doi:10.1006/jmbi.2000.3629 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **297**, 1145–1158

JMB

Structure and Thermodynamics of RNA-protein Binding: Using Molecular Dynamics and Free Energy Analyses to Calculate the Free Energies of Binding and Conformational Change

Carolina M. Reyes and Peter A. Kollman*

Department of Pharmaceutical Chemistry, University of California San Francisco San Francisco, CA 94122-0446, USA

**Corresponding author*

An adaptive binding mechanism, requiring large conformational rearrangements, occurs commonly with many RNA-protein associations. To explore this process of reorganization, we have investigated the conformational change upon spliceosomal U1A-RNA binding with molecular dynamics (MD) simulations and free energy analyses. We computed the energetic cost of conformational change in U1A-hairpin and U1A-internal loop binding using a hybrid of molecular mechanics and continuum solvent methods. Encouragingly, in all four free energy comparisons (two slightly different proteins, two different RNAs), the free macromolecule was more stable than the bound form by the physically reasonable value of ~10 kcal/mol. We calculated the absolute binding free energies for both complexes to be in the same range as that found experimentally.

Keywords: human U1A protein; RNA; molecular dynamics; continuum solvent models; MM/PBSA

Introduction

Structural and biochemical studies of RNA-protein interactions examine how a protein recognizes a specific RNA site, what effect it has on RNA structure, and how their interactions promote a specific function (Draper, 1995). For ribonucleoprotein (RNP) complexes, these functions include transcription, splicing, and translation, which are crucial in regulating gene expression. In many cases, RNA-protein complexes are formed by an ''adaptive binding'' mechanism, wherein both molecules undergo significant conformational changes upon binding (De Guzman *et al*., 1998). What is the contribution of structural reorganization in RNA-protein binding? The free energy cost of structural adaptation is experimentally difficult, if not impossible, to obtain, but amenable to theoretical studies. Theoretical investigations of RNA-protein binding can complement biochemical and structural studies and formulate insights into their dynamics and energetics. So, to better understand RNA-protein recognition, we computed the absolute binding free energies and the adaptation free energies of one of the best-characterized RNP complexes, U1A-RNA (Allain *et al*., 1996; Avis *et al*., 1996; Gubser & Varani, 1996; Nagai *et al*., 1990; Oubridge *et al*., 1994).

U1A-RNA is a component of the U1 small nuclear ribonucleoprotein that binds to the 5′ splice site of a primary transcript. The U1A protein contains two RNA-recognition motifs (RRM) or RNA-binding domains (RBD), which fold into βαβ-βαβ structure, found in many RNA-binding proteins (Lu & Hall, 1997; Nagai *et al*., 1995). The N-terminal RBD of U1A (U1A RBD1) has been studied extensively, both structurally and biochemically, which makes it an ideal system for computational studies (Avis *et al*., 1996; Hall & Kranz, 1995; Jessen *et al*., 1991; Kranz & Hall, 1998; Kranz *et al*., 1996; Nagai *et al*., 1990; Zeng & Hall, 1997).

U1A RBD1 binds a hairpin RNA and an internal loop RNA with subnanomolar affinities (Hall, 1994). Both the crystal structure of hairpin complex (Oubridge *et al*., 1994) and the NMR structure of the internal loop complex (Allain *et al*., 1996) show hydrophobic stacking of RNA loop bases with residues on the β-sheet. In both structures, illustrated

Abbreviations used: BADH, bond, angle, dihedral; elec, electrostatic; MD, molecular dynamics; MM, molecular mechanics; PB, Poisson-Boltzmann; RBD, RNA-binding domain; RNP, ribonucleoprotein; RRM, RNA recognition motif; vdW, van der Waals.

E-mail address of the corresponding author: pak@cgl.ucsf.edu

0022-2836/00/051145–14 $35.00/0

in Figure 1(a), the β2-β3 loop, connecting the β-sheet, protrudes through the RNA loop and forms hydrogen bonds with the closing loop base-pair. The protein loop insertion through the RNA loop suggests that the binding process requires extensive conformational rearrangement. Indeed, the NMR structures of the unbound protein (Avis *et al*., 1996) and the unbound internal loop RNA (Gubser & Varani, 1996) confirm this. As represented in Figure 1(b), the C-terminal helix, helix C, is parallel with the β-sheet and partially obstructing the RNA-binding site, whereas in the bound structures, helix C is perpendicular to the β-sheet. Similarly, the single-stranded region of the RNA is more compact and collapsed in the free structure, while the same region is splayed out in the complex (Figure 1(c)). Hence, a large conformational change accompanies the binding of U1A RBD1 to its RNA, as illustrated in Figure 1(d).

To probe the energetic contributions of conformational change upon U1A RBD1-RNA binding, we carried out molecular dynamics (MD) simulations and binding free energy analyses on the bound and unbound monomers. The approach takes snapshots of the macromolecule from the MD trajectory and calculates their conformational molecular mechanics (MM) energy and solvation free energy. The average binding free energy is the sum of the average MM energy, conformational entropy and solvation free energy (Srinivasan *et al*., 1998b). The conformational entropy, $T\Delta S$, is approximated with normal mode calculations (Kottalam & Case, 1990; McQuarrrie, 1976):

$$\Delta\langle G\rangle = \Delta\langle E_{EE}\rangle - \langle T\Delta S\rangle + \Delta\langle G_{solv}\rangle \qquad (1)$$

where:

$$\Delta\langle G_{solv}\rangle = \Delta\langle G_{PB}\rangle + \Delta\langle G_{nonpolar}\rangle \qquad (2)$$

The solvation free energy, ΔG_{solv}, is estimated as the sum of the electrostatic contribution, calculated by the finite difference Poisson-Boltzmann (PB) model (Sharp & Honig, 1990a) and the non-polar contribution, as a function of the solvent-accessible surface area (SA). This hybrid of molecular mechanics and continuum solvent (MM/PBSA) method has been applied to nucleic acids (Srinivasan *et al*., 1998a), analyzing the solvation of RNA hairpins and helices (Srinivasan *et al*., 1998b) and sequence-dependent solvation of DNA helices (Cheatham *et al*., 1998). In general, continuum solvent methods have been widely utilized to study protein-protein (Froloff *et al*., 1997; Honig & Nicholls, 1995), protein-ligand (Shen & Wendolowski, 1996; Zhang & Koshland, 1996), and protein-DNA recognition (Jayaram *et al*., 1999; Misra *et al*., 1994, 1998).

We employ the MM/PBSA method here to investigate the absolute binding free energies of U1A RBD1-RNA and to calculate adaptation free energy associated with binding. To calculate the free energy cost of adaptation, we compared the energies of the snapshots from the trajectories of the free monomers and the snapshots of the monomers from the complex trajectories. The differences in their average energies, therefore, represent the adaptation free energy.

We applied two methods to compute the absolute binding free energies. The simpler approximation analyzes only the trajectories of the complexes and extract from them the intermolecular energies of the protein and RNA. The second approach computes the binding free energy as the difference of the free energies of the complex trajectories and the free energies of the separate trajectories of the unbound monomers. The separate trajectory protocol more appropriately describes U1A-RNA binding, as it implicitly includes the energetic contribution of adaptation by sampling the conformations of the complex and the unbound monomers.

Results

MD trajectory analysis of the unbound monomers

The unbound proteins were adapted from the 116 residue NMR structure of U1A (Avis *et al*., 1996) truncated to match the proteins in the complex structures. The protein bound to the internal loop complex is U1A102 and the protein bound to the hairpin is U1A97. The root-mean-squared deviation (RMSD) plots for U1A102 are shown in Figure 2. The average backbone RMSD is 1.46 Å. We examine the C-terminal helix, helix C, which moves as a rigid body away from the surface upon binding of the RNA. Helix C has an average RMSD of 1.57 Å, while the rest of the protein, the non-helix C region, is 1.34 Å. In the 1.3 ns simulation of U1A102, the trajectory is fairly stable and there is no significant difference in RMSD between helix C and the rest of the protein.

