Molecular Dynamics Simulations of Protein RNA Complexes by Using an Advanced Electrostatic Model

Zhifeng Jing and Pengyu Ren*

Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712

* Corresponding author. Email: pren@mail.utexas.edu

ABSTRACT

Protein-RNA interactions are integral to the biological functions of RNA. It is well recognized that molecular dynamics (MD) simulations of protein-RNA complexes are more challenging than those of each component. The difficulty arises from the strong electrostatic interactions and the delicate balance between various types of physical forces at the interface. Previously MD simulations of protein-RNA complexes have predominantly employed fixed-charge force fields. Although force field modifications have been developed to address problems identified in the simulations, some protein-RNA structures are still hard to reproduce by simulations. Here we present MD simulations of two representative protein-RNA complexes using the AMOEBA polarizable force field. The van der Waals parameters were refined to reproduce accurate quantum-mechanical data of base-base and base-amino acid interactions. It was found that the refined parameters produced a more stable hydrogen-bond network in the interface. One of the complexes remained stable during the short simulations, whereas it could quickly break down in previous simulations using fixed-charge force fields. There were reversible breaking and formation of hydrogen bonds that are observed in the crystal structure, which may indicate the difference in solution and crystal structures. While further improvement and validation of the force fields are still needed, this work demonstrates that polarizable force fields are promising for the study of protein-RNA complexes.

1. INTRODUCTION

RNA is essential for many biological activities including gene expression and regulation and catalysis. RNA also has tremendous therapeutic potential for gene therapy, immunotherapy, and vaccines.[1-3] Most of biologically active RNAs interact with proteins. The interactions with proteins are crucial for the folding, stability, regulation, transport and localization of RNAs.[4-6] Therefore, understanding of the principles of RNA-protein interactions is useful for both fundamental research and practical applications. The mechanisms of protein-RNA recognition are highly diverse. Various types of interactions can be found at the protein-RNA interface, such as electrostatic interactions, hydrogen-bonding, salt bridge, pi-pi interactions and hydrophobic interactions.[7, 8] In protein-RNA complexes, proteins either exclusively interact with RNA backbone or have specific interactions with nucleobases. In addition, protein-RNA interactions often involve large conformation changes of RNAs.

Experimental techniques including X-ray crystallography and nuclear magnetic resonance (NMR) have been valuable for the study of protein-RNA complexes. The data generated by such experiments have enabled the development of software tools for the prediction of protein-RNA binding. Structural analyses have shown van der Waals (vdW) interactions are more prevalent than hydrogen bonds,[9] and interactions with the RNA backbone are more frequent than interactions with nucleobases.[10] There are also limitations in experimental studies of protein-RNA complexes. The biomolecules in solution may have more than one conformation, and they may be different from the crystal structure. NMR experiments are usually conducted in solution, but they are more ambiguous to interpret than X-ray crystallography.[11]

Molecular dynamics (MD) simulations have been widely used to study biomolecular structures and dynamics as they complement experiments by providing detailed pictures of the molecules and connecting microscopic structures with macroscopic observations. Some notable examples include protein structure prediction[12] and protein-ligand binding.[13, 14] However, MD simulations of protein-RNA complexes are still challenging. The simulation results are very sensitive to force field parameters[15] because of the conformational flexibility of RNAs and the

intricate network of interactions in the protein-RNA interface. There have been a lot of progress in nucleic acid force fields. Two major families of force fields for proteins and nucleic acids are AMBER and CHARMM.[15-17] AMBER force fields were first developed in the 1990s[18] and have been continuously refined.[19-23] A major focus has been on torsion parameters because the conformations of nucleic acids are sensitive to torsion parameters and the non-bonded parameters were found to be sufficient for describing gas-phase interaction energies and hydration free energies.[19] Recently, there have been more studies on refining non-bonded parameters to improve intramolecular hydrogen bonds, base stacking, protein-nucleic acid binding and other thermodynamics properties.[24-32] The revised non-bonded parameter for phosphate of Steinbrecher et al.[33] in combination with the OPC water model could alleviate the excessive binding excessive binding between 2'-hydroxyl groups and phosphate, while it has issues for other systems.[25, 34] Šponer and coworkers proposed hydrogen-bond terms (HBfix and gHBfix)[23-25] to selectively fine-tune the non-bonded interactions, which leads to general improvement for RNA simulations. Shaw and coworkers developed revised versions of AMBER force fields that include new parameters for nucleobase charge and vdW to improve base pairing and stacking interactions and the phosphate vdW parameters of Steinbrecher et al.,[33] which necessitates adjustments of torsion parameters. The parameters were used with the TIP4P-D water model developed by the same group.[27] Subsequently the phosphate charges and vdW parameters were modified to be compatible with the DES-Amber protein force field which includes rescaling of ionic charges.[35, 36] CHARMM force fields[37, 38] have also been improved over the years.[17] The latest version CHARMM36[37] improved the stability of base pairs by tuning 2'-OH torsion parameters. Common problems for standard fixed-charge force fields such as AMBER and CHARMM are the over-stabilization of protein-nucleic acid interactions and the difficulty in modeling ion-specific effects.[29] These problems have been partially addressed by special vdW parameters.[29, 30, 39]

Polarizable force fields provide more accurate description of electrostatics and balanced interactions in heterogenous environments, so they are promising in addressing the problems of fixed-charge force field for protein-RNA complexes.[40-42] CHARMM-Drude force field[43-48] and

AMOEBA[49-52] are two popular polarizable force fields. AMOEBA showed good accuracy for various nucleic acid systems, including DNA duplexes in different environment, RNA tetranucleotide, RNA tetraloops, RNA duplexes, and HIV-TAR.[49] There have been very few published studies on MD simulations of protein-RNA interfaces using polarizable force fields, which is likely due to the computational cost of polarizable force field and the difficulty in conformational sampling.

In this work, the vdW parameters in AMOEBA nucleic acids force field were refined by using high-quality quantum-mechanical data. MD simulations for representative protein-RNA systems were performed with both the original and the refined AMOEBA force field parameters.

2. METHODS

2.1 Molecular dynamics simulations

Two protein-RNA complex systems were selected for molecular dynamics simulations: U1A protein bound to RNA hairpin[53] (PDB code: 1urn) and *C. elegans Pumilio* FBF-2 protein bound to *gld-1* FBEa RNA[54] (PDB code: 3k5y). The PDB structure of 1urn contains three copies of the complex. The complex formed by chain B and chain Q was selected for simulations since it is the only one without missing residues. Both systems have been studied by Krepl et al,[5] and more recently by Tucker et al.[36] The crystal structure of U1A contains a 21-nucleotide RNA sequence. The FBF protein is responsible for the regulation of messenger RNA.[5] The FBF interface is composed of interaction between nucleobases and protein, while the RNA backbone is exposed to the solvent.[5]

The systems were prepared by using the tleap program in AmberTools.[55] All systems were neutralized and solvated in isometric truncated octahedron boxes with a minimum distance between solute and box edge of 15 Å. KCl ions were added to give a concentration of 150 mM. The solvated systems were minimized, gradually heated up from 50 K to 300 K in 500 ps NVT simulations with harmonic position restraint on protein and RNA, and equilibrated at the NPT ensemble with 300 K and 1 bar for 4 ns, and then the force constant of the position restraints was

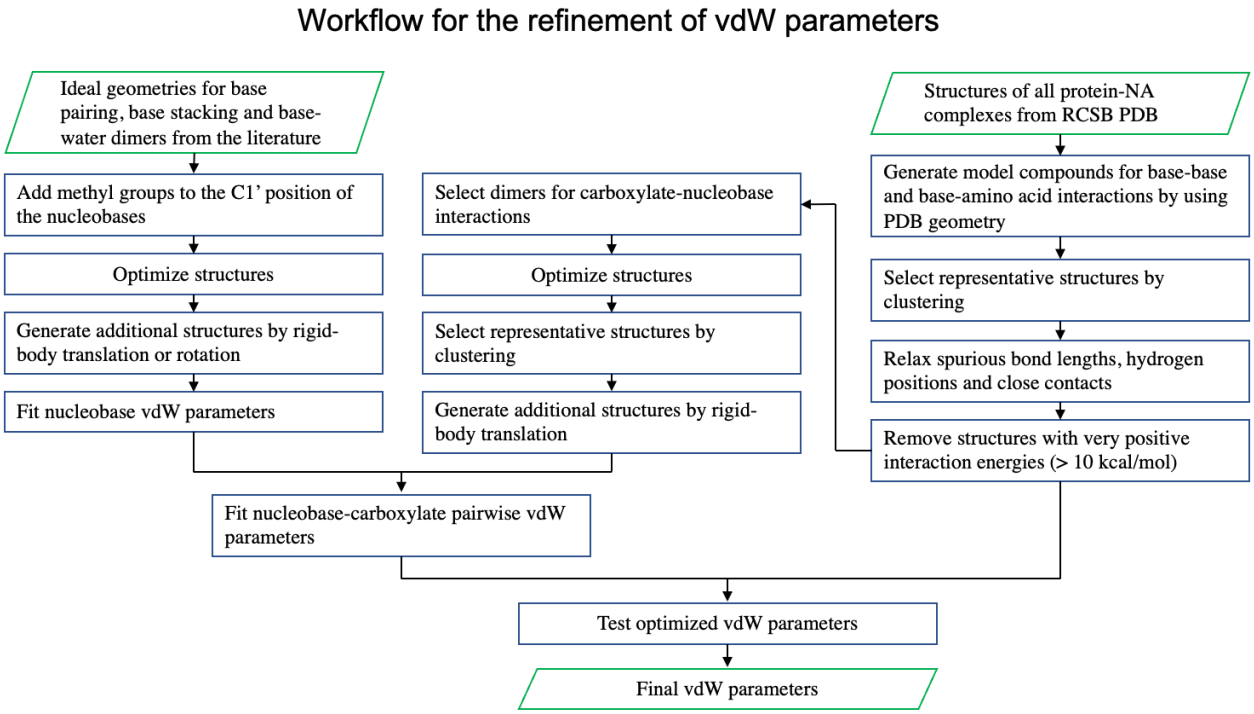
gradually reduced from 25 kcal/mol/Å^2 to zero in 6 ns NVT simulations. The production simulations were NVT simulations at 300 K. The RESPA integrator[56] was used with an outer time step of 2.0 fs. Temperature and pressure were maintained by Bussi thermostat[57] and Monte Carlo barostat,[58] respectively. The electrostatics was treated by PME with a real-space cutoff of 7.0 Å. The van der Waals (vdW) was truncated at 12.0 Å. Tinker-OpenMM program[59] was used for the simulations. Proteins and nucleic acids were modeled by the original amoebabio18 parameter and refined parameters as detailed below. Water and ions were modeled by the default parameters as described by Ren et al.[60] and Wang.[61]

