Felicia Pitici
David L. Beveridge
Anne M. Baranger
Chemistry Department and Molecular Biophysics Program, Wesleyan University, Middletown,CT 06459

Received 17 April 2002; accepted 9 July 2002

Molecular Dynamics Simulation Studies of Induced Fit and Conformational Capture in U1A–RNA Binding: Do Molecular Substates Code for Specificity?

Abstract: *Molecular dynamics (MD) simulations on stem loop 2 of U1 small nuclear RNA and a construct of the U1A protein were carried out to obtain predictions of the structures for the unbound forms in solution and to elucidate dynamical aspects of induced fit upon binding. A crystal structure of the complex between the U1A protein and stem loop 2 RNA and an NMR structure for the uncomplexed form of the U1A protein are available from Oubridge et al. (*Nature*, 1994, Vol. 372, pp. 432–438) and Avis et al. (*Journal of Molecular Biology*, 1996, Vol. 257, pp. 398–411), respectively. As a consequence, U1A–RNA binding is a particularly attractive case for investigations of induced fit in protein–nucleic acid complexation. When combined with the available structural data, the results from simulations indicate that structural adaptation of U1A protein and RNA define distinct mechanisms for induced fit. For the protein, the calculations indicate that induced fit upon binding involves a non-native thermodynamic substate in which the structure is preorganized for binding. In contrast, induced fit of the RNA involves a distortion of the native structure in solution to an unstable form. However, the RNA solution structures predicted from simulation show evidence that structures in which groups of bases are favorably oriented for binding the U1A protein are thermally accessible. These results, which quantify with computational modeling recent proposals on induced fit and conformational capture by Leuillot and Varani (*Biochemistry*, 2001, Vol. 40, pp. 7947–7956) and by Williamson (*Nature Structural Biology*, 2000, Vol. 7, pp. 834–837) suggest an important role for intrinsic molecular architecture and substates other than the native form in the specificity of protein–RNA interactions.* Biopolymers 65: 424–435, 2002

Keywords: *RNA-binding protein; U1A; RNA recognition motif; ribonucleoprotein domain; molecular dynamics; induced fit; free energy*

INTRODUCTION

The formation of many protein–nucleic acid complexes involves alterations of the structure of the protein and the nucleic acid relative to their structures in the uncomplexed state. This phenomena is referred to in the literature as "structural adaptation" or "induced fit."[1,2] In some cases the conformation found in the complex may

Corrrespondence to: Anne M. Baranger; email: abaranger@wesleya.edu
Contract grant sponsor: NIH
Contract grant number: GM-56857and GM-37909
Biopolymers, Vol. 65, 424–435 (2002)

be a minor component of the equilibrium mixture of conformers found in the free state. In this case, "conformational capture" more accurately describes the conformational changes occurring in the binding event.[1] An understanding of the contribution of induced fit to complex affinity and specificity is often hindered by the lack of structural information on the bound and unbound forms of the components of the complex. Even if these are known, the energetics and structural dynamics of induced fit can be difficult, if not impossible, to isolate in experiments.

Induced fit of both the nucleic acid and protein is particularly common in protein–RNA complexes.[1,2] The free RNA structure is often highly flexible, particularly in single-stranded regions, and can undergo substantial rearrangements upon binding. The binding of the U1A protein to RNA is attractive for investigations of induced fit in protein–RNA complexation. The crystal structure of the complex formed between the U1A protein and stem loop 2 of U1 snRNA has been determined,[3] and an NMR structure of uncomplexed U1A protein has been reported.[4] Although the structure of the unbound stem loop 2 RNA has not been determined, molecular dynamics (MD) simulations performed by several laboratories have proposed structures that are in qualitative agreement.[5–7] Collectively, these structural and theoretical investigations indicate that there are significant differences in the conformations of the free and bound forms of both the U1A protein and RNA. We report in this paper a series of MD simulations that are designed to probe the energetics and structural dynamics of induced fit in the formation of the U1A–RNA complex. When combined with the available structural data, the results from simulations indicate that structural adaptation of the protein and RNA define distinct mechanisms for induced fit. This study of the U1A–RNA system suggests an important role for molecular architecture and substates other than the native state in the specificity of protein–RNA interactions.

BACKGROUND

The U1A protein binds to RNA using a RNA recognition motif (RRM), one of the most common RNA binding domains.[8] The RRM, also called the ribonucleoprotein (RNP) domain or the RNA binding domain (RBD), is comprised of a $\beta\alpha\beta\beta\alpha\beta$ sandwich fold that forms a four-stranded antiparallel β-sheet supported by two α-helices.[9] RRMs bind to single-stranded RNA target sites within many different structural contexts. The U1A protein is a component of the U1 snRNP, a subunit of the spliceosome, which splices most eukaryotic pre-mRNA.[10] The U1A protein contains two RRMs, but the C-terminal RRM is not required for RNA binding.[11,12] The N-terminal RRM of the U1A protein binds to stem loop 2 of U1 snRNA and two adjacent internal loops in the 3′-untranslated region of its own pre-mRNA with high affinity and specificity.[12,13] All three target sites contain nearly identical sequences in the loop, AUUGCAC, closed by a CG base pair. Extensive biochemical and structural experiments have probed RNA recognition by the N-terminal RRM of the U1A protein, making this complex one of the best-characterized RRM–RNA complexes.[12–26]

The structure of the N-terminal RRM of the U1A protein (residues 2–118) has been solved by NMR spectroscopy,[4] and the structure of the complex of the N-terminal RRM of the U1A protein (residues 2–102) with stem loop 2 of U1 snRNA has been solved by x-ray crystallography.[3] The structures of the U1A protein found in the complex and the free protein are superimposed in Figure 1a. These structures show that a secondary structural element, helix C, is in significantly different positions in the free protein and the complex. The crystal structure of the complex structure reveals a neat complementation between the U1A protein and RNA, with helix C oriented away from the β-sheet. In the NMR solution structure of the U1A protein, helix C lies across the RNA binding motif, impeding the approach of the RNA. The x-ray and NMR structures of the U1A protein are subsequently referred to as the "open" and "closed" forms of helix C, respectively. The U1A protein must undergo a structural adaptation in which helix C moves off the surface of the β-sheet to provide access for the RNA to the binding site. The role of this helix in complex formation has been the subject of previous considerations by Mittermaier et al. based on dynamic NMR studies of the free protein and complex, which suggested that coordinated changes in conformation and dynamical processes occur upon binding.[27] Fluorescence studies have suggested that the helix is bound to the sheet in the free protein, but retains considerable flexibility.[28] Residues in helix C have been found to contribute to binding via cooperative interactions.[14,29] Thus, helix C has been identified, independent of the crystal and NMR structures, as an important contributor to the stability and specificity of the U1A–RNA complex. An estimate of the relative stabilities of the U1A protein with the open and closed helix C orientations in the absence of RNA is necessary to further understand the functional energetics of induced fit to complex formation.

