JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein−RNA Complexes?

M. Krepl,[†] M. Havrila,[†,‡] P. Stadlbauer,[†] P. Banas,[∥] M. Otyepka,[∥] J. Pasulka,[‡,§] R. Stefl,[‡,§] and J. Sponer*[,†,‡]

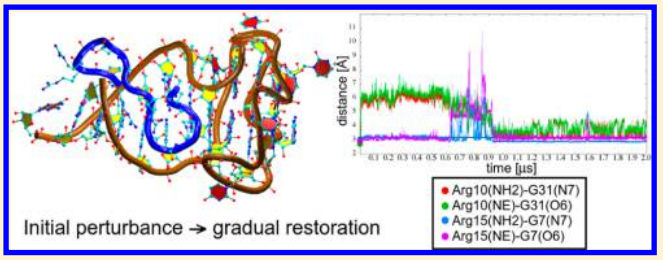
[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

[‡]Central European Institute of Technology (CEITEC), Masaryk University and [§]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[∥]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University, Tř. 17 Listopadu 12, 771 46 Olomouc, Czech Republic

Supporting Information

ABSTRACT: We report over 30 μs of unrestrained molecular dynamics simulations of six protein−RNA complexes in explicit solvent. We utilize the AMBER ff99bsc0χ_{OL3} RNA force field combined with the ff99SB protein force field and its more recent ff12SB version with reparametrized side-chain dihedrals. The simulations show variable behavior, ranging from systems that are essentially stable to systems with progressive deviations from the experimental structure, which we could not stabilize anywhere close to the starting experimental structure. For some systems, microsecond-scale simulations are necessary to achieve stabilization after initial sizable structural perturbations. The results show that simulations of protein−RNA complexes are challenging and every system should be treated individually. The simulations are affected by numerous factors, including properties of the starting structures (the initially high force field potential energy, resolution limits, conformational averaging, crystal packing, etc.), force field imbalances, and real flexibility of the studied systems. These factors, and thus the simulation behavior, differ from system to system. The structural stability of simulated systems does not correlate with the size of buried interaction surface or experimentally determined binding affinities but reflects the type of protein−RNA recognition. Protein−RNA interfaces involving shape-specific recognition of RNA are more stable than those relying on sequence-specific RNA recognition. The differences between the protein force fields are considerably smaller than the uncertainties caused by sampling and starting structures. The ff12SB improves description of the tyrosine side-chain group, which eliminates some problems associated with tyrosine dynamics.



INTRODUCTION

RNA molecules perform numerous functions in living organisms including gene expression, defensive response, and chemical catalysis. In vivo, almost all RNA molecules at all stages of their life interact with proteins, forming ribonucleoprotein complexes.[1−4] Their composition is diverse and highly dynamic. The interaction with proteins is essential to protect the RNA molecule from degradation, assist in the RNA folding, deliver the RNA to correct location in the cell, and regulate the gene expression.[5−9] The proteins can assist or actively participate in the RNA catalytic function (ribosome, spliceosome).[10,11] Insight into principles of protein−RNA interactions is therefore an important step in understanding of numerous biological processes.[1,2,4,6,12−14] Protein−RNA recognition mechanisms can include sequential and shape specificity.[15−17] Sequential specificity directly depends on the chemical identity of the interacting residues, (i.e., the amino acids side-chains) or a specific conformation of the main-chain, which can read out the unique chemical fingerprint of RNA bases. In contrast, the shape-specific binding involves main-chains and basic side-chains that recognize specific shape of the RNA sugar−phosphate backbone regardless of its sequence. Both recognition mechanisms are often combined in the protein−RNA complexes. Further, targeted chemical modifications (hydroxylation, methylation, etc.) of the individual RNA and protein moieties can further modulate the sequential specificity.

Structural recognition is given by the complementary molecular shapes of the participating molecules. In general, large interaction surface between the two molecules leads to thermodynamically more stable complexes. The size of protein−RNA interaction surface is often evaluated by the solvent accessible surface area (SASA) approximation.[18] With SASA, the interaction surface area of the complex is computed as the difference between the solvent accessible surface of free individual biomolecules and that of the complex.

Atomistic explicit solvent molecular dynamics (MD) simulations represent an important tool to study biomacromolecules and their molecular complexes. Although MD

Received: September 5, 2014
Published: January 20, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

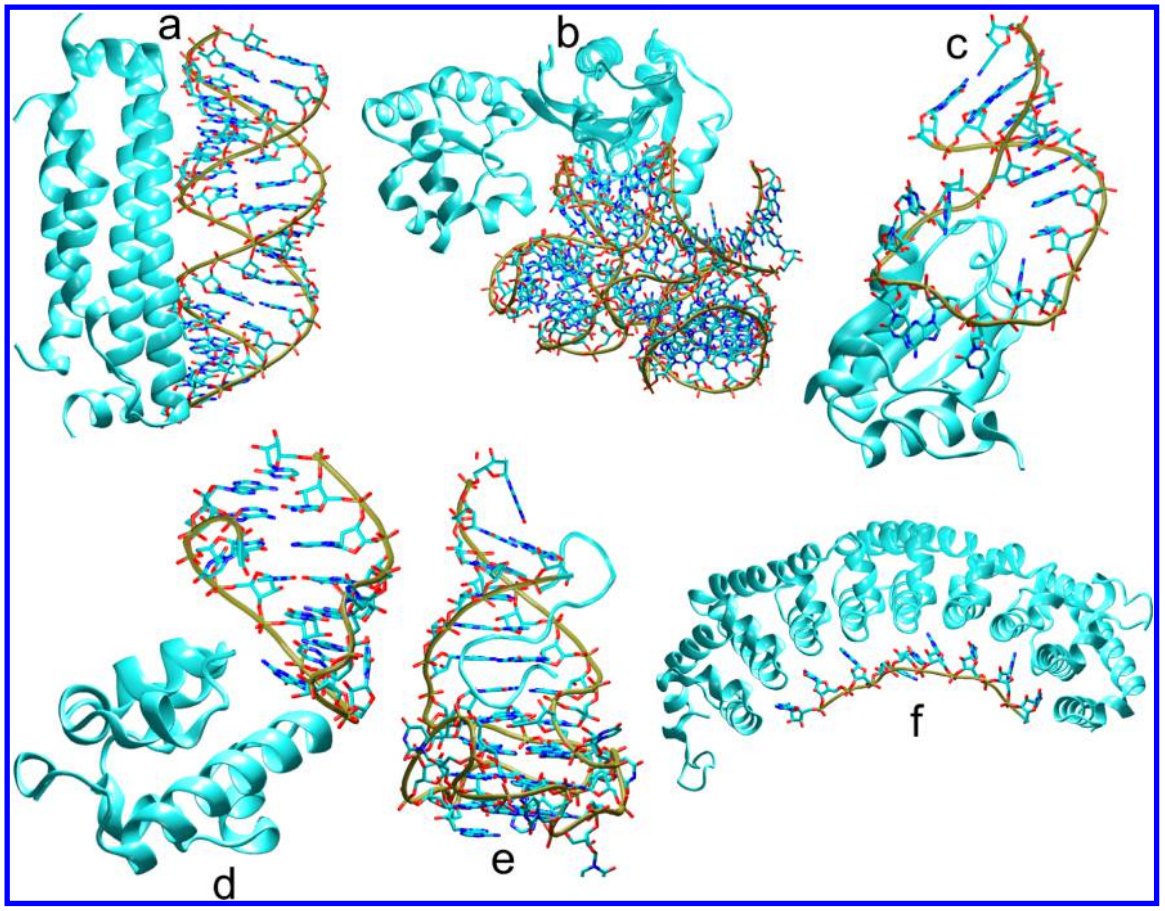


Figure 1. Studied protein−RNA complexes: (a) DsRNA and the B2 protein; (b) the L1 protein with the 23S rRNA fragment; (c) the U1A protein and an RNA hairpin; (d) the SAM domain of Vts1p with an RNA hairpin; (e) the FMRP RGG peptide with an RNA duplex−quadruplex junction; (f) the Pumilio FBF-2 protein with the gld-1 FBEa RNA. In the Supporting Information, we provide detailed pictures of the protein−RNA interfaces.

simulation studies are limited by affordable sampling (length of the simulations compared to real biomolecular dynamics) and force field approximations (difference between the real molecules and the potential energy model used to describe them), they can supply important insights that cannot be obtained by experiments.[19,20] The most important advantage of MD simulations is their capability to monitor all aspects of the time development of the 3D structure with picosecond-scale time resolution. However, the number of simulation studies of protein−RNA complexes in contemporary MD literature is surprisingly small. The published work includes, for example, investigations of the U1A protein−RNA complex[21−24] and some other systems.[25−32] Further, many of the published studies of protein−RNA complexes do not discuss details of the simulations and their accuracy. It is then not straightforward to judge which part of the results reflects the real properties of the studied system and which part limitations of the methodology, such as the effects of the starting structure, limited sampling, and force field approximations.[33,34] As shown in this paper, the accuracy of simulations of protein−RNA complexes is a delicate issue. When making unbiased (standard) MD simulations using good quality atomistic experimental structures as the start, we usually should not see any dramatic structural changes on the presently affordable simulation time scales, considering the k_{off} values of real biomolecular complexes. In other words, the simulated systems should be essentially structurally stable. As we pointed out elsewhere, it is very risky to attempt using unstable simulations to discuss, for example, the unfolding pathways, since when "unfolding" is caused by inaccurate force field or starting structure as the "denaturating condition", the whole trajectory may be entirely unrealistic.[20]

In contrast to numerous studies benchmarking force fields for RNA and protein simulations,[31,35−45] contemporary MD literature does not offer any study that would systematically address behavior of protein−RNA simulations. As shown below, description of protein−RNA complexes is affected by balance of the interactions at the protein−RNA interface, which cannot be judged from studies of RNA and proteins alone.

Clear requirement for successful MD simulations of complex biomolecular systems such as protein−RNA complexes are unambiguous atomistic experimental structures.[46] Those are obtained primarily by methods of X-ray crystallography and nuclear magnetic resonance (NMR). While both methods provide unique structural information, their limitations must be also considered, as discussed in ref 47. X-ray crystallography is primarily limited by the necessity to obtain a solid crystal of the studied biomolecule, which may differ from the in vivo environment. Packing of the biomolecules in the crystal lattice can lead to "crystal packing" effects, which can affect the resulting atomistic models. The X-ray structure represents a static averaged picture of a molecular complex that can be naturally quite dynamical. Further, X-ray structures of biochemically relevant protein−RNA complexes often do not reach the highest resolution. Still, the X-ray crystallography is

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

the leading tool in contemporary structural biology, and we do not have any better experimental data that could be used to initiate the simulations and to serve as benchmark for simulations of larger biomolecular systems. Experience gained by benchmark simulations of small model systems (RNA hairpins, short peptides, etc.) is not fully transferable to studies of larger biomolecular systems. NMR measurements are usually conducted in aqueous solution environment, thus bypassing the issues associated with biomolecule crystallization. Still, water environment differs from the crowded molecular environments occurring under in vivo conditions. NMR is more limited by the size of the biomolecule that can be studied at the atomistic level and the NMR atomistic models may be more ambiguous compared with the best resolved X-ray structures.

In our study, we explored six diverse protein−RNA complexes using MD simulations (Figure 1). The primary aim of our work was to assess capability of MD simulations to describe the protein−RNA complexes. Five complexes were selected from the database of high-resolution (better than 2.6 Å) X-ray structures. Generally, we excluded incomplete structures (a common feature of protein−RNA complexes determined by X-ray crystallography) and structures with a significant structural role of divalent ions, due to poor force field description of divalent ions.[19,20] The remaining structure has been solved by NMR. These six structures were selected from an original preselection of a few dozens of structures (see Supporting Information for the list), with the aim to simulate as diverse set of protein−RNA interactions as possible.

Our first system is an X-ray structure of Flock house virus (FHV) protein B2 in complex with dsRNA (PDB code 2az0; resolution 2.55 Å).[48] This protein functions as a powerful suppressor of the RNA-silencing defense of the host cell and is highly expressed in early stages of the infection.[49,50] The FHV B2 protein acts as a dimer, forming a four-helix bundle interacting with the backbone of the dsRNA molecule along its helical axis in a strictly sequence independent manner. Multiple FHV B2 protein dimers are able to coat the entire viral genome, protecting it from actions of the cellular Dicer complex. In addition to the genomic RNA, it is also able to bind already diced siRNAs and thus prevent RNA-silencing simultaneously on two levels. The B2 protein dimer in the 2az0 structure is interacting with an 18 nt long dsRNA helix. Since the RNA duplex is palindromic, all noted H-bonds interactions (with three exceptions) are present in the complex with a near complete symmetry.[48]

The second system is an X-ray structure of ribosomal L1 protein in complex with 80 nt fragment of 23S rRNA from *Thermus thermophilus* (3u4m; 2.0 Å).[51] This system constitutes the L1 stalk of the large ribosomal subunit. The L1 stalk has been implicated in release of the deacylated tRNA during the proteosynthesis.[52] It has been extensively studied by both experimental and theoretical methods.[31,52−55] The L1 is a two domain protein. It interacts with the 23S RNA in a highly specific way and with a large affinity.[54] The RNA molecule forms a highly intricate tertiary structure, which includes three A-RNA helices, two kink-turns, a tetraloop, and a structured loop. All conserved protein−RNA interactions are located between the domain I of the L1 protein and one of the RNA helices.[51] In addition, the L1 protein is also recognizing the kink-turn Kt-77, which has a unique 9 nt bulge.[31,56] These RNA segments are recognized based on their shapes and these parts of the protein−RNA interface lack sequence-specific H-bonds. Extensive structural description of this system can be found elsewhere.[31]

The third system is an X-ray structure of RNA-binding domain of U1A protein complexed with an RNA hairpin (1urn; 1.92 Å).[57] The U1A protein is a part of the U1 snRNP spliceosomal complex.[11] This protein−RNA system was extensively studied in the past by theoretical methods, including MD simulations.[21,23,58−60] The X-ray structure contains an N-terminal domain of the U1A protein complexed with 21 nt long RNA sequence that is representing one of the hairpins formed by the U1 snRNA.[57]

The fourth system is an X-ray structure of the SAM domain of Vts1p protein in complex with an RNA hairpin (2f8k; 2.0 Å).[61] Vts1p is a protein in *Sacharomyces cerevisiae* that influences the cell development by binding to 3′-UTR elements of specific transcripts and inducing their degradation.[62] This protein−RNA contact is facilitated through SAM (sterile α-motif) domain, which can interact with RNA stem−loop structures called SREs (stem−loop recognition elements).[63] The specific protein−RNA interactions are formed only with the loop part of the RNA which can be 5 nt or 4 nt long.[62]

The fifth protein−RNA system is an NMR structure of FMRP RGG peptide complexed with an RNA duplex−quadruplex junction (2la5).[64] FMRP protein is a regulatory protein in humans that binds with high affinity to guanine-rich RNA sequences capable of forming G-quadruplexes.[65] Loss of FMRP protein activity is a cause of X syndrome, which is one of the most common inherited mental diseases in humans.[66] The binding of FMRP to guanine-rich RNA is facilitated by the RGG box RNA binding domain, which is rich in arginine and glycine.[67] In the studied system, a 17 amino acids long RGG peptide is complexed with 36 nucleotide long guanine-rich RNA sequence. The RNA molecule is composed of a discontinuous three-tetrad guanine quadruplex, a mixed tetrad containing nonguanine bases, and a canonical RNA duplex. The peptide is interacting exclusively with the deep groove of the duplex. In addition, it is specifically recognizing the first two GC base pairs of the duplex (counted from the mixed tetrad) by forming an interaction between arginine side chains and the Hoogsteen edges of the guanines.[64] Note that the RNA is not structured in absence of the protein and vice versa.