2.2 Quantum mechanical calculations

Quantum mechanical (QM) calculations were used to generate reference data to refine the non-bonded interaction parameters for RNA and torsion parameters for protein. The choice of QM methods for force field parametrization significantly affects the performance of the resulting force field. Usually more accurate QM methods lead to better force fields (e.g. the DESRES RNA force field[27]), although due to error cancelation, lower-level QM methods could be better choices in some cases (e.g. AMBER RNA force fields parmbsc0,[22] OL3,[62] OL15[21]). Since AMOEBA incorporated more physical terms including polarization and atomic multipoles, it is less prone to error cancelation than non-polarizable force fields, and the parameterization of AMOEBA has relied on high-level QM methods such as MP2 with relatively large basis sets. MP2 with cc-pVTZ or aug-cc-pVTZ was used in the parameterization of the AMOEBA RNA force field. However, recent benchmark studies have found that MP2 suffers from large errors for dispersion interactions,[63, 64] which are prominent in base-base stacking. Various MP2-based method and DFT methods have been developed to address this issue. Based on results from previous work,[63, 65] DSD-BLYP-D3BJ, a double-hybrid DFT method with empirical dispersion correction, was chosen for QM calculations due to its excellent accuracy and computational efficiency.

2.3 Optimization of vdW parameters

Nucleobase vdW parameters



Scheme 1. Workflow for the refinement of nucleobase and nucleobase-carboxylate vdW parameters.

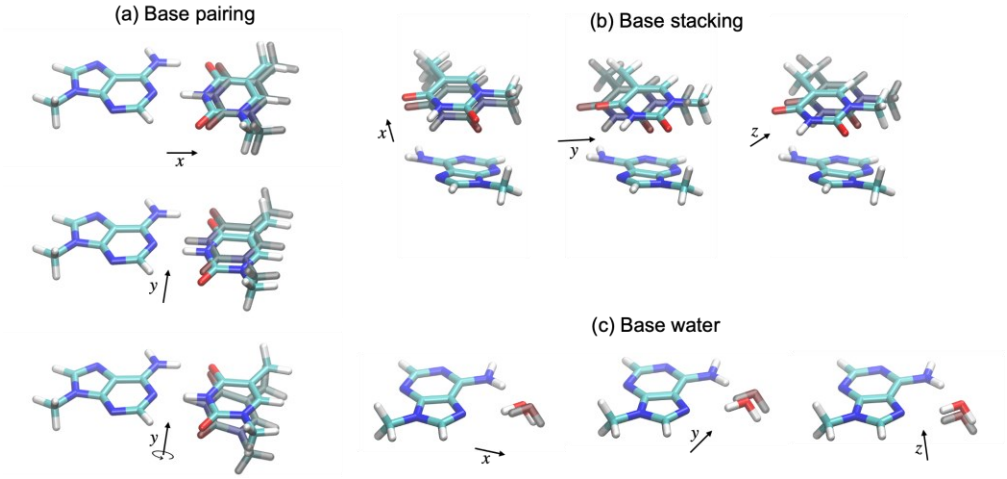


Figure 1. Illustration of additional structures for model compounds based on ideal geometry. (a) for base pairing, $\Delta x = \pm 0.20, \pm 0.10, \pm 0.05\ r_0$, $\Delta y = \pm 0.4$ Å, and $\Delta\theta_y = \pm 20°$, where r_0 is the equilibrium distance; (b) for base stacking, $\Delta x = \pm 0.20, \pm 0.10\ r_0$, $\Delta y = \Delta z = \pm 0.6, \pm 0.3$ Å; (c) for base water dimer, $\Delta x = \pm 0.20, \pm 0.10, \pm 0.05\ r_0$, $\Delta y = \Delta z = \pm 0.6, \pm 0.3$ Å.

Model compounds based on ideal geometries were generated for fitting nucleobase vdW parameters, and models based on PDB geometries were used for testing. The workflow is demonstrated in Scheme 1. The model compounds include model compounds including methylated nucleobases (A, T, U, G and C), N-methyl amide (NMA), ethanol (EtOH), acetate, methylammonium (MeNH3), imidazole and benzene. All interaction energies were calculated at the DSD-BLYP-D3BJ/def2-QZVP level of theory. The ideal geometries include canonical and noncanonical base pairing, base-base stacking and base-water dimers. The starting structures were taken from our previous work[50] with the addition of A-G cis-Watson-Crick/sugar-edge pair[66] from the JSCH benchmark database for non-covalent interactions (designated as "G…A 4").[67] The geometries were re-optimized by using TPSS-D3BJ/def2-TZVP. For each ideal geometry, additional structures were generated by translation or rotation, as shown in Figure 1. The *x*-axis was defined by the closest atom pair; the *y*-axis represents the principal axis of the interface, which was chosen manually for base pairing and defined by three atoms in each monomer that are closest to the interface; the *z*-axis is orthogonal to the *xy*-plane.

The PDB geometries include base-base and base-amino acid interactions. The procedure for selecting representative structures is shown in Scheme 1. A hierarchical clustering algorithm with average link and a RMSD cutoff distance of 2.5 Å was used in the clustering. Spurious bond lengths, as defined by AMOEBA bond energy greater than 10 kcal/mol, were relaxed by optimization using AMOEBA with position restraint. Short intermolecular distances (< 1.8 Å) were relaxed by rigid-body distance scan with TPSS-D3BJ/def2-TZVP/PCM. Hydrogen positions were optimized by TPSS-D3BJ/def2-TZVP/PCM or PM6/PCM. Starting from 5334 PDB structures of protein-NA complexes, 657 structures for 45 types of dimers were generated.

The interaction energies based on the ideal geometries were used to refine the nucleobase vdW parameters, while the PDB geometries were used for validation. The data points were assigned Boltzmann weights at 1000 K relative to the corresponding lowest energy points. The Trust Region Reflective algorithm in SciPy with bounds and quadratic restraint was used for the

parameter fitting. The bounds were ±5% for vdW diameter and ±20% for well depth within the initial parameters. The restraint was chosen based on 5-fold cross validation. Two sets of optimized vdW parameters were obtained: param1, in which the atom types are the same as in amoebabio18, and param2, in which Guanine does not share parameters with other nucleobases. The RMSEs of these parameters as well as the original amoebabio18 are summarized in Table 1. A breakdown of the mean error and root mean squared error (RMSE) to different types of base pairs is shown in Figure 2 and Figure 3. For both base-pairing and base-stacking interactions, amoebabio18 has generally positive errors. This is in part due to the deficiency of the MP2 method which was used in the parameterization of amoebabio18. The mean errors are significantly reduced in the param1 and param2 results. The improvement in RMSEs can be found in most types of base pairing and only a few types of base stacking (namely CU and GG), since the RMSEs of base stacking energy were relatively small. For param1, base pairs involving guanine have more positive errors than other base pairs, which was a motivation for the additional atom types for guanine in param2. Indeed, param2 has overall better accuracy than param1 for the training set (Table 1).

The two sets of parameters were evaluated on the interaction energy from the PDB geometries (Table 1), which was not used in the parameter fitting. The accuracy for base-base interactions were significantly improved in both sets of parameters. The accuracy for base-amino acid interactions were also improved, although the training set contains no amino acids. Contrary to the performance for the training set, param2 has slightly larger errors than param1 for the test set, despite having more adjustable parameters. The combination of small training error and large test error is characteristic of overfitting. Therefore, param1 was chosen for simulations.

