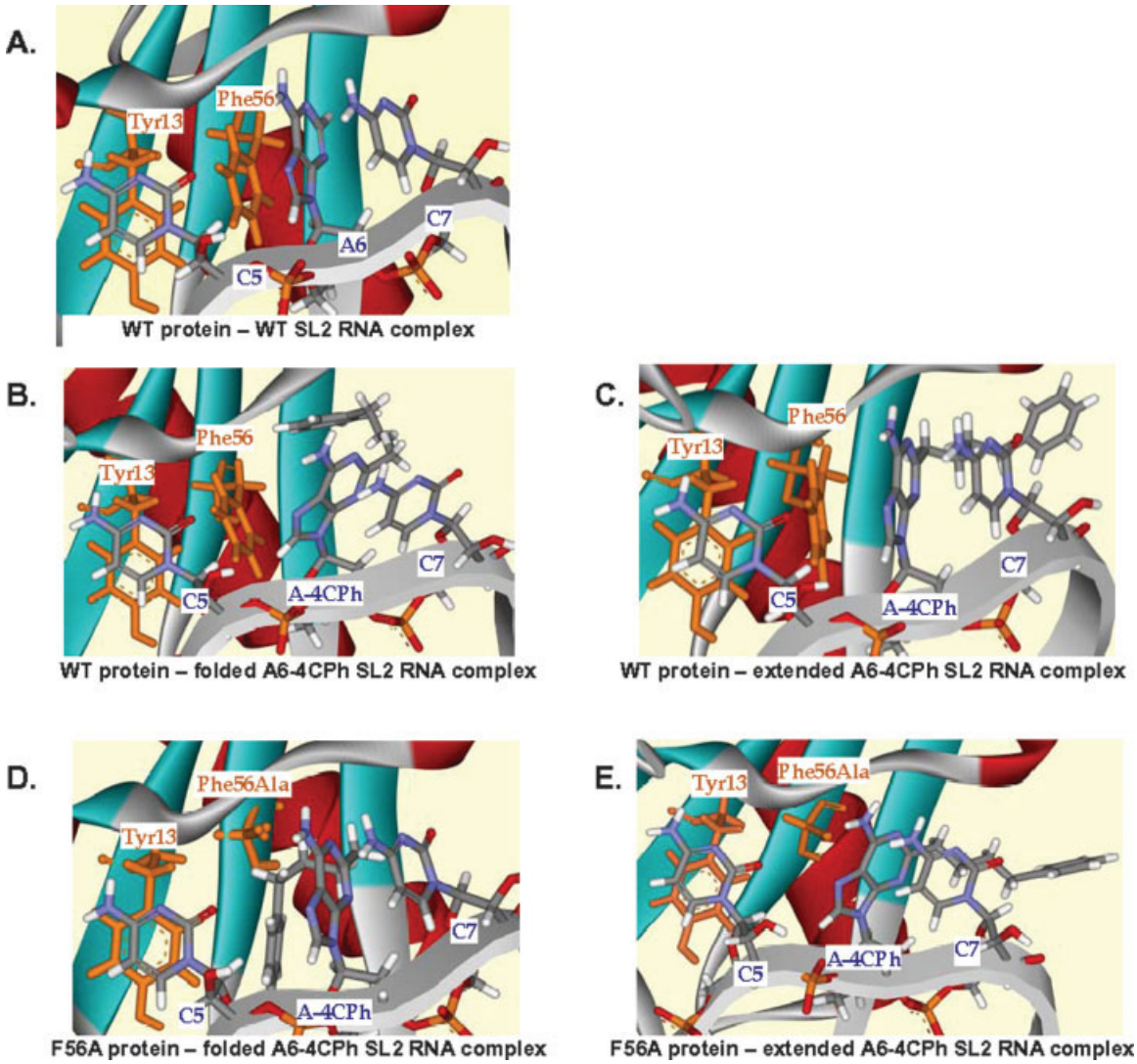

FIGURE 10 Average simulated structures of the U1A–SL2 RNA complexes from the last 2 ns of the simulations: (A) wild-type U1A–wild-type SL2 RNA complex, (B) wild-type U1A–folded A6-4CPh SL2 RNA complex, (C) wild-type U1A–extended A6-4CPh SL2 RNA complex, (D) Phe56Ala U1A–folded A6-4CPh SL2 RNA complex, (E) Phe56Ala U1A–extended A6-4CPh SL2 RNA complex.

