JCTC Journal of Chemical Theory and Computation

Letter

pubs.acs.org/JCTC

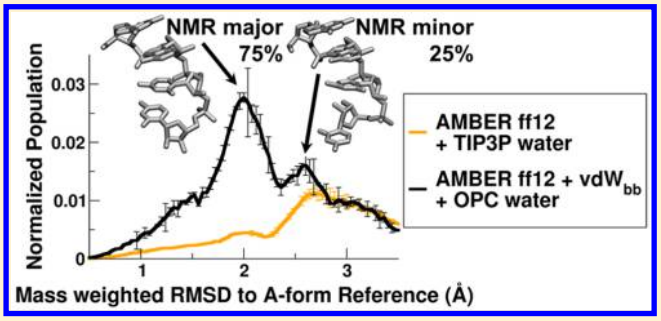
Improved Force Field Parameters Lead to a Better Description of RNA Structure

Christina Bergonzo and Thomas E. Cheatham III*

Department of Medicinal Chemistry, College of Pharmacy, University of Utah, Salt Lake City, Utah 84112, United States

Supporting Information

ABSTRACT: We compare the performance of two different RNA force fields in four water models in simulating the conformational ensembles r(GACC) and r(CCCC). With the increased sampling facilitated by multidimensional replica exchange molecular dynamics (M-REMD), populations are compared to NMR data to evaluate force field reliability. The combination of AMBER ff12 with vdW_{bb} modifications and the OPC water model produces results in quantitative agreement with the NMR ensemble that have eluded us to date.



Modeling nucleic acid structure using molecular dynamics simulations has become increasingly common, so the reliability of simulation with respect to experiment has become increasingly important.[1] RNA structure is dynamic and able to adopt multiple different secondary and tertiary conformations to perform a variety of functions, making parametrization of this molecule difficult. Recent changes to the AMBER RNA force fields used in molecular dynamics (MD) simulations have modified the glycosidic χ dihedral angle and the nonbonded terms in an attempt to improve agreement with both quantum calculations and experiment.[2−4]

Though reproducing the dynamics of a biomolecule is nontrivial, nucleic acids in particular are very sensitive to their ion environment. To provide appropriate balance between the models, recent ion parameter sets have been modified to be used with specific explicit water models, and results have been encouraging.[5,6] Due to the formal charge on RNA's phosphate group, it may help to think of nucleic acids more specifically as a polyanion, where there is every reason to expect the choice of water model *also* plays an important role in nucleic acid dynamics and the conformational ensemble that results from MD simulation.[7]

Water models, specifically rigid models of the three point (TIP3P[8] and SPC/E[9]) and four point (TIP4PEw[10]) variety, have been parametrized to reproduce various properties of liquid water, though no single model accurately describes all properties.[11] The more recently parametrized "optimal" point charge (OPC) water model was parametrized to capture a water molecule's charge hydration asymmetry[12] and the three lowest order multipole moments of water, and this model does a better job of accurately reproducing several experimental bulk properties of water.[13]

To test the effect of the water model on the conformational distribution of two small, though conformationally diverse, tetranucleotides, r(GACC) and r(CCCC), for which the structure distributions are known,[14,15] we ran multidimensional replica exchange MD (M-REMD) simulations to generate very highly converged structure ensembles for two force fields in combination with four water models. The standard AMBER ff for nucleic acids,[16,17] incorporating the Barcelona modifications to the α/γ backbone torsions[18] with the OL1 χ modifications,[4] is referred to as ff12. Additional van der Waals modifications to the phosphate oxygens[19] were applied to ff12, and the resulting force field is referred to as ff12 + vdW_{bb}. Water models TIP3P,[8] SPC/E,[9] TIP4PEw,[10] and OPC[13] were used. Neutralizing Na^+ ions were used, described by the Joung−Cheatham monovalent ion parameters which are optimized for each water model (with the exception of OPC).[6] For the systems simulated with OPC water, TIP4P-Ew-optimized ion parameters were used, as suggested by the authors of the OPC water model (AMBER14 manual).[20] M-REMD simulations were performed as described previously, combining temperature and scaled dihedral force constant Hamiltonian dimensions, totaling 192 replicas, where each replica ran on its own GPU.[21] All simulations were performed in duplicate using independently generated starting structure sets and were run on Blue Waters using GPU-accelerated *pmemd.cuda.MPI* available in the Amber14 suite of programs.[20,22,23]

Figure 1 shows histograms of the mass-weighted heavy atom RMSD of each set of M-REMD simulations to an A-form reference structure for the AMBER ff12 + vdW_{bb} force field with each of the different water models. Dashed gray lines denote the first two peaks which correspond to NMR major and NMR minor structures, respectively. The last highlighted peak corresponds to the intercalated structure[14,24] populated in MD simulation with these force fields, but not observed in NMR. In this structure, the C3 base is intercalated between the

