Molecular Dynamics Simulations of the Complex between Human U1A Protein and Hairpin II of U1 Small Nuclear RNA and of Free RNA in Solution

Yun Tang and Lennart Nilsson
Center for Structural Biochemistry, Department of Bioscience at Novum, Karolinska Institutet, S-141 57 Huddinge, Sweden

ABSTRACT RNA-protein interactions are essential to a wide range of biological processes. In this paper, a 0.6-ns molecular dynamics simulation of the sequence-specific interaction of human U1A protein with hairpin II of U1 snRNA in solution, together with a 1.2-ns simulation of the free RNA hairpin, is reported. Compared to the findings in the x-ray structure of the complex, most of the interactions remained stable. The nucleotide U8, one of the seven conserved nucleotides AUUGCAC in the loop region, was unusually flexible during the simulation, leading to a loss of direct contacts with the protein, in contrast to the situation in the x-ray structure. Instead the sugar-phosphate backbone of nucleotide C15 was found to form several interactions with the protein. Compared to the NMR structure of U1A protein complexed with the 3′-untranslated region of its own pre-mRNA, the protein core kept the same conformation, and in the two RNA molecules the conserved AUUGCAC of the loop and the closest CG base pair were located in very similar positions and orientations, and underwent very similar interactions with the protein. Therefore, a common sequence-specific interaction mechanism was suggested for the two RNA substrates to bind to the U1A protein. Conformational analysis of the RNA hairpin showed that the conformational changes of the RNA primarily occurred in the loop region, which is just involved in the sites of binding to the protein and in agreement with experimental observation. Both the loop and stem of the RNA became more ordered upon binding to the protein. It was also demonstrated that the molecular dynamics method could be successfully used to simulate the dynamical behavior of a large RNA-protein complex in aqueous solution, thus opening a path for the exploration of the complex biological processes involving RNA at a molecular level.

INTRODUCTION

RNA molecules play a central role in a wide range of biological processes, such as storage of genetic material, propagation of genetic information, protein biosynthesis, and enzymatic activity. The large number of RNA structures that have recently been elucidated at high resolution by x-ray crystallography (Holbrook and Kim, 1997) and NMR spectroscopy (Ramos et al., 1997) have advanced our understanding of RNA structure and function (Uhlenbeck et al., 1997). Usually, RNA molecules perform their functions in tight association with RNA-binding proteins rather than on their own, so RNA-protein interactions are essential to these biological processes involving RNA. As a number of RNA-protein complex structures have been determined (Nagai, 1996; Arnez and Cavarelli, 1997), it is possible to investigate the specific features of RNA-protein interactions at the molecular level. For example, protein-induced RNA conformational changes are common and substantial in RNA-protein interactions, which resemble protein-protein interactions rather than DNA-protein interactions (Draper, 1995; Varani, 1997; Frankel and Smith, 1998).

The most common structural motif in RNA-binding proteins is the ribonucleoprotein (RNP) motif, found in more than 200 proteins (Varani and Nagai, 1998). An RNP motif is composed of 90–100 amino acids, which form an RNA-binding domain (RBD) that is present in one or more copies in proteins to bind pre-mRNA, mRNA, pre-ribosomal RNA (rRNA), and small nuclear RNAs (snRNAs) (Burd and Dreyfuss, 1994; Nagai et al., 1995). These small RNP domains share a common $\beta\alpha\beta\beta\alpha\beta$ sandwich tertiary folding structure (Nagai et al., 1990) and are highly conserved, although they bind diverse RNA targets with different affinities and specificities, ranging from picomolar to micromolar. The two central β strands (β3 and β1) of the folded domain, named RNP1 and RNP2, respectively, are identified as the RNA binding sites.

The human U1A protein is a 282-amino acid protein associated with the U1 snRNP (small nuclear RNP), a large RNA-protein complex involved in pre-mRNA splicing (Lührmann et al., 1990). It contains two copies of the RNP motif, one at the N-terminus and one at the C-terminus, which are connected by a protease-sensitive polypeptide of ~100 residues (Sillekens et al., 1987). The 102-amino acid N-terminal RNP motif binds specifically to hairpin II of U1 snRNA (Lutz-Freyermuth et al., 1990), with binding sites exclusively located in the loop region of the hairpin, whereas the C-terminal RNP does not appear to associate with any RNA (Lu and Hall, 1995). It has been found that U1A protein also binds to the 3′-untranslated region (3′UTR) of its own pre-mRNA, where it can inhibit polyadenylation at the 3′ end and regulate its own translation (Boelens et al., 1993). The U1A protein can bind to the two distinct RNA targets with very high affinity and specificity

Received for publication 12 February 1999 and in final form 1 June 1999.
Address reprint requests to Dr. Lennart Nilsson, Center for Structural Biochemistry, Department of Bioscience at NOVUM, Karolinska Institutet, S-141 57 Huddinge, Sweden. Tel.: +46-8-6089228; Fax: +46-8-6089290; E-mail: lennart.nilsson@csb.ki.se.

0006-3495/99/09/1284/22 $2.00

(both with a subnanomolar dissociation constant). The two RNA molecules contain the same AUUGCAC heptanucleotide in the loop region (see Fig. 1), which is believed to be the main site of interaction with the protein. A common recognition strategy has thus been suggested for the two different RNA molecules to bind to the same protein (Jovine et al., 1996).

Recently, the structure of a complex between the human U1A protein (N-terminal RNP domain with 98 residues) and a 21-nucleotide RNA hairpin, representing the hairpin II of U1 snRNA, was elucidated by x-ray crystallography at 1.92-Å resolution (Fig. 1; Oubridge et al., 1994). Within the structure there are many direct and water-mediated hydrogen bonds and stacking interactions between the RNA bases and aromatic side chains of the protein, revealing the stereochemical basis for sequence-specific RNA recognition by the RNP domain. Another complex of the same protein domain (including 102 residues) with the 3′UTR of U1A pre-mRNA was determined structurally by NMR spectroscopy (Allain et al., 1996), which further confirms the structural basis of the RNA binding specificity of U1A protein (Allain et al., 1997) and provides us an opportunity to compare the recognition strategy of the two RNA targets. The structure of the free U1A protein (Nagai et al., 1990; Avis et al., 1996) and free 3′UTR (Gubser and Varani, 1996) has also been elucidated by x-ray crystallography and NMR spectroscopy, allowing us to investigate conformational changes during RNA-protein interactions. The two structures of U1A protein complexed with distinct RNA targets have clarified many important aspects of RNP-RNA

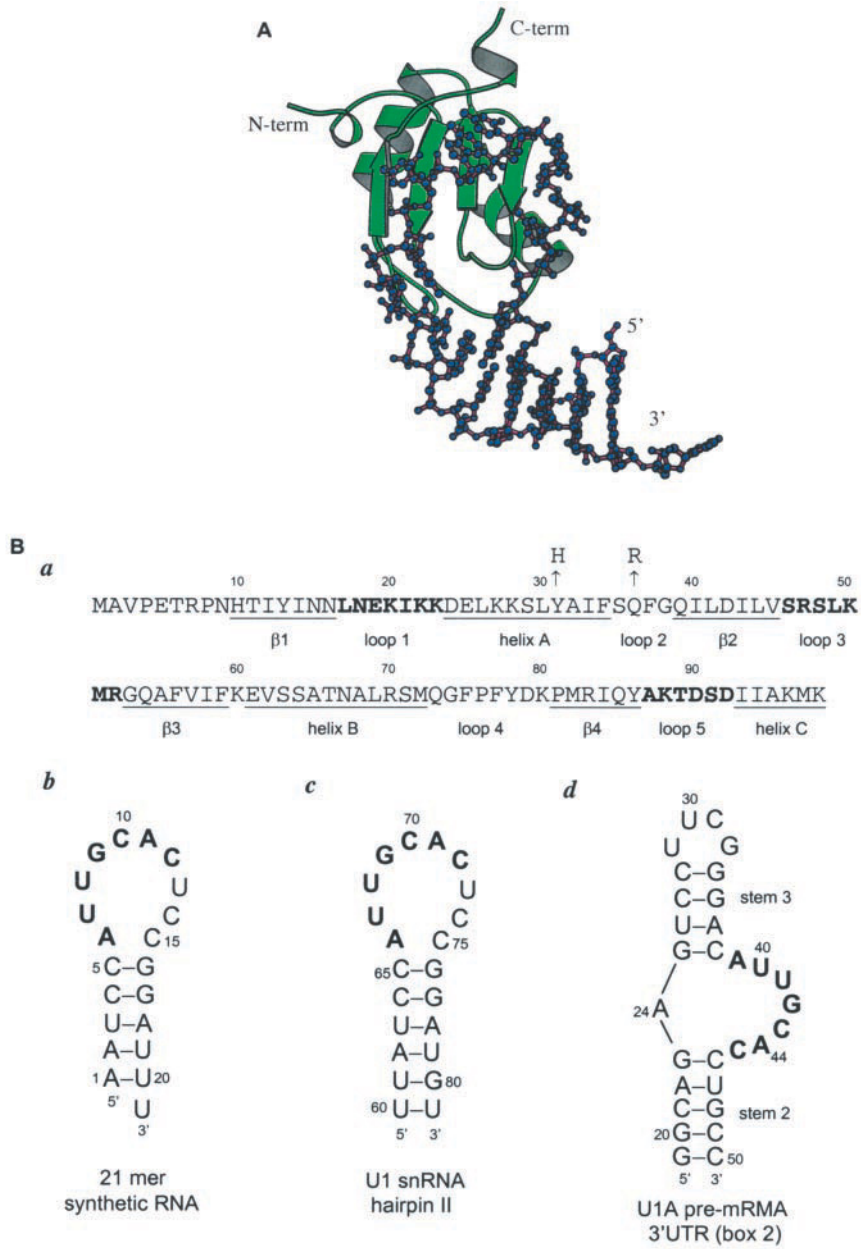


FIGURE 1 (*A*) Schematic representation of the x-ray structure of the U1A-RNA hairpin complex. The U1A protein is shown in ribbon, and the RNA is shown with a ball-and-stick model. This figure, together with Figs. 5 *A*, 5 *B*, 7, 8, 9, 11 *B*, and 11 *C*, was generated with the program MOLSCRIPT (Kraulis, 1991). (*B*) Sequence and number of related protein and RNA segments: (*a*) U1A protein; (*b*) the synthetic RNA hairpin used in the x-ray structure; (*c*) the hairpin II of U1 snRNA; (*d*) the 3′-untranslated region of U1A pre-mRNA (box 2) used in the NMR complex structure.

recognition. However, they remain static views, and they have raised intriguing questions concerning the molecular basis of specificity; for example, interactions involving the protein variable loops differ significantly (Allain et al., 1997). Therefore, it is necessary to study the specific interactions of U1A protein complexed with RNA from a dynamic and thermodynamic point of view.

Hall observed that chemical shifts of many protons from the bound U1 snRNA were substantially different from those of the free RNA, especially in the loop region of the hairpin (Hall, 1994). Allain and Varani studied the free structure of the RNA hairpin, which shows that the hairpin loop is largely unstructured in free RNA apart from the first three bases in the loop, which stack on each other (unpublished results, just mentioned in Oubridge et al., 1994). However, no structural data of this hairpin in the free state are available. Therefore, it is also necessary to run a conformational analysis on the RNA hairpin before and after its binding to the protein, to investigate how much the conformation of the RNA is changed by the binding of U1A protein, which could help us to better understand how the RNA performs its function.

Molecular dynamics (MD) simulation is a powerful tool for analyzing the structural and dynamic features of biomacromolecules (Karplus and Petsko, 1990; Tidor, 1997). It provides detailed information about atomic interactions as a function of time, which can both enhance and complement the experimental results. Water molecules, which have been shown to play an important role in both the affinity and specificity of protein-nucleic acid interactions (Schwabe, 1997), are easily incorporated into MD simulations to model the solvent effects. Besides proteins, there are an increasing number of MD simulations focusing on nucleic acids (Auffinger and Westhof, 1998), from which some conformational features of nucleic acids have been obtained (Cheatham and Kollman, 1997). For RNA, much of the simulation effort has been concentrated on UUCG tetraloop (Miller and Kollman, 1997), transfer RNA (Auffinger and Westhof, 1997; Auffinger et al., 1999), and hammerhead ribozyme (Hermann et al., 1998). Several MD simulations to date have also been performed on DNA-protein complexes (Eriksson et al., 1995; Tang and Nilsson, 1998; Nilsson, 1998). However, MD studies of RNA-protein interactions are still at an early stage (Reyes and Kollman, 1999). This is due, in part, to the fact that the structures of only a handful of RNA-protein complexes have been determined at high resolution, and the available RNA-protein complexes are often large and the RNA substrates are structurally more diverse than the DNA (Weeks, 1997), which makes RNA-protein interactions less tractable by simulation.

Pressure and temperature are thermodynamic variables, which may affect the simulation results. In biophysical chemistry, pressure has long been used as an environmental variable for probing the interactions of proteins with ligands and to study conformational equilibria, protein dynamics, and other properties of the native state of proteins. To mimic experimental conditions, the constant pressure and temperature (NPT) supercanonical ensemble (Kitchen et al., 1992; Ceccarelli and Marchi, 1997) was used in the present work. We performed an NPT MD simulation on the U1A protein complexed with the hairpin II of U1 snRNA and the free RNA hairpin in solution. The aim of this study is to investigate the dynamic features of the specific U1A-RNA interaction, to recognize the conformational changes of the RNA before and after binding to the protein, and to explore the structural basis of U1A protein sequence-specific binding to the RNA substrate.

METHODS

The solvated constant-pressure molecular dynamics simulations were performed on a DEC AXP 4100/300E 4CPU parallel computer and a four-node α cluster ($\sim$13 h per 10 ps for the complex and 6 h per 10 ps for the free RNA), using the program CHARMM (Brooks et al., 1983), version c26a2, and the all-atom version 22 force field (MacKerell et al., 1995, 1998). The TIP3P water model was used to simulate the solvent (Jorgensen et al., 1983).