key interactions such as the stacking interactions between the purine ring of A-4CPh and C7 and between Tyr13 and C5 and introduces new nonbonding interactions that may compensate for some of the local structural changes that destabilize the Phe56Ala U1A–wild-type SL2 RNA complex. For example, there is an off-centered stacking interaction between the tethered phenyl group and the purine ring of A-4CPh and an edge-to-face interaction between the tethered phenyl group and Tyr13. The analysis of intermolecular interactions suggests that several additional hydrogen bonds are formed in the complex of Phe56Ala U1A with the folded A6-4CPh SL2 RNA compared to the complex with the extended A6-4CPh SL2 RNA. Free energy component analysis supports the prediction of a folded conformation for A6-4CPh in the complex formed with Phe56Ala U1A, suggesting that the folded conformation is 1 kcal/mol more stable than the extended conformation.

The MD simulations predict that the complex formed between A6-4CPh SL2 RNA and wild-type U1A would be distorted if A-4CPh adopted a folded conformation. In the complex formed with the folded A6-4CPh SL2 RNA, the purine ring of A-4CPh, the

tethered phenyl group, and Phe56 are not parallel to each other and may interact using edge–face interactions. Although C7 still stacks with the purine ring of A-4CPh, both have moved from their positions in the wild-type complex. An analysis of intermolecular interactions between the U1A protein and SL2 RNA suggests that the complex formed with the folded A6-4CPh SL2 RNA is less precise than that formed with the wild-type RNA because several hydrogen bonds are no longer formed in this complex, while the number of nonbonded interactions has increased somewhat due to the close packing of the tethered phenyl group in the complex interface. Free energy component analyses predict that the complex formed with the extended A6-4CPh SL2 RNA is approximately 5 kcal/mol more stable than the complex formed with the folded A6-4CPh SL2 RNA. Taken together, these calculations support the initial design of A-4CPh, which placed the tethered phenyl group in the cavity formed by the Phe56Ala mutation in order to stabilize the complex formed with Phe56Ala U1A, but destabilize the wild-type complex.

The calculations indicate that there are two types of induced fit processes that reduce the stability of the complex of Phe56Ala U1A with A6-4CPh SL2 RNA. First, the tethered phenyl group is predicted to be in an extended conformation in the free A6-4CPh SL2 RNA. Thus, an energetic penalty must be paid to form the complexes with the folded A6-4CPh SL2 RNA, which will decrease the stability of the complex formed between Phe56Ala U1A and A6-4CPh SL2 RNA. Thus, the difference in stability between the complexes formed with the wild-type and Phe56Ala U1A proteins could be increased by altering the tethered phenyl group so that the folded conformation is preferred. This would reduce the energetic penalty of RNA conformational changes for the complex formed with the Phe56Ala U1A protein and would destabilize the complex with the wild-type U1A protein because the folded conformation of the tethered phenyl group is predicted to disrupt the well-ordered complex interface.

The second type of induced fit process required to form the complex of Phe56Ala U1A with A6-4CPh SL2 RNA involves structural changes throughout the loop of SL2 RNA. The predicted structures of the SL2 RNA with either the extended or folded A-4CPh are significantly different from that of the wild-type SL2 RNA, even at positions distal from the site of modification. The sugar pucker and AMIGOS analyses suggest that the structure of the wild-type free SL2 RNA is similar to the bound conformation of the RNA, while the A6-4CPh SL2 RNA must undergo energetically unfavorable conformational changes at several positions in the loop upon binding. The complexes formed between A6-4CPh SL2 RNA and both the wild-type and Phe56Ala U1A proteins will be destabilized by these induced fit processes. Although it is unclear how the structure of the modified base should be altered to favor the wild-type loop conformation, performing MD simulations on a number of design alternatives could suggest modifications to minimize required conformational changes upon binding.

