Assessing the Current State of Amber Force Field Modifications for DNA

Rodrigo Galindo-Murillo,[†,∥] James C. Robertson,[†,∥] Marie Zgarbová,[‡] Jiří Šponer,[‡,§] Michal Otyepka,[‡] Petr Jurečka,[‡] and Thomas E. Cheatham, III*,[†]

[†]Department of Medicinal Chemistry, University of Utah, 2000 East 30 South, Skaggs 105, Salt Lake City, Utah 84112, United States

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

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

S *Supporting Information*
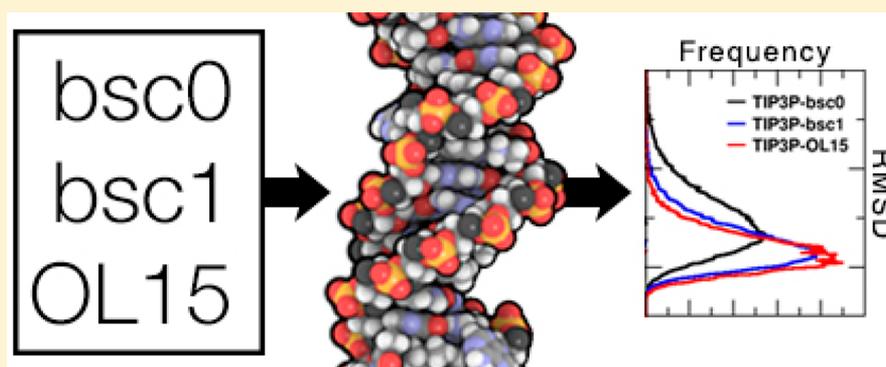
ABSTRACT: The utility of molecular dynamics (MD) simulations to model biomolecular structure, dynamics, and interactions has witnessed enormous advances in recent years due to the availability of optimized MD software and access to significant computational power, including GPU multicore computing engines and other specialized hardware. This has led researchers to routinely extend conformational sampling times to the microsecond level and beyond. The extended sampling time has allowed the community not only to converge conformational ensembles through complete sampling but also to discover deficiencies and overcome problems with the force fields. Accuracy of the force fields is a key component, along with sampling, toward being able to generate accurate and stable structures of biopolymers. The Amber force field for nucleic acids has been used extensively since the 1990s, and multiple artifacts have been discovered, corrected, and reassessed by different research groups. We present a direct comparison of two of the most recent and state-of-the-art Amber force field modifications, bsc1 and OL15, that focus on accurate modeling of double-stranded DNA. After extensive MD simulations with five test cases and two different water models, we conclude that both modifications are a remarkable improvement over the previous bsc0 force field. Both force field modifications show better agreement when compared to experimental structures. To ensure convergence, the Drew−Dickerson dodecamer (DDD) system was simulated using 100 independent MD simulations, each extended to at least 10 μs, and the independent MD simulations were concatenated into a single 1 ms long trajectory for each combination of force field and water model. This is significantly beyond the time scale needed to converge the conformational ensemble of the internal portions of a DNA helix absent internal base pair opening. Considering all of the simulations discussed in the current work, the MD simulations performed to assess and validate the current force fields and water models aggregate over 14 ms of simulation time. The results suggest that both the bsc1 and OL15 force fields render average structures that deviate significantly less than 1 Å from the average experimental structures. This can be compared to similar but less exhaustive simulations with the CHARMM 36 force field that aggregate to the ~90 μs time scale and also perform well but do not produce structures as close to the DDD NMR average structures (with root-mean-square deviations of 1.3 Å) as the newer Amber force fields. On the basis of these analyses, any future research involving double-stranded DNA simulations using the Amber force fields should employ the bsc1 or OL15 modification.

■ INTRODUCTION

The ability to simulate the structure and dynamics of nucleic acids, especially DNA at atomic resolution and over biologically relevant time scales, has led to new insights into the richness and complexity of dynamics on the submillisecond time scale.[1−7] This has been driven by improvements in hardware[8−10] and simulation methods[11−14] and continual develop-

ment and optimization of the underlying physical model that describes the system, more specifically, the force field.[15−23] As improvements to force fields are proposed, it is critically important to understand the strengths and weaknesses of the models and to assess and evaluate the force fields with simulations that sample well the expected properties within appropriate time scales.[24,25]

Although microsecond-length simulations of fully solvated atomistic duplex DNA are now routine,[3,26−29] only recently have convergence and reproducibility of the structure and dynamics of the internal portions of a DNA helix on the ∼1−5 μs time scale been convincingly demonstrated.[3,9,10,29] This has eliminated the previous limitations of not being able to achieve enough sampling (for modestly sized systems), noting that the community has still not reached the time scale (ms+) of being able to model internal base pair opening. Regardless, by eliminating the sampling problem—of the internal helix, neglecting internal base pair opening—the community can now focus on developing, assessing, and validating force fields as accurately as possible. The efforts of multiple research groups have led to two force fields that deserve close scrutiny in order to determine how well each reproduces experimental observables. Deeper exploration is required so that users in the community can determine the best force field to use to suit their needs. If past experiences serve, it is expected that longer simulations and usage by a larger community will inevitably uncover further deficiencies in the current force fields; nevertheless, it is prudent to learn as much about the current models as possible. Even on the microsecond time scale, simulations reveal limitations in the force fields that can provide insight into where to focus efforts for even further improvement. Since the original parm94 force field[30,31] was updated to parm98[32] and parm99,[33] the development of new parameters has progressed along two primary paths (Figure 1). One fork follows the efforts led by the Orozco group and is named for the Barcelona Supercomputing Center (BSC). The bsc0 modifications, published in 2007,[34] improved upon parm99 by updating the α and γ dihedrals. This corrected α overpopulated in gauche+ conformations and γ overpopulated in trans conformations. The most recent force field developed by this group, bsc1,[16] was released in 2015 and includes the bsc0 modifications and additional modifications to the sugar pucker, the χ glycosidic torsion, and the ε and ζ dihedrals. The other fork follows the collective work by research groups from the Czech Republic and includes "OL" in the name, referring to the city of Olomouc, where these parameters were generated. The number represents the version for that particular parameter; for example, χ_{OL4} represents the fourth version for the χ dihedral. Development along this path progressed in an incremental fashion as improvements from parm99 were made to the χ glycosidic torsion with χ_{OL4}.[17] The next OL improvements came when updates were made to the ε and ζ dihedrals.[23] At this point Amber 15 was released, and the recommended DNA force field was the combination bsc0 + χ_{OL4} + ε/ζ_{OL1}. These parameters resulted in improvements over bsc0 by increasing the populations of BII, increasing twist, and reducing the major groove width. Since then, the β dihedral has been parametrized to improve the ZI and ZII substates in Z-DNA, and the new ff-nucleic-OL15 (OL15) parameter set, consisting of the combination parm99 + bsc0 + χ_{OL4} + ε/ζ_{OL1} + β_{OL1}, has been released.[22] Thus, compared with the original 1995 force field of Cornell et al.,[30] the OL15 version represents a complete one-dimensional (uncoupled) parametrization of all
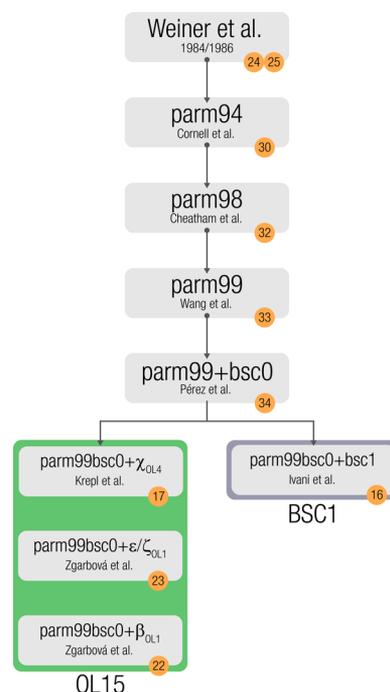
Figure 1. Historical flowchart of the two main forks of the AMBER force field for DNA. Until recently it was common that DNA simulations were run with the most recent force field available, but both OL15 and bsc1 were released in the past year and have shown improvements over previous iterations of the DNA force fields. The recent advances seen in two DNA force fields motivated the need for detailed comparisons of results from simulations. The combination of parameter modifications that now constitute OL15 is highlighted in green. References are presented in the orange circles.

of the DNA dihedral backbone potentials and thus may reach the accuracy limits obtainable by tuning the dihedral parameters of the Cornell et al. parameters.

In the present work, we present a detailed evaluation of the two most recent Amber force field modifications for DNA, bsc1 and OL15, which were developed to improve the accuracy of molecular dynamics (MD) simulations of double-stranded DNA on relatively long time scales. The systems tested include two solution NMR reference structures, which provide reliable comparison to experiment. In addition, two very high resolution X-ray structures of B-form DNA and a Z-DNA crystal structure were simulated in solution to assess the force field performance on a variety of DNA sequences. Although the results of solution-phase simulations cannot be directly compared to the crystal results, the simulations were performed to demonstrate that the force fields were not overtrained for the NMR structures and to highlight sequence-specific structural differences apparent in the current force fields compared with the earlier versions. The Z-DNA structure has been a particularly challenging system for previous and current force fields[17] because of the complex topology that includes left-handed helicity and alternating syn and anti nucleotide conformations, leading to less separation between backbone phosphates compared with B-DNA.[35−37] Thus, specific conditions are required in order to obtain Z-DNA in solution: high salt concentration, low humidity, and a specific sequence of CG steps.[37,38] The study of B−Z DNA transitions[39] is beyond the scope of this work, but this biologically important[40,41] molecule was included to show how the structure and dynamics of Z-DNA in solution are influenced

Table 1. Description of MD Simulations and Systems[a]

