

Molecular Dynamics Simulation of the RNA Complex of a Double-Stranded RNA-Binding Domain Reveals Dynamic Features of the Intermolecular Interface and Its Hydration

Tiziana Castrignanò,* Giovanni Chillemi,* Gabriele Varani,[†] and Alessandro Desideri[‡]

*Consorzio interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, University of Rome "La Sapienza", 00185 Rome, Italy, [†]Department of Biochemistry and Department of Chemistry, University of Washington, Seattle, Washington 98195-1700, USA, and [‡]Istituto Nazionale per la Fisica della Materia and Department of Biology, University of Rome "Tor Vergata", 00133 Rome, Italy

ABSTRACT The interaction between double-stranded RNA (dsRNA) and the third double-stranded domain (dsRBD) from *Drosophila* Staufen protein represents a paradigm to understand how the dsRBD protein family, one of the most common RNA-binding protein units, binds dsRNA. The nuclear magnetic resonance (NMR) structure of this complex and the x-ray structure of another family member revealed the stereochemical basis for recognition, but also raised new questions. Although the crystallographic studies revealed a highly ordered interface containing numerous water-mediated contacts, NMR suggested extensive residual motion at the interface. To address how interfacial motion contributes to molecular recognition in the dsRBD–dsRNA system, we conducted a 2-ns molecular dynamics simulation of the complex derived from Staufen protein and of the separate protein and RNA components. The results support the observation that a high degree of conformational flexibility is retained upon complex formation and that this involves interfacial residues that are critical for dsRBD–dsRNA binding. The structural origin of this residual flexibility is revealed by the analysis of the trajectory of motion. Individual basic side chains switch continuously from one RNA polar group to another with a residence time seldom exceeding 100 ps, while retaining favorable interaction with RNA throughout much of the simulation. Short-lived water molecules mediate some of these interactions for a large fraction of the trajectory studied here. This result indicates that water molecules are not statically associated with the interface, but continuously exchange with the bulk solvent on a 1–10-ps time scale. This work provides new insight into dsRBD–dsRNA recognition and builds upon a growing body of evidence, suggesting that short-lived dynamic interactions play important roles in protein–nucleic acid interactions.

INTRODUCTION

The double-stranded RNA (dsRNA)-binding domain is among the most common RNA-binding motifs (Varani, 1997; Varani and Nagai 1998) and is found in many proteins from all kingdoms of life involved in RNA processing, maturation, and localization (Green and Matthews, 1992; St Johnston et al., 1992). In vitro studies have shown that the double-stranded domain (dsRBD) proteins bind to dsRNA but not to dsDNA and only weakly to DNA–RNA hybrids (St Johnston et al., 1992; Bass et al., 1994; Clarke and Matthews, 1995; Bevilacqua and Cech, 1996). dsRBDs bind dsRNA of sufficient length (more than 12 base pairs) regardless of its base composition, and therefore represent general dsRNA binding units. Therefore, binding of the dsRBD to dsRNA represents an example of structure-specific but sequence-independent protein–RNA recognition that is distinct from that of other common RNA-binding motifs. Consistent with its biochemical activity, many dsRBD-containing proteins bind a wide variety of RNAs in the cell, although some dsRBDs exercise their function on a specific set of RNAs (Macdonald and Struhl, 1988; St

Johnston et al., 1991; Kim-Ha et al., 1991, 1993; Ferrandon et al., 1994; St Johnston, 1995; Broadus et al., 1998). How this domain can discriminate so effectively between double-stranded nucleic acids of similar structures, and how a domain with a general ability to bind dsRNA regardless of its sequence can function in the metabolism of specific mRNAs in the cell are outstanding questions that must be addressed if we are to understand the function of this ubiquitous RNA-binding module. These questions have begun to be addressed by the three-dimensional structure of dsRBDs (Bycroft et al., 1995; Kharrat et al., 1995) and especially by two complex structures (Ryter and Schultz, 1998; Ramos et al., 2000).

The x-ray structure of the complex of *Xenopus* Xlrbpa with an RNA duplex (Ryter and Schultz, 1998) and the nuclear magnetic resonance (NMR) structure of the third dsRBD from *Drosophila* Staufen protein in complex with a stem-loop (Ramos et al., 2000) revealed how the dsRBD binds dsRNA. Together, the two structures provide a satisfactory explanation for the ability of the domain to bind dsRNA and discriminate against dsDNA or hybrids. Discrimination in favor of dsRNA and against dsDNA is attributable to interactions mediated by loop 2 (which binds the RNA minor groove, making RNA-specific contacts with 2'-OH) and by loop 4 (where conserved basic side chains make structure-specific interactions with the phosphodiester backbone across the major groove from the site of loop 2-minor groove contacts). Interactions provided by helix α 1

Submitted May 9, 2002 and accepted for publication July 23, 2002.

Address reprint requests to A. Desideri, Dept. of Biology, Univ. of Rome Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy. Tel.: +39-06-72594376; Fax: +39-06-72594326; E-mail: desideri@uniroma2.it.

© 2002 by the Biophysical Society

0006-3495/02/12/3542/11 \$2.00

necessary for binding reinforce these interactions. Mutations within each of these three regions of the protein or in a conserved residue that positions loop 2 and loop 4 with respect to each other severely diminish or abolish the RNA-binding activity. Mutations of these same residues abolish the biological activity of at least one dsRBD-containing protein, *Drosophila* Staufen, establishing the functional significance of these structural observations.

Two significant differences between the NMR and crystal structure revealed the highly dynamic nature of the protein–RNA interface. The degree of order at the interface and the observation of well-ordered interfacial water molecules in the crystal structure reflect (at least in part) the lower quality of NMR structures when compared to high-resolution crystal structures. However, NMR studies of backbone dynamics also revealed unambiguously how regions of the protein critical for RNA recognition (loop 2 and loop 4) retain significant conformational flexibility in the RNA-bound protein (Ramos et al., 2000).

Residual flexibility is increasingly being appreciated as an important element in protein–nucleic acid recognition. However, quantitative properties of such motions have only recently begun to be addressed experimentally through novel NMR methods and computationally through molecular dynamics simulations. Although NMR has a unique ability to investigate experimentally motional properties of biomolecules, it is difficult to provide detailed insight into motional properties because the spectral density function is only sampled by the NMR experiment at a few frequencies (Peng and Wagner, 1992, 1994). Molecular dynamics (MD) techniques are simulations, but are capable of providing detailed insight into the motions that occur during molecular recognition and how motional properties of biomolecules change upon binding (Beveridge and Ravinshanker 1994; Cheatham et al., 1995, 1997; Young and Beveridge, 1998; Hermann and Westhof 1999; Reyes and Kollmann, 1999; Sen and Nilsson, 1999; Tang and Nilsson, 1999; Castri-gnanò et al., 2000; Tsui et al., 2000). The two methods synergistically reinforce each other when applied to the same system.

Here we report the analysis of 2-ns dynamics of the same complex derived from Staufen protein that was studied by NMR, and a comparison with the molecular dynamics simulation conducted on the separated protein and RNA components. A number of MD simulations have been dedicated to study human U1A protein and its RNA complex, which represent paradigms to understand how the RRM, the largest RNA-binding protein family, recognizes RNA (Reyes and Kollmann 1999, Hermann and Westhof 1999). However, the dsRBD represents a completely different paradigm in RNA recognition that has not yet been analyzed using this computational approach. The results add a new quantitative description of the residual motion present at the RNA interface of this important class of RNA-binding proteins and provides the observation that short-lived inter-

actions to contribute energetically important interactions during molecular recognition.