The sixth system is an X-ray structure of *C. elegans* Pumilio FBF-2 protein in complex with the gld-1 FBEa RNA (3k5y, 2.3 Å).[68] This protein is a member of the Pumilio/FBF eukaryotic protein family. These eukaryotic proteins are part of mRNA regulatory systems. They regulate cell development by activating, repressing, or destroying target mRNAs.[69] They are responsible for regulation of stem cell maintenance and cellular memory. By binding to the mRNA 3′-UTRs elements, they typically decrease (or rarely increase) the mRNA expression.[70] The individual Pumilio proteins are specific for different target mRNA sequences and participate in a great variety of biological processes. This specificity is established through a conserved eight-repeat RNA-binding domain. Each repeat contains three α-helices, which together form typical crescent shape of the Pumilio proteins.[71] The interactions with mRNA are facilitated by three amino acid residues from the second α-helix of each repeat. Two of these residues are interacting with base edges of RNA nucleotides and are responsible for the specificity of the Pumilio proteins toward different mRNAs. The third residue is usually stacked between two bases. There are no contacts between the protein and the mRNA backbone, which is fully exposed to the solvent.[68]

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

Table 1. List of Simulations

simulation name	PDB code	protein force field[a]	simulated structure	simulation length (ns)
complex of dsRNA and the B2 protein				
2az0_ff99	2az0	ff99SB	protein + RNA	1000
2az0_ff12	2az0	ff12SB	protein + RNA	1000
complex of L1 protein with 23S rRNA fragment				
3u4m_ff99	3u4m	ff99SB	protein + RNA	1000
3u4m_ff12	3u4m	ff12SB	protein + RNA	1000
3u4m_ff99_noMG[b]	3u4m	ff99SB	protein + RNA	500
3u4m_ff99_tyr1[c]	3u4m	ff99SB	protein + RNA	500
3u4m_ff99_tyr2[d]	3u4m	ff99SB	protein + RNA	500
3u4m_ff99_pr	3u4m	ff99SB	protein	1000
3u4m_ff12_pr	3u4m	ff12SB	protein	1000
3u4m_ff12_2[d]	3u4m	ff12SB	protein + RNA	100
3u4m_ff12_3[d]	3u4m	ff12SB	protein + RNA	100
3u4m_ff12_rst[d,e]	3u4m	ff12SB	protein + RNA	300
complex of domain of U1A protein with RNA hairpin				
1urn_ff99	1urn	ff99SB	protein + RNA	1000
1urn_ff12	1urn	ff12SB	protein + RNA	1000
1urn_ff99_2	1urn	ff99SB	protein + RNA	500
1urn_ff12_2	1urn	ff12SB	protein + RNA	500
1urn_ff99_3	1urn	ff99SB	protein + RNA	500
1urn_ff12_3	1urn	ff12SB	protein + RNA	500
complex of SAM domain of Vts1p with RNA hairpin				
2f8k_ff99	2f8k	ff99SB	protein + RNA	500
2f8k_ff12	2f8k	ff12SB	protein + RNA	1500
2f8k_ff12_GC[f]	2f8k	ff12SB	protein + RNA	500
2f8k_ff99_no5_GC[g]	2f8k	ff99SB	protein + RNA	1000
2f8k_ff99_no5_GC_2[g]	2f8k	ff99SB	protein + RNA	1000
2f8k_ff12_no5_GC[g]	2f8k	ff12SB	protein + RNA	1000
2f8k_ff12_no5_GC_2[g]	2f8k	ff12SB	protein + RNA	1000
complex of FMRP RGG peptide with RNA duplex−quadruplex junction				
2la5_ff99	2la5	ff99SB	protein + RNA	1000
2la5_ff12	2la5	ff12SB	protein + RNA	1000
2la5_ff12_3K[h]	2la5	ff12SB	protein + RNA	2000
2la5_ff12_3K_2[h]	2la5	ff12SB	protein + RNA	1000
2la5_RNA	2la5	–	RNA	1000
complex of Pumilio FBF-2 protein with gld-1a RNA				
3k5y_0[i]	3k5y	ff99SB	protein + RNA	1000
3k5y_ff99[j]	3k5y	ff99SB	protein + RNA	1000
3k5y_ff99_rst[j]	3k5y	ff99SB	protein + RNA	1000
3k5y_ff12[j]	3k5y	ff12SB	protein + RNA	1000
3k5y_ff12_rst[j]	3k5y	ff12SB	protein + RNA	1000
3k5y_prot_ff99[j]	3k5y	ff99SB	protein	500
3k5y_prot_ff12[j]	3k5y	ff12SB	protein	500

[a]Different force fields (ff99SB or ff12SB) were used for proteins. For RNAs, ff99bsc0$_{\chi OL3}$ force field was used (see Methods). [b]The simulation starts from the end point (1000 ns) of the 3u4m_ff99 simulation, all Mg^{2+} ions were removed from the system and replaced with K^+. [c]The Tyr7 hydroxyl group orientation was automatically assigned (CE1-CZ-OH-HH = 0°) by xleap during system building (the same as in the 3u4m_ff99 simulation). [d]The Tyr7 hydroxyl group was manually adjusted during system building to form an H-bond interaction. [e]The G2124(O2′)−Glu42(OE2) interaction was stabilized by distance harmonic restraint for the first 100 ns of the simulation. [f]The two AU base pairs at the RNA helix end were replaced with GC base pairs. [g]The same as[f] and, in addition, the 5′-end overhanging nucleotide was removed. [h]Additional K^+ ion was placed inside the mixed tetrad during system building. [i]Histidine 454 was ε-protonated. [j]Histidine 454 was double protonated.

METHODS

Simulation Protocol. We used the xLEaP module of AMBER 12 program package[72] to prepare starting topology and coordinates of all simulations. For RNA, we always used the bsc0χ_{OL3} force field,[43,73−76] which is the default force field for RNA in AMBER since 2010 (it is internally abbreviated as ff10, ff12, or ff14 RNA force field version, depending on the code version). The proteins were described either by the ff99SB[73,74,77] (the earlier AMBER default) or by the ff12SB[72] (the AMBER default at the time of our simulations) force field. The simulated systems were solvated by SPC/E[78] octahedral water boxes with minimal distance of 10 Å from the solute border. We used KCl (K^+: radius 1.5930 Å, well depth of 0.4297 kcal/mol. Cl^-: radius 2.7110 Å, well depth of 0.0128 kcal/mol)[79] excess salt with concentration of 150 mM (measured as the final concentration of the Cl^- ions).

All systems were minimized and equilibrated using a standard equilibration protocol.[31] An additional step to minimize atomic positions of hydrogens in the solute molecules was included.

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

Table 2. Overview of Simulation Time Development Including the Number of Native Intermolecular H-Bonds, Size of Protein−RNA Interaction Surface Area and the RMSD of the Protein−RNA Complexes[a]

	h-bonds[b]			interaction surface area [Å^2]			RMSD [Å]	
simulation name	init	sim	end	init	sim[c]	end[c]	sim[c]	end[c]
complex of dsRNA and the B2 protein								
2az0_ff99	14	13.3	13.3	2486	2523	2527	1.5	1.5
2az0_ff12		13.5	13.5		2600	2555	1.7	1.7
complex of L1 protein with 23S rRNA fragment								
3u4m_ff99	15	12	12.1	2986	3325	3621	2.3	2.4
3u4m_ff12		10.7	10.6		3232	3292	2.1	2.7
3u4m_ff99_tyr1		11.3	12.7		3449	3401	2.2	2.3
3u4m_ff99_tyr2		11.5	12.3		3271	3177	2.2	2.4
3u4m_ff12_2		9.6	8.9		2847	2773	2.3	2.5
3u4m_ff12_3		9.3	9.2		3254	3232	2.5	2.6
3u4m_ff12_rst		11.0	9.9		3349	3370	2.2	2.2
3u4m_ff99_noMG[d]	12	12.6	12.1	3794	3400	2838	2.2	2.4
complex of domain of U1A protein with RNA hairpin								
1urn_ff99	14	8.8	8.3	1861	1997	2047	2.9	3.5
1urn_ff12		10.3	10.7		1997	2014	2.0	2.2
1urn_ff99_2		11.6	10.3		2087	2085	2.2	2.0
1urn_ff12_2		8.1	7.0		2184	2788	2.7	3.2
1urn_ff99_3		10.6	8.8		2065	2173	2.2	2.8
1urn_ff12_3		8.4	7.8		1957	2164	2.4	2.6
complex of SAM domain of Vts1p with RNA hairpin								
2f8k_ff99	8	4.9	1.9	933	1134	1286	2.6	3.5
2f8k_ff12		7.3	7.2		1290	1294	2.3	2.5
2f8k_ff12_GC		2.3	1.5		1245	1481	3.7	3.7
2f8k_ff99_no5_GC		7.1	7.3		1004	986	1.4	1.4
2f8k_ff99_no5_GC_2		7.0	7.4		958	1039	1.5	1.3
2f8k_ff12_no5_GC		7.2	6.7		1028	1088	1.6	1.7
2f8k_ff12_no5_GC_2		7.0	6.2		996	1152	1.7	2.0
complex of FMRP RGG peptide with RNA duplex−quadruplex junction								
2la5_ff99	4	2.0	1.9	1779	1716	1695	3.2	3.2
2la5_ff12		1.9	2.0		1500	1562	3.0	3.0
2la5_ff12_3K		2.4	2.9		1723	1760	3.2	3.3
2la5_ff12_3K_2		2.3	2.0		1677	1662	3.4	3.5
complex of Pumilio FBF-2 protein with gld-1a RNA								
3k5y_0	18	5.6	6.8	2537	1985	2023	3.5	3.6
3k5y_ff99		8.9	7.9		2198	2160	3.0	3.3
3k5y_ff99_rst		10.7	10.5		2404	2345	2.8	3.3
3k5y_ff12		7.5	4.8		2235	2019	2.9	3.2
3k5y_ff12_rst		10.9	10.0		2379	2242	2.4	2.4

[a]The values from the initial structure (init), the entire simulation (sim), and the last 50 ns (end) are shown. [b]Only the H-bonds present in the initial experimental structure (native H-bonds) were considered in this analysis. The presence of each experimental H-bond (Table 3) was ascertained at every frame of the simulation trajectory and recorded in a binary fashion (present = 1, absent = 0). The H-bond was recorded as being present (populated) when its heavy atom distance was less than 3.2 Å (see the Methods). The sum of the recorded numbers gave the total population of all native H-bond interactions at each frame (every 1 ps) of the simulation. The sum was then averaged over the respective trajectory portion (the whole trajectory or the last 50 ns) to give the averaged population of the experimentally observed H-bonds in the course of the simulation. [c]Median values were used to express the size of protein−RNA interface and the RMSD of the complex in the course of the simulation. [d]The simulation initial structure was based on the simulation end structure of the 3u4m_ff99 simulation.

Production simulation runs were performed with the pmemd module of AMBER 12[72] with CUDA implementation for calculations on graphic cards.[80,81] The particle mesh Ewald (PME)[82,83] was used for calculations of the electrostatic interactions. The cutoff distance of the nonbonded Lennard-Jones interactions was 9 Å. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[84] Periodic boundary conditions, a 2 fs integration step, and 300 K temperature (Berendsen weak-coupling)[85] were used. The size of the simulation water boxes was selected to have at least 10 Å distance between the solute and box borders. This setup prevented any direct interactions between the solute images, which we also monitored along the simulation trajectories. Analyses of the resulting trajectories were performed using ptraj and cpptraj modules[86] of AMBER. VMD[87] and PyMOL[88] programs were used for visualization. Gnuplot and Raster3D[89] were used to produce graphs and figures, respectively.

The specific recognition was assessed by monitoring behavior of the interface H-bonds. We analyzed all H-bonds present in the experimental structure as well as any new H-bonds formed during the simulations. In our figures, we indicate a heavy atom distance of 3.2 Å as an approximate H-bond threshold, though

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

Table 3. List of Experimental Protein−RNA Interface H-bond Interactions[a]

H-bond	distance
Pumilio FBF-2 protein with gld-1 FBEa RNA (3k5y)	
U1(O4)-Gln504(NE2)	2.8
U1(N3)-Asn500(OD1)	2.9
U1(O2)-Asn500(ND2)	2.9
U1(O4)-Lys557(NZ)	3.1
G2(N1)-Glu457(OE2)	2.6
G2(N2)-Ser453(OG)	2.9
U3(O4)-Gln419(NE2)	2.7
U3(N3)-Asn415(OD1)	2.9
U3(O2)-Asn415(ND2)	2.9
G4(N2)-Cys363(SG)	3.3
C5(N3)-Arg364(NH2)	3.0
C5(O2)-Arg364(NE)	2.9
A7(N1)-Gln291(NE2)	2.9
A7(N6)-Gln291(OE1)	3.0
U8(O2)-Asn244(ND2)	3.0
U8(N3)-Asn244(OD1)	2.8
U8(O4)-Gln248(NE2)	3.1
A9(O2′)-Lys201(NZ)	2.7
FMRP RGG peptide with an RNA duplex−quadruplex junction (2la5)	
Arg10(NH2)-G31(N7)	3.0
Arg10(NE)-G31(O6)	2.8
Arg15(NH2)-G7(N7)	3.0
Arg15(NE)-G7(O6)	3.0
Domain of U1A protein with RNA hairpin (1urn)	
G9(O6)-Asn16(N)	3.1
C10(N3)-Lys88(N)	3.1
C12(N3)-Asp92(N)	3.0
C12(N4)-Asp90(O)	2.9
G16(OP)-Leu49(N)	2.9
G9(O2′)-Lys50(O)	2.6
A11(N1)-Ser91(OG)	2.9
C10(N4)-Tyr86(O)	3.1
G9(N7)-Asn15(ND2)	3.0
U7(N3)-Glu19(OE)	2.8
U8(N3)-Asn16(OD1)	2.9
G16(N7)-Arg52(NH2)	3.0
A6(N1)-Arg52(NH1)	2.7
U8(O4)-Lys80(NZ)	2.9
L1 protein with 23S rRNA fragment (3u4m)	
G2123(N2)-Glu42(OE1)	3.1
G2124(O2′)-Glu42(OE2)	2.6
G2124(O2′)-Thr217(OG1)	2.8
G2124(N2)-Thr217(O)	3.1
L1 protein with 23S rRNA fragment (3u4m)	
C2174(O2′)-Met218(SD)	3.2
C2175(O2′)-Gly219(O)	2.7
G2121(N2)-Asp166(OD2)	2.9
A2176(OP1)-Tyr7(OH)	2.5
G2125(OP1)-Thr40(OG1)	2.7
C2128(OP1)-Lys36(N)	2.7
U2130(OP1)-Lys5(N)	2.8
G2131(OP1)-Lys5(NZ)	2.7
C2175(OP1)-His3(ND1)	2.6
C2177(OP1)-Ser221(OG)	2.6
C2178(OP1)-Ser211(OG)	2.7
SAM domain of Vts1p with RNA hairpin (2f8k)	
Lys461(NZ)-U(−3)(OP1)	3.0
His466(NE2)-C(−2)(OP2)	2.9
Lys467(NZ)-U(−1)(OP1)	2.8
Lys467(NZ)-U1(OP2)	3.3
Leu496-U2(OP2)	3.1
Gly497(N)-U2(OP2)	3.1
G3(N2)-Arg464(O)	3.2
Arg500(NH2)-G3(OP2)	3.0
DsRNA and the B2 protein (2az0)[b]	
Arg54(NE)-G6(OP1)	2.9
Lys47(NZ)-C8(OP2)	3.3
Asn40(ND2)-G5(N3)[c]	3.0
G5(O2′)-Asn40(OD1)[c]	2.7
Lys62(NZ)-U16(O2′)	2.7
Arg36(NH1)-A15(OP1)	2.7
Asn33(ND2)-U16(OP1)	2.6
Arg54*(NH2)-G6(OP1)	3.1
Lys47*(NZ)-C8(OP2)	3.2
Asn40*(ND2)-G5(N3)[c]	3.1
G5(O2′)-Asn40*(OD1)[c]	2.7
Lys62*(NZ)-A15(O2′)	2.9
Arg36*(NH1)-U16(OP1)	2.7
Asn33*(ND2)-G17(OP1)	2.8

[a]Heavy atoms donor/acceptor distances, in Ångstrom. [b]The asterisk refers to the second chain of the dimeric B2 protein. [c]We assume, based on the simulation behavior and the NQ-flipper[92] report, that both Asn40 side-chain rotamers may be incorrectly positioned in the X-ray structure. Therefore, the distance values in the Table refer to the presumably correct rotamer state obtained by flipping the side-chain position by 180°.

any H-bond threshold definition is necessarily arbitrary. We also monitor directionality of all interactions, although we do not report this data for space reasons. We have further monitored the basic structural dynamics of the RNA molecules (including the backbone dihedral angle behavior) and of the proteins using standard tools. We have also analyzed cation interactions and all trajectories were extensively visually monitored to spot any significant developments in the simulation. Note that due to the extent of simulations and size of the systems we report these results only when particularly relevant.

We used the LCPO method[90] to analyze the interaction surface area of the protein−RNA complexes. We used an in-house implementation of the adaptive biasing force (ABF) method[91] to calculate free energy profiles of several dihedrals. For the NMR structure, we also computed violations of experimental NOE distances in our simulations using a $(r^{-6})^{(-1/6)}$ weighted average.