PDB entry	sequence	aggregated simulation time per force field/water combination (ms)	experimental details	resolution (Å)	type	ref
1BNA/1NAJ	d(CGCGAATTCGCG)	1	X-ray/NMR	1.9	B-DNA	2, 48
1FZX	d(GGCAAAAAACGG)	0.5	NMR	–	B-DNA	49
1SK5	d(CTTTTAAAAG)	0.1	X-ray	0.89	B-DNA	50
3GGI	d(CCAGGCCTGG)	0.1	X-ray	0.6	B-DNA	51
1I0T	d(CGCGCG)	0.1	X-ray	1.74	Z-DNA	52

[a]All of the simulations were performed with net-neutralizing ions and excess NaCl to reach a concentration of ~200 mM using the Joung–Cheatham ion model and with the TIP3P and OPC water models. All of the simulations were of duplex DNA, although only single strands are listed in the table. It should be noted that all of the sequences are palindromic except for PDB entry 1FZX.

by different force field parameters, even though Z-DNA in solution under these conditions (with relatively short MD simulation time scales compared with that for B−Z transitions and low salt concentrations) is likely metastable. Less extensive sampling, but still aggregating ~90 μs of MD simulation with the CHARMM 36 force field, was also performed on one of the NMR structures, the Drew−Dickerson dodecamer (DDD),[2] although it was not analyzed as deeply. The different approaches to developing the force fields, and the resulting parameter sets, demonstrate that an accurate force field is a difficult problem to solve. The updated parameters do show incremental improvement over previous versions and in general demonstrate increased agreement with experiment. Nucleic acids and protein simulations are routinely performed in explicit solvent, mainly using the TIP3P,[42] SPC/E,[43] and TIP4P/Ewald[44] water models. Few differences in the DNA structures have been observed (unpublished observations). However, given the development of the optimal point charge (OPC) water model in 2014,[45] which has been shown to improve the agreement with experiment for some systems,[46] we decided to explore this model further in the current work. This model was parametrized to capture the charge asymmetry of a water molecule, and this leads to improvement in simulating the properties of bulk water.[47] The assessment and validation with the OPC water model led, somewhat surprisingly, to a slight overall improvement in all of the tested systems, although with a considerable hit in the simulation speed (~30% slower). Our extended simulations, totaling 14.4 ms of combined sampling time on five tested systems, confirm improvements of simulated DNA with both the bsc1 and OL15 force fields compared to high-resolution experimental X-ray structures and NMR spectroscopy structures. Both bsc1 and OL15 perform in a remarkably similar manner, and only very detailed and specific point differences were detected. This has led to the recommendation with the release of the Amber 16 code base that both force fields be considered as recommended for simulation of DNA over other available force fields.

■ METHODS

The systems considered in order to evaluate the different force field modifications are presented in Table 1. All of the systems were modeled using three different force fields: the current parm99 with the bsc0 correction (denoted as bsc0), the recent bsc1, and OL15 (a combination of ff99bsc0 with the modifications ε/ζ_{OL1}, χ_{OL4}, and β_{OL1}). Additionally, we include in Tables S1−S5 in the Supporting Information the results of simulations of the test systems using the $\varepsilon/\zeta_{OL1}+\chi_{OL4}$ modifications (without the β dihedral adjustment) that have been in use since 2013 by a large community of users (since this was the default force field in Amber previously), who may

want to better understand the implications for their own work. The total sampling of the five DNA systems in the eight combinations of force field and water model totaled 14.4 ms. In addition to the previously published extended sampling of DNA using the CHARMM 36 force field, which provides some benchmarking against the Amber force field with a different sequence,[3,29] in this work we also uncovered some older MD trajectories on the DDD system from our lab and have included root-mean-square deviation (RMSD) comparisons of average structures to the DDD NMR structures.

Crystallographic water molecules and counterions were removed in each case. This starting structure was then solvated with either the TIP3P[42] or OPC[45] water model in a truncated octahedral box using a 10 Å buffer distance between the solute and the edges of the box. Sodium counterions were added to neutralize the charge using the Joung−Cheatham[53,54] model, and an excess of NaCl was added to achieve a final excess salt concentration of ~200 mM. Ten individual copies were created for systems 1SK5, 3GGI, and 1I0T, each copy with a total MD simulation time of at least 11 μs. For the systems 1BNA and 1FZX, 100 individual copies were built, each copy with a total MD simulation time of at least 11 μs for 1BNA and 6 μs for 1FZX. In all copies of each system, the ions were randomized using CPPTRAJ: a random water molecule was swapped for an ion at least 4.0 Å from each other and no closer than 6 Å from solute atoms. Initial equilibration for each copy was achieved using incremental minimization steps in which the solute was kept fixed with a harmonic restraint of 5 kcal mol^{-1} Å^{-2} for 1 ns. The restraint was decreased to values of 0.5 and 0.1 for subsequent 1 ns equilibration time, and a final unrestrained 1 ns simulation was performed. Equilibration was performed using an integration time step value of 1 fs. Production simulations were run in the *NPT* ensemble at 300 K using Langevin dynamics (collision frequency value of 1 ps^{-1}) for temperature control.[55,56] Constant pressure was monitored using the Berendsen barostat (pressure relaxation time set to 1 ps).[57] The SHAKE methodology was used to restrain hydrogen atoms (tolerance of 0.0000001). Hydrogen mass repartitioning was used in all of the simulations, allowing an integration time step of 4 fs.[58] Periodic boundary conditions were used, and the long-range electrostatics was treated using the particle mesh Ewald methodology with a cutoff value of 10 Å and default parameters.[59,60] Aggregated trajectories used to perform the analysis were created by deleting the first 1000 ns of sampling time for each copy and concatenating the remaining frames into a single trajectory file. An example of the CPPTRAJ analysis script for the DDD system is available in Table S6. In addition to the scripts in the Supporting Information, the topologies, the raw (solvent- and ion-stripped and aggregated) trajectories, and
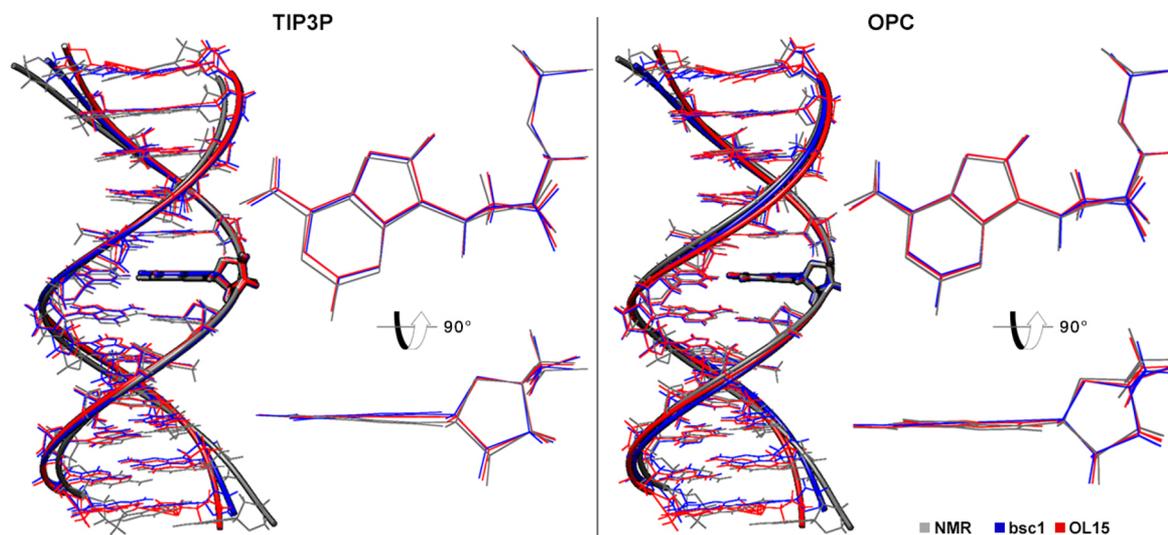
Figure 2. DDD average structures calculated from the aggregated trajectories aligned to the NMR average structure using only the heavy atoms. The detail represents A6 (shown in a slightly thicker representation in the full helix). The RMSDs of the heavy atoms of the internal eight base pairs from the MD average structures with respect to the NMR average structure are less than 1 Å for both force fields in both water models.

all of the analysis files are available for download at http://www.amber.utah.edu/FF-DNA-bsc1-OL15/.

For the CHARMM 36[15] runs, the simulation inputs were built using CHARMM c37b2 with the CHARMM 36 force field that had been altered to match Amber atom naming conventions to facilitate direct comparisons and also to add the Joung–Cheatham ion parameters where appropriate. The resulting PSF and coordinate files were converted into Amber format using the chamber program in Amber, noting that if one were to do an equivalent conversion today, use of the parmed.py program would be recommended. Similar protocols were followed except for using Amber 12 without hydrogen mass repartitioning. Two sets of simulations were performed, each with 100 independent production MD simulations from equilibrated systems using the ABC equilibration protocols starting from randomized initial ion distributions (as above)[7] with CHARMM 36, its modified TIP3P water model, and either default NaCl parameters from CHARMM or Joung–Cheatham ion parameters at ∼200 mM excess salt.[53,54] Each of the 100 independent production MD simulations on DDD for the two ion parameter sets was run for at least 1.1 μs, and the first 200 ns were omitted prior to aggregation of the trajectories. The MD simulations used 2 fs integration time steps, SHAKE on hydrogen atoms with a tolerance of 0.000001, 300 K with Langevin temperature control $(1\ ps^{-1})$, particle mesh Ewald with a 9 Å cutoff and default parameters, and constant pressure (5.0 ps coupling time).