METHODS

The simulations of the free RNA, the free protein and the protein–RNA complex were carried out using the NMR structure closest to the average structures to provide the starting coordinates. The experimental structures consisted of 16, 14, and 16, respectively, structural bundles consistent with the experimental NMR-derived structural constraints. The macromolecules were immersed in rectangular boxes of the following dimensions: $76 \times 70 \times 50 \text{ \AA}^3$ (complex) $46 \times 49 \times 62 \text{ \AA}^3$ (RNA), and $63 \times 43 \times 47 \text{ \AA}^3$ (protein), filled with TIP3P water molecules (Jorgensen et al., 1983). A minimum solute-wall box distance of 10 \AA was imposed. The systems were neutralized by addition of Na^+ cations or Cl^- anions using the AMBER leap module. The three simulated systems are composed of a total of 14,196 (RNA), 13,090 (protein) and 25,038 (complex) atoms.

Molecular dynamic simulations were conducting on a cluster of Compaq ES40 workstations by modeling each system with the AMBER95 all atom force field (Cornell et al., 1995) and using periodic boundary conditions. A cutoff radius of 9 \AA was introduced for nonbonded interactions, updating the neighbor pair list every 10 steps. The electrostatic interactions were calculated with the Particle Mesh Ewald method (Darden et al., 1993). The SHAKE algorithm (Ryckaert et al., 1977) was used to constrain all bond lengths involving hydrogens. Optimization and relaxation of solvent and ions were performed at first by keeping the solute atoms constrained to their initial position with progressively decreasing force constants of 500, 25, 15, and 5 Kcal/mol \AA^2 , respectively. The systems were minimized thereafter without any additional constraints and progressively heated up to the temperature at which the simulation was conducted. Density, volume, and overall potential energy of the system were monitored and observed to reach convergence within the first 100 ps. The three 2-ns simulations were carried out at a constant temperature (Berendsen et al., 1984) of 300 K and at a constant pressure of one atmosphere with a 2-fs time step. Pressure and temperature coupling constants were 0.5 ps.

Atomic coordinates were saved every 0.1 ps for analysis during the production run (from 0.5 to 2 ns). The AMBER *carnal* and *ptraj* modules were used to analyze structural properties of the complexes and individual molecules (root mean square deviation (RMSD), hydrogen bonds, etc.). The time dependence of the RMSD from the starting structure, as indication of phase-space accessibility, has been calculated by

$$\text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i^{\min}(t) - \mathbf{r}_i(t_0)|^2}, \quad (1)$$

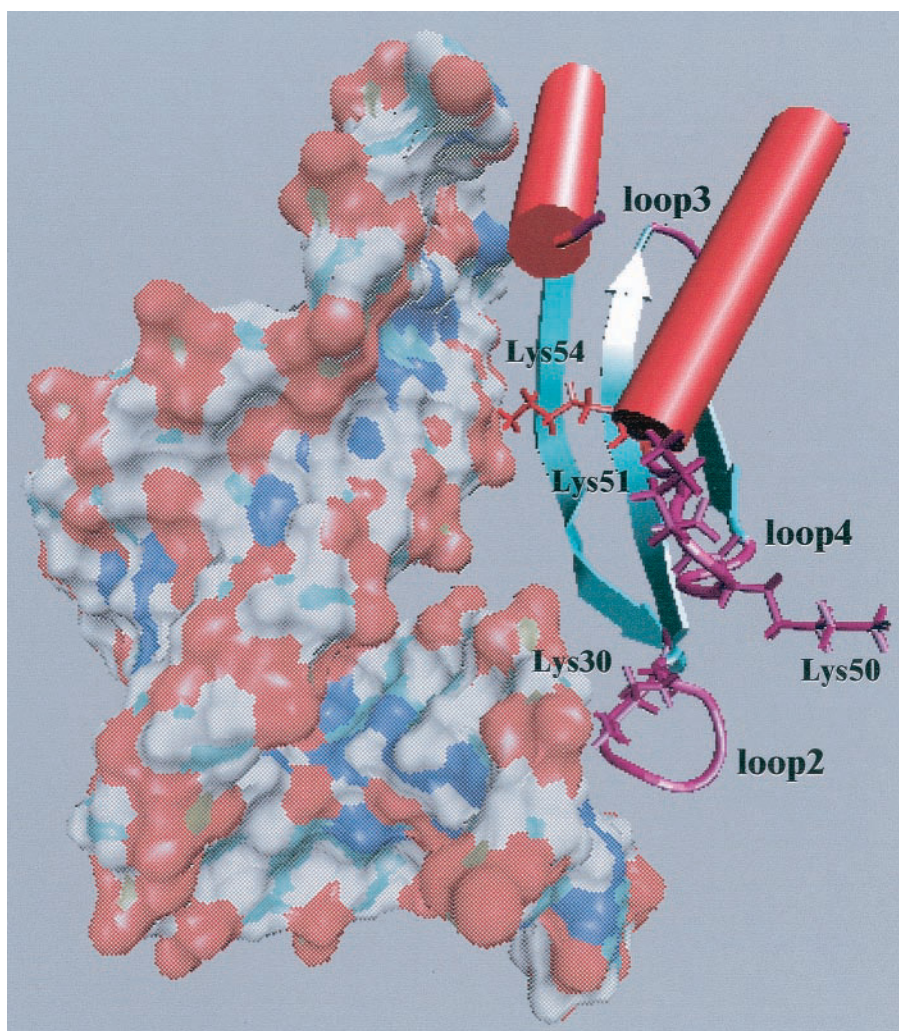
where the coordinates $\{\mathbf{r}_i^{\min}(t)\}$ are obtained optimally superimposing the instantaneous configurations at time t with the starting structure ($t = t_0$). The atomic root mean square fluctuations (RMSF) have been computed by using the definition,

$$\text{RMSF}_i = \sqrt{\sum_{\alpha=1}^3 \langle (r_{i,\alpha}^{\min}(t) - \bar{r}_{i,\alpha})^2 \rangle_{\text{MD}}}, \quad (2)$$

where the averages have been computed over the equilibrated MD trajectory.

The criterion for the occurrence of a hydrogen bond was a maximum donor–acceptor distance of 3.5 \AA and a minimum donor–proton–acceptor angle of 120° . The solvent accessible surface area of RNA and protein sites was evaluated with the program NACCESS (Hubbard and Thornton, 1993), using the default probe size of 1.4 \AA .

FIGURE 1 View of the structure of the dsRBD–dsRNA complex used as the starting structure for the simulation. The lateral chain of the lysines, which have been found to contact RNA during the simulation, have been explicitly represented.



RESULTS AND DISCUSSION

Stability of the simulation

A 2-ns MD simulation of the complex between Staudenmann dsRBD3 and an RNA stem loop containing a 13 base-pairs double-helical region and capped by a tetraloop (Fig. 1) was conducted as described in the methodological section. An analogous set of simulations was also conducted for the separate protein and RNA components. In each case, NMR-derived coordinates were used as starting structures for the simulation. The RMSD of the dsRBD–dsRNA complex relative to the starting structure is reported in Fig. 2 *A*. The results show that the simulation is stable over the entire 2 ns over which it was conducted, with the RMSD reaching a plateau after ~300 ps and oscillating around 3 Å for the remainder of the trajectory. The analysis of the results reported throughout the manuscript refers to the last 1.5 ns of the simulations, because inspection of all RMSD plots indicate that all three systems are all well equilibrated after the first 500 ps of the simulation.