The simulation of U1A97, however, exhibits a different behavior. The backbone average RMSD, plotted in Figure 3, is 1.7 Å. The average helix C RMSD is 3.53 Å, while the non-helix C region RMSD is stable at 1.39 Å. Helix C deviates from the starting structure early in the simulation and with larger fluctuations than helix C of U1A102. At about 2 ns, the RMSD moves to 5 Å and remains there for the rest of the simulation. Figure 4 displays the initial structure (gray) superimposed with the minimized average structure from the last 500 ps (green). Figure 4 clearly illustrates the flexible regions, β2-β3 loop and helix C, during the simulation. NMR backbone dynamics studies by Mittermaier *et al*. (1999) show that β2-β3 loop is more flexible in the free protein compared to the complex and highlight the conformational change in helix C upon complex formation. If we match the non-helix C region and measure the angle represented by the helical axis of helix C in its initial and final average structures, we calculate a difference of about 35°. Over the 3 ns MD simulation of U1A97, we observe higher RMS fluctuations in

helix C compared to that found in U1A102. Helix C of U1A102 is five residues longer than that of U1A97 with greater H-bonding and van der Waals contacts with the β-sheet and thus maintains a more stable helix C. Also, the observed increased flexibility of the U1A97 helix C is in accord with the results of studies by Hall and co-workers (Kranz & Hall, 1998; Zeng & Hall, 1997). They found that the shorter helix C, truncated at residue 95, leads to a binding free energy that is 2 kcal/mol less than that found with the protein truncated at residue 101. However, they showed that the binding of U1A wt97 decreased by only 0.3 kcal/mol relative to U1A101 (Kranz & Hall, 1998). Furthermore, Zeng & Hall (1997) noted that the shorter helix is more dynamic and less helical in structure. Our simulated U1A97 is indeed more flexible, but the helical conformation is maintained throughout the simulation. Therefore, if a structured helix C is important in binding, and if truncation at residue 97 decreased binding by only 0.3 kcal/mol, then we can infer that U1A97 should have a helical structure, as was observed in our MD simulation. Nevertheless, it is interesting to observe that the β2-β3 loop and the helix C regions, which we know undergo large conformational rearrangements upon binding, have an inherent flexibility in the MD simulation. The flexibility of adaptive regions is even more apparent in the MD simulations of the unbound RNAs.

The RMSD plots for the unbound internal loop and hairpin RNAs are shown in Figure 5. The average RMSD for the internal loop (top) is 3.84 Å. Fitting on the single-stranded loop, upper stem and lower stem regions give average RMSDs of 3.59 Å, 1.26 Å, and 1.45 Å, respectively. Both the upper and lower stems have low RMSDs, while the loop RMSD parallels that of the overall RMSD. The loop heptanucleotide AUUGCAC fluctuations, therefore, contribute significantly to the overall RMSD of the internal loop.

Similarly, the RMSD of the hairpin RNA, 4.00 Å, couples with its loop RMSD, 4.48 Å, over the 2.4 ns MD simulation, while the stem region yields a low RMSD of 1.75 Å. The free hairpin structure is not known but it is believed to be highly unstructured (Hall, 1994; Oubridge *et al*., 1994). Without a starting structure, we modeled the free hairpin by running an MD simulation starting from the bound RNA. During the simulation, the loop bases, which were initially splayed out and solvent-exposed, collapsed inward and away from the solvent. We observed this inward movement, or collapse, of the loop bases in the first 500 ps and they remained removed from solvent for the rest of the simulation. Figure 6 illustrates the difference between the starting structure and the final 500 ps minimized average structure. Interestingly, Tang & Nilsson (1999) have observed similar loop flexibility in their MD simulation of the unbound hairpin.

Adaptation energy of the monomers

Our free and bound trajectories allow us to examine the conformational change or adaptation free energy associated with binding. The adaptation energy is calculated as the BOUND minus FREE total energies. A comparison of the average energy components of the bound and unbound U1A102 is given in Table 1.

For U1A102, the bond, angle, dihedral (BADH) energies are favorable by 12 kcal/mol. Both $\langle \Delta E(\text{vdW}) \rangle$ and $\langle \Delta E(\text{elec}) + \Delta G(\text{PB}) \rangle$ disfavor reorganization by 10 and 13 kcal/mol, respectively. The total electrostatics energy, $\langle \Delta E(\text{elec}) + \Delta G(\text{PB}) \rangle$, is the sum of the MM electrostatics energy and ΔG(PB). The total average adaptation energy change is 12.6 kcal/mol, so the conformation of the free monomer is more favored. As shown in Table 2, the calculations of bound and free U1A97 give similar results. Its adaptation

Table 1. Adaptation energy for U1A102

U1A102	Bound	Free	Adaptation energy
$\langle E(\text{BADH}) \rangle$	1741.55	1753.22	−11.67
	(2.54)	(2.53)	(3.59)
$\langle E(\text{vdW}) \rangle$	−387.75	−397.66	9.90
	(1.27)	(1.66)	(2.09)
$\langle E(\text{elec}) \rangle$	−2179.35	−2315.93	136.58
	(6.16)	(5.52)	(8.27)
$\langle E(\text{MM total}) \rangle$	−825.56	−960.37	134.81
	(5.84)	(5.99)	(8.36)
$\langle G(\text{PB}) \rangle$	−2587.66	−2463.82	−123.84
	(5.52)	(5.62)	(7.88)
$\langle E(\text{elec}) + G(\text{PB}) \rangle$	−4767.00	−4779.80	12.80
	(1.60)	(1.53)	(2.22)
$\langle G(\text{non-polar}) \rangle$	38.77	37.13	1.65
	(0.05)	(0.06)	(0.08)
$\langle G(\text{total}) \rangle + T\Delta S$	−3374.44	−3387.06	**12.62**
	(2.27)	(2.49)	(3.37)

Average over 100 snapshots from the last 1 ns trajectory; standard error of the mean in parentheses.