The nucleotide bases of the single-stranded loop region of stem loop 2 are splayed exterior to the loop

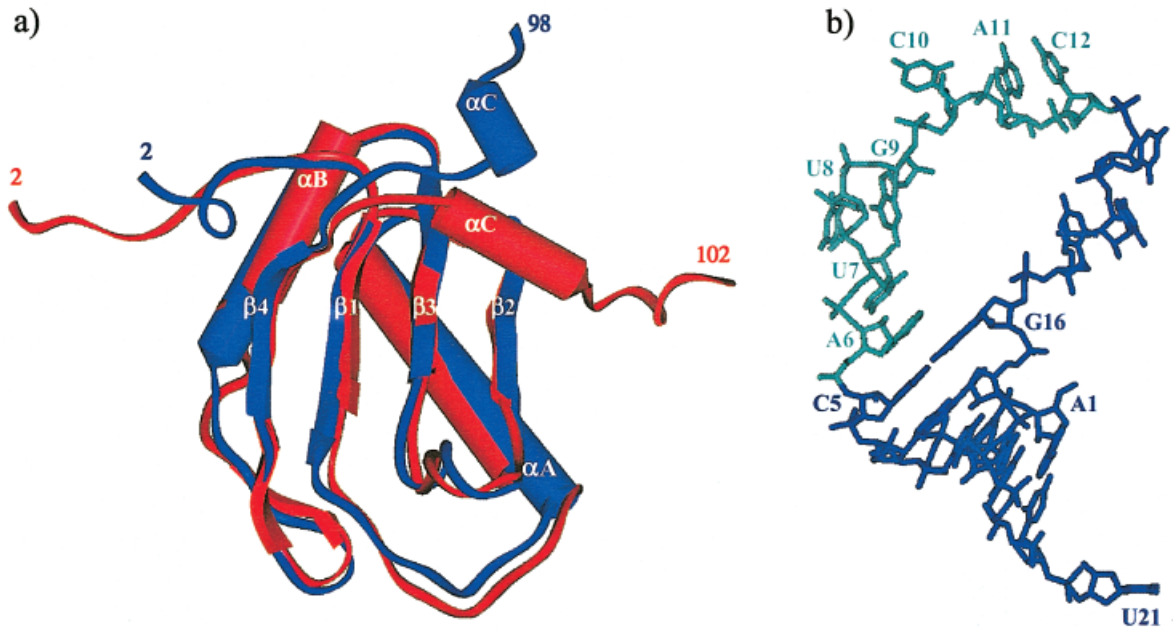

FIGURE 1 Experimentally determined structures of the U1A protein and the stem loop 2 RNA: (a) Superposition of the crystal structure of the bound form of the U1A protein (blue) and the NMR structure of the protein free in solution (red); and (b) crystal structure of the bound form of the RNA.[3,4] The nucleotides in the loop that contact the U1A protein are shown in cyan.

to contact the U1A protein in the crystal structure of the complex (Figure 1b). The corresponding structure of the unbound form is not known, but from an NMR structure of one of the internal loop target sites and MD simulations on stem loop 2, [5–7,30] the bases would be expected to be interior to the loop for the RNA free in solution. Thus a considerable structural adaptation or induced fit likely occurs in the RNA as well as in the U1A protein upon complex formation.

Additional information on induced fit is not readily obtained by experiment because it is difficult to investigate the conformational change from the free to bound structure of either the RNA or the U1A protein in the absence of the other component of the complex. However, these transitions can be probed using MD simulations. Recent developments in MD have resulted in improved capabilities for obtaining accurate all-atom models of the dynamical structure of bound and unbound forms of protein–nucleic acid complexes.[31,32] Several previous MD studies on the U1A–RNA complex have been reported[5–7,33–36] but do not specifically address the structural dynamics of induced fit at the level presented herein.

METHODS

Systems

The atomic coordinates for MD starting structures of the open and closed forms of the U1A protein were obtained from the crystal structure of the U1A–stem loop 2 complex[3] and the NMR solution structure of the unbound U1A protein, respectively.[4] Model 5 of the NMR data set was selected as representative for the unbound form of the U1A protein using *NMRCLUST 1.2*[37] with a root mean square (RMS) criterion for superimposing the C_α atoms of residues 10:98. For the stem loop RNA, the crystal structure of the bound form was taken as the MD starting structure. The x-ray and NMR structures were produced using U1A protein constructs comprised of residues 2:98 and 2:117, respectively. To enable the comparison with biochemical assays,[17,19] which used constructs comprised of residues 2:102, the model systems for simulations were adjusted for length to terminate at position 102. Six C-terminal residues were added to the fragment obtained from the crystal structure by homology modeling to a NMR structure of the complex formed by the U1A 2:102 protein with an internal loop RNA target site.[38] The modeling procedure included the selection of a template and the generation of new coordinates for the targeted region. The template was model 13 from the set of NMR structures, and it was selected with *NMRCLUST 1.2*[37] using a RMS fit for the C_α atoms of residues 93:102. Internal coordinates were transferred for the last six residues of the target, and side-chain atoms beyond C_β were built for Lys96, which was not well resolved in the crystal. Two point mutations, H31Y and R36Q, were also introduced into the model of the open form to revert its sequence to wild type. The fragment obtained from the NMR solution structure was truncated and capped with a carboxyl group.

Solvation effects were modeled by a periodic representation of rectangular cells that contain explicit TIP3 water

molecules.[39] The symmetry cells were chosen to ensure a minimal distance of 12 Å between the protein atoms and each face of the prism. Neutralizing Cl^- ions were assigned iteratively to sites of minimum electrostatic potential, and salt was subsequently added by randomly placing Na^+ and Cl^- ions more than 6 Å away from the solute or other ions. The resulting systems for the open/closed forms of the U1A protein included 101 amino acids, 7452/7773 water molecules, and (33, 40)/(35, 42) sets of (Na^+, Cl^-) ions, for a total of 23,571/24,526 atoms. For the stem loop 2 RNA, the system comprised 21 bases, 5266 water molecules, and (44, 24) sets of (Na^+, Cl^-) ions, for a total of 16,528 atoms.

Molecular Dynamics

MD simulations were carried out using the *AMBER 5.0* force field[40] with the parm96 set of parameters.[41] Energy terms were calculated using Ewald sums for long-range electrostatics[42] and a 9 Å cutoff for the direct part of the sums and for van der Waals interactions. High frequency motions involving hydrogen atoms were constrained with *SHAKE* at a tolerance of 10^{-4}Å.[43] Global rotations and translations were removed every 100 steps, and the corresponding energy was accounted for by scaling the atomic velocities. The list of nonbonded atom pairs was updated at 10 step intervals during MD, and every step during minimization.