The RMSDs with respect to their respective starting structures of the RNA free and in complex with dsrbd3 protein are shown in Fig. 2 *B* as a function of the simulation time. The RMSD of the free RNA reaches the equilibrium more slowly than it does in the complex and higher fluctuations are also observed. A similar plot for the protein (Fig. 2 *C*) indicates that the conformational space sampled by the protein is lower in the complex than in the free protein, the plateau being ~3.5 and 2.5 Å for the bound and the free protein, respectively. The molecular origin of these differences in RNA and protein dynamic behavior is discussed below.

The relatively small RMSDs and overall stability of the simulation indicate that all structures simulated here are well maintained over the course of the simulation. Because NMR structures are generated as structural bundles consistent with the experimental data (14 for the complex, 16 for the RNA, 14 for the protein), we are able to compare the conformational distribution represented by the experimental structures and by the MD simulation. Table 1 compares the

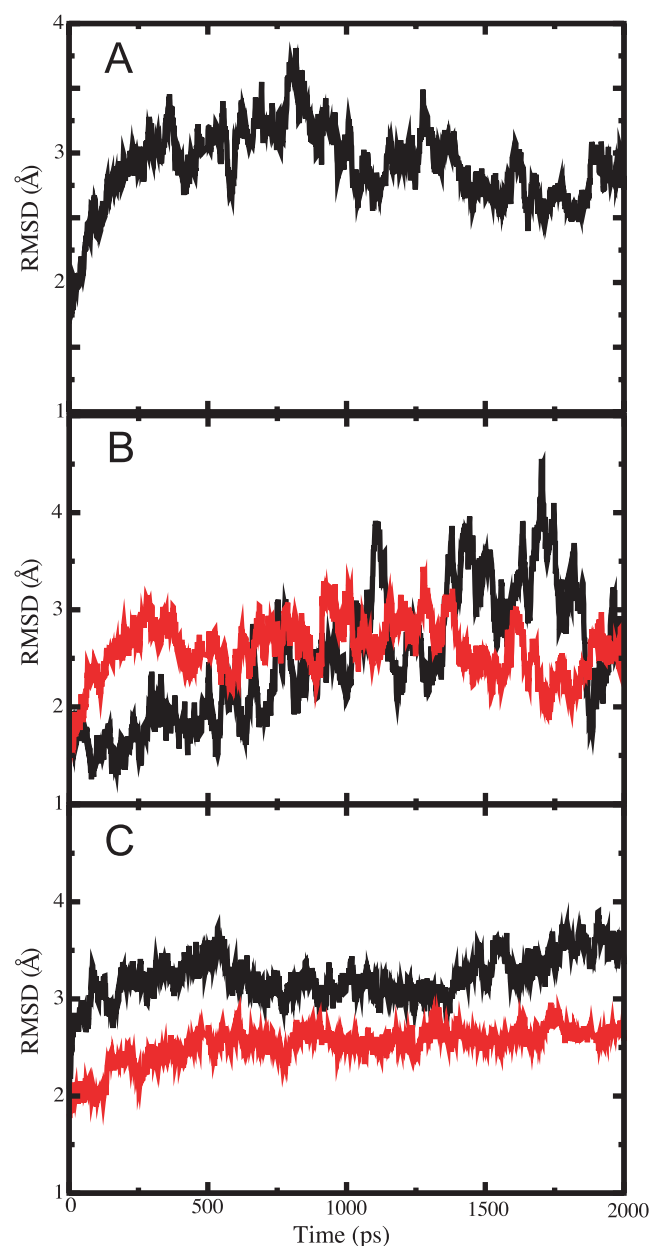


FIGURE 2 Time evolution of the root mean square deviations relative to the starting coordinates. (A) dsRBD–dsRNA complex. (B) Free (black) and bound (red) dsRNA. (C) Free (black) and bound (red) dsRBD protein domain.

TABLE 2 Minimum RMSDs (in Å) of the MD trajectory relative to each NMR structure

Structure Number	Complex	Protein in Complex	RNA in Complex	Protein Free	RNA Free
1	4.18	3.07	3.28	2.93	4.88
2	3.79	3.96	1.86	2.68	5.43
3	3.12	3.13	1.20	2.52	2.14
4	3.56	3.01	2.67	2.55	1.91
5	3.37	2.70	2.93	2.59	2.60
6	3.36	3.25	2.74	2.50	4.96
7	4.60	4.57	2.59	3.26	3.34
8	3.89	3.09	3.12	2.34	3.18
9	4.41	4.09	3.11	3.16	4.02
10	5.03	5.18	3.63	2.46	2.75
11	2.81	3.09	1.95	2.78	3.26
12	3.56	3.06	2.21	2.45	4.12
13	3.81	2.69	3.96	2.74	3.57
14				2.82	3.72
15				2.94	3.44
16				2.72	3.72
17				2.67	

The NMR structure used as a starting frame for the MD simulation is not reported.

average, minimum, and maximum RMSDs within the NMR structures (relative to the starting structure); these values are comparable to the RMSDs of the MD trajectories. The minimum RMSD of the MD trajectory (relative to each experimental NMR structure) is reported in Table 2. These results indicate that both the complex and the free protein and RNA sample conformational spaces during the MD simulation are close to those observed by NMR.

Analysis of the conformational fluctuations and their changes upon complex formation

A detailed picture of the local motions that occur on a residue-by-residue basis within the protein and the RNA in their free and bound forms was obtained by analyzing RMSFs averaged over each protein and RNA residue. Figure 3A reports the per residue protein RMSFs calculated over the C- α atoms. The RMSFs calculated from blocks of 300 or 500 ps give comparable values. From this analysis, we estimate the error on the reported values to be lower than 10%. The plot reveals several regions of the protein that are

TABLE 1 Average, minimum, and maximum pair-wise RMSDs (in Å) between the (A) NMR structures and (B) of the MD trajectory relative to the starting NMR structure

	(A) NMR versus NMR			(B) MD versus NMR		
	Average	Minimum	Maximum	Average	Minimum	Maximum
Complex	4.4	3.8	7.4	2.8	2.5	3.6
Protein in complex	3.4	2.2	5.2	2.6	2.3	3.0
RNA in complex	3.8	2.7	7.8	2.6	1.9	3.4
Protein free	2.4	1.4	3.0	3.3	2.8	4.0
RNA free	4.7	1.8	8.6	2.8	1.6	4.6

