Mechanism and Thermodynamics of Binding of the Polypyrimidine Tract Binding Protein to RNA†

Nathan Schmid, Bojan Zagrovic, and Wilfred F. van Gunsteren*

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, CH-8093 Zürich, Switzerland

Received December 20, 2006; Revised Manuscript Received March 21, 2007

ABSTRACT: The polypyrimidine tract binding protein (PTB) is involved in many physiological processes, including alternative splicing, internal ribosomal entry side (IRES)-mediated initiation of translation, and polyadenylation, as well as in ensuring mRNA stability. However, the role of PTB in these processes is not fully understood, and this has motivated us to undertake a computational study of the protein. PTB RNA binding domains (RBDs) 3 and 4 and their complexes with oligopyrimidine RNAs were simulated using the GROMOS simulation software using the GROMOS 45A4 force field. First, the stability and fluctuations of the tertiary fold and of the secondary structural elements in individual domains, the combined RBD34 domain, and their complexes with RNA were studied. Second, the simulation results were validated against the experimental NMR NOE data. The analysis of hydrogen bonding patterns, salt bridge networks, and stacking interactions of the RNA to the binding pockets of the protein domains showed that binding is not sequence-specific and that many RNA fragments can bind to them successfully. Further calculations of the relative free energy of binding for different polypyrimidine sequences were carried out using the thermodynamic integration (TI) and single-step perturbation (SSP) methods. It is was not possible to calculate the relative free energies with high accuracy, but the obtained results do give qualitative insights into PTB's affinity for different RNA sequences. Furthermore, the low-energy conformations of the complexes that were found provided additional information about the mechanism of binding.

Polypyrimidine tract binding protein 1 (PTB1) is a eukaryotic RNA binding protein, and it plays a role in several aspects of RNA metabolism: alternative splicing regulation, internal ribosomal entry site (IRES)-mediated translation initiation, and 3′-end processing, as well as in ensuring mRNA stability. In the process of splicing, intron sequences are cut out of the RNA precursor (pre-RNA) by two transesterification reactions. Although the chemistry of the process is rather simple, the recognition of the intron−exon boundaries is not. It takes a complex splicing machinery (general splicing factor and associated proteins) and many *cis*-acting elements to recognize the intron−exon border efficiently (*1*). A gene can be spliced differently in different cell types (alternative splicing). PTB plays a key role in the alternative splicing of several genes, including c-src, α-actinin, FGF-R2, Calcitonin/CGRP, GABA $\gamma 2$, and α-tropomyosin (*2*). The c-src gene contains three exons (E3, E4, and EN) which are spliced alternatively in neural and non-neural cells. In non-neural cells, PTB binds the RNA and loops out the EN exon which is cut out as an intron by the splicing machinery (U1 and U2). In neural cells, another factor (Fox1) is present which prevents PTB binding and looping of the EN exon. Thus, the EN exon is included in the spliced c-src pre-RNA (*3*). For a review of PTB's role in alternative splicing, see ref *2*.

Initiation of translation of an mRNA usually involves a 5′-cap-dependent recruitment of the ribosome and associated initiation factors (IF) to the mRNA. The ribosome then scans the mRNA until it finds the start codon, and decoding can begin (*4*). The translation of a small subset of genes is initiated in a different way. These genes contain internal ribosomal entry sites (IRESs) which enable 5′-end-independent initiation of translation to occur. The IRES complex binds several protein factors, including PTB, and binds to the ribosome and other factors to start the translation. PTB plays an important role in the stabilization of the IRES RNA structure (*5*).

In 3′-end processing, the pre-mRNA's 3′-end is first cleaved and then a poly(A)-tail is added. The addition of the 3′-tail is mediated by polyadenylation signals (pA signals) which are pyrimidine-rich sequences in the pre-mRNA. PTB also binds to these sequences and represses polyadenylation (*6*).

Finally, it has been shown that cytosolic PTB can stabilize mRNA. For example, it was shown that PTB deficient cells produce fewer insulin secretory granules, because the mRNA of granule proteins is translated to a lesser degree as its stability is reduced (*7*).

PTB1 is a 58 kDa protein and has four RNA binding domains (RBDs) which are connected by flexible interdomain linkers. The structures of RBD1 and RBD2 (*8*) and RBD3 and RBD4 (*9*) were determined by NMR spectroscopy. Furthermore, it was shown by NMR that PTB recognizes sequences through specific interactions (*10*). More recently, the structures of all four RBDs complexed to

† Financial support from the National Center of Competence in Research (NCCR) and the Structural Biology of the Swiss National Science Foundation (SNSF) is gratefully acknowledged.

* To whom correspondence should be addressed. Phone: +41 44 632 5501. Fax: + 41 1 632 1039. E-mail: wfvgn@igc.phys.chem.ethz.ch.

10.1021/bi6026133 CCC: $37.00
Published on Web 05/12/2007

Table 1: Identity and Sequences of the Human Polypyrimidine Tract Binding Protein 1 (hPTB1) Polypeptide Chain and the RNA of the Simulated Systems

identity	protein residues	RNA	no. of solute atoms	no. of water molecules
RBD3	324−443	−	1179	12187
RBD4	444−531	−	882	5083
RBD34	324−531	−	2058	25980
RBD3C	324−443	CUCUCU	1325	12134
RBD4C	444−531	CUCUCU	1028	5464
RBD34C	324−531	2 × CUCUCU	2350	25900
RBD34C_6C	324−531	2 × CCCCCC	2356	25781

RNA_{CUCUCU} have been determined by solution NMR (*11*). All domains bind the RNA via the β-sheet surface as expected from other structures. RBD1, -2, and -4 bind three nucleotides. RBD3 binds five nucleotides and is thought to be the strongest RNA binder of the four domains (*12*, *15*). RBD3 and -4 interact with each other via their α-helices, which is unusual for domains of the RBD type (*11*). In our project, we have studied RBD3 and -4 and have focused on three general questions. First, we were interested in examining the quality of the GROMOS 45A4 atomistic force field for studying protein−RNA complexes. Historically, not much attention has been given to this topic by the simulation community. Second, we were interested in critically appraising the description of the RNA−PTB1 interface that was based on the static NMR model structure. In particular, we wanted to see how dynamic and fluctuating or, alternatively, how static and permanent the key contacts between the two molecules are. Finally, our aim was to analyze the binding process from a more quantitative thermodynamic perspective by computationally calculating relative free energies of binding of different RNA ligands to the protein. Our results confirm the validity of the force field that was used and reveal several novel features of the binding process. Most importantly, our results suggest that the binding is significantly less specific and the binding interface is significantly more fluctuating than was suggested on the basis of the NMR analysis (*11*).

MATERIALS AND METHODS

Simulation Setup. All simulations were performed using the GROMOS biomolecular simulation software and the 45A4 GROMOS force field (*16*). Initial coordinates of the protein and RNA_{CUCUCU} were taken from the NMR structure deposited in the Protein Data Bank (PDB) as entry 2ADC (*11*) (first structure of the bundle). The RNA_{CCCCCC} coordinates were generated from the RNA_{CUCUCU} coordinates by changing the appropriate atom types and adjusting the corresponding bond lengths. All missing hydrogens were generated by the *progch* GROMOS++ (*17*) program. The RBD34 domain was artificially separated at a halfway point in the middle of the interdomain region (Table 1) to produce the individual subsystems RBD3 and RBD4. The protein, the RNA, or the protein−RNA complexes were solvated in cubic boxes containing a given number (Table 1) of simple point charge (SPC) water (*18*) molecules. Periodic boundary conditions were applied. All simulations were initiated with the following equilibration scheme. First, the initial velocities were randomly generated from a Maxwell−Boltzmann distribution at 50 K. All solute atom positions were restrained to their positions in the NMR model structure through a harmonic potential energy term with force constant of 2.5×10^4 kJ mol^{-1} nm^{-2}. The system was simulated with these settings for 20 ps. Second, the temperature was increased in 50 K steps during five additional 20 ps equilibration steps, with the positional restraints being reduced by 5×10^3 kJ mol^{-1} nm^{-2} at each step. Next, a production simulation was performed. The temperature of 300 K and atmospheric pressure were kept constant using a weak-coupling approach (*19*) with relaxation times τ_T of 0.1 ps and τ_p of 0.5 ps and an isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$. Nonbonded interactions were calculated using a triple-range cutoff scheme. The interactions within a cutoff distance of 0.8 nm were calculated at every step from a pair list which was updated every fifth time step. At this point, interactions between atoms (of charge groups) within 1.4 nm were also calculated and were kept constant between updates. To account for the influence of the dielectric medium outside the cutoff sphere of 1.4 nm, a reaction-field force based on a relative dielectric constant ϵ of 61 (*20*) was added.