The protocol in each instance involved an energy minimization of the initial structure (2000 steps), heating to 298 K (10 ps), equilibration at 298 K and 1 atm (50 ps), and the production run (to 5 ns). A round of optimization included 100 steps of steepest descent and 400 steps of the conjugated gradient method. Both minimization and equilibration were conducted gradually by releasing initial harmonic constraints on the protein (25 kcal/mol) and on neutralizing ions (20 kcal/mol). The schedules for removing the restraints involved decrements of 10 kcal/mol (5 kcal/mol in the end) every 500 steps or every 10 ps. The first 10 ps of equilibration were conducted at constant (E, V), after which the system was coupled to a (T, p) reservoir at exchange intervals of 0.2 ps and with different scaling factors for velocities of solute and solvent atoms.[39] The production run was continued to 5 ns under weaker coupling conditions, of 0.5 ps for temperature and 1 ps for pressure. The integration step for simulation was 2 fs.

Free Energy Analysis

Free energy component analysis is a computational method of making free energy estimates for macromolecular structures from a sum of contributions from electrostatic effects, van der Waals interactions, the hydrophobic effect, and various entropic terms, each either calculated as well as possible from force fields or obtained semiempirically from experimental measures.[44] The theoretical basis of this approach is described fully in a series of recent papers on protein DNA complexes and a theoretical account of the method based on statistical thermodynamics has been provided along with considerations on the capabilities and limitations of the calculations, including the uncertainties and approximations associated with estimates of the individual terms.[44,45] The method provides a computational model for the diverse contributions to net free energies, but the propagation of uncertainties in the summation process suggests caution in making nuanced interpretation of the results. In general, the results on dynamical structure from MD should be considered more reliable than the calculated free energies. The implementation of free energy component analysis in this project is similar to that utilized in a previously published calculation on the U1A–RNA complex, except in this project the analysis was performed with the ensembles of states generated from MD simulations.[5]

RESULTS

The results of the MD simulations on the U1A protein and stem loop 2 RNA are presented here in terms of two measures: the root mean square distance (RMSD) between optimally aligned structures and the superposition of a set of structures representative of the Boltzmann ensemble produced in the course of a MD trajectory. The RMSD is presented in two forms: (a) as a time series that shows the deviation of the MD structures from the starting structure as the simulation progresses (1D-RMSD), and (b) as 2D-RMSD plots that show the RMSD of all MD structures from all others during the course of the simulation. On the 2D-RMSD plots, the various shadings indicate structures with a given similarity in overall RMSD values, with darker regions on the maps associated with lower RMSD and more similar structures.

U1A Protein

Two simulations were performed, one beginning with the U1A protein structure found in the complex in which helix C is oriented away from the β-sheet, the open form, and one beginning with the U1A protein structure of the free protein in which helix C contacts the surface of the β-sheet, the closed form.[3,4] Both simulations were carried out for 5 ns. The 1D- and 2D-RMSD analyses of the MD simulations beginning with the open form are shown in Figure 2. A set of snapshots from this trajectory are shown in Figure 3a. Here helix C shows the largest dynamical range of motion in the protein, but is observed to generally remain away from the surface of the β-sheet, as observed in the crystal structure of the complex.

The 1D- and 2D-RMSD analyses of the MD simulations beginning with the closed form of the free U1A protein are shown in Figure 2 and a superposition of a set of MD structures is shown in Figure 3b.

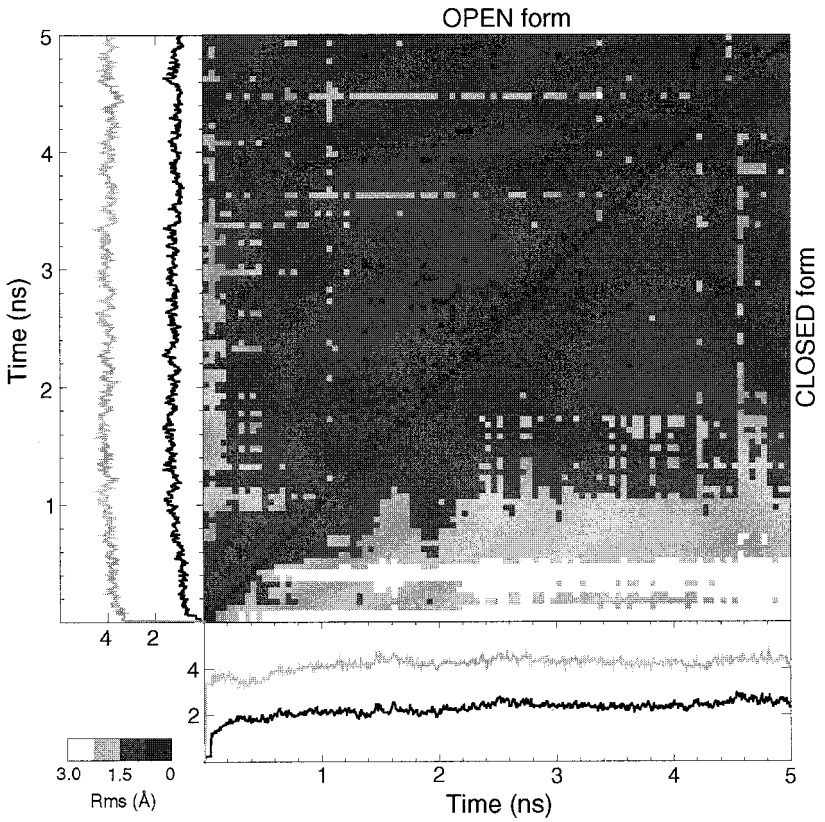

FIGURE 2 RMSD as a function of time computed from MD simulations on the U1A protein. Upper left *xy* plot and triangle of matrix: 1D- and 2D-RMSD for the MD simulation initiated with the open form of the U1A protein found in the crystal structure of the complex,[3] Lower right *xy* plot and triangle of matrix: 1D- and 2D-RMSD for MD simulation initiated with the closed form of the U1A protein found in the NMR structure of the free protein.[4] In both 1D-RMSD *xy* plots, the black lines indicate the RMSD of the simulated system from the initial structure used for that simulation (i.e., MD beginning at open form and compared with the open form starting structure) and the gray lines indicate the RMSD of the simulated system from the initial structure used for the other simulation (i.e. MD beginning at open form but compared with the closed form NMR structure). The gray scale in the 2D RMSD plots ranges from 0 to 3.0 Å, with darker regions associated with more similar structures. Calculations include the Cα atoms of residues 10:98.