The MD simulations provide a possible explanation for the lower affinity of the Phe56Ala U1A protein for A6-3CPh SL2 RNA compared to A6-4CPh SL2 RNA. A-4CPh and A-3CPh differ by one methylene group in the linker region (Figure 2). The complex formed between Phe56Ala U1A and the A6-3CPh SL2 RNA is only stabilized by 0.6 kcal/mol compared to the complex formed with the wild-type RNA, while the complex formed with A4-4CPh SL2 RNA is stabilized by 1.8 kcal/mol. If the folded A6-3CPh SL2 RNA interacts with the Phe56Ala U1A protein in a similar manner as the folded A6-4CPh SL2 RNA, the propyl linker of A-3CPh would be able to interact with Ala56. However, the propyl linker is too short to allow the tethered phenyl ring reach the position of the phenyl ring found in the structure of the complex formed with A6-4CPh SL2 RNA complex. Without the interaction between the tethered phenyl group and Tyr13, the complex of Phe56Ala U1A with A6-3CPh SL2 RNA would be expected to be less stable than the complex formed with A6-4CPh SL2 RNA. Additionally, the contribution of solvation effects to binding would be expected to be smaller to form the complex with the modified RNA containing the shorter methylene linker.

CONCLUSION

The calculations presented in this article predict the tethered phenyl group of A6-4CPh SL2 RNA is placed in a folded conformation in the complex with Phe56Ala U1A and an extended conformation in the complex with the wild-type protein. The simulations of the free RNA suggest that the wild-type SL2 RNA, although flexible, is preorganized for binding to the U1A protein, while A6-4CPh RNA must undergo conformational changes upon binding. Thus, the affinity of A6-4CPh SL2 RNA for Phe56Ala U1A could be increased by altering the modified base to favor the wild-type SL2 RNA structure, while the selectivity of A6-4CPh SL2 RNA for Phe56Ala U1A over the wild-type U1A

could be improved by preorganizing the tethered phenyl group into a folded conformation. Thus, the MD simulations and the resulting computational model of the complex formed between Phe56Ala U1A and A6-4CPh SL2 RNA supports the initial design hypothesis, but have revealed that altered protein–RNA contacts and structural adaptation of the RNA may contribute significantly to the binding. The insight gained from these simulations will be applied to refining molecular designs aimed at altering RRM–RNA affinity and specificity.

Funding was provided by the NIH to AMB, GM-56857, and to DLB, GM-37909. YZ was supported by a NIH Training Grant in Molecular Biophysics (GM-08271). BLK was supported by a NIH Postdoctoral Fellowship (F32 GM072345). This work was partially supported by the National Center for Supercomputing Applications under LRAC/NRAC grant number MCA94P011 utilizing the Xeon Linux Cluster and by the Texas Advanced Computing Center utilizing the Cray-Dell PowerEdge Xeon Linux Cluster.