With the exception of the CHARMM 36 runs, which were run with Amber 12 pmemd, all of the MD simulations were performed using the CPU and GPU version of pmemd as available in Amber 14.[9,10,61−63] Analysis was performed using CPPTRAJ version 16.[64] Average structures were calculated by best-fitting the DNA to the first frame followed by straight coordinate averaging over all DNA atoms over each of the aggregated trajectories. Molecular graphics were rendered in VMD,[65] and principal component pseudotrajectories visualized with the help of the Normal Mode Wizard plugin.[66]

RESULTS AND DISCUSSION

The DDD sequence is the reference benchmark system regarding B-form DNA duplex structure and dynamics because of the availability of very high resolution NMR data in the solution phase.[48] For this reason, DDD has commonly been studied as new force field parameters have been developed and optimized.[16,17,22,23,34] The average structures calculated from our simulations show strong quantitative agreement with the NMR reference (Figures 2 and S1), especially in simulations with the OPC water model, where the differences in structure among bsc1, OL15, and experiment are rather small. This is evidenced by the RMSDs of the internal eight base pairs (considering all heavy atoms but neglecting the two terminal base pairs on each end, which tend to fray on the microsecond time scale) of the average structures from MD simulations with respect to the NMR average structure (created by best-fitting and averaging the conformations that make up the NMR ensemble from the PDB file), which are less than 1 Å for bsc1 and OL15 (Tables 2 and S7). The sub-1 Å deviations of the average structures calculated from the aggregated trajectories omitting the terminal two base pairs on each end of the helix in all cases are remarkable, especially considering that the instantaneous deviations on the picosecond to nanosecond time scale are considerably larger (as a result of thermal

Table 2. Root-Mean-Square Deviations (RMSDs) (in Å) of Average Structures from Molecular Dynamics Simulations with Respect to the NMR Reference Structure of the Drew–Dickerson Dodecamer (DDD)[a]

	bsc0	bsc1	OL15	CHARMM36	CHARMM36-JC[b]
TIP3P	1.00	0.64	0.53	1.29	1.30
OPC	0.91	0.61	0.44		

[a]The average structure from simulations for each system was calculated over the full aggregated trajectory for that system; the DDD NMR reference was an average of the models in the 1NAJ structure. RMSDs were calculated over all heavy atoms of the internal eight base pairs. [b]CHARMM36-JC refers to the simulations with the Joung-Cheatham ion parameters.
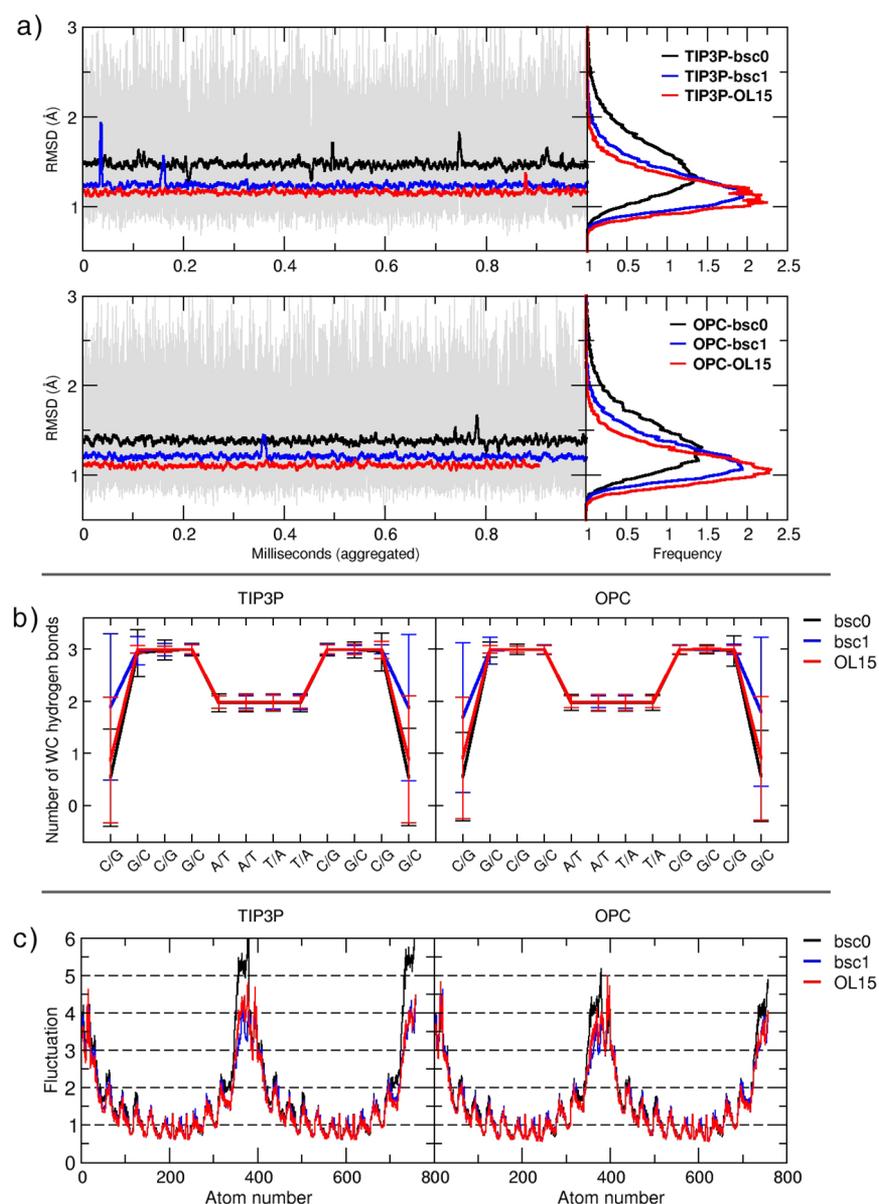
Figure 3. Root-mean-square deviations (RMSDs), Watson−Crick (WC) hydrogen bonds, and root-mean-square fluctuations (RMSFs) for DDD in both the TIP3P and OPC water models for three force fields. (a) The 1 μs running averages of the RMSD are shown in dark, solid-colored lines. while data from individual frames (every 2 ns) are shown in gray for all systems. RMSD histograms are also shown. The first and second base pairs at each end of the DNA sequence were omitted for RMSD calculations. RMSD measurements used the 1NAJ average structure as a reference. (b) Average number of WC hydrogen bonds for each base pair (canonical values are 2 for AT and 3 for GC) following the 3DNA[81] framework in CPPTRAJ using the full aggregated trajectory; error bars refer to standard deviations. It should be noted that numbers of hydrogen bonds less than 2 are possible as a result of fraying events. (c) RMSF (in Å) using the entire aggregated trajectory with respect to the average structure for each system.

fluctuations, as observed in the RMSD vs time plots in Figure 3). The RMSDs of the average structures from TIP3P MD with respect to the NMR reference are 1.00, 0.64, and 0.53 Å for bsc0, bsc1, and OL15, respectively, and those for OPC are 0.91, 0.61, and 0.44 Å for bsc0, bsc1, and OL15, respectively. The RMSD over time also shows improvements with bsc1 and OL15 compared with bsc0, noting that the instantaneous RMSD values of the MD snapshots are higher than the values reported above as a result of thermal fluctuations; the improvement occurs with both water models (Figure 3a). The RMSDs of the average structures (also omitting the two terminal base pairs on each end) from the CHARMM 36 simulations, with either the CHARMM ion parameters or the Joung−Cheatham ion parameters for NaCl at ∼200 mM excess,

with respect to the NMR reference are 1.29 and 1.30 Å, respectively (Table 2 and Figure S2). All of the helical parameters for the CHARMM simulations are shown in Table S8.

Visual inspection of the trajectory files shows stable duplex structures throughout the entire simulation time, and consistent with previous MD simulation reports, transient fraying events of the first and second base pairs are observed on either side of the DNA chain. These fraying events are evidenced by the short-lived bumps in the RMSD versus time plots[3] and have been well-characterized in previous work, where it was observed that the fraying of the terminal base pairs can cause long-lived noncanonical structure conformations that affect the end results (see, e.g., Figure 6 in the article by Zgarbová et al.[23] and

Table 3. Average Structural Parameters for the DDD System Obtained by Averaging Values Calculated from the Trajectory Snapshots from the 1 ms Aggregated Trajectories for Each Combination of Force Field and Water Model[a]