Thermodynamic Integration. To compute differences in free energy (ΔG_{AB}), the two Hamiltonians of the two states, A and B, are coupled with the coupling parameter λ. The λ-dependent Hamiltonians $H(\lambda)$ are chosen such that $H(\lambda_A)$ corresponds to the Hamiltonian in state A (H_A) and $H(\lambda_B)$ in state B (H_B). For this perturbation, special molecular-topology building blocks (Figure 1B) were used. It can be shown (*21*) that the *Helmholtz* free energy difference between the two states is given by the formula (*22*)

$$\Delta G_{AB} = G(\lambda_A) - G(\lambda_B) = \int_{\lambda_B}^{\lambda_A} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \qquad (1)$$

Simulations are carried out for a set of λ points, and the ensemble average of the Hamiltonians $\{\langle \partial H(\lambda)/\partial\lambda \rangle_\lambda\}$ is computed. Integration of the $\langle \partial H(\lambda)/\partial\lambda \rangle_\lambda$ curve gives the free energy difference between states A and B. From the thermodynamic cycle, the difference in the free energy of binding ($\Delta\Delta G_{AB}^{binding}$) between ligands A and B can be computed by performing one set of simulations in water and one set in the protein environment. The following equation gives the difference in the binding free energy of two ligands:

$$\Delta\Delta G_{AB}^{binding} = \Delta G_{AB}^{complex} - \Delta G_{AB}^{water} \qquad (2)$$

Single-Step Perturbation. The free energy difference between a real state (A) and a reference state (R) can be calculated from a simulation of state R using the perturbation formula (*23*)

$$\Delta G_{AR} = -k_B T \ln\langle \mathrm{e}^{-(E_A - E_R)/k_B T} \rangle_R \qquad (3)$$

where E_A and E_R are the potential energies of the system in states A and R, respectively, k_B is Boltzmann's constant, and T is the absolute temperature. The angled brackets indicate averaging over the simulation ensemble for state R (*24*). The (unphysical) reference state R is sampled in two molecular dynamics simulations, once in water and once in complex with the protein. Applying eq 3 and following the thermodynamic cycle, we carried out the calculation of relative free

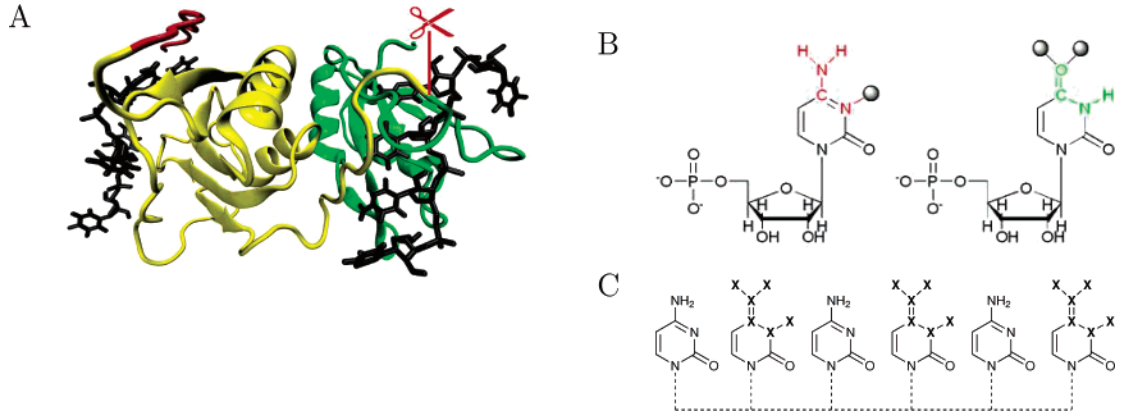

FIGURE 1: (A) Structure of RBD34 in complex with RNA_{CUCUCU} (both domains) as taken from ref *11*. RBD3 is colored yellow, its interdomain linker to RBD2 red, and RBD4 green. The pair of scissors marks the location where the domains were clipped. The bound RNA is shown using ball-and-stick representation. (B) Structure of molecular topology building blocks of cytosine (red) and uracil (green) which were used for thermodynamic integration. The gray spheres represent atoms that are not present in the corresponding pyrimidine (dummy atoms). (C) Topology of the reference ligand (RNA_{CXCXCX}) for single-step perturbation. Soft-core atoms are indicated with an X. For the sake of simplicity, the sugar and phosphate backbone are not shown but are indicated with dashed lines.

energies of binding ($\Delta\Delta G_{AB}^{binding}$) for various ligands using eq 4:

$$\Delta\Delta G_{AB}^{binding} = (\Delta G_{AR}^{complex} - \Delta G_{AR}^{water}) - (\Delta G_{BR}^{complex} - \Delta G_{BR}^{water}) \quad (4)$$

If the ensemble of state A overlaps extensively with the ensemble of state R, perturbation formula 3 will give a reasonably accurate estimate of the free energy difference. The advantage of eq 3 is that only one simulation of state R may yield many free energies ΔG_{AR}, ΔG_{BR}, etc., while using eq 1 simulations at different λ values are required per free energy ΔG_{AR}. However, due to finite sampling, the estimate is generally less accurate than that of eq 1. To improve sampling, perturbed atoms in the reference ligand are made soft, meaning that other atoms can fully overlap with them with finite probability. For this purpose, a small offset α_{LJ} is added to the interatomic distance r_{ij} in the Lennard-Jones potential-energy function:

$$V_{LJ}(r_{ij}) = \left(\frac{C_{12}}{\alpha_{LJ}\lambda^2 C_{126} + r_{ij}^6} - C_6\right)\frac{1}{\alpha_{LJ}\lambda^2 C_{126} + r_{ij}^6}, \quad (5)$$

making its $\lim_{r_{ij}\to 0} V_{LJ}(r_{ij})$ finite. In eq 5, C_6 and C_{12} indicate the corresponding van der Waals interaction parameters ($C_{126} = C_{12}/C_6$), λ is the coupling parameter of the Hamiltonians, and r_{ij} is the interatomic distance of atoms i and j. For charged atoms, the electrostatic term

$$V_{CRF}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1}\left[\frac{1}{(\alpha_C\lambda^2 + r_{ij}^2)^{1/2}} - \frac{\frac{1}{2}C_{rf}r_{ij}^2}{(\alpha_C\lambda^2 + R_{rf}^2)^{3/2}} - \frac{1 - \frac{1}{2}C_{rf}}{R_{rf}}\right] \quad (6)$$

was softened by a similar offset parameter α_C. Here, q_i and q_j are the charges, $\epsilon_1 = 1$, and C_{rf} and R_{rf} are reaction-field parameters (coefficient and cutoff, respectively). The softness parameters ($\alpha_{LJ} = 1.5$ and $\alpha_C = 0.5$ nm^2) were chosen as used previously (*25*). The topology of the reference ligand used in this study is given in Figure 1C.

Analysis. The atom-positional root-mean-square deviation (rmsd) between the indicated atoms of two structures was calculated after superposition of the indicated atoms. Root-mean-square fluctuations (rmsf) of atoms around their average positions were calculated after superposition of the indicated atoms with the energy-minimized NMR structure. Secondary structure of the protein was assigned according to the rules defined by Kabsch and Sander's DSSP (*26*). The presence of a hydrogen bond was determined by geometric criteria. If the hydrogen−acceptor distance was less than 0.25 nm and the donor−hydrogen−acceptor angle was at least 135°, the hydrogen bond was considered to be present. Salt bridges were identified with a donor−acceptor distance of <0.6 nm (*27*). Ring systems were considered to stack if the distance between the centers of geometry of the rings was less than 0.5 nm and the angle between the planes through the two rings was less than 30°. Proton−proton distances were compared to upper bounds derived from NMR spectra (*11*). Proton−proton distances were averaged using $1/r^6$ averaging [$\bar{r} = (\langle r^{-6}\rangle)^{-1/6}$]. Positions of protons that were not treated explicitly by the force field were calculated from standard configurations (*28*). In cases where the NOE upper bounds corresponded to more than one proton, a pseudoatom approach (*29*) with the standard GROMOS corrections (*30*) was applied. No additional multiplicity corrections (*31*) were added.

Software and Hardware. All simulation and energy minimization computations were carried out using GROMOS XX 0.2.3 (*17*). For analysis, either GROMOS++ 0.2.4 (*17*) or *esra*[1] was used. Additional analysis, conversion, and batch programs were written in either Perl, C++, or Java. The Java esra analysis programs for the computation of salt bridges and stacking interactions[2] were published on the esra CVS server.

Visualization was done with Visual Molecular Dynamics (VMD) (*32*).

[1] Java analysis package written by Mika A. Kastenholz and Vincent Kräutler.

[2] programs.saltbridge and programs.stacking.