Simulation details

The starting coordinates of the U1A-RNA complex were extracted from the B/Q monomer in the x-ray trimer structure at 1.92-Å resolution (Oubridge et al., 1994), PDB entry code 1URN (Bernstein et al., 1977), including 769 protein heavy atoms, 436 RNA heavy atoms, and 157 water oxygen atoms. In the x-ray structure, the C-terminus of U1A was poorly ordered beyond residue 96, and residues Met^{1} and Lys^{98} were omitted. Two mutations, $Tyr^{31}\rightarrow$His and $Gln^{36}\rightarrow$Arg, which were not directly involved in RNA recognition, were contained in the protein structure (see Fig. 1; Oubridge et al., 1994). Bases U13, C14, C15 and the 3′ end of the RNA chain were also poorly ordered. All of the hydrogen atoms were added by the CHARMM subroutine HBUILD (Brünger and Karplus, 1988). The complex structure together with the 157 crystal water molecules was minimized in vacuo for 1000 steps, using the adopted basis Newton-Raphson (ABNR) method, to remove the unfavorable contacts, keeping harmonic constraints with a force constant of 20.0 kcal/(mol·Å^2) on heavy atoms of the complex.

The minimized complex structure, together with the 157 solvent water molecules, was then inserted into the center of a water box, keeping any complex atoms at least 10.0 Å away from the boundary and leading to a box size of $84 \times 56 \times 52$ Å^3. Water molecules closer than 1.8 Å from any solute or crystallographic water atoms were deleted from the water box. Thirteen sodium counterions were added at random positions into the system, at least 3.0 Å from any complex atoms, to make the system electroneutral. The final system contained 24,805 atoms, including the 2250 solute atoms. The solvated U1A-RNA complex system was min-

imized again, first with the steepest descent method for 100 steps, then with the ABNR method for 1000 steps, to adjust the water molecules and counterions locally and eliminate any residual geometrical strain, keeping the heavy atoms of the complex fixed. The minimized solvated system was served as the initial structure of the subsequent molecular dynamics simulation.

After the initial structure was prepared, the molecular dynamics simulation was begun with an initial and equilibration stage, followed by a 400-ps production run, giving a total simulation time of 600 ps. All of the complex atoms were released during the whole process. The initial atomic velocities were assigned from a Gaussian distribution corresponding to a temperature of 300 K. The nonbonded energies and forces were smoothly shifted to zero at 12.0 Å (Steinbach and Brooks, 1994), and a constant dielectric ($\epsilon = 1$) was used for electrostatic interactions. The nonbonded list, including neighboring atoms within a 14.0-Å distance, was updated every 20 steps. All hydrogens were treated explicitly; a time step of 2 fs was used to integrate the equations of motion with the Leapfrog Verlet algorithm. All bonds involving hydrogens were constrained with the SHAKE algorithm (Ryckaert et al., 1977). Coordinates and energies were saved every 100 time steps for further analysis.

In the energy minimization and the molecular dynamics simulation, periodic boundary conditions were applied in all directions. The solvent and counterion images were updated every 20 steps. The NPT ensemble was implemented, using the weak coupling scheme (Berendsen et al., 1984), with a pressure coupling time of 5.0 ps and a temperature coupling time of 5.0 ps. The system temperature was set at 300 K, and the reference pressure of the system was set at 1.0 atm. The isothermal compressibility was set at 4.63×10^{-5} atm^{-1}, and the value was approximated from experimental data for water.

The starting coordinates of the free RNA hairpin were extracted from the complex, followed by the same pretreatment and molecular dynamics process as mentioned above, with a water box of $64 \times 46 \times 40$ $Å^3$. Twenty sodium counterions were placed 3.6 Å from the phosphorous atom on the O═P═O bisectors. The whole system contained 11,800 atoms. The molecular dynamics simulation consisted of a 0.6-ns initial stage followed by a 0.6-ns production run, giving a total simulation time of 1.2 ns.

Analysis of the simulation

The analyses of the simulations focused on the production stages. Four structures averaged from different short periods of the production stage of the complex, together with the average structure from the whole production run, were used in the analysis of the complex. All of these structures were minimized with the ABNR method for 500 steps with a 20.0 kcal/(mol·$Å^2$) harmonic force constant on the heavy atoms, and the water molecules and counterions were deleted. The average conformations of the RNA hairpin free and in complex created from the corresponding production stage, with a similar treatment, were used in the analysis of RNA.

The root mean square deviations (RMSDs) of the selected atoms were calculated from the trajectory at 0.2-ps intervals, with the initial structure as the reference. Average RMSD values of the backbone and the side chains or bases were then calculated for each residue or nucleotide. The root mean square fluctuation (RMSF) of each residue was calculated similarly.

The protein-RNA, protein-solvent, and RNA-solvent nonbonded interaction energies were calculated from the trajectory at 1.0-ps intervals, using CHARMM force field.

The electrostatic potentials of the protein and RNA were calculated from the average dynamics structure of the complex and shown on the solvent-accessible surface, using the program GRASP (Nicholls, 1992).

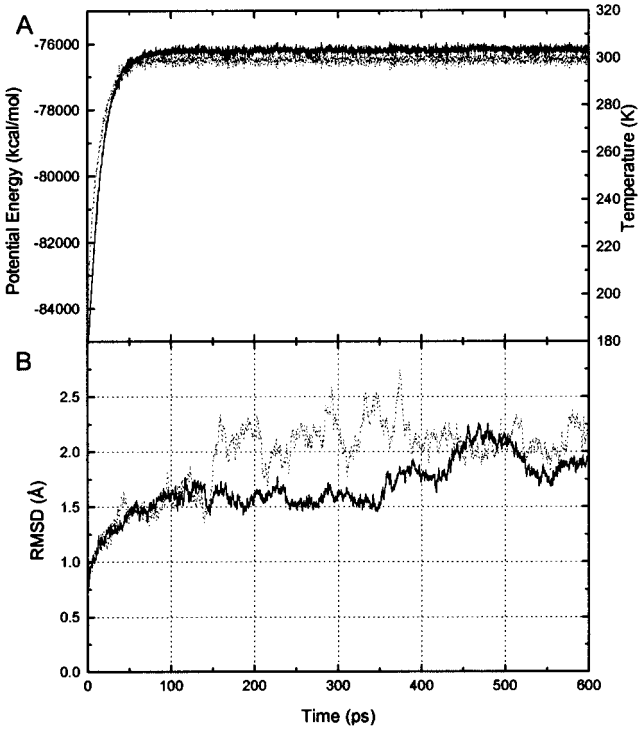
The solvent-accessible surface areas were calculated from the trajectory at 6.0-ps intervals, using the definition of Lee and Richards (1971), with a probe of radius 1.4 Å. The Lennard-Jones radii values were used for the complex atoms.

The hydrogen bonds were analyzed from the production trajectories with 0.2-ps and 1.0-ps time resolutions. A hydrogen bond (A···H−D) was defined by an A-H distance of less than 2.5 Å and an A-H-D angle of more than 120°. The percent occupancy of a hydrogen bond was defined as the number of frames with the hydrogen bond present divided by the number of total analysis frames. The lifetime of a specific incarnation of a hydrogen bond was calculated as the time elapsed from its first appearance until it was first broken. The average lifetime of a hydrogen bond during the simulation was then calculated as the average of all of its incarnations, with the option of excluding those with a lifetime shorter than a time cutoff of 0.1, 1.0, and 5.0 ps, separately. The water-mediated hydrogen bonds were analyzed similarly. A water bridge occurred between the protein and RNA atoms, which formed hydrogen bonds with a common water molecule.

The most probable positions for the water molecules and sodium ions around the solute were calculated from the production trajectories at 1.0-ps intervals, by the use of a three-dimensional histogram with a $(2\ Å)^3$ bin size, by orienting every coordinate set to be superimposed on the average structure of the solute. Only probabilities larger than 40% for the water and 10% for the sodium were considered in the density map.

RESULTS

The potential energy of the system, the temperature, and the RMSD values of the protein and RNA atoms from the initial x-ray structure remained quite stable during the production run (shown in Fig. 2), except that the RMSD of the protein underwent a small increase at 450–500 ps. The global RMSD of the complex was below 2.5 Å after a 600-ps simulation.



Interaction energies and solvent-accessible surface

The U1A protein–RNA hairpin nonbonded interaction energies were stable in the simulation (Fig. 3 *A*). The van der Waals interaction energy was constant, at about −90 kcal/mol. The major component of the total interaction energy was the electrostatic (including hydrogen bonding) interaction energy at about −310 kcal/mol, with some fluctuation. In particular, there was a decrease at 450–500 ps, indicating that the electrostatic interaction was more favorable during this period. These interaction energies did not include the effects of the solvent.

The origins of the favorable electrostatic interaction energies are clearly illustrated by the electrostatic potential calculated from the average structure of the complex. A strongly positive electrostatic potential region was found on the solvent-accessible surfaces of loops 1 and 3 of the U1A protein, whereas the major groove surface of the RNA double-helical stem was a strongly negative potential region (Fig. 4). The positive and negative potential regions faced each other and matched very well.

The solvent-accessible surface area of the complex increased slightly with time (Fig. 3 *B*). In the first 250 ps of the production run (200–450 ps), the surface area of the complex was constant, but after this it increased a little and

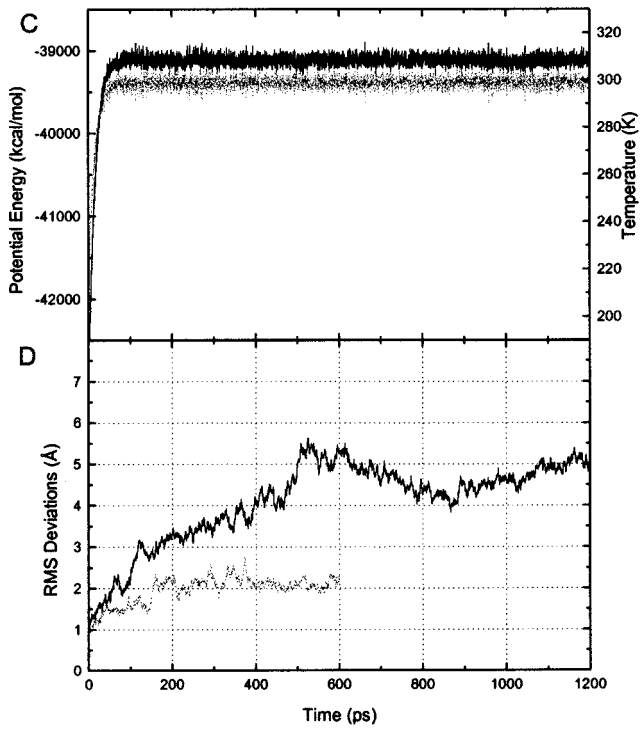


FIGURE 2 Time evolution of the potential energy of the system (*solid line*) and temperature (*dotted line*) of (*A*) the complex and (*C*) the free RNA. Time evolution of the root mean square (RMS) deviations of (*B*) U1A protein (*solid line*) and RNA hairpin (*dotted line*) from the initial structure, and (*D*) the RNA free (*solid line*) and in complex (*dotted line*).

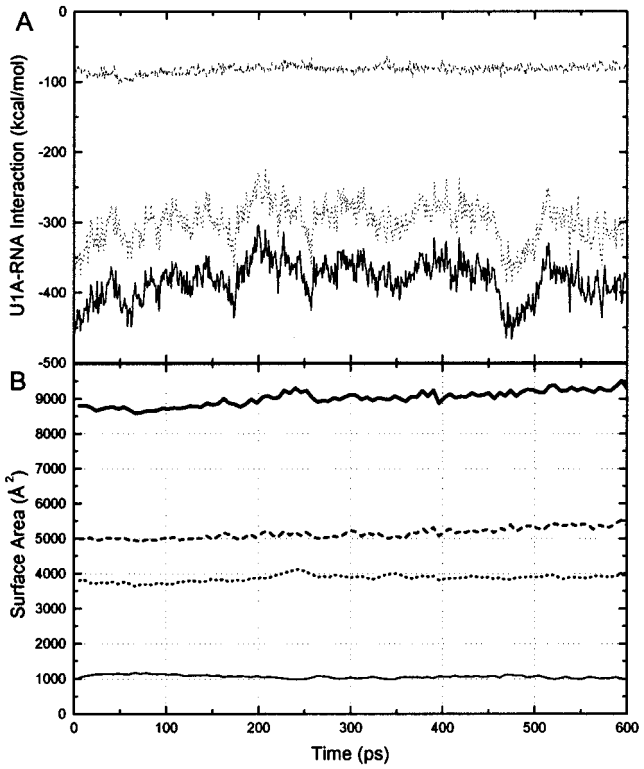


FIGURE 3 (*A*) Time-dependent RNA-protein nonbonded interaction energies. *Solid line*: total energy; *dashed line*: van der Waals energy; *dotted line*: electrostatic energy. (*B*) Time-dependent solvent-accessible surface areas of the complex. *Thick solid line*: whole complex; *dashed line*: protein; *dotted line*: RNA; *thin solid line*: buried area.