	exptl		TIP3P			OPC		
	NMR	X-ray	bsc0	OL15	bsc1	bsc0	OL15	bsc1
shear/Å	0.0	0.0	0.0 ± 0.11	0.0 ± 0.10	0.0 ± 0.11	0.0 ± 0.11	0.0 ± 0.10	0.0 ± 0.10
stretch/Å	−0.34	−0.23	−0.2 ± 0.04	−0.02 ± 0.04	−0.02 ± 0.04	−0.02 ± 0.04	−0.03 ± 0.04	−0.02 ± 0.04
stagger/Å	−0.12	0.10	0.06 ± 0.16	0.0 ± 0.16	0.05 ± 0.16	0.07 ± 0.16	0.02 ± 0.15	0.06 ± 0.15
buckle/deg	0.02	0.2	−0.02 ± 4.92	0.03 ± 4.64	0.01 ± 4.52	−0.04 ± 4.59	0.04 ± 4.37	0.01 ± 4.32
propeller/deg	−17.58	−13.3	−12.7 ± 3.09	−12.26 ± 2.72	−10.38 ± 2.89	−12.59 ± 2.85	−11.87 ± 2.62	−9.91 ± 2.76
opening/deg	−1.10	1.31	0.3 ± 1.19	0.1 ± 1.76	−0.41 ± 1.70	0.18 ± 1.73	0.08 ± 1.66	−0.46 ± 1.64
shift/Å	0.0	0.0	0.0 ± 0.19	0.0 ± 0.19	0.0 ± 0.20	0.0 ± 0.17	0.0 ± 0.18	0.0 ± 0.19
tilt/deg	0.01	0.3	0.0 ± 1.26	0.0 ± 1.23	0.01 ± 1.26	0.0 ± 1.21	0.01 ± 1.20	0.01 ± 1.22
slide/Å	−0.21	0.07	−0.4 ± 0.28	0.0 ± 0.26	−0.27 ± 0.25	−0.46 ± 0.26	−0.12 ± 0.25	−0.33 ± 0.24
rise/Å	3.2	3.3	3.31 ± 0.07	3.32 ± 0.06	3.31 ± 0.07	3.30 ± 0.07	3.31 ± 0.06	3.30 ± 0.6
roll/deg	3.03	1.98	3.5 ± 2.13	2.8 ± 1.99	2.28 ± 1.89	2.91 ± 2.06	2.10 ± 1.98	2.10 ± 1.84
twist/deg	35.7	34.2	32.86 ± 1.59	35.21 ± 1.32	34.64 ± 1.43	33.37 ± 1.55	35.55 ± 1.34	34.73 ± 1.43
X displacement/Å	−0.81	−0.23	−1.48 ± 0.66	−0.56 ± 0.54	−0.94 ± 0.57	−1.46 ± 0.62	−0.61 ± 0.53	−1.01 ± 0.56
Y displacement/Å	0.0	0.1	0.0 ± 0.37	0.0 ± 0.28	0.0 ± 0.30	0.0 ± 0.32	−0.01 ± 0.26	0.0 ± 0.29
helical rise/Å	3.18	3.29	3.18 ± 0.14	3.29 ± 0.10	3.26 ± 0.12	3.18 ± 0.13	3.27 ± 0.10	3.24 ± 0.11
helical inclination/deg	5.0	4.0	6.72 ± 4.11	4.88 ± 3.44	4.26 ± 3.40	5.79 ± 3.89	3.68 ± 3.42	3.97 ± 3.32
tip/deg	0.0	−0.7	0.0 ± 2.48	0.0 ± 2.12	−0.01 ± 2.22	−0.01 ± 2.29	−0.01 ± 2.05	−0.02 ± 2.15
helical twist/deg.	36.0	34.6	34.07 ± 1.50	36.19 ± 1.27	35.61 ± 1.38	34.47 ± 1.44	36.43 ± 1.28	35.64 ± 1.38
major width/Å	19.56	19.12	19.88 ± 0.33	19.34 ± 0.26	19.51 ± 0.29	19.95 ± 0.28	19.41 ± 0.25	19.58 ± 0.26
minor width/Å	12.2	12.2	12.51 ± 0.20	12.26 ± 0.15	12.37 ± 0.17	12.54 ± 0.17	12.27 ± 0.14	12.37 ± 0.14
pucker/deg	137.1	129.5	130.5 ± 8.21	148.3 ± 6.02	149.7 ± 6.94	130.7 ± 7.78	148.5 ± 5.83	149.7 ± 6.43
α/deg	298.8	299.2	288.6 ± 4.60	289.5 ± 4.37	285.6 ± 6.03	289.3 ± 4.69	290.5 ± 4.32	286.1 ± 6.18
β/deg	172.4	175.7	168.1 ± 3.92	167.3 ± 5.16	164.4 ± 6.20	168.4 ± 4.02	169.6 ± 5.27	165.5 ± 6.07
γ/deg	50.28	56.52	57.3 ± 4.32	54.0 ± 3.58	60.1 ± 7.68	57.4 ± 4.53	53.8 ± 3.50	59.7 ± 7.58
δ/deg	126.7	122.8	121.0 ± 4.63	131.6 ± 3.47	134.6 ± 3.87	120.8 ± 4.50	131.4 ± 3.38	134.4 ± 3.70
ε/deg	188.5	190.4	197.7 ± 7.57	199.0 ± 7.56	201.3 ± 5.34	194.8 ± 7.23	194.8 ± 7.90	199.8 ± 5.49
ζ/deg	257.1	251.3	254.6 ± 9.18	244.6 ± 7.52	245.6 ± 6.64	256.6 ± 8.49	247.5 ± 7.59	245.9 ± 6.70
χ/deg	249.2	243.9	241.7 ± 5.09	251.3 ± 3.77	247.5 ± 3.86	241.0 ± 4.82	250.9 ± 3.66	247.4 ± 3.75

[a]The two terminal base pairs on each side of the DDD were excluded for the average value calculation. The standard deviations are shown alongside the averages. NMR data are average values from the PDB 1NAJ structure, and the X-ray data were calculated from the PDB 1BNA structure.

additional work by Dršata et al.[67]). No fraying or internal base pair opening is observed on the time scales of the simulations where two Watson−Crick (WC) hydrogen bonds are observed in the central AATT region and three WC bonds are observed with G4C21 and C9C19 (Figure 3b). This is consistent for the three force fields and expected on the basis of the low RMSD values and visual inspection showing stable duplexes. Maintenance of the internal base pairs is expected since internal base pair opening times are on the order of milliseconds.[68−70] The terminal base pairs, however, show decreased numbers of average WC hydrogen bonds instead of the canonical value of three. This is caused by multiple fraying events that break canonical hydrogen bonds and lead to distorted conformations such as trans WC/sugar edge (tWC/SE), stacked, and other observed configurations.[67,71] Although fraying is more frequent with bsc0 and least frequent in bsc1, it is difficult to ascribe less frequent terminal base pair opening to "better" behavior since the characteristics and frequency of terminal base pair openings are not well-characterized in experiments on these short time scales. We did, however, calculate two of the primary structures observed in fraying, tWC/SE and WC. OL15 was found to populate WC structures in ~23% of the frayed frames and tWC/SE in ~64% of the frayed frames. Conversely, bsc1 populated tWC/SE in ~16% and WC in ~60% of the frayed frames. As fraying tends to distort the canonical structure, for MD simulations on time scales that are currently routine (ns to µs), better representation of the expected structure can be anticipated

from simulations that fray less. The fraying frequency trend with the various force fields is observed regardless of the water model applied. However, more variation is detected for the TIP3P water model, even up into the third base pair from each end of the DNA. As this affects the overall structure (in the absence of more complete sampling of the terminal base pairs) of the backbone and the base pairing, perhaps simulations with OPC may be recommended in shorter simulations (where we cannot completely sample); however, this involves a trade-off due to the increased costs. In the short term, to avoid end effects from incomplete sampling of fraying events, weak restraints can be applied to maintain the base-pair hydrogen bonds on the termini. Nevertheless, realistically in the longer term, since terminal base pair fraying is real and could influence results, for example in comparison with SAXS data, where the frayed bases could contribute to the SAXS signature, sampling should ultimately be sufficient to capture and reliably represent terminal base pair fraying.

The root-mean-square fluctuations (RMSFs), representing the fluctuations about the average structure for each tested force field, are presented in Figure 3c and show nearly complete agreement for the inner base steps. A small increase in atomic fluctuations is detected for the bsc0 force field at the terminal base pairs; this is in agreement with increased fraying as measured by the number of WC bonds for each base pair (as previously discussed). The more frequent fraying events observed with the TIP3P water model for both bsc1 and OL15 could partially explain why the simulations using OPC
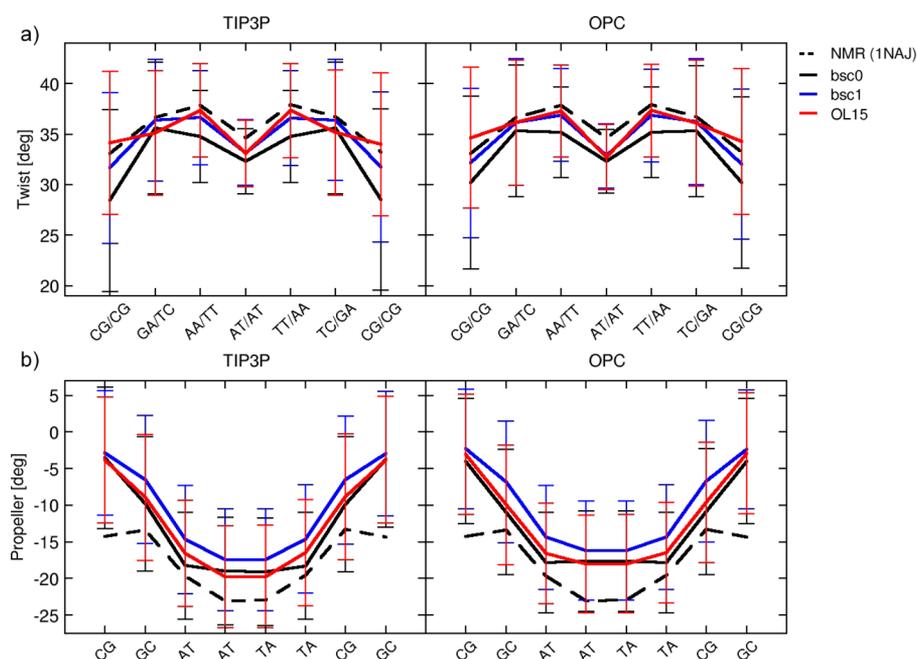
Figure 4. (a) Twist values for individual base-pair steps and (b) propeller values at each base pair for the DDD system. MD simulations using the different force field modifications are compared with the experimental NMR structure (dashed line). The entire aggregated trajectories (1 ms) were used, and the last two terminal base pairs at each end were not considered for the analysis. Values from simulations are averaged over the aggregated trajectory, and error bars show the standard deviations.

present an overall better agreement when compared to experimental structures.

Average values for structural, interhelical, and intrahelical properties from our simulations using bsc0 and the optimized bsc1 and OL15 force fields using two water models for the DDD are presented alongside NMR and X-ray reference values in Table 3. Overall, and as already suggested, both the bsc1 and OL15 force field modifications generate structures that are in better agreement with experimental structural properties than the earlier bsc0 force field. The most notable differences from the experimental values are observed in the propeller values. It is important to note the sensitivity of the helical parameters to the overall geometry: the average structures obtained with bsc1 and OL15 are less than 1 Å apart in RMSD (0.63 and 0.81 Å, respectively, with TIP3P and OPC over all atoms), yet some of the helical parameters differ to a larger degree than perhaps expected on the basis of the similarity of the average structures. This could in part reflect the larger standard deviations observed in the values that tend to vary. To ensure that the differences were not due to averaging of values over snapshots compared with calculating helicoidal parameters from average structures, we calculated the helicoidal parameters from the aggregate average structures over the trajectory, and these agreed with the reported values with an r^2 of 0.99.