H-Bond Definition and Nomenclature. The side chains of the glutamate and aspartate can rotate during the simulation, switching the H-bond acceptor between the two oxygen atoms. Similarly, arginine side chain can rotate to switch the H-bond donor between its two NH_2 groups or the NE atom. For RNA, the two phosphate nonbridging oxygen atoms can alternate as H-bond acceptors. Therefore, in the case of these residues, we measured distance between all possible H-bond forming atom couples. For the sake of clarity, our graphs always show only the shortest distance and thus do not reveal the fluctuational switches involving the above-noted atoms. However, their flips were monitored by us. Thus, in the graphs, the atoms Glu(OE),

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

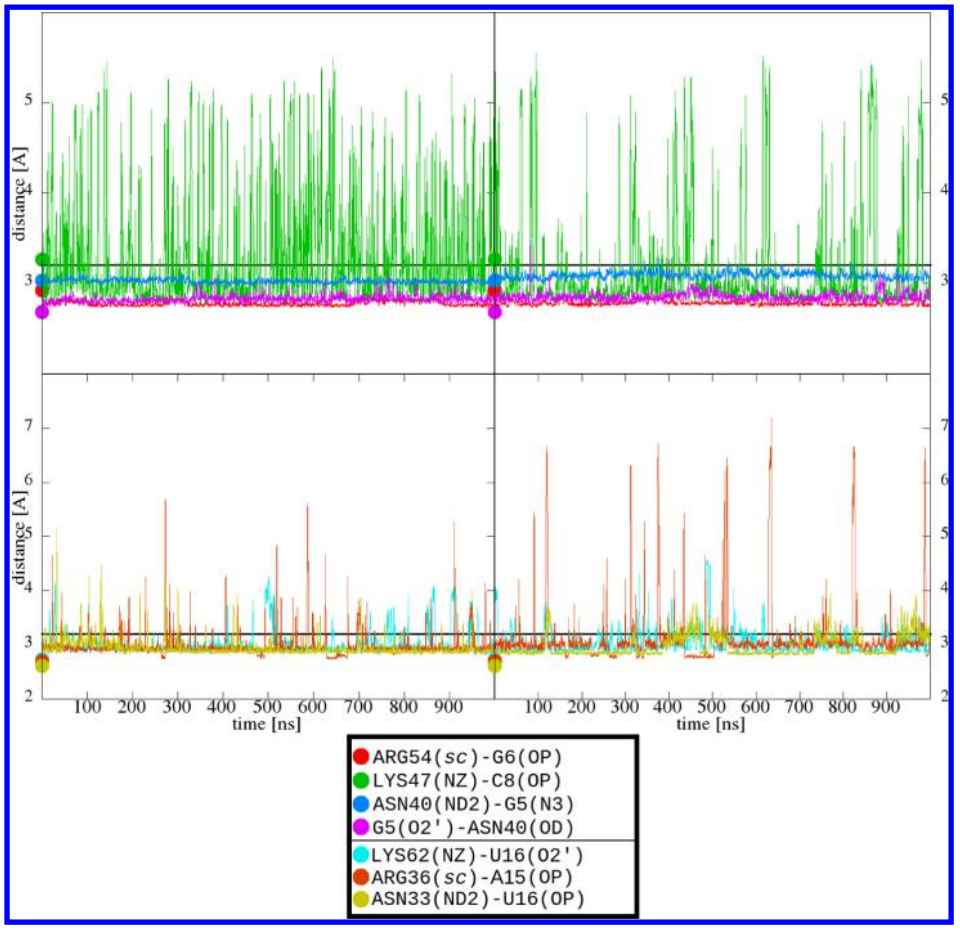


Figure 2. Time development of H-bond heavy atom distances (see the legend) in 2az0_ff99 (left) and 2az0_ff12 (right) simulations. The colored points indicate the initial values. The black horizontal line marks 3.2 Å. The symmetrical interactions are shown in Supporting Information Figure S8. We assume that the Asn40 is incorrectly oriented in the initial X-ray structure and our reference colored point is referring to distances after flipping the side-chain end segment by 180°.

Asp(OD), and Nucleotide(OP) stand for interaction involving any of the two relevant oxygen atoms. Arginine is marked as Arg(*sc*), where *sc* stands for the *side-chain* atoms NH1, NH2, or NE (Supporting Information Figure S1).

Details of the selection of the studied systems and details of the initial systems building (including the protonation states) are discussed in the Supporting Information.

RESULTS

We conducted MD simulations of six protein−RNA complexes (Table 1 and Figure 1). Unless stated otherwise, the starting structures included all experimentally observed biomolecules, ions, and water molecules. We did not include ligands and other small molecules that were used for preparation or crystallization of the experimental systems (e.g., glycerol). Table 2 summarizes the main analyses—the development of number of native H-bonds (specific recognition), size of the protein−RNA interface surface area (shape recognition), and root-mean-square deviation (RMSD) (internal stability) in our simulations. In addition, graphs of time development of all these individual parameters are present either in the main text (individual H-bonds) or in the Supporting Information (complete RMSD data, complete interface surface area data, and the rest of H-bonds analyses). Complete list of all experimental intermolecular H-bonds is in Table 3. A graphic overview of these interactions in the context of the entire protein−RNA interfaces is in Supporting Information Figures S2−S7. Statistical analysis of number of intermolecular atomic contacts is in Supporting Information Table S1. Note that RMSD is a very ambiguous descriptor of the simulation behavior, and thus, the main emphasis in our analyses is placed on monitoring specific H-bonds.

Short Summary of the Results. In the following, we present detailed description of extended simulations of six protein−RNA complexes. As the individual systems are very different and also their simulation behavior is highly variable, we were unable to find a unified way of presentation of the studied systems, and thus, we present the results separately for each complex. The systems are roughly ordered from the most stable to the least stable complex. For all systems, we always first discuss two initial 1-μs simulations, one with the ff99SB and the other with the ff12SB protein force field. Then, based on the initial results, we perform additional simulations. The results can be briefly summarized as follows.

DsRNA−B2 protein complex simulations are entirely stable, albeit sampling non-native (or unobserved) proximal contacts as fluctuations to certain extent.

In the case of the L1 protein−23S rRNA fragment complex, the simulations provide stable results for its region with shape-specific recognition. We obtain reasonable results also for the part of the complex that relies on sequence-specific H-bonds, albeit the simulations are not capable to simultaneously keep all

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

observed H-bonds. There is some difference between the ff99SB and ff12SB protein force fields.

For the U1A protein−RNA hairpin complex, no simulation is able to simultaneously keep all observed H-bonds, with no difference between the two protein force field variants. In all simulations, some H-bonds are lost at the beginning of the simulations. The simulation then stabilizes with this perturbed H-bonded network with no further deterioration. Several different specific simulation scenarios are observed for this complex.

For the Vts1p SAM domain−RNA hairpin complex, some simulations remain in essentially complete agreement with the experimental structure while others lead to complete loss of the protein−RNA interface. The results do not depend on the protein force field version and the probability of getting stable trajectories can be greatly increased by modifying the RNA part of the molecule, which is not interacting with the protein.

For the FMRP RGG peptide−RNA duplex−quadruplex junction, we identify critically important monovalent ion binding site in the area of the duplex−quadruplex junction that has not been found in the original NMR study. All simulations are initially unstable, but when manually intervening into the starting structure and placing the predicted ion into the right position, we could get a trajectory with essentially a full repair of the experimental structure on the microsecond time scale.

No stable trajectories could be obtained for the Pumilio FBF-2 protein/gld-1 FBEa RNA complex, for which all our simulations show progressive loss of its protein−RNA interface.

The subsequent paragraphs contain a detailed description of the individual systems followed by discussion explaining the likely reasons of the very variable simulation outcomes revealed in our study.

Complex of dsRNA and the B2 Protein Was Stable in Simulations. This protein−RNA complex has a relatively large interaction interface area (Table 2) that is complemented by 14 H-bonds (Supporting Information Figure S2). All these H-bonds involve the sugar−phosphate backbone, suggesting a sequence-independent binding. The dimeric nature of the complex doubles the number of interactions. Both chains of the protein are interacting via identical amino acid residues with identical or nearby dsRNA atoms.

The simulated system was stable with both protein force fields. The interaction surface area was essentially unchanged (Table 2) and the protein−RNA H-bonds were preserved (Figure 2) albeit the interaction of Lys47(NZ) with the phosphate group of C8 was reversibly fluctuating. These fluctuations, however, typically brought the lysine side chain into similar H-bond interaction with phosphate groups of neighboring upstream or downstream nucleotides.

We also noted flipping of both Asn40 side chains immediately after the simulation start (Figure 3). We assume that orientation of these side chains in the original X-ray structure may be incorrect as the new interaction pattern is more consistent. Incorrectness of the initial structures is also strongly suggested by the NQ-Flipper program.[92] Interestingly, the flips would replace the originally proposed nonspecific interaction by a sequence-specific interaction. Since the authors of the experimental structure explicitly discussed the Asn40−RNA interaction as important,[48] we have decided not to flip the Asn40 side chains beforehand in our simulations. Because the Asn40 side chains flipped in few picoseconds and the system was otherwise very stable, it did not introduce any error into the simulations. We did not observe any difference between the ff99SB and ff12SB simulations. The RNA part of the system was fully stable in both simulations, including the terminal base pairs. Also, the protein part was stable, except of genuine fluctuations of the short unpaired N- and C-ends which had no impact on system stability.

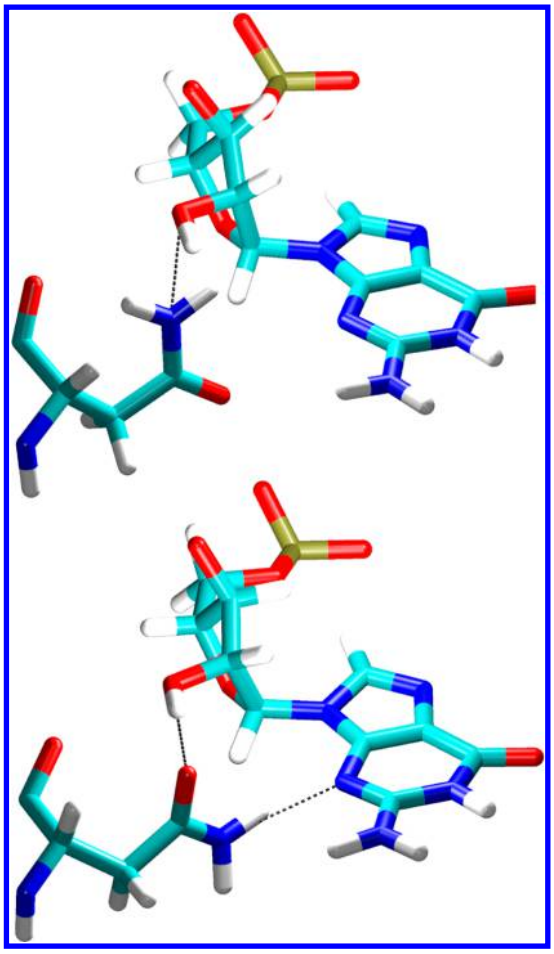


Figure 3. Top: the G5(O2′)-Asn40(ND2) interaction suggested by the X-ray structure. Bottom: the side chain of Asn40 immediately flipped in simulations to form the G5(O2′)-Asn40(OD1) and sequence-dependent Asn40(ND2)-G5(N3) interactions instead.

Complex of L1 Protein with 23S rRNA Fragment (3u4m) Reveals Some Differences between the ff99SB and ff12SB Protein Force Fields and Effect of the Starting Structure. In the starting structure, we used all experimentally observed ions and waters which in this particular case are structurally important. The C2111 was N3-protonated to satisfy the RNA tertiary interactions.[31]

Shape-Recognized Part of the Complex Is Stable. A portion of the L1 protein−23S rRNA fragment interaction interface involves nonspecific interactions between the protein and RNA without any sequence-specific H-bonding (see the Introduction and Supporting Information Figure S3). These protein−RNA contacts mostly involve nonspecific van der Waals overlaps of amino-acid carbon atoms and the RNA sugar rings. This interface was stable in our simulations except of occasional short fluctuations caused by protein or RNA dynamics (not shown).

Specific Interactions at the Protein−RNA Interface. MD description of the specific interactions was rather satisfactory, although we noticed some reproducible differences between the

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

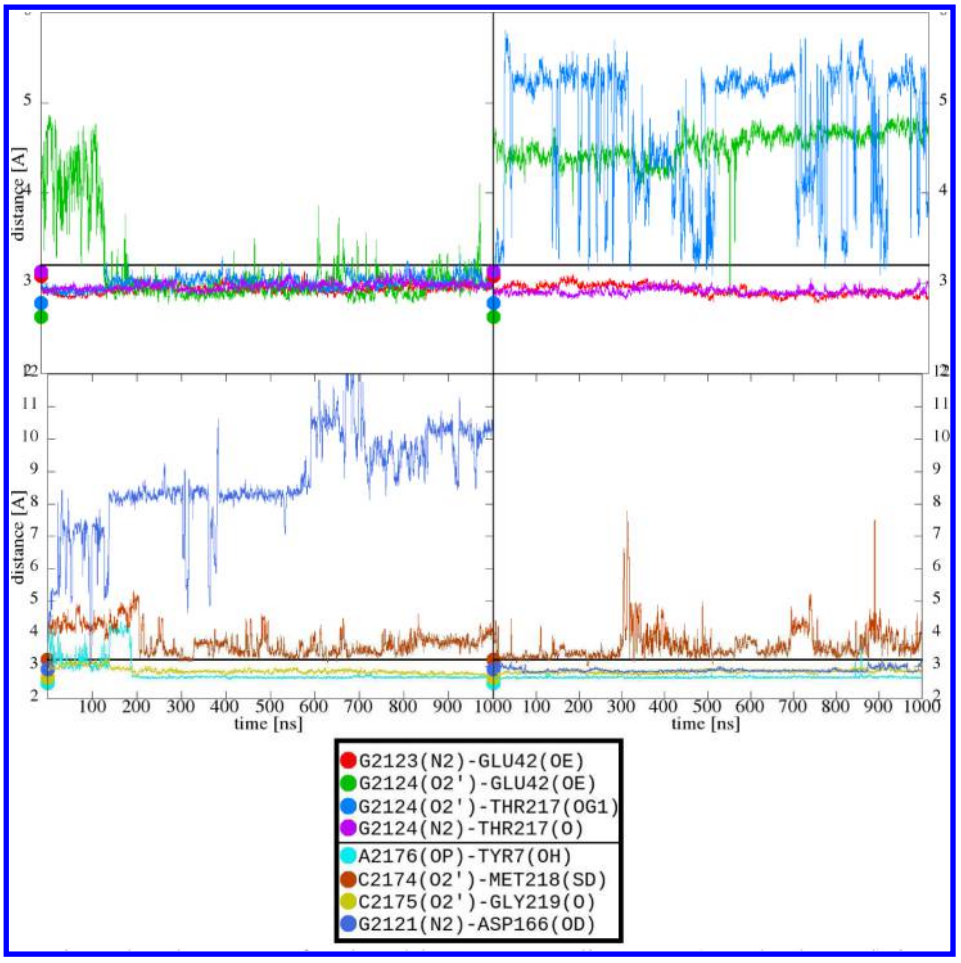


Figure 4. Time development of H-bond heavy atom distances (see the legend) in 3u4m_ff99 (left) and 3u4m_ff12 (right) simulations. The colored points indicate the initial values. The black horizontal line marks 3.2 Å.

ff99SB and ff12SB force fields and effect of the starting structure (Figure 4).

In the 3u4m_ff99 simulation, we observed immediate and permanent loss of the G2121(N2)-Asp166(OD) interaction. This was observed in all subsequent ff99SB simulations (Supporting Information Figures S9 and S10) and was likely caused by the protein force field subtly shifting the entire protein loop containing the Asp166 residue with respect to the RNA position (Supporting Information Figure S11). This interaction was stable with ff12SB, however, in the 3u4m_ff12 simulation, the G2124(O2′)-Thr217(OG1) and G2124(O2′)-Glu42(OE) interactions were lost instead.

To confirm that the difference was caused by the force field variant and not by random effects of sampling, we conducted two additional 100 ns simulations using the ff12SB force field (Table 1). Both simulations confirmed the main 3u4m_ff12 simulation (Supporting Information Figures S12 and S13). The G2124(O2′)-Glu42(OE) and G2124(O2′)-Thr217(OG1) interactions were lost immediately after the start and on a scale of tens of nanoseconds, respectively. It should be noted that the G2124(O2′)-Glu42(OE) interaction was initially unstable also in the ff99SB simulations, but with this force field, it was always later restored. We suggest that the behavior was caused by several factors. First, there could be an MM (molecular mechanics) energy strain in the starting structure around the G2124(O2′)−Glu42(OE) interaction, as evidenced by its initial instability in all simulations, regardless of the used force field. The interaction was later restored with ff99SB but not with ff12SB, which could be related to quite different dihedral profiles of glutamate side-chain in ff99SB and ff12SB (Supporting Information Figure S14).

The destabilization of the G2124(O2′)−Thr217(OG1) interaction could be a secondary effect caused by the previous loss of the G2124(O2′)−Glu42(OE) interaction, which resulted in fluctuations of the G2124(O2′) hydroxyl group orientation. Nevertheless, we observed loss of the G2124-(O2′)−Thr217(OG1) interaction even in a subsequent ff12SB simulation where the G2124(O2′)−GLU42(OE) interaction was restrained (see Table 1 and Supporting Information Figure S15). We found out that, in addition to the protein−RNA G2124(O2′)−Glu42(OE) interaction instability, there was a subtle ff12SB-specific shift of one of the protein loops (Supporting Information Figure S16). Therefore, behavior of this region was different in ff99SB and ff12SB protein force fields. In addition, even when initially restraining the G2124-(O2′)−Glu42(OE) interaction with ff12SB (during the first 100 ns, see Table 1), the interaction was lost after the restraint was lifted, further supporting the view that there is some real difference between the ff99SB and ff12SB force field descriptions of this region.