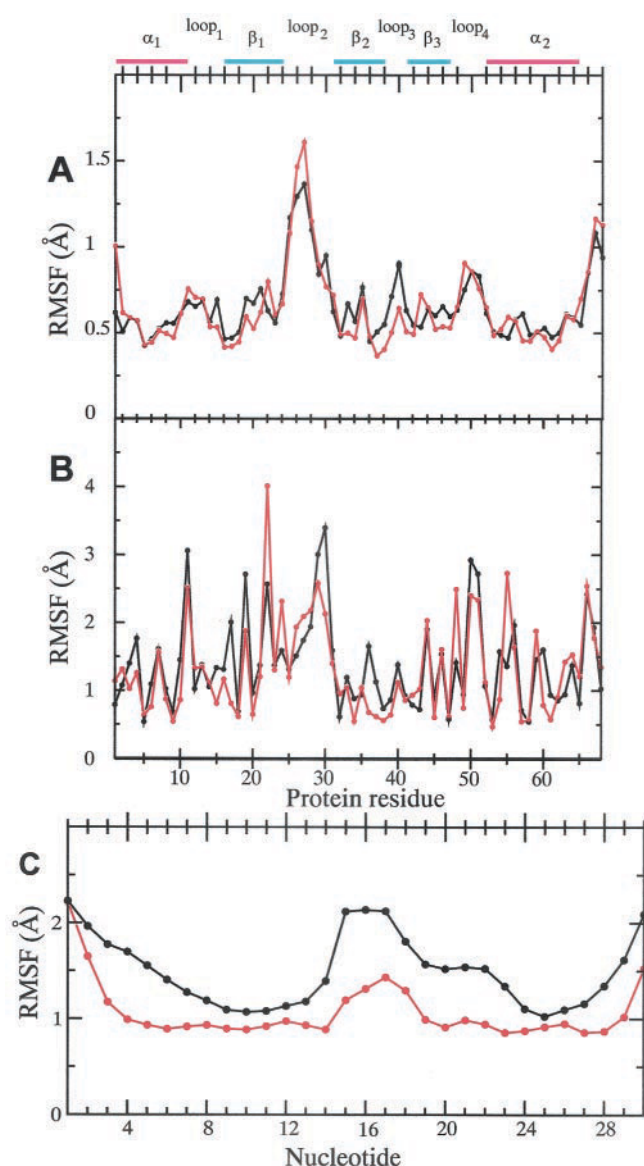


FIGURE 3 Root mean square fluctuations observed in the course of the simulation: (A) calculated for each protein residue for the free (black) and bound (red) dsRBD over the C- α atoms; (B) averaged for each protein residue for the free (black) and bound (red) dsRBD over the side chain; (C) averaged over each RNA base for the free (black) and bound (red) dsRNA. The secondary structure of the protein is represented in the upper part of the figure.

considerably more mobile than the well-ordered secondary structured regions of the dsRBD. In the free protein, these are the C-terminal amino acids, which are disordered as expected, and, significantly, loops 2, 3, and 4 (residues 25–30, 38–40, and 47–51, respectively). Surprisingly, flexibility of loop 3 is quenched in the complex, although this loop is not in direct contact with the RNA. The excess mobility (compared to the well-ordered hydrophobic core of the protein) of residues within loop 2 is reduced for some residues (i.e., 25, 30) but increased for others (i.e., 26, 27),

while residues within loop 4 retain similar mobility in the presence or absence of RNA. These results indicate that loops 2 and 4, two of the three regions of the protein that participate to the RNA interface, are more highly mobile than the remainder of the protein both in the free and in the RNA-bound form of the protein. In fact, loop 2 appears to be the region of highest mobility, as evidenced by the value of the RMSF for residues within this critical loop. These results are consistent with the conclusions of NMR relaxation measurements (Ramos et al., 2000) that reported enhanced levels of flexibility within loop 2 and loop 4. Moreover, several amino acids could not be quantitatively analyzed in the complex because of line broadening indicative of a motion on a slower time scale than sampled by the heteronuclear NOE measurements. Finally, NOE interactions involving many side chain protons were quenched due to motion on the nanosecond time scale (A. Ramos and G. Varani, et al., unpublished results).

The quenching of the loop 3 fluctuations must be due to long-range effects, because this region is not directly involved in interactions with RNA. An intriguing possibility is suggested by the observation that interactions involving loop 2 and loop 4 must be precisely spaced to allow for recognition of dsRNA. Loop 4 binds across the major groove from the site of interaction with loop 2 and the minor groove (Ryter and Schultz, 1998; Ramos et al., 2000), and it is very likely that this spacing (which differs between A-form RNA and B-form DNA) is important for discrimination between DNA and RNA double helices (Ramos et al., 2000). In support of this suggestion, mutation of Phe-32, a universally conserved residue that fixes the relative position of the two loops, abolishes RNA binding, though this residue does not contact RNA at all. Perhaps the increased rigidity of loop 3 reflects the long-range consequences of the establishment of interactions involving both loop 2 and loop 4, which leads to increased structural rigidity in the protein.

To further characterize conformational flexibility, we calculated side chain RMSFs for free and bound protein as well (Fig. 3 B). The plot indicates that the positively charged side chains, i.e., lysines and arginines, have the largest fluctuations, as expected from their long length. Those that localize within loops 2 and 4 (e.g., Lys-30, Lys-50, and Lys-51) and that contact the RNA retain comparable high levels of mobility in the free and bound forms. However, positive side chains that do not reside within these loops have large fluctuations. Overall, it appears that surface accessibility and chain length dictates side-chain mobility, but it is relevant and consistent with other results discussed here that RNA binding does not affect significantly the mobility of side chains at the RNA–protein interface.

The RNA per-base RMSFs weighted per mass unit over all the atoms is reported in Fig. 3 C. We attribute the higher fluctuations observed in the RNA to relatively long-range motions (bending and twisting of the RNA helix) and there-

fore to the absence of long-range interactions (other than base pairing) between RNA residues equivalent to the packing interactions observed in the hydrophobic core of the protein. The plot of Fig. 3 indicates that conformational fluctuations are higher in the free than in the bound RNA. In both cases, the highest fluctuations are localized to the helix termini (as expected) and to the tetraloop. This result was unexpected, because UUCG tetraloops are very stable structural units (Cheong et al., 1990), although it is perhaps to be expected that it is more rigid than an uninterrupted stretch of perfect A-form RNA. As a matter of fact, bases 15, 16, and 17, the most mobile ones, have a smaller number of structural restraints when compared to the other ones. Protein binding causes a strong quenching of the fluctuations for almost all the RNA bases, including the tetraloop, which remains nonetheless the most mobile region of the structure. Interestingly, base 15, among the most mobile residues in the free RNA, shows the largest quenching in fluctuations, probably because it becomes involved in direct and water-mediated protein interaction (see below).

Direct hydrogen bonds at the protein–RNA interface form a network of short-lived, continuously exchanging interactions

The intermolecular dsRBD–RNA interface is small compared to other RNA–protein complexes and involves relatively few protein residues (Ryter and Schultz, 1998; Ramos et al., 2000). The buried area of 1450 Å² obtained from the experimental results is confirmed from the average value observed in our simulation (1550 Å²). This value does not change significantly over the entire trajectory (data not shown). Clearly, this protein–RNA complex displays a relatively small number of long-lasting contacts when compared to other protein–RNA or protein–DNA complexes (Eriksson et al., 1995; Nilsson, 1998; Tang and Nilsson 1998; Reyes and Kollman, 1999; Chillemi et al., 2001). A schematic representation of the direct protein–RNA hydrogen bonds present for more than 30% of the simulation time, a cutoff used in MD simulation studies of similar systems (Tang and Nilsson 1998), is shown in Fig. 4. The figure indicates that dsRBD and dsRNA do not form a tight complex with an extensive and intricate intermolecular interface, but rather conserve a relatively high degree of freedom with respect to each other. Among the residues involved in the RNA interaction (primarily loops 2 and 4 and helix α , plus a few interactions involving residues from the β 1 strand), only loop 2 and loop 4 residues maintain direct hydrogen bonds with RNA atoms for a long percentage of time. In helix α 1, only His-6 forms a hydrogen bond for a relatively long percentage (38%) of the simulation time. This residue is one of only three amino acids involved in base-specific interactions, the other being Pro-26 and Lys-30 within loop 2. All other contacts involve nonse-

quence-specific interactions with the sugar–phosphate backbone of the double-stranded RNA.