To further investigate the similarities and differences of the force fields when compared to each other and to the NMR reference, we calculated p values from a two-sample t test for every backbone dihedral angle, base pair, and base-pair step helical property of DDD. This statistical analysis was performed because many of the helical properties showed very close agreement between bsc1 and OL15, and we wanted to be able to say with confidence whether the differences in the mean values obtained from our analysis were significantly different. The results are summarized in Table S9. Overall the p values indicate that the differences in the mean values of the dihedral

angles and helical parameters determined from MD with the bsc1 and OL15 force fields are statistically significant. Additionally, when the force field results are compared to the NMR data, both bsc1 and OL15 have p values showing that shear, shift, and tilt have mean values that are *not* significantly different from the NMR data. In this respect bsc1 outperformed OL15 in stagger and opening, but OL15 had better agreement with experiment for the α, γ, and δ dihedrals as well as slide. Each force field had generally higher p values in OPC, which agrees with the overall better agreement seen in the RMSDs of average structures. Plots showing all of the DDD structural parameters can be found in Figures S3−S6.

Underestimation of twist has been a concern since the Cornell et al. conception of the Amber DNA force field and the subsequent bsc0 modification.[72] The twist values predicted by both bsc1 and OL15 are on average closer than the bsc0 values to the observed experimental values (35.7° for NMR and 34.2° for X-ray). Twist of the eight internal base pairs shows better agreement for OL15 in TIP3P, with a value of 35.2° compared with 34.6° for the bsc1 modification. Regardless of the water model, and despite these small differences, both force field modifications represent improvement over bsc0 with respect to the twist structural parameter for DNA; this will have significant implications for simulation of DNA circles.[73−75] Base-step detail of the twist value provides further information (Figure 4a). The data show that bsc0 is consistently off in almost every step compared with both bsc1 and OL15. Symmetrical behavior of the per-base-step twist value around the central step is seen, characteristic of a palindromic sequence. This symmetrical feature is obtained only in a fully converged ensemble of simulations, as presented in this work. The twist value in CpG steps is underestimated by bsc1 and overestimated by OL15 regardless of the water model used but is closer to the reference value in both cases compared with bsc0. GpA base steps stand in very good agreement for both
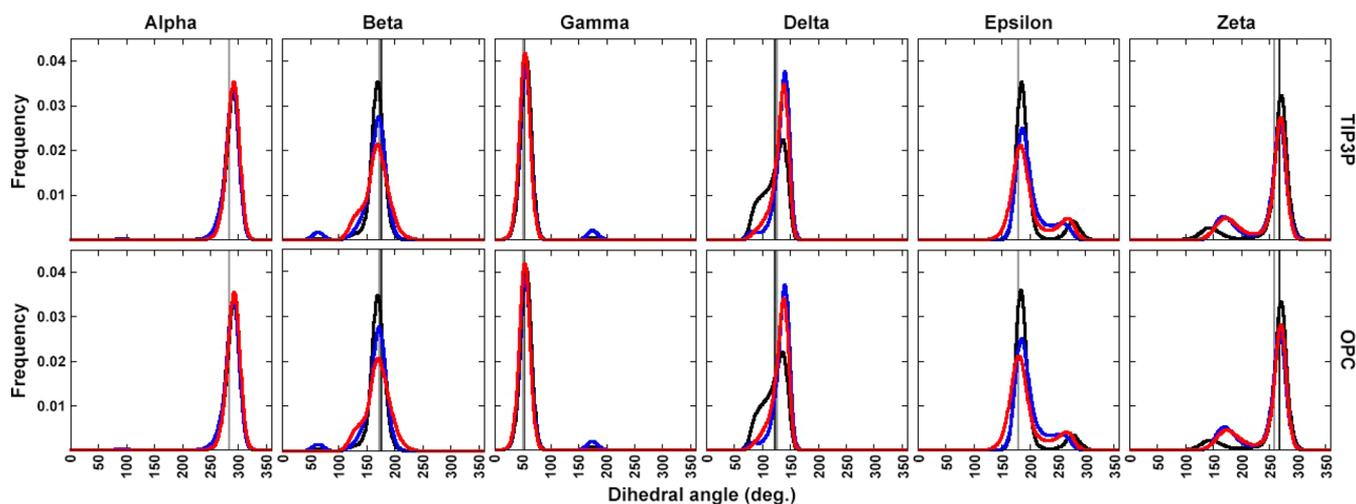
Figure 5. Populations of the six DNA backbone dihedrals of DDD DNA using the full 1 ms aggregated trajectories. Only the inner eight base pairs are included in the analysis. Black, blue, and red lines correspond to bsc0, bsc1, and OL15, respectively. The X-ray and NMR references are shown as gray and black vertical lines, respectively.

force fields, especially with the OPC water model, where the values are within 0.3° of the reference value. The same behavior is seen in ApA base steps, where the values for bsc1 and OL15 for OPC are off by 0.1° and 0.3°, respectively. For ApT base steps, the bsc1 modification has a value of 33.2° and OL15 a value of 33.1° with TIP3P (the OPC values are 32.8° and 32.7°, respectively), which makes them nearly identical yet smaller than the NMR experimental value of 34.6°. Propeller twist in the DDD system is consistently off by ∼5° for OL15 and ∼7° for bsc1 (assuming the NMR reference). A per-base-pair comparison shows a general overestimation (Figure 4b), especially for base pairs near the end of the DNA chain. The high propeller twist near the termini is likely due to fraying effects influencing the structure and modifying the base pairing. Improvement is also obtained in both grooves with differences of ∼0.1−0.3 Å for bsc1 and OL15 regardless of water model used. This is expected after the improvements in the majority of the helical parameters. Comparisons with NMR observables as presented in Table S10 are consistent with our current discussion, as fewer nuclear Overhauser effect (NOE) violations are rendered by the updated force field modifications in comparison with bsc0 statistics.

Distributions of backbone dihedrals for DDD are displayed in Figure 5. Close agreement with the experimental value and between the compared force fields is observed for the α and γ dihedrals. The OL15 modifications that involve the ε/ζ, χ, and β dihedrals appear to have completely reduced the low trans population in the γ dihedral that is evident in bsc1 (in a very low population, however). The bsc1 modification for the β dihedral is in good agreement with the NMR value of 172.4° and shows an increased population of close to 60° in comparison with bsc0. The β dihedral representation for OL15 is also in good agreement with the experimental value and displays the low population at ∼130°, which was observed in previous work[22] and aids in the BI/BII substate balance.[76] The δ dihedral, which is part of the furanose ring (ν_3 dihedral), presents improvement over the earlier bsc0 version in both bsc1 and OL15, although it is slightly overestimated by 5° and 8°, respectively. Good agreement with experimental values is also detected in the ε/ζ dihedrals for bsc1 and OL15, and both display increased populations in the gauche− region for ε and

trans for ζ, corresponding to an increased percentage of the DNA BII state. The fraction of the ensemble in the BII substate was calculated by taking the $\varepsilon - \zeta$ difference; frames that had $\varepsilon - \zeta > 0$ were considered to be in the BII state. The fractions of BII population for the DDD systems are shown in Table 4 and

Table 4. Comparison of the Average BII Fractions for the DDD, 1FZX, 1SK5, and 3GGI Systems; The Values for the 1I0T System Represent the Fractions of the ZI State As Measured for 5′-GpC-3′ Steps[a]

	TIP3P			OPC		
	bsc0	OL15	bsc1	bsc0	OL15	bsc1
DDD	0.12	0.23	0.22	0.10	0.21	0.21
1FZX	0.09	0.18	0.17	0.08	0.15	0.16
1SK5	0.15	0.20	0.17	0.11	0.18	0.15
3GGI	0.20	0.28	0.21	0.15	0.27	0.22
	ZI Population					
1I0T	0.29	0.79	0.08	0.24	0.80	0.06

[a]ZI is defined as follows: ζ(dG) is between 240 and 360° and β(dC) is between 205 and 300°. All frames were considered for BII and ZI analysis.

display an almost identical increment in the BII population for both bsc1 and OL15 in both water models used. Per-base-step analysis (Figure 6) shows similar BII estimates for OL15 and bsc1. Improvements over bsc0 at the CpG and GpC steps at both ends of the DDD are evident, although under-representation of the central AT, TT, and TC is still present in both bsc1 and OL15.

To directly compare the modes of motion present in the simulations, principal component analysis (PCA) over all heavy atoms of the internal eight base pairs of each trajectory was performed and the principal component projections are shown in Figure 7. The combined PCA consisted of calculating the covariance matrix from both trajectories of the two force fields being compared (an example of a CPPTRAJ analysis script can be found in Bergonzo et al.[13]). This technique provides insight into the collective dynamics that are sampled during a simulation and is able to rank the contributions of the individual modes of motion.[29] Each of the histograms
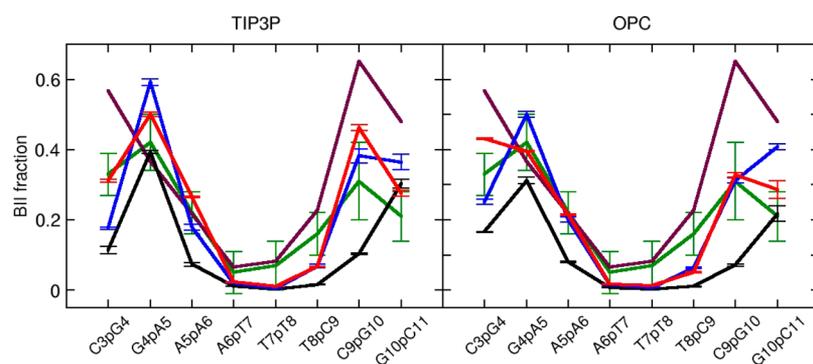
Figure 6. Comparison of the BII fractions for different base steps for the DDD system using bsc0 (black), bsc1 (blue), and OL15 (red) with NMR data (green, Tian et al.;[82] maroon, Schwieters and Clore;[83] N_e = 8).
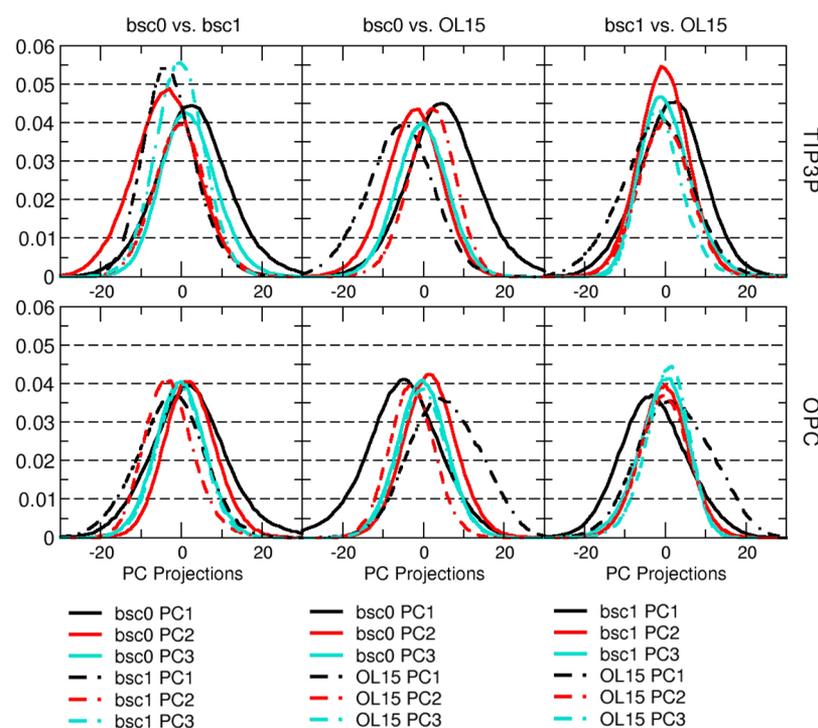
Figure 7. Principal component projections for the internal eight base pairs of the DDD simulations and both water models. Refer to the main text for a detailed discussion.