Another difference between the two protein force fields was detected for the A2176(OP)−Tyr7(OH) interaction, albeit it was only temporary and was partially related to the starting structure. The stability of the A2176(OP)−Tyr7(OH)

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

interaction depended on correct initial position of the hydroxyl group of Tyr7. The default assignment of the xleap module of AMBER gave an initially incorrect hydroxyl orientation. With ff99SB, the hydroxyl could not rotate into the correct position for about 200 ns due to a quite high torsional barrier for the hydroxyl rotation. The ff12SB force field reduced this energetic barrier and the Tyr7 hydroxyl reoriented on scale of few ns, so that the interaction was stable since the beginning of the simulation (Figure 4). To further examine this issue, we conducted two additional simulations using the ff99SB force field where the Tyr7 hydroxyl group was initially either in the default xleap assigned (incorrect) or in manually adjusted (correct) position (see Table 1). The first simulation again temporarily lost the A2176(OP)−Tyr7(OH) interaction, while it was completely stable in the second simulation. Thus, we suggest that the observed different initial behavior of the A2176(OP)−Tyr7(OH) interaction was related to the default incorrect initial orientation of the Tyr7 hydroxyl group coupled with a too high barrier for its rotation in the ff99SB force field (Supporting Information Figures S9 and S10). Note that although the A2176(OP)−Tyr7(OH) interaction was ultimately repaired with the ff99SB force field, the 200 ns time scale of the repair was longer than the length of most protein−RNA simulations published until now. Coupled to the A2176(OP)−Tyr7(OH) interaction was behavior of the C2174(O2′)−Met218(SD) interaction. It was initially unstable in ff99SB simulations before the Tyr7 hydroxyl rotated into the correct position. The larger heavy-atom distance of this interaction is due to the sulfur atom size.[93]

The system also contains several RNA phosphates/protein interactions, which are analyzed in the Supporting Information.

Dynamics of the L1 Protein. The simulations captured reversible interdomain motions of the two domains of the L1 protein that may contribute to the biochemically relevant L1 stalk dynamics.[31,52] The interdomain movement was particularly noticeable in two simulations of the isolated L1 protein (see Table 1). Both visual analysis and Interactive Essential Dynamics (IED) method[94] showed that the primary movement was opening and rotation of the protein domains with respect to each other (Supporting Information Figure S18).

Secondary structures of the individual segments of the L1 protein were stable with exception of the helix α5 (residues 117−133) of the domain II. It was reversibly disrupted in the middle (residues 123−125) and formed a short β sheet-like structure that was flanked by remains of the helix on both sides (Figure 5). This was accomplished by a temporary shift of φ/ψ dihedrals of residues 123−125 from values typical for α-helices to values typical for β-sheets. The fluctuations occurred on ∼100 ns time scale with both force fields. The fluctuations were localized and did not have any impact on the other parts of the system. In our earlier shorter simulations of the L1 stalk,[31] which included a bound tRNA molecule, we noted the same dynamics. In those simulations, the changes led to formation of additional contacts between this segment of the L1 protein and the bound tRNA molecule. Thus, the folding−unfolding dynamics of the α5 helix may contribute to the tRNA binding.

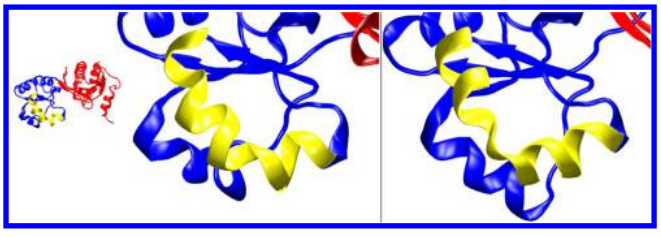


Figure 5. Changes in the secondary structure of the α5 helix of the L1 protein (left, starting structure; right, 700 ns of the 3u4m_ff99 simulation). Domain I is colored red, domain II is blue, and the α5 helix is yellow.

Dynamics of the 23S rRNA Fragment. The rRNA molecule of the L1 stalk fragment was quite stable. We observed some specific local dynamics, entirely consistent with earlier shorter simulations of the isolated RNA segment.[31] The RNA intrinsic dynamics did not influence the native interactions with the L1 protein. In simulations, the RNA helix end was often reversibly interacting with the protein. This explains the at-first-sight counterintuitive increase of the overall interaction surface area of this complex (Table 2).

Structural Ions at the Protein−RNA Interface. The simulations confirm importance of two experimentally observed buried Na^+ ions (replaced by K^+ in the simulations) that are trapped between the protein and RNA and form part of the protein−RNA interface. They lack any interactions with the solvent and their coordination shells are composed entirely of the protein and RNA atom groups. These coordination shells are fully maintained in all our simulations, form an integral part of this molecular complex, and extend the specific binding between the helix 77 and the protein. This indicates that such buried structural monovalent ions can be well described by MD simulations. In the Supporting Information we also provide details of the Mg^{2+} dynamics.

Complex of the U1A Protein with the RNA Hairpin (1urn) Reveals Diverse Dynamics in Different Simulations. The X-ray structure contains three protein−RNA complexes in the asymmetric unit, but we simulated only one of them (chains B and Q) as all three structures have similar protein−RNA interface H-bond network. The 3′-end overhanging nucleotide was removed.

Stability of the Protein−RNA Interface. The U1A complex contains a dense network of H-bond interactions (Supporting Information Figure S4) while having a relatively small interaction surface area. Shape recognition is therefore less pronounced in comparison to the two preceding systems and local H-bond instabilities are having a larger overall effect. The recognition is strikingly different from, for example, the dsRNA−B2 protein complex.

In the initial two simulations, the ff12SB force field performed better than the ff99SB. Specifically, the G9(O6)−Asn16(N), G16(OP)−Leu49(N), U8(N3)−Asn16(OD1), and U8(O4)−Lys80(NZ) interactions were better maintained with ff12SB as their disruptions were reversible (Figure 6). They were not stable with ff99SB. On the other hand, the ff12SB did not maintain the A11(N1)−Ser91(OG) interaction while ff99SB did so. The U7(N3)−Glu19(OE) interaction was disrupted shortly after the start with both force fields.

Additional Simulations Reveal a Sampling Issue. To verify the results, we conducted four additional simulations (see Table 1) that ultimately revealed that the differences between the first two simulations were not systematic and were rather related to sampling (Supporting Information Figure S22 and S23). For example, the third ff12SB simulation showed very similar behavior as the first ff99SB simulation. In other words, the simulations did not capture any really discernible differences between the ff99SB and ff12SB protein force fields. The simulation behavior was rather random, and none of the

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

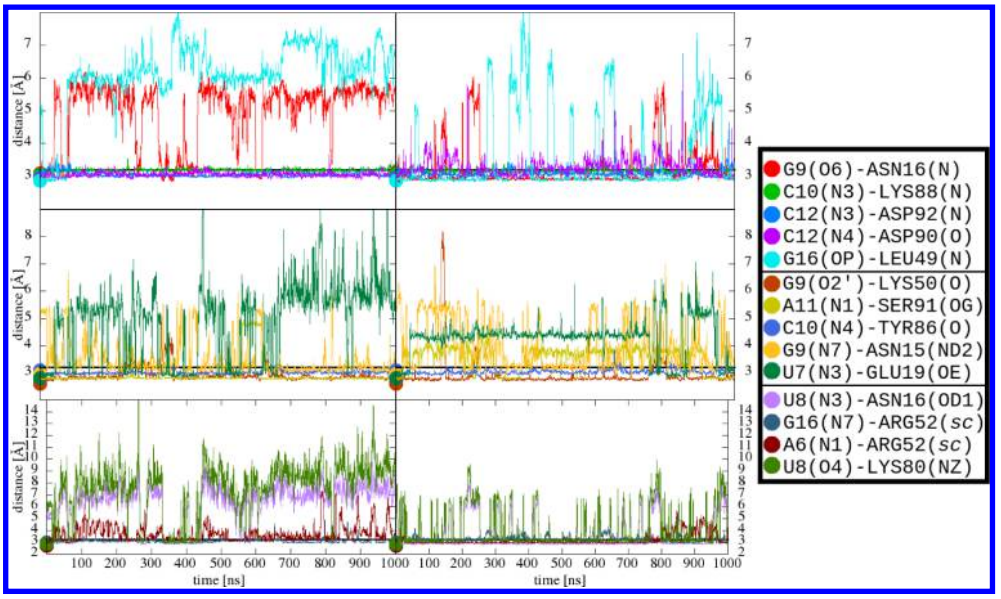


Figure 6. Time development of H-bond heavy atom distances (see the legend) in 1urn_ff99 (left) and 1urn_ff12 (right) simulations. The colored points indicate the initial values. The black horizontal line marks 3.2 Å.

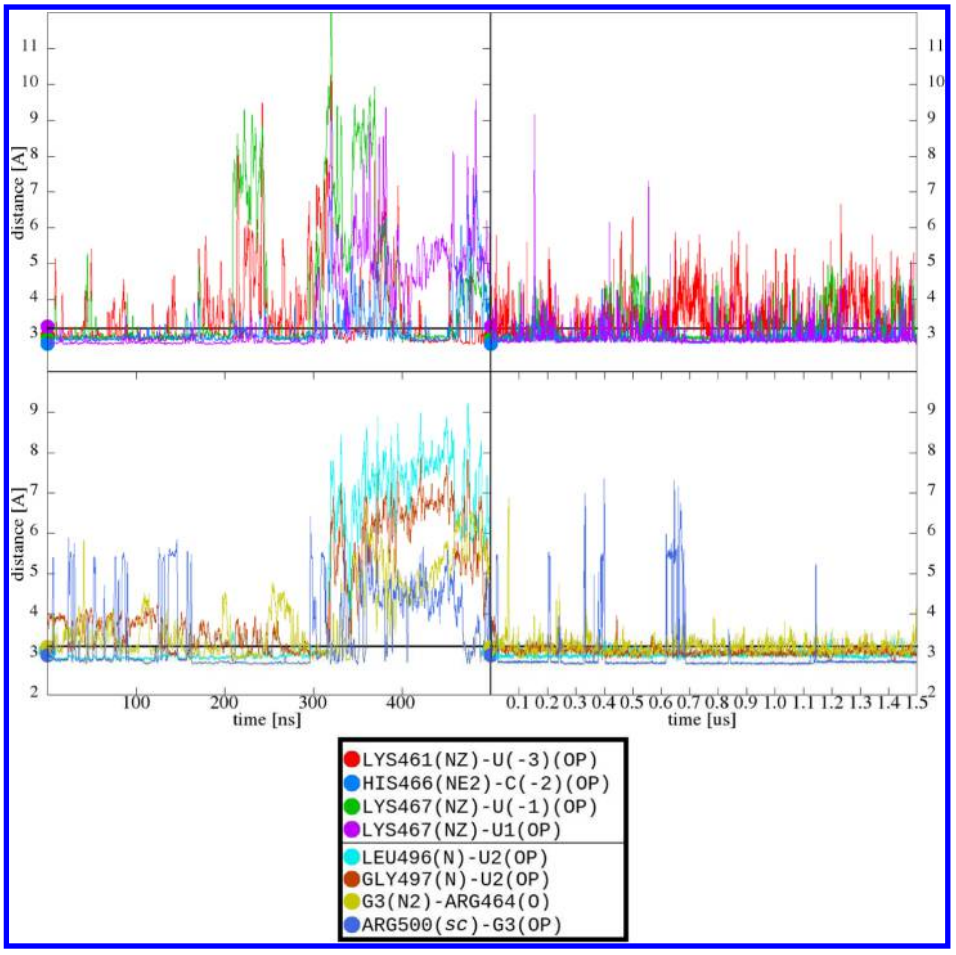


Figure 7. Time development of H-bond heavy atom distances (see the legend) in 2f8k_ff99 (left) and 2f8k_ff12 (right) simulations. The colored points indicate the initial values. The black horizontal line marks 3.2 Å.

simulations was able to fully reproduce the crystallographically observed protein−RNA interface although the simulations did not progressively deviate from the starting structure. We made the following observations (cf. all the figures): The simulation behavior was predetermined by the early stages of the individual simulation runs. The simulations were not capable to simultaneously keep all the observed H-bonds, due to some strain caused either by the force field or by some hidden

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

problem of the starting structure that we did not identify. After initially losing some of the H-bonds, the simulated structures stabilized and did not show any signs of further significant developments. The relatively best behaving simulation was the second ff99SB simulation (Table 2, Supporting Information Figure S22 left) with only three protein−RNA H-bond interactions substantially fluctuating.

In summary, none of our simulations was capable to keep stable all H-bonds. Many of the individual H-bonds showed large fluctuations, although in most cases they were capable to temporarily return back to the starting values. However, the simulations had evident difficulty to simultaneously preserve all the native interactions. This may be a hallmark of problems in simulations of protein−RNA complexes possessing dense networks of H-bonds that are mutually affecting each other.

Dynamics of the Protein and the RNA Hairpin. The secondary structure of the U1A protein was stable in all simulations, with no changes in its overall fold. Similarly, the RNA hairpin structure was quite stable. We noticed short fraying events of the terminal base pair in some simulations with no observable effect on the rest of the structure. The specific shape of the hairpin part of the RNA was visually reasonably maintained in the simulations. Developments at the protein−RNA interface can also indirectly affect the internal structure of the RNA. Consequently, those simulations better preserving the protein−RNA interface had lower overall RMSD of the RNA part of the system (Supporting Information, RMSD graphs). The instabilities at the protein−RNA interface were mainly associated with structural changes of the individual amino acid side-chains.

Behavior of Ions in the Simulations. We observed interference of the K^+ ions with several protein−RNA interactions. The ions were repeatedly entering the area between the G9 nucleotide and the protein, interfering with the G9(N7)−Asn15(ND2), U8(N3)−Asn16(OD1), and U8(O4)−Lys80(NZ) interactions (Supporting Information Figure S24). Whether the ions actively contributed to the destabilization of the protein−RNA interface or just opportunistically utilized the created space could not be confidently clarified.

Complex of SAM Domain of Vts1p with RNA Hairpin (2f8k) Reveals Clear Sampling Issue. *Stability of the Protein−RNA Interface in the Initial Simulations.* As the C5 nucleotide was not visible in the X-ray structure due to disorder, we have modeled it. All protein−RNA H-bond interactions except one occurred between amino acid side chains and phosphate groups. In our initial ff12SB simulation, all interactions were essentially stable, with occasional local and reversible disruptions (Figure 7). However, in our initial ff99SB simulation, almost all original protein−RNA interactions were permanently lost since ∼320 ns. This was also coupled with changes of the RNA hairpin topology and changes in its relative position to the protein, leading to a complete loss of the experimental structure (Supporting Information Figure S25).

The ff99SB simulation revealed some more subtle fluctuations already in the first part of the trajectory, giving an impression that the simulation acquired some structural destabilization at the beginning, which propagated further as the simulation progressed. We have noticed that the destabilization may be related to the RNA helix end fraying (opening of the RNA helix terminal base pairs).[45] There are three AU base pairs at the helix end making this structure particularly prone to end fraying. Furthermore, there is unpaired overhanging nucleotide at the 5′-end of the RNA, which is extensively fluctuating in the simulations (Supporting Information Figure S5). Perturbations of RNA helix ends occurred in both above simulations but were structurally different (Supporting Information Figure S26). Since the RNA force field was identical in both simulations, the different RNA end-effects should reflect sampling difference. It is well established that a microsecond time scale is not sufficient to obtain converged sampling of end fraying of canonical helices.[45,95] The nucleotides directly affected by the end fraying did not participate in the protein−RNA interface. However, they could destabilize it by increasing the fluctuations of the adjacent part of the RNA hairpin. The terminal or frayed nucleotides could also form new interactions with the protein and thus introduce unfavorable changes in the protein structure. Some such interactions indeed occurred in the ff99SB simulation and led to increased surface area of the protein−RNA interface (Table 2 and Supporting Information).

Additional Simulations. We conducted another ff12SB simulation (2f8k_ff12_GC, see Table 1) where we replaced the two terminal AU base pairs with more stable GC base pairs. In this simulation, the end-effects of the RNA helix were eliminated. Nevertheless, the disruption of the protein−RNA complex occurred after 120 ns, now with the ff12SB force field and in absence of the end fraying (Supporting Information Figure S27). Thus, the loss of the protein−RNA interface in the initial ff99SB simulation was caused neither by the ff99SB force field nor by the RNA helix end fraying. Yet another possibility was that the unpaired overhanging 5′-end nucleotide was disrupting the protein−RNA interface. Indeed, in the 2f8k_ff12_GC simulation, the terminal nucleotide contacted the protein during the first 120 ns, and after the loss of the protein−RNA interface, it established extensive interactions with the protein. Still, the contacts occurring during the first part of the trajectory portion did not seem to be extensive enough to be the primary cause of the extensive structural deformation. Instead, the newly formed interactions could be its consequence.

We have then carried out 1-μs ff99SB simulation with the GC base pairs and with removal of the overhanging nucleotide (see Table 1, the 2f8k_ff99_no5_GC simulation). This complex was entirely stable (Supporting Information Figure S28). This simulation finally proved that the above-described results were not reflecting a protein force field difference, because both force fields could result in stable simulations. To further improve sampling, we did one ff99SB and two ff12SB 1-μs simulations of the same modified system. In all cases, the complex was stable (Supporting Information Figure S28−S31).