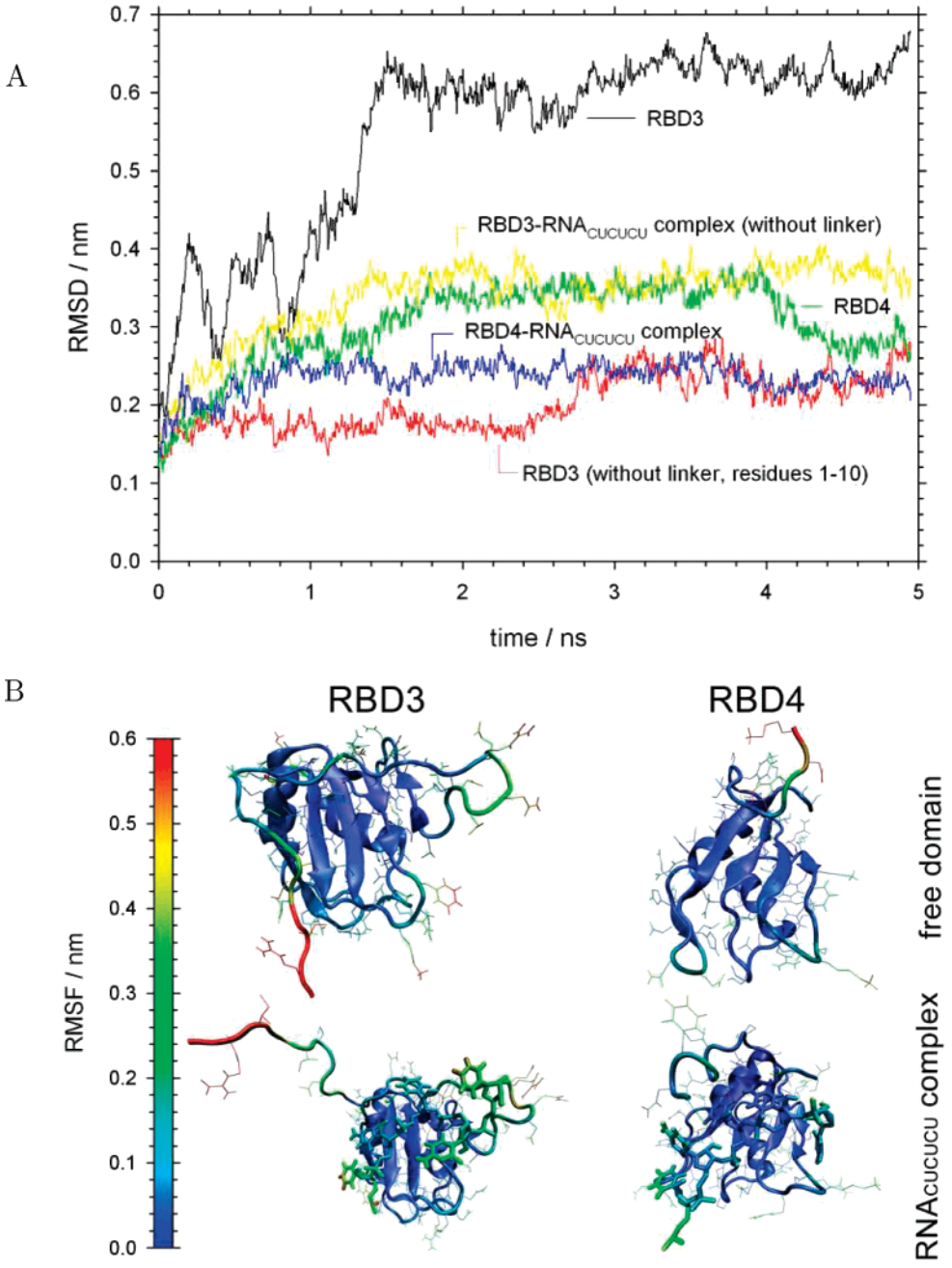

FIGURE 2: (A) Atom-positional rmsd with respect to the energy-minimized NMR structure of RBD3 and RBD4 without and with bound RNA$_{CUCUCU}$. (B) Atom-positional rmsf mapped onto the NMR structure.

RESULTS AND DISCUSSION

Structural Description of the Complexes

Simulation of RBD3 and RBD4 and the Combined RBD34 Domain. RBD3 and RBD4 (Table 1) were simulated for 5 ns each at 300 K and atmospheric pressure in the absence of RNA. In Figure 2A, the atom-positional root-mean-square deviation (rmsd) of the backbone atoms with respect to the energy-minimized NMR structure is shown. Although their structure was determined as the combined RBD34 domain, the individual domains are stable in solution beyond 1−2 ns. The N-terminal linker in RBD3 (residues 1−10) connecting RBD3 to RBD2 in the wild-type protein is very flexible, moving around extensively. The rmsd without linker residues 1−10 is much smaller. In Figure 2B, the root-mean-square fluctuations (rmsfs) of all atoms are mapped onto the structures of the domains. The mobility of linker residues 1−10 is evident. Also, the residues connecting RBD3 and RBD4 show enhanced flexibility.

The protein domains as the combined RBD34 entity were simulated for 5 ns in their native state. The backbone atom-positional rmsd of several parts of the molecule with respect to the energy-minimized NMR structure is shown in Figure 3A. RBD3 and RBD4 on their own as well as the combined RBD34 domain are stable. Once more, the flexibility of linker residues 1−10 of RBD3 is clearly seen. In Figure 3B, the secondary structure contents of the molecule are shown, as obtained from the MD trajectories and the 20 structures of the NMR bundle. The percentage of residues in secondary structural elements remains constant over time and is in agreement with the NMR data. Further, the simulation trajectory was validated against the available NMR distance

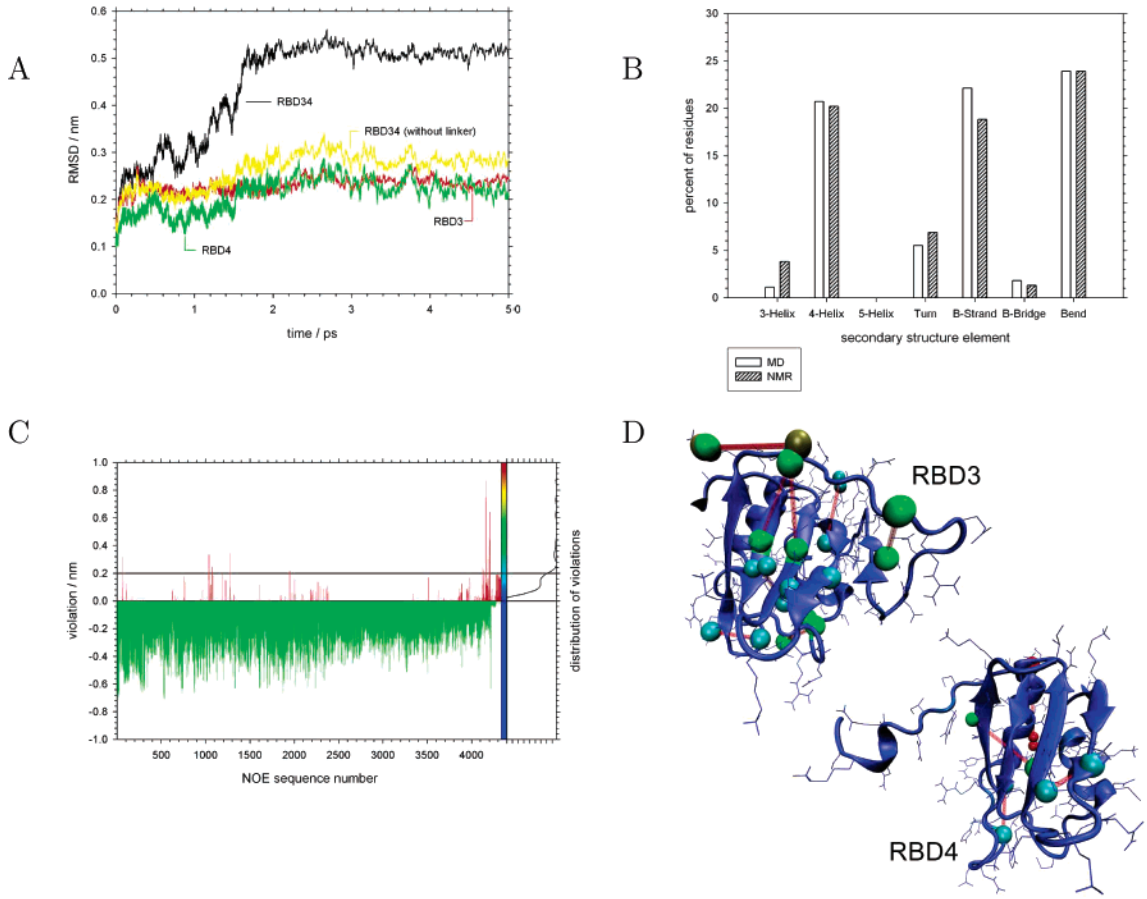

FIGURE 3: Analysis of RBD3 and RBD4 in complex with each other (RBD34). (A) Backbone atom-positional rmsd of the MD trajectory structures with respect to the energy-minimized NMR structure. (B) Secondary structure analysis of the complex (DSSP). The white bars were computed from the simulation trajectory coordinates and the striped ones from the NMR bundle. (C) Violations of NOE upper bounds (4336 analyzed). Violated upper bounds (195, average violation of 0.096 ± 0.014) are colored red and the fulfilled ones green. The distribution histogram of the violations is shown at the right. (D) NOE violations mapped onto the starting structure. Atoms involved in large violations (violation of >0.2 nm, red lines) of the NOE bounds are shown as spheres. The color indicates the size of the violation.

upper bounds which were used for structure refinement (*11*). In total, 4336 NOE upper bounds were analyzed by calculating the distances and averaging (see Figure 3C). One hundred ninety-five of the NOE upper bounds are violated by an average of only 0.096 ± 0.014 nm. In Figure 3, the violations are displayed on the starting structure. The largest violation is 0.86 nm. The distance between the two atoms (Asn 154 HD and Glu 179 CG) in the structure after the equilibration period was 0.845 nm. In the simulation, the Asn 154 side chain moves out into the water environment and the distance between the atoms of the pair is further increased. This results in the large violation of the NOE upper bound. However, the helices do not move against each other. Taken together, this may be considered as an indication that only minor changes in the structure are induced by the force field that is used. There is no systematic deviation affecting just one particular part of the structure.

Free RNA. The (free) RNA_{CUCUCU} was simulated in water for 2 ns. Atom-positional rmsd and rmsf analysis showed that the (free) molecule is structurally heavily fluctuating (data not shown). As expected, the bases show enhanced movement compared to the RNA's phosphate backbone: this effect comes mostly from the rotation of the bases around their long axis. Finally, the overall structure is compacting to a somewhat more globular shape (decreasing radius of gyration) which is expected due to a lack of stabilization by the protein and the interaction of the bases with the phosphate backbone and ribose moieties via hydrogen bonds.