represents the projection of an individual mode of motion for each force field tested. For clarity, only the first three projections of the principal components are shown. These three projections contribute ~80% of the eigenvalues and hence ~80% of the overall motion of the system. Trajectories that explore equivalent dynamical processes will have overlapping PC projections. The best overlap for the TIP3P systems is observed with bsc1 and OL15, which is expected since bsc1 and OL15 consist of newer and updated parameters and on the basis of the agreement with average properties shown in Table 3. However, the poor overlap between the bsc0/bsc1 and bsc0/OL15 projections led us to explore the principal modes in the individual force field trajectories and compare the top modes through visualization of the pseudotrajectories. These motions can be observed in the video file available in the Supporting Information. We found that the top mode for each force field was twisting. The bsc0 and bsc1 projections showed agreement with the second mode, which involves bending toward the major groove (asymmetric bend), while this motion was the

third-ranked mode for OL15. Again, bsc0 and bsc1 exhibited agreement with the third mode, another bending mode that is more symmetric than the previous one. OL15 demonstrated this mode as well, but it was the second-ranked mode rather than the third-ranked one. The eigenvalue fractions were 0.39, 0.35, and 0.40 for the twisting mode, 0.21, 0.21, and 0.16 for the asymmetric bend; and 0.19, 0.19, and 0.22 for the symmetric bending mode for bsc0, bsc1, and OL15, respectively. The top three modes showed nearly equivalent dynamics in the three force fields, as shown in the supporting video, but the ranking of the bending modes for OL15 differed from that for the other parameter sets.

Additional systems were chosen to evaluate the ability of bsc0, bsc1, and OL15 to accurately model duplex DNA with different sequences and to show that the parameter sets have not been overtrained to reproduce DDD structural data. We found two B-form duplex DNA X-ray structures with sub-1 Å resolution (1SK5 and 3GGI) and a Z-DNA structure (1I0T) with sub-2 Å resolution in the Protein Data Bank. It is
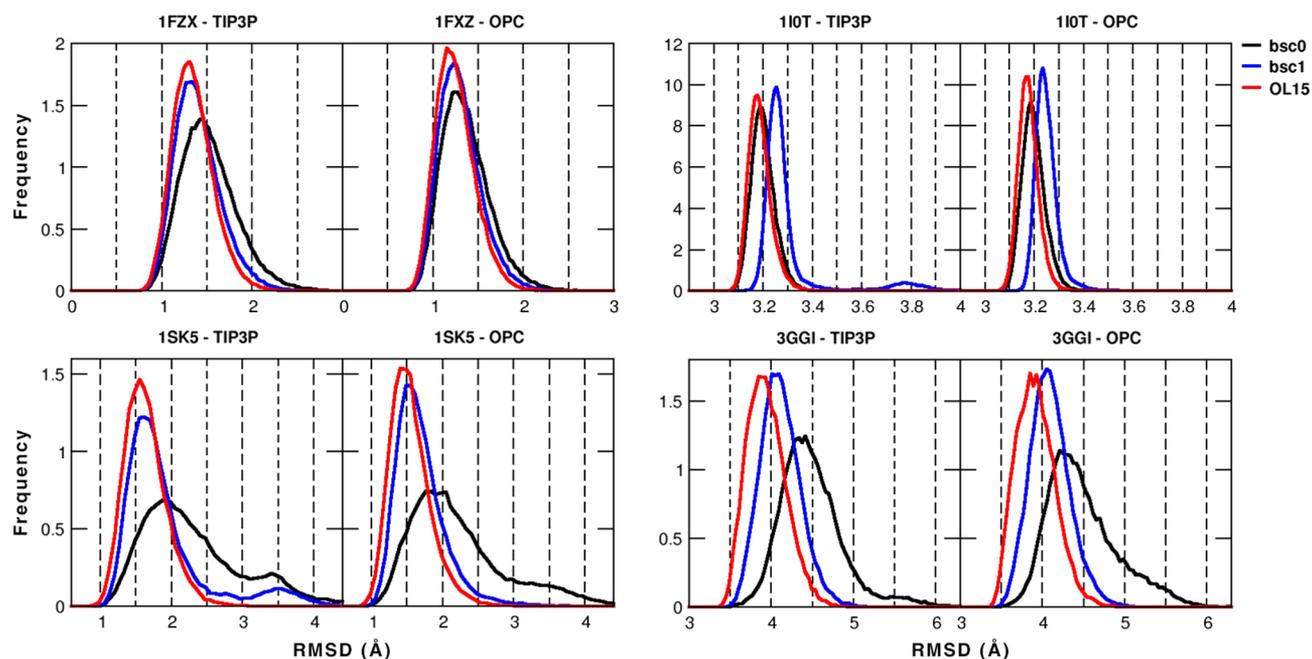
Figure 8. RMSD histograms for three B-form DNA duplexes (PDB entries 1FZX, 1SK5, and 3GGI) and one Z-form DNA duplex (PDB entry 1I0T). Values were calculated using the aggregated trajectories for each system, not considering the first 100 ns for each individual copy. The RMSD values were obtained using the original experimental structure as the reference and were calculated over the internal base pairs omitting a single terminal base pair on each end.

important to mention that solution-phase MD simulations should not necessarily reproduce crystal structures perfectly because of the absence of crystal packing and crystallization conditions,[50,77] which will lead to small but noticeable variations in the conformation; with that in mind, we also included the 1FZX model that was obtained by NMR spectroscopy and provides a better comparison to MD simulations run in solution than structures from X-ray crystallography. In the same way as for the DDD NMR structure already mentioned, we calculated an average structure from the 10 submitted conformers of 1FZX available in the PDB and used this average structure as a reference. RMSD histograms for each of the tested systems are shown in Figure 8. Overall, as previously detected with the DDD analysis, both bsc1 and OL15 have lower RMSD values, which translate to better agreement with the reference experimental structure. Average structure information for all of the systems is available in Tables S2−S5, and per-base parameters are available in Figures S7−S22.

For the case of the 1FZX system (NMR structure), both bsc1 and OL15 have lower RMSDs of the MD average structure with respect to the NMR reference compared with bsc0 (see Table S7). The agreement with experiment (as evidenced by a slight shift in the RMSD histograms in Figure 8 compared with Figure 3A) does not appear to be as good as with the high-resolution DDD structure; this could be due to the fact that the DDD PDB structure (1NAJ) was solved with considerably more NMR residual dipolar coupling and ^{31}P chemical shift anisotropy restraints, which arguably leads to a higher-resolution structure. With the 1FZX system, the OPC water model increases the performance of bsc0, which is now close to bsc1 (Figure 8). This appears to be caused by increased twist values in the central base pairs, effectively reducing the population of structures with RMSDs greater than ∼2 Å. Propeller twist does not appear to be as influenced by the OPC

water model as the helical twist (Figure 9 and the Supporting Information). As previously discussed in the DDD case, an underestimation of the population of the BII state with bsc0 has been observed. This DNA substate is increased considerably and in a similar fashion for bsc1 and OL15 (Table 4).

For the 1SK5 bsc1 and OL15 trajectories, the average RMSDs of the MD snapshots are within ∼1.4−1.6 Å of the reference regardless of the water model (Figure 8), and the OPC simulations considerably reduced the population of structures with RMSDs of ∼3.5 Å for bsc1 seen in the TIP3P trajectories. In general, the RMSDs of the MD average structures with respect to the X-ray reference were reduced in OPC relative to TIP3P, except in the case of the bsc0 force field (Table S7). The twist value shows improvement over bsc0 for TT and AA base steps and an overestimation of TA base steps of ∼4°. Propeller twist is not improved by bsc1 (Figure 9), which shows underestimation for TA and AT base pairs (in both water models). The bsc0 and OL15 force fields render values very close to each other and also a general under-estimation.

