Molecular dynamics studies of U1A-RNA complexes

CAROLINA M. REYES and PETER A. KOLLMAN

Department of Pharmaceutical Chemistry, University of California–San Francisco,
San Francisco, California 94143-0446, USA

ABSTRACT

The U1A protein binds to a hairpin RNA and an internal-loop RNA with picomolar affinities. To probe the molecular basis of U1A binding, we performed state-of-the-art nanosecond molecular dynamics simulations on both complexes. The good agreement with experimental structures supports the protocols used in the simulations. We compare the dynamics, hydrogen-bonding occupancies, and interfacial flexibility of both complexes and also describe a rigid-body motion in the U1A–internal loop complex that is not observed in the U1A–hairpin simulation. We relate these observations to experimental mutational studies and highlight their significance in U1A binding affinity and specificity.

Keywords: molecular dynamics simulations; nucleic acids; PIE-RNA; RNA–protein interactions

INTRODUCTION

RNA–protein interactions govern many cellular processes: transcription, splicing of the primary transcript to mRNA, transport of mRNA outside the nucleus, and its translation to proteins. RNA-binding proteins containing the RNA-binding domain (RBD) participate in many of these interactions. The RBD, which is also commonly known as the RNA-recognition motif (RRM) or RNP consensus sequence (RNP-CS), is a sequence of about 90 amino acids (Nagai et al., 1995). This sequence folds into a single domain composed of two α helices packed against a four-stranded β -sheet. Residues in the hydrophobic core and on the two central β strands are conserved across more than 100 RBDs (Birney et al., 1993).

One of the best-characterized RBDs, both biochemically and structurally, is the N-terminal RBD of the U1A protein (Scherly et al., 1991). The U1A protein is a component of the U1 small nuclear ribonucleoprotein (snRNP), which recognizes the 5' exon of the pre-mRNA and initiates spliceosomal assembly. In the U1 snRNP, U1A protein binds to stem-loop II of U1 RNA. U1A also binds to its own 3'UTR and regulates its own expression by inhibiting polyadenylation, forming what is known as the polyadenylation inhibition element RNA (PIE-RNA) complex (Boelens et al., 1993). Interest-

ingly, U1A binds to both RNAs with picomolar affinities (Hall, 1994).

The crystal structure of U1A bound to stem-loop II of U1A RNA has been solved (Oubridge et al., 1994) and the NMR structure of U1A bound to its own 3' UTR internal loop RNA (PIE-RNA) has been determined (Allain et al., 1996). The protein β -sheet surface contacts the AUUGCAC loop sequence of the hairpin and the internal loop through hydrophobic stacking, hydrogen-bonding (H-bonding), and electrostatic interactions. In both structures, the U1A β 2- β 3 turn protrudes through the AUUGCAC loop of the RNA. Figure 1 shows the secondary structures of both RNAs and their experimental structures bound to U1A.

In the U1A–hairpin complex, the RNA is composed of a 10-nt loop and a 5-bp stem region. Most of the protein contacts occur in the loop region where the nucleotides are turned out to stack with the hydrophobic residues on the protein β -sheet, except for last three loop nucleotides, which are solvent exposed. The internal loop RNA is a 30-mer composed of the AUUGCAC single-stranded loop flanked by the upper and lower stem regions. Similar RNA–protein interactions are found in the U1A–internal loop structure as in the hairpin complex, with additional electrostatic interactions on the phosphate backbone in the stem regions of the internal loop complex (Allain et al., 1997).

In both structures, the extensive contacts of the unpaired bases with the β -sheet surface seem to imply that recognition of the AUUGCAC sequence determines the binding specificity of U1A (Moras & Potersz-

Reprint requests to: Peter A. Kollman, Department of Pharmaceutical Chemistry, Box 0446, University of California–San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0446, USA; e-mail: pak@cgl.ucsf.edu.

FIGURE 1. Sequence and secondary structures for hairpin and internal loop RNA (top); starting structures for the U1A-hairpin and U1A-internal loop simulations (bottom).

man, 1995). However, the β -sheet contains the highly conserved RNP domain in the two central strands (Burd & Dreyfuss, 1994). Moreover, another snRNP protein, the U2B' protein, uses essentially the same residues for the recognition of single-stranded nucleotides, but does not bind the U1A hairpin RNA (Scherly et al., 1991).

What are the molecular determinants that enable U1A to bind with such high affinity ($K_d = 10^{-11} \text{ M}^{-1}$) for its RNA? How well ordered is the protein-RNA interface? Are there differences in the dynamics of the U1A-hairpin and U1A-internal loop complexes that suggest the roles of conformational flexibility in recognition?