The 1D RMSD results indicate that the protein equilibrates after ~1.5 ns of simulation to a stable dynamical structure and the 2D-RMSD plot (lower triangle of matrix) shows that the dynamical structure from MD also oscillates in the vicinity of a single structural form for the duration of the simulation. The MD structures are observed to maintain the closed form of helix C and end up 2.3 Å RMS from the starting NMR solution structure. The most significant implication of these results is that there is no thermal interconversion of the open and closed forms of the U1A protein in either MD simulation, indicating that at ambient temperature and over the 5 ns of MD, the two forms of helix C in the free U1A protein correspond to two different substates in the underlying potential energy surface, separated on the average by 4.6 Å RMSD. A more detailed comparison of the results of the two separate MD simulations on the U1A protein is shown in Figure 4, in which the RMSD fluctuation by residue for the two forms is plotted. Both structures are quite stable in the regions of $\beta\alpha\beta\beta\alpha\beta$ secondary structure. The N-terminal and C-terminal regions show the largest fluctuations. The loop 3 residues 45–52 appear to be a somewhat flexible hinge in the protein, consistent with experimental observations.[27,46]

Stem Loop 2 RNA

The RMSD plots for a 5 ns MD simulation on free stem loop RNA in solution beginning with the structure found in the complex are shown in Figure 5. The

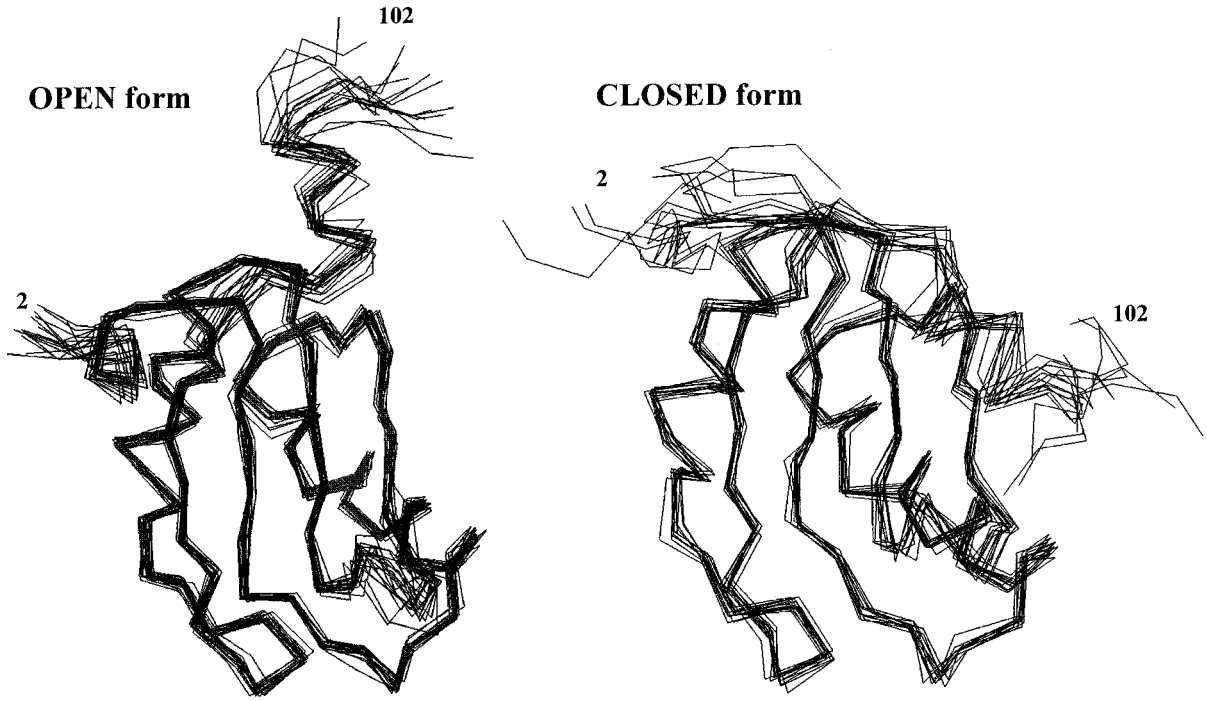

FIGURE 3 Superposition of MD snapshots from the two MD simulations on the U1A protein: (a) the MD simulation of the open form of the U1A protein and (b) the MD simulation of the closed form of the U1A protein. The clusters include snapshots extracted at a frequency of 250 ps from the production part of each trajectory. The structures were oriented by superimposing the Cα atoms of residues 10:98.

1D-RMSD shows that during equilibration the structure moves rapidly from the initial form to a state ~4 Å RMSD away. Changes thereafter are localized in the more flexible loop region. The initial structure and a superposition of structures produced during the MD trajectory are shown in Figure 6. In the bound form of the RNA,[3] which served as the initial structure in the simulation, the nucleotide bases involved in protein–RNA contacts are oriented toward the exterior of the loop. In the dynamical structure of the equilibrated form of the stem loop RNA many of these bases are oriented toward the interior. This MD prediction of the solution structure of stem loop RNA is similar to the structures found independently by MD simulations[6,7] and is consistent with the solution structure of a related internal loop target site of the U1A protein.[30] In summary of the stem loop 2 RNA results, the MD indicates that the bound form of stem loop 2 RNA is an unstable structure that in the absence of protein relaxes rapidly to the (predicted) solution structure form, which is distinctly different than the bound form in the loop region. This indicates that U1A–RNA induced fit in the case of the stem loop 2 RNA is essentially a matter of simple distortion of the dynamical structure of the RNA free in solution.

Examination of the dynamical structure of the RNA more closely reveals additional detail relevant to the protein–RNA binding event. The time series of MD structures of the RNA indicates that the nucleotide bases of the loop region are not always interior to the loop as expected, but there is a dynamic equilibrium between structures with bases interior and exterior to the loop (Figure 7). Comparing the MD structures in Figure 7 to the crystal structure of the bound form of RNA, the bases C10, A11, and C12 are well positioned to make contacts with the U1A protein upon binding. Although the structures from simulations differ from that observed in the crystal structure of the bound form (5.7 Å RMSD difference), the results indicate a possible case for some thermally accessible preorganization of the RNA bases prior to complexation.