Evaluation of the 2f8k simulations was not straightforward. While in some simulations the protein RNA interface was lost, other simulations were nicely stable. The results almost certainly did not reflect differences between the ff99SB and ff12SB force fields, as we obtained stable and unstable simulations with both force fields.

Some data indicated that the instabilities could be related to the end fraying and interactions between the flanking 5′ terminal nucleotide and the protein. When the stems were stabilized by the AU → GC substitutions together with removal of the 5′-end overhang, all simulations were stable. Of course, we cannot guarantee a full convergence of this particular result. However, even this did not explain all the results, since we had also one 1.5-μs simulation of the unmodified system (2f8k_ff12) that was perfectly stable. Perhaps, stability of this

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

simulation was a consequence of a combination of a specific pattern of end fraying that allowed fixation of the overhanging nucleotide by a specific interaction with the protein, which did not perturb the complex (Supporting Information Figure S32). Thus, the outcome of the individual simulations may reflect a complex combination of random end-fraying effects of the RNA hairpin helix, random dynamics of the overhanging 5′-end nucleotide, and their interactions with the protein. This issue cannot be fully resolved using microsecond-scale simulations.

However, the 2f8k simulations may reflect yet another and more general problem related to sampling. One of the likely problems in protein−RNA simulations is the fact that the initial experimental structures are necessarily locally unrelaxed. Even when the structure is correctly refined from the experimental point of view, the force field can still face, at the beginning of the simulation, locally unrelaxed molecular contacts, backbone geometries, etc. We reiterate that we do not suggest any errors in the experimental structure, however, even correct experimental structure is not sensed as relaxed by the force field and is always high in force field energy. In addition, high energy hot-spots in the initial structure can also be created by crystal packing and its loss in the simulation. In other words, the starting experimental structure never corresponds to the potential energy minimum of the simulation force field and the molecule must first somehow relax to lead to a stable simulation. Such relaxation is the basic goal of the presimulation equilibration phase. However, we hypothesize that for complex protein−RNA systems the standard equilibration protocols may not be robust enough and the system enters the production phase not fully relaxed. Then, in some simulations, due to their stochastic nature, the initial high energy of the system can result in perturbation at the beginning of the production simulations. The initial structural perturbations may sometimes be barely noticeable and may result in visible geometrical change later on in the simulation once the system crosses the energy barrier (cf. Figure 7, left). Such simulations then appear as unstable. In some other simulations, the structure can shake off the initial stress without being perturbed and then can continue stably.

Dynamics of the Protein and RNA. The secondary structure of the SAM domain of Vts1p protein was fully stable in all simulations. Salient feature of the initial 2f8k_ff99 and 2f8k_ff12 simulations was the gradual disintegration of the three terminal AU base pairs and complex dynamics of the 5′-end unpaired overhanging nucleotide U(−6), which is bulged into the solvent in the initial structure. In both simulations, the paring of the first two AU base pairs was permanently lost after tens to hundreds of nanoseconds (Supporting Information Figure S26), but the bases adopted different orientations. In the 2f8k_ff99 simulation, the bases from the disrupted base pairs interacted with the protein, while in the 2f8k_ff12 simulation they formed a base triad at the helix end and did not interact with the protein. We suspected that the instability of protein−RNA interactions in the first ff99SB simulation was at least partially related to the disruptions introduced by the nucleotides affected by the end-fraying, as explained above. In the other simulations, the end-fraying was suppressed by using the GC base pairs.

Complex of the FMRP RGG Peptide and the RNA Duplex−Quadruplex Junction (2la5) Can Be Stabilized Only after Manual Placement of a Structural Ion. We used the first frame from the NMR ensemble as the initial structure. As our simulations assume neutral pH, both ends of the peptide chain were charged accordingly. All frames in the NMR ensemble are quite similar and the differences are seen in the loops of the quadruplex part of the molecule, far away from the protein−RNA interface (Supporting Information Figure S33). The NMR restraints were not utilized in our production simulations. Nevertheless, we computed average violations of the experimental intermolecular NOE distances as an additional way to ascertain the stability of this protein−RNA complex in our simulations (Table 4).

Table 4. List of Violations of Intermolecular NOE Distances in Simulations of the 2la5 Structure[a]

	whole simulation		last 50 ns	
simulation name	No. of violations	median value of violation	No. of violations	median value of violations
2la5_ff99	22	1.1	37	1.5
2la5_ff12	49	1.8	57	2.5
2la5_ff12_3K	15	1.1	14	0.8
2la5_ff12_3K_2	25	0.7	35	2.6

[a]The number of violations (out of 88 intermolecular NOEs) and the median of these violations (in Å) are indicated for the entire simulation and for the last 50 ns.

Stability of the Protein−RNA Interface. There are only four protein−RNA H-bond interactions in the original NMR structure, with Arg15 and Arg10 side chains H-bonded with the Hoogsteen edges of the G7 and G31, respectively (Figure 8). In both initial simulations, the Arg10/G31 interaction was permanently lost shortly after the start. The Arg15/G7 interaction was locally deformed, with the arginine partially interacting with the downstream U8 base. The simulations revealed numerous short-lived protein−RNA interactions due to the dynamic nature of the short unstructured peptide chain (see below).

Dynamics of the RNA Duplex−Quadruplex Junction. The quadruplex part of the molecule was fully stable in both simulations. Both K^+ ions initially placed inside the channel formed by the three guanine tetrads maintained their position. The bases of the quadruplex loops fluctuated and were mostly exposed to the solvent.

The mixed tetrad has been significantly rearranged in the 2la5_ff99 simulation after 50 ns (Figure 9), while it was more stable in the 2la5_ff12 simulation, albeit there were still fluctuations of the A17 nucleotide position and a conformational change of the A17(N6)-U8(O4) H-bond. The 2la5_ff99 simulation also revealed permanent loss of stem duplex G7=C30 and A33-U3 canonical base pairs since 300 and 540 ns, respectively.

Dynamics of the FMRP RGG Peptide. According to the NMR structure, the peptide lacks any specific secondary structure. It displays highly dynamical behavior in simulations with a backbone average RMSD of 2.7 Å (Supporting Information, RMSD graphs).

Structural Ions. The NMR structure did not resolve any ions, however, two monovalent ions were standardly placed between the three G-tetrads. Besides that, three highly occupied ion-binding sites developed spontaneously in the simulations (Figure 10). In the 2la5_ff99 simulation, there were three K^+ ion sites with nearly 100% occupation time. The first ion site was coordinated by U28(O2), G29(O6, N7), Arg15(O), C30(N3), and A17(N1) atoms. A K^+ ion partially moved to this site immediately after the start and permanently settled

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

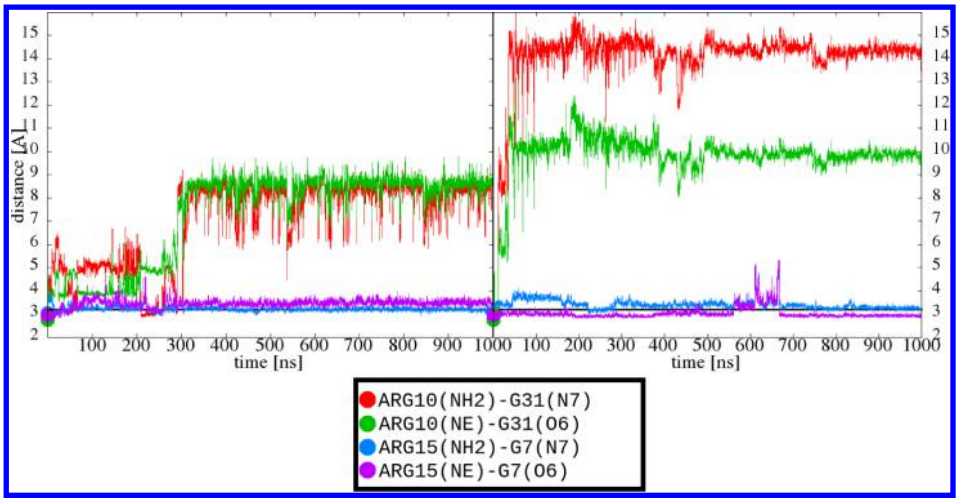


Figure 8. Time development of H-bond heavy atom distances (see the legend) in 2la5_ff99 (left) and 2la5_ff12 (right) simulations. The colored points indicate the initial values. The black horizontal line marks 3.2 Å.

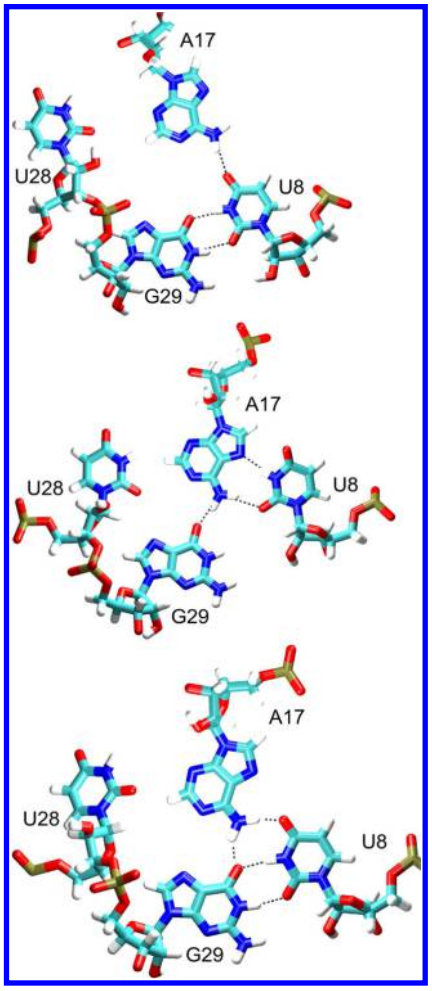


Figure 9. Mixed junction tetrad in the NMR structure (top), in the 2la5_ff99 simulation (middle) and in the 2la5_ff12 simulation (bottom).

there after 300 ns. This event appeared to be an unsuccessful (not completed) attempt of the K^+ ion to reach the center of the mixed tetrad (see below). The second site contained ion coordinated by G16(N3, O2′), and G18(N3, O4′) atoms with ion entrance after 100 ns. The third site involved the G21(N3,

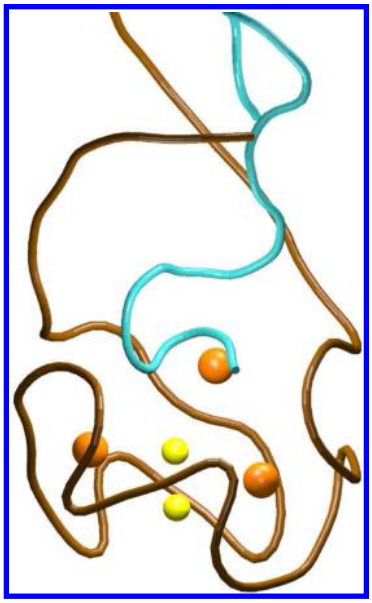


Figure 10. Long-residency highly occupied K^+ ion sites of the FMRP RGG peptide (blue)−RNA duplex−quadruplex (brown) complex in the 2la5_ff12 simulation. Yellow spheres, quadruplex ions; orange spheres, the other ion binding sites discovered by the simulations.

O2′) and G26(N3, O2′) atoms. The first 2la5_ff12 simulation revealed a modest difference in the first ion site, as the ion was coordinated by U28(O2′), G29(O1P), Arg15(O), and Gly16-(O) atoms and exchanged several times with the bulk solvent. The other two ion sites developed similarly in both simulations.

The K^+ ions also often shortly bound to the peptide main chain carbonyl atoms. These atoms normally participate in secondary structure interactions in most proteins. Due to the lack of secondary structure, they were largely exposed to the solvent in our system.

Placing a Channel Ion into the Mixed Tetrad Stabilizes the System. The two initial simulations indicated that ions were unsuccessfully attempting to reach the mixed tetrad. Spontaneous complete insertion of the ion from the bulk solvent would require a partial unfolding of the structure to create a path for the ion. This assumption was supported by the RNA-only simulation (2la5_RNA) where the ion moved into the

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

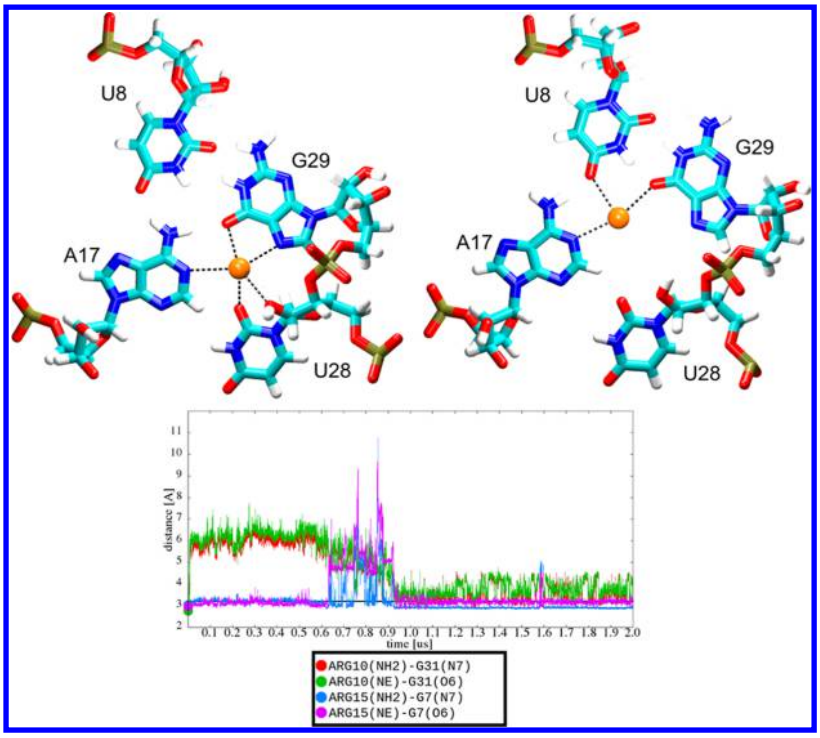


Figure 11. During the 2la5_ff12_3K simulation, the K^+ ion (orange sphere) was initially inserted near the mixed tetrad. It then moved into the tetrad plane, being coordinated by the A17(N1), G29(O6), U28(O2), U28(O2′), and G29(N7) atoms from 25 to 620 ns (top left) and by the U8(O4), A17(N1), and G29(O6) atoms from 620 ns until the end (top right). The coordination is indicated by black dotted lines. This was followed by gradual almost complete restoration of the experimental structure (bottom). The graph indicates time development of H-bond heavy atom distances (see the legend). The colored points indicate the initial values. The black horizontal line marks 3.2 Å. While the Arg10/G31 pair still fluctuated, there was always at least one stable H-bond contact between the two residues after 950 ns.

mixed tetrad shortly after start, passing directly through space normally occupied by the peptide chain. We have therefore built another protein−RNA system where we initially placed an additional K^+ ion between the last G-tetrad and the mixed tetrad. In the 2-μs simulation (2la5_ff12_3K), the ion moved after 25 ns into the plane of the mixed tetrad. It remained coordinated by A17(N1), G29(O6), U28(O2), U28(O2′), and G29(N7) atoms until ∼620 ns when it moved to its final position with coordination to U8(O4), A17(N1), and G29(O6) (Figure 11). While the mixed junctional tetrad initially showed the same deformations as those observed during the initial 2la5_ff12 simulation (Figure 9), the experimental NMR arrangement was restored when the K^+ reached the final position (after 620 ns). While the K^+ ion still occasionally fluctuated after that, the mixed tetrad arrangement was restored permanently. Amazingly, the protein−RNA H-bond interactions started gradual reparation after 620 ns, reaching nearly the original arrangement around 1 μs and then being maintained until the end of the simulation (Figure 11). This simulation thus gave us two lessons. First, although not apparent from the experimental structure, the mixed tetrad was predicted to be a major ion-binding site by MD. Second, although the protein−RNA interface was initially heavily perturbed, the simulation was capable on a ∼1-μs time scale to essentially relax back to the experimental arrangement. However, it required an appropriate manual initial placement of the ion. This is a striking illustration of the challenges imposed by the starting structures and simulation time scale in simulations of protein−RNA complexes. This is further substantiated by another simulation (2la5_ff12_3K_2) where we used a nearly identical initial structure as in the 2la5_ff12_3K simulation except that the third K^+ ion was initially placed into a slightly different position inside the mixed tetrad. In this 1-μs simulation, the mixed tetrad and the protein−RNA interface were not restored (Supporting Information Figure S34).

Comparison with the Experimental NMR Data. We have calculated average violations of the experimental NOE intermolecular distances along the simulation trajectories (see the Methods).[64] The results indicate qualitative agreement with our other analyses. Specifically, the best behaved 2la5_ff12_3K simulation showed the smallest number of NOE violations with the lowest median (Table 4). The tendency for gradual restoration of the structure was visible, with almost complete agreement with the experiment achieved at the end of this simulation. Specifically, at the end, only seven of the remaining NOE violations were over 1 Å and six of those involved inherently flexible parts of the molecule (such as chain ends). The one remaining important NOE violation was between Arg15(HB) and G18(H1) atoms. It was caused by different conformation of the Arg15 side chain, which resulted from several simulation flips of its χ dihedrals. It could be indicative of some remaining inaccuracy, as it was coupled with minor perturbation of the H-bond interactions between Arg15 and the G7 (see above).