To begin to address these questions, we have performed nanosecond molecular dynamics (MD) simulations on both the U1A-hairpin crystal structure and the U1A-internal loop NMR structure in water. MD allows for the observation of atomic motion that may be hidden by the time-averaged nature of a NMR experiment or damped by crystal packing forces. Recent methodological advances in force-field development and the improved treatment of long-range interactions with Particle-Mesh Ewald (PME) have enabled stable, fully solvated simulations of proteins and nucleic acids

(Darden et al., 1993; Cheatham et al., 1995), including nonhelical RNAs (Auffinger & Westhof, 1998). The PME method accurately accounts for long-range electrostatic interactions beyond the cutoff in a periodic system with marginal computational overhead by using a fast-Fourier transform. Our work here shows that these advances have made it possible to study RNA-protein complexes.

To our knowledge, this is the first unrestrained MD simulation of an RNA-protein complex in explicit water. Although a number of simulations of DNA-protein complexes have been done (Eriksson & Nilsson, 1995; Miaskiewicz & Ornstein, 1996; Roxström et al., 1998; Tang & Nilsson, 1998), relatively few of these have employed the full periodic boundary conditions used here. Our simulations of both the U1A-hairpin and U1A-internal loop complexes agree well with experimental structures, showing that we have successfully simulated and established a protocol for an RNA-protein system. We have completed a nanosecond simulation of each complex. Analyses of the trajectories, including differences in the structure and dynamics of the two complexes, are presented below.

RESULTS

U1A-hairpin complex

For the U1A-hairpin complex, Figure 2 presents the average root-mean-square deviation (RMSD) of the trajectory from the starting crystal structure fitted on all the atoms, the protein backbone, and the RNA hairpin. The average RMSD of the protein backbone is 1.30 Å while the average RNA RMSD is 1.31 Å. Both values are within the range observed for MD simulations of individual proteins (Fox & Kollman, 1996) and RNA (Miller & Kollman, 1997). The average all-atom RMSD is 1.64 Å, which is higher as it includes all atoms, including those of the flexible U118 at the 3' end of the RNA. The RNA loop (residues 103–112) region has an average RMSD of 1.16 Å, and the stem region (residues 98–102, 113–117) has a lower average RMSD of 0.88 Å. Hence, we are able to obtain a stable simulation of the U1A-hairpin complex comparable to previous MD simulations of proteins and nucleic acids.

As an additional comparison with experiment, the atomic fluctuations averaged over the trajectory were calculated and related to crystallographic B-factors. Figure 3 shows the average per-residue experimental and MD-derived B-factors. The outlying point in the figure corresponds to U118, at the 3' end of the RNA stem. During the simulation, U118 is highly mobile because it is unpaired and only interacts with solvent.

We observe lower B-factors for nucleotides U110, C111, and C112 compared to the experimental B-factors. These 3 nt, in the 3' end of the hairpin loop, are not interacting with the protein. During the simulation, they

FIGURE 2. U1A-hairpin simulation RMSDs calculated with fitting to all atoms, to the protein backbone, to the RNA heavy atoms, and to the loop and stem regions of the RNA relative to the initial structure.

form stable stacking interactions with each other, slowly exchanging stacking of C111 with U110 and with C112. This may explain their lower fluctuations compared to experiment. However, they do have higher fluctuations compared to the loop nucleotides interacting with the protein. In general, the calculated B-factors agree reasonably well with the experimental B-factors, further validating our protocol for a stable simulation of an RNA-protein complex.

U1A-internal loop complex

Figure 4A shows the RMSD for the simulation of the U1A-internal loop complex relative to its initial NMR structure. The average all-atom RMSD is 4.52 Å. The average RMSD for the protein backbone is 2.40 Å, and the average RMSD for the RNA is 5.04 Å. Based on the similarly large RMS deviations of the RNA and all atoms, we suspect that the large deviation of the complex is mainly due to the RNA.

FIGURE 3. Averaged B-factors per residue in the U1A-hairpin simulation (dashed line) with the experimental values (solid line).

To focus on the RNA, we calculated the RMSDs of subregions of the internal loop. The subregions were defined as the loop region (residues 121–127), the upper stem (residues 109–120), and the lower stem (res-

FIGURE 4. A: U1A-internal loop RMSD relative to the starting structure. B: RMSDs of the internal loop subregions defined as the single-stranded RNA loop (AUUGCAC), the upper stem, and the lower stem.

TABLE 1. Average (AVE) root-mean-square deviations (RMSDs) of the MD trajectory relative to the initial structure, relative to other NMR structures, and pairwise between the 25 NMR structures. Maximum (MAX) and minimum (MIN) values are also calculated for all the atoms, protein backbone (BB) of residues 7–98, the RNA atoms, the RNA upper stem residues 109–120, the RNA lower stem residues 103–107 and 128–132, and the RNA loop region (residues 121–127).

RMSD	MD to starting NMR structure (Å)			MD to other NMR structures ^a (Å)			Within NM structures ^b (Å)		
	AVE	MAX	MIN	AVE	MAX	MIN	AVE	MAX	MIN
All-atom	4.52	6.63	0.54	4.88	8.32	2.72	4.49	8.35	2.25
Protein BB	2.41	3.09	0.59	2.87	3.77	2.11	2.11	3.26	1.44
RNA	5.04	7.96	0.42	4.77	9.35	2.10	4.02	7.15	1.64
RNA upper	1.86	3.04	0.29	2.10	3.69	0.96	1.53	2.70	0.72
RNA lower	1.15	1.57	0.32	1.63	2.50	0.96	1.18	1.84	0.62
RNA loop	1.79	2.72	0.42	1.89	2.87	1.18	1.58	2.03	0.85

^aRMSD of the trajectory to the other 24 NMR structures.