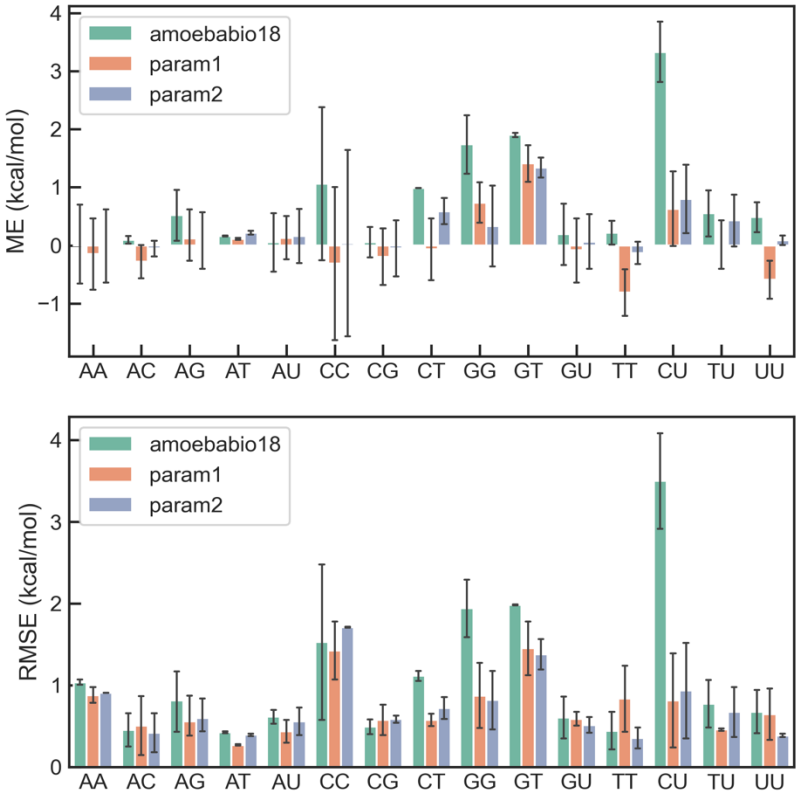


Figure 2. Distribution of errors of AMOEBA force fields for base stacking interaction in each type of base-stacking pairs.

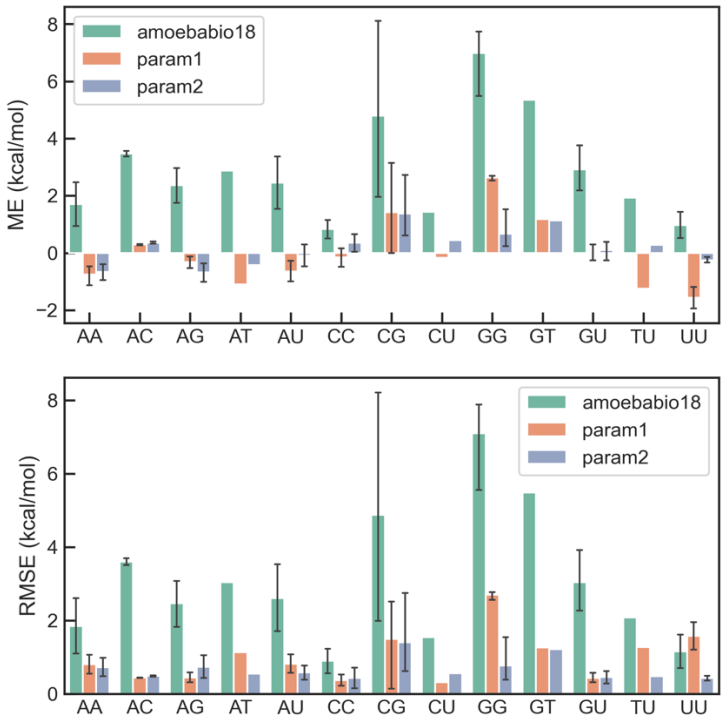


Figure 3. Distribution of errors of AMOEBA force fields for base pairing interaction in each type of base pairs.

Table 1: Weighted RMSE of AMOEBA parameters on base-base and base-amino acids interactions (kcal/mol).

Model	Training set		Test set	
	Base stacking	Base pairing	Base	Base-Amino Acid
Amoebabio18	1.466	3.775	1.895	1.309
Param1	0.860	1.317	0.829	1.193
Param2	0.887	1.033	0.856	1.194

The optimized vdW parameters are listed in Table 2. The changes in parameters are less than 5% from amoebabio18. The vdW radii are smaller in all optimized parameters than in amoebabio18, since the base-pair interactions in amoebio18 are too weak. In param2, the vdW radii of guanine are also smaller than those of the corresponding atoms in other nucleobases.

Table 2: Nucleobase vdW sigma (Å) and epsilon (kcal/mol) in amoebabio18 and optimized parameters.

Vdw type	Atom	amoebabio18	Param1	Param2
54	U/T N3, C N1	3.700 0.127	3.600 0.124	3.700 0.127
(154)	G N1			3.560 0.127
50	A N3/N1, C N3	3.640 0.127	3.530 0.124	3.560 0.120
(150)	G N3			3.500 0.125
49	A N7	3.640 0.127	3.530 0.124	3.560 0.120
(149)	G N7			3.500 0.125
58	C O2, U/T O2/O4	3.350 0.129	3.250 0.124	3.250 0.124
(158)	G O6,			3.200 0.120
53	A H6, C H4	2.650 0.020	2.620 0.018	2.650 0.020
(153)	G H2			

Nucleobase-protein vdW parameters

The hydrogen bonds between Asp/Glu side chains and nucleobases are a common motif in protein-nucleic acid interfaces. It was found that even with the optimized nucleobase vdW parameters, some of these hydrogen bond interactions still have relatively large errors. This may be attributed to the partial covalent bonding nature of these interactions. As in our previous work on phosphate-binding proteins,[68] pairwise vdW parameters between carboxylate O and NH were included to improve the accuracy for these interactions. The training set structures were obtained by geometry optimization (TPSS-D3BJ/def2-TZVP) of model compounds from the PDB geometries (Scheme 1). For each optimized structure, two additional structures were generated by rigid-body translation, resulting in intermolecular distances of 0.9 and 1.1 times the equilibration distance. The performance for the training set and the test set are shown in Figure 4.

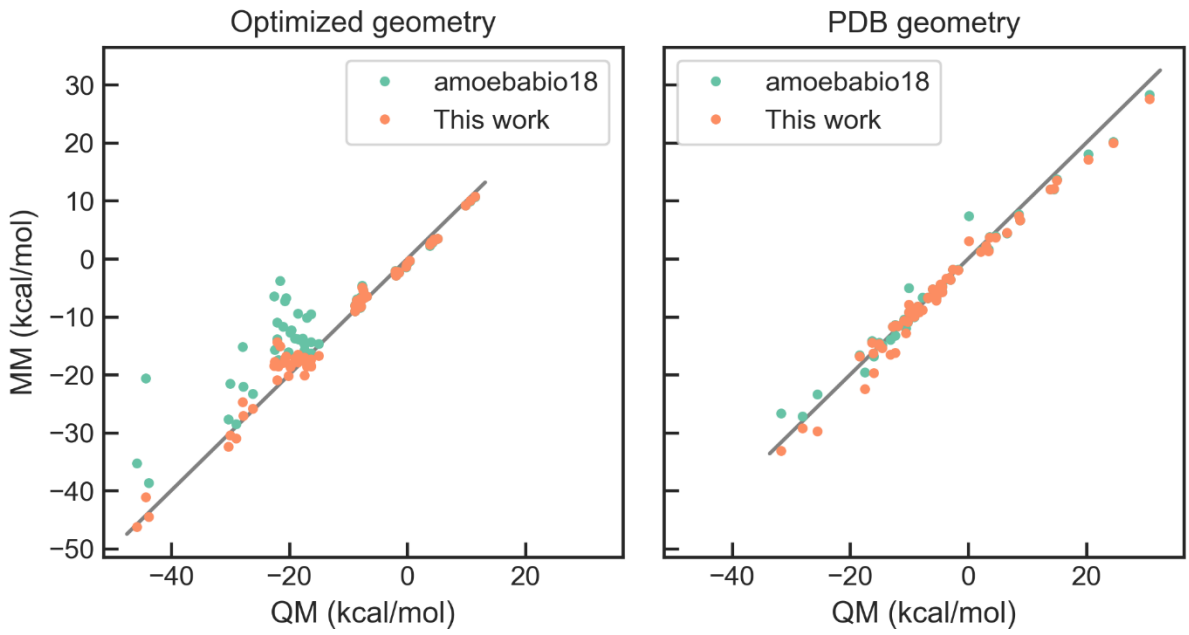


Figure 4. Correlation between QM and MM interaction energies for nucleobase-carboxylate interactions.

2.4 Optimization of Asn torsion parameters

During initial MD simulations of the U1A protein-RNA interface using amoebabio18, it was found that the rotamer conformation of some Asn residues in the interface was different from those in the crystal structures, which disrupted the hydrogen-bond network. Therefore, the torsion parameters of Asn were revisited. Conformations were generated by (1) one-dimensional torsion scan of Asn χ1 and χ2 torsions at 30 deg interval, with backbone torsion fixed at either alpha-helix or beta-sheet conformation, which leads to 48 conformations in total, and (2) 3-by-3 two-dimensional scan of χ1 and χ2 torsions at -30, 0 and 30 deg based on each of the 11 rotamers from the backbone dependent rotamer library,[69] for a total of 99 conformations. All the conformations were optimized at the ωB97Xd/6-311++g(2d,2p) level of theory with backbone and side chain torsions frozen, and the energies were calculated by DSD-BLYP-D3BJ/def2-QZVP. The Trust Region Reflective algorithm was used in the least-square fitting. The RMSE was reduced from 2.06 kcal/mol for amoebabio18 to 1.60 kcal/mol for the optimized parameters (param1). The

improvement is more noticeable when tested on the conformations sampled from the initial MD simulations of U1A. As shown in Figure 5, amoebabio18 has large errors in the relative energies of these conformations; specifically, it predicted similar energies for the low-energy and high-energy conformers, while the two conformers differ by about 4 kcal/mol according to QM (DSD-B3LYP-D3BJ/def2-QZVP). The optimized parameters could correctly rank the low-energy and high-energy conformers. The side chain torsion parameters of the original AMOEBA protein force field were fitted to MP2/CBS energy of one-dimensional torsion, with backbone fixed at alpha-helix and beta-sheet conformations.[70] Both MP2 and DSD-B3LYP-D3BJ are very accurate for conformation energies.[63, 65] The main difference between this work and the original parameterization is in the conformers. The training set structures were generated by using a general procedure rather than tailored to the issue of U1A simulations, so the optimized parameters are expected to be a general improvement over the original parameters.