Protein−RNA Complexes. Both protein domains (RBD3 and RBD4) were simulated in a complex with the CUCUCU hexanucleotide for 5 ns. The backbone atom-positional rmsd of both complexes with respect to the energy-minimized NMR structure is shown in Figure 2A and reaches a value of 0.3 nm (excluding the linker residues). The rmsd indicates that the complexes are stable in solution. In Figure 2B, the all-atom rmsf was mapped onto the starting structure. The rmsf values for the bound RNAs were compared to the ones obtained in the free RNA simulation (data not shown). Rotational movement of some of the bases has disappeared. RBD3 binds three bases of the hexanucleotide tightly (bases 3−5), while RBD4 binds four (bases 3−6) and two of them (bases 4 and 5) tightly. Here, "tight binding" was identified as low fluctuations (low rmsf).

The RBD34−RNA_{CUCUCU} complex was simulated for a total of 4.9 ns (Figure 4). The backbone atom-positional rmsd of the whole protein and of single-domain complexes with respect to the energy-minimized NMR structure is shown in Figure 4A. The convergence in rmsd indicates that the complex is stable. Furthermore, most of the rmsd increase is due to the movement of the linker, as mentioned earlier. The movement of the interdomain linker is not inhibited by the bound RNA. The atom-positional rmsd for the RNA

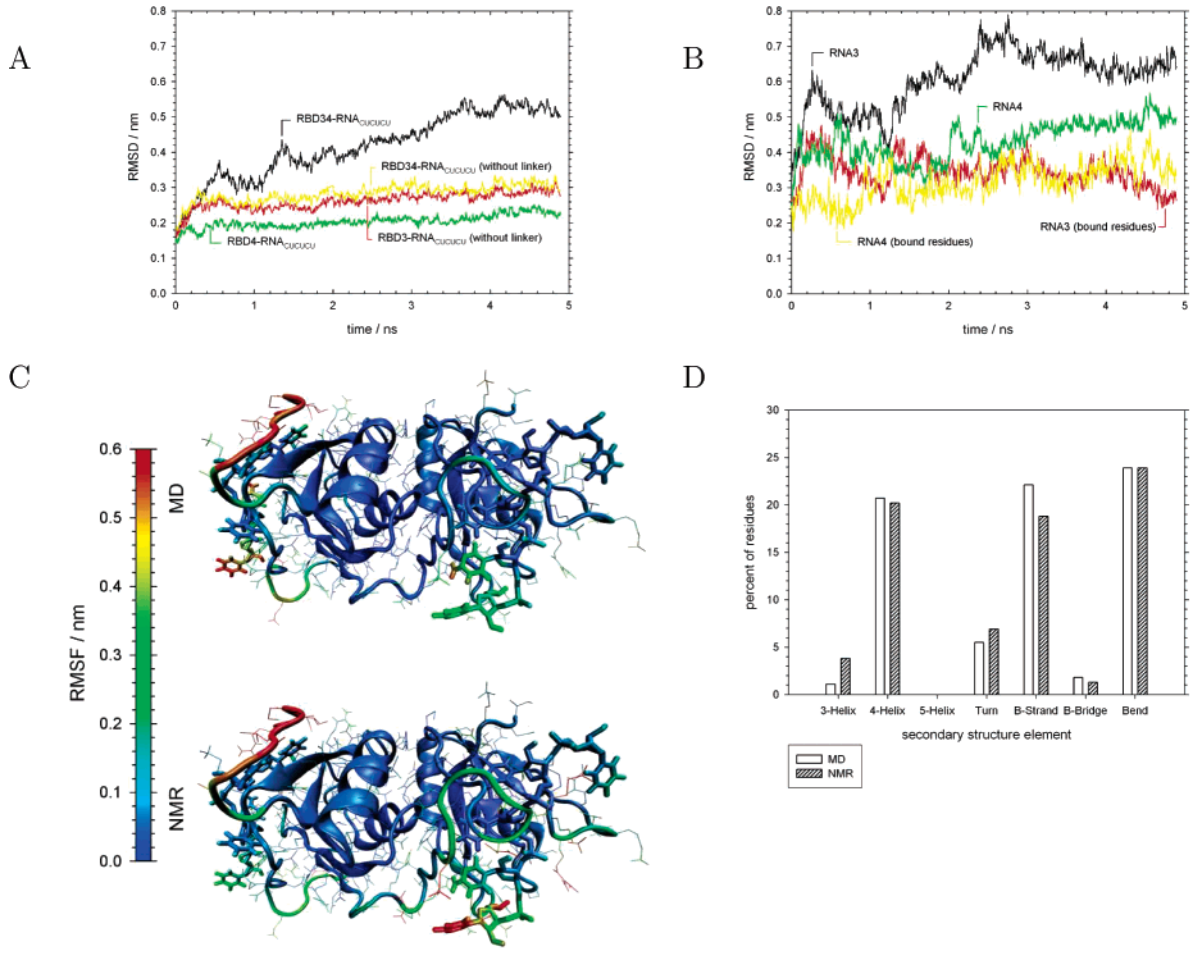

FIGURE 4: Analysis of the RBD34−RNA$_{CUCUCU}$ complex. (A) Backbone atom-positional rmsd of MD trajectory structures with respect to the energy-minimized NMR structure. (B) Atom-positional rmsd of all RNA atoms. (C) Values of rmsf, calculated from simulation trajectory coordinates (top) and structures of the NMR bundle (bottom), mapped onto the starting structure. (D) Secondary structure analysis. The white bars were computed from the simulation trajectory coordinates and the striped ones from the NMR bundle.

atoms is shown in Figure 4B. The rmsd of the bound residues reaches 0.3 nm, which is lower than the rmsd of all residues due to the movement and rotation of the unbound residues. The rmsfs calculated from the trajectory coordinates and the structures of the NMR bundle were mapped onto the starting structure (Figure 4C). The largest fluctuations are associated with the movement of the interdomain linker on RBD3, the interdomain linker between RBD3 and RBD4, and the unbound nucleotides. The rmsf from the simulation trajectory is very well correlated ($R = 0.82$) with the "mock fluctuations" calculated for the NMR bundle. The percentage of residues adopting different secondary structural elements is shown in Figure 4D, both for the MD trajectory and for the structures of the NMR bundle. The MD and NMR results agree well. Only minor changes in secondary structural elements are seen during the simulation period (Figure 5). As in the case of the unbound domains, the simulation was validated against the NMR data. A total of 4609 NOE upper bounds were analyzed, 312 of which are violated by an average of 0.11 ± 0.011 nm (Figure 6A). The atoms participating in violated upper bounds are distributed over the whole complex (Figure 6B). Although the overall violation average is low, there are a few sizable violations: the participating atoms cluster mostly in RBD4 and its binding pocket. As in the protein-only simulation, the Asn 154 side chain moves from the protein interior to the solvent in the equilibration period and causes the two largest violations. At the edge of a β-strand, the Ser 136 side chain shows a similar behavior: it flips out into the water and causes two large violations of the NOE upper bounds. The aromatic residues His 134 and Phe 164 separate in the simulation and begin to stack with the bases of the RNA ligand, which results in another sizable violation. Thus, a few of the measured interactions cannot be reproduced on the simulation time scale which may due to the force field inadequacy or to inaccuracy of the particular NOE bound derived from the NMR data. Yet, in the simulation, the RNA is tightly associated with RBD4, most of the intermolecular NOE upper bounds being only slightly violated. This can be taken as an indication that the specificity of binding of RBD4 is rather low.

The binding pockets of the two domains were analyzed in greater detail (Figure 7). Hydrogen bonds to the bases, salt bridges, and stacking interactions and their fluctuations were monitored over time. To recognize different RNA sequences, it is likely that the protein has to make specific hydrogen bonds to the bases. It is easily seen in Figure 7B that these specific hydrogen bonds are not very stable and do break during the simulation. Furthermore, some specific contacts in the interface described on the basis of NMR structure (*11*) were not reproduced by the simulation. In particular, in RBD3, the hydrogen bond from the protein (hydroxyl hydrogen HG of Thr 84) to Ura 2 is missing. The intramolecular RNA contact through the hydrogen bond

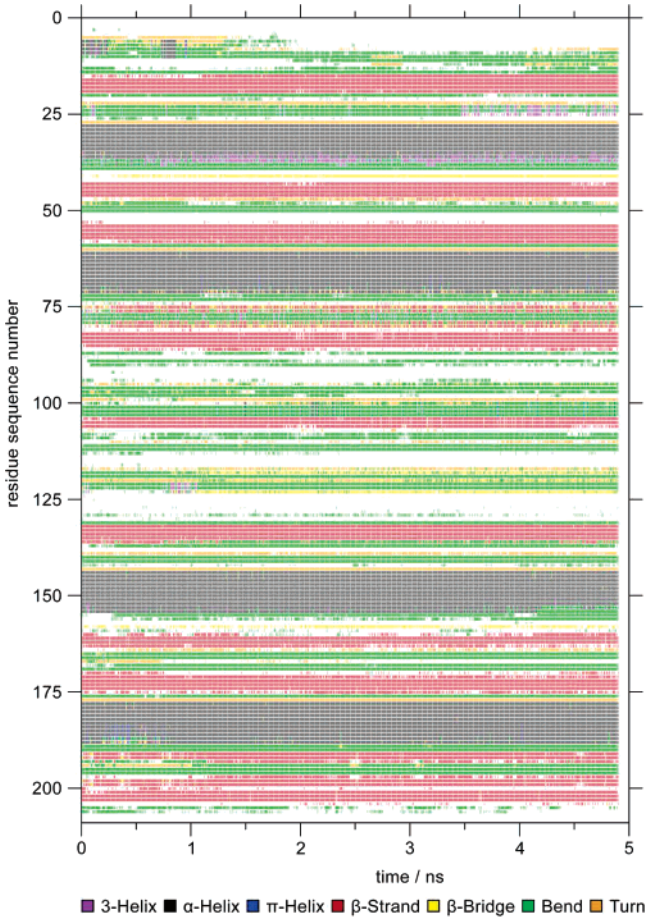

FIGURE 5: Time series of the secondary structure for all residues in the RBD34−RNA_{CUCUCU} simulation. Secondary structure types are defined in ref *26*.