^bRMSD of all 25 NMR structures to each other.

idues 103–107, 128–132), according to Allain et al. (1996). In Figure 4B, the average RMSDs of the fitted loop region, fitted upper stem, and fitted lower stem are 1.79, 1.86, and 1.15 Å, respectively. The lower RMS deviations for the subregions implied that the local structures are well maintained over the course of the simulation even though the global RNA RMSD is 5 Å.

There are 25 structures in the reported NMR results (Allain et al., 1996); the RMSD of the MD trajectory relative to each member of the family was calculated to determine whether the simulation sampled the conformational distribution represented by the experimental structures. Table 1 lists the average RMSDs of the MD trajectory relative to the initial structure, to the other 24 NMR structures, and the RMSD of the NMR structures to each other. The trend is similar in all three calculations. The average all-atom RMSDs are comparable to the RNA RMSDs (4–5 Å), while the subregions (upper stem, lower stem, loop) maintain low RMSDs (1–2 Å). This shows that the deviations within the simulation concur with the range of RMSDs in the NMR family.

Based on the lower RMS deviations of the subregions, we surmised that the large global RMS for the RNA may be because of interdomain motion. To investigate this further, the interhelical angle within the RNA molecule, defined as the angle between the upper stem helix axis and the lower stem helix axis, was calculated over the course of the simulation, as shown in Figure 5. The angle equilibrates to about 135° and then decreases by roughly 35° between 400 and 800 ps. The interhelical angles, calculated over the course of simulation, fall between 91° and 141°, roughly within the 100°–170° range found in the NMR family of structures. Although the interhelical angles in the experimental structures are not well defined because of the lack of Nuclear Overhauser Effect (NOE) distance constraints between the protons of the upper stem and the lower stem, we maintain that the simulated range of interhelical angles overlaps the angle distribution in the NMR structures. Furthermore, we observe a rigid-body mo-

tion represented by the interhelical angle change as the stem regions move closer together and away from the initial structure. Even though local structures are well conserved, this global motion gives the high overall RMSD seen for the RNA.

For a more detailed analysis of the RNA dynamics and to try to localize the observed global motion, the RNA backbone torsional angles and sugar puckers are calculated over the course of the trajectory and compared with experimental constraints. Plotting the torsion angles over the trajectory show torsional transitions indicative of crankshaft motions in the loop region and in the tetraloop of the upper stem.

In the loop region, crankshaft motions are observed within the first 500 ps for the ϵ and ζ torsions of U123 and the α and β torsions of G124 and C125. At 310 ps, four connected torsional transitions occur: U123 ϵ and ζ switched from 280° to 210° and from 90° to 170° and G124 α and β torsions changed from –150° to –70°

FIGURE 5. Interhelical angle of the internal loop RNA traced over the simulation; the dashed lines represent the maximum and minimum in the range of interhelical angles found in the family of NMR structures in 1NUM.

and 170° to 110°, respectively. For C125, there were two transitions at 47 ps, when α changed from 80° to 150° and β changed from 150° to 80°, and then at 243 ps, when both torsions switched back.

In the upper stem's tetraloop region, more frequent transitions occurred for ϵ and ζ torsions of G116 and the α , β , and γ torsions of G117. Similar tetraloop crankshaft transitions have also been observed in the MD simulation of UUCG tetraloop (Miller & Kollman, 1997). We did not detect any transitions in the stem or base-pairing regions of the RNA and only observed them in the loop and tetraloop regions where there are very few NMR torsional constraints. Upon comparison with the available experimental torsional constraints, we find that of the total 106 NMR torsional constraints, only one is violated. This corresponds to the G117 γ torsion that is involved in a crankshaft motion with G117 β .

The sugar puckers in the helical regions remain C3'-endo during the simulation while U114 and C115 in the tetraloop and U123 and C125 in the loop region remain C2'-endo. All the calculated puckers agree with experimental observations. The only sugar repuckering that occurred during the simulation is in G124 for which an unambiguous sugar conformation could not be assigned experimentally (Allain et al., 1996). For G124, we observe a pucker flip from C3'-endo to O4'-endo at 670 ps and then back again to C3'-endo at 810 ps. Based on the lower deviations of the individual sub-regions and the agreement with the torsional and sugar pucker constraints, we surmise that the large RMS deviation for the RNA is because of interdomain motion.

A more stringent test and validation of the unrestrained simulation is to determine how well it fits the experimentally observed NOE distances. Table 2 tallies the NOE violations within the protein, the RNA, and the RNA-protein interface. The interproton distances are calculated by $\langle r_{ij}^{-6} \rangle^{-1/6}$ averaging over the trajectory. Violations greater than 0.2 Å are reported as the experimental upper bounds minus the MD-derived distances. The average violations of the calculated interproton distances are also listed.