Free Energy Analysis

Free energy component analysis was used to estimate the relative stability of the open and closed forms of the U1A protein based on the dynamical structures obtained from simulations. In this calculation, structures were extracted from the MD trajectories at 2 ps

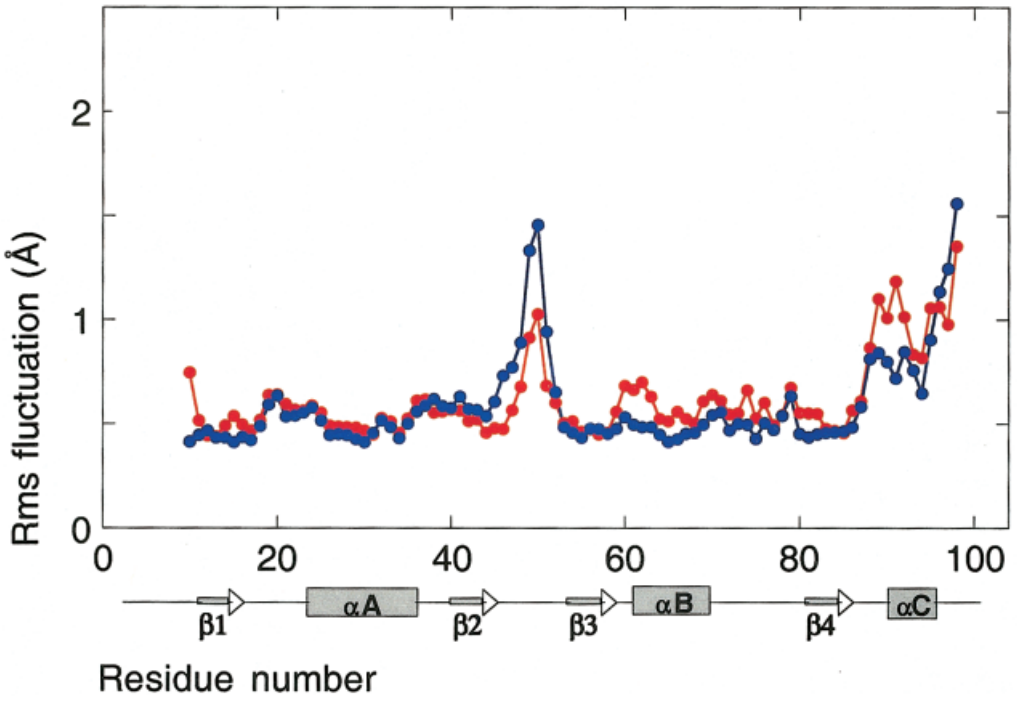

FIGURE 4 RMSD by amino acid residue of the average positions reached in the MD simulations for the open (blue) and closed (red) forms of the U1A protein relative to the starting structures. The trajectory frames were oriented relative to the starting structure using a least-squares fit of the Cα atoms of residues 10:98. The secondary structural elements of the protein are displayed along the *x* axis.

intervals and used as a basis for free energy analysis using protocols applied to a series of protein–DNA binding studies, described in detail elsewhere.[44,45] The calculated conformational free energy of the two forms including both solute and solvent components is shown in Figure 8a. The results for the two forms of the U1A protein show substantial overlap, as can be seen in the cumulative distributions in Figure 8b. The computed trend shows a preferential stability for the open form of the uncomplexed protein in contrast to the NMR results, which favor the closed form in solution. However, the intrinsic uncertainties in the free energy estimates do not permit us an unequivocal claim in this matter, and the considerable overlap in the calculated distributions is consistent with the free energies of the open and closed forms being fairly close. Full numerical results of the free energy component analysis are provided in Table I.

DISCUSSION

The results described above show that stable MD trajectories were obtained for the U1A protein and stem loop 2 RNA in solution. Comparing the behavior of the MD simulations beginning with the structures of the components in the complex as determined by crystallography and the structure of the free U1A protein in solution as determined by NMR,[3,4] we find that the protein and nucleic acid use significantly different mechanisms for induced fit, illustrated schematically in Figure 9. A quantitative description of the molecular substates involved will be reported along with further studies on mutants elsewhere.[47] The MD simulation of the U1A protein starting with the structure found in the complex relaxes from the initial structure C_{U1A} to a nearby substate B_{U1A}, in which helix C is still oriented away from the surface of the β-sheet. The MD simulation of the U1A protein starting with the NMR structure relaxes from the initial structure N_{U1A} to an equilibrated form of the solution structure, F_{U1A}, in which helix C remains across the surface of the β-sheet. These forms are not thermally interconvertable under the conditions of the simulations. Therefore, B_{U1A} and F_{U1A} are indicated to be separate substates that are separated by 4.6 Å by the MD, the difference being almost entirely due to the position of helix C. If the MD results are accurate, induced fit of the U1A protein upon complex formation involves a form of conformational capture.[1]

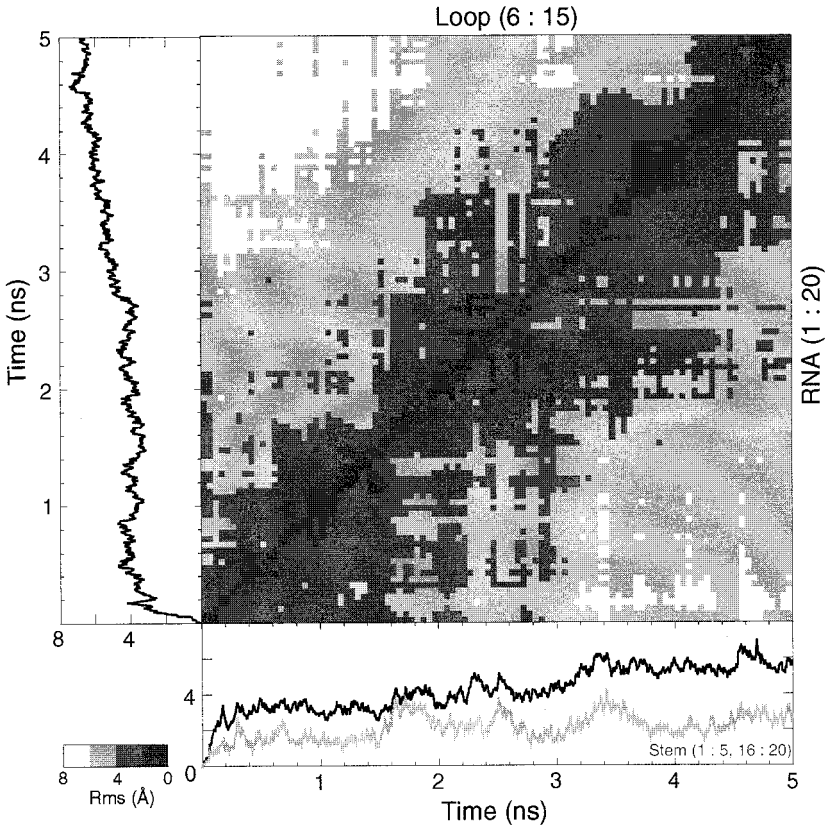

FIGURE 5 RMSD as a function of time computed from MD simulations on stem loop 2 RNA. Upper triangle of matrix: 1D- and 2D-RMSD vs time for RNA loop atoms only (6:15); Lower triangle of matrix: 1D RMSD for RNA all atoms (1:20, black line) and for stem atoms only (1:5 and 16:20, gray line) and 2D RMSD for all RNA atoms. The gray scale in the 2D RMSD plots ranges from 0 to 8.0 Å, with darker regions indicative of more similar structures.