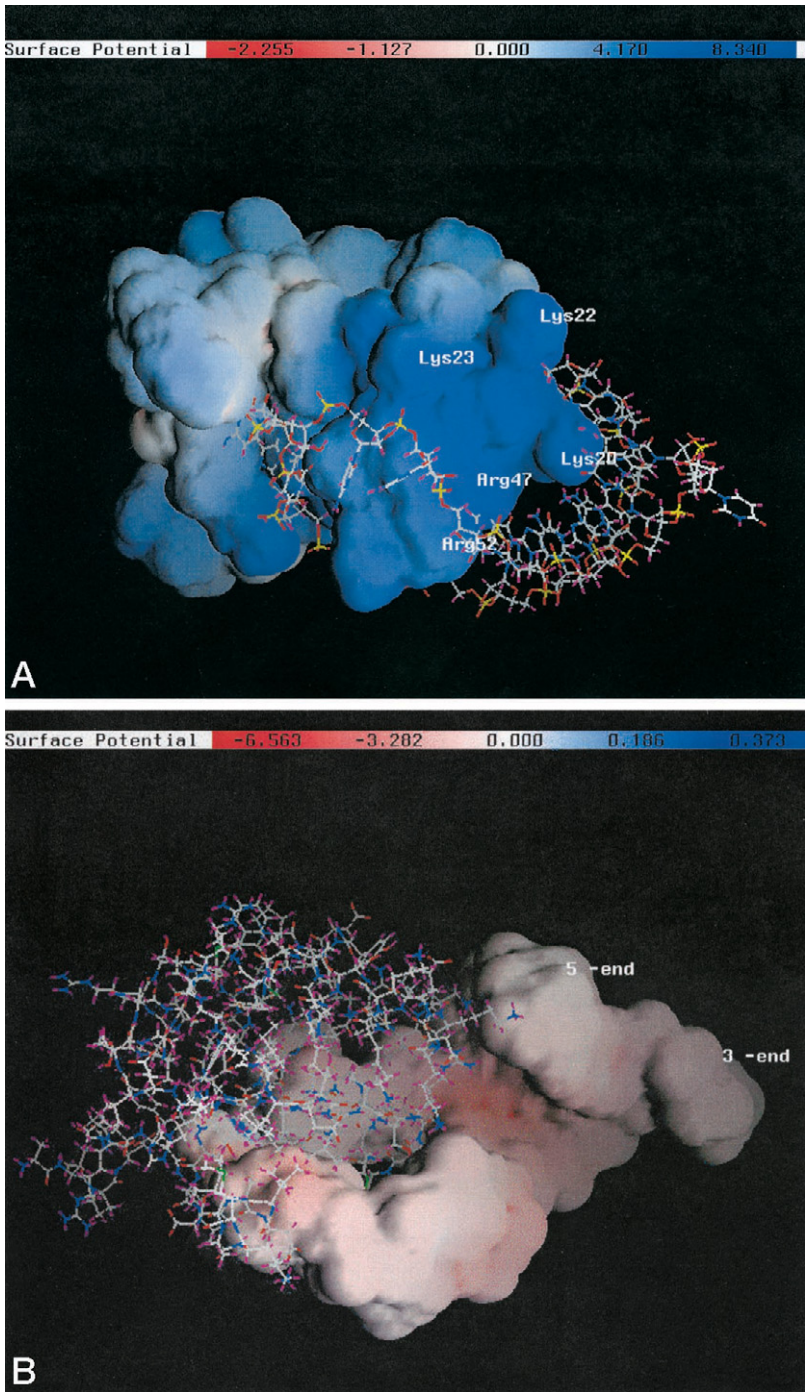
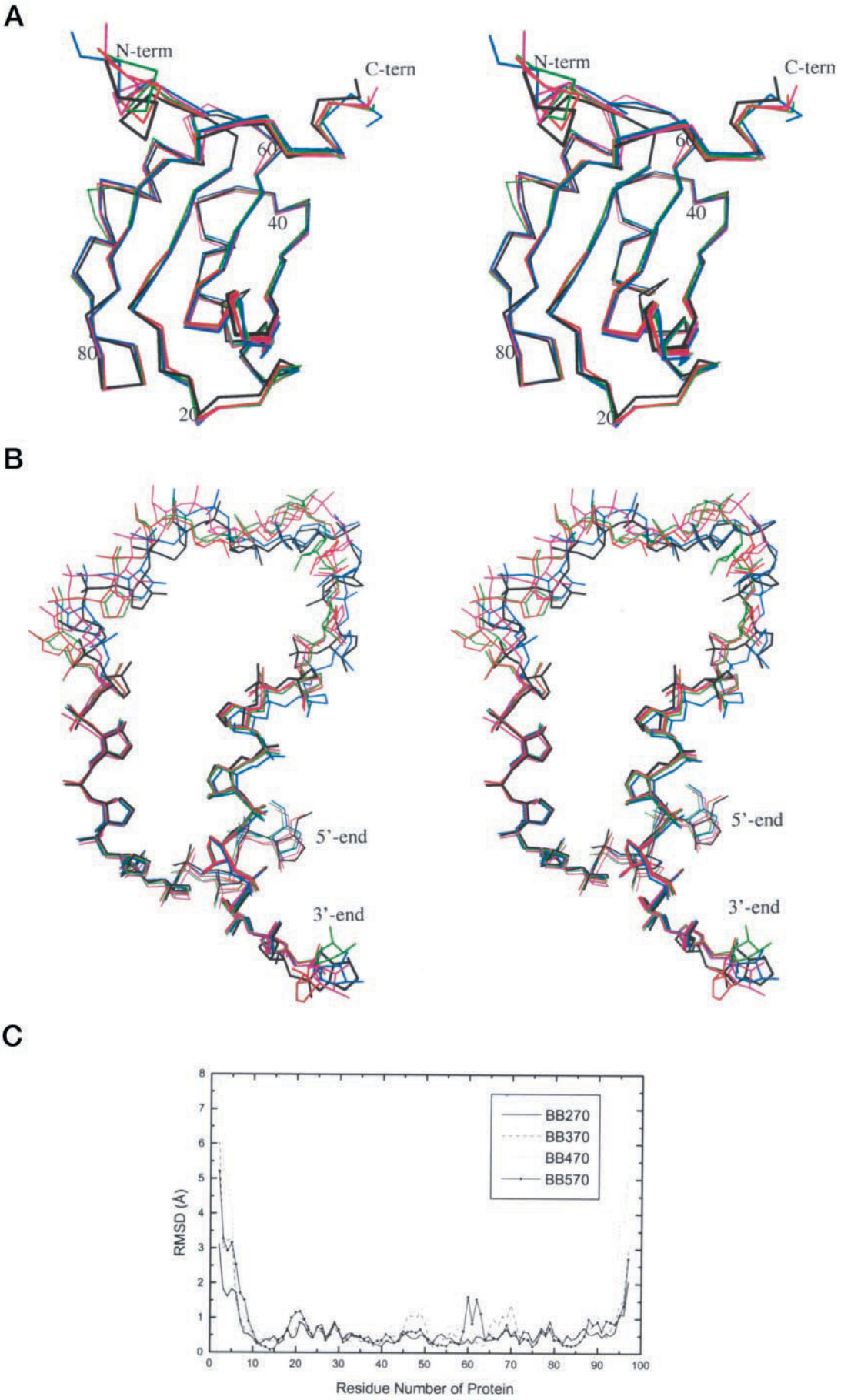


FIGURE 4 Electrostatic potentials of (*A*) the U1A protein and (*B*) the RNA hairpin are shown on the solvent-accessible surface, produced with the program GRASP (Nicholls, 1992). The surface charge distribution is color coded as follows: dark blue, positive ($\sim$8.5*kT*); red, negative ($\sim -$6.5*kT*); white, neutral. *k* is the Boltzmann constant and *T* is the temperature.

remained constant again after 500 ps, at a 300 Å^2 higher level. The increased surface area of the complex came from the protein, and the exposed area of the RNA was constant. However, the buried surface area remained constant, at $\sim$1000 Å^2 for each part, during the simulation. The buried surface area was $\sim$17% for the protein and $\sim$22% for the RNA.

Deviations and fluctuations

From the results above, it seemed that something had happened on the U1A protein at 450–500 ps. Average structures from four short windows, 270–300, 370–400, 470–500, and 570–600 ps, together with the x-ray structure, were superimposed to investigate the changes (Fig. 5). Compared to the x-ray structure of the complex, most of the folded domain of the protein (α-carbon atoms only) overlapped very well, except that the N- and C-terminals were in severe disorder. In particular, there were structural deviations in loop 3 and the end of helix B for the second structure; in loops 1 and 3 and helix B for the third structure; and in loop 1 and the beginning of helix B for the last structure. The RNA hairpin (backbone only) appeared flex-


A
N-term
C-tern
60
40
80
20
N-term
C-tern
60
40
80
20
B
5'-end
3'-end
5'-end
3'-end
C
BB270
BB370
BB470
BB570
RMSD (Å)
Residue Number of Protein
0
1
2
3
4
5
6
7
8
0
10
20
30
40
50
60
70
80
90
100

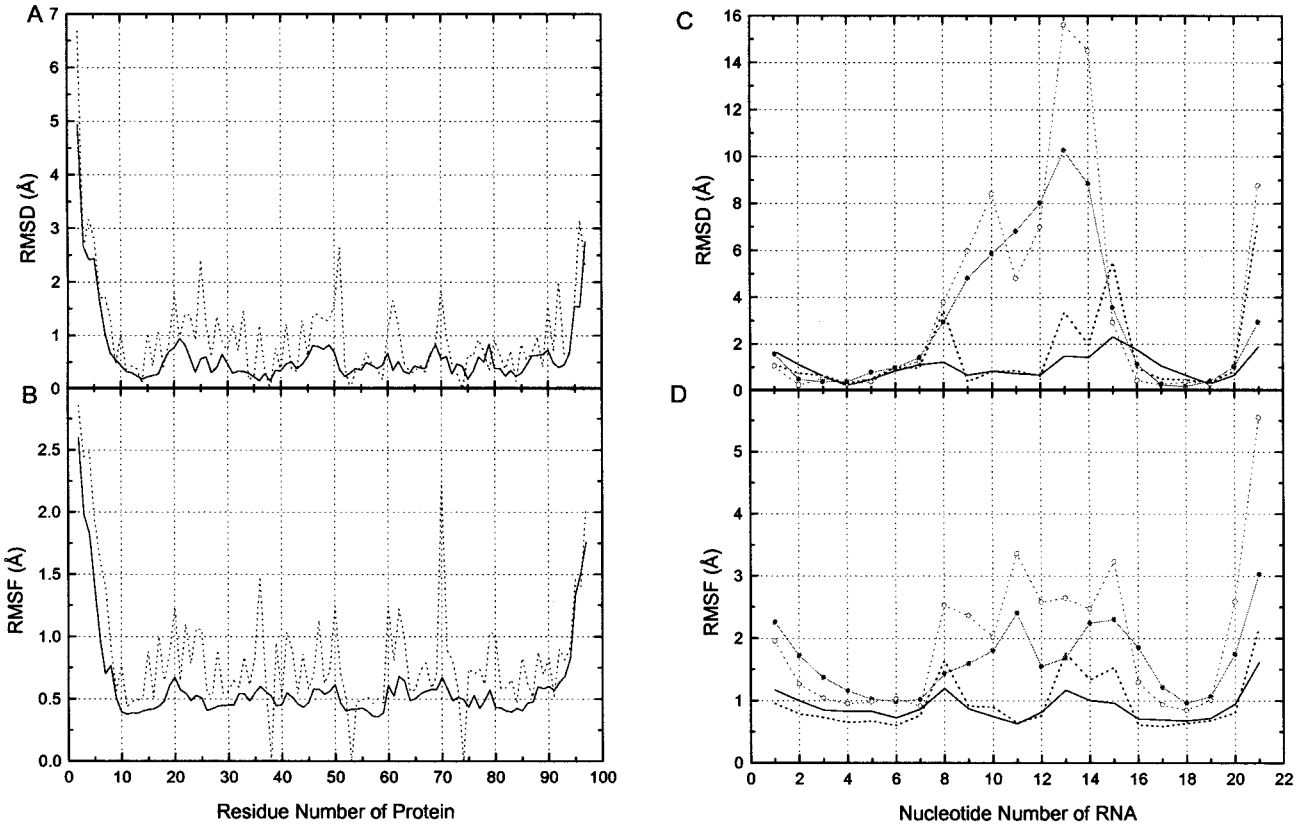


FIGURE 6 The RMS deviations of (*A*) each residue in the U1A protein and (*C*) each nucleotide in the RNA hairpin, as well as the RMS fluctuations of (*B*) each residue in the U1A protein and (*D*) each nucleotide in the RNA hairpin from the production run, compared with the initial structure of the complex. ——, Backbone; ·····, side chains or bases. —●—, backbone of free RNA; ···○···, bases of free RNA.

ible in the loop from U8 to C15, but highly stable in the double-helix stem and A6, U7.

The detailed local motions of the complex structure during the final simulation were reflected quantitatively from the RMSD and RMSF values of each residue in the complex. Fig. 6 shows a uniform mobility for most of the residues in the U1A protein in the production stage, with ~0.5 Å for the RMSD and RMSF values for the backbone atoms, except for the residues at the two termini, especially the N-terminal. The backbone of residues Glu^{19} to Lys^{22} of loop 1 had higher RMSD values close to 1.0 Å, whereas the RMSF values of these residues remained stable, indicating that these four residues changed their conformations during the simulation. The backbone of residues Ser^{46} to Lys^{50} of loop 3 had RMSD values a little higher than the average, but RMSF values very little higher than the average, which demonstrated that these residues had also changed their conformations somewhat. Notably, the long positively charged side chain of Arg^{52} remained highly stable with very low RMSD and RMSF values, indicative of the important structural role played by Arg^{52} in the RNA-protein interaction. Interestingly, residues Leu^{69} and Arg^{70} of helix B, which had no direct contacts with RNA, also had higher than average RMSD values. The side chain of Arg^{70} in particular was highly flexible and deviated from the original position.

Fig. 6, *C* and *D*, shows that the sugar-phosphate backbone of nucleotides U13 to G16 had RMSD values of ≥1.5 Å in the complex, whereas the backbone of nucleotides U8, U13, C14, and U15 had RMSF values of ≥1.0 Å, except for the two ends. The bases of U8, U13, and C15 had unusually high RMSD and RMSF values, besides the 3′-end nucleotide U21. C14 also had a high RMSF value, but this was weaker than that of U8, U13, and C15. Contrasted to the high mobility of U8, nucleotides A6, U7, and G_9CAC_{12} in the loop region were quite stable during the simulation, the backbone as well as the bases, which is very similar to the nucleotides within the double-helix stem.