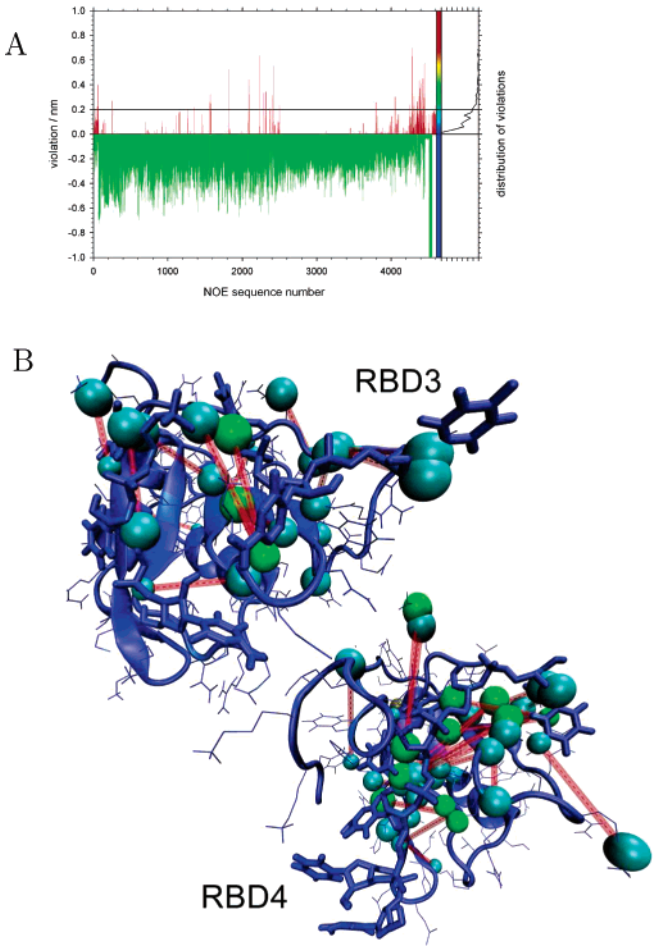

FIGURE 6: NOE analysis of the RBD34−RNA_{CUCUCU} complex. (A) Violations of NOE upper bounds (4609 analyzed)). Violated bounds (312, average of 0.11 ± 0.011 nm) are colored red and fulfilled ones green. The distribution histogram of the violations is shown at the right. (B) NOE violations mapped onto the starting structure. Atoms involved in large violations (violation of >0.2 nm, red lines) of the NOE bounds are shown as spheres. The color indicates the size of the violation.

between amino hydrogens (H4) of Cyt 3 and the carbonyl oxygen (O2) of Ura 2 is lost right from the start. For Cyt 3, the hydrogen bond from hydroxyl hydrogen HG of Ser 86 to the carbonyl acceptor (O2) of the base is unstable. Ura 4 is bound more specifically, and the hydrogen bonds to this base are present. The amide hydrogen (H) of Gln 92 recognizes the carbonyl acceptor (O2) on the base. This contact is made more specific by a second hydrogen bond from the amide hydrogen (H3) of the base to the backbone carbonyl (O) of Asn 90. It is worth mentioning that the stacking interaction of His 88 with Ura 4 is very stable, which also contributes to the stabilization of the specific hydrogen bond cluster of the base. Cyt 5 and Ura 6 show no specific hydrogen bonds to the protein domain: the amino hydrogens (H4) of Cyt 5 do not contact the phosphate of Ura 4, and thus, an intramolecular RNA contact is once more not seen. For RBD4, the binding appears to be even less specific. Only Cyt 3 forms stable hydrogen bonds to the protein domain. The position of Ura 2 is stabilized by a salt bridge between Arg 200 and the phosphate. This stabilization is not strong enough to strengthen the hydrogen bond between the amide hydrogen (H3) of the base and the hydroxyl oxygen (OG) of Ser 202. Again, the intramolecular contact of the RNA through the hydrogen bond from the amino hydrogens (H4) of Cyt 4 to the carbonyl acceptor (O2) of Ura 2 is not seen. Cyt 3 is recognized specifically by two very stable hydrogen bonds from the amino hydrogens (H4) of the base to the carbonyl oxygen (O) of Phe 203 and from the hydroxyl hydrogen (HG) of Ser 204 to the carbonyl acceptor (O2) of the base. This recognition is made even more specific by a third hydrogen bond from Lys 205's amino hydrogen (H) to the amide nitrogen (N3) of the base. A stacking interaction of His 134 is contributing to the recognition, but most of the time, the residues are not stacking, indicating that this interaction is less important. The salt bridge between Lys 205 and the phosphate of Cyt 3, deemed important on the basis of the NMR structure, was not observed. The hydrogen bonds of Ura 4 are very unstable and break and re-form regularly; the carboxyl end of the polypeptide chain is stabilized by a salt bridge-like contact to Lys 162. However, this contact is not stable enough to stabilize the hydrogen bonds from Lys 162 to the base. Finally, in our simulation, we do not see that Ura 4 makes any specific contacts to the protein. It appears that the description of the interface, as inferred from the NMR data, is different and also contains only a few specific contacts as one can see in the right panel of Figure 7B. Here, we should mention that the interactions of both protein domains with each other, especially the salt bridge, are preserved during the simulation (data not shown).

The same simulation setup as in the RBD34−RNA_{CUCUCU} case was used to analyze another sequence, namely the

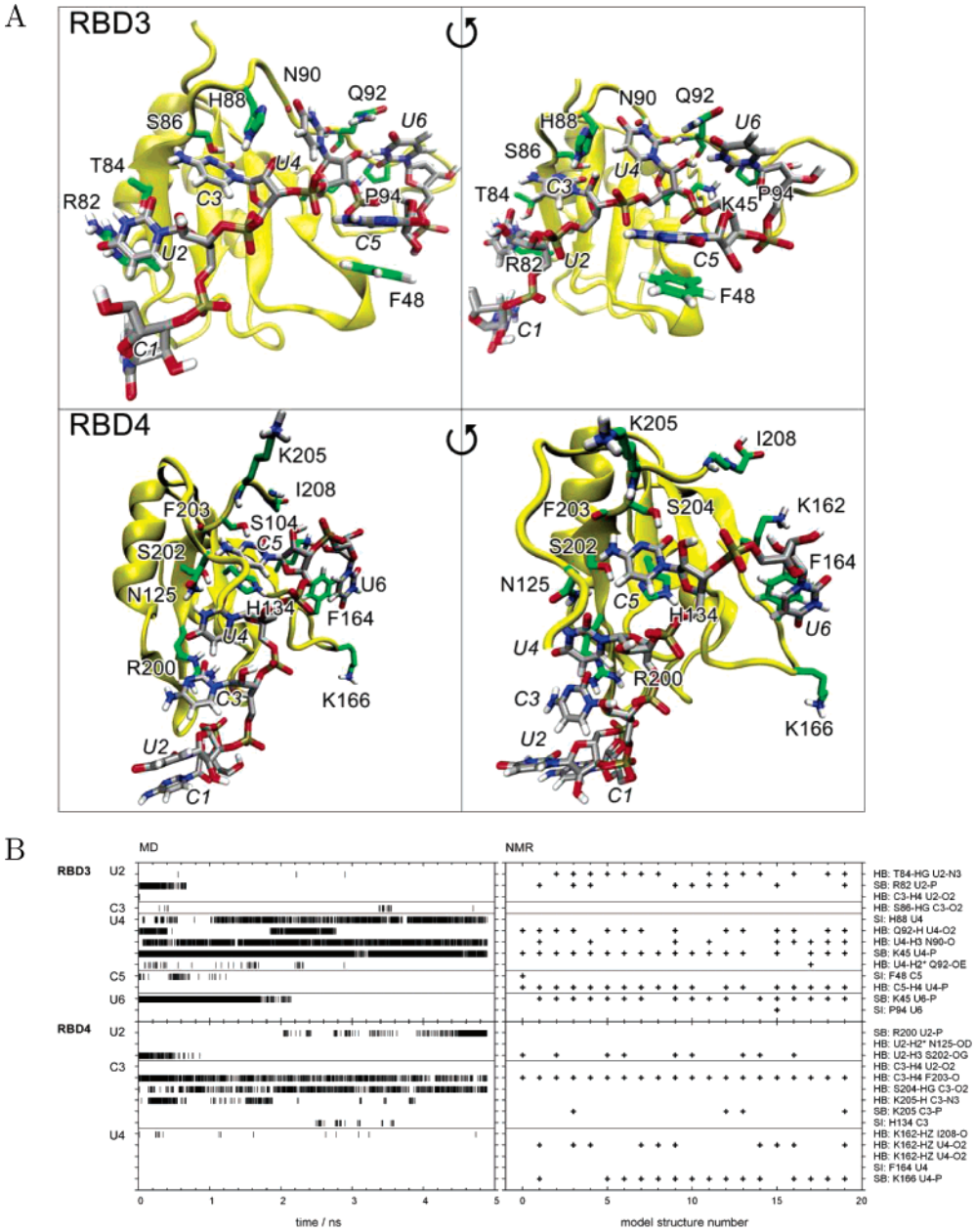

FIGURE 7: Analysis of the RBD34−RNA_{CUCUCU} interface. (A) Illustration of the binding pockets. The protein domains are colored yellow as a secondary structure cartoon. Side chain or backbone atoms of residues involved in the binding of the RNA as well as the RNA are shown in a ball-and-stick representation. (B) Time series of binding interactions (HB, hydrogen bond; SB, salt bridge; SI, stacking interaction). In RBD3 only U4 is specifically bound while in RBD4 only C3 is. The same computation using the 20 model structures of the NMR bundle gave the results shown at the right.