Received: May 13, 2015
Published: August 7, 2015

DOI: 10.1021/acs.jctc.5b00444
J. Chem. Theory Comput. 2015, 11, 3969−3972

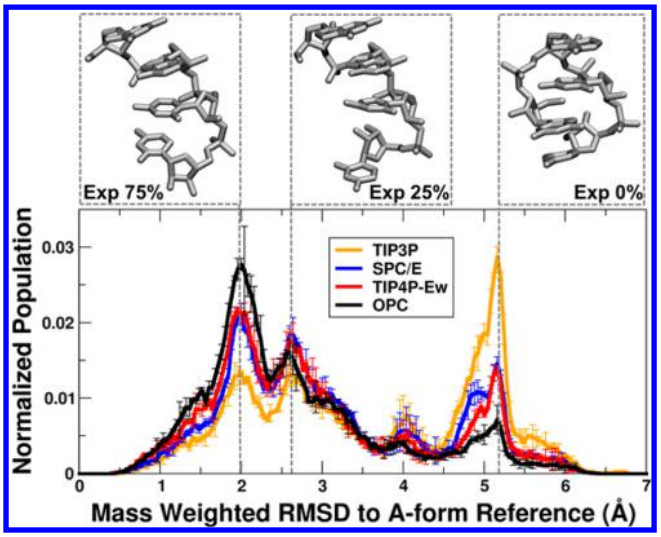


Figure 1. Population analysis showing the distribution percentage of structures at specific RMSD values from an A-form reference structure for AMBER ff12 + vdW$_{bb}$ in four water models. Mass weighted RMSD histograms of the unbiased replicas at 277 K are averaged between two runs. Error bars represent standard deviation between two runs. Independent sets of starting structures were used for each simulation. The population analysis for ff12 is shown in Supporting Figure 1.

G1 and A2 bases, and the G1 base is intercalated between the C3 and C4 bases. Base−backbone and backbone−backbone hydrogen bonds stabilize the structure. As the population of NMR major structure increases from TIP3P to SPC/E and TIP4PEw to OPC, the population of this artificial structure decreases, summarized in Table 1 as "largest non-NMR". We can attribute a 2× increase in the NMR major population to using the vdW$_{bb}$ force field modifications (ff12 + TIP3P to ff12 + vdW$_{bb}$ + TIP3P) and another 2× increase in the NMR major population from employing the OPC water model. Interestingly, the population of NMR minor structure is similar in the ff12 + vdW$_{bb}$ TIP3P and OPC simulations, while the population of non-NMR structures decreases from 37.2% of the ensemble to 7.5%. This indicates the OPC water model plays a significant role in penalizing the overstacked and over-hydrogen-bonded conformations in favor of the NMR major conformation specifically, and not just more "extended" conformations in general.

The ratios of NMR major to NMR minor described in experiment, as well as in each of the force field + water combinations tested here, are reported in Supporting Table 1. The combination of ff12 with all water models yield, at highest, a ratio of 1.6:1, compared to the experimental ratio of 3:1. Using the vdW$_{bb}$ modifications improves the NMR major to minor ratio for all water models, though the four-point models do better than the three-point models, and OPC does better than TIP4PEw, most accurately reproducing the experimental ratio out of the models and force fields tested (at a ratio of major to minor of 3.16:1).

For the r(CCCC) tetranucleotide, simulations were carried out in each force field, but water models were limited to TIP3P and OPC based on the results in the r(GACC) system. Figure 2

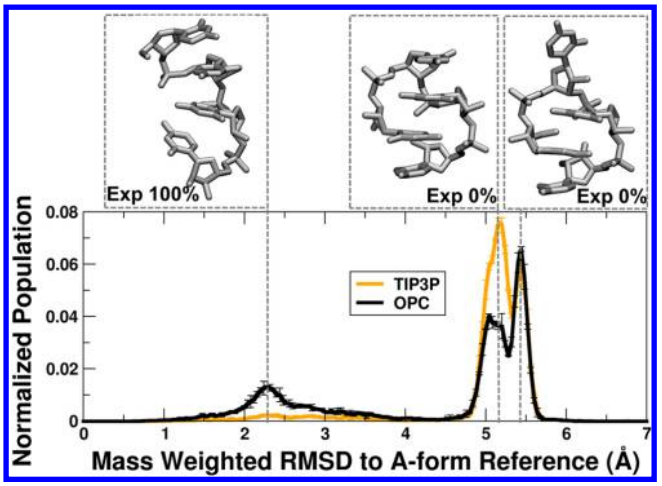


Figure 2. Population analysis showing the distribution percentage of structures at specific RMSD values from an A-form reference structure for AMBER ff12 + vdW$_{bb}$ in TIP3P and OPC water models. Mass weighted RMSD histograms of the unbiased replicas at 277 K are averaged between two runs. Error bars represent standard deviation between two runs. Independent sets of starting structures were used for each simulation. The population analysis for ff12 is shown in Supporting Figure 2.