From the total of 1,720 protein-protein NOE distances, 153 were violated. The violations are not localized to a particular region, but distributed throughout the protein. The majority of them (129 out of 153 violations) are associated with equivalent or nonstereospecifically assigned -CH₃ or -CH₂ hydrogens, which are averaged to their center of geometry. There are no such equivalent hydrogens in the RNA-RNA NOE distances and lower violations are obtained: from the total of 588 RNA-RNA NOE distances, 17 were violated. These violations are localized in the unpaired regions of the loop, the tetraloop and A107. At the RNA-protein interface, there are 82 NOE distance constraints. Thirteen were violated, nine of which involve equivalent protein hydrogens. The average distance violations are 0.11 Å for the protein, 0.02 Å for the RNA, and 0.18 Å

TABLE 2. NOE violations greater than 0.2 Å with total NOE distances.

	Violations >0.2 Å	Total NOE distances ^a	Average ^b violations
Protein total	153	1,720	0.11
Strong (0.0–2.3 Å)	7	34	0.27
Medium (0.0–2.9 Å)	7	82	0.08
Weak (0.0–3.5 Å)	25	168	0.14
3D Strong (0.0–5.0 Å)	87	957	0.11
3D Weak (0.0–7.0 Å)	27	479	0.08
RNA total	17	588	0.02
Strong (0.0–2.3 Å)	0	0	—
Medium (0.0–3.0 Å)	0	55	0.00
Weak (0.0–4.0 Å)	3	134	0.01
3D Strong (0.0–5.0 Å)	6	247	0.01
3D Weak (0.0–7.0 Å)	8	154	0.04
RNA-protein interface	13	82	0.18
Medium (0.0–2.9 Å)	1	1	1.02
Weak (0.0–3.5 Å)	4	14	0.29
3D Strong (0.0–5.0 Å)	5	39	0.19
3D Weak (0.0–7.0 Å)	3	28	0.08

^a Experimental NOE distances were provided by Varani and Allain.

^b Sum of violations divided by total number of NOE distance constraints.

for the RNA-protein interface. Overall, out of the total 2,390 NOE distances, 183 violations were observed, mostly from equivalent protons in the protein and single-stranded regions of the RNA.

Comparison of U1A-hairpin and U1A-internal loop dynamics

The above analyses addressed how well the simulations reproduced their respective experimental structures. We now compare the simulations of the two complexes, focusing on average fluctuations and H-bonding occupancies. The average per-residue fluctuations, $\langle \Delta r^2 \rangle$, for the U1A protein in both the hairpin complex (residues 2–97) and the internal loop complex (residues 2–102) are shown in Figure 6A. In general, both structures follow the same trend: the average fluctuations of the β -sheet, which contacts the single-stranded loop region of the RNA, are less than those of the helices. Figure 6B shows the RNA fluctuations in both structures where the loop bases involved in the protein contacts have lower fluctuations compared to the stem regions. Also, the increased fluctuations in the internal loop RNA upper stem (residues 109–120) and lower stem (residues 103–107, 128–132) support the observed stem motion in the MD simulation. Hence, the decreased fluctuations in the protein β -sheet and the RNA loop in both simulations suggest that the binding interfaces form relatively rigid core.

Since the U1A protein presents the same recognition surface to both RNAs, we calculated the protein-RNA H-bonding occupancies and compared the aromatic

FIGURE 6. **A:** Average per-residue fluctuations of the protein in the U1A–hairpin (solid line) and U1A–internal loop (dashed line) simulation. **B:** Average per-base fluctuations in the RNA of the U1A–hairpin and U1A–internal loop simulation.

stacking interactions of the two simulated complexes. Table 3 lists the H-bonding pairs that are greater than 70% occupied during the simulations. There are more intermolecular high-occupancy H-bonding pairs in the hairpin complex simulation than in the internal loop complex simulation, particularly in the $\beta 2$ – $\beta 3$ protein loop (residue 47–52), which protrudes through the single-stranded AUUGCAC sequence.

The AUUGCAC loop corresponds to residues 103–109 in the hairpin complex and to 121–127 in the internal loop complex. H-bonding between the protein and the single-stranded bases of the RNA, retained in both complexes, are: Asn16 with U105_{hp} (in the hairpin complex) and with U123_{il} (in the internal loop complex); Glu19 with U104_{hp} and U122_{il}; Tyr 86 and Lys88 with C107_{hp} and C125_{il}; Thr89 with A108_{hp} and A126_{il}; Ser91 with A108_{hp} and A126_{il}; Asp92 with C109_{hp} and C127_{il}. These U1A residues are at the tail end of $\beta 1$ and the beginning of helix C. $\beta 1$ contains the conserved RNP region, whereas Helix C has been observed to contribute to RNA binding affinity (Zeng & Hall, 1997).