cytosine hexanucleotide (RNA_{CCCCCC}) (5 ns). For analysis, backbone atom-positional rmsd's of the RBD34−RNA complex, the backbone of single-domain complexes, and the RNA_{CCCCCC} with respect to the minimized NMR structure were calculated (data not shown). The rmsd value converges to ~0.5 nm for the whole complex (again, largest contribution from the flexible linker) and to ~0.22 nm for the single-domain complexes, which can be taken as an indication of the stability of the complex. From atom-positional rmsf calculations (data not shown), it was found that a greater number of nucleotides show enhanced movement than in the RNA_{CUCUCU} case, indicating that the RNA_{CCCCCC} is less tightly bound to RBD4 than RNA_{CUCUCU}. The last residues of the RNA are unbound and move freely, which was not the case in the RNA_{CUCUCU} complex.

To explain the different binding affinities of the two ligands, the binding pockets (interface) were analyzed in greater detail (Figure 8). For RBD3, it is seen that the most permanent contacts are three stable salt bridges and several hydrogen bonds. Sequence-specific binding is mainly achieved by recognition of the bases by hydrogen bonds to the protein backbone: the carbonyl oxygen of Cyt 2 (O2) is recognized by the hydroxyl hydrogen (HG) of Thr 84. One of the amino hydrogens (H41) of Cyt 3 makes specific contacts with the backbone carbonyl oxygen of Lys 87. This specific contact is stabilized by two less specific fluctuating hydrogen bonds from Gln 92's amide hydrogen to the carbonyl (O2) and nitrogen (N3) in the aromatic ring system of the base. Cyt 4 is bound by three hydrogen bonds from and to the backbone of the protein. All of these hydrogen bonds are not very stable

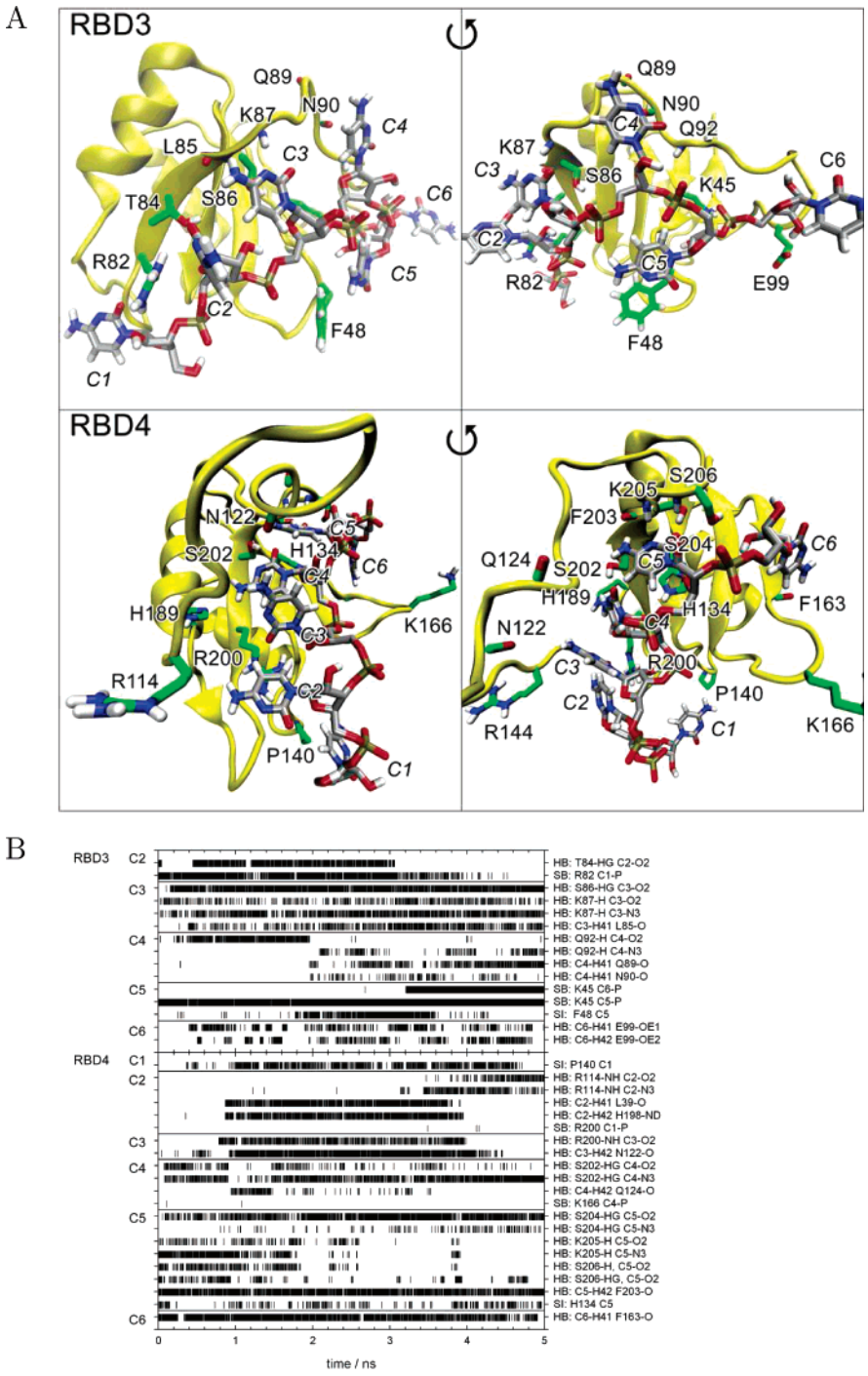

FIGURE 8: Analysis of the RBD34−RNA_{CCCCCC} interface. (A) Illustration of the binding pockets. The protein domains are colored yellow as a secondary structure cartoon. Side chain or backbone atoms of residues involved in the binding of the RNA as well as the RNA are shown in a ball-and-stick representation. (B) Time series of binding interactions (HB, hydrogen bond; SB, salt bridge; SI, stacking interaction). Most of the nucleotides are bound in both domains by rather unspecific hydrogen bonds.

and break and re-form regularly. A very stable salt bridge between Lys 45 and Cyt 5's phosphate is seen. This salt bridge is further stabilized by a second salt bridge to the phosphate of Cyt 6. This is an indication that the RNA_{CCCCCC} is bent around this lysine residue. A rather stable stacking interaction between Cyt 5 and Phe 48 contributes to the stabilization. Finally, Cyt 6's amino hydrogens (H41 and H42) make contacts to the carboxy group of Glu 99. For RBD4, none of the salt bridges seem to be stable. Here most of the permanent contacts are hydrogen bonds to the bases and the backbone. Cyt 1 and 5 are making stable stacking interactions which increase the overall stability of the complex. Cyt 2 makes a series of hydrogen bonds with its amino hydrogens (H41 and H42) and its carbonyl (O2) acceptor to several residues (Arg 114, Leu 39, and His 198). The specific recognition of Cyt 3 is achieved by two hydrogen bonds: Arg 200's amino hydrogens (NH) contact the base's carbonyl acceptor (O2), and the amino hydrogens (H41 and H42) of the base interact extensively with the backbone carbonyl of Asn 122. A hydrogen bond (whose donor and acceptor are present for both Cyt and Ura) between Ser 202 and Cyt 4's acceptor atoms (O2 and N3) contributes to the binding of Cyt 4. Furthermore, Cyt 5 is recognized very specifically by two hydrogen bonds: one between Ser

Table 2: Nonbonded Energies (kilojoules per mole) of RNA−Protein and RNA−Solvent Interactions[a]

	RNA_{CUCUCU}	RNA_{CCCCCC}	ΔE
	RBD3		
RNA−protein			
E_{nb}	−1870	−2010	−140
E_{LJ}	−400	−394	6
E_{el}	−1460	−1620	−160
RNA−solvent			
E_{nb}	−2200	−2110	90
E_{LJ}	−150	−137	13
E_{el}	−2050	−1970	80
	RBD4		
RNA−protein			
E_{nb}	−1590	−1920	−330
E_{LJ}	−386	−437	−51
E_{el}	−1210	−1480	−270
RNA−solvent			
E_{nb}	−2200	−2260	−60
E_{LJ}	−150	−97	53
E_{el}	−2050	−2170	−120

[a] The subscript nb stands for nonbonded, LJ for Lennard-Jones, and el for electrostatic energy. For both domains, RBD3 and RBD4, RNA_{CCCCCC} has more favorable interactions with the protein than RNA_{CUCUCU}.