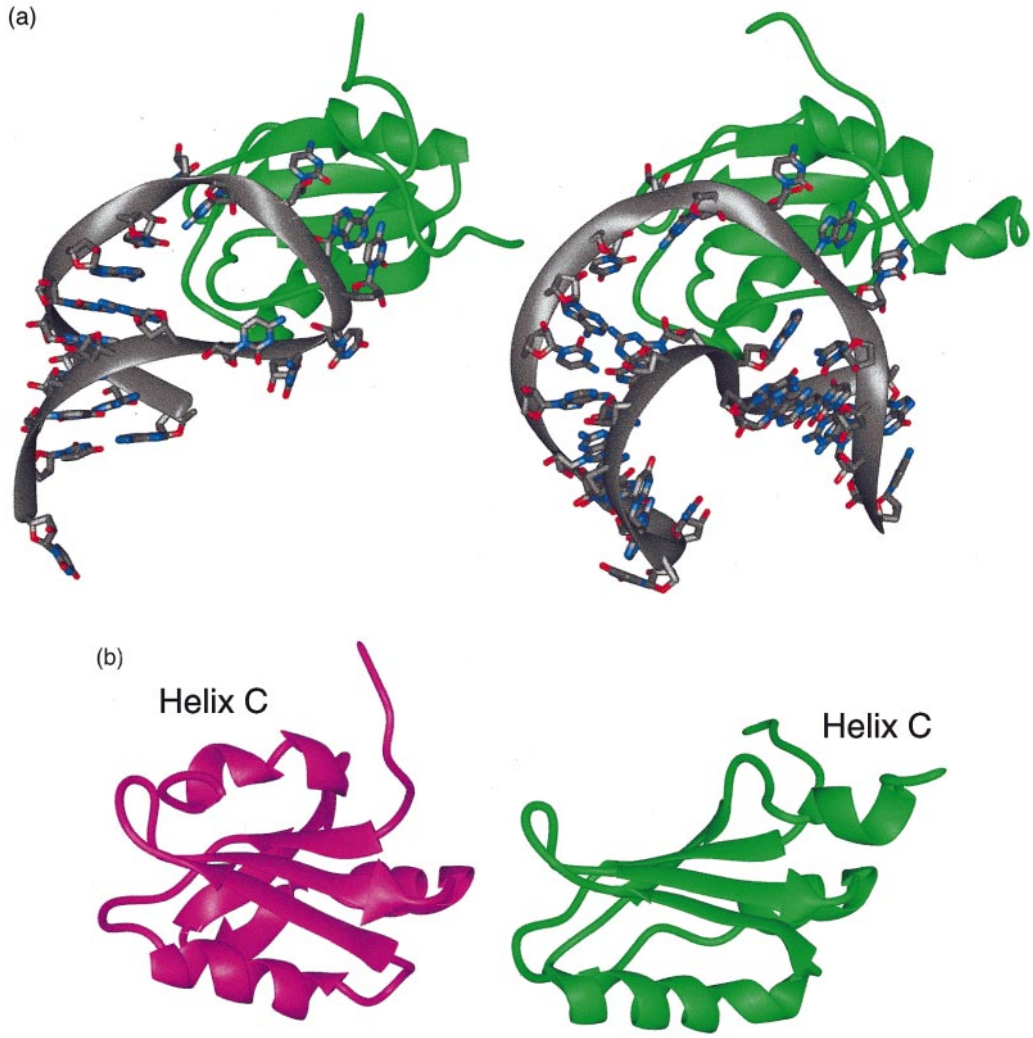

Figure 1 (*legend shown opposite*)

energy is 11.3 kcal/mol. The $\langle \Delta E(\text{BADH}) \rangle$ is −14.5 kcal/mol, while $\langle \Delta E(\text{vdw}) \rangle$ and $\langle \Delta E(\text{elec}) + \Delta G(\text{PB}) \rangle$ are unfavorable by 9.0 kcal/mol and 11.8 kcal/mol, respectively. In both conformations of bound U1A97 and U1A102, helix C is perpendicular to the β-sheet surface, exposing the conserved RNP region of the β-sheet to interact with the RNA loop. In the unbound protein, helix C lies on top of the β-sheet stabilized by hydrophobic interactions. Therefore, the 11-12 kcal/mol difference can be attributed to the change in orientation of helix C relative to the β-sheet RNA-binding surface.

Comparing the bound and free internal loop in Table 3, we also find that the free conformation is more favorable by 7.9 kcal/mol. The magnitudes of the energy contributions, however, decompose differently. The $\langle \Delta E(\text{BADH}) \rangle$ and $\langle \Delta E(\text{vdW}) \rangle$ are unfavorable by 14 kcal/mol and 24 kcal/mol, while the $\langle \Delta E(\text{elec}) + \Delta G(\text{PB}) \rangle$ contribution is favorable by 32.7 kcal/mol. Analogous to the internal loop, the energy analysis of the bound and free hairpin RNA in Table 4 yields a mean adaptation energy of 8.7 kcal/mol. With the hairpin RNA, both $\langle \Delta E(\text{BADH}) \rangle$ and $\langle \Delta E(\text{vdW}) \rangle$ are unfavorable by 12 kcal/mol and 5 kcal/mol, while the $\langle \Delta E(\text{elec}) + \Delta G(\text{PB}) \rangle$ contribution is favorable by 9.5 kcal/mol. The absolute values of the average bound and free hairpin energies are similar, since the free hairpin is modeled after the bound crystal structure. Nevertheless, for the change in orientation of the loop bases, we calculate the adaptation energy penalty of 8-9 kcal/mol. In the bound internal loop and hairpin RNAs, the loop bases are splayed out into solvent and the phosphate backbone is more extended. Hence, the more ''open'' bound conformation permits more interactions with water, resulting in more favorable solvation. With such highly charged systems as RNAs, solvation plays a significant role in conformational and binding free energies.

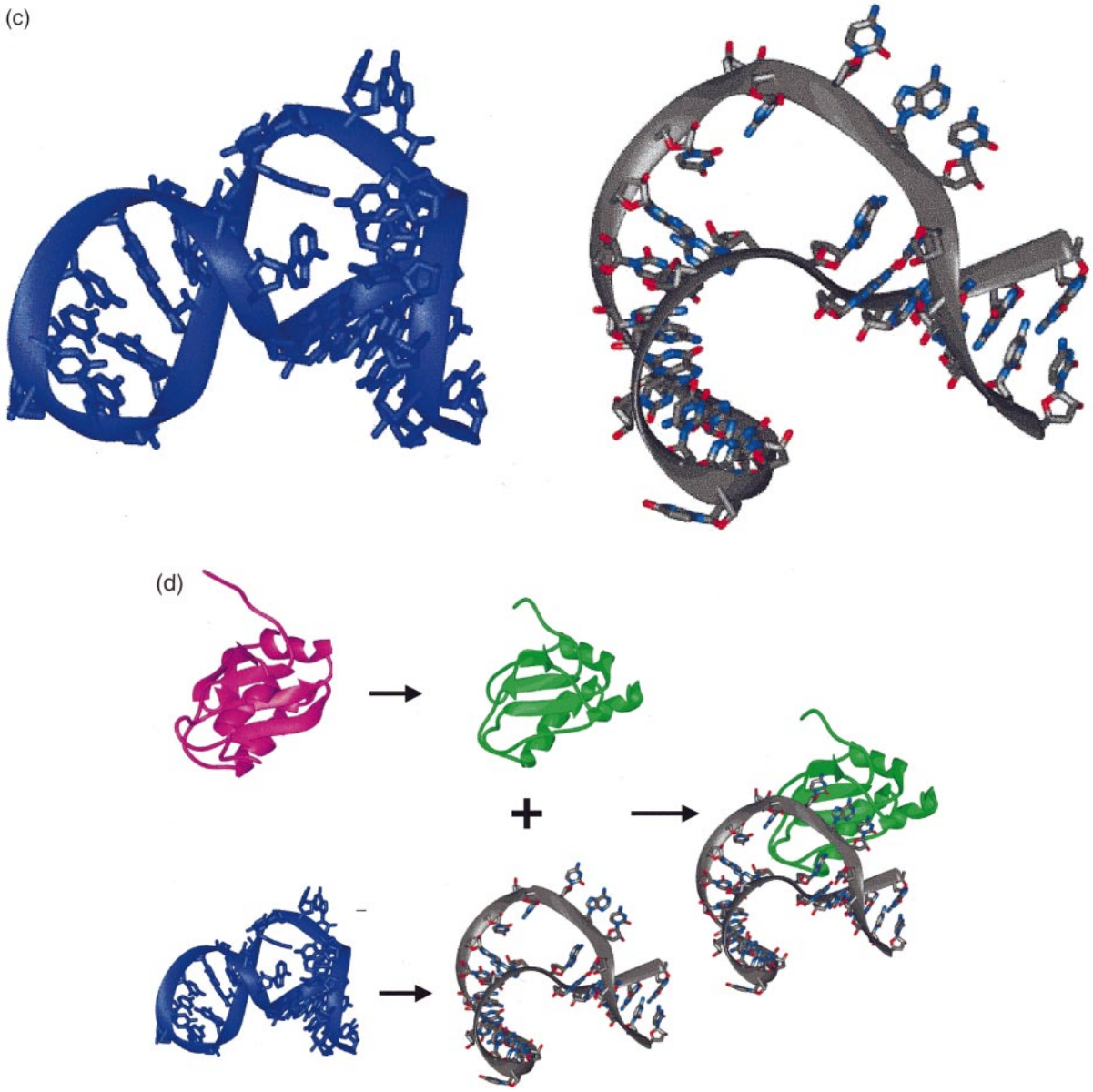

Figure 1. (a) The U1A protein (green) bound to the hairpin RNA (left) and to the internal loop RNA (right). (b) The conformational change of U1A102 upon binding; in the free form (magenta), helix C sits on top of the β-sheet, while in the bound form (green), helix C is oriented away from the β-sheet. (c) The internal loop RNA undergoes significant structural change upon binding. The free internal loop (blue) is more collapsed, while the bound internal loop has its loop bases extruding out to interact with hydrophobic residues on the protein's β-sheet. (d) Adaptive binding of U1A-RNA.