The simulation suggests that interactions involving Lys-30 and Lys-54 have a major role in the recognition process, because these amino-acid side chains form numerous direct hydrogen bonds with RNA atoms for a high percentage of the simulation time. Both Lys-30 and Lys-54 are very highly conserved in dsRBD proteins, and mutations of either residue strongly affect RNA binding (Ramos et al., 2000). However, a detailed analysis of the interactions formed by these two residues in the course of the simulation reveals that individual amino-acid side chains undergo large amplitude motions that lead to contacts with the RNA remaining present for only a few tens of picoseconds before a new set of contacts is established. However, these protein side chains remain associated with the RNA for a high percentage of the simulation by continuously switching among different interactions even if individual contacts are short lived. An illustration of the time-dependent network of interactions observed at the interface during the simulation is reported in Fig. 5, where we show how the side chain of Lys-54 rotates quickly, continuously changing its hydrogen atoms involved in the interaction with the O3' atom of G19. Much more remarkable is the observation that each of the two Lys amino groups contacts a different oxygen atom of the RNA backbone in the course of the simulation. This is shown in Fig. 6, where five different snapshots of the Lys-54–RNA contacts are shown. In detail, three of them (*A*, *B*, and *C*) show the occurrence of different direct Lys-54–RNA contacts, whereas the remaining two (*D* and *E*) show the occurrence of different Lys-54–RNA water-mediated contacts. Similar observations could be made for the other lysines that contact the RNA, namely K30, K50, and K51. We conclude that, rather than forming single well-defined interactions with the RNA, these side chains continuously switch from one polar interaction to the other on a very fast time scale, yet retain a nearly continuous association with RNA polar groups.

Residual interfacial flexibility may be responsible for the relatively high per-residue RMSF (Fig. 3, *A* and *B*) and provides a molecular explanation for it. This result also supports the conclusion that the apparent disorder highlighted by the bundle of NMR structures reflects a true physical property of the interface. It is interesting to compare the behavior of K50, K51, and K54 observed in the MD trajectory with that described by NMR (Ramos et al., 2000). Figure 3 *A* indicates that the backbone of Lys-54 is rigid, and indeed this residue belongs to an α -helix, whereas Lys-50 and Lys-51 are highly mobile even at the backbone level and are found on the loop connecting the third strand of the β -sheet with the terminal α -helix. However, all these side chains have high levels of flexibility (Fig. 3 *B*), although they all participate in RNA recognition during the simulation and in the

Direct hydrogen bonds

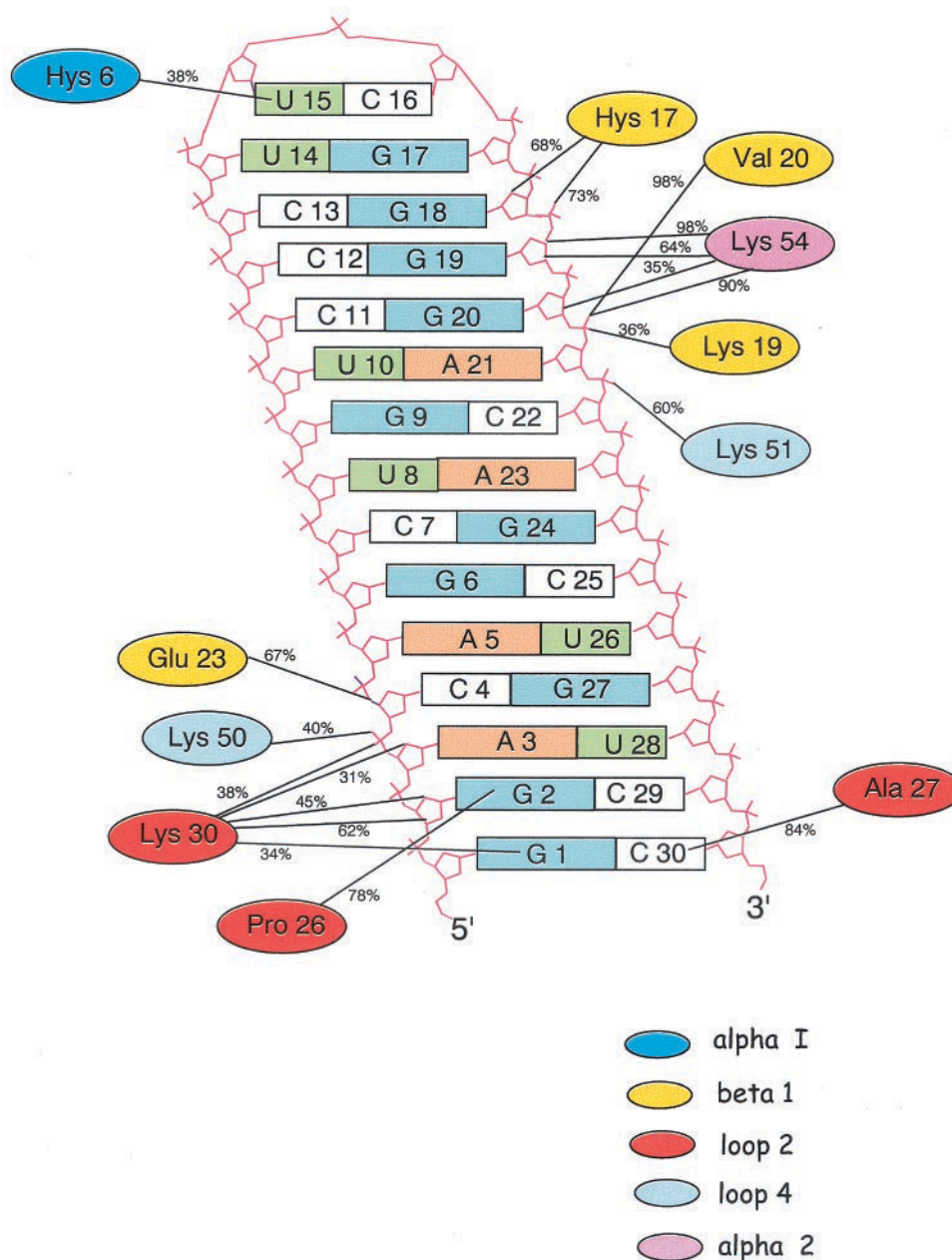
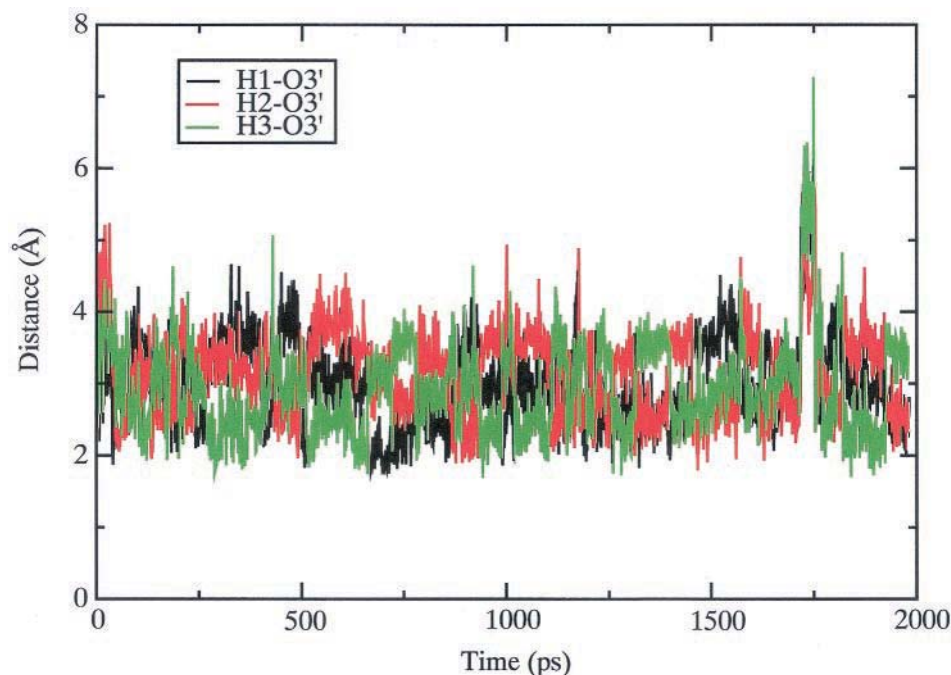