Compared to the RNA in complex, the free RNA hairpin had a quite different conformation for the loop region, from

FIGURE 5 Stereo view of the superposition of four structures of (*A*) U1A protein and (*B*) RNA hairpin averaged from 270–300 ps (*red*), 370–400 ps (*green*), 470–500 ps (*blue*), and 570–600 ps (*purple*), together with the x-ray structure (*black*). (*C*) The backbone RMS deviations of the four average protein structures, compared with the x-ray structure.

U8 to C15, indicated by the RMSD values much higher than 2.0 Å and RMSF higher than 1.5 Å (Fig. 6, *C* and *D*). In most of these nucleotides, the bases were deviating and fluctuating more than the backbone, especially U13 and C14, the bases of which deviated from the starting structure by more than 14 Å. The double-helix stem and the first two nucleotides of the loop remained stable, with RMSDs less than 1.0 Å. Nucleotides A1, A2, G16, and U20 had RMSF values greater than 1.5 Å for the backbone.

Overall structure of the complex

The overall structure of the complex was very similar to the initial x-ray structure, with RMSD values of 0.92 Å and 1.68 Å for the backbone and side chains of the U1A protein, and 1.35 Å and 2.38 Å for the backbone and bases of the RNA hairpin, respectively. For the U1A protein in the average structure, the four β-sheets were in an antiparallel arrangement on the surface. Hydrophobic residues in these sheets were packed on each other and formed a hydrophobic core, which linked the four β-sheets together (Fig. 7 *A*). Among these residues, only Tyr^{13}, Leu^{44}, and Phe^{56} were exposed to bind to the RNA. Therefore, the four β-sheets, together with helices A and B, were quite stable and almost completely overlapped with those in the x-ray structure. Loop 3 between β2 and β3 of the protein protruded from the surface and was confronted with the N- and C-terminals. There was a large hydrophilic and charged region on the outside surface of loop 3, loop 1, and part of helix A. Together with loop 5 between β4 and helix C, which is also highly polar, these three polar loops were involved in extensive contacts with the RNA hairpin (Fig. 7 *A*).

The RNA hairpin is attached to the surface of the protein in an open conformation. The 10-nucleotide loop of the RNA hairpin encircles loop 3 of the protein and falls into the hydrophobic saddle region between loop 3 and the C-terminal of the protein (Fig. 7 *A*). The seven conserved nucleotides, $A_6UUGCAC_{12}$, were splayed out on the protein surface. The bases of A6 and U7 were stacked with the bases of the double-helix stem. They and the base of G9 were located in the inside of the loop, whereas the bases of U8, C10, A11, and C12 were on the outside of the loop. The base plane of C10 was approximately perpendicular to that of A11, whereas the bases of A11 and C12 were stacked on each other. The bases of G_9CAC_{12} were buried in the protein hydrophobic core, but U8 had no obvious contacts with the protein, which was different from the situation in the x-ray structure, although the related three residues Asn^{16}, Lys^{80}, and Arg^{83} remained stable during the simulation. The last three nucleotides in the loop, UCC, had no obvious contacts with the protein, with bases extending into the solution and the backbone pointing toward the protein. Their bases were partially stacked on each other. The double-helical stem of the hairpin faced toward the protein through the major groove but had no obvious contacts with the surface of the protein. The 5′ end of the stem was very close to several positively charged residues of the protein, such as Lys^{20}, Lys^{22}, and Arg^{47}.

Contacts at the RNA-protein interface

There are several nucleotides contacting with U1A protein at the RNA-protein interface, such as A6, U7, G9, C10, A11, C12, and G16. The residues in U1A protein contacting RNA bases included Tyr^{13}, Asn^{15}, Glu^{19}, Leu^{44}, Arg^{52}, Gln^{54}, Phe^{56}, Gln^{85}, Tyr^{86}, Lys^{88}, Thr^{89}, Asp^{90}, Ser^{91}, and Asp^{92}. These nucleotides and amino acid residues form a variety of electrostatic, hydrophobic, and hydrogen-bonding interactions at their interface.

From Fig. 4 we can see that the positive electrostatic potential on the outside of the U1A protein loops 1 and 3 was just opposed to the negative electrostatic potential on the major groove surface of the RNA hairpin. In the positively charged region of the protein, residues Lys^{20} and Lys^{22} formed electrostatic interactions with the phosphate groups of nucleotides A2, U3, and C4 of the RNA (Table 1). Arg^{47} also formed some electrostatic interactions with these nucleotides, but with low occupancy, and so is not listed in Table 1. The phosphate groups of C10 and G16 formed electrostatic interactions with the side chains of Lys^{88} and Arg^{52}, respectively. Moreover, we found an electrostatic interaction between the phosphate group of C15 and the side chain of Lys^{23}, which was not observed in the x-ray structure.

The bases of G9, C10, A11, and C12, four of the seven conserved nucleotides in the loop region, were buried in the hydrophobic core (Fig. 7), so a wide range of hydrophobic contacts was found between these nucleotides and the protein (Table 2). The side chain of Gln^{54} was parallel to the base plane of G9 and underwent a hydrophobic interaction. The aromatic plane of Tyr^{13} and the long side chain of Lys^{88} were parallel to the base of C10 on both sides and underwent stacking interactions with the base plane of C10. The aromatic plane of Phe^{56} was stacked with the base plane of A11. The CD2 atom of Leu^{44} underwent a hydrophobic interaction with the C2 atom of A11. One side of the base of C12 was stacked with the base of A11, and on the other side, it had hydrophobic contacts with the side chain of Asp^{92}. In contrast to the x-ray structure, the base of C12 did not undergo stacking interactions with the side chain of Asp^{92}. The sugar rings of G9, C10, and A11 also made partial hydrophobic contacts with Gln^{54} and Phe^{56}, respectively. The side chain of Leu^{49} underwent several hydrophobic interactions with both the sugar and base of G16, and it interacted with the base of G6. In addition, the CB atom of Ser^{48} underwent hydrophobic interactions with the sugar backbone of C15. This was the second interaction related to C15.

Besides hydrophobic contacts, the four buried nucleotides G_9CAC_{12} also formed hydrogen-bonding interactions with the U1A protein (Fig. 7).

A number of direct and water-mediated hydrogen bonds were monitored at the RNA-protein interface during the

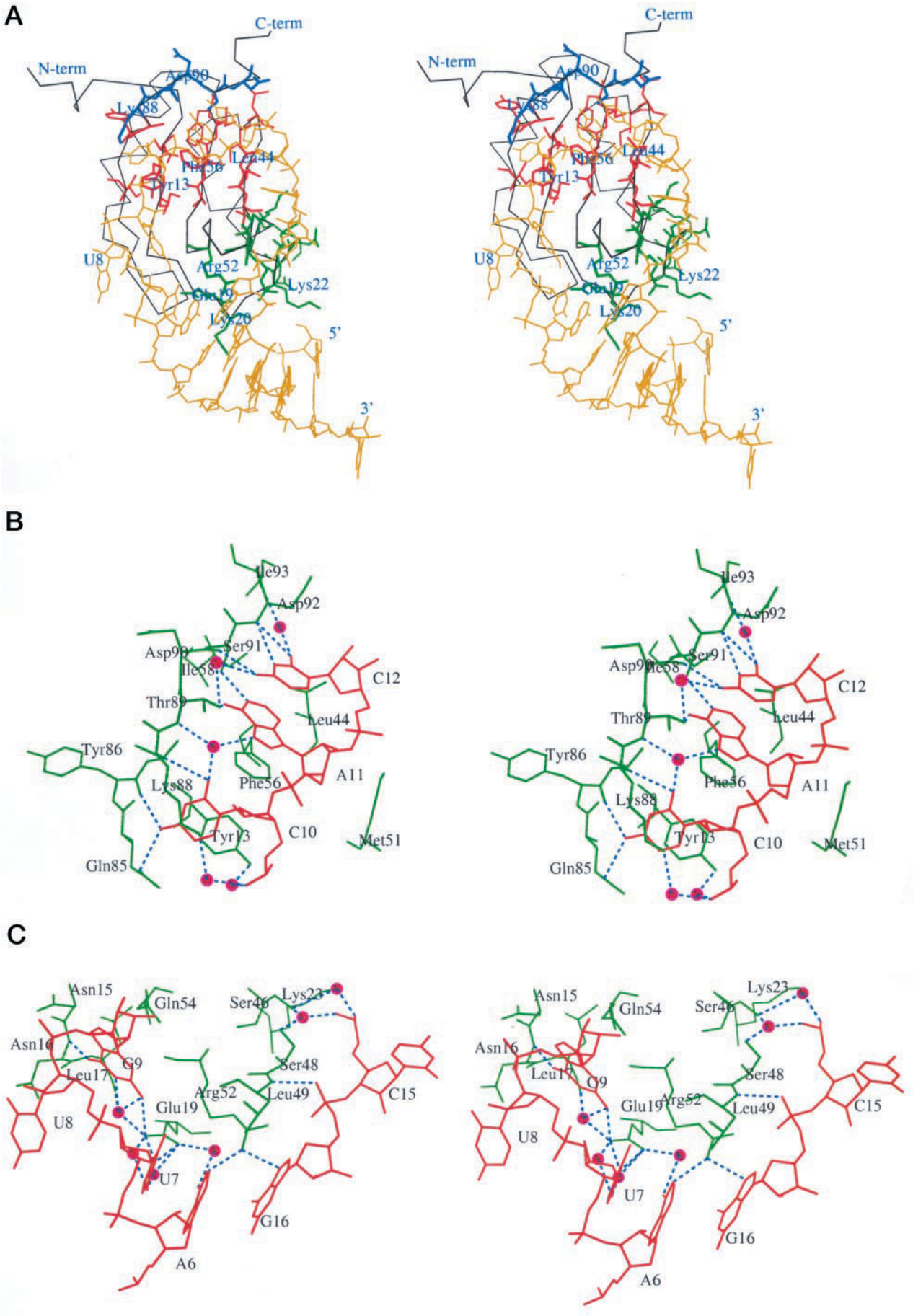

A
C-term
N-term
Asp90
Lys88
Phe56
Leu44
Tyr13
U8
Arg52
Lys22
Glu19
Lys20
5'
3'
B
Ile93
Asp92
Ser91
Asp90
Ile58
C12
Thr89
Leu44
Tyr86
Lys88
Phe56
A11
Tyr13
C10
Met51
Gln85
C
Asn15
Gln54
Ser46
Lys23
Asn16
G9
Leu17
Ser48
Arg52
Leu49
C15
Glu19
U8
U7
G16
A6

TABLE 1 Electrostatic interactions at the RNA-protein interface

Nucleotide atoms	Amino acid atoms	Average distance (Å)	RMS fluctuation (Å)
A2 P	Lys^{22} NZ	6.94	1.43
U3 P	Lys^{20} NZ	5.97	1.35
C4 P	Lys^{20} NZ	6.42	0.76
C10 P	Lys^{88} NZ	5.97	0.63
C15 P	Lys^{23} NZ	6.62	0.72
G16 P	Arg^{52} NH2	5.24	0.48

TABLE 2 Hydrophobic contacts at the RNA-protein interface

Nucleotide atoms	Amino acid atoms	Average distance (Å)	RMS fluctuation (Å)
A6 C2	Leu^{49} CD1	3.93	0.32
G9 C1′	Gln^{54} CD	4.14	0.23
G9 C2′	Gln^{54} CD	3.96	0.26
A11 C1′	Phe^{56} CE1	3.69	0.21
A11 C1′	Phe^{56} CZ	3.85	0.24
A11 C2	Leu^{44} CD2	4.18	0.65
C12 C2	Asp^{92} CB	3.89	0.22
C15 C5′	Ser^{48} CB	4.02	0.27
G16 C1′	Leu^{49} CG	3.99	0.22
G16 C1′	Leu^{49} CD1	3.98	0.26
G16 C4	Leu^{49} CD1	4.11	0.30
G16 C8	Leu^{49} CD1	3.83	0.27
G9 base	Gln^{54} side chain	3.99	
C10 base	Tyr^{13} side chain	3.74	
C10 base	Lys^{88} side chain	4.06	
A11 base	Phe^{56} side chain	3.92	

simulation (Table 3). These hydrogen bonds were formed not only between the backbone of RNA and the backbone of protein, but also between backbone and side chain, bases and backbone, and bases and side chain of RNA and protein, respectively. Eight of the direct hydrogen bonds and five water-mediated ones were formed between RNA bases and protein side chains. In particular, the side chain of residue Glu^{19} formed three direct and four water-mediated hydrogen bonds with the bases of A6, U7, and G9, which indicated that Glu^{19} was important to the specificity. Direct hydrogen bonds between atoms A6 N1 and Arg^{52} HH12, U7 H3 and Glu^{19} OE1, G9 O6 and Asn^{16} HN, C10 H41 and Tyr^{86} O, A11 N1 and Ser^{91} HG1, C12 H41 and Asp^{90} O, and G16 N7 and Arg^{52} HH11 had high occupancies (more than 80%) and were also found in the x-ray structure.

Eleven of the 15 direct hydrogen bonds were consistent with those observed in the x-ray structure. All of these hydrogen bonds were present more than half of the production run time. Four of these with more than 80% occupancy formed between bases of RNA and side chain of protein, namely A6 N1 and Arg^{52} HH12, U7 H3 and Glu^{19} OE1, A11 N1 and Ser^{91} HG1, and G16 N7 and Arg^{52} HH11, indicating that these four hydrogen-bonding interactions were critical for the specific RNA-protein interaction. Among the 12 water-mediated hydrogen bonds, in which three bridged more than two residues, only two were present in the x-ray structure, possibly because the water-mediated hydrogen bonds were less stable than the direct ones.

In the x-ray structure, it was found that most of the hydrogen bond donors and acceptors of G_9CAC_{12} of the loop were involved in hydrogen bonding interactions with the protein residues. However, in the dynamics simulation some of these hydrogen bonds were present for less than one-third of the simulation, so they were not listed in Table 3. Some of the direct hydrogen bonds became water-mediated. The base of U8 did not form any hydrogen bonds with the protein.