Simulations of the Complex of Pumilio FBF-2 Protein with gld-1 FBEa RNA Could Not Be Stabilized. *Stability of the Protein−RNA Interface.* The initial ff99SB and ff12SB simulations were very similar and showed a progressive loss of the protein−RNA interface (Figure 12). This process began from both RNA strand ends. Nucleotide A9 became flexible immediately. It flipped away from the protein and was usually

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

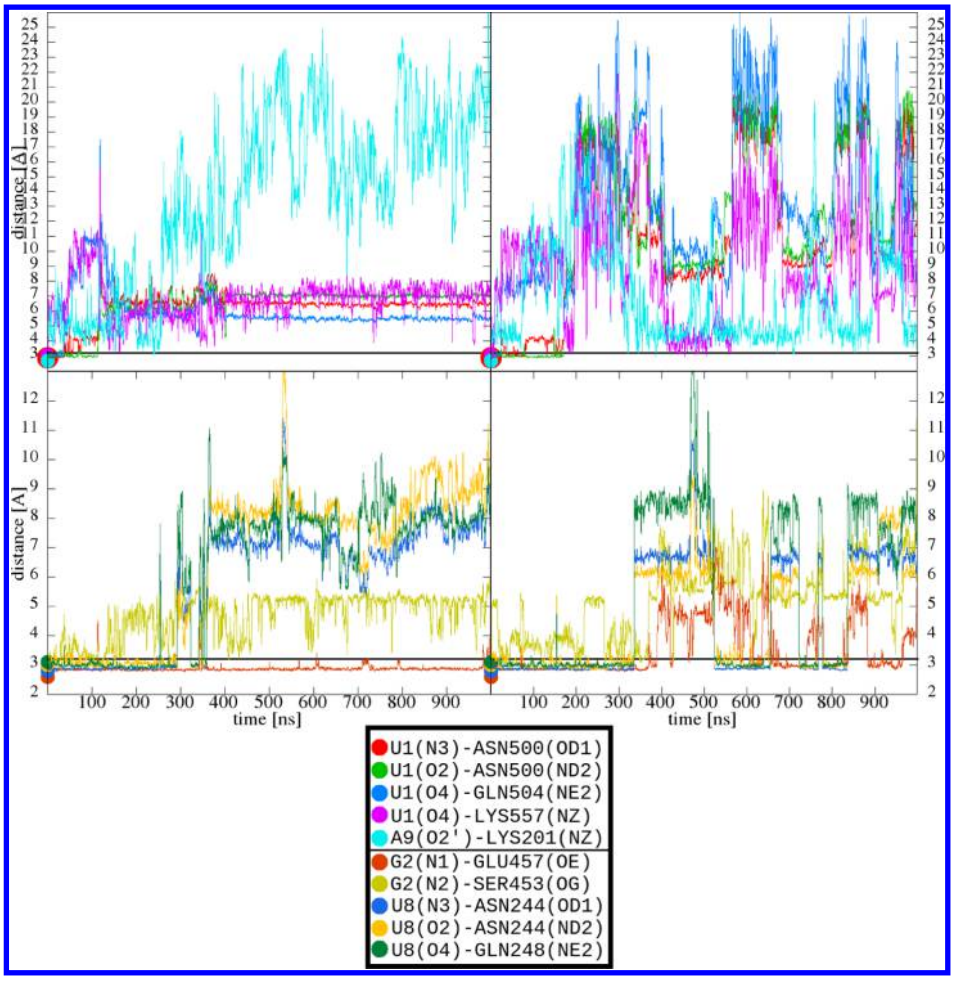


Figure 12. Time development of H-bond heavy atom distances of two 5′- and 3′-end terminal RNA nucleotides in the 3k5y_ff99 (left) and 3k5y_ff12 (right) simulations (for the remaining nucleotides see the Supporting Information Figure S35).

exposed to the solvent for the rest of the simulations, without forming any stable interactions with the protein.

The highly conserved[68] protein−RNA interactions involving U1 were irreversibly lost after dozens of nanoseconds (Figure 12). Similarly to the A9 residue, the U1 was then usually entirely flexible. As the simulations progressed, the loss of the protein−RNA contacts propagated from the strand ends to the neighboring residues (G2 and U8; see Figures 12 and 13). We then tried to stabilize the system by using distance restraints on the conserved protein−RNA H-bond interactions of U1 and U8 nucleotides (there are no suitable conserved contacts of the A9 that could be effectively restrained). However, even in these simulations, we were unable to sufficiently stabilize the protein−RNA interface (Supporting Information Figures S36 and S37). In summary, we have carried out five simulations of this complex (Table 1), but we did not see any sign of its stabilization. We suggest that the behavior of the simulations may be significantly influenced by the crystal packing effects that are affecting the structure and are absent in the simulations (Supporting Information Figure S38). In principle, we could try simulations including the crystal lattice. However, such simulations are presently difficult even for simple systems such as canonical A-RNA double helices,[96] so we do not assume they would be straightforwardly applicable for our system.

Dynamics of the FBF-2 Protein. The internal structure of the Pumilio FBF-2 protein was fairly stable in our protein−RNA simulations. The loss of the protein−RNA interactions was primarily facilitated by conformational changes of the RNA. Nevertheless, we observed changes in the simulations of the isolated protein molecule (Table 1). Specifically, the residues normally associated with the bound RNA molecule gradually shifted away from their starting positions. Especially, the hydrophobic amino acid residues had a tendency to reduce their exposure to the solvent by moving inward, closer to the protein. Similar changes were observed also in the protein−RNA simulations once the intermolecular interactions were lost. Since our starting coordinates for the isolated protein simulations were obtained by removal of the bound RNA molecule, changes in the vicinity of the former protein−RNA interface during the simulation were not unexpected. It suggests that the Pumilio FBF-2 RNA binding domain does not have a permanent stable RNA binding interface and is structured upon arrival of the RNA molecule via induced fit at least at the level of side chains, which allows it to bind different RNA sequences. We speculate that the delicate balance of forces at the protein−RNA interface may be difficult for the force field description.

Histidine Protonation. In the Pumilio complex, His454 interacts with the RNA only by stacking between the bases and the histidine ring. We usually select the histidine protonation state based on maximizing the number of H-bonds in the

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

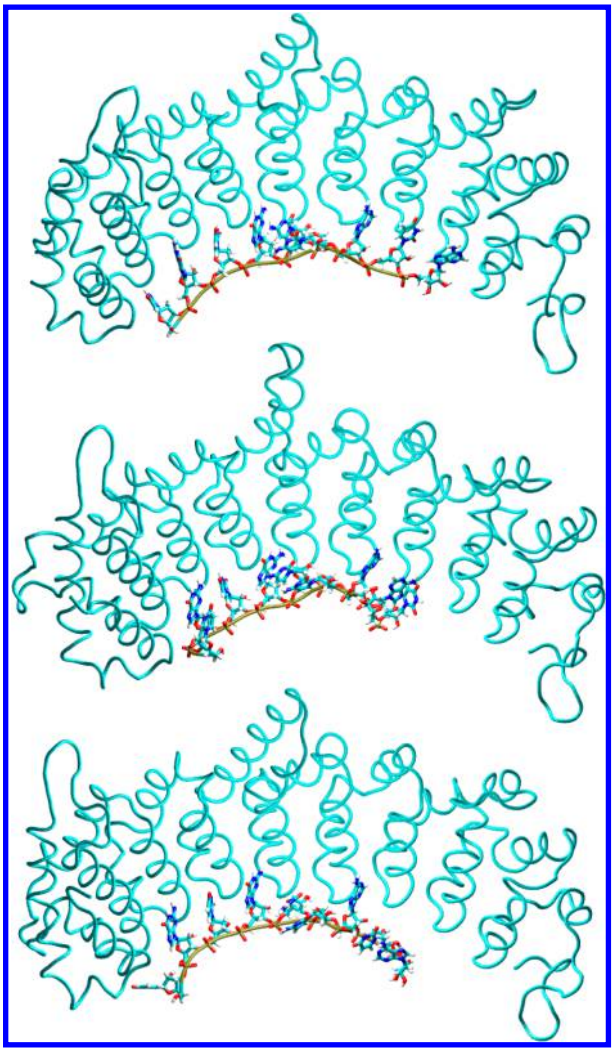


Figure 13. Progressive loss of the protein−RNA interactions of the Pumilio complex in the 3k5y_0 simulation − 0 ns (top), 500 ns (middle), and 1000 ns (bottom). The ssRNA moves away from the protein.

experimental structure. However, this could not be applied for His454 having no hydrogen bonding that would indicate the protonation state. Thus, the assignment was done in a trial and error fashion. Our first guess (ε-protonation) was most likely wrong (3k5y_0 simulation, see Supporting Information Figure S39 and S40). In the initial structure, the histidine ring was only interacting with bases G2 and U3 by stacking. However, in the simulation, it formed a stable G2(N2)-His454(ND1) interaction. This was facilitated by change in the χ-torsion of the G2 nucleotide after ∼190 ns of the simulation (Supporting Information Figure S41). Note that we had no ultimate proof that such development was incorrect. Nevertheless, we suspected that it might have been aiding the process of breakage of the protein−RNA contacts (see above). In all subsequent simulations, the His454 was therefore double-protonated, thus preventing formation of the G2(N2)−His454(ND1) interaction (Table 1). Nevertheless, the simulations could not be permanently stabilized irrespective of the His454 protonation state.

Tyrosine Ring and Hydroxyl Group Rotation. We noted problematic behavior of the tyrosine amino acids residues when using the ff99SB force field. There is an excessive penalization of rotation of the tyrosine hydroxyl group (Supporting Information Figure S42). In simulations, it took often hundreds of nanoseconds before this rotation occurred (compared with rotation of the whole tyrosine aromatic ring, which readily occurred on picosecond time scale unless obstructed by the environment − see below). Our free energy calculations indicated that there is a barrier of ∼8 kcal/mol for the rotation of the tyrosine hydroxyl group in the ff99SB force field (Supporting Information Figure S42). This resulted into time scale of hundreds of nanoseconds for rotation of the hydroxyl group. A consequence in the simulations was inability of the tyrosine hydroxyl group to quickly respond to structural changes in its vicinity (the hydroxyl group can act as both H-bond donor and acceptor). The conformation of the tyrosine hydroxyl group may also be locked in the initial automated hydrogen atoms assignment during the system building by xleap. In cases where the assignment by xleap was wrong, we noticed instabilities in the structures (the Pumilio, L1 stalk, and U1A complexes) that sometimes became irreversible before the hydroxyl group could properly reorient. Sometimes the whole tyrosine aromatic ring flipped over before the hydroxyl group could rotate. Note that given the high energy barrier, the position of the hydroxyl group could not be corrected by the equilibration. Another ff99SB issue resulted from the lack of specific dihedral parameter for rotation of the tyrosine aromatic ring (Supporting Information Figure S42). Its dihedral profile was determined by the other parameters. In our opinion, the tyrosine aromatic ring rotation might be excessively flexible. For example, in the Tyr/base stacks of the Pumilio simulations, the dynamics of the stacks was entirely dominated by the nucleobase movements. In principle, this issue also applies to the phenylalanine aromatic ring, which has the same dihedral parameters as tyrosine.

Ff12SB increased and reduced the energy barrier for ring and hydroxyl group rotation, respectively. Although it still did not prevent the breakage of the protein−RNA interactions in the Pumilio complex, it eliminated the problematic behavior of tyrosines in the L1 stalk and some minor issues in the other systems.

DISCUSSION AND CONCLUSIONS

We carried out over 30 μs of MD simulations on six diverse protein−RNA complexes in explicit solvent. The main aim of our work was to investigate applicability of the MD technique for protein−RNA systems, an issue that has not been systematically addressed in the literature so far. As explained in the Introduction, although we should not always expect a perfect agreement between the experimental structure and the MD simulation behavior, microsecond simulations using good quality experimental structures as the start should not dramatically deviate from the starting structures and should preserve basic features of the observed protein−RNA recognition patterns.

Simulation Behavior Is Highly Variable. Among the six complexes, one system (dsRNA−B2 protein) provides stable trajectories in full agreement with the experimental structure, except that the simulations suggest incorrectly refined X-ray position of the Asn40 side chain (Figure 3). The Asn40 side chain flip leads to base-specific Asn40(ND2)−G5(N3) interaction while the original experimental work suggested sequence-independent RNA binding.[48] Our simulations indicate that G or A could be preferred as the fifth nucleotide in

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

the RNA recognition pattern. Miss-assignments of side chain rotamers of asparagine and glutamine are not uncommon.[92]

For another system (Pumilio FBF-2), we could not find any approach to stabilize the trajectories close to the experimental structure. This behavior can be affected by crystal packing forces straining the starting structure (see the Results and Supporting Information for a full description).

For the remaining four systems, we report highly diverse simulation behavior in the individual trajectories.

The complex between the L1 stalk and rRNA shows an area of shape recognition between the kink-turn 77 of the RNA and the protein, lacking sequence-specific hydrogen bonding. This recognition is essentially preserved in all simulations. The complex further contains a binding area between the helix 77 and the protein, containing H-bonds. These H-bond interactions are only partially reproduced by the simulations. Further, we report a modest reproducible difference between the ff99SB and ff12SB protein force fields, including improved description of the tyrosine side chain in the ff12SB force field, which eliminates the high rotational barrier of tyrosine hydroxyl group present in the ff99SB version. The simulations confirm that two fully buried monovalent ions contribute to the protein−RNA recognition. We also observe fluctuations in the secondary structure of the protein. Namely, part of the α5 helix of the L1 protein is reversibly changing into β-sheet-like structure that could serve as an additional binding region for the deacylated tRNA molecule (Figure 5).

In summary, the L1 protein/23S rRNA segment complex is quite well described by the simulations albeit not all observed H-bonds could be kept simultaneously. The complex appears to be primarily stabilized by the shape-recognition, in line with earlier site-directed mutagenesis studies reporting largest drop in stability when nucleotides constituting the shape of the rRNA molecule (i.e., the kink-turns) were mutated.[53] On the other hand, mutations that abolished specific protein−RNA H-bond interactions had smaller impact.

The U1A−RNA hairpin complex contains an intricate network of H-bonds, and none of our simulations was capable to keep all of them simultaneously stable. Nevertheless, the simulated structure did not progressively deteriorate. Many individual native interatomic contacts (H-bonds) showed large fluctuations, but they were capable to temporarily return back to the starting values. However, the simulations had evident difficulty to simultaneously preserve all the native interactions. We suggest that it may be a hallmark of problems in simulations of protein−RNA complexes possessing dense networks of H-bonds that are mutually affecting each other. The full set of simulations did not reveal any noticeable difference between the ff99SB and ff12SB descriptions. The observed variability in the simulation behavior was attributed to the insufficient sampling of the microsecond-scale simulations. We have further noticed that the behavior of the individual simulations was predetermined by their early stages. The simulations always quickly lost some H-bonds, due to some strain caused either by the force field or by some hidden problem of the starting structure that we did not identify. Once the simulated structures lost some of the H-bonds, the simulations became steady. We identified several typical simulation behaviors, but the presently affordable simulation time scale did not allow us to obtain a fully converged picture for this system.

Simulations of the complex between the SAM domain of Vts1p and RNA hairpin (2f8k) nicely illustrate complexity of protein−RNA complex simulations. While in some simulations the protein−RNA interface collapses, other simulations are stable. The results do not reflect differences between the ff99SB and ff12SB force fields, as both force fields can provide stable as well as unstable simulations. The simulations appear to be substantially stabilized when (i) replacing the two terminal AU base pairs by more stable GC base pairs and (ii) removing the 5′-overhang nucleotide. This part of RNA does not directly participate in the protein−RNA interface in the X-ray structure. However, the modification prevents excessive RNA end fraying and interactions between the frayed and terminal nucleotides with the protein. Such interactions appear, at least in some cases, to destabilize the complex. They may reflect force field limitation of the RNA description that propagates to the protein−RNA interface. Nevertheless, one stable microsecond trajectory has been obtained also for the unmodified system. Thus, we suggest that the individual simulations may reflect a combined effect of random end-fraying of the RNA helix, random dynamics of the overhanging 5′-end nucleotide, and their interactions with the protein. Further, the simulations may reflect yet another and more general problem related to sampling. The initial experimental structures are necessarily locally unrelaxed, for a variety of reasons discussed in the Results section. Then, in some simulations, due to their stochastic nature, the initially high energy of the system can result in perturbation at the beginning of the production simulations. The initial structural perturbations may sometimes be barely noticeable at the beginning and may result in visible geometrical changes only in later stages of the simulation, once the system crosses relevant energy barriers (cf. Figure 7, left). Other simulations of the same system can shake off the initial stress through different rearrangements that avoid the initial perturbation, and then, they can remain in agreement with the experimental structure.