FIGURE 4 Schematic representation of direct protein-RNA direct contact observed for more than 50% (*thick line*) and between 30% and 50% (*thin line*) of the total simulation time. Amino acids are color coded according to different regions of the protein, as shown at the bottom.

NMR structure. Consistent with this observation, superposition of the NMR structures shows a spread of structures for these side chains (Fig. 6 of Ramos et al., 2000). Unfortunately this conclusion, based on superposition of structures, cannot be corroborated with direct observations of local mobility, because there are no effective methods available to monitor side-chain mobility (other than for methyl groups; see e.g., Mittermaier et al., 1999). The behavior described herein may be responsible for the high levels of fast local dynamics observed in ^{15}N NMR relaxation exper-

iments (Ramos et al., 2000). Although the two methods are distinct, NMR and MD simulations may both have monitored a fundamental property of the dsRBD-dsRNA interface. An attractive explanation for these phenomena is that retaining residual flexibility upon binding reduces entropic costs associated with the rigidification of long basic side chains. Because favorable electrostatic interactions are retained in the different hydrogen bonding conformations observed in the simulation, enthalpic gains can be achieved at relatively small entropic costs through this mechanism.

FIGURE 5 Time evolution of the distance between each of the three hydrogen atoms of the amino group of Lys-54 and the O3' atom of G19.



Water-mediated hydrogen bonds at the protein–RNA interface are a short-lived yet important component of the interface

Analysis of the simulation trajectories allowed us to analyze protein–RNA interactions mediated by water molecules as well. As for the direct interactions discussed in the previous section, we only considered water-mediated protein–RNA hydrogen bonds that are present for more than 30% of the simulation time. Water molecules mediate, at least transiently, interactions that are key to recognition (Fig. 6). As reported in Fig. 7, water-mediated interactions are numerous, comparable to the number of direct hydrogen bonds reported in Fig. 3, yet lower than reported for protein–DNA complexes (Chillemi et al., 2001). Because of their number, these interactions are likely to play a thermodynamic and structural role comparable to that of direct interactions discussed in the previous section.

Water-mediated hydrogen bonds involve His-17, Phe-18, Lys-19, and Glu-23 within strand β 1, Lys-30 within loop 2, Lys-50 and Lys-51 within loop 4, and Lys-54 within helix α 2. Phe-18, Lys-19, Lys-50, Lys-51, and Lys-54 have water-mediated RNA interactions for more than 50% of simulation time, suggesting that they are preferred hydration sites. These residues are important for binding, because mutation of Phe-18, Lys-30, Lys-50, and Lys-51 to Ala all reduce binding affinity (Ramos et al., 2000). None of the water molecules mediating the interactions are single long-lasting water molecules, but rather they all represent transient water molecules that are in continuous fast (typically 1–10-ps) exchange with the bulk water. For example, the water-mediated interaction between the NH group of

Phe-18 and the O5' atom of G19, which is present for 87% of the simulation time, is mediated by 147 different individual water molecules throughout the simulation, and the longest-lasting water molecule bridges the two atoms for 67 ps. The observation that all water-mediated contacts involve water molecules continuously exchanging with solvent is consistent with the high flexibility observed at the interface.

It is very unlikely that any of the water-mediated contacts observed in the simulation could have been detected by NMR using current methods, due to their very short residence time and to difficulties in distinguishing signals from RNA 2'-OH from the water signals (Otting et al., 1991; Kubinec and Wemmer, 1992; Liepinsh et al., 1992). In contrast, the crystallographic structure of the dsRBD from *Xenopus laevis* Xlrpba protein bound to a dsRNA duplex (Ryter and Schultz, 1998) reports 18 water-mediated contacts. This number is close to what is observed in our simulation (15), but only three of these contacts are common between experimental work and simulation. This difference can, in large part, be explained based on sequence divergence between the two dsRBDs: the only common residues involved in water-mediated contacts are Arg-143, Lys-163, and Lys-164 of Xlrpba (these correspond to Lys-30, Lys-50, and Lys-51 of Staufen). Both Lys-50 and Lys-51 are involved in water-mediated contacts analogous to those observed in the crystal structure. A second significant difference concerns the residence time of these water molecules. In the simulation, the water molecules bound to these residues are not fixed but rather in fast exchange with the solvent. For example, the contact between Lys-50 and the phosphate oxygens of A13 is mediated by 396 different