Some hydrogen bonds changed their donors or acceptors. For example, the phosphate group of G16 formed a hydrogen bond with the HN atom of Arg^{47} in the x-ray structure, but we found that the phosphate group of G16 formed a hydrogen bond with the HN atom of Leu^{49} with ~63% occupancy.

Some new hydrogen bonds were found from the simulation. In particular, most of the water-mediated hydrogen bonds were new, some of them with more than 70% occupancy, which suggested that these water-mediated bonds contribute to the specific interactions and should not be ignored. It was noted that the phosphate group of C15 underwent two water-mediated hydrogen-bonding interactions with the side chains of Ser^{46} and Ser^{48}, and the phosphate group of U13 also interacted with the side chain of Asp^{92} through a water bridge.

Conformation of the RNA hairpin

As indicated in Fig. 6, *C* and *D*, the loop of the free RNA hairpin was highly flexible and disordered, which was also reflected by the superimposition of the six snapshots taken from the production simulation with an interval of 0.1 ns (Fig. 8 *A*). The double-helix stem of the RNA, as well as the first two nucleotides in the loop, was superimposed very well. Only the sugar-phosphate backbone of nucleotides A1, U7, G16, and U20 exhibited some fluctuation.

Superimposing the average conformations of the free and bound RNA (Fig. 8 *B*) shows that the double-helix stem of the RNA hairpin in the two conformations overlapped quite well. The RMS deviation was just 0.36 Å for all the atoms of the 10 nucleotides in the stem. Both of them were close to the canonical A-RNA conformation. However, the conformation of the stem in complex was closer to the canonical A-RNA than that in free RNA, in which the terminal

FIGURE 7 Interactions of the U1A protein with the RNA hairpin. (*A*) An overall stereo view of the complex. The hydrophobic core is shown in red, the polar loops 1 and 3 in green, and the polar loop 5 in blue. (*B* and *C*) Detailed interaction stereoviews. Residues in the U1A protein are green, nucleotides in the RNA red, hydrogen bonds blue, and the purple spheres represent water molecules.

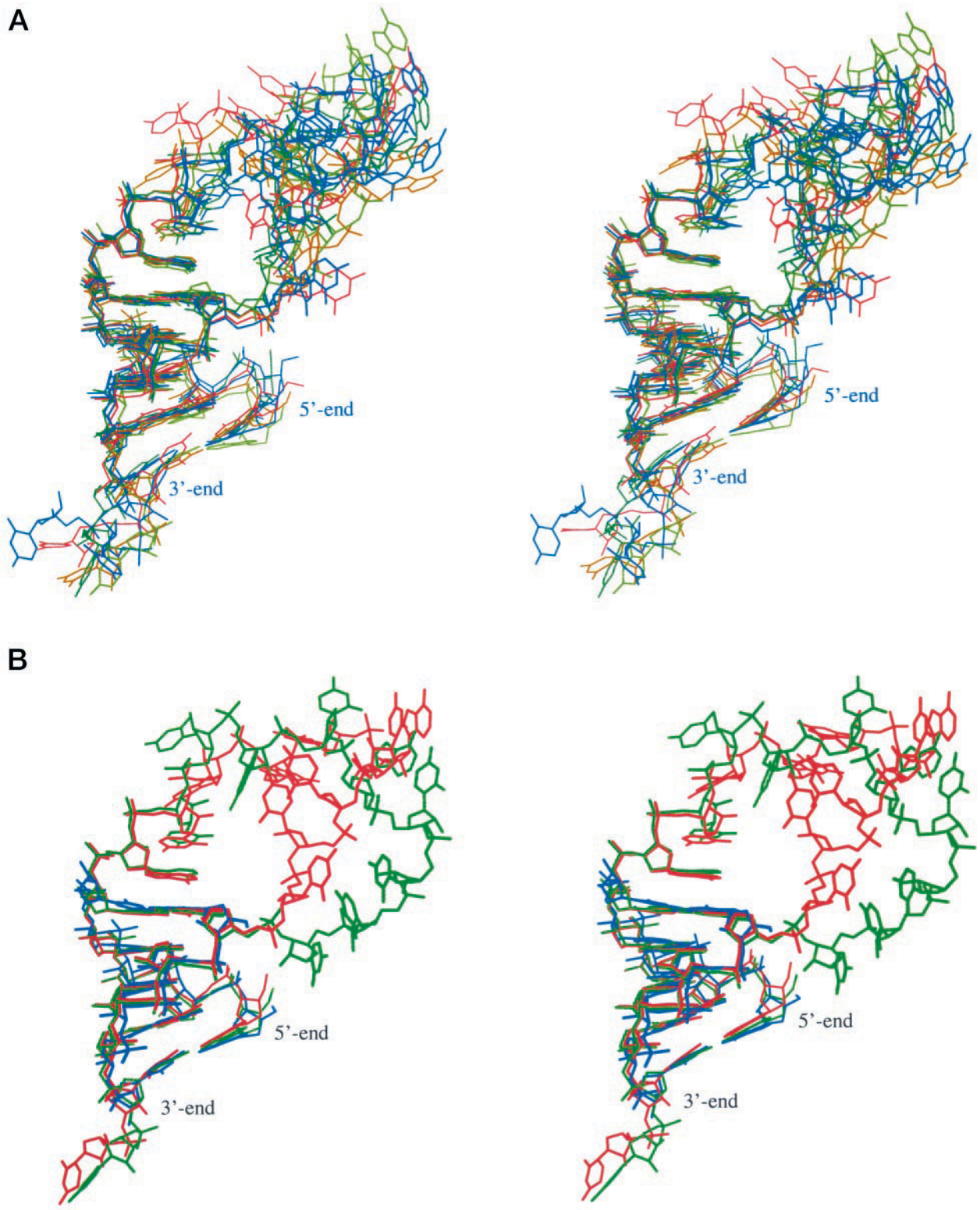


FIGURE 8 Stereo view of superimposition of (*A*) six snapshots of the free RNA hairpin, taken from the production run with an interval of 0.1 ns, colored from red to blue as time evolution. (*B*) The two conformations of the RNA free (*red*) and in complex (*green*), as well as a canonical A-RNA double helix (*blue*) with the same sequence as the hairpin stem.

base pair A1U20 was more mobile. The significant differences between the two conformations were located in the loop region, from U8 to C15, as well as the 3′-end U21. Among the 10 loop nucleotides, A6, U7, G9, C10, U13, and C14 were located on the inside of the loop, whereas U8, A11, C12, and C15 were on the outside, so the space

enclosed by the loop was small. However, the space enclosed by the loop was enlarged in the complex, in which only A6, U7, and G9 remained on the inside of the loop. Nucleotides U13 and C14 had the most deviations in the complex from the free conformation of the RNA.

In the x-ray structure of the U1A-RNA hairpin complex, only nucleotides U8, C10, and U13 were found to have C2′-endo pucker (U13 was not mentioned in Oubridge et al., 1994), and the other ones had C3′-endo pucker. However, this situation changed during the simulations. Besides nucleotides U8, C10, and U13 with C2′-endo pucker in the complex, the sugar puckers of U7 and U21 changed from C3′-endo to C2′-endo spontaneously. The other nucleotides maintained a C3′-endo pucker during the simulation. In the free state, nucleotides C12, C14, C15, and U21 changed their ribose pucker from C3′-endo to C2′-endo spontaneously in the course of the simulation, in addition to U8, C10, and U13 keeping their C2′-endo puckers. However, U7 in the unbound state kept its C3′-endo conformation. Among the 10 nucleotides in the loop region, interestingly, only the three purines, namely A6, G9, and A11, maintained the C3′-endo pucker both in the free state and in the complex, whereas the pyrimidines preferred C2′-endo pucker to C3′-endo.

Interactions maintaining the RNA conformation

Table 4 lists all of the hydrogen-bonding interactions occurring within the RNA hairpin in the two different states, calculated from the corresponding production run with 0.2-ps resolution. From Table 4 *A*, we can see that most of the direct hydrogen bonds present in the free conformation were base pairing interactions in the double-helix stem, with four of these base pairs having almost 100% occupancy and a long average lifetime, which indicated that these base pairs were quite stable during the simulation. However, the ending base pair A1U20 only had about half occupancy for the two hydrogen bonds, indicating that this base pair was less well defined during the simulation. Among the four hydrogen bonds involving the 2′-OH group, which were found only in the loop region, three were formed with their own O3′ atoms with less than 50% occupancy and a very short average lifetime, and the fourth, C12 H2′, was formed with the phosphate oxygen atom O2P of the next nucleotide with more than 50% occupancy and a longer average lifetime. In addition, the 5′-end hydrogen A1 H5T formed a hydrogen bond with A2 O1P.

In contrast to these, more hydrogen bonds were present within the RNA in the complex. Almost all of the hydrogen-bonding interactions between the five base pairs were 100% occupied, including the first one, A1U20. Only the occupancy of hydrogen bond between U3 O4 and A18 H61 was slightly lower. In addition to the three hydrogen bonds between the 2′-OH group and the O3′ atom of the same nucleotide present in the free state, four more such interactions appeared in the three loop nucleotides C12, U13, and C15, and in U19, with about one-third occupancy. Another hydrogen bond was found to connect the bases of U7 and G9. The hydrogen bond connecting the backbone of G12 and U13 in the free state disappeared in the complex. Besides these hydrogen bonds within the RNA, a large number of hydrogen bonds formed between the RNA and the U1A protein have been described above.

In contrast to the normal O(N)-H···O(N) hydrogen-bonding interactions, a number of unusual C-H···O hydrogen bonds (Wahl and Sundaralingam, 1997) were found in the RNA hairpin (Table 4 *B*). Most of these hydrogen bonds were present between the base H6 or H8 atom and the backbone O5′ atom within the same nucleotide with a C3′-endo sugar pucker. Most of these hydrogen bonds were formed in the double-helix stem, such as A2, U3, C4, C5, G17, A18, and U19, both in the free state and in complex, with about half occupancy and a very short average lifetime. In the complex, G16 had a C-H···O hydrogen bond between the H8 atom and O2P instead of the O5′ atom of the backbone, with higher occupancy than the others, showing that the phosphate group of G16 was a little different from the others in the stem. Only a few C-H···O hydrogen bonds were present in the loop nucleotides, except A6 and U7 in the free conformation and A6 in the complex, which were extensions of the double-helix stem. Nevertheless, a C-H···O hydrogen bond was found in C12 of the loop in the complex. Another strange C-H···O hydrogen bond was found to connect the backbones of C5 and A6 in the complex, with about one-third occupancy.

The water-mediated hydrogen bonds were also analyzed from the production run (Table 4 *C*). From Table 4 *C*, we see that in both conformations most of the adjacent phosphate groups were bridged to form hydrogen bonds through water molecules with about half occupancy, especially in the double-helix stem. The 2′-OH group of U13 and the phosphate group of C12, the base of G17 and bases of G16 and A18 formed water-mediated hydrogen bonds, but no bridging water was seen between the backbones of G16 and G17. Nevertheless, differences were obvious between the two states, particularly in the loop region. In the free state, the backbone and base of G9 formed water-mediated hydrogen bonds with the base and backbone of C10, respectively. The phosphate groups of C10 and C12, and C12 and C14 also formed water-mediated hydrogen-bonding interactions. However, these interactions disappeared in the complex, and instead the sugar of U8 and base of G9, bases of C10 and A11, phosphate groups of C10 and A11, C12 and U13 formed water-mediated hydrogen bonds. No water-mediated C-H···O hydrogen bonds were found during the simulations.

Besides the hydrogen-bonding interactions, there were also some base stacking interactions within the RNA hairpin. In addition to the base stacking interactions in the double-helix stem, the bases of A6 and U7, G9 and C10, A11 and C12, and U13 and C14 were stacked separately on each other in pairs in the free state. In the complex, this situation was changed. A6 and U7 kept the stacking inter-

TABLE 3 Hydrogen bonds at the RNA-protein interface (with occupancy ≥ 33.3%)

Nucleotide atoms · · · Amino acid atoms	Occupancy (%)	Average lifetime (ps)		
		a	b	c
(A) Direct hydrogen bonds				
A6 N1 · · · Arg^{52} HH12*	96.1	6.0	9.5	28.8
U7 H3 · · · Glu^{19} OE1*	82.2	7.6	31.9	105.7
U7 H3 · · · Glu^{19} OE2	41.4	0.6	7.0	42.0
G9 H21 · · · Glu^{19} OE2*	51.7	2.2	3.8	15.3
G9 O6 · · · Asn^{16} HN*	87.0	2.5	5.6	26.2
C10 H41 · · · Tyr^{86} O*	96.6	7.9	9.8	24.7
C10 H42 · · · Gln^{85} OE1*	67.1	1.4	4.0	19.1
C10 O2 · · · Lys^{88} HN	62.2	0.9	2.3	8.1
A11 H62 · · · Thr^{89} OG1*	79.7	1.3	2.5	12.2
A11 N1 · · · Ser^{91} HG1*	100.0	199.9	199.9	199.5
C12 H41 · · · Asp^{90} O*	92.9	3.1	4.2	18.9
C12 N3 · · · Asp^{92} HN*	70.8	0.8	2.0	9.4
C12 O2 · · · Asp^{92} HN	54.4	0.6	1.7	6.8
G16 N7 · · · Arg^{52} HH11*	88.8	2.2	3.8	14.9
G16 O1P · · · Leu^{49} HN	63.0	1.9	4.9	16.3
(B) Water-mediated hydrogen bonds				
A6 H61 · · · w · · · Glu^{19} OE1	77.4	4.5	7.9	19.5
G9 H1 · · · w · · · Leu_{17} O	52.6	2.0	4.4	18.1
G9 H1 · · · w · · · Glu_{19} OE2	79.1	2.1	4.6	17.4
U7 O4 · · · w · · · Glu^{19} OE1	52.0	0.9	2.2	7.8
G9 H21 · · · w · · · Glu^{19} OE2	34.3	0.7	2.9	13.0
C10 O1P · · · w · · · Tyr^{13} OH	65.4	1.4	3.2	15.7
C10 O2 · · · w · · · Thr^{89}HN*	43.6	1.2	3.4	17.5
A11 N7 · · · w · · · Thr^{89}HN*	38.6	0.9	2.4	12.0
C10 O1P · · · w · · · Lys^{88} HZ2	36.8	1.6	3.5	7.5
C12 H42 · · · w · · · Thr^{89} HG1	76.9	1.2	2.7	11.2
C12 O2 · · · w · · · Ile^{93} HN*	78.4	1.9	3.6	14.6
C15 O1P · · · w · · · Ser_{46} HG1	76.5	2.5	5.1	17.5
C15 O1P · · · w · · · Ser^{48} HG1	64.4	1.2	3.1	10.4
U13 O1P · · · w · · · Asp^{92} OD1	33.4	2.2	3.8	9.3
C15 O2P · · · w · · · Ser^{46} OG	37.6	0.8	2.7	9.0

*The hydrogen bond is present in the x-ray structure. a: time cutoff 0.1 ps, resolution 0.2 ps; b: time cutoff 1.0 ps, resolution 0.2 ps; c: time cutoff 5.0 ps, resolution 1.0 ps.

action together with the stem, the bases of C10, A11, and C12 underwent partially stacking interactions, and U13, C14, and C15 were also partially stacked. More importantly, the bases of G_9CAC_{12} were buried in the hydrophobic core of the protein and formed stacking interactions with the protein separately.