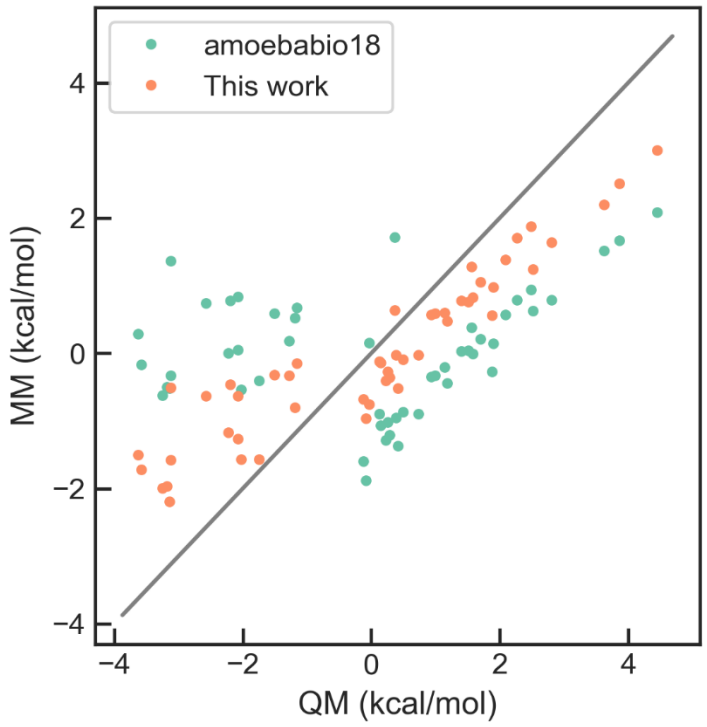


Figure 5. Correlation between QM and MM relative energies for the Asn conformers sampled from MD simulations of U1A.

3. Results and Discussions

3.1 U1A simulations

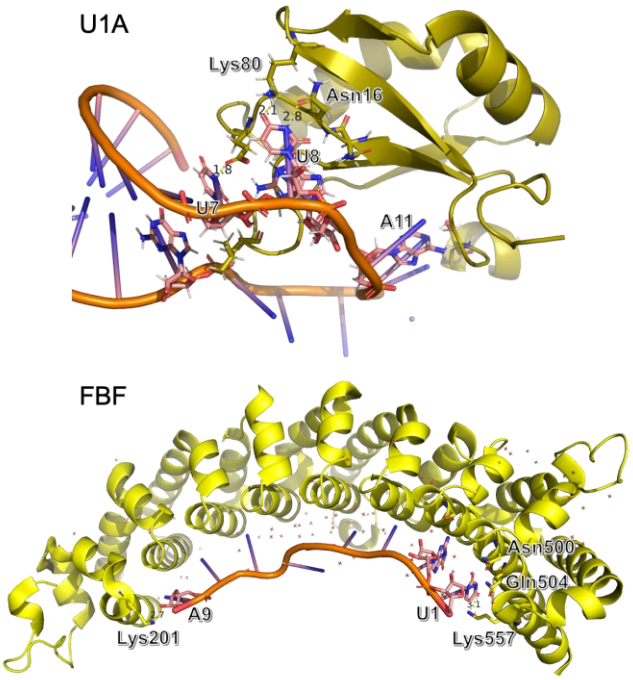


Figure 6. Crystal structure of the U1A (PDB code: 1urn) and FBF (PDB code: 3k5y) protein-RNA complexes.

The U1A protein-RNA complex has been well studied by experiments and MD simulations.[53, 71-73] Its interface consists of two parts, one between the stem region of the RNA hairpin (including U7, U8, G9 and G16) and two loops of the protein, and the other between the loop region of the RNA (including C10, A11 and C12) and a loop of the protein. The stability of the protein-RNA interface during the simulations is measured through the distances between several hydrogen-bond donor and acceptors (Figure 7), and the hydrogen bond is considered to be maintained if the distance is below 4 Å. During the simulations using the amoebabio18 parameters, the hydrogen bond between U8(N3) and Asn16(OD1) was quickly lost within 20 ns in four out of five simulations. This is accompanied by the rotation of the Asn16 χ1 torsion and the loss of the

hydrogen bond between U8(O4) and Lys80. This can be clearly seen in Figure 7 where the average distance for the two hydrogen bonds in the last 20 ns of 60 ns simulations are around 6 Å. The two hydrogen bonds with U8 could reform briefly and then break multiple times during extended 100-ns simulations (Figure 8), which indicates that there is only a small free energy difference between the native structure and the disrupted structures. Among the conformations with disrupted hydrogen bond between U8(O4) and Lys80(NZ), there are water-mediated interactions when the distance is between 4 and 5 Å. The other part of the protein-RNA interface represented by the hydrogen bond between A11(N1) and Ser91(OG) remained relatively stable in the simulations.

Preliminary simulations using the optimized vdW parameters suffered from similar issues for the hydrogen bonds around Asn16 (see Supporting Information). The stability of the interface was significantly improved after the introduction of the optimized Asn torsion parameters. The two hydrogen bonds with U8 were well maintained in all five simulations (Figure 7). The hydrogen bond between A11(N1) and Ser91(OG) was also stable in the majority of simulations, although there was a larger variation between simulations compared to the simulations with amoebabio18. The variations in the hydrogen bonds with A11 was largely due to the conformations of the C-terminal residues. The C-terminal of the U1A protein consists of a short five-residue alpha-helix, followed by unstructured residues Met97 and Lys98. Short alpha-helices are generally not very stable in aqueous solution. In addition, the last residue Lys98 is missing in the crystal structure, and is not included in this work and previous work.[74] The C-cap residue could affect the electrostatic interactions with RNA.

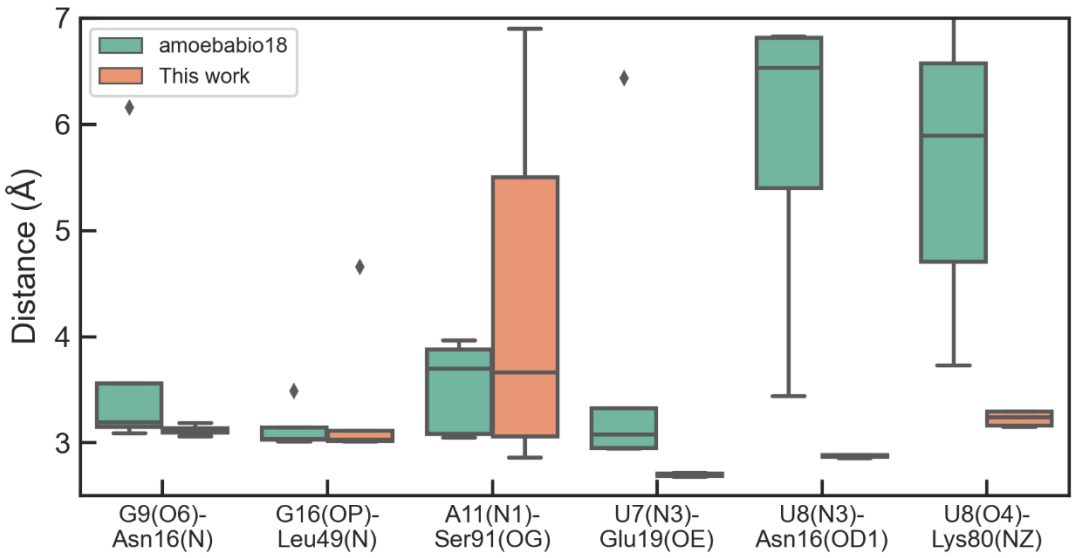


Figure 7. Average distances between selected hydrogen-bonds donors and acceptors in the U1A complex during MD simulations. The average is calculated over the last 20 ns of each of the five 60-ns simulations.

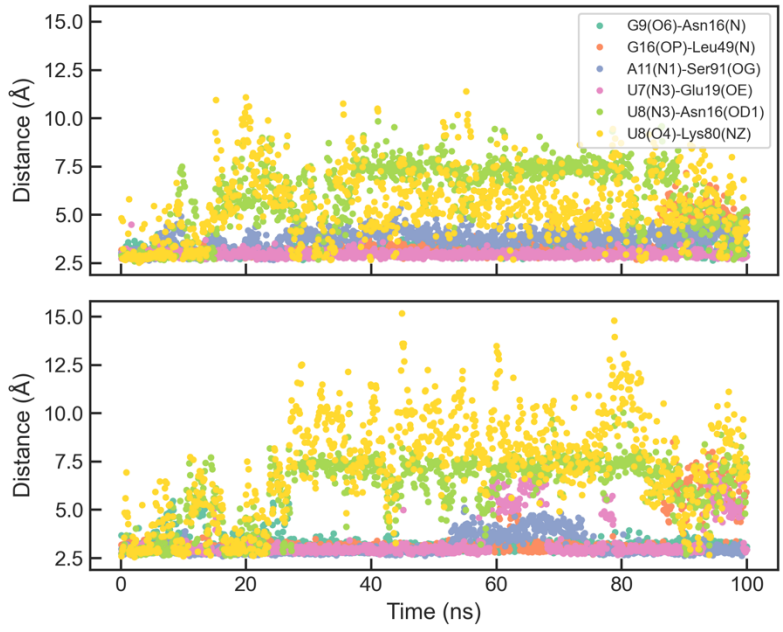


Figure 8. Evolution of selected hydrogen-bond distances in two extended simulations of U1A with the original AMOEBA parameters.