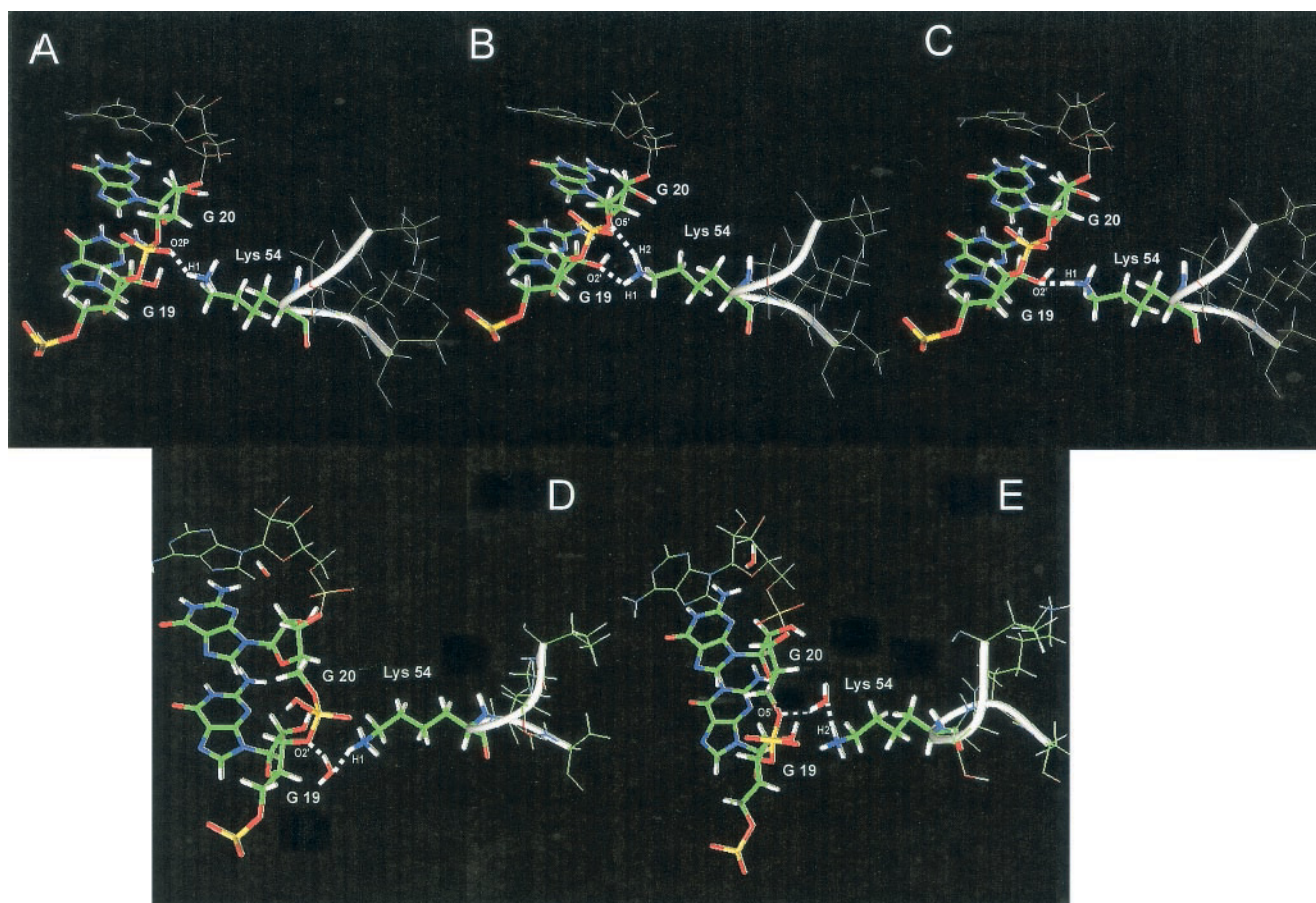


FIGURE 6 Five snapshots of the MD simulation showing the variety of RNA contacts explored by the lateral chain of Lys-54. (A, B, C) Direct protein–RNA contacts. (D, E) Water-mediated protein–RNA contacts.

water molecules during the simulation. Although the longest-lasting molecule bridges the two chemical groups for 29 ps, this position is occupied by a water molecule for 79% of the simulation time. Perhaps this large fractional occupation allows its identification through x-ray diffraction. A similar behavior is found for the water molecules close to residue Lys-30 and Lys-51. These results indicate that the water molecules seen through x-ray diffraction are found in a specific position with high probability, but are nonetheless in fast exchange with the solvent. This behavior may again confer flexibility to the protein–RNA complex and stabilize the complex by providing favorable hydration while reducing entropic costs.

CONCLUSIONS

The MD simulations of the complex between Staufen dsRBd3 and RNA and of the separate protein and RNA components have allowed us to examine the role of conformational flexibility and of water-mediated interactions in protein–RNA recognition. The simulations indicate that the protein retains a substantial flexibility in the complex. In

particular, loops 2 and 4, the regions of the protein that mediate recognition, maintain a high degree of flexibility upon binding, with critical basic protein side chains retaining favorable interactions with the RNA but migrating from one contact to another with typical residence times of a few tens of picoseconds. The simulation also highlights the crucial role of water molecules in mediating recognition, because water-mediated contacts are comparable in number to direct hydrogen-bonding interactions. The presence of many water-mediated contacts was observed in the x-ray structure of a related dsRBD–dsRNA complex. However, the MD simulation indicates that contacts mediated by molecules are in fast exchange with the solvent (every few picoseconds typically), thus contributing to the substantial flexibility of the intermolecular interface. This is in contrast to the static picture provided by x-ray diffraction.

A high degree of flexibility at the protein–RNA interface is likely to provide entropic advantages, because it reduces the entropic loss due to rigidification of protein side chains, as observed instead in other examples of RNA–protein recognition. A strong correlation between interfacial order and specificity was observed in a study of sequence-specific

Water mediated Protein-RNA hydrogen bonds

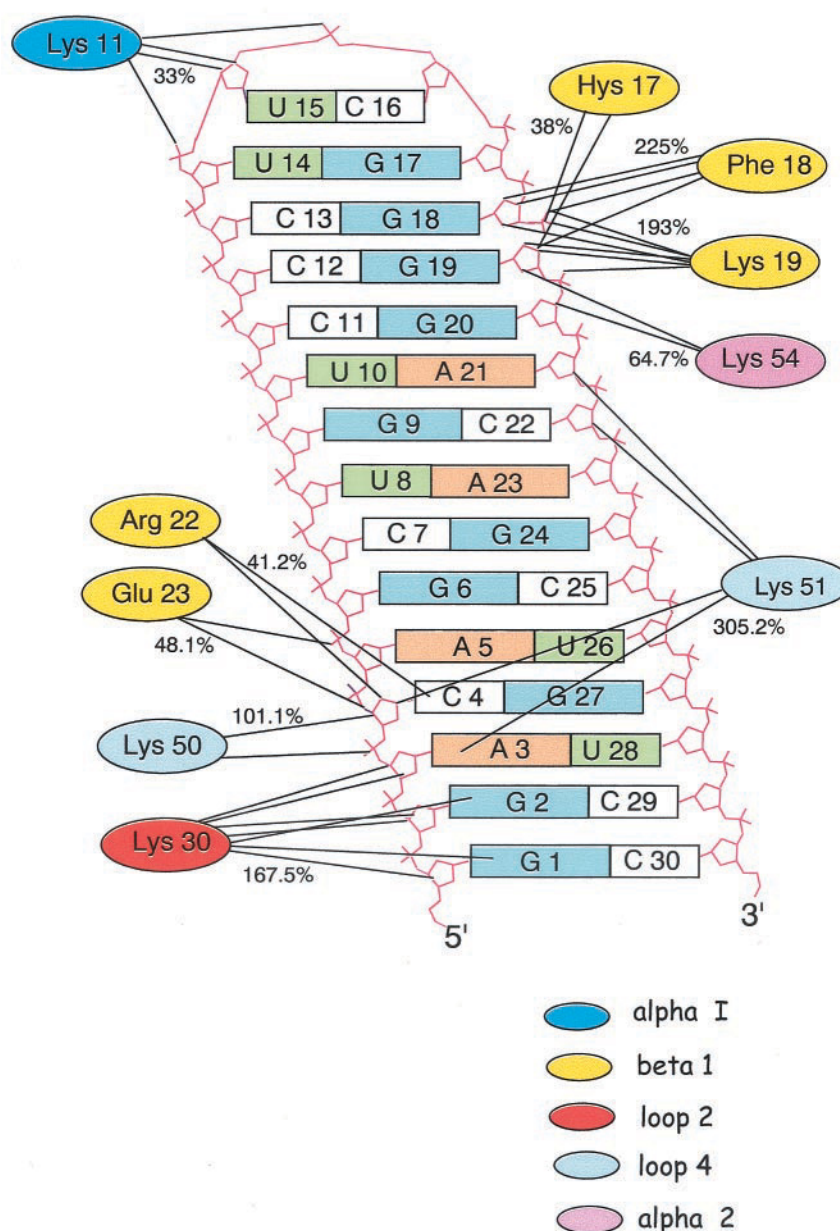


FIGURE 7 Schematic representation of water-mediated contacts observed for more than 50% (*thick line*) and between 30% and 50% (*thin line*) of the total simulation time. Amino acids are color coded according to different regions of the protein, as shown at the bottom.