Perhaps the most interesting insight can be obtained from simulations of the complex of the FMRP RGG peptide and the RNA duplex−quadruplex junction (2la5). There are only four protein−RNA H-bond interactions in the original NMR structure (Figure 8). Our initial simulation attempts could not reproduce this H-bond pattern. However, we noticed that a monovalent ion is attempting to penetrate into the mixed RNA tetrad forming the quadruplex−duplex junction, which was obviously not apparent from the experimental NMR structure. However, binding of the ion would require temporary separation of the protein−RNA complex, which did not occur spontaneously on our simulation time scale. Thus, we manually added the ion somewhere between the mixed tetrad and the adjacent G-tetrad. One of these two simulations, after usual initial perturbation of the structure, spontaneously found an ion position that after more than 0.5 μs resulted in reparation of the mixed tetrad and of the protein−RNA interface. Good agreement with the NMR data has then been achieved (Table 4). Thus, our simulations of the 2la5 NMR structure provided an amazing illustration of the sensitivity of the simulations to the details of the starting structures as well as of the time scale limitations. To obtain one stable simulation that keeps the experimental structure, we had to first recognize that there is a key structural ion binding site not apparent from the experimental data. Then, we had to add the ion to the predicted area manually, since the ion was unable to spontaneously reach its final destination on the microsecond time scale. Finally, even when starting simulations with the added structural cation, one simulation was not capable to

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

repair the starting structure while the other required more than 0.5 μs to relax the ion position and repair the protein−RNA interface.

In summary, we investigated six protein−RNA complexes, and we observed wide range of simulation behaviors (Table 5).

Table 5. Final Overview of the Simulated Protein−RNA Complexes

protein−RNA complex	PDB code	interface size[a] [Å^2]	dissociation constant (K_D)[b]	structural stability in simulations[c]
B2−dsRNA complex	2az0	2486	∼1 nM[48]	+++
L1−23S rRNA fragment	3u4m	2986	∼2 pM[51]	++
U1A−RNA hairpin	1urn	1861	∼20 pM[97]	++
Vts1p SAM−RNA hairpin	2f8k	933	∼25 nM[61]	+
FMRP RGG motif−RNA duplex−quadruplex junction	2la5	1779	∼4 nM[64]	+
Pumilio FBF-2−gld-1a RNA	3k5y	2537	∼30 nM[68]	−

[a]Size of the protein−RNA interface of the initial structures was calculated using the LCPO method.[90] [b]Experimentally determined dissociation constants of the individual protein−RNA complexes. [c](+++) perfect structural stability; (++) overally stable with modest structure perturbations; (+) unstable unless the starting structure was modified; (−) unstable.

For one system (2az0), we easily obtained fully stable microsecond trajectories. For another system (3k5y), all trajectories resulted into a progressive loss of the protein−RNA interface. For two additional systems (3u4m and 1urn) we obtained more or less globally stable complexes though none of our simulations was capable to simultaneously stabilize all the experimentally observed H-bonds. The most intricate were the remaining systems (2f8k and 2la5). For these systems, although many trajectories were unstable, we obtained at least one long trajectory which stabilized the experimentally observed protein−RNA H-bonding interface. In case of the 2la5 system, to get the stable simulation required a manual intervention into the starting structure. Even after this, it required more than 0.5 μs for the system to start to relax back to the starting structure after the initial perturbation.

Can We Make Stable Simulations of Protein−RNA Complexes? The Starting Structure Issue. Let us finally come to our initial question. The above data suggest that the starting structures represent the most critical issue in protein−RNA simulations. Each protein−RNA complex should be simulated after careful initial inspection of the experimental structure and preliminary simulation tests to identify potential sources of instabilities.[20] Obviously, correct simulations require appropriate histidine (and sometimes nucleobase, cf. the L1 stalk system) protonation states, which are not always unambiguous. With the ff99SB force field, initial positioning of the tyrosine hydroxyl group should be inspected. In specific cases, modification of potentially unstable segments of the molecules not directly involved in the recognition may be considered, for example, to reduce RNA helix end fraying and other end effects that may derail the simulation. In other cases, manual placement of monovalent ions not resolved by the experiment may stabilize the trajectories.

However, even when the structure is correctly refined from the experimental point of view and all the above-noted issues are properly addressed, we may still obtain unstable simulations. The starting structures may still affect the simulation in a random manner in a way that is not common for small homogeneous systems. The reason is as follows: Even when using in all aspects correct (within the resolution limits) starting structure, the structure still does not correspond to an energy minimum as defined by the force field and is therefore potentially strained. It contains unrelaxed covalent structure, locally unrelaxed molecular contacts, sugar−phosphate backbone geometries, etc. Crystal packing effects can also contribute to deformation of the starting structure. The starting structure thus can be, for a number of reasons, high in potential energy and needs to relax. Standard equilibration protocols that are sufficient for small systems may be not robust enough for full relaxation of complicated protein−RNA complexes, and then, a large part of the relaxation process may be shifted to the production simulations. Then, in some simulations, due to their stochastic nature, the initially high energy of the system can result in random toxic structural perturbation at the beginning of the production simulations, which may immediately or with some delay derail the simulations. Approximations of the force field may further complicate the simulations. However, independent simulations of the same system may result into well-behaving trajectories. Obviously, with a correct force field, the simulation should be capable to eventually repair the initial perturbations. However, it is not guaranteed that such stabilization is always achieved on the presently affordable time scale. Such systems require performing multiple and long simulations. For example, the 2f8k complex can be very satisfactorily simulated albeit many individual simulation attempts result in unstable trajectories.

Comments on the Force Fields. The ff99SB and ff12SB variants of the Cornell et al. protein AMBER force field have similar performance in describing the protein−RNA interface. The ff12SB variant, nevertheless, improves some simulation features. The results do not mean, however, that the simulations are not limited by the protein force field, because both tested variants share the same electrostatics and are based on the same pair-additive approximation. Also, the RNA force field is known to have accuracy limitations, as reviewed elsewhere.[20] We have used the ff99bsc0χ_{OL3} variant, which most likely is presently the best choice for RNA.[20]

Specific Recognition vs Shape Recognition. Our data indicate that the capability of the MD technique to describe the protein−RNA complexes also depends on the nature of the protein−RNA recognition, with some recognition patterns suppressing and others amplifying the force field inaccuracies. It looks that systems having dense network of H-bonds at the protein−RNA interface (as exemplified by the 1urn structure) may be challenging for force field description, as the closely spaced H-bonds (their force field description) may mutually interfere with each other. On the other hand, the shape recognition part of the L1 stalk complex lacking sequence-specific H-bonds is perfectly stable. We notice a trend that the larger is the role of "overlap" shape recognition between specifically preformed shapes of both interacting partners, the more stable the simulations appear. Also, those complexes where both the protein and RNA are structured independently may be more stable in simulations than those where the RNA and protein substantially affect each other's structures via induced fit recognition. An important role may be played not only by the overall stability (the balance between unfolded and folded monomer structures in the thermodynamics sense) but

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

also by the properties of the RNA and protein conformational basins utilized in the recognition. A narrow well-defined native conformational basins of the monomers that are not accompanied by easily (on the simulation time scale) accessible competing alternative conformations are likely to simplify the force field description of the protein−RNA interface.

In the B2 protein−dsRNA and L1 stalk complexes, both RNA and proteins appear to be well structured even in isolation. The kink-turns (present in the L1 stalk RNA) are known to require protein for folding; however, once folded, they adopt a very salient shape.[30] In addition, this specific kink-turn is recognized mostly in a shape recognition mode. The RNA hairpins participating in the U1A and SAM domain complexes are intrinsically more plastic (and challenging for force fields), which may also complicate the description of the protein−RNA interface.

The FMRP RGG peptide−RNA duplex−quadruplex junction complex may at first sight deviate from the trends, as both molecules are unfolded in isolation. They require to have their partners to be structured.[64] However, the quadruplex stems, once formed, are known to be exceptionally stiff.[46] Thus, the quadruplex−duplex junction, once properly structured, represents a very well-defined stiff scaffold for the short flexible peptide, which in addition establishes a quite simple network of just four H-bonds with the RNA.

The most problematic system is the Pumilio FBF-2/gld-1a complex. Its simulations result in a gradual decay of protein−RNA interactions. It starts from strand ends on nanosecond time scale and continues with penultimate nucleotides within tens of nanoseconds (Figure 12). This could not be prevented even by restraining the protein−RNA interactions involving the terminal bases. The gradual loss of terminal nucleotides in protein−RNA interactions could be a force field issue similar to the excessive end-fraying observed in terminal DNA−RNA duplex base pairs.[95] In this particular case, neither of the subsystems is well structured in absence of its partner. The isolated RNA single strand would obviously be very flexible and also simulations of the isolated Pumilio protein clearly show rather fast reorientations of the surface amino acids. Nevertheless, this system can also be affected by extensive crystal packing interactions between the RNA strand ends and the neighboring protein−RNA complex in the crystal lattice. In the crystal lattice, the RNA strand end nucleotides are interacting with the protein from both sides. This stabilizes the inherently dynamical RNA strand ends with additional interactions that are missing in the solution simulations (Supporting Information Figure S38). It is possible that the problems in Pumilio complex simulations are a combination of force field issues and a realistic response of the system to the loss of crystal packing interactions in the simulations.

Interestingly, we observed changes in the size of the protein−RNA interface for several of the systems in the simulations (Table 2). Each of the changes was caused by a different factor. In the 3u4m simulations, there was an RNA helix end randomly interacting with the protein and increasing the interface size. Due to overall structural collapse, the interface was enlarging and shrinking in the unmodified 2f8k and 3k5y simulations, respectively. However, we observed a trend for slight increase (~5%) of the protein−RNA interface size even in otherwise completely stable systems (e.g., 2az0 and modified 2f8k simulations). We hypothesize that this may be a result of the loss of the crystal packing interactions in the solvated system.

Concluding Remarks−Methodological Suggestions How to Perform and Interpret Atomistic Protein−RNA Simulations. We demonstrate that different types of protein−RNA complexes are captured by the simulation technique with a highly variable degree of reliability. Capability of the simulation technique to reproduce the experimental protein−RNA interfaces should be evaluated case by case.

The MD technique provides entirely stable trajectories for some protein−RNA complexes (2az0 and part of 3u4m). A more common outcome is that the simulation achieves reasonable agreement with the experiment but not all details of the structure are reproduced (as exemplified by the 1urn and 3u4m systems). There are also experimental structures for which stable unrestrained simulations cannot be achieved, with a progressive loss of the protein−RNA interface (3k5y).

The structural stability of the simulated systems does not correlate with the size of buried interaction surface and the experimentally determined binding affinities (i.e., the thermodynamic stability).

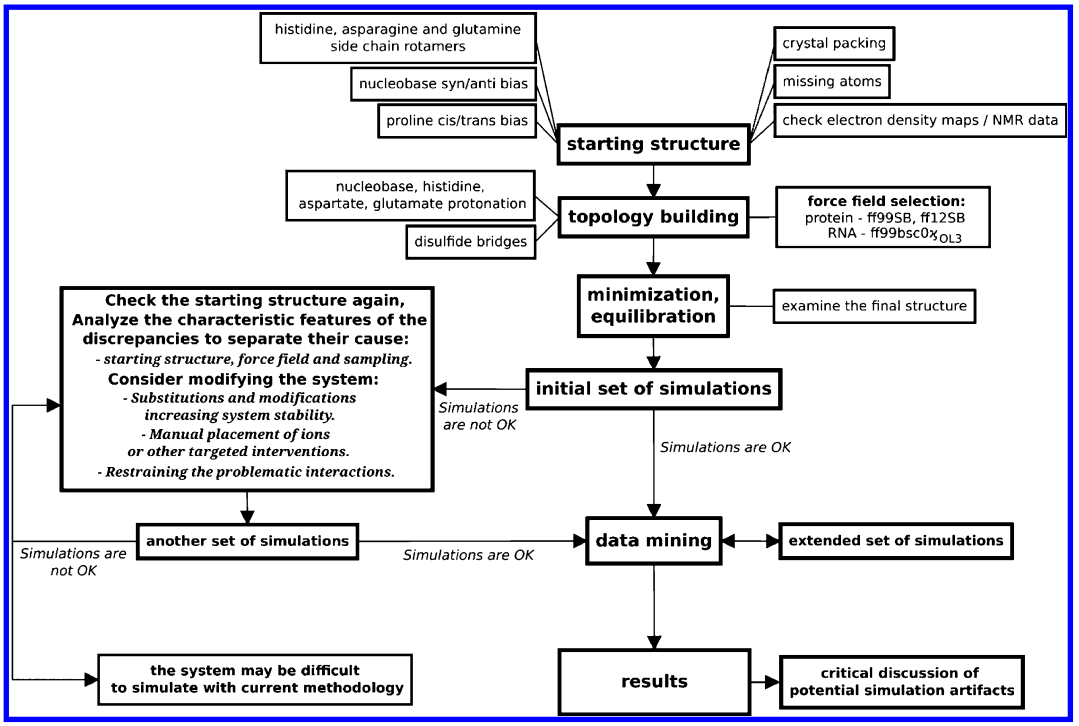
The simulations are affected by the type of protein−RNA recognition. Segments relying on shape-specific recognition of RNA are more easily described by the simulation methodology. Segments relying on sequence-specific recognition via H-bonding are more difficult. Several of the studied complexes indicate that the force fields are sometimes incapable to simultaneously stabilize all the native H-bonds, especially when having a dense network of H-bonds at the interface. We suggest that it can indicate coupling of subtle inaccuracies of the force field description of the individual interactions, resulting in conflicts between (the description of) the individual H-bonds. Typical sign of such behavior is loss of a few H-bonds in early stages of the simulation with subsequent steady simulation behavior (1urn and 3u4m). The simulation does not further deteriorate, but also, it does not show any tendency to return back to the full set of native contacts, despite being temporarily able to form the individual specific interactions at non-overlapping time periods.

For some systems, a limited amount of shorter simulations can indicate false unstable behavior despite that, for these systems, the force field in principle would be capable to provide a correct description. It this case, the initial instability is rather related to the starting structure. Longer simulations accompanied eventually by manual interventions into the starting structures based on in-depth simulation analysis can ultimately lead to stabilization of such complexes in amazing agreement with the experiment. Simulations of 2f8k and 2la5 complexes show practical protocols how such calculations could be executed. Our study supports a view that properties of the starting structures is presently the most critical factor affecting behavior of the subsequent simulations of protein−RNA complexes.

Our simulations also provide an indication that the standard equilibration protocols that are sufficient to prepare production simulations of isolated medium-sized RNA molecules may be less robust for simulations of at least some protein−RNA complexes. In other words, we suggest that even a starting experimental structure that is in all aspects essentially correct may destabilize the subsequent simulations. It is because the starting structure is necessarily high in force field potential energy. Although we cannot prove this suggestion directly, simulation behavior of several systems supports this view, as in detail explained for the 2f8k system. Typical behavior in this case is capability of some simulations to remain stable at the

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

Scheme 1. Example of Work Flow Scheme Highlighting a Possible Best-Way Approach for Executing Simulations of Protein−RNA Complexes[a]



[a]Important issues to consider at the individual steps are emphasized. Note that the scheme is only indicative and some systems may require additional actions to be taken. The essence is to emphasize that simulations of protein−RNA complexes require a very careful preparation and interpretation, which need to be done case by case. Relying on popular simple analytical tools such as principal component analysis or cross-correlation diagrams without full understanding of the trajectories may easily lead to false-positive results, when simulation artifacts caused by the starting structures and force fields are overlooked or even misinterpreted as signs of biochemically relevant dynamics.

experimental structure while other independent but otherwise equivalent simulations are unstable. This scenario is also consistent with capability of the system to return back to the starting structure in later stages of the simulations. To investigate this simulation scenario is obviously exceptionally demanding in terms of the simulation time.

Our study strongly suggests that multiple microsecond-scale simulations should become a standard for protein−RNA simulations. For several of the presented systems, more economical simulation time scale would provide entirely misleading picture (cf. mainly the data for 2la5, 2f8k, and 1urn). We obviously do not claim that performing multiple microsecond-scale simulations provides entirely converged results.

Our work provides a set of benchmark simulations for diverse protein−RNA complexes that can be used for assessment of other biomolecular force field parametrizations. Note that properties of complex protein−RNA systems are different from properties of small isolated RNAs and protein molecules that are usually used to test performance of biomolecular force fields. Problems in MD simulations of protein−RNA complexes in contemporary literature may be underreported.

Finally, we present a short workflow diagram that summarizes the problems and issues encountered during our simulations. It can serve as template for conducting simulations of other protein−RNA complexes (Scheme 1).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Comments on structure selection and protonation assignment; extensive graphical overviews of the protein−RNA interfaces of the complexes; figures of time developments of H-bond interactions; dihedral profile energy analysis of several residues; discussion of protein−RNA phosphate interactions and role of Mg^{2+} ion in the L1 stalk complex; graphs of time developments of RMS deviations of all simulations; graphs of time developments of protein−RNA complex interaction surface areas in all simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sponer@ncbr.muni.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grant P305/12/G034 from the Grant Agency of the Czech Republic. Institutional funding was further obtained by project "CEITEC—Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund and by the Operational Program Research and Development for Innovations—European Regional Development Fund (CZ.1.05/2.1.00/03.0058). We thank Ms. Marie Sarazova for help with initial selection of the structures.