Similar aromatic stacking interactions are found and maintained in both simulations. In the hairpin complex,

TABLE 3. Hydrogen-bonding pairs for U1A–hairpin and U1A–internal loop complexes that are greater than 70% occupied over the course of the trajectory.

U1A–hairpin H-bonding pair	% occupied	U1A–internal loop	% occupied
Asn 15 ND2HN22–G 106 O6	79.2	Tyr 13 OHH–C 125 O2P	94.9
Asn 16 NH–G 106 O6	100.0		
Asn 16 OD1–U 105 N3H3	100.0	Asn 16 OD1–U 123 N3H3	92.4
Glu 19 OE1–U 104 N3H3	99.6	Glu 19 OE1–U 122 N3H3	89.8
Glu 19 OE2–U104 N3H3	89.8	Glu 19 OE2–122 N3H3	80.4
Glu 19 OE2–G 106 N1H1	86.9		
Glu 19 OE2–G 106 N2H21	99.8		
Arg 47 NH1 HH12–A 99 O1P	99.8		
Arg 47 NH2HH12–A 99 O1P	95.3		
Ser 48 HGOG–C 112 O1P	72.2		
Leu 49 NH–G 113 O1P	90.8		
Leu 49 NH–G 113 O2P	92.2		
Lys 50 O–G 106 O2' HO'2	92.6		
Arg 52 NH1HH1–A 103 N1	96.9		
Arg 52 NH1HH12–G 113 N7	79.8		
Arg 52 NH1HH12–G 113 O6	95.3		
Arg 52 NH2HH22–G 113 N7	76.5		
Tyr 86 O–C 107 N4H41	100.0	Tyr 86 O–C 125 N4H41	99.3
Lys 88 NH–C 107 N3	97.3	Lys 88 NH–C 125 N3	99.6
Lys 88 NH–C 107 O2	95.7	Lys 88 NH–C 125 O2	100.0
Thr 89 O–A 108 N6H62	94.9	Thr 89 O–A 126 N6H62	81.4
Asp 90 O–C 109 N4H41	100.0	Asp 90 O–C 127 N4H41	89.0
Ser 91 OG–A 108 N6H61	99.8		
Ser 91 OGHG–A 108 N1	100.0	Ser 91 OGHG–A 126 N1	85.2
Asp 92 NH–C 109 N3	100.0	Asp 92 NH–C 127 N3	95.5
Asp 92 NH–C 109 O2	98.7	Asp 92 NH–C127 O2	99.4

A103 stacks with U104, U105 with G106, and Tyr13 with C107. A108 is sandwiched between Phe56 and C109. In the crystal structure (Oubridge et al., 1994), C109 makes hydrophobic contacts with Asp92; during the simulation; however Asp92 moves away from C109 and forms electrostatic interactions with Lys96 instead. In the internal loop complex, partial stacking is observed for A121 with U122 and U123 with G124. Similar to U1A-hairpin, Tyr13 stacks with C125 and A126 lies between Phe56 and C127.

DISCUSSION

Interdomain motion in the internal loop complex simulation

We obtain good agreement with the experimental structures, which validates the protocols used in our simulations. Although a large RMS deviation is observed for the internal loop RNA in the simulation, we believe that a rigid domain motion accounts for this based on the decrease in the RNA interhelical angle. In the U1A-internal loop system, the upper stem and the lower stem move closer together as rigid units and produce a higher global RMSD for the RNA while maintaining low local RMSDs as individual domains. This is further supported by greater fluctuations in the upper and lower stem regions compared to the loop region. Interestingly, however, there are neither NOE distance violations nor torsional constraint violations in the helical regions of the upper and lower stems, even though they show the greatest fluctuations. Hence, the increased helical fluctuations due to the hinge motion detected in the dynamics are still compatible with the experimental constraints.

The RNA interhelical angles are not directly determined by experiment because proton distances between the upper and lower stems are beyond the short-ranged accuracy of NOEs and are therefore indirectly defined by NMR torsion and distance constraints. In the NMR structures, the angles range from 100° to 170°, whereas the simulated angle changes from 135° to 100°. The simulated interhelical angle is based on intramolecular, nonbonded, and electrostatic forces as well as the inclusion of explicit water interactions.

We also consider that the crankshaft transitions in loop nucleotides U123 and G124, which occur at 310 ps, may be related to the decrease in interhelical angle. We observe a similar interhelical angle change in the MD simulation of the unbound internal loop (Reyes & Kollman, in prep.) at an even higher frequency, which further supports the hypothesis that this angle change or hinge motion is indeed a dynamical property. It is not clear, however, whether the difference in the method of structure determination (crystallography versus NMR)

is a factor or whether the difference in RNA structure (hairpin versus internal loop) contributes to the difference in dynamics.

This is a good example of how MD studies with full periodic boundary conditions, explicit water, and long-range electrostatics can complement NMR structure determination and refinement. MD studies can resolve dynamical information that is difficult to obtain experimentally. More NMR dynamics studies and order parameter calculations are currently underway (G. Varani, pers. comm.), which may allow evaluation of the predicted interdomain motion.