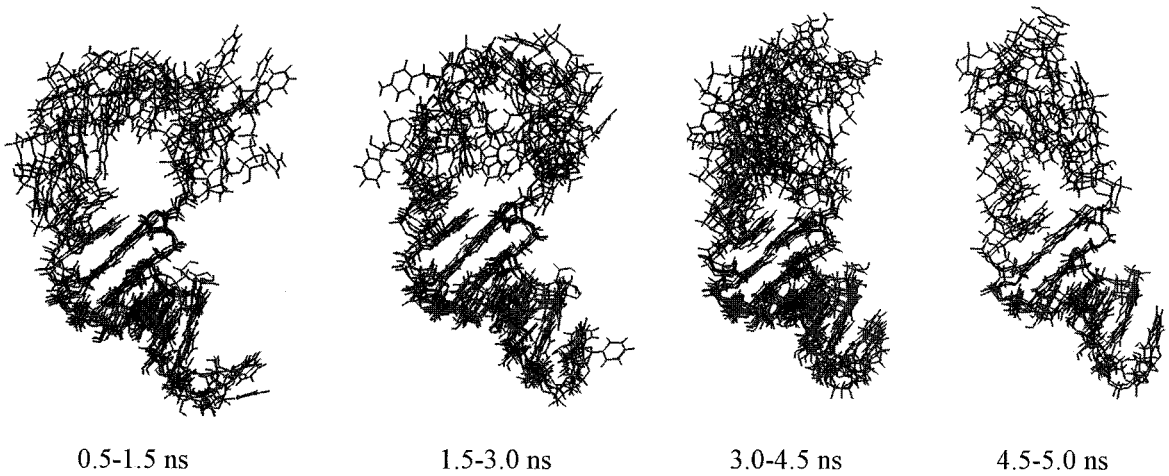

FIGURE 6 MD snapshots for stem loop 2 RNA as a function of time over the course of a 5 ns trajectory. The clusters include snapshots extracted at a frequency of 250 ps from the production part of each trajectory. Bundles of structures are superpositions over the segments of the trajectory noted on the figure. The structures were oriented by superimposing the heavy atoms of bases in the stem region.

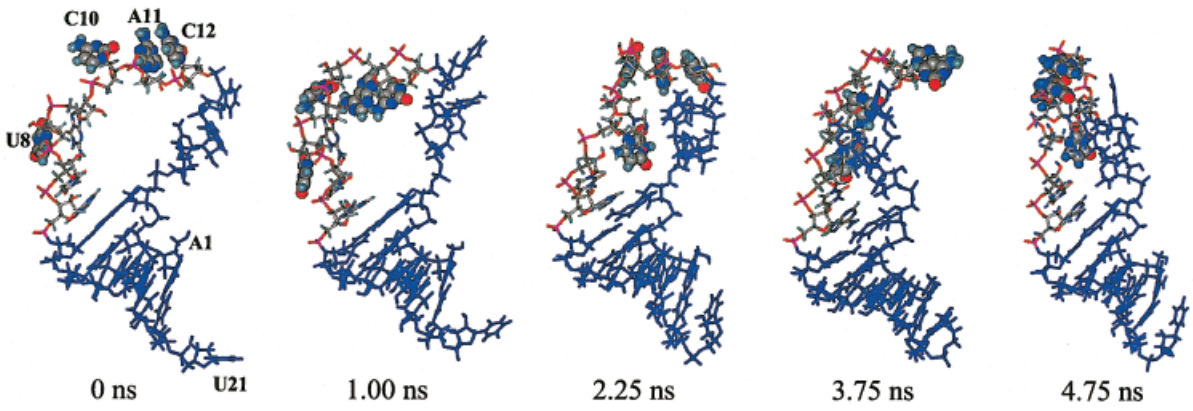

FIGURE 7 MD structures of stem loop 2 RNA at time points that are the midpoints of the bundles shown in Figure 6. Bases U8, C10, A11, and C12 are shown.

Rather than a simple distortion to an otherwise unstable state, a transition to an intermediate B_{U1A} that is preorganized for binding occurs, followed by the binding event in a sequential or concerted mechanism. Although the NMR solution structure of the U1A protein places helix C in the closed position, there is experimental evidence that helix C is dynamic and weakly bound to the surface of the β-sheet.[27,28] NMR experiments showed that residues within helix C, the junction between helix C and the end of the β-sheet, and the surface of the β-sheet undergo significant conformational exchange.[27] The interaction of helix C with the surface of the β-sheet was also studied using fluorescence experiments, in which a Trp was substituted for Phe56, one of the residues on the surface of the β-sheet that contacts helix C.[28] These experiments suggested that helix C binds to the β-sheet, removing Trp56 from the solvent, but that this interaction is dynamic on a nanosecond or longer time scale. Both the NMR dynamics experiments and the fluorescence experiments were performed on the U1A protein comprised of residues 2–102 that was used our computational study. However, the solution structure was obtained with a longer peptide that was comprised of residues 2–118. It has been suggested that the additional C-terminal amino acids stabilize the interaction of helix C with the surface of the β-sheet; however, equilibrium binding experiments have shown that the two fragments of U1A protein bind with identical affinity and specificity to stem loop 2.[29]

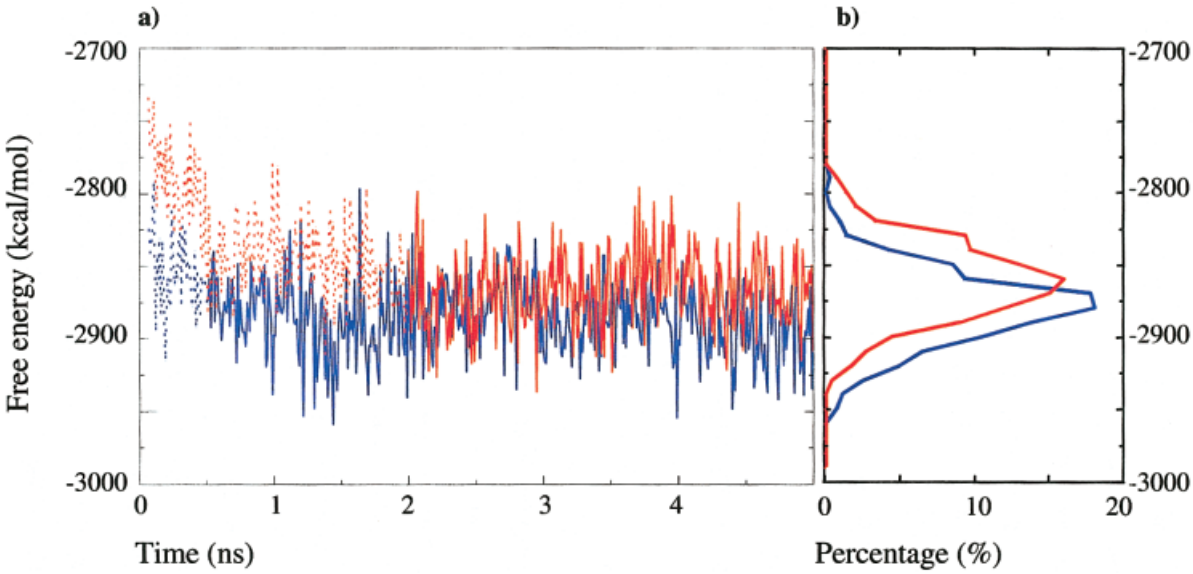

FIGURE 8 Results of free energy component analysis applied to the MD structures for the open (blue) and closed (red) forms of the U1A protein: (a) time series and (b) integrated distributions.