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

REFERENCES

(1) Varani, G.; Nagai, K. RNA Recognition by RNP Proteins during RNA Processing. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 407−445.

(2) Chen, Y.; Varani, G. Protein Families and RNA Recognition. *Febs J.* **2005**, *272*, 2088−2097.

(3) Hogg, J. R.; Collins, K. Structured Non-Coding RNAs and the RNP Renaissance. *Curr. Opin. Chem. Biol.* **2008**, *12*, 684−689.

(4) Änkö, M.-L.; Neugebauer, K. M. RNA Protein Interactions In Vivo: Global Gets Specific. *Trends Biochem. Sci.* **2012**, *37*, 255−262.

(5) Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G. RNA-binding Proteins and Post-Transcriptional Gene Regulation. *FEBS Lett.* **2008**, *582*, 1977−1986.

(6) Stefl, R.; Skrisovska, L.; Allain, F. H. T. RNA Sequence- and Shape-Dependent Recognition by Proteins in the Ribonucleoprotein Particle. *EMBO Rep.* **2005**, *6*, 33−38.

(7) Jarmoskaite, I.; Russell, R. RNA Helicase Proteins as Chaperones and Remodelers. *Annu. Rev. Biochem.* **2014**, *83*, 697−725.

(8) Rinn, J. L.; Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145−166.

(9) Dreyfuss, G.; Kim, V. N.; Kataoka, N. Messenger RNA-Binding Proteins and the Messages They Carry. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 195−205.

(10) Schmeing, T. M.; Ramakrishnan, V. What Recent Ribosome Structures Have Revealed About the Mechanism of Translation. *Nature* **2009**, *461*, 1234−1242.

(11) Wahl, M. C.; Will, C. L.; Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **2009**, *136*, 701−718.

(12) Chen, Y.; Varani, G. Engineering RNA-Binding Proteins for Biology. *Febs J.* **2013**, *280*, 3734−3754.

(13) Lunde, B. M.; Moore, C.; Varani, G. RNA-Binding Proteins: Modular Design for Efficient Function. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 479−490.

(14) Cléry, A.; Blatter, M.; Allain, F. H. T. RNA Recognition Motifs: Boring? Not Quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290−298.

(15) Draper, D. E. Themes in RNA−Protein Recognition. *J. Mol. Biol.* **1999**, *293*, 255−270.

(16) Nadassy, K.; Wodak, S. J.; Janin, J. Structural Features of Protein−Nucleic Acid Recognition Sites. *Biochemistry* **1999**, *38*, 1999−2017.

(17) Iwakiri, J.; Tateishi, H.; Chakraborty, A.; Patil, P.; Kenmochi, N. Dissecting the Protein−RNA Interface: The Role of Protein Surface Shapes and RNA Secondary Structures in Protein−RNA Recognition. *Nucleic Acids Res.* **2011**, DOI: 10.1093/nar/gkr1225.

(18) Connolly, M. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221*, 709−713.

(19) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe. *Acc. Chem. Res.* **2010**, *43*, 40−47.

(20) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771−1782.

(21) Blakaj, D. M.; McConnell, K. J.; Beveridge, D. L.; Baranger, A. M. Molecular Dynamics and Thermodynamics of Protein−RNA Interactions: Mutation of a Conserved Aromatic Residue Modifies Stacking Interactions and Structural Adaptation in the U1A-Stem Loop 2 RNA Complex. *J. Am. Chem. Soc.* **2001**, *123*, 2548−2551.

(22) Kormos, B. L.; Benitex, Y.; Baranger, A. M.; Beveridge, D. L. Affinity and Specificity of Protein U1A-RNA Complex Formation Based on an Additive Component Free Energy Model. *J. Mol. Biol.* **2007**, *371*, 1405−1419.

(23) Kormos, B. L.; Pieniazek, S. N.; Beveridge, D. L.; Baranger, A. M. U1A Protein-Stem Loop 2 RNA Recognition: Prediction of Structural Differences from Protein Mutations. *Biopolymers* **2011**, *95*, 591−606.

(24) Law, M. J.; Linde, M. E.; Chambers, E. J.; Oubridge, C.; Katsamba, P. S.; Nilsson, L.; Haworth, I. S.; Laird-Offringa, I. A. The Role of Positively Charged Amino Acids and Electrostatic Interactions in the Complex of U1A Protein and U1 Hairpin II RNA. *Nucleic Acids Res.* **2006**, *34*, 275−285.

(25) Xue, Q.; Zhang, J.-L.; Zheng, Q.-C.; Cui, Y.-L.; Chen, L.; Chu, W.-T.; Zhang, H.-X. Exploring the Molecular Basis of dsRNA Recognition by Mss116p Using Molecular Dynamics Simulations and Free-Energy Calculations. *Langmuir* **2013**, *29*, 11135−11144.

(26) Lazar, P.; Lee, Y.; Kim, S.; Chandrasekaran, M.; Lee, K. W. Molecular Dynamics Simulation Study for Ionic Strength Dependence of RNA−Host factor Interaction in *Staphylococcus aureus* Hfq. *Bull. Korean Chem. Soc.* **2010**, *31*, 1519−1526.

(27) Reblova, K.; Spackova, N.; Koca, J.; Leontis, N. B.; Sponer, J. Long-residency Hydration, Cation Binding, and Dynamics of Loop E/Helix IV rRNA-L25 Protein Complex. *Biophys. J.* **2004**, *87*, 3397−3412.

(28) Ye, W.; Yang, J.; Yu, Q.; Wang, W.; Hancy, J.; Luo, R.; Chen, H.-F. Kink Turn sRNA Folding upon L7Ae Binding Using Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2013**, *15*, 18510−18522.

(29) Li, W.; Sengupta, J.; Rath, B. K.; Frank, J. Functional Conformations of the L11-ribosomal RNA Complex Revealed by Correlative Analysis of Cryo-EM and Molecular Dynamics Simulations. *RNA* **2006**, *12*, 1240−1253.

(30) Spackova, N.; Reblova, K.; Sponer, J. Structural Dynamics of the Box C/D RNA Kink-Turn and Its Complex with Proteins: The Role of the A-Minor 0 Interaction, Long-Residency Water Bridges, and Structural Ion-Binding Sites Revealed by Molecular Simulations. *J. Phys. Chem. B* **2010**, *114*, 10581−10593.

(31) Krepl, M.; Reblova, K.; Koca, J.; Sponer, J. Bioinformatics and Molecular Dynamics Simulation Study of L1 Stalk Non-Canonical rRNA Elements: Kink-Turns, Loops, and Tetraloops. *J. Phys. Chem. B* **2013**, *117*, 5540−5555.

(32) Estarellas, C.; Otyepka, M.; Koča, J.; Banáš, P.; Krepl, M.; Šponer, J. Molecular Dynamic Simulations of Protein−RNA Complexes: CRISPR/Csy4 Endoribonuclease. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, DOI: 10.1016/j.bbagen.2014.10.021.

(33) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. Extensive Molecular Dynamics Simulations Showing That Canonical G8 and Protonated A38H(+) Forms Are Most Consistent with Crystal Structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010**, *114*, 6642−6652.

(34) Banáš, P.; Mládek, A.; Otyepka, M.; Zgarbová, M.; Jurečka, P.; Svozil, D.; Lankaš, F.; Šponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *8*, 2448−2460.

(35) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E.; Case, D. A. Simulations of a Protein Crystal with a High Resolution X-ray Structure: Evaluation of Force Fields and Water Models. *J. Phys. Chem. B* **2010**, *114*, 12811−12824.

(36) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.

(37) Li, D.-W.; Brüschweiler, R. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.

(38) Yildirim, I.; Kennedy, S. D.; Stern, H. A.; Hart, J. M.; Kierzek, R.; Turner, D. H. Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iGiC Base Pairs. *J. Chem. Theory Comput.* **2011**, *8*, 172−181.

(39) Stadlbauer, P.; Trantirek, L.; Cheatham, T. E., III; Koca, J.; Sponer, J. Triplex Intermediates in Folding of Human Telomeric Quadruplexes Probed by Microsecond-Scale Molecular Dynamics Simulations. *Biochimie* **2014**, *105C*, 22−35.

(40) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E., Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7*.

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

(41) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(42) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. Dependence of A-RNA Simulations on the Choice of the Force Field and Salt Strength. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701−10711.
(43) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(44) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. Simulations of A-RNA Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.
(45) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankaš, F.; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177−3189.
(46) Sponer, J.; Cang, X. H.; Cheatham, T. E. Molecular Dynamics Simulations of G-DNA and Perspectives on the Simulation of Nucleic Acid Structures. *Methods* **2012**, *57*, 25−39.
(47) Daubner, G. M.; Cléry, A.; Allain, F. H. T. RRM−RNA Recognition: NMR or Crystallography... and New Findings. *Curr. Opin. Struct. Biol.* **2013**, *23*, 100−108.
(48) Chao, J. A.; Lee, J. H.; Chapados, B. R.; Debler, E. W.; Schneemann, A.; Williamson, J. R. Dual Modes of RNA-Silencing Suppression by Flock House Virus Protein B2. *Nat. Struct. Mol. Biol.* **2005**, *12*, 952−957.
(49) Li, H.; Li, W. X.; Ding, S. W. Induction and Suppression of RNA Silencing by an Animal Virus. *Science* **2002**, *296*, 1319−1321.
(50) Lu, R.; Maduro, M.; Li, F.; Li, H. W.; Broitman-Maduro, G.; Li, W. X.; Ding, S. W. Animal Virus Replication and RNAi-Mediated Antiviral Silencing in *Caenorhabditis elegans*. *Nature* **2005**, *436*, 1040−1043.
(51) Tishchenko, S.; Gabdulkhakov, A.; Nevskaya, N.; Sarskikh, A.; Kostareva, O.; Nikonova, E.; Sycheva, A.; Moshkovskii, S.; Garber, M.; Nikonov, S. High-Resolution Crystal Structure of the Isolated Ribosomal L1 Stalk. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 1051−1057.
(52) Trabuco, L. G.; Schreiner, E.; Eargle, J.; Cornish, P.; Ha, T.; Luthey-Schulten, Z.; Schulten, K. The Role of L1 Stalk-tRNA Interaction in the Ribosome Elongation Cycle. *J. Mol. Biol.* **2010**, *402*, 741−760.
(53) Said, B.; Cole, J. R.; Nomura, M. Mutational Analysis of the L1 Binding-Site of 23S Ribosomal-RNA in *Escherichia coli*. *Nucleic Acids Res.* **1988**, *16*, 10529−10545.
(54) Drygin, D.; Zimmermann, R. A. Magnesium Ions Mediate Contacts between Phosphoryl Oxygens at Positions 2122 and 2176 of the 23S rRNA and Ribosomal Protein L1. *RNA* **2000**, *6*, 1714−1726.
(55) Reblova, K.; Sponer, J.; Lankas, F. Structure and Mechanical Properties of the Ribosomal L1 Stalk Three-Way Junction. *Nucleic Acids Res.* **2012**, *40*, 6290−6303.
(56) Schroeder, K. T.; McPhee, S. A.; Ouellet, J.; Lilley, D. M. J. A Structural Database for K-Turn Motifs in RNA. *RNA* **2010**, *16*, 1463−1468.
(57) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Crystal-Structure at 1.92-Å Resolution of the RNA-Binding Domain of the U1A Spliceosomal Protein Complexed with an RNA Hairpin. *Nature* **1994**, *372*, 432−438.
(58) Guallar, V.; Borrelli, K. W. A Binding Mechanism in Protein−Nucleotide Interactions: Implication for U1A RNA Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3954−3959.
(59) Reyes, C. M.; Kollman, P. A. Structure and Thermodynamics of RNA−Protein Binding: Using Molecular Dynamics and Free Energy Analyses to Calculate the Free Energies of Binding and Conformational Change. *J. Mol. Biol.* **2000**, *297*, 1145−1158.
(60) Kurisaki, I.; Takayanagi, M.; Nagaoka, M. Combined Mechanism of Conformational Selection and Induced Fit in U1A−RNA Molecular Recognition. *Biochemistry* **2014**, *53*, 3646−3657.
(61) Aviv, T.; Lin, Z.; Ben-Ari, G.; Smibert, C. A.; Sicheri, F. Sequence-Specific Recognition of RNA Hairpins by the SAM Domain of Vts1p. *Nat. Struct. Mol. Biol.* **2006**, *13*, 168−176.
(62) Aviv, T.; Lin, Z.; Lau, S.; Rendl, L. M.; Sicheri, F.; Smibert, C. A. The RNA-binding SAM Domain of Smaug Defines a New Family of Post-transcriptional Regulators. *Nat. Struct. Mol. Biol.* **2003**, *10*, 614−621.
(63) Gavis, E. R.; Lunsford, L.; Bergsten, S. E.; Lehmann, R. A Conserved 90 Nucleotide Element Mediates Translational Repression of Nanos RNA. *Development* **1996**, *122*, 2791−2800.
(64) Phan, A. T.; Kuryavyi, V.; Darnell, J. C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; et al. Structure−Function Studies of FMRP RGG Peptide Recognition of an RNA Duplex−Quadruplex Junction. *Nat. Struct. Mol. Biol.* **2011**, *18*, 796−804.
(65) Darnell, J. C.; Jensen, K. B.; Jin, P.; Brown, V.; Warren, S. T.; Darnell, R. B. Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function. *Cell* **2001**, *107*, 489−499.
(66) Bassell, G. J.; Warren, S. T. Fragile X Syndrome: Loss of Local mRNA Regulation Alters Synaptic Development and Function. *Neuron* **2008**, *60*, 201−214.
(67) Darnell, J. C.; Fraser, C. E.; Mostovetsky, O.; Darnell, R. B. Discrimination of Common and Unique RNA Binding Activities among Fragile X Mental Retardation Protein Paralogs. *Hum. Mol. Genet.* **2009**, *18*, 3164−3177.
(68) Wang, Y.; Opperman, L.; Wickens, M.; Hall, T. M. T. Structural Basis for Specific Recognition of Multiple mRNA Targets by a PUF Regulatory Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20186−20191.
(69) Wickens, M.; Bernstein, D. S.; Kimble, J.; Parker, R. A PUF Family Portrait: 3′-UTR Regulation as a Way of Life. *Trends Genet.* **2002**, *18*, 150−157.
(70) Suh, N.; Crittenden, S. L.; Goldstrohm, A.; Hook, B.; Thompson, B.; Wickens, M.; Kimble, J. FBF and Its Dual Control of gld-1 Expression in the *Caenorhabditis elegans* Germline. *Genetics* **2009**, *181*, 1249−1260.
(71) Wang, X.; McLachlan, J.; Zamore, P. D.; Hall, T. M. T. Modular Recognition of RNA by a Human Pumilio-Homology Domain. *Cell* **2002**, *110*, 501−512.
(72) Case, D.A.; T. A., D; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M., et al. *AMBER 12*, University of California: San Francisco, 2012.
(73) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(74) Wang, J. M.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential(RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(75) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(76) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(77) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(78) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(79) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243

(80) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(81) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(82) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(83) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(84) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints—Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(85) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(86) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(87) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(88) *The PyMOL Molecular Graphics System*, Version 1.4.1.; Schrodinger, LLC: New York, 2010.
(89) Merritt, E. A.; Bacon, D. J., Raster3D: Photorealistic Molecular Graphics. In *Macromolecular Crystallography, Part B*, Carter, C. W., Sweet, R. M., Eds.; Elsevier Academic Press Inc.: San Diego, 1997; Vol. *277*, pp 505−524.
(90) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217−230.
(91) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, -.
(92) Weichenberger, C. X.; Sippl, M. J. NQ-Flipper: Recognition and Correction of Erroneous Asparagine and Glutamine Side-Chain Rotamers in Protein Structures. *Nucleic Acids Res.* **2007**, *35*, W403−W406.
(93) Zhou, P.; Tian, F.; Lv, F.; Shang, Z. Geometric Characteristics of Hydrogen Bonds Involving Sulfur Atoms in Proteins. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 151−163.
(94) Mongan, J. Interactive Essential Dynamics. *J. Comput. Aided Mol. Des.* **2004**, *18*, 433−436.
(95) Dršata, T.; Pérez, A.; Orozco, M.; Morozov, A. V.; Šponer, J.; Lankaš, F. Structure, Stiffness, and Substates of the Dickerson−Drew Dodecamer. *J. Chem. Theory Comput.* **2013**, *9*, 707−721.
(96) Liu, C.; Janowski, P. A.; Case, D. A. All-Atom Crystal Simulations of DNA and RNA Duplexes. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, DOI: 10.1016/j.bbagen.2014.09.018.
(97) Hall, K. B.; Stump, W. T. Interaction of N-Terminal Domain of U1A Protein with an RNA Stem/Loop. *Nucleic Acids Res.* **1992**, *20*, 4283−4290.

DOI: 10.1021/ct5008108
J. Chem. Theory Comput. 2015, 11, 1220−1243