Previous simulations had difficulty in keeping all the native hydrogen bonds in the U1A interface. Simulations with AMBER ff99bsc0χOL3 for RNA and ff99SB or ff12SB for protein led to an altered hydrogen-bond network.[5] Simulations with ff99SB did not maintain several hydrogen bonds with U7, U8, G9 and G16. Simulations with ff12SB which includes a reparameterization of side-chain torsion parameters did not bring significant differences in the simulations. It is worth noting that four simulations with ff12SB behaved similarly as those with ff99SB; in two other simulations, the disruption of hydrogen bonds with U8, G9 and G16 were reversible and the A11(N1)-Ser91(OG) bond was disrupted.[5] Coincidently, the two ff12SB simulations were similar to the AMOEBA param1 simulations (Figure 7), except that U7(N3)-Glu19(OE) bond was very stable in AMOEBA simulations. Šponer and coworkers identified some deficiencies of the AMBER force fields through QM/MM calculations, especially for the solvent-exposed hydrogen bonds G9(N7)-Asn15(ND) and U8(O4)-Lys80(NZ).[74] A structure-specific force field term for the hydrogen bonds (HBfix) could help stabilize the interface.[24] This idea was further explored for simulations of other RNA systems.[23, 25] In the simulations using AMBER ff12SB χOL3 with HBfix, the overall interface remained stable during 1000-ns simulations, and there were reversible breaking of G9(N7)-Asn15(ND), U7(N3)-Glu19(OE) and U8(O4)-Lys80(NZ), which were also observed in some of the AMOEBA simulations.

Tucker et al. reported 50-microsecond simulations of several protein-nucleic acid complexes including U1A and FBF, using modified versions of AMBER force field.[36] It was found that a reparameterization of nonbonded parameters for the nucleic acid backbone (mainly 2'-hydroxyl and phosphate oxygen) based on experimental data substantially improved the simulations of some protein-nucleic acid complexes. However, the re-parameterized force field (DES-Amber 3.20) and their previous force field (DES-Amber) have similar results for U1A. The overall RMSD went up to 4-5 Å after a few microseconds, indicating that some hydrogen bonds were possibly disrupted at a very long timescale.

Electrostatic interactions also play roles in the association and stability of the U1A complex. It was shown by mutagenesis experiments that while Lys20, Lys22 and Lys23 all affect

the association rate, Lys20 has minimal effect on the complex stability.[73] In the crystal structure, Lys22 has a salt bridge with the backbone phosphate group, and Lys20 and Lys23 have no direct contact with the RNA. The interactions of these Lys residues during the AMOEBA simulations are generally consistent with the crystal structure: Lys20 had transient salt bridges with the phosphate group of U3 or G4, and Lys23 remained away from the RNA. In a previous simulation with AMBER ff99 force field, Lys20 and Lys23 formed stable salt bridges with phosphate groups.[75] So the fixed-charge force field seems to overestimate the electrostatic interactions between Lys and phosphate, while AMOEBA has more balanced electrostatic interactions.

3.2 FBF simulations

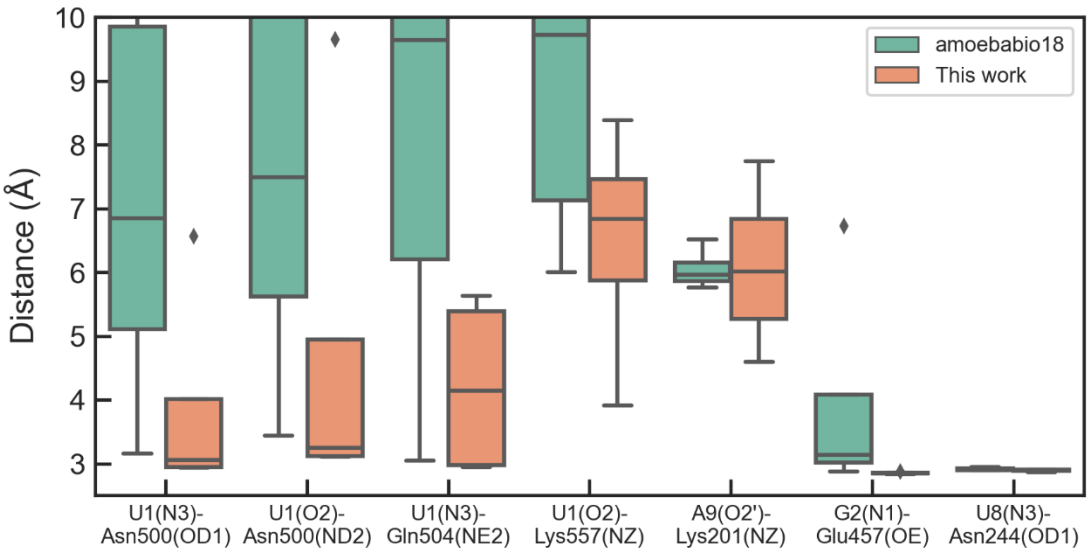


Figure 9. Average distances between selected hydrogen-bonds donors and acceptors in the protein-RNA interface during MD simulations of FBF. The average is calculated over the last 40 ns of each of the four 120-ns simulations.

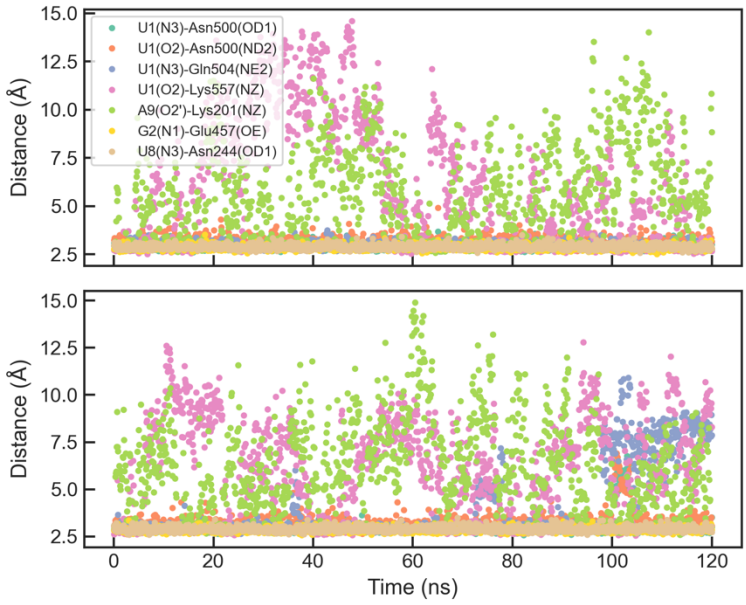


Figure 10. Evolution of selected hydrogen-bond distances in two representative simulations of FBF with the optimized parameters.

The FBF protein-RNA complex has an elongated interface composed of interactions between nucleobases and protein, while the RNA backbone is exposed to the solvent (Figure 6). The distances between hydrogen bond donors and acceptors at the two ends (U1, G2 and A9) during the simulations are shown in Figure 9. The optimized AMOEBA parameter has a significant improvement over amoebabio18 for the hydrogen bonds at the 5'-terminal of the RNA strand, including U1(N3)-Asn500(OD1), U1(O2)-Asn500(ND2), U1(O4)-Gln504(NE2), U1(O4)-Lys557(NZ) and G2(N2)-Ser453(OG). Even with the optimized parameters, there were rapid reversible breaking of the terminal hydrogen bonds, particularly U1(O4)-Lys557(NZ) and A9(O2')-Lys201(NZ) (Figure 10). Both hydrogen bonds are solvent-exposed and associated with flexible Lys side-chains, so their existence in the crystal structure might be due to the crystal packing effect. Apart from the two hydrogen bonds with Lys, other hydrogen bonds were fairly stable during simulations with the optimized parameters. For both the original and optimized parameters, the 5'-terminal (U1) interface is less stable than the 3'-terminal (A9) interface. An

explanation is that the A9(O2')-Lys201(NZ) bond is stronger than the U1(O4)-Lys557(NZ) bond due to the electrostatic interaction between Lys201 and the RNA backbone, and the stacking between A9 and Tyr245 is also stronger than the stacking between U1 and Tyr501.

Previous simulations with AMBER ff99bsc0χOL3 for RNA and ff99SB or ff12SB for protein could not stabilize the interface.[5] The terminal hydrogen bonds with U1 and A9 were lost within 50 ns, followed by the loss of hydrogen bonds with G2 and U8 after 300 ns. The interruption of the interface was mainly caused by the conformational changes of the RNA.[5] Furthermore, the interface could not be stabilized by simulations with distance restraints on hydrogen bonds with U1 and U8. Based on these results, Krepl et al.[5] concluded that crystal packing might play a role in the stabilization of the structure. In the simulations with DES-Amber and DES-Amber 3.20 force fields, the FBF complex was stable throughout 50-microsecond simulations. The backbone RMSDs were between 2-3 Å and there was no sign of dissociation. Since the optimized AMOEBA parameters and DES-Amber 3.20 achieved higher stability of the protein-RNA interface compared to previous versions of force fields, the crystal structure of the FBF complex could be either stable or metastable in solution.