Table I Free Energy Contributions to the Stability of U1A

Component		Closed	Open	Closed–Open
Intramolecular energy	H_{int}	−1316.4 (1.7)	−1134.3 (1.1)	−182.1 (2.8)
Bond and torsion	$H_{int}^{bd,ts}$	1653.1 (0.8)	1651.4 (0.5)	1.7 (1.3)
Electrostatic	H_{int}^{el}	−2589.8 (1.7)	−2378.1 (1.1)	−211.7 (2.8)
Van der Waals	H_{iar}^{vdWI}	−379.7 (0.5)	−407.6 (0.3)	27.9 (0.8)
Solvation free energy	ΔG_{solv}	−1547.3 (1.4)	−1749.0 (0.9)	201.7 (2.3)
Electrostatic	ΔG_{solv}^{el}	−1582.3 (1.4)	−1782.3 (0.9)	200.0 (2.3)
Added salt	$\Delta G_{solv}^{el:salt}$	−13.6 (0.0)	−14.3 (0.0)	0.7 (0.0)
Van der Waals	ΔG_{solv}^{vdW}	−268.8 (0.1)	−262.9 (0.1)	−5.9 (0.2)
Cavity	ΔG_{solv}^{cav}	317.4 (0.2)	310.5 (0.1)	6.9 (0.3)
Total free energy	ΔG	−2863.7 (1.0)	−2883.3 (0.7)	19.6 (1.7)

[a]Individual terms for the closed and open forms represent averages over the trajectory intervals sampled at 2 ps after 0.5 and 3 ns of simulation, respectively. The last column includes differences between the free energies of the closed and open forms. Average values and standard errors are expressed in kcal/mol.

The induced fit of stem loop 2 RNA upon complexation falls into the category illustrated schematically in Figure 9. The mechanism followed is that of a simple distortion of the native form of the RNA in which most bases are interior to the loop, N_{RNA}, to a form, C_{RNA}, in which the bases are exterior to the loop. C_{RNA} would be unstable in the absence of protein. The evidence for this mechanism is the rapid relaxation of the protein-bound form of the RNA to the solution structure that was observed in the MD simulations and is shown in Figures 5–7. The native form of the RNA in solution calculated by MD shows that structures with bases exterior to the loop region are thermally accessible and are in dynamical equi-

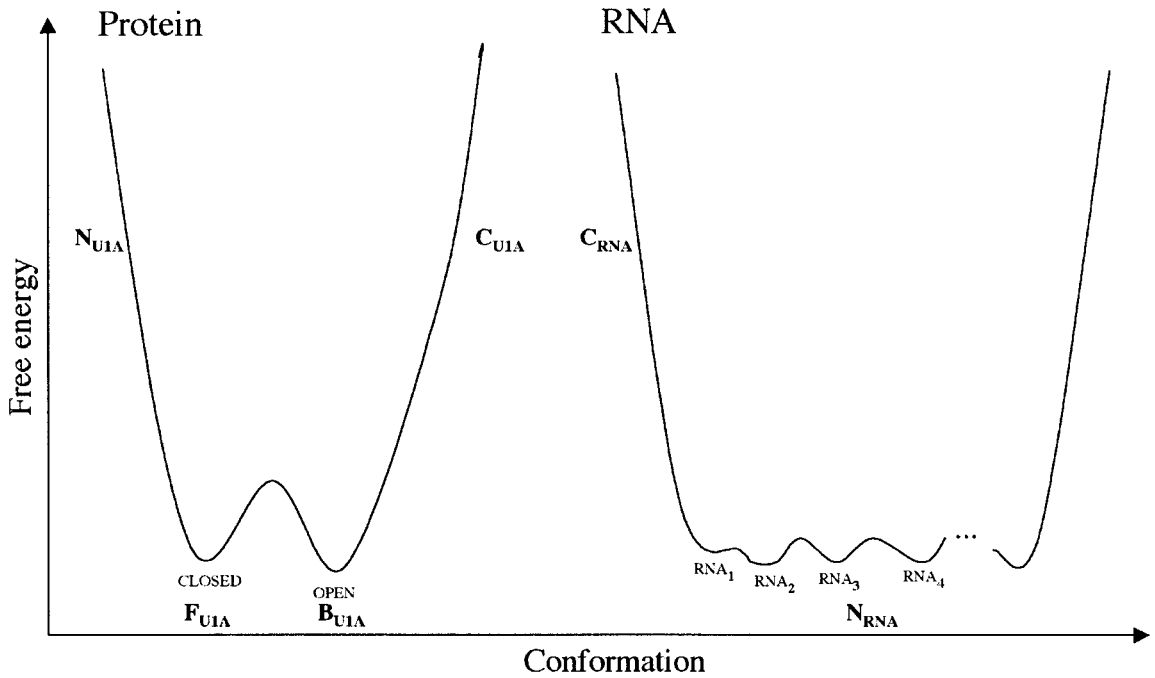

FIGURE 9 Schematic diagram of induced fit in the complexation of the U1A protein and stem loop 2 RNA implied by the MD simulations. C_{U1A} is the open form of the U1A protein found in the crystal structure of the complex,[3] B_{U1A} is the MD structure of the open form of the U1A protein, N_{U1A} is the closed form of the U1A protein found in the NMR structure of the free protein,[4] F_{U1A} is the MD structure of the closed form of the U1A protein, C_{RNA} is the structure of stem loop 2 RNA found in the crystal structure of the complex, and N_{RNA} is the MD structure of stem loop 2 RNA.

librium with structures with the bases interior to the loop. It has been suggested that the flexibility of the RNA may enable a more intimate interaction of the RNA and protein, and may enable the RNA to bind to other protein target sites.[1,2] The difference between the mechanism of induced fit in the protein and nucleic acid in the U1A–RNA complex is essentially a matter of thermal accessibility of substates preorganized for complexation and lends overall support to the idea that specificity and biological functions of protein–RNA complexes, as noted independently by Leuillot and Varani[1] and by Williamson,[2] can be controlled by induced fit and conformational capture mechanisms.