RNA–protein recognition in the RRM family of RNA-binding proteins (Mittermaier et al., 1999). From the few examples where dynamic processes have begun to be studied at RNA–protein surfaces, sequence-specific recognition requires a complex distribution of dynamics, with some side chains being highly ordered and others retaining high degree of flexibility. In contrast, high levels of interfacial flexibility may be prevalent in cases, such as the present, of nonsequence-specific recognition. The present work highlights the important role of flexibility in molecular recognition and the ability of MD simulations to reveal quantitative features of the underlying dynamic processes.

REFERENCES

- Bass, B. L., S. R. Hurst, and J. D. Singer. 1994. Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr. Biol.* 4:301–314.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Comput. Phys.* 81:3684–3690.
- Bevilacqua, P. C., and T. R. Cech. 1996. Minor groove recognition of the double-stranded RNA binding domain from the RNA-activated protein kinase PKR. *Biochemistry.* 35:9983–9994.
- Beveridge, D. L., and G. Ravinshanker. 1994. Molecular dynamics study of DNA. *Curr. Opin. Struct. Biol.* 4:246–255.

- Broadus, J., S. Fuestenberg, and C. Q. Doe. 1998. Staufen-dependent localization of *prospero* mRNA contributes to neuroblast daughter-cell fate. *Nature*. 391:792–795.
- Bycroft, M., S. Grunert, A. G. Murzin, M. Proctor, and D. St Johnston. 1995. NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.* 14:3563–3571.
- Castrignano, T., G. Chillemi, and A. Desideri. 2000. Structure and hydration of BamHI DNA recognition site: a molecular dynamics investigation. *Biophys. J.* 79:1263–1272.
- Chillemi, G., T. Castrignano, and A. Desideri. 2001. Structure and hydration of the DNA-Human Topoisomerase I covalent complex. *Biophys. J.* 81:490–500.
- Clarke, P. A., and M. B. Matthews. 1995. Interaction between the double-stranded RNA binding motif and RNA: definition of the binding site for the interferon-induced protein kinase DAI (PKR) on adenovirus VA RNA. *RNA*. 1:7–20.
- Cheatham, T. E., J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular dynamics simulation on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. *J. Am. Chem. Soc.* 117:4193–4194.
- Cheatham, T. E., J. L. Miller, T. I. Spector, P. Cieplak, and P. A. Kollman. 1997. Molecular dynamics simulations on nucleic acid systems using the Cornell et al. force field and particle mesh Ewald electrostatics. In *Molecular Modeling of Nucleic Acids*. N. B. Leontis, editor. American Chemical Society, Washington, DC. 285–302.
- Cheong, C., G. Varani, and I. Tinoco, Jr. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature*. 346:680–682.
- Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, M. Kenneth, J. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 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., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an $N \cdot \log(n)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
- Eriksson, M. A. L., T. Hard, and L. Nilsson. 1995. Molecular dynamics simulation of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402–426.
- Ferrandon, D., L. Elphick, C. Nusslein-Volhard, and D. St Johnston. 1994. Staufen protein associates with the 3'UTR of *bicoid* mRNA to form particles that move in a microtubule-dependent manner. *Cell*. 79:1221–1232.
- Green, S. R., and M. B. Matthews. 1992. Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Gene Dev.* 6:2478–2490.
- Hermann, T., and E. Westhof. 1999. Simulation of the dynamics at an RNA protein interface. *Nature Struct. Biol.* 6:540–544.
- Hubbard, S. J., and J. M. Thornton. 1993. "NACCESS", Computer Program. Department of Biochemistry and Molecular Biology, University College, London.
- Kharrat, A., M. J. Macias, T. J. Gibson, M. Nilges, and A. Pastore. 1995. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* 14:3572–3584.
- Kim-Ha, J., J. L. Smith, and P. M. Macdonald. 1991. *Oskar* mRNA is localized to the posterior pole of *Drosophila* oocyte. *Cell*. 66:23–35.
- Kim-Ha, J., P. Webster, J. L. Smith, and P. M. Macdonald. 1993. Multiple RNA regulatory elements mediate distinct steps in localization of *oskar* mRNA. *Development*. 119:169–178.
- Kubinec, M. G., and D. E. Wemmer. 1992. NMR evidence for DNA bound water in solution. *J. Am. Chem. Soc.* 114:8739–8740.
- Jorgensen, W. L., J. Chandrasekhar, J. Madura, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem Phys.* 79:926–935.
- Liepinsh, E., G. Otting, and K. Wuthrich. 1992. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. *Nucleic Acids Res.* 20:6549–6553.
- Macdonald, P. M., and G. Struhl. 1988. *Cis*-acting sequences responsible for anterior localization of *bicoid* mRNA in *Drosophila* embryos. *Nature*. 336:595–598.
- Mittermaier, A., L. Varani, D. R. Muhandiram, L. E. Kay, and G. Varani. 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.
- Nilsson, L. 1998. Protein–nucleic acid interactions. In *Encyclopedia of Computational Chemistry*. P.V.R. Schleyer, editor. John Wiley and sons, New York. 2220–2229.
- Otting, G., E. Liepinsh, and K. Wuthrich. 1991. Protein hydration in aqueous solution. *Science*. 254:974–980.
- Peng, J. W., and G. Wagner. 1992. Mapping of the spectral densities of N–H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry*. 31:8571–8586.
- Peng, J. W., and G. Wagner. 1994. Investigation of protein motions via relaxation measurements. *Methods Enzymol.* 239:563–596.
- Ramos, A., S. Grunert, J. Adams, D. R. Micklem, M. R. Proctor, S. Freund, M. Bycroft, D. St Johnston, and G. Varani. 2000. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.* 19:997–1009.
- Reyes, C. M., and P. A. Kollman. 1999. Molecular dynamics study of U1A–RNA complexes. *RNA*. 5:235–244.
- Ryckaert, J.-P., G. Cicciotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.
- Ryter, J. M. and S. C. Schultz. 1998. Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* 17:7505–7513.
- Sen, S., and L. Nilsson. 1999. Structure, interaction, dynamics and solvent effects on the DNA–*EcoRI* complex in aqueous solution from molecular dynamics simulation. *Biophys. J.* 77:1782–1800.
- St Johnston, D. 1995. The intracellular localization of messenger RNAs. *Cell*. 81:161–170.
- St Johnston, D., D. Beuchle, and C. Nusslein-Volhard. 1991. Staufen, a gene required to localized maternal RNAs in the *Drosophila* egg. *Cell*. 66:51–63.
- St Johnston, D., N. H. Brown, J. G. Gall, and M. Jantsch. 1992. A conserved double-stranded RNA binding domain. *Proc. Natl. Acad. Sci. U.S.A.* 89:10979–10983.
- Tang, Y., and L. Nilsson. 1998. Interaction of human SRY protein with DNA: a molecular dynamics study. *Proteins*. 31:417–433.
- Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and Hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305.
- Tsui, V., I. Radhakrishnan, P. E. Wright, and D. A. Case. 2000. NMR and molecular dynamics studies of the hydration of a zinc finger–DNA complex. *J. Mol. Biol.* 302:1101–1117.
- Varani, G. 1997. RNA–protein intermolecular recognition. *Acc. Chem. Res.* 30:189–195.
- Varani, G., and K. Nagai. 1998. RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* 27:407–445.
- Young, A. Y., and D. L. Beveridge. 1998. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* 281:675–687.