4. CONCLUSIONS

The AMOEBA RNA force field was refined to improve the accuracy for modeling protein-RNA interfaces. Accurate DFT methods were chosen as a reference for the intermolecular interactions between nucleobases, water and small organic molecules. The nucleobase parameters were optimized by fitting to a large QM dataset including both optimized and non-ideal geometries and validated by cross-validation and separate test sets. Special vdW pair parameters were included to better represent hydrogen-bond interactions between nucleobases and carboxylate groups. The Asn torsion parameters were optimized based on conformations from torsional scan and rotamer library. The optimized AMOEBA parameters achieved better accuracy for the problematic protein-RNA interfaces identified in previous work. Since the reparameterization was based on a large QM dataset, the optimized parameters should represent a general improvement in

describing protein-RNA interactions. Nevertheless, further tests on more systems are needed. Adjustment in torsion parameters might also be necessary to be compatible with the new vdW parameters.

Simulations with the optimized parameters correctly reproduced the important hydrogen bonds in the U1A protein-RNA interface, which was recently achieved by using AMBER force field with a structure-specific force field term. The AMOEBA simulations are also more realistic for the interactions between Lys side-chains and backbone phosphate groups. For the FBF protein-RNA complex, the AMOEBA simulations produced a stable binding interface except for reversible disruptions of some solvent-exposed hydrogen bonds between terminal nucleotides and Lys side-chains. Overall, the AMOEBA force field is promising for modeling the complex hydrogen-bond network and electrostatic interactions in protein-RNA interfaces.

5. Associated Content

Additional text for the simulation methods for DNA and RNA systems; Summary of model compounds with PDB geometry (Table S1); comparison of original and optimized parameters for DNA and RNA systems (Table S2); Correlation of QM and MM energies in Asn torsion fitting (Figure S1); Evolution of selected hydrogen-bond distances in simulations of U1A and FBF with different parameters (Figure S2-4); parameter file in Tinker format.

6. Acknowledgements

This work was supported by the National Institutes of Health (Nos. R01GM106137 and R01GM114237), National Science Foundation (CHE-1856173), and CPRIT (RP160657). The authors thank Prof. Jiří Šponer for suggestions of the simulation systems.

7. References

(1) Smith, C. I. E.; Zain, R. Therapeutic Oligonucleotides: State of the Art. *Annual Review of Pharmacology and Toxicology* **2019**, *59* (1), 605-630. DOI: 10.1146/annurev-pharmtox-010818-021050.
(2) Sahin, U.; Karikó, K.; Türeci, Ö. mRNA-based therapeutics — developing a new class of drugs. *Nature Reviews Drug Discovery* **2014**, *13* (10), 759-780. DOI: 10.1038/nrd4278.
(3) Jing, Z.; Qi, R.; Thibonnier, M.; Ren, P. Molecular Dynamics Study of the Hybridization between RNA and Modified Oligonucleotides. *J. Chem. Theory Comput.* **2019**, *15* (11), 6422-6432. DOI: 10.1021/acs.jctc.9b00519.
(4) Dreyfuss, G.; Kim, V. N.; Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular Cell Biology* **2002**, *3* (3), 195-205. DOI: 10.1038/nrm760.
(5) Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Sponer, J. Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein–RNA Complexes? *J. Chem. Theory Comput.* **2015**, *11* (3), 1220-1243. DOI: 10.1021/ct5008108.
(6) Lunde, B. M.; Moore, C.; Varani, G. RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology* **2007**, *8* (6), 479-490. DOI: 10.1038/nrm2178.
(7) 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 Research* **2012**, *40* (8), 3299-3306. DOI: 10.1093/nar/gkr1225.
(8) Draper, D. E. Themes in RNA-protein recognition. *J. Mol. Biol.* **1999**, *293* (2), 255-270. DOI: 10.1006/jmbi.1999.2991.
(9) Jones, S.; Daley, D. T. A.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. Protein–RNA interactions: a structural analysis. *Nucleic Acids Research* **2001**, *29* (4), 943-954. DOI: 10.1093/nar/29.4.943.
(10) Ellis, J. J.; Broom, M.; Jones, S. Protein–RNA interactions: Structural analysis and functional classes. *Proteins* **2007**, *66* (4), 903-911, https://doi.org/10.1002/prot.21211. DOI: 10.1002/prot.21211.
(11) Daubner, G. M.; Cléry, A.; Allain, F. H. T. RRM–RNA recognition: NMR or crystallography…and new findings. *Current Opinion in Structural Biology* **2013**, *23* (1), 100-108. DOI: https://doi.org/10.1016/j.sbi.2012.11.006.
(12) Perez, A.; Morrone, J. A.; Brini, E.; MacCallum, J. L.; Dill, K. A. Blind protein structure prediction using accelerated free-energy simulations. *Science Advances* **2016**, *2* (11), e1601274. DOI: doi:10.1126/sciadv.1601274.
(13) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **2015**, *137* (7), 2695-2703. DOI: 10.1021/ja512751q.
(14) Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.; Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N.; et al. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Central Science* **2021**, *7* (3), 467-475. DOI: 10.1021/acscentsci.1c00039.
(15) Gallardo, A.; Bogart, B. M.; Dutagaci, B. Protein–Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2022**, *62* (12), 3079-3089. DOI: 10.1021/acs.jcim.2c00121.

(16) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836-3849. DOI: 10.1021/ct100481h.
(17) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; et al. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chemical Reviews* **2018**, *118* (8), 4177-4338. DOI: 10.1021/acs.chemrev.7b00427.
(18) 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 Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117* (19), 5179-5197. DOI: 10.1021/ja00124a002.
(19) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: a refined force field for DNA simulations. *Nature Methods* **2016**, *13* (1), 55-58. DOI: 10.1038/nmeth.3658.
(20) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8* (7), 2506-2520. DOI: 10.1021/ct300275s.
(21) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11* (12), 5723-5736. DOI: 10.1021/acs.jctc.5b00716.
(22) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817-3829. DOI: 10.1529/biophysj.106.097782.
(23) Mlýnský, V.; Kührová, P.; Kühr, T.; Otyepka, M.; Bussi, G.; Banáš, P.; Šponer, J. Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J. Chem. Theory Comput.* **2020**, *16* (6), 3936-3946. DOI: 10.1021/acs.jctc.0c00228.
(24) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *WIREs RNA* **2017**, *8* (3), e1405. DOI: 10.1002/wrna.1405.
(25) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15* (5), 3288-3305. DOI: 10.1021/acs.jctc.8b00955.
(26) Liebl, K.; Zacharias, M. Tumuc1: A New Accurate DNA Force Field Consistent with High-Level Quantum Chemistry. *J. Chem. Theory Comput.* **2021**, *17* (11), 7096-7105. DOI: 10.1021/acs.jctc.1c00682.
(27) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (7), E1346-E1355. DOI: 10.1073/pnas.1713027115.
(28) Janeček, M.; Kührová, P.; Mlýnský, V.; Otyepka, M.; Šponer, J.; Banáš, P. W-RESP: Well-Restrained Electrostatic Potential-Derived Charges. Revisiting the Charge Derivation Model. *J. Chem. Theory Comput.* **2021**, *17* (6), 3495-3509. DOI: 10.1021/acs.jctc.0c00976.

(29) Yoo, J.; Winogradoff, D.; Aksimentiev, A. Molecular dynamics simulations of DNA–DNA and DNA–protein interactions. *Current Opinion in Structural Biology* **2020**, *64*, 88-96. DOI: https://doi.org/10.1016/j.sbi.2020.06.007.
(30) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **2016**, *12* (1), 430-443. DOI: 10.1021/acs.jctc.5b00967.
(31) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (42), 16820-16825. DOI: doi:10.1073/pnas.1309392110.
(32) He, W.; Naleem, N.; Kleiman, D.; Kirmizialtin, S. Refining the RNA Force Field with Small-Angle X-ray Scattering of Helix–Junction–Helix RNA. *The Journal of Physical Chemistry Letters* **2022**, *13* (15), 3400-3408. DOI: 10.1021/acs.jpclett.2c00359.
(33) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405-4412. DOI: 10.1021/ct300613v.
(34) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, *5* (21), 3863-3871. DOI: 10.1021/jz501780a.
(35) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *J. Chem. Theory Comput.* **2020**, *16* (4), 2494-2507. DOI: 10.1021/acs.jctc.9b00251.
(36) Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E. Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA–Protein Complexes. *J. Phys. Chem. B* **2022**. DOI: 10.1021/acs.jpcb.1c10971.
(37) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell Jr, A. D. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929-1943.
(38) MacKerell Jr, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21* (2), 105-120.
(39) Yoo, J.; Aksimentiev, A. Improved Parametrization of Li+, Na+, K+, and Mg2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *The Journal of Physical Chemistry Letters* **2012**, *3* (1), 45-50. DOI: 10.1021/jz201501a.
(40) Savelyev, A. Assessment of the DNA partial specific volume and hydration layer properties from CHARMM Drude polarizable and additive MD simulations. *Phys. Chem. Chem. Phys.* **2021**, *23* (17), 10524-10535, 10.1039/D1CP00688F. DOI: 10.1039/D1CP00688F.
(41) Savelyev, A.; MacKerell, A. D. Competition among Li+, Na+, K+, and Rb+ Monovalent Ions for DNA in Molecular Dynamics Simulations Using the Additive CHARMM36 and Drude Polarizable Force Fields. *J. Phys. Chem. B* **2015**, *119* (12), 4428-4440. DOI: 10.1021/acs.jpcb.5b00683.
(42) Baltrukevich, H.; Bartos, P. RNA-protein complexes and force field polarizability. *ChemRxiv* **2022**. DOI: 10.26434/chemrxiv-2022-hvnx1
(43) Lemkul, J. A.; MacKerell Jr., A. D. Polarizable force field for RNA based on the classical drude oscillator. *J. Comput. Chem.* **2018**, *39* (32), 2624-2646. DOI: https://doi.org/10.1002/jcc.25709.