The 3GGI system saw improved results with bsc1 and OL15 compared with bsc0, although the RMSDs with respect to the reference structure remained higher than in the previously discussed systems. The averaged RMSDs of snapshots from the aggregate trajectories for bsc1 and OL15 with respect to the experimental structure are in the range of ∼3.8−4.1 Å for both TIP3P and OPC (Figure 8). Similar trends were observed in the RMSDs of the average structures from MD with respect to the X-ray reference (Table S7), where the lowest RMSDs were seen with OL15 for both TIP3P and OPC. Twist is underpopulated in bsc0 as previously discussed, with both bsc1 and OL15 in better overall agreement with experiment. GG and GC steps are off by ∼5° in both cases and with both water models (Figure 9). Propeller, on the other hand, is slightly over-represented by bsc1 and under-represented by
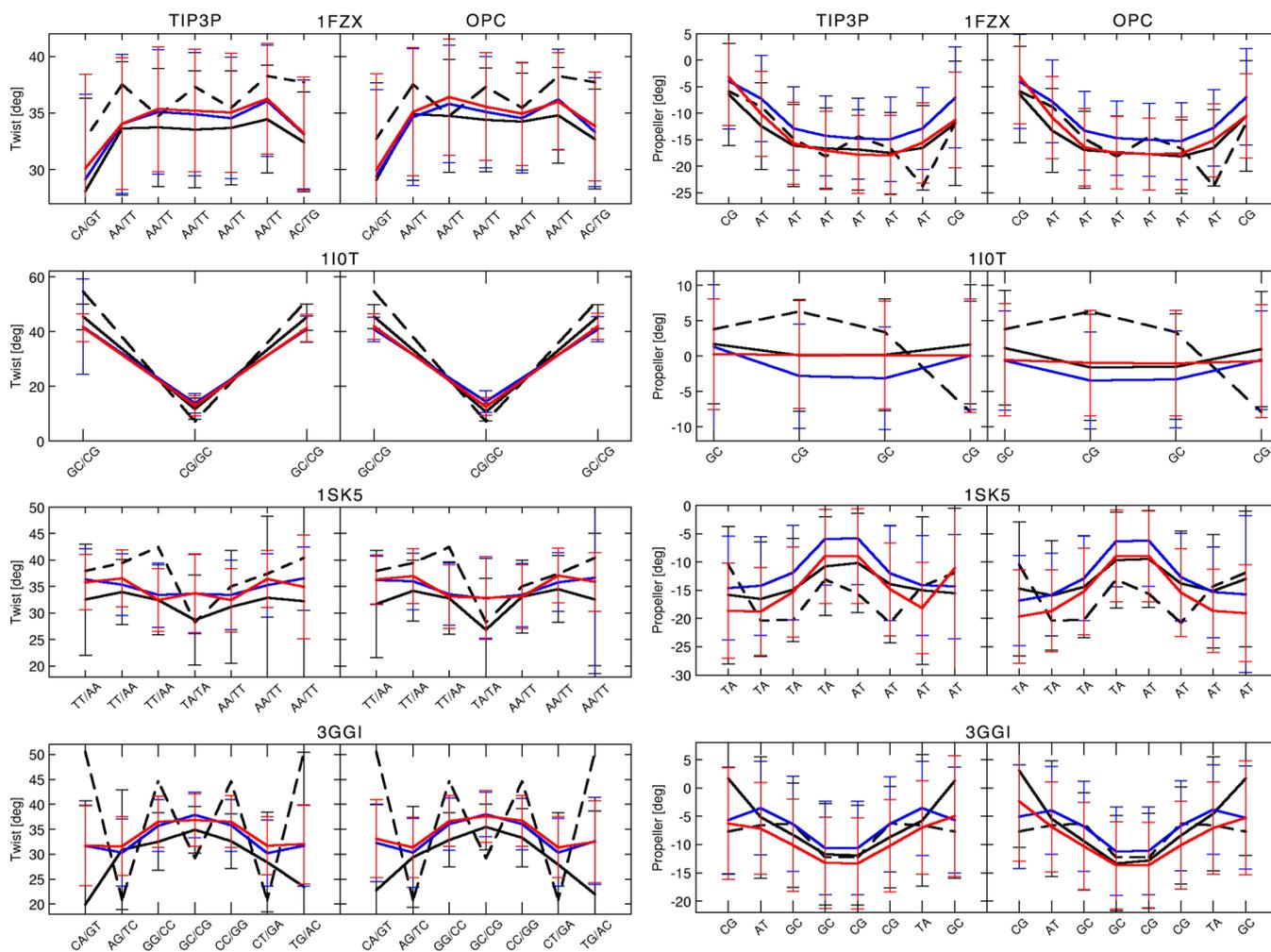
Figure 9. Twist and propeller for individual base-pair steps for the 1FZX, 1I0T, 1SK5, and 3GGI systems. TIP3P and OPC are shown in the left and right columns, respectively. The black dashed line is the X-ray reference in each case (NMR for 1FZX); the black, blue, and red solid lines are for bsc0, bsc1, and OL15, respectively. The internal seven dinucleotide steps of 1FZX, 1SK5, and 3GGI and the internal three steps of 1I0T were considered for twist calculations. The internal seven base pairs of 1FZX, 1SK5, and 3GGI and the internal four base pairs of 1I0T were considered for propeller calculations.

OL15, but still, both force field modifications are within ∼2° of the observed experimental value in each base pair. The BII population is almost unchanged in going from bsc0 to bsc1 in TIP3P, while OL15 showed an increase (Table 4). The BII fraction dropped for bsc0 in OPC but remained nearly the same for bsc1 and OL15 in the two water models.

The last system we included in the analysis is the Z-DNA structure 1I0T, which, as mentioned in the Introduction, is notoriously difficult for current MD simulations and force fields.[17] The structure we tested began in the Z form, and no transitions to B form were observed, even though the salt concentration was not ideal for Z-DNA. Unbiased molecular dynamics simulations provide a limited picture of this uncommon configuration because of the large energy barriers between Z-DNA and B-DNA.[78,79] Hence, simulations of Z-DNA by current nonpolarizable all-atom force fields should be considered with caution. Regardless, the RMSD histograms show that bsc0 and OL15 sampled configurations closer to the reference structure than bsc1 (Figure 8). The RMSDs of the MD average structures with respect to the X-ray reference for this system (Table S7) were unusual because bsc0 and OL15 showed almost perfect agreement (both sub-1 Å), rather than an improvement for OL15 as observed in all of other systems,

and bsc1 performed worse than bsc0. This demonstrates the difficulty of modeling Z-DNA. The twist values are in good agreement with the experimental reference, showing a difference of ∼2° for the central CG/GC step (Figure 9), which is not the case with the propeller twist. The internal CG and GC values are off by ∼5−6° for the three tested force fields, which helps explain the high RMSDs observed before. One of the main goals of the last torsion adjustment for OL15 (β_{OL1}) was to influence the ZI/ZII equilibrium.[22] Values for these ZI substates are presented in Table 4. The OL15 modification generates a significant increase in the ZI population compared with bsc0, while bsc1 performs poorly with the Z-DNA system. This indicates that the β refinement, which is entirely absent in bsc1, may significantly improve the description of some noncanonical DNA and may be potentially important for deformed structures.

■ CONCLUSIONS

We have assessed the performance of two recent modifications to the Amber force field that were designed to enhance the representation of double-stranded DNA. The individual modifications reviewed here (bsc1 and OL15) improve the

overall performance of DNA simulations and are to be considered a general force field for most DNA systems. No force field is perfect, and as experience over the years has suggested, we doubt that there will be a universal force field capable of representing the enormous structural diversity of biopolymers. For example, inclusion of polarization terms that allow the representation of the dynamic redistribution of electronic charge over time may lead to an improved force field, but at an increased computational cost.[80] Regardless, sub-1 Å, and in some cases sub-0.5 Å, agreement with the average NMR structures is remarkable. This work concludes that both the OL15 and bsc1 force field modifications increase the accuracy of representing averaged structures of DNA compared with the earlier models and recommends that users switch to either the bsc1 or OL15 force field for MD simulations of DNA with the Amber force field. Benchmarking of the CHARMM 36 force field[15] with DNA has been previously explored by our group with a different sequence[29] and is reported here with DDD and two ion parameter sets, each with sampling times extending to over 90 μs of aggregated simulations. The results show RMSDs of the MD average structures with respect to the NMR reference of ~1.3 Å, which is good agreement but less optimal than the sub-1 Å deviations of the average structures with bsc1 and OL15. Notable differences compared with experiment for the CHARMM 36 results are twist values below 33°, larger roll, increased inclination, reduced major groove width ,and shifts in the ε and ζ dihedrals. Consideration of the results suggests that these latest Amber force fields better match the NMR DDD structure. The millisecond-length aggregated trajectories for DDD in each force field/water model combination presented in this work, combined with four additional systems also with extensive sampling, demonstrate an exhaustive collection of data to compare the latest DNA force fields for Amber. The force fields show a convincing improvement over bsc0 and strong overall structural agreement with each other, but with many finite and subtle differences in average helical parameters that suggest a recommendation of one force field over the other is not obvious at present.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00186.

> Tables of DNA structural data, an example analysis script, and figures of DNA helicoidal parameters for all systems (PDF)
> Animation of the first three principal components from the PCA of d(CGCGAATTCGCG) (MPG)

AUTHOR INFORMATION

Corresponding Author
*E-mail: tec3@utah.edu.

Author Contributions
||R.G.-M. and J.C.R. contributed equally.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Dans, P. D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M. Unraveling the Sequence-Dependent Polymorphic Behavior of d(CpG) Steps in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 11304−11320.

(2) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179−2183.

(3) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III. On the Absence of Intra-Helical DNA Dynamics on the μs to ms Timescale. *Nat. Commun.* **2014**, *5*, 5152.

(4) Robertson, J. C.; Cheatham, T. E., III. DNA Backbone BI/BII Distribution and Dynamics in E2 Protein-Bound Environment Determined by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119*, 14111−14119.

(5) Pasi, M.; Maddocks, J. H.; Beveridge, D. L.; Bishop, T. C.; Case, D. A.; Cheatham, T. E., III; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; Mitchell, J.; Osman, R.; Orozco, M.; Pérez, A.; Petkevičiūtė, D.; Spackova, N.; Sponer, J.; Zakrzewska, K.; Lavery, R. μABC: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272−12283.

(6) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps. *Biophys. J.* **2004**, *87*, 3799−3813.

(7) Beveridge, D. L.; Cheatham, T. E., III; Mezei, M. The ABCs of Molecular Dynamics Simulations on B-DNA, circa 2012. *J. Biosci.* **2012**, *37*, 379−397.