204's HG and the carbonyl acceptor of the base (O2) and one between the base's amino hydrogen (H42) and the backbone carbonyl oxygen of Phe 163. This specific binding is further stabilized by the stacking of the base's aromatic system with His 134. Finally, Cyt 6 is recognized by a very stable hydrogen bond of the amino hydrogen of the base (H4) to the backbone carbonyl oxygen of Phe 163.

Again, it was found that RBD3 and RBD4 bind RNA_{CCCCCC} not as specifically as one may infer from the static description of the binding pocket in ref *11*. Both RBDs contain several loops which encompass the centrally located β-sheets. The backbone of these loops is accessible from the binding pocket on the β-strand, and it donates (by amide hydrogens) and accepts (by carbonyl oxygens) hydrogen bonds to the RNA. As these loops also exhibit increased flexibility (which was seen in the rmsf analysis), the overall shape of the binding pocket is not well-defined. It is thus reasonable to speculate that both RBDs bind a whole series of short polypyrimidine sequences. The simulations show that the specificity of binding is indeed low and hydrogen bond contacts to the bases can be made to both sequences. Further, it was also seen for both sequences that most of the hydrogen bonds from the protein to the RNA are made to the phosphate backbone or the sugar's hydroxyl oxygen atoms. In particular, hydrophobic contacts between the sugar's carbon skeleton, the bases, and the protein strengthen the binding.

Thermodynamics of Binding

Nonbonded Interactions. The energies obtained in the two RNA_{CUCUCU}− and RNA_{CCCCCC}−RBD34 complex simulations were analyzed in more detail. If entropic effects are neglected, the free energy difference of binding ($\Delta\Delta G^{binding}$) can be estimated from the enthalpy which is mainly given by the protein−ligand and water−ligand nonbonded energies. Table 2 lists the nonbonded energies of RNA_{CUCUCU} and RNA_{CCCCCC} and their differences. It is seen for both domains and RNA sequences that the electrostatic energy E_{el} contributes most to the nonbonded energy. Lennard-Jones interactions (E_{LJ}) appear to be less important. The differences between the energies of the two RNA sequences provide insights into the specificity of binding. On the one hand, RBD3 appears to have slightly less Lennard-Jones interaction with RNA_{CCCCCC}, and on the other hand, it has a stronger electrostatic interaction. This finally results in a lower nonbonded energy of the RBD3−RNA_{CCCCCC} complex. The situation is slightly different for RBD4. Here not only the electrostatic energy but also the Lennard-Jones energy is lower. The interactions of RNA_{CUCUCU} are energetically less favorable than the interactions of RNA_{CCCCCC}. These calculations give qualitative insight into the mechanism of binding: electrostatic interactions, including salt bridges and hydrogen bonds, are the most important interactions in the recognition of the RNA in the protein's binding pocket.

Table 3: Results (kilojoules per mole) of the Thermodynamic Integration for Two Perturbations[a]

perturbation	$\Delta G_{AB}^{complex}$	ΔG_{AB}^{water}	$\Delta\Delta G_{AB}^{binding}$
$RNA_{CCCCCC} \rightarrow PNA_{CUCUCU}$	414	420	−6
$RNA_{CUCUCU} \rightarrow PNA_{UCUCUC}$	501	508	−7

[a] First, RNA_{CCCCCC} appears to bind slightly less strongly to RBD3 than RNA_{CUCUCU}. Second, RNA_{UCUCUC} binds more strongly to RBD3 than RNA_{CUCUCU}.

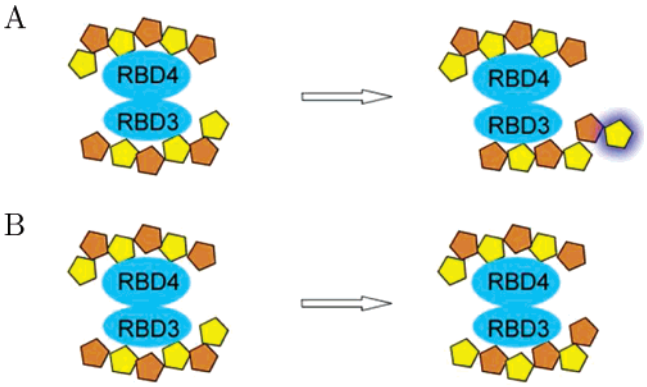

FIGURE 9: Shifting of the reading frame. (A) In the experiment, the shifting of the reading frame involves a physical movement of the whole ligand. The unbound nucleotide is indicated by the purple glow. (B) In the simulation, the shifting of the reading frame was realized using a perturbation of the first sequence (CUCUCU) into the second (UCUCUC).

Thermodynamic Integration. The free energy difference of binding of RNA_{CUCUCU} and RNA_{CCCCCC} was calculated using thermodynamic integration (TI). For the perturbation of RNA_{CCCCCC} to RNA_{CUCUCU} in the complex and in water, 26 λ points were sampled for 400 ps each. Integration of the $\langle \partial H/\partial\lambda \rangle_\lambda$ curves (data not shown) using the trapezoidal method and subtraction gave the relative binding free energy (Table 3). RNA_{CUCUCU} appears to bind better to RBD3 ($\Delta\Delta G = -6$ kJ/mol), although the free energy difference is small. However, the qualitative binding affinity of the two sequences is in agreement with experiment. It was shown using systematic evolution of ligands by exponential enrichment (SELEX) that alternating cytosine-uracil sequences $[(CU)_n]$ show enhanced affinity for PTB (*33*). In a second thermodynamic integration calculation, the CUCUCU sequence was frame-shifted to give the UCUCUC sequence. The perturbation of the first sequence into the second needed again special building blocks, including dummy atoms (Figure 1B). For

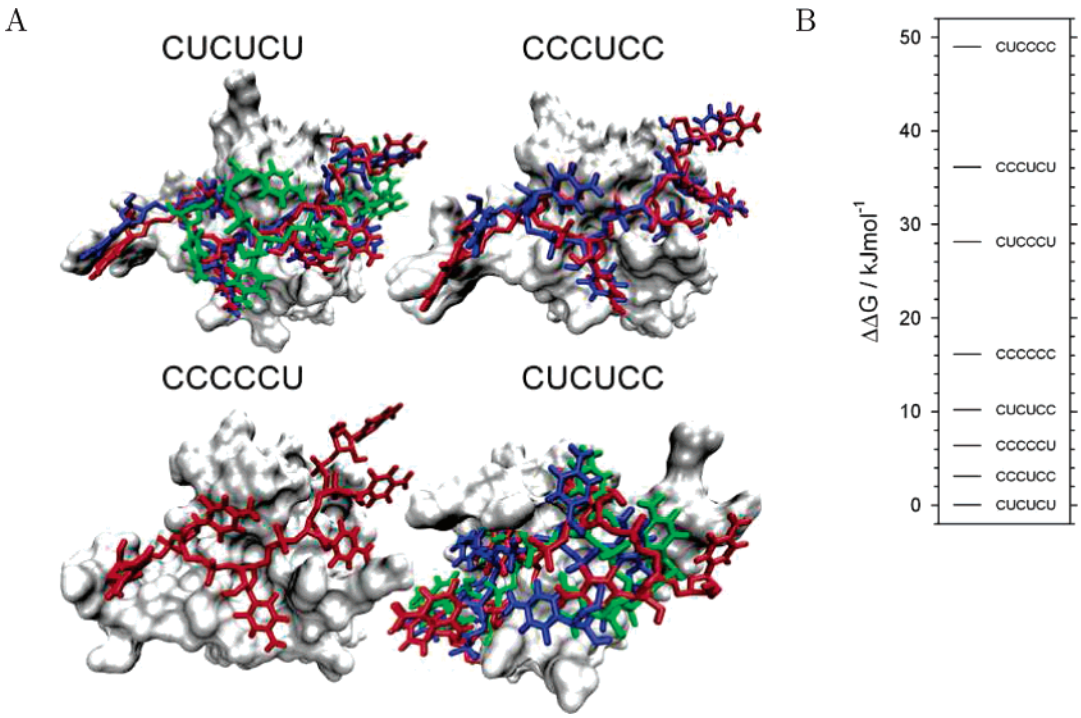

FIGURE 10: (A) Lowest-energy structures for the four best binding sequences. The surface of RBD3 is colored white. The ball-and-stick model represents the different RNA conformations. RBD4 and its bound RNA are not shown for the sake of simplicity. For most of the sequences, the obtained low-energy conformations differ significantly from each other. (B) Relative binding free energy of all uracil-cytosine combinations obtained in the single-step perturbation calculation. Comparison of the sequences shows that U_4 and U_6 contribute most to the free energy of binding.

the perturbation, 26 λ points (for the complex with RBD3 and in water) were simulated (Table 3). The finding that RNA_{UCUCUC} binds more strongly to RBD3 than RNA_{CUCUCU} ($\Delta\Delta G = -7$ kJ/mol) was surprising, because it appears to be in contradiction with the experimental NMR NOE data, where frame-shifting was not observed, but Figure 9 illustrates that in the experimental setting the frame-shifting event may be energetically less favorable than in our simulation. In the experiment, the shifting of the reading frame involves a physical movement of the whole ligand. The overall orientation of the RNA toward the protein changes, and the binding of one nucleotide to the protein becomes looser (Figure 9A). In our simulation, the shifting of the reading frame was realized using a perturbation of the first sequence (CUCUCU) into the second (UCUCUC). Much less physical movement and no unbound residues are involved in this process (Figure 9B). Finally, the finding that RNA_{UCUCUC} binds to RBD3 and that the difference in free energy is rather small indicates that the domain may move along UC-rich sequences. This enables the domain to search a RNA molecule for the best position, where all RBDs are bound to the most favorable sequences.