Counterions and hydration

The distributions of water molecules and counterions around the complex during the simulation are shown in Fig. 9 *A*, according to the probability density map. Fig. 9 *A* illustrates that most of the water molecules with large probability were distributed uniformly around the folded domain of the protein and part of the loop of the RNA, whereas the water molecules around the two terminals of the protein and the RNA double-helix stem were more randomly distributed with low probability density. During the simulation the most probable positions of the sodium ions were close to the major groove of the RNA stem and outside of the protein C-terminal helix. The former region was the negative electrostatic region, and the latter one had two negative residues, Asp^{90} and Asp^{92}.

For the free RNA hairpin, the sodium counterions were initially placed 3.6 Å from the phosphorous atoms. In the course of time, the counterions gradually moved away from the RNA atoms. In particular, four ions were detained in the major groove of the hairpin stem and stayed within 6.0 Å of the phosphate groups, as seen from the most probable distribution region of the ions shown in Fig. 9 *B*. The probability densities of the counterions were higher (up to 45%) in the free state than in the complex (<30%). Fig. 9 *B* also shows the probability density map of water molecules around the free RNA, in which water molecules with a probability of more than 40% were distributed around the hairpin stem.

TABLE 4 Hydrogen bonds within the RNA hairpin (with occupancy ≥ 33.3%)

Hydrogen bonds	Free RNA		RNA in complex	
	Occupancy (%)	Average lifetime (ps)	Occupancy (%)	Average lifetime (ps)
(A) Normal direct hydrogen bonds				
A1 H5T · · · A2 O1P	44.0	2.2	34.4	1.4
A1 H61 · · · U20 O4	66.6	9.5	97.6	12.6
A1 N1 · · · U20 H3	51.1	4.4	91.6	5.8
A2 H61 · · · U19 O4	96.4	6.2	95.2	5.1
A2 N1 · · · U19 H3	99.6	54.3	99.9	99.9
U3 H3 · · · A18 N1	99.9	119.8	99.8	99.9
U3 O4 · · · A18 H61	96.4	6.3	87.0	2.3
C4 H41 · · · G17 O6	97.2	7.9	97.0	11.4
C4 N3 · · · G17 H1	99.9	199.8	99.0	44.0
C4 O2 · · · G17 H21	99.6	49.8	99.7	57.0
C5 H41 · · · G16 O6	99.6	49.8	99.3	28.4
C5 N3 · · · G16 H1	99.9	199.9	99.9	99.9
C5 O2 · · · G16 H21	98.9	18.5	99.0	18.9
U7 H2′ · · · U7 O3′	35.4	0.3	33.4	0.4
G9 H2′ · · · G9 O3′	45.5	0.4	44.6	0.4
A11 H2′ · · · A11 O3′	37.1	0.4	46.0	0.4
C12 H2′ · · · U13 O2P	58.5	3.7		
U7 O2 · · · G9 H22			53.2	1.1
C12 H2′ · · · C12 O3′			34.2	0.3
U13 H2′ · · · U13 O3′			38.0	0.4
C15 H2′ · · · C15 O3′			44.0	0.4
U19 H2′ · · · U19 O3′			33.8	0.3
(B) Unusual direct C—H · · · O hydrogen bonds				
A2 H8 · · · A2 O5′	49.6	0.5	45.2	0.5
U3 H6 · · · U3 O5′	52.6	0.5	53.4	0.5
C4 H6 · · · C4 O5′	48.1	0.5	54.4	0.5
C5 H6 · · · C5 O5′	52.9	0.5	45.3	0.4
A6 H8 · · · A6 O5′	58.5	0.5	44.8	0.4
G17 H8 · · · G17 O5′	59.4	0.6	56.4	0.5
A18 H8 · · · A18 O5′	48.2	0.5	56.9	0.5
U19 H6 · · · U19 O5′	39.9	0.5	45.2	0.5
U7 H6 · · · U7 O5′	63.6	0.6		
C5 H2″ · · · A6 O4′			36.4	0.4
C12 H6 · · · C12 O5′			54.6	0.5
G16 H8 · · · G16 O2P			64.2	0.9

In both states of the RNA hairpin, most of the hydrogen-bonding donors and acceptors formed hydrogen bonds with water molecules during the simulations, indicating the high degree of hydration of the RNA hairpin. The water-mediated hydrogen-bonding interactions (listed in Table 4 *C*) were only a part of the hydration. From the production runs, we found that almost all of the phosphate group oxygen atoms were hydrated with three water molecules in both states, except that in the complex O1P of C15 and both O1P and O2P of G16 formed no more than two hydrogen bonds to water molecules. For the sugar and base atoms, the hydration degrees are listed in Table 5, in which almost all atoms in the free RNA have at least one hydrogen bond to water 50% of the time. Apart from a few differences, most of the atoms in the double-helix stem had very similar hydrations in both states. All O_2' atoms had occupancies of ~140%; that is, they were hydrated by about one or two water molecules. On the contrary, for the hairpin loop atoms, the bases of G9, C10, A11, and C12 had high occupancy for the free RNA, but low occupancy in the complex, because of the hydrophobic interactions with the protein.

The interaction energies between the complex and the solvent are shown in Fig. 10. Both the protein and the RNA had very small van der Waals interaction energies with water molecules, so most of interactions between the solute and solvent came from electrostatics and remained stable during the simulation. The protein had lower electrostatic interaction energy with water molecules than the RNA hairpin. Both of these solute-solvent interaction energies were much lower than the RNA-protein interaction energy (Fig. 3 *A*).

Comparison with the NMR structure of U1A-3′UTR complex

The complex of U1A protein with 3′UTR of its own pre-mRNA was determined by NMR spectroscopy (Allain et al.,

TABLE 4—Continued

Hydrogen bonds	Free RNA		RNA in complex	
	Occupancy (%)	Average lifetime (ps)	Occupancy (%)	Average lifetime (ps)
(C) Water-mediated hydrogen bonds				
A2 O1P · · · w · · · U3 O1P	54.8	1.4	46.2	0.9
A2 O2P · · · w · · · U3 O1P	58.9	2.7	35.0	2.4
U3 O1P · · · w · · · C4 O1P	46.9	0.8	51.1	0.8
U3 O5′ · · · w · · · C4 O1P	35.0	0.5	46.8	0.6
C4 O5′ · · · w · · · C5 O1P	46.0	0.6	45.3	0.7
C5 O5′ · · · w · · · A6 O1P	56.4	0.7	41.6	0.6
C5 O2P · · · w · · · A6 O1P	44.1	1.5	48.4	1.8
A6 O2P · · · w · · · U7 O1P	78.8	5.6	86.4	5.6
U7 O2P · · · w · · · U8 O1P	64.2	2.6	55.2	4.0
U8 O2P · · · w · · · G9 O2P	48.5	4.6	38.1	4.1
G9 O1P · · · w · · · C10 O1P	61.4	1.9	47.2	1.3
A11 O5′ · · · w · · · C12 O1P	60.3	1.4	63.8	0.8
C12 O2P · · · w · · · U13 O2′	34.4	1.7	77.7	5.4
G16 N7 · · · w · · · G17 N7	49.1	1.2	45.5	0.8
G17 N7 · · · w · · · A18 H62	42.0	0.6	50.4	0.7
G17 O2P · · · w · · · A18 O1P	49.0	2.4	56.8	2.5
A18 O2P · · · w · · · U19 O1P	52.1	2.9	55.0	2.2
U3 O2P · · · w · · · C4 O1P	41.3	1.7		
C5 O1P · · · w · · · A6 O1P	46.8	0.6		
G9 O1P · · · w · · · C10 H42	44.8	0.8		
G9 H22 · · · w · · · C10 O4′	38.3	0.6		
C10 O2P · · · w · · · C12 O1P	50.6	3.1		
C10 O2P · · · w · · · C12 O2P	40.3	3.5		
C12 O1P · · · w · · · U13 O2′	43.7	3.1		
C12 O2P · · · w · · · C14 O1P	82.9	5.1		
C15 O1P · · · w · · · G16 O1P	45.6	6.2		
A6 O1P · · · w · · · U7 O1P			35.2	0.8
U8 O3′ · · · w · · · G9 N7			39.0	0.8
C10 O2 · · · w · · · A11 N7			74.2	1.0
C10 O2P · · · w · · · A11 O2P			87.3	8.3
A11 N3 · · · w · · · A11 O2′			40.6	0.6
C12 O1P · · · w · · · U13 O2P			52.5	1.4
C12 O2P · · · w · · · U13 O2P			55.4	1.9
A18 O5′ · · · w · · · U19 O1P			34.4	0.6
U19 O2P · · · w · · · U20 O1P			49.4	1.8
U20 O2P · · · w · · · U21 O4′			35.8	1.3

Time cutoff 0.1 ps, resolution 0.2 ps.

1996), and detailed interactions between the protein and the RNA have been described based on this NMR structure (Allain et al., 1997).

From a superposition of the two complexes (Fig. 11), it was obvious that some parts of the complexes were superimposed very well, especially the U1A protein, with only 1.0 Å of RMSD for the backbone atoms from residues 10 to 90. For the RNA, the C5G16 base pair and $A_6UUGCAC_{12}$ in the hairpin overlapped well with the C38G25 base pair and $A_{39}UUGCAC_{45}$ of 3′UTR, except that the base of U8 deviated slightly; hence we might anticipate that the corresponding nucleotide (U41) in the U1A − 3′UTR complex is also flexible. C15 of the hairpin occupied a position similar to that of A24 of 3′UTR. These common residues were just the main sites of binding of the two RNA molecules to the same protein. Comparing our Tables 1–3, respectively, with tables IV, II, and III in Allain et al. (1997), it can easily be seen that the electrostatic, hydrophobic, and hydrogen-bonding interactions undergone by these nucleotides and the protein were very similar. In particular, most of the base–side chain stacking and base–side chain hydrogen-bonding interactions were the same.

However, some differences were also found. One difference was that the base of C45 underwent a stacking interaction with the side chain of Asp^{92} in the NMR structure, just like the case in the x-ray structure, but in our simulation, the corresponding nucleotide C12 did not undergo a stacking interaction with Asp^{92}. Instead it made a local hydrophobic contact with the CB atom of Asp^{92}. U41 in 3′UTR formed one hydrogen bond with Glu^{19}, but the corresponding U8 had no obvious interactions with the protein in the simulation. Besides the common residues, three other nucleotides took part in electrostatic interactions in the stem of 3′UTR; so did three other nucleotides in the hairpin stem, but in different positions. The base of A24 of 3′UTR underwent

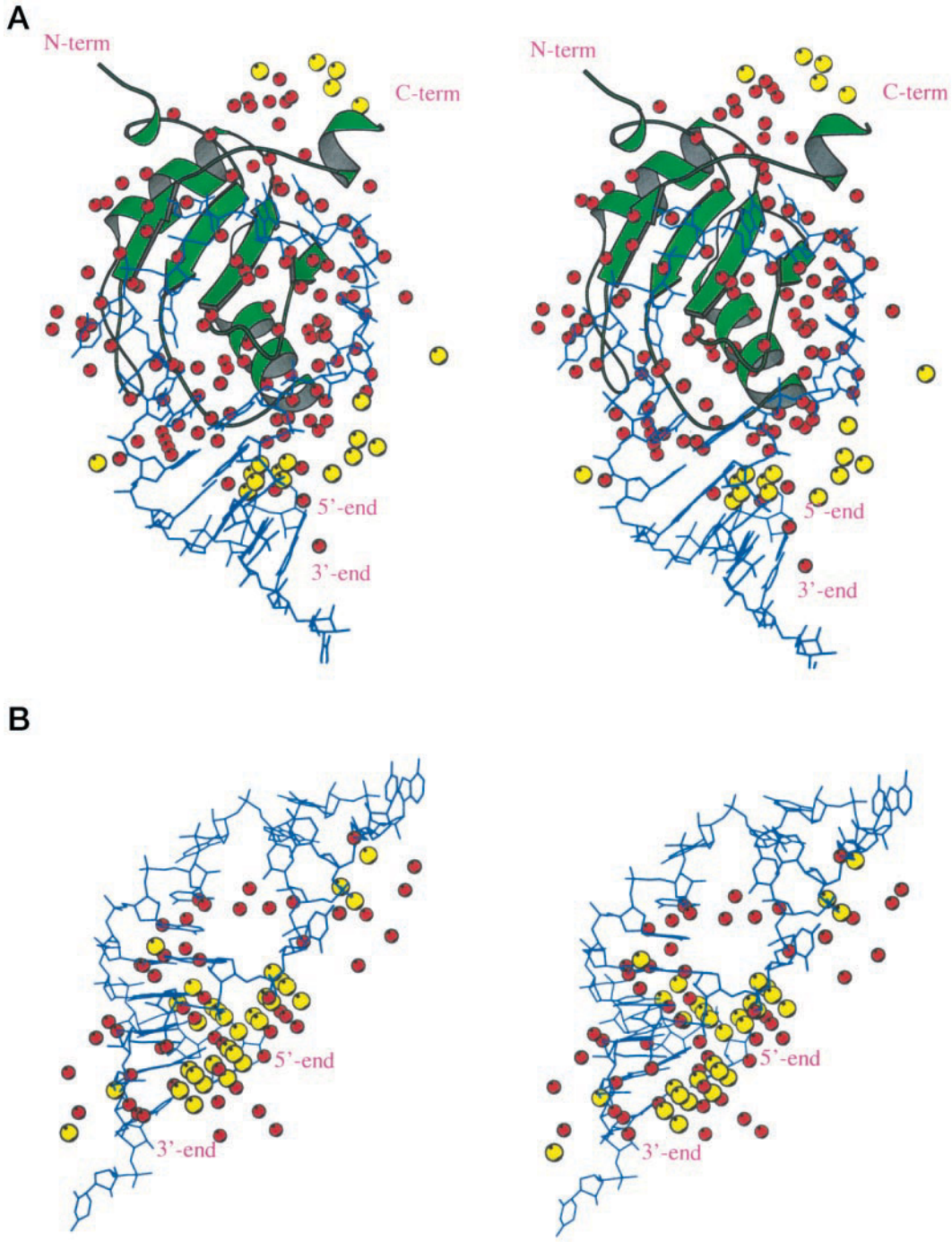


FIGURE 9 Stereo view of the sodium ions (*yellow spheres*) and water molecules (*red spheres*): distribution around (*A*) the RNA-protein complex and (*B*) the free RNA, according to the probability density map.

interactions with the protein, but the corresponding base of C15 did not.