Binding free energies

The calculated average binding free energies of U1A RBD1-RNA are reported in Table 5. In general, the MM energies and the electrostatic solvation energies compensate each other; both are large numbers, but opposite in sign. For example, extended conformations expose lots of surface charges to solvent, lowering the solvation energy. However, the distance between the charges is now increased, increasing E(elec) and E(vdW). The binding of U1A RBD1 to its RNA is an energetic balance of the vdW energies and the total electrostatic energies (sum of MM electrostatic and PB terms). In all four calculations, the vdW contributions to binding are favorable by 82-116 kcal/mol, but the total electrostatic contributions are unfavorable by 40-60 kcal/mol. The non-polar solvation term, solvent-accessible surface area-dependent term, is always favorable by 10-14 kcal/mol. The sum of the MM energies, $\langle \Delta E(\text{MM total}) \rangle$, and the solvation free energies, $\langle \Delta G(\text{PB}) \rangle + \langle \Delta G(\text{non-polar}) \rangle$, yield favorable binding for both hairpin complex (−69 kcal/mol for single trajectory and −48.4 kcal/mol for separate trajectory) and

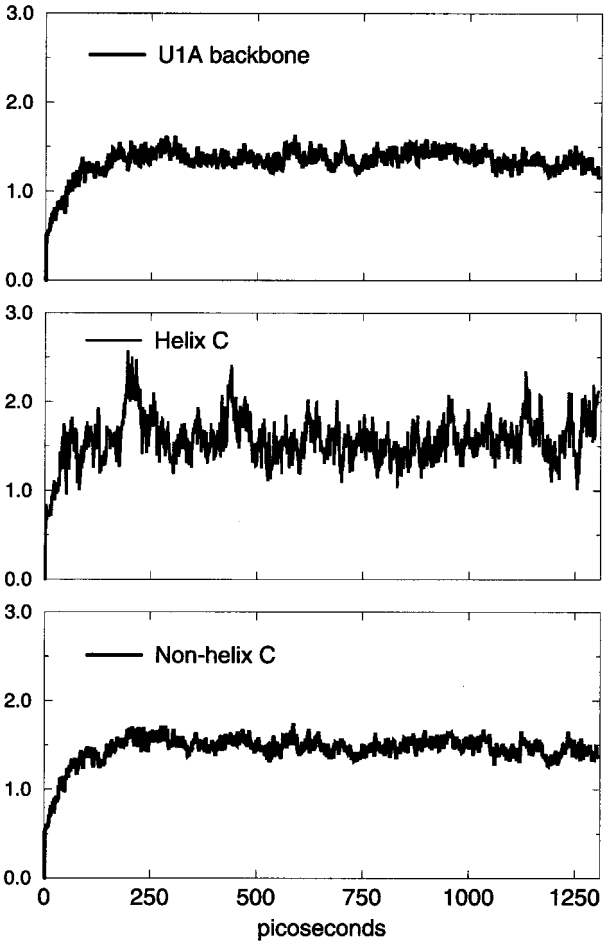

Figure 2. RMSD plots for U1A102 relative to its initial structure; top, protein backbone; middle, helix C region; and bottom, the non-helix C region.

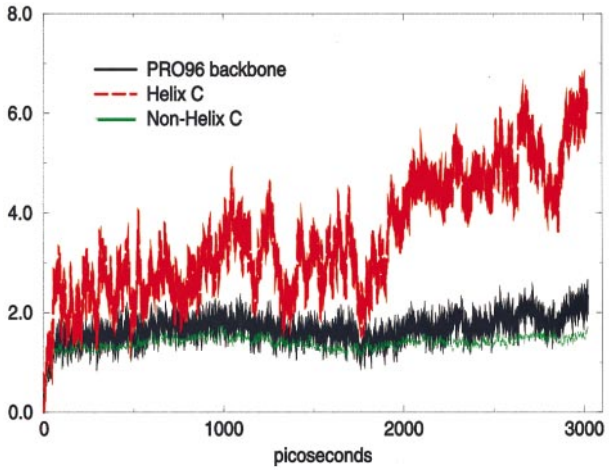

Figure 3. RMSD plot for U1A97 relative to its initial structure. Fitting to the protein backbone (black) and the non-helix C region (green) give stable RMSD while the helix C region (red) shows greater mobility.

internal loop complex (−72 kcal/mol for single trajectory and −51.5 kcal/mol separate trajectory). The ∼20 kcal/mol energy difference between the single trajectory and separate trajectory protocols stems from the conformational adaptation of the monomers in binding.

The computed sum of the MM energies and solvation free energies are significantly more favorable than experimentally determined binding free energies, since it does not account for the entropic contributions to binding. The entropy calculations were performed on the experimental structures of the complexes and unbound monomers. We also computed the entropies of the bound monomers minimized from their experimental complex structures. For hairpin binding, the average solute entropy loss is 51.1 kcal/mol. The entropy loss for internal loop binding is 39.6 kcal/mol. Note (Table 5) that both averages are based on two independent estimates (bound and unbound mono-

Table 2. Adaptation energy for U1A97

U1A97	Bound	Free	Adaptation energy
$\langle E(\text{BADH})\rangle$	1652.27	1666.77	−14.5
	(2.75)	(2.01)	(3.40)
$\langle E(\text{vdW})\rangle$	−387.66	−396.69	9.03
	(1.53)	(1.58)	(2.20)
$\langle E(\text{elec})\rangle$	−2356.65	−2512.49	155.84
	(3.62)	(5.21)	(6.34)
$\langle E(\text{MM total})\rangle$	−1092.03	−1247.42	155.96
	(4.32)	(5.07)	(6.66)
$\langle G(\text{PB})\rangle$	−2268.45	−2124.44	−144.01
	(3.30)	(4.80)	(5.82)
$\langle E(\text{elec}) + G(\text{PB})\rangle$	−4625.09	−4636.9	11.81
	(1.50)	(1.79)	(2.34)
$\langle G(\text{non-polar})\rangle$	34.12	34.21	−0.09
	(0.03)	(0.05)	(0.06)
$\langle G(\text{total})\rangle + T\Delta S$	−3326.36	−3337.64	**11.28**
	(2.46)	(2.13)	(3.25)

Average over 100 snapshots from the last 1 ns trajectory; standard error of the mean in parentheses.

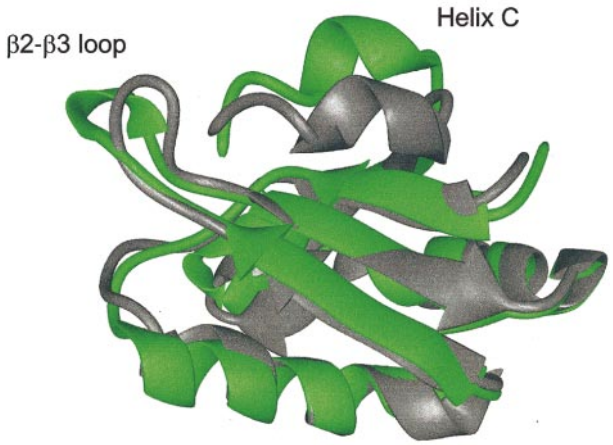

Figure 4. Superimposing the backbone of the initial (gray) and averaged final structure (green) of U1A97 reveals the shifted positions of helix C and β2-β3 loop.

mers) that differ by 10-20 kcal/mol for these systems. These entropy values are based on simple harmonic approximations to the normal modes of the two minimized initial structures and are crude estimates of the solute entropy, since they do not necessarily represent the trajectory-averaged solute entropies. We report 45.3 kcal/mol as the average of these four sets of entropy calculations, $\langle -T\Delta S \rangle$, as broad ballpark estimates of the entropic contributions to the binding free energies.