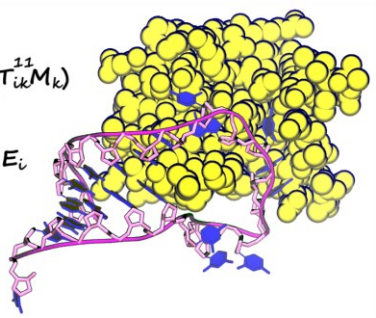
(44) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116* (9), 4983-5013. DOI: 10.1021/acs.chemrev.5b00505.
(45) Savelyev, A.; MacKerell, A. D. Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. *J. Phys. Chem. B* **2014**, *118* (24), 6742-6757. DOI: 10.1021/jp503469s.
(46) Salsbury, A. M.; Lemkul, J. A. Recent developments in empirical atomistic force fields for nucleic acids and applications to studies of folding and dynamics. *Current Opinion in Structural Biology* **2021**, *67*, 9-17. DOI: https://doi.org/10.1016/j.sbi.2020.08.003.
(47) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement Using Quantum Mechanical Base Stacking and Conformational Energetics. *J. Chem. Theory Comput.* **2017**, *13* (5), 2053-2071. DOI: 10.1021/acs.jctc.7b00067.
(48) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J. Chem. Theory Comput.* **2017**, *13* (5), 2072-2085. DOI: 10.1021/acs.jctc.7b00068.
(49) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14* (4), 2084-2108. DOI: 10.1021/acs.jctc.7b01169.
(50) Zhang, C.; Bell, D.; Harger, M.; Ren, P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J. Chem. Theory Comput.* **2017**, *13* (2), 666-678. DOI: 10.1021/acs.jctc.6b00918.
(51) Zhang, C.; Lu, C.; Wang, Q.; Ponder, J. W.; Ren, P. Polarizable Multipole-Based Force Field for Dimethyl and Trimethyl Phosphate. *J. Chem. Theory Comput.* **2015**, *11* (11), 5326-5339. DOI: 10.1021/acs.jctc.5b00562.
(52) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48* (1), 371-394. DOI: 10.1146/annurev-biophys-070317-033349.
(53) 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* (6505), 432-438. DOI: 10.1038/372432a0.
(54) 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* (48), 20186-20191. DOI: 10.1073/pnas.0812076106.
(55) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton. *J. Mol. Biol.* **1999**, *285* (4), 1735-1747. DOI: 10.1006/jmbi.1998.2401.
(56) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular-Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990-2001. DOI: 10.1063/1.463137.
(57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. DOI: 10.1063/1.2408420.
(58) Daan, F.; Berend, S. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press, Inc., 1996.
(59) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. *J. Comput. Chem.* **2017**, *38* (23), 2047-2055. DOI: 10.1002/jcc.24853.

(60) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933-5947. DOI: 10.1021/jp027815+.
(61) Wang, Z. Polarizable Force Field Development, and Applications to Conformational Sampling and Free Energy Calculation. Ph. D. Dissertation, Washington University in St. Louis, St. Louis, Missouri, 2018.
(62) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, 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* (9), 2886-2902. DOI: 10.1021/ct200162x.
(63) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144* (21), 214110. DOI: 10.1063/1.4952647.
(64) Fink, R. F. Spin-component-scaled Møller–Plesset (SCS-MP) perturbation theory: A generalization of the MP approach with improved properties. *J. Chem. Phys.* **2010**, *133* (17), 174113. DOI: 10.1063/1.3503041.
(65) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19* (48), 32184-32215, 10.1039/C7CP04913G. DOI: 10.1039/C7CP04913G.
(66) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7* (4), 499-512. DOI: 10.1017/S1355838201002515.
(67) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985-1993, 10.1039/B600027D. DOI: 10.1039/B600027D.
(68) Qi, R.; Jing, Z.; Liu, C.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Elucidating the Phosphate Binding Mode of Phosphate-Binding Protein: The Critical Effect of Buffer Solution. *J. Phys. Chem. B* **2018**, *122* (24), 6371-6376. DOI: 10.1021/acs.jpcb.8b03194.
(69) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins* **2000**, *40* (3), 389-408. DOI: 1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2.
(70) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046-4063. DOI: 10.1021/ct4003702.
(71) Hermann, T.; Westhof, E. Simulations of the dynamics at an RNA–protein interface. *Nature Structural Biology* **1999**, *6* (6), 540-544. DOI: 10.1038/9310.
(72) 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 Research* **2006**, *34* (1), 275-285. DOI: 10.1093/nar/gkj436.
(73) Law, M. J.; Rice, A. J.; Lin, P.; Laird-Offringa, I. A. The role of RNA structure in the interaction of U1A protein with U1 hairpin II RNA. *RNA* **2006**, *12* (7), 1168-1178. DOI: 10.1261/rna.75206.
(74) Pokorná, P.; Kruse, H.; Krepl, M.; Šponer, J. QM/MM Calculations on Protein–RNA Complexes: Understanding Limitations of Classical MD Simulations and Search for Reliable

Cost-Effective QM Methods. *J. Chem. Theory Comput.* **2018**, *14* (10), 5419-5433. DOI: 10.1021/acs.jctc.8b00670.
(75) Qin, F.; Chen, Y.; Wu, M.; Li, Y.; Zhang, J.; Chen, H.-F. Induced fit or conformational selection for RNA/U1A folding. *RNA* **2010**, *16* (5), 1053-1061. DOI: 10.1261/rna.2008110.

$$\mu_i^{ind} = \alpha_i \left(\sum_{j \neq i} T_{ij}^{1} M_j + \sum_{k \neq i} T_{ik}^{11} M_k \right)$$

$$U_{elec}^{ind} = -\frac{1}{2} \left(\sum_i (\mu_i^{ind})^T E_i \right)$$



Molecular Dynamics Simulations of Protein RNA Complexes by Using an Advanced Electrostatic Model

Zhifeng Jing and Pengyu Ren*

Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712

* Corresponding author. Email: pren@mail.utexas.edu

Simulations of RNA and DNA systems

The optimized parameters were tested on several RNA and DNA systems following the procedure in our previous work.[1] Three 1000-ns simulations were performed for tetranucleotides AAAA and GACC. Two 500-ns simulations were performed for 2JXQ, 2KOC and 1NAJ. The AAAA simulations had slightly lower percentage of A-form conformation compared to the simulations with the original AMOEBA parameters, while the GACC system had slightly higher percentage of A-form. Simulations of an RNA duplex, an RNA hairpin and a DNA duplex resulted in slightly higher RMSDs to the crystal structure than simulations with the original parameters.

Table S1. Numbers of model dimer structures for protein-nucleic acid interactions with PDB geometry

	A	T	U	G	C
A	15				
T	14	10			
U	29	2	15		
G	35	19	34	25	
C	25	15	25	21	16
NMA	19	17	17	16	19
EtOH	7	8	9	9	9
Acetate	10	12	11	11	12
MeNH3	6	4	4	5	4
Imidazole	14	12	12	14	12
Benzene	16	17	17	18	16

Table S2. Root mean squared deviations (Å) of DNA and RNA systems in AMOEBA simulations.

		RMSD(amoebabio18)		RMSD(this work)	
Type	PDB	Non-terminal	Heavy-atom	Non-terminal	Heavy-atom
RNA duplex	2JXQ	1.25	1.98	1.588	2.598
		1.41	2.38	1.416	2.385
RNA hairpin	2KOC	1.72	1.78	1.93	1.983
		1.9	1.95	1.901	1.977
B-DNA duplex	1NAJ	1.29	2.54	1.439	2.17
				1.532	2.343

For each system, the results of two independent simulations are listed.

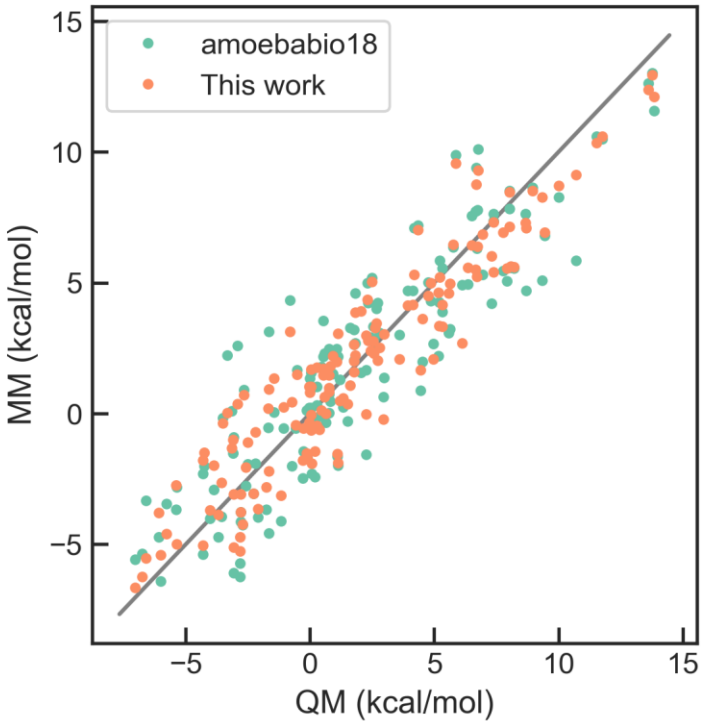


Figure S1. Correlation between QM and MM relative energies for the Asn conformers generated from one-dimensional torsion scan and rotamer library.