shows histograms of the mass-weighted RMSD from an A-form reference structure for the AMBER ff12 + vdW$_{bb}$ modifications in TIP3P (orange) and OPC (black) water models. Much less conformational variability than the r(GACC) tetranucleotide is sampled by the M-REMD simulations for r(CCCC), and only three representative clusters were found. The histogram peak around 2.25 Å represents the extended NMR structure, where a significant increase in population is observed using OPC water. As in the r(GACC) system, a non-negligible percentage of the ensemble populates the intercalated structure, or a variant where the C2 base is unstacked, shown as representative structures corresponding to the histogram peaks at 5.2 and 5.8 Å, respectively. Unlike the r(GACC) system, the combination of ff12 + vdW$_{bb}$ and OPC water does not shift the ensemble to

Table 1. Percent NMR Major, Minor, and Other Structures Found by Each Force Field/Water Model Combination

force field	water model	NMR major	NMR minor	all non-NMR	largest non-NMR
experiment		~75 ± 5	~25 ± 5	0	n/a
ff12	TIP3P	15.8 ± 0.8	18.6 ± 1.2	39.5 ± 1.1	20.3 ± 0.5
	SPC/E	24.3 ± 1.0	27.8 ± 0.1	25.8 ± 2.2	7.1 ± 0.9
	TIP4PEw	30.7 ± 0.1	27.4 ± 1.0	19.9 ± 2.2	6.1 ± 1.5
	OPC	39.5 ± 2.1	24.3 ± 0.6	11.7 ± 1.7	3.1 ± 0.3
ff12 + vdW$_{bb}$	TIP3P	31.6 ± 1.0	17.7 ± 0.1	37.2 ± 9.5	19.0 ± 0.3
	SPC/E	40.2 ± 0.4	23.1 ± 0.4	20.2 ± 0.5	6.9 ± 0.1
	TIP4PEw	48.5 ± 1.0	19.7 ± 1.5	15.0 ± 0.5	6.0 ± 0.4
	OPC	58.5 ± 0.5	18.5 ± 0.4	7.5 ± 0.7	2.4 ± 0.1

DOI: 10.1021/acs.jctc.5b00444
J. Chem. Theory Comput. 2015, 11, 3969−3972

quantitatively reproduce the experimental NMR; both intercalated structures still populate a majority of the ensemble.

Table 2 shows the percent of NMR, intercalated and intercalated 2-unstacked structures found. With ff12 in TIP3P water, the NMR structure is not found. Employing the ff12 + vdW_{bb} modifications shifts the population of the NMR structure to roughly 2% of the ensemble. However, switching to the OPC water model increases the percentage of NMR structure to 12.6% with ff12; this population almost doubles to 20.4% when both OPC and vdW_{bb} modifications are applied. Notably, in both force fields, use of the OPC water model destabilizes the intercalated structure in favor of the extended NMR structure and also shifts some population to the intercalated 2-unstacked cluster. Both of these structures can be characterized as more "extended" compared to the intercalated structure.

Table 2. Percent NMR, Intercalated, and Intercalated 2-Unstacked Structures Found by Each Force Field/Water Model Combination

force field	water model	NMR	intercalated	intercalated 2-unstacked
experiment		~100 ± 5	0	0
ff12	TIP3P	not found	52.4 ± 0.9	28.3 ± 0.6
	OPC	12.6 ± 0.1	23.7 ± 1.6	33.3 ± 1.2
ff12 + vdW_{bb}	TIP3P	2.1 ± 0.1	52.0 ± 0.4	26.5 ± 0.6
	OPC	20.4 ± 0.4	27.4 ± 0.3	28.6 ± 0.4

The improvement in each of the simulated ensembles when the combination of ff12 + vdW_{bb} and OPC is used is due to a combination of contributing factors. The increased vdW radii of the backbone phosphate oxygen atoms destabilize the intercalated structure in favor of the extended NMR major structure. This has more of an effect in the r(GACC) system where it is aided, in part, by the heterogeneity of the tetranucleotide; r(GACC) intercalated structure is more easily disrupted than its r(CCCC) intercalated counterpart.[25] Though the r(CCCC) ensemble does not quantitatively reproduce experiment, it should be noted that the ff12 + vdW_{bb} + OPC combination increases the population of NMR structure by an order of magnitude, making it at least isoenergetic with the other two intercalated structures. Additionally, the OPC model more faithfully reproduces the higher multipole moments of water and is better able to represent the octopole and model the effect of the unpaired electrons, important for water−ion interactions.[26]