All the above terms added together ($\langle \Delta E(\text{MM total}) \rangle - \langle T\Delta S \rangle + \langle \Delta G(\text{PB}) \rangle + \langle \Delta G(\text{nonpolar}) \rangle$), yield the average calculated binding free energies of U1A RBD1-RNA. The experimental binding free energies are −14.2 kcal/mol for the hairpin (Williams & Hall, 1996) complex and −12.6 kcal/mol for the internal loop complex (Gubser & Varani, 1996). In the single trajectory calculation of hairpin binding, the average binding is −23.7 kcal/mol, while the separate trajectory

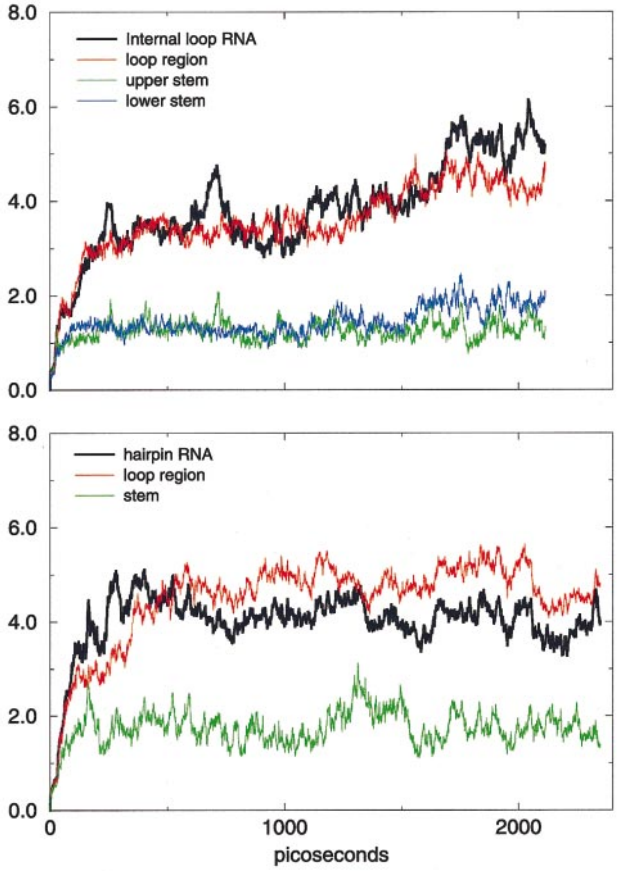

Figure 5. RMSD plots for the internal loop RNA (top) fitting all RNA (black), the loop region (red), upper stem (green) and lower stem (blue). RMSD plots for the hairpin RNA simulation (bottom), fitting all RNA (black), loop region (red) and the stem (green).

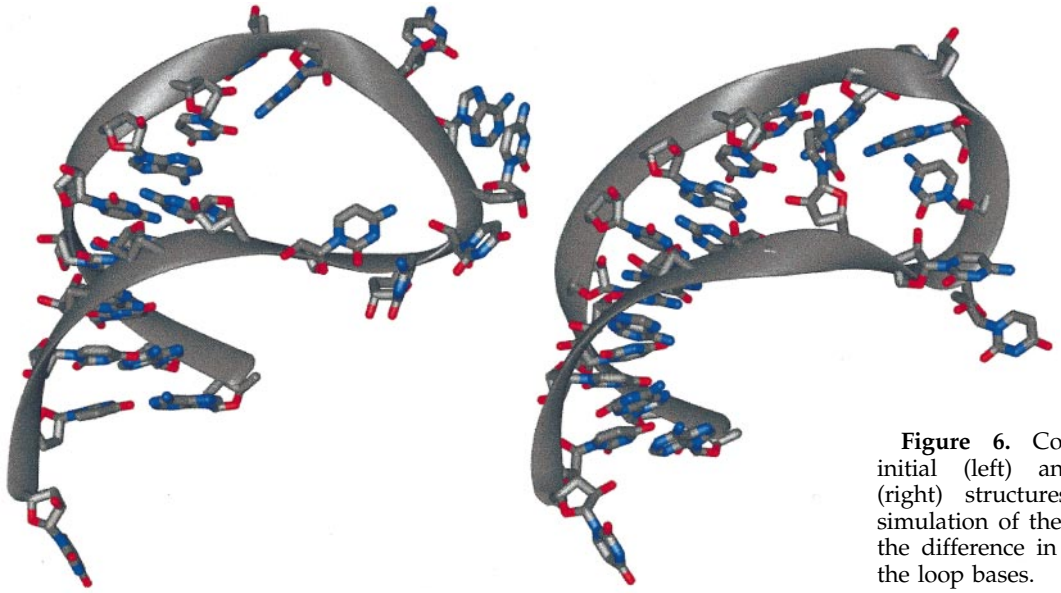

Figure 6. Comparison of the initial (left) and final average (right) structures from the MD simulation of the hairpin illustrates the difference in the orientation of the loop bases.

protocol yields −3.1 kcal/mol. For the internal loop complex, the single trajectory average binding is −26.7 kcal/mol and the separate trajectory average is −6.2 kcal/mol. The separate trajectory calculation of the internal loop complex, whose MD simulations started from known experimental structures of the free monomers, gives the best agreement with experiment. However, given the approximations in the entropy calculations, the absolute binding free energies are less accurate and they do not include salt contributions to binding.

Salt effects

We now investigate electrostatic effects in solvation; more specifically, the effect of ionic strength to U1A-hairpin binding with non-linear Poisson-Boltzmann (PB) calculations (Sharp & Honig, 1990b). The average contribution of 150 mM ionic strength to binding (150 mM salt − no salt) is 28.3(±0.7) kcal/mol for the single trajectory protocol and 26.5(±1.3) kcal/mol for the separate trajectory protocol. This is indeed a significant contribution, but we cannot confirm the accuracy of this absolute value with experimental results. However, we can compare the relative contribution of increasing salt concentration to binding with the results of studies by Hall & Stump (1992). Relative to 150 mM ionic strength, as was measured experimentally, Table 6 compares the calculated and experimental effects of added salt. For 500 mM salt concentration, binding decreases by 2.7 kcal/mol for the single trajectory and 2.5 kcal/mol for the separate trajectory. Both compare well with the experimentally determined 3.1 kcal/mol (Hall & Stump, 1992). Hall and Stump extrapolated their results to 1 M salt and suggested that binding would decrease by 6.2 kcal/mol. We calculate that at 1.0 M salt concentration, U1A RBD1-hairpin binding would decrease by only 3.9 kcal/mol for the single trajectory and 3.6 kcal/mol for the separate trajectory. The accuracy of the PB calculations decreases in the limit of high ionic strength (Sharp, 1995; Fixman, 1979), which may be the reason for the discrepancy at 1 M ionic strength. In any case, we computed that at increasing ionic strength, the binding free energy becomes significantly less favorable, as found experimentally.

Discussion

Total electrostatics opposes binding

MD simulations (Hermann & Westhof, 1999) recently examined the microscopic picture of salt-induced U1A-hairpin destabilization. With explicitly added 1 M Na^+ and Cl^- counterions, MD studies showed increased fluctuations on the RNA-protein binding site and on the RNA compared to 100 mM salt simulations. Hermann and Westhof concluded that the decrease in U1A-hairpin binding at higher salt concentration is due to the increased dynamics at the binding interface.

The total electrostatics is computed as the sum of the MM electrostatic energies (E(elec)) and the Poisson-Boltzmann (ΔG(PB)) solvation term. We calculated a ΔG total electrostatics of ~60 kcal/mol for U1A-internal loop and 40-60 kcal/mol for U1A-hairpin. Although these are obtained by adding two large terms with standard errors of the order of 6-20 kcal/mol, their sums are well converged, with standard errors of 1-3 kcal/mol. This is due to compensatory effects of the PB term with the MM electrostatic energies. In our calculations of the total electrostatics, the more favorable PB solvation term for the highly charged unbound RNAs relative to the complex is a significant contributor to the overall electrostatics. The favorable electrostatic solvation of the unbound RNA dominates and opposes binding.