REFERENCES

1. Maris, C.; Dominguez, C.; Allain, F. H. T. FEBS J 2005, 272, 2118.
2. Zhao, Y.; Baranger, A. M. J Am Chem Soc 2003, 125, 2480–2488.
3. Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Nature 1994, 372, 432–438.
4. Allain, F. H. T.; Howe, P. W. A.; Neuhaus, D.; Varani, G. EMBO J 1997, 16, 5764–5774.
5. Allain, R. H. T.; Bouvet, P.; Dieckmann, T.; Feigon, J. EMBO J 2000, 19, 6870–6881.
6. Varani, L.; Gunderson, S. I.; Mattaj, I. W.; Kay, L. E.; Neuhaus, D.; Varani, G. Nature Struct Biol 2000, 7, 329–335.
7. Johansson, C.; Finger, L.; Trantirek, L.; Mueller, T.; Kim, S.; Laird-Offringa, I.; Feigon, J. J Mol Biol 2004, 337, 799–816.
8. Handa, N.; Nureki, O.; Kuimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. Nature 1999, 398, 579–585.
9. Price, S. R.; Evans, P. R.; Nagai, K. Nature 1998, 394, 645–650.
10. Deo, R. C.; Bonanno, J. B.; Sonenberg, N.; Burley, S. K. Cell 1999, 98, 835–845.
11. Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. J. Genes Dev 1999, 13, 1102–1115.
12. Wang, X.; Tanaka Hall, T. M. Nature Struct Biol 2001, 8, 141–145.
13. Mazza, C.; Segref, A.; Mattaj, I.; Cusack, S. EMBO J 2002, 21, 5548–5557.
14. Green, M. R. Annu Rev Cell Biol 1991, 7, 559–599.
15. Stark, H.; Dube, P.; Lührmann, R.; Kastner, B. Nature 2001, 409, 539–542.
16. Tsai, D. E.; Harper, D. S.; Keene, J. D. Nucleic Acids Res 1991, 19, 4931–4936.
17. Hall, K. B. Biochemistry 1994, 33, 10076–10088.
18. Nolan, S. J.; Shiels, J. C.; Tuite, J. B.; Cecere, K. L.; Baranger, A. M. J Am Chem Soc 1999, 121, 8951–8952.
19. Shiels, J. C.; Tuite, J. B.; Nolan, S. J.; Baranger, A. M. Nucleic Acids Res 2002, 30, 550–558.
20. Cramer, C. J. Essentials of Computational Chemistry; Wiley: New York, 2002.
21. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab Initio Molecular Orbital Theory; Wiley: New York, 1986.
22. Szabo, A.; Ostlund, N. S. Modern Quantum Chemistry; McGraw-Hill: New York, 1982.
23. Hariharan, P. C.; Pople, J. A. Theor Chim Acta 1973, 28, 213–222.
24. Hehre, W. J.; Ditchfield, R.; Pople, J. A. J Chem Phys 1972, 56, 2257–2261.
25. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269–10280.
26. Cieplak, P.; Cornell, W. D.; Bayly, C. I.; Kollman, P. A. J Comput Chem 1995, 16, 1357–1377.
27. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. J Am Chem Soc 1993, 115, 9620–9631.
28. Pitici, F.; Beveridge, D. L.; Baranger, A. M. Biopolymers 2002, 65, 424–435.
29. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. AMBER 6.0 University of California: San Francisco, 1999.
30. Ravishanker, G.; Wang, W.; Beveridge, D. Molecular Dynamics Tool Chest 2.0; Wesleyan University, Middletown, CT, 1998.
31. Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. Nucleic Acids Res 1997, 25, 4940–4945.
32. Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. AMBER 8.0 University of California: San Francisco, 2004.
33. Gohlke, H.; Case, D. A. J Comput Chem 2004, 25, 238–250.
34. Ajay, A.; Murcko, M. A. J Med Chem 1995, 38, 4953–4967.
35. Jayaram, B.; McConnell, K.; Dixit, S. B.; Beveridge, D. L. J Comput Phys 1999, 151, 333–357.
36. Jayaram, B.; McConnell, K.; Dixit, S. B.; Das, A.; Beveridge, D. L. J Comput Chem 2002, 23, 1–14.
37. Kombo, D. C.; Jayaram, B.; McConnell, K. J.; Beveridge, D. L. Molecular Simulation 2002, 28, 187–211.
38. Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Biophys J 2004, 86, 67–74.

39. Onufriev, A.; Bashford, D.; Case, D. A. J Phys Chem B 2000, 104, 3712–3720.
40. Onufriev, A.; Bashford, D.; Case, D. A. Proteins 2004, 55, 383–394.
41. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978–1988.
42. Connolly, M. L. J Appl Cryst 1983, 16, 548–558.
43. Kranz, J. K.; Lu, J. R.; Hall, K. B. Protein Sci 1996, 5, 1567–1583.
44. Tang, Y.; Nilsson, L. Biophys J 1999, 77, 1284–1305.
45. Reyes, C. M.; Kollman, P. A. J Mol Biol 2000, 297, 1145–1158.
46. Williams, D. J.; Hall, K. B. J Mol Biol 1996, 257, 265–275.
47. Harvey, S. C.; Prabhakaran, M. J Am Chem Soc 1986, 108, 6128–6136.
48. Altona, C.; Sundaral. M. J Am Chem Soc 1972, 94, 8205–8212.
49. Duarte, C. M.; Pyle, A. M. J Mol Biol 1998, 284, 1465–1478.

Reviewing Editor: J. A. McCammon