(8) Shaw, D. E.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Lerardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Deneroff, M. M.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91−97.

(9) 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.

(10) 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*, 1542−1555.

(11) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(12) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

(13) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., III. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492−499.

(14) Roe, D. R.; Bergonzo, C.; Cheatham, T. E., III. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2014**, *118*, 3543−3552.

(15) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; Mackerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.

(16) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13*, 55−58.

(17) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E., III; Jurečka, P.; Sponer, 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*, 2506−2520.

(18) Langley, D. R. Molecular Dynamic Simulations of Environment and Sequence Dependent DNA Conformations: The Development of the BMS Nucleic Acid Force Field and Comparison with Experimental Results. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487−509.

(19) Lavery, R.; Zakrzewska, K.; Sklenar, H. JUMNA (Junction Minimisation of Nucleic Acids). *Comput. Phys. Commun.* **1995**, *91*, 135−158.

(20) Soares, T. A.; Hünenberger, P. H.; Kastenholz, M. A.; Kräutler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; Van Gunsteren, W. F. An Improved Nucleic Acid Parameter Set for the GROMOS Force Field. *J. Comput. Chem.* **2005**, *26*, 725−737.

(21) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., III; 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*, 2886−2902.

(22) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E., III; 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*, 5723−5736.

(23) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E., III; Otyepka, M.; Jurečka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.

(24) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(25) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

(26) Šponer, J.; Mládek, A.; Šponer, J. E.; Svozil, D.; Zgarbová, M.; Banáš, P.; Jurečka, P.; Otyepka, M. The DNA and RNA Sugar-Phosphate Backbone Emerges as the Key Player. An Overview of Quantum-Chemical, Structural Biology and Simulation Studies. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257−15277.

(27) 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.

(28) Dans, P. D.; Danilāne, L.; Ivani, I.; Dršata, T.; Lankaš, F.; Hospital, A.; Walther, J.; Pujagut, R. I.; Battistini, F.; Gelpí, J. L.;

Lavery, R.; Orozco, M. Long-Timescale Dynamics of the Drew-Dickerson Dodecamer. *Nucleic Acids Res.* **2016**, *44*, 4052−4066.

(29) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III. Convergence and Reproducibility in Molecular Dynamics Simulations of the DNA Duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1041−1058.

(30) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(31) Kollman, P. A.; Dixon, R.; Cornell, W. D.; Fox, T.; Chipot, C.; Pohorille, A. The Development/Application of the "Minimalist" Organic/Biochemical Molecular Mechanic Force Field Using a Combination of ab Initio Calculations and Experimental Data. In *Computer Simulations of Biomolecular Systems*; Van Gunsteren, W. F., Wilkinson, A. J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; pp 83−96.

(32) Cheatham, T. E., III; 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*, 845−862.

(33) 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*, 1049−1074.

(34) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E., III; 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*, 3817−3829.

(35) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, *282*, 680−686.

(36) Pohl, F. M.; Jovin, T. M. Salt-Induced Co-Operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly (dG-dC). *J. Mol. Biol.* **1972**, *67*, 375−396.

(37) Pohl, F. M. Polymorphism of a Synthetic DNA in Solution. *Nature* **1976**, *260*, 365−366.

(38) Pohl, F. M.; Jovin, T. M. Salt-Induced Co-Operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with poly(dG-dC). *J. Mol. Biol.* **1972**, *67*, 375−396.

(39) Fuertes, M. A.; Cepeda, V.; Alonso, C.; Pérez, J. M. Molecular Mechanisms for the B-Z Transition in the Example of poly[d(G-C) X d(G-C)] Polymers. A Critical Review. *Chem. Rev.* **2006**, *106*, 2045−2064.

(40) Rich, A.; Nordheim, A.; Wang, A. H. The Chemistry and Biology of Left-Handed Z-DNA. *Annu. Rev. Biochem.* **1984**, *53*, 791−846.

(41) Rich, A.; Zhang, S. Timeline: Z-DNA: The Long Road to Biological Function. *Nat. Rev. Genet.* **2003**, *4*, 566−572.

(42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(43) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

(44) 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*, 9665−9678.

(45) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.

(46) Bergonzo, C.; Cheatham, T. E., III Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11*, 3969−3972.

(47) 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. *J. Phys. Chem. B* **2012**, *116*, 9776−9783.

(48) Wu, Z.; Delaglio, F.; Tjandra, N.; Zhurkin, V. B.; Bax, A. Overall Structure and Sugar Dynamics of a DNA Dodecamer from Homo- and Heteronuclear Dipolar Couplings and ^{31}P Chemical Shift Anisotropy. *J. Biomol. NMR* **2003**, *26*, 297−315.

(49) MacDonald, D.; Herbert, K.; Zhang, X.; Polgruto, T.; Lu, P. Solution Structure of an A-Tract DNA Bend. *J. Mol. Biol.* **2001**, *306*, 1081−1098.

(50) Han, G. W.; Langs, D.; Kopka, M. L.; Dickerson, R. E. The Ultra-High Resolution Structure of d(CTTTTAAAAG)$_2$: Modulation of Bending by T-A Steps and Its Role in DNA Recognition (PDB entry 1SK5). http://www.rcsb.org/pdb/explore/explore.do?structureId=1sk5 (accessed June 13, 2016).

(51) Maehigashi, T.; Hsiao, C.; Woods, K. K.; Moulaei, T.; Hud, N. V.; Williams, L. D. B-DNA Structure Is Intrinsically Polymorphic: Even at the Level of Base Pair Positions. *Nucleic Acids Res.* **2012**, *40*, 3714−3722.

(52) Tereshko, V.; Wilds, C. J.; Minasov, G.; Prakash, T. P.; Maier, M. A.; Howard, A.; Wawrzak, Z.; Manoharan, M.; Egli, M. Detection of Alkali Metal Ions in DNA Crystals Using State-of-the-Art X-Ray Diffraction Experiments. *Nucleic Acids Res.* **2001**, *29*, 1208−1215.

(53) 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*, 9020−9041.

(54) Joung, I. S.; Cheatham, T. E., III. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, *113*, 13279−13290.

(55) Shih, C. C.; Georghiou, S. Large-Amplitude Fast Motions in Double-Stranded DNA Driven by Solvent Thermal Fluctuations. *Biopolymers* **2006**, *81*, 450−463.

(56) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409−1419.

(57) Ryckaert, J.-P. J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(58) 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*, 1864−1874.

(59) Darden, T. A.; 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.

(60) 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.

(61) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(62) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER14*; University of California: San Francisco, 2014.

(63) Cheatham, T. E., III; Case, D. A. Using AMBER to Simulate DNA and DNA. In *Computational Studies of DNA and RNA*; Sponer, J., Lankaš, F., Eds.; Springer: Dordrecht, The Netherlands, 2006; pp 45−72.

(64) Roe, D. R.; Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(65) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(66) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575−1577.

(67) 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.

(68) Guéron, M.; Kochoyan, M.; Leroy, J. L. A Single Mode of DNA Base-Pair Opening Drives Imino Proton Exchange. *Nature* **1987**, *328*, 89−92.

(69) Leijon, M.; Gräslund, A. Effects of Sequence and Length on Imino Proton Exchange and Base Pair Opening Kinetics in DNA Oligonucleotide Duplexes. *Nucleic Acids Res.* **1992**, *20*, 5339−5343.

(70) Moe, J. G.; Russu, I. M. Proton Exchange and Base-Pair Opening Kinetics in 5′-d(CGCGAATTCGCG)-3′ and Related Dodecamers. *Nucleic Acids Res.* **1990**, *18*, 821−827.

(71) 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.

(72) Pérez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Towards a Molecular Dynamics Consensus View of B-DNA Flexibility. *Nucleic Acids Res.* **2008**, *36*, 2379−2394.

(73) Harris, S. A.; Sands, Z. A.; Laughton, C. A. Molecular Dynamics Simulations of Duplex Stretching Reveal the Importance of Entropy in Determining the Biomechanical Properties of DNA. *Biophys. J.* **2005**, *88*, 1684−1691.

(74) Harris, S. A.; Laughton, C. A.; Liverpool, T. B. Mapping the Phase Diagram of the Writhe of DNA Nanocircles Using Atomistic Molecular Dynamics Simulations. *Nucleic Acids Res.* **2008**, *36*, 21−29.

(75) Lankas, F.; Lavery, R.; Maddocks, J. H. Kinking Occurs during Molecular Dynamics Simulations of Small DNA Minicircles. *Structure* **2006**, *14*, 1527−1534.

(76) Čech, P.; Kukal, J.; Černý, J.; Schneider, B.; Svozil, D. Automatic Workflow for the Classification of Local DNA Conformations. *BMC Bioinf.* **2013**, *14*, 205.

(77) Dickerson, R. E.; Goodsell, D. S.; Kopka, M. L.; Pjura, P. E. The Effect of Crystal Packing on Oligonucleotide Double Helix Structure. *J. Biomol. Struct. Dyn.* **1987**, *5*, 557−579.

(78) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. Reaction Path Ensemble of the B-Z-DNA Transition: A Comprehensive Atomistic Study. *Nucleic Acids Res.* **2013**, *41*, 33−43.

(79) Kastenholz, M. A.; Schwartz, T. U.; Hünenberger, P. H. The Transition between the B and Z Conformations of DNA Investigated by Targeted Molecular Dynamics Simulations with Explicit Solvation. *Biophys. J.* **2006**, *91*, 2976−2990.

(80) Savelyev, A.; MacKerell, A. D. All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.

(81) Lu, X.-J. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31*, 5108−5121.

(82) Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L. J.; Hatcher, M. E. ^{31}P NMR Investigation of Backbone Dynamics in DNA Binding Sites. *J. Phys. Chem. B* **2009**, *113*, 2596−2603.

(83) Schwieters, C. D.; Clore, G. M. A Physical Picture of Atomic Motions within the Dickerson DNA Dodecamer in Solution Derived from Joint Ensemble Refinement against NMR and Large-Angle X-Ray Scattering Data. *Biochemistry* **2007**, *46*, 1152−1166.