Table 3. Adaptation energy for internal loop RNA

Internal loop	Bound	Free	Adaptation energy
$\langle E(\text{BADH})\rangle$	1418.03	1403.91	14.12
	(2.02)	(2.09)	(2.91)
$\langle E(\text{VDW})\rangle$	−227.81	−252.20	24.39
	(1.29)	(1.40)	(1.90)
$\langle E(\text{elect})\rangle$	1768.88	2458.59	−689.71
	(12.22)	(17.99)	(21.75)
$\langle E(\text{MM total})\rangle$	2959.10	3610.30	−651.20
	(12.42)	(17.50)	(21.46)
$\langle G(\text{PB})\rangle$	−9767.64	−10423.89	656.24
	(11.65)	(16.94)	(20.56)
$\langle E(\text{elec}) + G(\text{PB})\rangle$	−7998.00	−7965.30	−32.70
	(1.38)	(1.78)	(2.25)
$\langle G(\text{non-polar})\rangle$	35.15	32.26	2.89
	(0.03)	(0.04)	(0.05)
$\langle G(\text{total})\rangle + T\Delta S$	−6773.40	−6781.32	**7.93**
	(1.95)	(1.82)	(2.67)

Average over 100 snapshots from the last 1 ns trajectory; standard error of the mean in parentheses.

Theoretical studies of DNA-protein complexes have found unfavorable electrostatic contributions to binding, due to favorable solvation of the unbound nucleic acids (Jayaram *et al.*, 1999; Misra *et al.*, 1998). Moreover, experimental estimates, based on the number of released ions upon U1A-hairpin association, assess that the binding is primarily due to non-electrostatic contributions (Hall & Stump, 1992). Similarly, other RNA-protein binding studies estimate that as much as 80 % of binding come from non-electrostatic contributions (Ryan & Draper, 1989). Our calculations reveal that ΔE(vdW) interactions are very favorable, essentially driving the association of U1A-RNA. The structures of the complexes do reveal hydrophobic stacking and H-bonds on the interface.

However, we do not imply that electrostatic interactions are insignificant in U1A-RNA complexes. The net contribution is small, but the large magnitude of these terms means that ionic strength can significantly adjust the balance of forces in the U1A-RNA interaction, as we have found. As an alternative to the interpretation offered by Hermann & Westhof (1999), we propose that the salt effects on U1A-RNA binding are primarily due to bulk electrostatic (ionic strength) effects that change the relative stability of free and bound RNA, rather than alterations in the entropy or dynamics of the complex.

Conformational preorganization free energy

For U1A-RNA association, we approximate the conformational change upon binding as the sum of the adaptation energies of the monomers. Experimentally, this value is difficult to obtain but we do know, based on the structures, that U1A-RNA binding is accompanied by significant conformational change. The adaptation free energy emerges from our MM/PBSA protocols of calculating the separate monomer trajectories and the monomers in the complex trajectories. The energetic difference between the two protocols rep-

Table 4. Adaptation energy for the hairpin

Hairpin	Bound	Free	Adaptation energy
$\langle E(\text{BADH})\rangle$	975.85	963.97	11.88
	(1.88)	(1.67)	(2.51)
$\langle E(\text{vdW})\rangle$	−127.58	−132.45	4.87
	(0.93)	(1.00)	(1.37)
$\langle E(\text{elec})\rangle$	370.23	348.53	21.70
	(3.51)	(4.88)	(6.01)
$\langle E(\text{MM total})\rangle$	1218.50	1180.05	38.45
	(3.73)	(4.75)	(6.04)
$\langle G(\text{PB})\rangle$	−5643.72	−5612.55	−31.17
	(3.12)	(4.66)	(5.61)
$\langle E(\text{elec}) + G(\text{PB})\rangle$	−5273.49	−5264.00	−9.49
	(1.06)	(1.19)	(1.60)
$\langle G(\text{non-polar})\rangle$	26.39	24.93	1.45
	(0.02)	(0.25)	(0.25)
$\langle G(\text{total})\rangle + T\Delta S$	−4398.83	−4407.56	**8.73**
	(1.86)	(1.50)	(2.39)

Average over 100 snapshots from the last 1 ns trajectory; standard error of the mean in parentheses.

Table 5. Calculated average binding free energy

	U1A RBD1 hairpin binding		U1A RBD1 internal loop binding	
	Single trajectory	Separate trajectory	Single trajectory	Separate trajectory
$\langle E(\text{BADH})\rangle$	0.0	2.2	0.0	2.5
	–	(2.9)	–	(4.5)
$\langle E(\text{vdW})\rangle$	−112.0	−98.1	−116.2	−82
	(0.6)	(2.7)	(0.6)	(2.8)
$\langle E(\text{elec})\rangle$	−3090.6	−2913.0	−4375.4	−4928.5
	(6.0)	(6.4)	(17.8)	(20.8)
$\langle E(\text{MM total})\rangle$	−3203.6	−3008.9	−4491.6	−5008.0
	(6.0)	(6.9)	(18.1)	(20.8)
$\langle G(\text{PB})\rangle$[a]	3147.4	2972.2	4433.7	4966.1
	(6.0)	(5.9)	(18.3)	(20.0)
$\langle E(\text{elec}) + G(\text{PB})\rangle$	56.7	59.2	58.3	38.1
	(2.1)	(2.9)	(1.1)	(3.4)
$\langle G(\text{non-polar})\rangle$	−12.8	−11.7	−14.1	−9.6
	(0.0)	(0.1)	(0.1)	(0.1)
$\langle \boldsymbol{G}\textbf{(total)}\rangle + \langle \boldsymbol{T\Delta S}\rangle$	**−69.0**	**−48.4**	**−72.0**	**−51.5**
	(0.9)	(3.9)	(1.0)	(4.2)
$-T\Delta S$	57.6	44.6	49.4	29.6
$-\langle T\Delta S\rangle$	45.3	45.3	45.3	45.3
$\langle \Delta \boldsymbol{G}_{\textbf{bind}}\rangle$	**−23.7**	**−3.1**	**−26.7**	**−6.2**
	(0.9)	(3.9)	(1.0)	(4.2)
Experimental binding[b]	**−14.2**	**−14.2**	**−12.6**	**−12.6**

Calculated binding energies and standard errors in parenthesis averaged from 100 snapshots from the last 1 ns in kcal/mol.
MM/PBSA: $\Delta G_{\text{bind}} = \Delta E_{\text{MM total}} - T\Delta S + \Delta G_{\text{PB}} + \Delta G_{\text{non-polar}}$.
[a] Zero ionic strength.
[b] (Williams & Hall, 1996; Gubser & Varani, 1996).

resents the conformational adaptation that occurs with U1A-RNA binding.

In the complex trajectories, the bound monomers are preorganized for optimal interactions; hence, their binding free energies are overestimated and more favorable than the unbound monomer trajectories. Therefore, it should be emphasized that with macromolecular association accompanied by large conformational change, theoretical calculations based on the complex structures alone can overestimate the experimental binding free energy. Binding is indeed always a thermodynamic balance of the bound and unbound constituents.

Our estimation of 20 kcal/mol (∼half protein and ∼half RNA) is within the range of values that have been proposed for the conformational adaptation or strain free energy (Chipot & Pohorille, 1998; Bostrom *et al.*, 1998; Vorobjev & Hermans, 1999; Vorobjev *et al.*, 1998; Jayaram *et al.*, 1999). More specifically, our estimation of 11-12 kcal/mol for the change in orientation of the U1A helix C relative to the hydrophobic β-sheet (parallel *versus* perpendicular) closely corresponds to the potential of mean force (PMF) calculation made by Chipot & Pohorille (1998). They calculated a free energy difference of 4-13 kcal/mol for a 90° change in orientation of a polyleucine α-helix relative to a water-hexane interface. Other theoretical studies of binding free energies with MM-continuum solvent hybrid methods have explored the conformational energy penalties with macromolecular binding. For small protein ligands, the strain energy was computed to be around 3 kcal/mol (Bostrom *et al.*, 1998). For the extreme case of preorganization, protein folding, Hermans and co-workers have assessed the free energy difference between folded and misfolded protein models to be of the order of 11-190 kcal/mol (Vorobjev *et al.*, 1998). For DNA-protein association of the *Eco*RI-DNA complex, Beveridge and co-workers have found the adaptation free energy to be 63 kcal/mol (Jayaram *et al.*, 1999).