DISCUSSION

Physical basis of the U1A-RNA interaction

As suggested by the x-ray structure (Oubridge et al., 1994) and the simulation, electrostatic recognition is strongly indicated to be the basis of the U1A-RNA interactions. There were several electrostatic interactions at the U1A-RNA interface (Table 1), and electrostatic energy was the dominant source of the RNA-protein interaction energy during the simulation (Fig. 3 *A*). Hence the electrostatic interactions contributed substantially to the binding affinity. The electrostatic potential study also demonstrated that a region with positive electrostatic potential was present at the protein

TABLE 5 Hydrogen-bonding occupancy (≥50%) with water for each donor or acceptor atom of RNA sugar and base

		Occupancy				Occupancy				Occupancy	
Nucleotide	Atom	Free	Complex	Nucleotide	Atom	Free	Complex	Nucleotide	Atom	Free	Complex
A1	O5′	146.3	79.6	G9	O2′	129.4	114.8	C15	O4′	99.9	54.4
	O2′	134.7	36.8		N3	81.2	70.4		O2′	145.8	140.1
	H2′	52.4	54.4		H1	92.6	89.3		O2	205.9	200.6
	N7	92.9	93.9		H21	84.6			N3	113.8	115.8
	N1	77.1			H22	85.4			H41	66.6	66.8
	N3	76.7	72.6		O6	141.3	64.0		H42		63.2
	H62	79.1	80.2		N7	171.7	91.0	G16	O5′	68.7	84.0
A2	O4′	52.3		C10	O4′	66.5			O4′	62.7	
	O2′	132.3	136.6		O2′	148.2	51.0		O2′	143.1	140.4
	N7	64.4	72.3		H2′		63.9		H2′	53.8	56.1
	N3	84.1	78.0		O2	188.3	95.8		N3	77.1	79.8
	H62	77.5	84.2		N3	98.9			H22	79.9	79.6
U3	O4′	58.9	54.6		H41	69.5			O6	91.1	73.8
	O2′	144.5	138.9		H42	77.6			N7	181.1	76.4
	O2	88.5	80.4	A11	O5′	64.7	74.4	G17	O4′	52.7	
	O4	71.5			O4′	59.8			O2′	143.7	139.8
C4	O5′	51.4	50.4		O2′	138.4	141.6		H2′	55.1	52.6
	O2′	137.9	137.6		N7	78.7	82.6		N3	65.3	70.4
	O2	68.1	59.4		N1	103.5			H22	81.0	79.2
	H42	77.8	82.8		N3	78.7			O6	85.4	84.0
C5	O5′	57.9			H61	56.4			N7	153.5	144.8
	O4′		53.0		H62	68.8		A18	O4′	52.0	
	O2′	135.8	138.2	C12	O4′	65.9	62.2		O2′	141.8	135.0
	H2′		57.2		O3′	56.3			H2′	51.0	50.5
	O2	61.4	53.0		O2′	136.3	142.8		N7	77.2	66.4
	H42	76.8	81.0		H2′		51.5		N3	78.7	73.9
A6	O2′	141.2	142.6		O2	203.9	103.0		H62	83.9	87.0
	H2′	54.8	61.9		N3	105.0		U19	O4′	59.4	83.5
	N7		54.0		H41	61.7			O2′	132.4	136.7
	N1	119.5			H42	74.4	87.3		O2	85.9	84.4
	N3	71.1	50.4	U13	O5′		52.8		O4	95.8	83.5
	H61	83.9	80.3		O4′	97.1	113.4	U20	O4′	65.4	54.5
	H62	81.8	77.3		O2′	114.9	119.8		O2′	144.6	140.8
U7	O4′		77.1		O2	86.0	110.0		H2′	54.1	54.2
	O2′	149.9	118.2		H3	95.9	95.8		O2	118.3	97.4
	H2′	51.3			O4	153.7	151.2		O4	86.7	82.8
	H3	96.8		C14	O5′	78.9	96.3	U21	O5′	53.7	54.7
	O4		171.2		O4′		68.8		O4′	80.6	84.2
U8	O4′	90.9	101.4		O2′	147.6	123.1		O3′	117.0	112.0
	O2′	154.5	152.2		H2′	50.6			O2′	130.4	137.6
	H2′	60.8	59.8		O2	202.4	162.6		H2′	81.1	74.5
	O2	134.9	130.4		N3	106.7	89.0		H3T	68.1	62.8
	H3	95.3	94.3		H41	69.7	72.7		O2	128.2	130.2
	O4	97.4	153.4		H42	66.6	79.0		H3	95.2	93.2
G9	O4′	55.2		C15	O5′	83.7	101.9		O4	149.8	141.4

solvent-accessible surface, whereas the major groove surface of the RNA double-helix stem had a negative electrostatic potential (Fig. 4). The counterions were present in the major groove of the RNA stem, with high probability density during the simulation, which also confirmed the negative electrostatic character of the major groove. The RNA stem faced the protein through the major groove and remained stable, possibly induced by the electrostatic interactions.

From the simulation, we observed a conformational transition in the U1A protein structure at 450–500 ps. Before this period, the structural deviations occurred in loop 3 and the end of helix B; after that, the structural deviations moved to loop 1 and the beginning of helix B (Fig. 5), although the whole folded domain remained stable. Thus the RMSD value of the protein underwent a small increase at 450–500 ps, leading to a slight increase in the solvent-accessible surface area of the protein (Fig. 3 *B*). However, during this period the RNA remained stable, which indicated that the conformational transition was independent on RNA. During the transient state, electrostatic interactions were more favorable (Fig. 3 *A*), suggesting that the interactions with the RNA were better for this conformation of the protein.

The buried solvent-accessible surface area remained constant during the simulation (Fig. 3 *B*), which meant that the hydrophobic interactions at the RNA-protein interface were stable. The stacking interactions between the bases and the

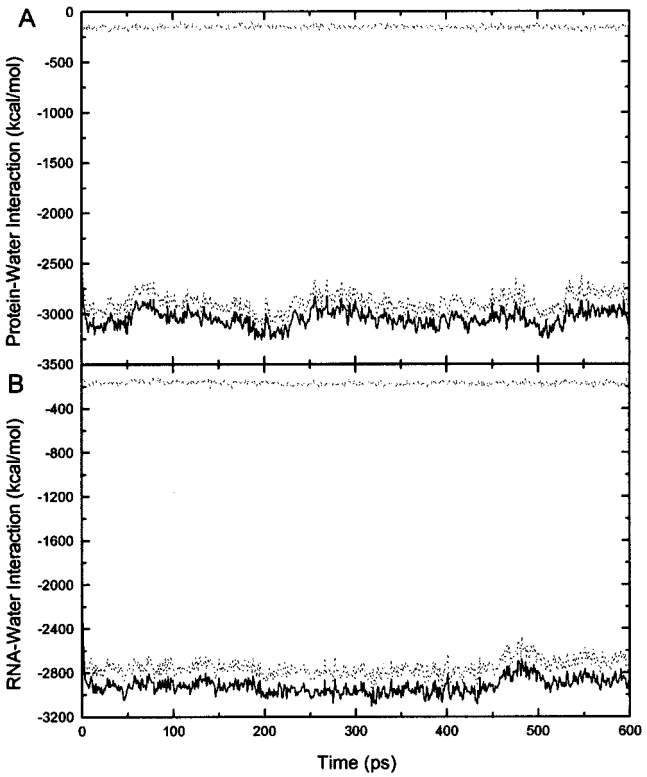


FIGURE 10 Time-dependent (*A*) protein-water and (*B*) RNA-water nonbonded interaction energies. *Solid line*: total energy; *dashed line*: van der Waals energy; *dotted line*: electrostatic energy.

side chains and other hydrophobic interactions mainly contributed to the binding affinity and not to the specificity.

Loops 1, 3, and 5 of the U1A protein were important to the specificity of the RNA-U1A interactions. Except for the two termini, the whole protein was very stable during the simulation; these three loops in particular were unexpectedly stable, with only a few deviations from the x-ray structure. All three loops included many charged and polar residues and were involved in a number of hydrogen-bonding and electrostatic interactions with the RNA hairpin, especially Glu^{19} and Arg^{52}. Residue Glu^{19} of loop 1 formed several base–side chain hydrogen bonds with nucleotides A6, U7, and G9, indicating that it was very important to the specificity. However, Glu^{19} changed its conformation during the simulation, still keeping in touch with these nucleotides, suggesting that the mobile hydrogen-bonding interactions are important for the specific RNA-U1A interactions. On the other hand, residue Arg^{52} of loop 3 was quite stable during the simulation, forming two base–side chain hydrogen bonds with A6 and G16 with a high occupancy of ~90%, so it was important to the specificity of RNA-U1A interactions. In contrast to the x-ray structure of the complex, Asp^{92} did not form stacking interactions with C12; instead Asp^{92} extended farther into the solvent and was highly hydrated.

Among the seven conserved nucleotides, $A_6UUGCAC_{12}$, of the loop in the RNA hairpin, only G_9CAC_{12} were involved in all of the electrostatic, hydrophobic, and hydrogen-bonding interactions, whereas A6, U7, and G16 were mainly included in hydrogen-bonding interactions and U8 underwent no apparent interactions with the protein. Therefore, we thought that G_9CAC_{12} contributed not only to the high binding affinity, but also to the high specificity of the RNA hairpin to the U1A protein, whereas A6, U7, and G16 were primarily responsible for the high specificity.

Conformational changes in the RNA hairpin

Most of conformational changes in the RNA hairpin in response to the binding of the U1A protein were in the loop region rather than in the double-helix stem, in agreement with Hall's observation (Hall, 1994). From the analysis of the results, it can be seen that the binding of the protein not only ordered the random loop nucleotides, but also made the double-helix stem closer to canonical A-RNA, although the action on the stem was weak because of the absence of direct RNA-protein contacts there.

The reason for the ordering of the stem should be attributed to electrostatic interactions with the U1A protein. During the RNA-protein interaction process, several positively charged residues of the protein, such as Lys^{20}, Lys^{22}, and Arg^{47}, formed electrostatic interactions with the phosphate groups of A2, U3, and C4, thus stabilizing the double-helix stem.

The ordering of the loop could originate from a wide range of interactions upon binding to the protein. Most of the intramolecular hydrogen-bonding and base-stacking interactions in the free conformation remained, and more hydrogen bonds were added in the loop region. A series of hydrogen-bonding, electrostatic, and hydrophobic interactions also took place at the RNA-protein interface, mainly in G_9CAC_{12}, making the RNA conformation more stable, even in the loop region.

The ordering of the conformation of the RNA reduces the entropy of the system, as does the complex formation in itself, offsetting the entropy increase due to release of water molecules and ions from the RNA and protein, as seen in the RNA loop, where the hydration degree of the nucleotides decreased because of hydrophobic interactions with the protein.