RNA-protein interface

We observe lower fluctuations in the β -sheet residues of the protein and in the single-stranded region of the RNA which together form the RNA-protein binding interface. This implies that the RNA-protein binding surface constitutes a relatively rigid core within the bound structure. Two factors may explain this difference in mobility relative to noninterfacial regions. First, the protein-RNA interface is tightly packed, with many strong van der Waals contacts. Second, there is an extensive H-bonding network on the interface that contributes to the decrease in mobility.

This observation of a rigid interface suggests two possible pictures of the binding thermodynamics: either the separate protein and RNA are both highly preorganized for binding, resulting in a lower entropy penalty for binding, or both species must undergo considerable ordering and possibly reorganization during the binding event. Indeed, NMR structures of the unbound U1A protein (Avis et al., 1996) and internal loop RNA (Gubser & Varani, 1996) confirm that a large conformational change must occur upon binding. In the unbound U1A, Helix C resides on the β -sheet surface, blocking part of the binding surface, while the single-stranded AUUGCAC bases are turned towards the loop interior in the free RNA; the binding surfaces are buried in both monomers. Moreover, thermodynamic studies by Hall & Kranz (1995) indicate that association of the U1A protein with the hairpin occurs with an unfavorable entropy of -17 eu. The burial of hydrophobic surfaces usually results in a positive entropy change because of the release of water molecules. NMR dynamics studies of the free protein (Lu & Hall, 1997) suggest that the $\beta 2$ - $\beta 3$ loop exhibits conformational flexibility. In this case, the negative entropy is consistent with restraining the conformation of the binding interface despite the positive contribution from water release upon U1A-RNA binding. The decreased flexibility of the binding interface in the simulations of the bound complexes concurs with this thermodynamic data. More comprehensive analyses of the dynamics and flexibility of the unbound RNA are under way.

Recognition elements

A comparison of the intermolecular contacts of the complexes can provide insights on the key protein residues and RNA bases that contribute to the binding affinity of U1A for both RNAs. For the U1A–hairpin complex, extensive mutagenesis data are available, which we compare to the high-occupancy H-bonds observed in the hairpin complex simulation (Table 3). A review of mutational studies reveals that an Asn15-to-Val15 substitution abolishes RNA binding, whereas Asn16-to-Val16 reduces RNA binding (Jessen et al., 1991). The high H-bonding occupancies observed for Asn15 and Asn16 in the simulations are consistent with these experimental observations. In the hairpin, G106, located in the single-stranded region, makes H-bonds with Asn15, Asn16, Glu19, and Lys50. Mutating this guanine to adenine results in a reduction of 4.9 kcal in binding free energy (Hall & Kranz, 1995). This may be because of the loss of H-bonds between G106 and residues Asn15, Asn16, and Glu19, and also to nonadditive energetic contributions that are propagated through the binding site.

Residues 48–52 in the $\beta 2$ – $\beta 3$ loop, which protrudes through the RNA hairpin, show multiple H-bonding contacts with the RNA. Leu49 has been proposed (Laird-Offinga & Belasco, 1995) to lock the protein onto the RNA like “a button in a button hole.” Hydrogen bonds between the backbone NH of Leu49 and the phosphates of G113 are observed over the course of the simulation. Arg52 also makes multiple hydrogen bonds with the RNA, and mutation of Arg52 to Gln completely abolishes binding (Nagai et al., 1990). Varani and coworkers (Allain et al., 1997) have proposed that the residues in the $\beta 2$ – $\beta 3$ loop are additional specificity determinants for U1A. We find that in the hairpin complex, the $\beta 2$ – $\beta 3$ loop residues are involved in extensive contacts with the phosphate backbone in the stem region. However, we do not find high-occupancy (>70% occupied) intermolecular H-bonding contacts in the same $\beta 2$ – $\beta 3$ region of the U1A–internal loop complex as highlighted by Allain et al. (1997), but we do find lower occupancy contacts in that region.

High H-bonding occupancies are also detected in residues 86–92 in the hairpin complex. Reduced binding affinity for a Ser91-to-Ala mutation and for protein truncation at this position have been reported (Zeng & Hall, 1997). In general, the high-occupancy H-bonding obtained from the MD simulation of the U1A–hairpin complex agrees with mutational studies on U1A–hairpin binding.

Mutational studies on the internal loop complex are limited. The high-occupancy contacts found in the MD simulation can suggest possible residues for mutational analysis that may significantly affect the binding of U1A to the internal loop complex. Even though only

a nanosecond of simulation was performed, breaking and formation of H-bonding, hydrophobic contacts, and electrostatic interactions are observed during the simulation, which add dynamical insights into U1A recognition.

CONCLUSIONS

We have presented the first unrestrained molecular dynamics studies of RNA–protein complexes with full periodic boundary conditions. Both simulations compared well with experimental structural data, and a rigid-body motion was detected for the U1A–internal loop complex. Extensive H-bonding contacts were maintained with the single-stranded AUUGCAC loop common to both RNAs, indicative of its contribution to a high binding affinity. Moreover, a less flexible RNA–protein interface (relative to noninterfacial regions) was observed in both simulations, also in agreement with previous thermodynamic data.