Overall, the MM/PBSA method is able to qualitatively reproduce the absolute binding free energy of U1A RBD1 to its RNA target. As noted above, the greatest source of error in these calculations is the solute entropy term, which is obtained from normal mode calculations. A simple harmonic approximation of the normal modes will not adequately describe the internal entropies of macro-

Table 6. Effect of salt concentration in kcal/mol relative to 150 mM salt

Salt concentration	Single trajectory hairpin	Separate trajectory hairpin	Experimental[b]
150 mM	0	0	0
500 mM	2.7	2.5	3.1
1.0 M	3.9	3.6	6.2

Calculated and experimental effects of ionic strength to binding relative to 150 mM in kcal/mol.
[a] Hall & Stump (1992).

molecules, but it is not clear that a quasi-harmonic methods (Vorobjev & Hermans, 1999; Vorobjev *et al*., 1998) will be much better, given the limited number of local minima sampled in nanosecond simulations. Moreover, MM/PBSA does not explicitly include solvent entropy, but includes it implicitly in the change in solvent-accessible surface area upon binding.

Interestingly, however, if we simply look at the non-polar solvation term, it is the smallest of all the other terms and yet gives a pretty good approximation of the overall observed binding for U1A RBD1. However, the non-polar solvation term alone does not resolve or explain binding specificity. Based on our calculations of different U1A RBD1-hairpin mutations (Reyes & Kollman, 2000), the non-polar solvation term is fairly constant and insensitive to the single amino acid or base mutations.

Our theoretical studies reveal regions, that undergo substantial conformational change upon U1A RBD1-RNA binding and have inherent flexibility in MD simulations. Secondly, this preorganization or adaptation free energy of the protein and RNA upon complex formation is of the order of 20 kcal/mol. We have described the molecular dynamics of the unbound monomers of the U1A RBD1-RNA complexes. The ability to observe dynamic functional regions is an important feature of MD simulations, as it can allow us to predict functional sites or a mechanism of binding for a given complex structure.

Furthermore, we investigated the energetics of conformational adaptation and binding of U1A RBD1-RNA by the MM/PBSA method. We find that the total electrostatics oppose binding while van der Waals contributions favor binding. Our protocol of using separate trajectories of the monomers, starting from their unbound experimental structures, yields the best agreement with experimental binding free energies. This implies that free energy analyses based on the complex structure alone may overestimate binding by as much as 20 kcal/mol, since they ignore the energetic penalty of conformational rearrangement upon binding.

In summary, with the caveats mentioned duly noted, there are two very exciting results from these calculations. First, the calculated absolute binding free energies of both U1A-hairpin RNA and U1A-internal loop RNA ($\Delta G \sim -5$ kcal/mol) are close to those found experimentally ($\Delta G \sim -12$ to -14 kcal/mol), despite the fact that these calculated free energies are the difference between free energies of thousands of kcal/mol. This encouraging energy balance has been found in a number of other examples, most notably a hairpin loop monomer dimer equilibrium (Srinivasan *et al*., 1998b). Secondly, to our knowledge this is the first calculation of the free energy difference between free and bound macromolecules (adaptation free energy) where a significant conformational change has taken place. All four adaptation free energies are of the order of ~10 kcal/mol, which is a physically reasonable magnitude.

Materials and Methods

MD simulations

MD simulations for the U1A-RNA complexes have been described by Reyes & Kollman (1999). All MD simulations were performed using AMBER5 (Case *et al*., 1997) with the Cornell *et al*. (1995) force field. They included explicit TIP3P water molecules (Jorgensen *et al*., 1983), neutralizing counterions, periodic boundary conditions, and full electrostatics with particle-mesh Ewald (Darden *et al*., 1993). The simulations ran with a constant temperature of 300 K and a constant pressure of 1 atm (101,325 Pa). Table 7 enumerates the MD simulation methods for the complexes and the unbound monomers.

The starting structures for the unbound U1A protein simulations were adapted from one of the NMR family of structures of the unbound 116 residue U1A (Avis *et al*., 1996). The NMR structure was truncated to a 96 residue protein (U1A97) corresponding to the crystal structure of U1A bound to the hairpin (Oubridge *et al*., 1994) and to a 101 residue protein (U1A102) corresponding to the NMR structure of U1A bound to the internal loop (Allain *et al*., 1996). Hydrogen atoms were added in EDIT, and their positions were optimized by minimization. After careful equilibration with slow gradual heating to 300 K over 60 ps, the simulations were continued for 1.3 ns for U1A102 and for 3 ns for U1A97. U1A97 was run longer because its calculated free energy (equation (1)) continually decreased until the last nanosecond, at which point its average free energy was stable.

One of the NMR structures of the unbound internal loop was taken as the initial structure (Gubser & Varani, 1996). The experimental structure for the unbound hairpin is not available. We modeled the free hairpin RNA from the crystal structure of the complex by removing the protein. Both RNAs were equilibrated to 300 K in

Table 7. MD simulations

System	Water box size (Å^3)	Neutralizing counterions	Total number of atoms	Length of simulation (ps)
U1A-hairpin	$84 \times 63 \times 59$	13 Na^+	28,369	1600
U1A97	$66 \times 51 \times 57$	8 Cl^-	19,844	3000
Free hairpin	$71 \times 53 \times 46$	20 Na^+	14,371	2300
U1A-internal loop	$99 \times 77 \times 63$	21 Na^+	42,845	2200
U1A102	$64 \times 64 \times 63$	8 Cl^-	22,636	2100
Free internal loop	$79 \times 64 \times 54$	29 Na^+	23,435	2200

Comparison of methods for all simulations: water box size (Å^3), number of added neutralizing counterions, total number of atoms, length of simulation (picoseconds).

60 ps. The internal loop simulation was carried out to 2.1 ns, while the unbound hairpin ran to 2.4 ns.

Snapshots for energy analyses were obtained from the MD trajectories of the unbound proteins, unbound RNAs, and both complexes. The water molecules and counterions were deleted for energy analyses. We monitored the change in total energies over time and the RMSD from the initial structures to assess the equilibration of the trajectories. We then selected 100 snapshots from the last nanosecond of each equilibrated trajectory in 10 ps intervals. The RMSD to the initial starting structure was calculated using CARNAL, the trajectory analysis suite of AMBER. Snapshots from the trajectories were visually examined and illustrated with MidasPlus (Ferrin *et al.*, 1988).

Energy analyses

The MM energies for the protein, RNA and complex were computed using the ANAL module in AMBER5 with the Cornell *et al.* (1995) force field. All solute pairwise interactions were included and represented as bond + angle + dihedrals (BADH), van der Waals, electrostatics, and total energies. In the single trajectory analysis, the energy calculations were performed on the monomers in the context of their complex structures, whereas in the separate trajectory method, analyses were executed on shapshots from the respective MD simulations of the complex and the monomers.

The electrostatic component of the solvation free energy for each snapshot was computed by the finite difference Poisson-Boltzmann method (Sharp & Honig, 1990a,b) as implemented in the Delphi II program (Nicholls *et al.*, 1990). This was approximated as the reaction field energy of taking a solute from a vacuum dielectric ($\varepsilon = 1$) to an aqueous dielectric ($\varepsilon = 80$). A probe radius of 1.4 Å was used to define the molecular surface, represented by atom-centered spheres with radii adapted from the PARSE parameter set (Massova & Kollman, 2000; Sitkoff *et al.*, 1994; Srinivasan *et al.*, 1998a). Atomic charges were taken from the Cornell *et al.* (1995) force field in order to have a consistent set of charges for calculating the total electrostatic energies (sum of MM electrostatics and PB) (Srinivasan *et al.*, 1998a). For each computed structure, an 80% boxfill cubic lattice with a grid resolution of 0.5 Å/grid point was applied and 1000 linear iterations were required for energy convergence. To calculate the effects of salt concentration on the hairpin complex binding, 1000 linear iterations and an additional 1000 non-linear iterations were performed at specified 150 mM, 500 mM and 1.0 M ionic strengths. Explicit monovalent or divalent counterions were not included in the salt calculations, but the bulk ionic strength was specified and treated with the non-linear PB calculations (Sharp & Honig, 1990b).