In the x-ray structure nucleotide U8, the third of the seven conserved nucleotides AUUGCAC in the RNA loop, is included in extensive interactions with the protein, just like the other six nucleotides (Oubridge et al., 1994). However, our simulation demonstrated that this nucleotide was unexpectedly flexible, which is very similar to the behavior of the last three nucleotides UCC in the loop but different from the other six nucleotides, resulting in no obvious contacts with the protein. The seven conserved nucleotides AUUGCAC of the loop constitute the site of binding of the RNA hairpin to the U1A protein. Among the seven nucleotides, A6 and U7 were continuously stacked with the stem base pairs, whereas G9, C10, A11, and C12 were buried in the

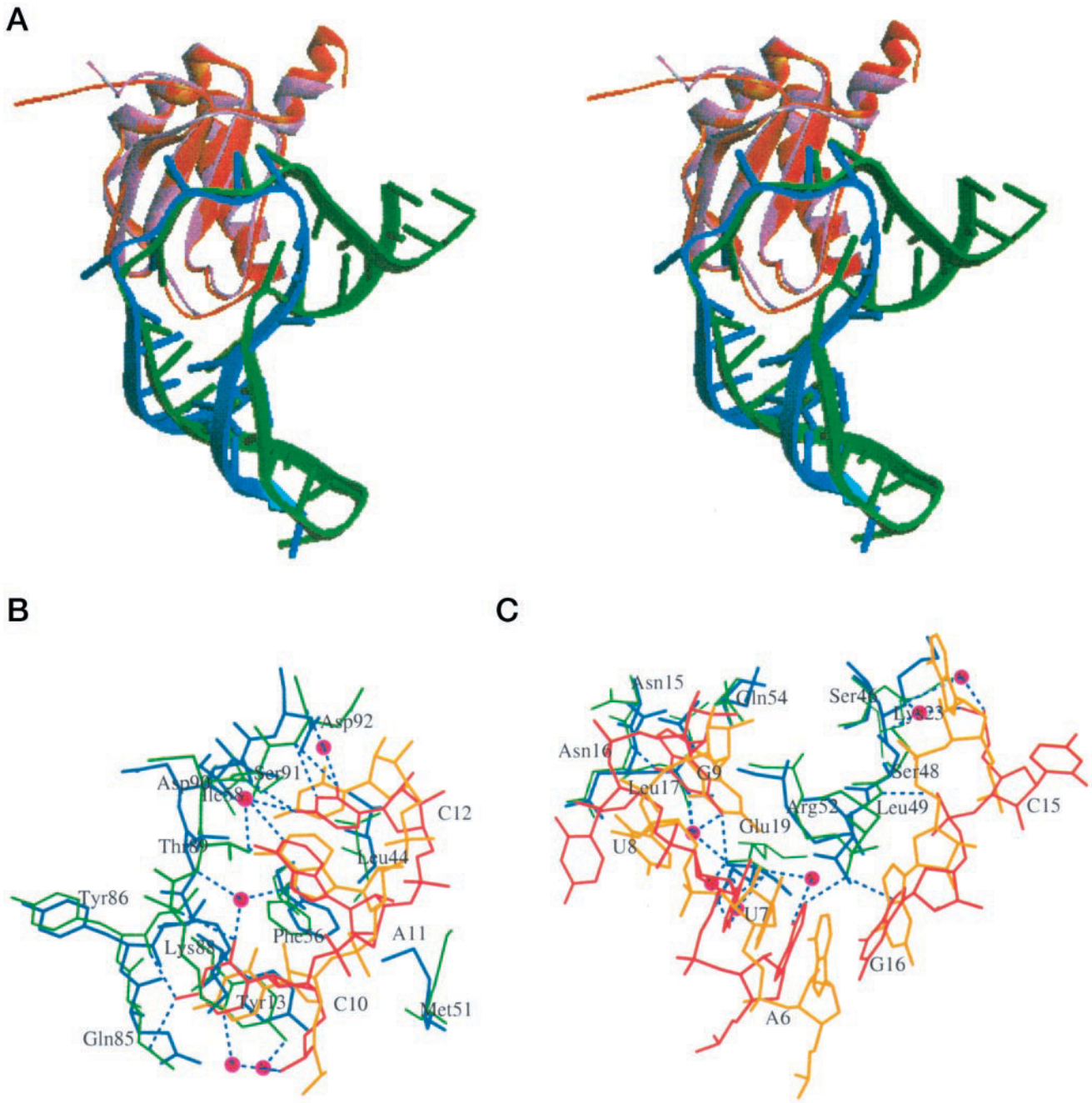


FIGURE 11 Comparison of the dynamic average structure with the NMR structure of U1A protein complexed with the 3′-untranslated region of U1A pre-mRNA. The latter was taken from the Protein Data Bank (Bernstein et al., 1977), PDB code 1AUD, model 15. (*A*) Stereo view of the superimposition of the two complex structures. The NMR structure is shown in red and green. (*B* and *C*) Comparison of the common interaction details. The dynamic average structure is shown in green and red.

hydrophobic core of the protein. These six nucleotides were restrained and quite stable. Only U8 was freely mobile on the outside of the protein surface. Even if in the crystal structure the base of U8 forms a partial stacking interaction with the base of G9 and has three water-mediated and direct hydrogen bonds with the protein residues, these interactions were too weak to restrain U8 in the simulation. In the MD simulation of the free RNA hairpin, U8 was also quite flexible, as were the other nucleotides from G9 to C15 of the loop. Therefore, it may be that the flexibility of U8 fits a biological need and would be beneficial to the insertion of nucleotides G_9CAC_{12} into the hydrophobic core of the protein. The changed sugar conformation of U7 from 3′-endo to 2′-endo during the simulation could be connected to the high mobility of U8.

Similarly, nucleotides U13, C14, and C15, located on the other side of the four buried ones in the loop, also appeared to be highly mobile and had no evident contacts with the

protein, which should be advantageous to the insertion of G_9CAC_{12}. The poor order of these nucleotides in the x-ray structure confirmed their mobility. However, from the simulation the backbone of C15 was detected to undergo four interactions with the protein, namely one electrostatic interaction with the side chain of Lys^{23}, one hydrophobic interaction with the Cβ atom of Ser^{48}, and two water-mediated hydrogen bonds with the side chains of Ser^{46} and Ser^{48}. The backbone of U13 also formed a weak water-mediated hydrogen bond with the protein. Obviously, these interactions were the extension of the interactions between G16 and the protein.

During the simulations purines kept the C3′-endo pucker and pyrimidines preferred C2′-endo pucker.

Suggestion for U1A protein specifically binding to RNA

How the U1A protein sequence-specifically binds to RNA substrate has long been studied by x-ray crystallography and NMR spectroscopy. Here we give a molecular dynamics view of the specific basis of U1A binding to RNA substrates.

Experiments have shown that the free U1A protein has a packed conformation, with helix C covering the hydrophobic surface of the β-sheet (Avis et al., 1996). The free RNA hairpin is highly flexible in the loop region (mentioned in Oubridge et al., 1994; simulated here). After the initial complex formation, helix C of the protein moves up to uncover the hydrophobic surface, allowing the loop of the RNA hairpin to take its place, while loop 3 of the protein also protrudes through the loop of the RNA (Oubridge et al., 1994; Allain et al., 1996). This is the so-called induced-fit mechanism.

Compared to the NMR structure of U1A protein complexed with 3′UTR of its own pre-mRNA (Allain et al., 1996), which is also confirmed to adopt an induced-fit mechanism, there are very similar intermolecular interactions (Allain et al., 1997) between the two complexes. It is reasonable to expect that there is a common interaction mechanism for the two RNA substrates binding to the U1A protein, namely, a common structural basis for the U1A protein binding to its RNA substrates (Jovine et al., 1996).

Based on the knowledge of the simulation and experiments, we suggest such a common interaction mechanism for the two RNA substrates binding to the U1A protein. The mechanism includes three steps. The first is electrostatic recognition, to make RNA and U1A approach each other in the correct position and orientation. Then, RNA is anchored at residue Arg^{52}, on the edge of the positively charged region of the protein, by forming two base–side chain hydrogen-bonding interactions with A6 and G16, respectively. Meanwhile, the negatively charged Glu^{19} also forms a base–side chain hydrogen bond with U7. Third, the protein changes its conformation in response to the approach of RNA, followed by conformational changes of RNA. That is to say, loop 3, containing the crucial residue Arg^{52}, protrudes through the loop of RNA to force the helix C to recover the hydrophobic core surface, and the loop of RNA takes the place of helix C to cover the hydrophobic surface of the protein. Then a series of specific interactions occurs between their interfaces, the bases of G_9CAC_{12}, and some side chains of the protein.

This suggestion is supported by the following circumstances. At first, the above-mentioned electrostatic potentials showed the electrostatic recognition to be the basis of U1A-RNA interaction. Meanwhile, several other important positively charged residues in the free protein, including Lys^{50}, Lys^{80}, and Lys^{98}, were located around the positive region, with their long side chains pointing toward the outside. These residues have no direct contacts with RNA after binding, but mutation of any of these residues reduces the binding affinity dramatically (Nagai et al., 1990). So these residues must be important in the first step of RNA recognizing the protein. In the second step, the RNA would anchor to the surface through hydrogen bonds from A6 and G16 to Arg^{52}. A6 and G16 are located in the middle of the RNA hairpin and at the edge of the negative charge region. To anchor A6 and G16 to the protein before any of the other nucleotides would help to keep the position and orientation of the loop stable. Arg^{52} was very stable during the simulation; both the backbone and side chain had very low RMSD and RMSF values (see Fig. 6). Arg^{52} also formed a salt bridge with the phosphate group of G16. It is known experimentally that the specific RNA binding is abolished in the $Arg^{52}\rightarrow$Gln mutant (Nagai et al., 1990). After RNA anchored to the surface of the protein, nucleotides U7 and C15 were also fixed at their positions. However, U8 and U13 stayed flexible, which is important to the insertion of G_9CAC_{12}, while the protein changed its conformation.

The conformational changes in helix C and loop 3 of the protein occur in response to the correct recognition of the RNA.

This work was supported by the Swedish Natural Science Research Council.

REFERENCES

Allain, F. H.-T., C. C. Gubser, P. W. A. Howe, K. Nagai, D. Neuhaus, and G. Varani. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature.* 380:646–650.

Allain, F. H.-T., P. W. A. Howe, D. Neuhaus, and G. Varani. 1997. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* 16:5764–5774.

Arnez, J. G., and J. Cavarelli. 1997. Structures of RNA-binding proteins. *Q. Rev. Biophys.* 30:195–240.

Auffinger, P., S. Louise-May, and E. Westhof. 1999. Molecular dynamics simulations of solvated yeast $tRNA^{Asp}$. *Biophys. J.* 76:50–64.

Auffinger, P., and E. Westhof. 1997. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated $tRNA^{Asp}$ anticodon hairpin. *J. Mol. Biol.* 269:326–341.

Auffinger, P., and E. Westhof. 1998. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* 8:227–236.

Avis, J. M., F. H.-T. Allain, P. W. A. Howe, G. Varani, K. Nagai, and D. Neuhaus. 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, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Bernstein, F. C., T. F. Koetzle, G. J. B. Willians, E. F. Mayer, J. M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Boelens, W. C., E. J. Jansen, W. J. van Venrooij, R. Stripecke, I. W. Mattaj, and S. I. Gunderson. 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell.* 72:881–892.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Brünger, A., and M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins.* 4:148–156.

Burd, C. G., and G. Dreyfuss. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science.* 265:615–621.

Ceccarelli, M., and M. Marchi. 1997. Simulation of a protein crystal at constant pressure. *J. Phys. Chem. B.* 101:2105–2108.

Cheatham, T. E., III, and P. A. Kollman. 1997. Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* 119: 4805–4825.

Draper, D. E. 1995. Protein-RNA recognition. *Annu. Rev. Biochem.* 64: 593–620.

Eriksson, M. A. L., T. Härd, and L. Nilsson. 1995. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402–426.

Frankel, A. D., and C. A. Smith. 1998. Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell.* 92: 149–151.

Gubser, C. C., and G. Varani. 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, K. B. 1994. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry.* 33:10076–10088.

Hermann, T., P. Auffinger, and E. Westhof. 1998. Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* 27: 153–165.

Holbrook, S. R., and S.-H. Kim. 1997. RNA crystallography. *Biopolymers.* 44:3–21.

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

Jovine, L., C. Oubridge, J. M. Avis, and K. Nagai. 1996. Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy. *Structure.* 4:621–631.

Karplus, M., and G. A. Petsko. 1990. Molecular dynamics simulations in biology. *Nature.* 347:631–639.

Kitchen, D. B., L. H. Reed, and R. M. Levy. 1992. Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry.* 31: 10083–10093.

Kraulis, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.

Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

Lu, J., and K. B. Hall. 1995. An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol.* 247:739–752.

Lührmann, R., B. Kastner, and M. Bach. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta.* 1087:265–292.

Lutz-Freyermuth, C., C. C. Query, and J. D. Keene. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA.* 87:6393–6397.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

MacKerell, A. D., Jr., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.

Miller, J. L., and P. A. Kollman. 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.

Nagai, K. 1996. RNA-protein complexes. *Curr. Opin. Struct. Biol.* 6:53–61.

Nagai, K., C. Oubridge, N. Ito, J. Avis, and P. Evans. 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., C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature.* 348:515–520.

Nicholls, A. 1992. GRASP: Graphical Representation and Analysis of Surface Properties. Columbia University, New York.

Nilsson, L. 1998. Protein-nucleic acid interactions. *In* Encyclopedia of Computational Chemistry. P. V. R. Schleyer, editor. John Wiley and Sons, New York. 2220–2229.

Oubridge, C., N. Ito, P. R. Evans, C.-H. Teo, and K. Nagai. 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.

Ramos, A., C. C. Gubser, and G. Varani. 1997. Recent solution structures of RNA and its complexes with drugs, peptides and proteins. *Curr. Opin. Struct. Biol.* 7:317–323.

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

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

Schwabe, J. W. R. 1997. The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* 7:126–134.

Sillekens, P. T. G., W. J. Habets, R. P. Beijer, and W. J. van Venrooij. 1987. CDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.* 6:3841–3848.

Steinbach, P. J., and B. R. Brooks. 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* 15:667–683.

Tang, Y., and L. Nilsson. 1998. Interaction of human SRY protein with DNA: a molecular dynamics study. *Proteins.* 31:417–433.

Tidor, B. 1997. Molecular dynamics simulations. *Curr. Biol.* 7:R525–R527.

Uhlenbeck, O. C., A. Pardi, and J. Feigon. 1997. RNA structure comes of age. *Cell.* 90:833–840.

Varani, G. 1997. RNA-protein intermolecular recognition. *Acc. Chem. Res.* 30:189–195.

Varani, G., and K. Nagai. 1998. RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* 27:407–445.

Wahl, M. C., and M. Sundaralingam. 1997. C-H···O hydrogen bonding in biology. *Trends Biochem. Sci.* 22:97–102.

Weeks, K. M. 1997. Protein-facilitated RNA folding. *Curr. Opin. Struct. Biol.* 7:336–342.