The results suggest that the specific combination of water model and nucleic acid force field can drastically influence the final conformational ensemble. Here we show that the OPC water model with the ff12 + vdW_{bb} RNA force field modifications appear to better represent the RNA structural ensemble. r(GACC) results quantitatively reproduce the experimental NMR, but fall short for the r(CCCC) system. These results help us pinpoint future changes to the force field, which will need to address the overstabilization of RNA−RNA hydrogen bonding. Moving beyond r(GACC) and r(CCCC), our preliminary results suggest that with each of these model systems this force field combination appears to improve the conformational ensemble with respect to experiment (*unpublished data*). Though more in-depth investigation of more diverse RNA systems is necessary, the net gain in fidelity to experiment cannot be ignored. Though the OPC model's extra point reduces the achievable nanosecond per day by roughly 28% compared to the SPC/E three-point model (Supporting Table 2), we feel this increased simulation cost can be ameliorated by using hydrogen mass repartitioning with a 4 fs time step[27] and by taking advantage of the water model's availability in AMBER's GPU-accelerated MD engine.[22]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00444.

Population analyses, ratios of major to minor NMR results, convergence anlysis, simulation details, and analysis scripts (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tec3@utah.edu.

Funding

This research is enabled by the Blue Waters sustained-petascale computing project (NSF OCI 07-25070 and PRAC OCI-1440031), the NSF Extreme Science and Engineering Discovery Environment (XSEDE, OCI-1053575) and allocation MCA01S027P, the Center for High Performance Computing at the University of Utah. Research funding comes from NIH R01-GM098102 (TEC).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank James Roberson for careful reading of this manuscript.

■ ABBREVIATIONS

MD, molecular dynamics; M-REMD, multidimensional replica exchange molecular dynamics; OPC, optimal point charge; RNA, ribonucleic acid; vdW, van der Waals

■ REFERENCES

(1) Cheatham, T. E., III; Case, D. A. Twenty-Five Years of Nucleic Acid Simulations. *Biopolymers* **2013**, *99* (12), 969−977.

(2) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA Chi Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **2010**, *6* (5), 1520−1531.

(3) 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.

(4) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., 3rd; 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.

(5) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733−2748.

(6) Joung, I. S.; Cheatham, T. E., III. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020−9041.

(7) Sklenovský, P.; Florová, P.; Banáš, P.; Réblová, K.; Lankaš, F.; Otyepka, M.; Šponer, J. Understanding RNA Flexibility Using Explicit

DOI: 10.1021/acs.jctc.5b00444
J. Chem. Theory Comput. 2015, 11, 3969−3972

Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs a. *J. Chem. Theory Comput.* **2011**, *7*, 2963−2980.
(8) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6665−6670.
(9) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269−6271.
(10) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665−9678.
(11) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954−9960.
(12) Mukhopadhyay, A.; Fenley, A. T.; Tolokh, I. S.; Onufriev, A. V. Charge Hydration Asymmetry: The Basic Principle and How to Use It to Test and Improve Water Models Charge Hydration Asymmetry. *J. Phys. Chem. B* **2012**, *116* (32), 9776−9783.
(13) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.
(14) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. Benchmarking AMBER Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised X Torsions. *J. Phys. Chem. B* **2011**, *115* (29), 9261−9270.
(15) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-Handed Helix, and Revised Parameters for AMBER Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry* **2013**, *52* (6), 996−1010.
(16) Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. A Modified Version of the Cornell et Al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845−862.
(17) Wang, J.; 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* (12), 1049−1074.
(18) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/gamma Conformers. *Biophys. J.* **2007**, *92* (11), 3817−3829.
(19) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405−4412.
(20) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*, 2014.
(21) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., 3rd Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10* (1), 492−499.
(22) 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* (9), 3878−3888.
(23) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542−1555.
(24) Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., 3rd Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117* (15), 4014−4027.
(25) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., III. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21* (9), 1−14.
(26) Tan, M.-L.; Lucan, L.; Ichiye, T. Study of Multipole Contributions to the Structure of Water around Ions in Solution Using the Soft Sticky Dipole-Quadrupole-Octupole (SSDQO) Model of Water. *J. Chem. Phys.* **2006**, *124* (17), 174505.
(27) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long Time Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864−1874.

DOI: 10.1021/acs.jctc.5b00444
J. Chem. Theory Comput. 2015, 11, 3969−3972