The non-polar contribution to solvation free energy was estimated with Sanner's MSMS software (Sanner *et al.*, 1996), which calculates the solvent-accessible surface area. The SA term relates to the non-polar solvation term as: $\Delta G_{\text{non-polar}} = \gamma\ (\text{SASA}) + \beta$, where $\gamma = 0.00542$ kcal/Å and $\beta = 0.92$ kcal/mol (Cheatham *et al.*, 1998).

Solute entropies were approximated with normal mode calculations (Srinivasan *et al.*, 1998b). This entailed at least 100,000 steps of steepest descent and conjugate gradient minimizations, with a distance dependent dielectric in the absence of solvent, followed by Newton-Raphson minimizations until the root-mean-square of the elements of gradient vector is less than 10^{-4} kcal/mol. The rotational, translational, and vibrational entropies were computed at 298 K. Ideally, one would average the entropies calculated for each snapshot in the same way as the total energies and solvation free energies were calculated. However, since normal mode calculations give inaccurate harmonic approximations of solute entropy and are computationally expensive, we performed only single calculations of the minimized starting structures. We minimized the experimental structures of both complexes and unbound monomers (U1A102, U1A97, and internal loop) and the 500 ps snapshot of the unbound hairpin simulation for which an experimental structure is not available. The minimized structures have moved away from the initial structures (1.0 to 2.8 Å RMSD) more dramatically at the terminal base-pairs of the RNA helices. We also computed the entropies of the bound monomers, adapted from the experimental structures of both complexes. Since normal mode calculations are imperfect estimates for the solute entropies (distorted minimized structures, harmonic approximation, single calculation), we consider them here to give us the magnitude of the entropic contribution to the binding free energy. Therefore, we reported the average entropy difference in binding of the monomers. As discussed (Cheatham *et al.*, 1998) normal mode analyses give only qualitative estimates of the solute entropy.

Acknowledgments

We thank Dr Jed Pitera and Dr Gabriele Varani for helpful discussions. We are grateful to the NSF National Partnership for Advanced Computational Infrastructure (NPACI) for supercomputer allocations. P.A.K. acknowledges the NIH (CA-25644) for support and C.M.R. is grateful to the Ford Foundation

References

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

Avis, J. M., Allain, F. H. T., Howe, P. W. A., 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.

Bostrom, J., Norrby, P. O. & Liljefors, T. (1998). Conformational energy penalties of protein-bound ligands. *J. Computer-Aided Mol. Des.* **12**, 383-396.

Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. S., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., *et al.* (1997). *AMBER 5*, University of California, San Francisco, CA.

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

Chipot, C. & Pohorille, A. (1998). Folding and translocation of the undecamer of poly-L-leucine across the water-hexane interface. A molecular dynamics study. *J. Am. Chem. Soc.* **120**, 11912-11924.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.

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

De Guzman, R. N., Turner, R. B. & Summers, M. F. (1998). Protein-RNA recognition. *Biopolymers,* **48**, 181-195.

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

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13-27.

Fixman, M. (1979). The Poisson-Boltzmann equation and its application to polyelectrolytes. *J. Chem. Phys.* **70**, 4995-5005.

Froloff, N., Windemuth, A. & Honig, B. (1997). On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. *Protein Sci.* **6**, 1293-1301.

Gubser, C. C. & 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, K. B. (1994). Interaction of RNA hairpins with the human U1a N-terminal RNA binding domain. *Biochemistry,* **33**, 10076-10088.

Hall, K. B. & Kranz, J. K. (1995). Thermodynamics and mutations in RNA-protein interactions. *Methods Enzymol.* **259**, 261-281.

Hall, K. B. & Stump, W. T. (1992). Interaction of N-terminal domain of U1a protein with an RNA stem loop. *Nucl. Acids Res.* **20**, 4283-4290.

Hermann, T. & Westhof, E. (1999). Simulations of the dynamics at an RNA-protein interface. *Nature Struct. Biol.* **6**, 540-544.

Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science,* **268**, 1144-1149.

Jayaram, B., McConnell, K. J., Dixit, S. B. & Beveridge, D. L. (1999). Free energy analysis of protein-DNA binding: the *Eco*RI endonuclease-DNA complex. *J. Comput. Phys.* **151**, 333-357.

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

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991). Identification of molecular contacts between the U1-a small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447-3456.

Kottalam, J. & Case, D. A. (1990). Langevin modes of macromolecules: applications to crambin and DNA hexamers. *Biopolymers,* **29**, 1409-1421.

Kranz, J. K. & Hall, K. B. (1998). RNA binding mediates the local cooperativity between the beta-sheet and the C-terminal tail of the human U1A RBD1 protein. *J. Mol. Biol.* **275**, 465-481.

Kranz, J. K., Lu, J. R. & Hall, K. B. (1996). Contribution of the tyrosines to the structure and function of the human U1a N-terminal RNA binding domain. *Protein Sci.* **5**, 1567-1583.

Lu, J. & Hall, K. B. (1997). Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry,* **36**, 10393-10406.

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

McQuarrie, D. A. (1976). *Statistical Mechanics*, Harper and Rowe, New York.

Misra, V. K., Hecht, J. L., Sharp, K. A., Friedman, R. A. & Honig, B. (1994). Salt effects on protein-DNA interactions - the lambda-CI repressor and *Eco*RI endonuclease. *J. Mol. Biol.* **238**, 264-280.

Misra, V. K., Hecht, J. L., Yang, A. S. & Honig, B. (1998). Electrostatic contributions to the binding free energy of the lambda cl repressor to DNA. *Biophys. J.* **75**, 2262-2273.

Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. & Varani, G. (1999). Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.* **294**, 967-979.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein-A. *Nature,* **348**, 515-520.

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.

Nicholls, A., Sharp, K. A. & Honig, B. (1990). *DelPhi*, Department of Biochemistry and Molecular Biophysics, Columbia University, NY.

Oubridge, C., Ito, H., Evans, P. R., Teo, C. H. & 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.

Reyes, C. M. & Kollman, P. A. (1999). Molecular dynamics studies of U1A-RNA complexes. *RNA,* **5**, 235-244.

Reyes, C. M. & Kollman, P. A. (2000). Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* **295**, 1-6.

Ryan, P. C. & Draper, D. E. (1989). Thermodynamics of protein RNA recognition in a highly conserved region of the large-subunit ribosomal RNA. *Biochemistry,* **28**, 9949-9956.

Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). Reduced surface - an efficient way to compute molecular surfaces. *Biopolymers,* **38**, 305-320.

Sharp, K. A. (1995). Polyelectrolyte electrostatics: salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson-Boltzmann model. *Biopolymers,* **36**, 227-243.

Sharp, K. A. & Honig, B. (1990a). Electrostatic interactions in macromolecules - theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301-332.

Sharp, K. A. & Honig, B. (1990b). Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* **94**, 7684-7692.

Shen, J. & Wendoloski, J. (1996). Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson-Boltzmann equation: assessment of its accuracy. *J. Comput. Chem.* **17**, 350-357.

Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978-1988.

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

Srinivasan, J., Miller, J., Kollman, P. A. & Case, D. A. (1998b). Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dynam.* **16**, 671-682.

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

Vorobjev, Y. N. & Hermans, J. (1999). ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* **78**, 195-205.

Vorobjev, Y. N., Almagro, J. C. & Hermans, J. (1998). Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins: Struct. Funct. Genet.* **32**, 399-413.

Williams, D. J. & Hall, K. B. (1996). RNA hairpins with non-nucleotide spacers bind efficiently to the human U1a protein. *J. Mol. Biol.* **257**, 265-275.

Zeng, Q. Y. & Hall, K. B. (1997). Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA,* **3**, 303-314.

Zhang, T. & Koshland, D. E. (1996). Computational method for relative binding energies of enzyme-substrate complexes. *Protein Sci.* **5**, 348-356.

Edited by B. Honig

(Received 18 October 1999; received in revised form 14 February 2000; accepted 18 February 2000)