The recognition elements highlighted here form intermolecular contacts observed in the simulations of the bound complexes. Binding, however, is a balance between the free energies of the complex and the free monomers. MD simulations of the bound complexes show different dynamics. The hairpin complex is less flexible and forms more RNA–protein hydrogen bonds, whereas the internal loop complex shows greater flexibility and interdomain motion. Given that both complexes have similar binding free energies, our data prompt the speculation that they do not necessarily have the same energetics or driving forces. The observed flexibility in the bound internal loop complex suggests that binding of U1A with the internal loop may have low internal entropy penalty given that the bound complex is also flexible. Molecular dynamics simulations of the unbound RNAs are in progress to compare the differences in dynamics and intermolecular contacts with the complex simulations.

Despite the crucial role in so many cellular processes, few RNA–protein interactions are understood in precise detail. We have shown that molecular dynamics simulation is now a viable tool for studying these interactions. In particular, MD simulation provides microscopic dynamic information that complements structure determination and mutagenesis to provide a clearer picture of the behavior and thermodynamics of RNA–protein complexes.

MATERIALS AND METHODS

All simulations were carried out using the SANDER module in AMBER 5.0 (Case et al., 1997) and the Cornell et al. (1995) force field.

The 1.92-Å-resolution coordinates of the hairpin complex (Oubridge et al., 1994), ID code 1URN, were obtained from

the Brookhaven Protein Data Bank (PDB; Abola et al., 1997) and used as the starting structure for the U1A-hairpin simulation. Ninety-three crystallographic water molecules were included and 13 Na⁺ ions were added to neutralize the system. Using the program EDIT, a rectangular box of 8,609 TIP3P water molecules was added surrounding the complex. The final system consisted of 28,369 atoms within box dimensions of approximately $84 \times 63 \times 60 \text{ \AA}^3$.

For the internal loop complex, one of the 25 NMR structures (Allain et al., 1996) in the PDB (ID code 1NUM) was arbitrarily chosen as the starting structure; 21 Na⁺ ions were added to neutralize the system. We added 13,399 TIP3P water molecules, for a box size of $99 \times 76 \times 62 \text{ \AA}^3$. The solvated internal loop complex totaled 42,845 atoms.

Both systems were minimized and equilibrated with the same protocol. Initially, the complex was minimized for 1,000 steps, followed by 40 ps of dynamics of the water molecules to let the water relax around the solute. The whole system was then minimized for 3,000 more steps. Using separate solvent-solute Berendsen temperature coupling (Berendsen et al., 1984), the system was heated up in stages of 10 ps at 50 K, 20 ps at 150 K, then 30 ps at 300 K prior to production runs. This gradual heating serves to equilibrate the kinetic energy of the system, evenly distributing it among the many degrees of freedom. Following the equilibration phase, we continued each simulation at 300 K for 1,000 ps.

SHAKE was applied to all bonds involving hydrogen atoms. The timestep was 2 fs and a residue-based, 10-Å, nonbonded cutoff was used. The PME approach was applied for a more accurate treatment of electrostatics. The simulations of both complexes were performed on 16 processors on a Cray T3E at the San Diego Supercomputer Center. For the U1A-hairpin complex, a nanosecond simulation took a total of approximately 4,200 CPU hours, while a nanosecond simulation for the U1A-internal loop complex took approximately 5,200 CPU hours.

CARNAL was used for the trajectory analyses of RMSDs, protein-RNA H-bonding occupancies, torsion angles, and atomic fluctuations. The atomic fluctuations, $\langle \Delta r^2 \rangle$, are related to the crystallographic B-factors by $B = 8\pi^2 \langle \Delta r^2 \rangle / 3$. The B-factors for the U1A-hairpin simulation were calculated and averaged over all non-hydrogen atoms in a given residue. We also compared the per-residue fluctuations of both simulations by averaging the fluctuations of all atoms in a given residue. The RNA interhelical angle was obtained by defining the upper stem and lower stem helix axis vectors, based on the center of geometry of the base-pairing residues in the upper stem (residues 109–112, 117–120) and the lower stem (residues 104–107, 128–131), and then computing the angle between the two vectors. Interproton distances were also calculated over the course of the trajectory, averaged by $\langle r_{ij}^{-6} \rangle^{-1/6}$. Nonstereospecific or equivalent hydrogens were averaged to their center of geometry. The calculated interproton distances were compared to the experimental NOE distances. The average violations are calculated as the sum of violations divided by the total number of constraints. Varani and Allain provided us with an earlier set of experimental torsional and distance constraints used in determining the structure reported in Allain et al. (1996). UCSF MidasPlus (Ferrin et al., 1988) was used to display the molecules and the NOE distance violations.