CONCLUSIONS

The MD simulations reported in this paper suggest that the conformation of the U1A protein that binds to RNA is a stable substate of the protein structure, while the conformational change of the RNA upon binding the U1A protein is an energetically unfavorable distortion of the stable solution structure. These different mechanisms of structural adaptation would be expected to have considerably different impacts on binding affinity and specificity.[1,2] The stem loop RNA is dynamic in the absence of the U1A protein and upon binding undergoes not only an energetically unfavorable conformational distortion, but an entropically unfavorable conformational restriction. Both of these factors would be expected to hinder binding. This destabilization would be countered by the direct interaction energy of protein–RNA binding, solvent release, and an increase in vibrational entropy from the low frequency modes unique to the complex. By contrast, the existence of a U1A protein substate preorganized for binding cognate RNA, while costly per se, is also paid for by complexation energy and vibrational entropy. It is interesting to note that this mechanism for induced fit is rooted in the internal architecture of the protein and RNA and therefore, must be established during evolutionary development. In this sense, the presence of functional non-native substates in the molecular architecture argues in favor of the idea that substates code for specificity in this system. If the binding of the U1A protein to a noncognate RNA requires a different U1A protein conformation and this different conformation is not a stable substate of the U1A protein, then binding to the noncognate RNA would be less favorable than binding to cognate RNA. Thus, the particular mechanism of induced fit proposed for the U1A protein in this study would improve the specificity of protein–RNA binding.

Further studies, both experimental and theoretical, are in progress on this system. We have completed corresponding MD simulations on four mutants for which binding affinities have been determined. These simulations are being analyzed with respect to predicted structural changes on mutation of bound and unbound species and for differential changes in the dynamics of substates.[47]

Funding was provided by the NIH to AMB, GM-56857, and to DLB, GM-37909. AMB is an Alfred P. Sloan Research Fellow.

REFERENCES

1. Leulliot, N.; Varani, G. Biochemistry 2001, 40, 7947–7956.
2. Williamson, J. R. Nature Struct Biol 2000, 7, 834–837.
3. Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Nature 1994, 372, 432–438.
4. Avis, J. M.; Allain, F. H.-T.; Howe, P. W. A.; Varani, G.; Nagai, K.; Neuhaus, D. J Mol Biol 1996, 257, 398–411.
5. Blakaj, D. M.; McConnell, K. J.; Beveridge, D. L.; Baranger, A. M. J Am Chem Soc 2001, 123, 2548–2551.
6. Tang, Y.; Nilsson, L. Biophys J 1999, 77, 1284–1305.
7. Reyes, C. M.; Kollman, P. A. J Mol Biol 2000, 297, 1145–1158.
8. Varani, G.; Nagai, K. Ann Rev Biophys Biomol Struct 1998, 27, 407–445.
9. Burd, C. G.; Dreyfuss, G. Science 1994, 265, 615–621.
10. Stark, H.; Dube, P.; Lührmann, R.; Kastner, B. Nature 2001, 409, 539–542.
11. Lu, J.; Hall, K. B. J Mol Biol 1995, 247, 739–752.
12. Scherly, D.; Boelens, W.; van Venrooij, W. J.; Dathan, N. A.; Hamm, J.; Mattaj, I. W. EMBO J 1989, 8, 4163–4170.
13. van Gelder, C. W. G.; Gunderson, S. I.; Jansen, E. J. R.; Boelens, W. C.; Polycarpu-Schwarz, M.; Mattaj, I. W.; van Venrooij, W. J. EMBO J 1993, 12, 5191–5200.
14. Kranz, J. K.; Hall, K. B. J Mol Biol 1998, 275, 465–481.
15. Kranz, J. K.; Hall, K. B. J Mol Biol 1999, 285, 215–231.
16. Katsamba, P. S.; Myszka, D. G.; Laird-Offringa, I. A. J Biol Chem 2001, 276, 21476–21481.
17. Shiels, J. C.; Tuite, J. B.; Nolan, S. J.; Baranger, A. M. Nucleic Acids Res 2002, 30, 550–558.
18. Luchansky, S. J.; Nolan, S. J.; Baranger, A. M. J Am Chem Soc 2000, 122, 7130–7131.
19. Nolan, S. J.; Shiels, J. C.; Tuite, J. B.; Cecere, K. L.; Baranger, A. M. J Am Chem Soc 1999, 121, 8951–8952.

20. Hall, K. B. Biochemistry 1994, 33, 10076–10088.
21. Rimmele, M.; Belasco, J. RNA 1998, 4, 1386–1396.
22. Grainger, R. J.; Murchi, A. I. H.; Norman, D. G.; Lilley, D. M. J. J Mol Biol 1997, 273, 84–92.
23. Laird-Offringa, I. A.; Belasco, J. G. Proc Natl Acad Sci USA 1995, 92, 1859–11863.
24. Jessen, T.-H.; Oubridge, C.; Teo, C. H.; Pritchard, C.; Nagai, K. EMBO J 1991, 10, 3447–3456.
25. Tsai, D. E.; Harper, D. S.; Keene, J. D. Nucleic Acids Res 1991, 19, 4931–4936.
26. Bentley, R. C.; Keene, J. D. Mol Cell Biol 1991, 11, 1829–1839.
27. Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. J Mol Biol 1999, 294, 967–979.
28. Jean, J. M.; Clerte, C.; Hall, K. B. Protein Sci 1999, 8, 2110–2120.
29. Zeng, Q.; Hall, K. B. RNA 1997, 3, 303–314.
30. Gubser, C. C.; Varani, G. Biochemistry 1996, 35, 2253–2267..
31. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182–196.
32. Cheatham, T. E.; Young, M. A. Biopolymers 2001, 56, 232–256.
33. Hermann, T.; Westhof, E. Nat Struct Biol 1999, 6, 540–544.
34. Reyes, C. M.; Kollman, P. A. RNA 1999, 5, 235–244.
35. Reyes, C. M.; Kollman, P. A. J Mol Biol 2000, 295, 1–6.
36. Olson, M. A. Biophys J 2001, 81, 1841–1853.
37. Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. Protein Eng 1996, 9, 1063–1065.
38. Allain, F. H.; Howe, P. W. A.; Neuhaus, D.; Varani, G. EMBO J 1997, 16, 5764–5774.
39. Jorgensen, W. L.; Chadrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926–935.
40. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. AMBER: Version 5; 5.0 ed.; University of California: San Francisco, 1997.
41. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179–5197.
42. Darden, T. A.; York, D. M.; Pedersen, L. G. J Chem Phys 1993, 98, 10089–10092.
43. van Gunsteren, W. F.; Berendsen, H. J. C. Angew Chem 1990, 29, 992–1023.
44. Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. J Comp Phys 1999, 151, 333–357.
45. Jayaram, B.; Mcconnell, K.; Dixit, S. B.; Das, A.; Beveridge, D. L. J Comput Chem 2002, 23, 1–14.
46. Lu, J.; Hall, K. B. Biochemistry 1997, 36, 10393–10405.
47. Pitici, F.; Baranger, A. M.; Beveridge, D. L. Manuscript in preparation.