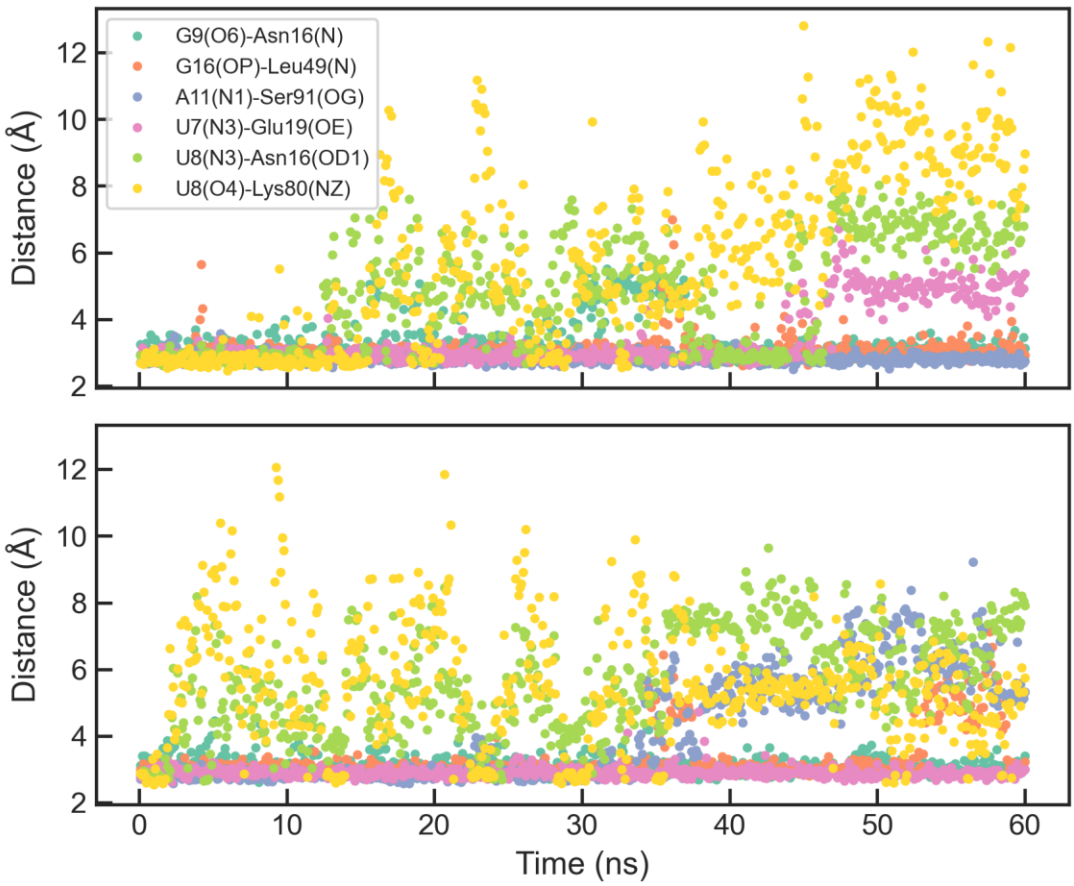


Figure S2. Evolution of selected hydrogen-bond distances in simulations of U1A with optimized vdW parameters.

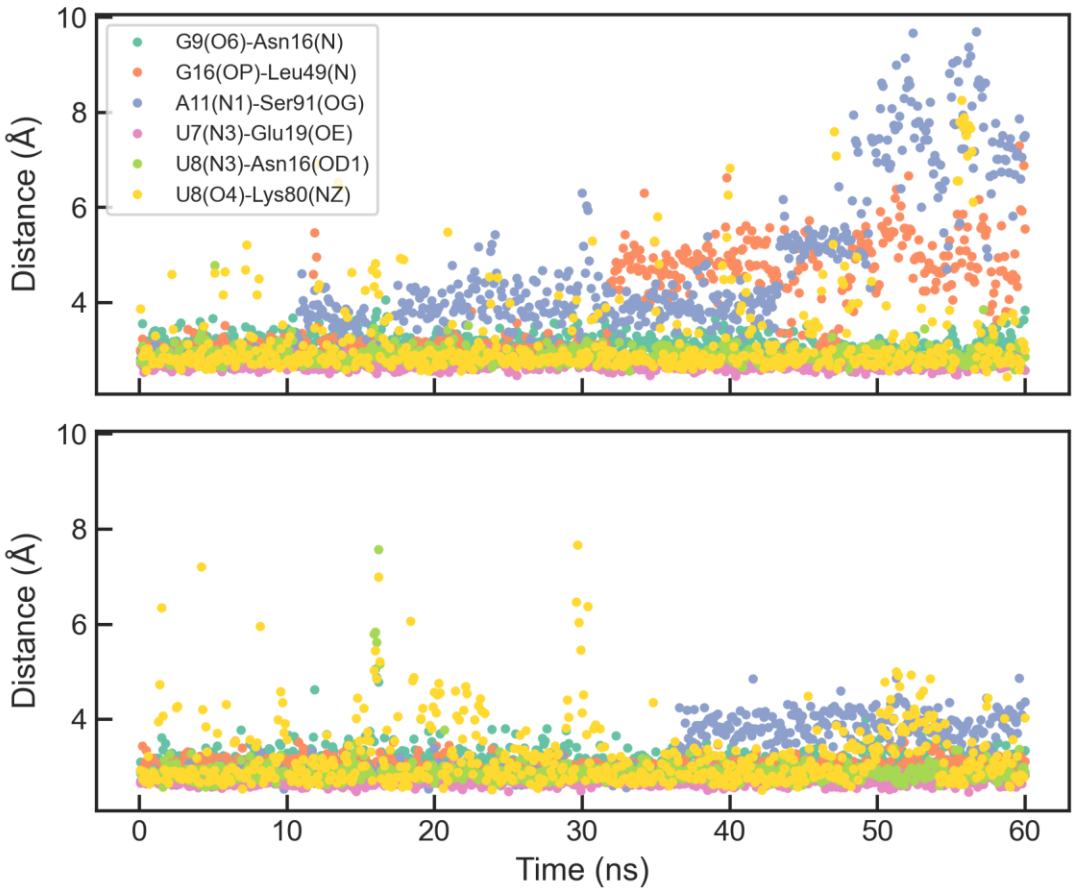


Figure S3. Evolution of selected hydrogen-bond distances in simulations of U1A with optimized vdW and Asn side-chain torsion parameters.

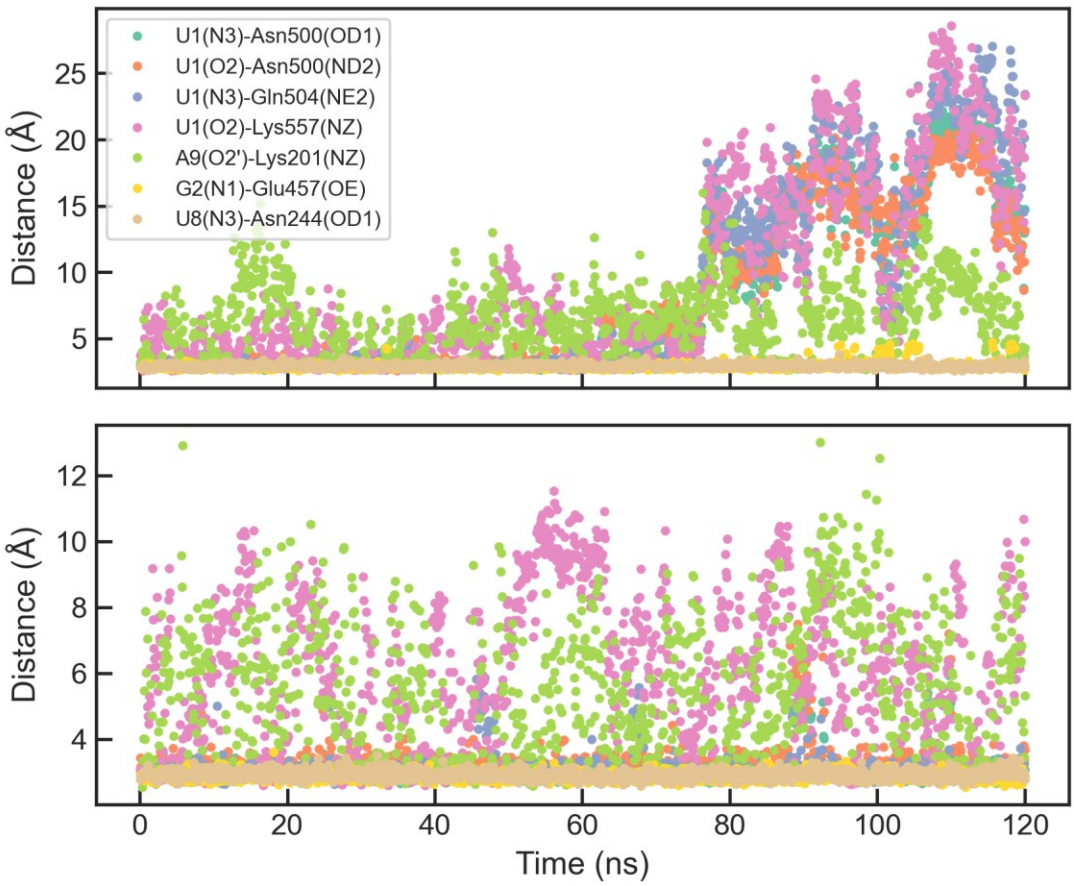


Figure S4. Evolution of selected hydrogen-bond distances in simulations of FBF with original AMOEBA parameters.

References

(1) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14* (4), 2084-2108. DOI: 10.1021/acs.jctc.7b01169.