NOTE ADDED IN PROOF

We extended the simulation of the U1A-internal loop complex to 2.2 ns and observed an interhelical angle transition from 100° to 135° at around 1.5 ns. This further supports our observation of interdomain motion in the nanosecond time-scale for the U1A-internal loop complex.

ACKNOWLEDGMENTS

We thank Dr. Gabriele Varani for his helpful discussions and Dr. Frederic Allain for providing the NMR restraints. P.A.K. acknowledges the NIH (CA-25644 and GM-29072) for support. C.M.R. is grateful to the Ford Foundation for financial support and also thanks Drs. E. Meng, J. Miller, J. Pitera, and U. Schmitz for their suggestions.

Received September 9, 1998; returned for revision October 2, 1998; revised manuscript received October 26, 1998

REFERENCES

- Abola EE, Sussman JL, Prilusky J, Manning NO. 1997. Protein data bank archives of three-dimensional macromolecular structures. *Methods Enzymol* 277:556–571.
- Allain FHT, Gubser CG, Howe PWA, Nagai K, Neuhaus D, Varani G. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* 380:646–650.
- Allain FHT, Howe PWA, Neuhaus D, Varani G. 1997. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J* 16:5764–5774.
- Auffinger P, Westhof E. 1998. Simulations of the molecular dynamics of nucleic acids. *Curr Opin Struct Biol* 8:227–236.
- Avis JM, Allain FHT, Howe PWA, Varani G, Nagai K, Neuhaus D. 1996. Solution structure of the N-terminal RNP domain of U1A protein: The role of C-terminal residues in structure stability and RNA binding. *J Mol Biol* 257:398–411.
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
- Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domain: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 21:5803–5816.
- Boelens WC, Jansen EJR, van Venrooij WJ, Stripecke R, Mattaj IW, Gunderson SI. 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* 72:881–892.
- Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265:615–621.
- Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling CS, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, Weiner PK, Kollman PA. 1997. *AMBER 5* (software). University of California, San Francisco.
- Cheatham TE III, Miller JL, Fox T, Darden TA, Kollman PA. 1995. Molecular dynamics simulations on solvated biomolecular systems: The Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J Am Chem Soc* 117:4193–4194.
- Cornell WD, Cieplak P, Bayly CI, Kollman PA. 1995. A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc* 117:5179–5197.
- Darden T, York D, Pedersen L. 1993. Particle Mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
- Eriksson MAL, Nilsson L. 1995. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys J* 68:402–26.
- Ferrin TE, Huang CC, Jarvis LE, Langridge R. 1988. The MIDAS display system. *J Mol Graphics* 6:13–27, 36–37.
- Fox T, Kollman PA. 1996. The application of different solvation and electrostatic models in molecular dynamics simulation of ubiqui-

- tin: How well is the X-ray structure "maintained"? *Proteins Struct Funct Genet* 25:315–34.
- Gubser CG, Varani G. 1996. Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry* 35:2253–2267.
- Hall KB. 1994. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* 33:10076–10088.
- Hall KB, Kranz JK. 1995. Thermodynamics and mutations in RNA-protein interactions. *Methods Enzymol* 259:261–281.
- Jessen TH, Oubridge C, Teo CH, Pritchard C, Nagai K. 1991. Identification of molecular contacts between U1A small nucleoprotein and U1 RNA. *EMBO J* 10:3447–3456.
- Laird-Offinga IA, Belasco JG. 1995. Analysis of RNA-binding protein by in vitro genetic selection: Identification of an amino acid residue important for locking U1A onto its RNA target. *Proc Natl Acad Sci USA* 92:11859–11863.
- Lu J, Hall KB. 1997. Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry* 36:10393–10405.
- Miaskiewicz K, Ornstein RL. 1996. DNA binding by TATA-box binding protein (TBP): A molecular dynamics computational study. *J Biomol Struct Dynam* 13:593–600.
- Miller JL, Kollman PA. 1997. Theoretical studies of an exceptionally stable RNA tetraloop: Observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J Mol Biol* 270:436–450.
- Moras D, Poterszman A. 1995. Diverse modes of recognition. *Curr Biol* 5:249–251.
- Nagai K, Oubridge C, Ito N, Avis J, Evans P. 1995. The RNP domain: A sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem Sci* 20:235–240.
- Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348:515–520.
- Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the u1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372:432–438.
- Roxström G, Velázquez I, Paulino M, Tapia O. 1998. Molecular dynamics simulation of a Zif268-DNA complex in water. Spatial patterns and fluctuations sensed form a nanosecond trajectory. *J Phys Chem B* 102:1828–1832.
- Scherly D, Kambach C, Boelens W, van Venrooij WJ, Mattaj IW. 1991. Conserved amino acid residues within and outside the N-terminal ribonucleoprotein motif of U1A small nuclear ribonucleoprotein involved in u1 RNA binding. *J Mol Biol* 219:577–584.
- Tang Y, Nilsson L. 1998. Interaction of human SRY protein with DNA: A molecular dynamics study. *Proteins Struct Funct Genet* 31:417–433.
- Zeng Q, Hall K. 1997. Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA* 3:303–314.