Single-Step Perturbation. Relative free energies of binding for a series of ligands were calculated using the single-step perturbation approach. Every second nucleotide in the CCCCCC hexamer was perturbed to an unnatural reference state with properties between those of uracil and cytosine. This was achieved by coupling of the Hamiltonians of uracil (H_{URA}) and cytosine (H_{CYT}) with the coupling parameter λ set to 0.5. The differing atoms and bonds between uracil and cytosine were softened as described earlier (see Materials and Methods) using the following parameters: $\alpha_{LJ} = 1.5$ and $\alpha_C = 0.5$ nm^2. This reference state bound to RBD3 was sampled for 2 ns, while the free one in water was sampled for 4 ns. The difference in free energy of sequences A and B was calculated using eq 4. The energy was calculated from the trajectory coordinates and the topologies for different sequences. For most sequences, several low-energy conformations were sampled but not for RNA_{CCCCCC} and RNA_{CCCCCU}. For these two sequences, no lower-energy conformation after the initial structure could be found, as shown by smoothly rising cumulative ΔG_{AR} curves (*34*). Figure 10A shows a superposition of the lowest-energy conformations for each of the four best binding sequences. For the RNA_{CCCCCU} hexanucleotide, no other low-energy conformation besides the starting structure was sampled. For RNA_{CCCUCC}, two low-energy conformations are shown; for RNA_{CUCCCC}, RNA_{CUCUCC}, and RNA_{CUCUCU}, three are shown. The different low-energy conformations do differ significantly from each other. This may also be an indication of the coexistence of many different binding patterns to the RBD3 domain and of the fact that many small interactions contribute to the free energy of binding. The relative binding free energies are given in Figure 10B. Again it is seen that the RNA_{CUCUCU} hexanucleotide binds best to RBD3. Furthermore, the sequences with uracil nucleotides at positions 4 and 6 appear to bind better than the ones with cytosine. A possible explanation may be that U_4 and U_6 contribute most to the free energy of binding. However, we note that the values obtained from the approximate single-step perturbation calculation are not in agreement with the ones from the more accurate thermodynamic integration calculation (Table 3).

CONCLUSIONS

The results obtained in this study show that the GROMOS 45A4 set of force field parameters is reasonably successful in the simulation of RNA−protein complexes. All the

simulated complexes were stable, and most of the experimental data could be reproduced.

The analysis of the RNA binding interface is a time-consuming undertaking: many different interactions, including salt bridge networks, hydrogen bond patterns, and hydrophobic and stacking interactions, must be taken into consideration. It was shown that the specificity of binding of RBD3 and RBD4 to RNA is low. The described weak and unspecific binding affinity is indeed needed for the recognition of many different polypyrimidine motifs. One may speculate that the recognition of the RNA with the polypyrimidine tract binding protein is a cooperative process which involves not only the binding of single domains to RNA but also a combination of all interactions of all domains with the RNA in its single-stranded form. It may also be possible that PTB simply stabilizes RNA−protein complexes needed for alternative splicing like it does with the IRES complex in IRES-mediated initiation of translation (*5*). It is clear that alternative splicing is a complex process and further research has to be done to understand it in more detail.

The accomplished thermodynamic calculations of the free energy of binding show results which are in agreement with the literature. Although the agreement with experimental data is qualitatively good, the obtained quantitative results of the calculations may be inaccurate. Accurate calculation of the relative free energy using the thermodynamic integration and single-step perturbation methods needs sampling of large conformational ensembles. Given the size of the RBD34−RNA complex, this is currently not feasible.

The qualitative relative free energies calculated from both methods gave further insight into the mechanism of binding: it was shown that the free energy differences between the different polypyrimidine sequences are low and that frame-shifting may be energetically favorable. This suggests a mechanism of sliding of the domains along the RNA, which gives PTB a new dynamic role in the regulation of alternative splicing.

ACKNOWLEDGMENT

We thank Florian C. Oberstrass and Frédéric H.-T. Allain for discussions and for providing the NMR refinement data.

REFERENCES

1. Smith, C. W., and Valcarcel, J. (2000) Alternative pre-RNA splicing: The logic and combinatorial control, *Trends Biochem. Sci. 25*, 381−388.
2. Wagner, E. J., and Gracia-Blanco, M. A. (2001) Polypyrimidine tract binding protein antagonizes exon definition, *Mol. Cell. Biol. 10*, 3281−3288.
3. Chou, M.-Y., Underwood, J. G., Nikolic, J., Luu, M. H. T., and Black, D. L. (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing, *Mol. Cell 5*, 949−957.
4. Hershey, J. W. B., and Merrick, W. C. (2000) *The pathway and mechanism of initiation of protein synthesis*, Cold Spring Harbor Laboratory Press, Plainview, NY.
5. Hellen, C. U. T., and Sarnow, P. (2001) Internal ribosome entry sites in eukaryotic mRNA molecules, *Genes Dev. 15*, 1593−1612.
6. Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. (2004) Polypyrimidine tract binding protein modulates efficiency of polyadenylation, *Mol. Cell. Biol. 24*, 4174−4183.
7. Knoch, K.-P., Bergert, H., Borgonovo, B., Saeger, H.-D., Altkrüger, A., Verkade, P., and Solimena, M. (2004) Polypyrimidine tract-binding protein promotes insulin secretory granule biogenesis, *Nat. Cell Biol. 6*, 207−214.
8. Simpson, P. J., Monie, T. P., Szendröi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S., and Matthews, S. (2004) Structure and RNA interactions of the N-terminal RRM domains of PTB, *Structure 12*, 1631−1643.
9. Conte, M. R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., and Curry, S. (2000) Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RMM fold, *EMBO J. 19*, 3132−3141.
10. Yuan, X., Davydova, N., Conte, M. R., Curry, S., and Matthews, S. (2002) Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein, *Nucleic Acids Res. 30*, 456−462.
11. Oberstrass, F. C., Auweter, S. D., Erat, M., and Allain, F. H.-T. (2005) Structure of PTB bound to RNA: Specific binding and implications for splicing regulation, *Science 309*, 2054−2057.
12. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*, *Nat. Struct. Biol. 2*, 537−547.
13. Allain, F. H., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin, *EMBO J. 19*, 6870−6881.
14. Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., and Feigon, J. (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target, *J. Mol. Biol. 337*, 799−816.
15. Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., and Allain, F. H. (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1, *EMBO J. 25*, 163−173.
16. Soares, T. A., Hünenberger, P. H., Kastenholz, M. A., Kräutler, V., Lenz, T., Lins, R. D., Oostenbrink, C., and van Gunsteren, W. F. (2005) An improved nucleic acid parameter set for the GROMOS force field, *J. Comput. Chem. 26*, 725−737.
17. Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Kräutler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and van Gunsteren, W. F. (2005) The GROMOS software for biomolecular simulation: GROMOS05, *J. Comput. Chem. 26*, 1719−1751.
18. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) *Intermolecular Forces* (Pullmann, B., et al., Eds.) Reidel, Dordrecht, The Netherlands.
19. Berendsen, H. J. C., Postma, J. P. M., DiNola, A., van Gunsteren, W. F., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684−3690.
20. Heinz, T. N., van Gunsteren, W. F., and Huenenberger, P. H. (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations, *J. Chem. Phys. 115*, 1125−1135.
21. van Gunsteren, W. F., Daura, X., and Mark, A. E. (2002) Computation of free energy, *Helv. Chim. Acta 85*, 3113−3129.
22. Kirkwood, J. G. (1935) Statistical mechanics of fluid mixtures, *J. Chem. Phys. 3*, 300−313.
23. Zwanzig, R. W. (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys. 22*, 1420−1426.
24. Beveridge, D. L., and DiCapua, F. M. (1989) Free energy via molecular simulation: Applications to chemical and biomolecular systems, *Annu. Rev. Biophys. Biophys. Chem. 18*, 431−492.
25. Schäfer, H., van Gunsteren, W. F., and Mark, A. E. (1999) Estimating relative free energies from a single ensemble: Hydration free energies, *J. Comput. Chem. 20*, 1604−1617.
26. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577−2637.
27. Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and van Gunsteren, W. F. (2007) Configurational entropy elucidates the role of salt-bridge networks in protein thermostability, *Protein Sci*. (in press).
28. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular simulation: The GROMOS96 manual and user guide*, Hochschulverlag AG, ETH, Zurich.
29. Billeter, M., Braun, W., and Wüthrich, K. (1983) Pseudo-structures for the 20 common amino-acids for use in the studies of protein

conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance, *J. Mol. Biol. 169*, 949−961.
30. Oostenbrink, C., Soares, T. A., van der Vegt, N. F. A., and van Gunsteren, W. F. (2005) Validation of the 53A6 GROMOS force field, *Eur. Biophys. J. 34*, 273−284.
31. Fletcher, C. M., Jones, D. N. M., Diamond, R., and Neuhaus, D. (1996) Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures, *J. Biomol. NMR 8*, 292−310.
32. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 3338.
33. Perez, I., Lin, C. H., McAfee, J. G., and Patton, J. G. (1997) Mutation of PTB binding sites causes misregulation of alternative 3′ splice site selection in vivo, *RNA 3*, 764−778.
34. Oostenbrink, C., and van Gunsteren, W. F. (2004) Free energy of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation, *Proteins 54*, 237−226.

BI6026133