All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution

ALEXANDER D. MACKERELL, JR., NILESH K. BANAVALI

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

Received 19 March 1999; accepted 30 August 1999

ABSTRACT: Molecular dynamics simulations based on empirical force fields can greatly enhance knowledge of DNA and RNA structure and dynamics in solution. Presented are results on simulations of three DNA sequences and one RNA sequence using the new all-atom CHARMM27 force field for nucleic acids presented in the accompanying manuscript (Foloppe, MacKerell, J Comput Chem, this issue). Data are reported on structural, dynamic, and hydration properties including dihedral angle, sugar puckering, and helicoidal parameter probability distributions. Also presented are calculations of a DNA hexamer in 0 and 75% ethanol starting from both the canonical A and B forms. Analysis of RMS differences with respect to the canonical A and B forms of DNA show a highly anticorrelated behavior indicating that the force field samples the equilibrium between the A and B forms of DNA. Proper stabilization of B form DNA in aqueous solution and A form DNA in 75% ethanol show that this equilibrium can be perturbed by environmental contributions. Success of the force field in reproducing a variety of experimental data for duplex DNA and RNA indicates that it is of general use for computational investigations of nucleic acids as well as nucleic acids in complexes with proteins and lipids.
 J Comput Chem 21: 105–120, 2000

Keywords: CHARMM; force field; molecular dynamics; parametrization; DNA; RNA

Correspondence to: A. D. MacKerell; e-mail: amackere@rx.umaryland.edu

Contract/grant sponsor: NIH; contract/grant number: GM51501

This article includes Supplementary Material available from the authors upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/21/105 or http://journals.wiley.com/jcc/

CCC 0192-8651 / 00 / 020105-16

Introduction

Successful molecular dynamics (MD) simulations of duplex DNA and RNA in solution have only recently become possible.[1–5] Examples of MD simulation of nucleic acids include the influence of environment on DNA structure,[6–9] alternate chemical forms of DNA,[10–12] structures of RNA and DNA beyond duplexes,[13–16] and nucleic acid–protein complexes.[17] Initially, it was suggested that the use of Ewald sums[18] to treat long-range electrostatic interactions was necessary to perform stable simulations of duplex DNA,[2, 19] however, studies using atom truncation also yielded stable structures,[5, 7, 20] indicating that the ability to perform stable simulations was, in part, due to advances in the force fields used in the calculations. Despite the improvements in force fields, limitations in the widely used AMBER[21] and CHARMM[22] all-atom nucleic acid force fields exist.[23–25] These limitations motivated improvements in both force fields[25, 26] and the development of the Bristol–Myers Squibb (BMS) force field for nucleic acids.[27]

Presented are results from MD simulations of duplex DNA and RNA in solution using the new CHARMM27 all-atom force field for nucleic acids, named based on the version of the program CHARMM[28, 29] with which they will initially be released. The goal of the present study is to characterize the properties of the CHARMM27 force field with respect to experimental data. Such a characterization is necessary to validate the force field for use in solution simulations, define the applicability of the force field, and allow users to identify novel results vs. force field-dependent behavior when applying the force field. Analysis of the MD simulations concentrates on intrinsic characteristics of DNA and RNA, including dihedral angles, sugar puckering, and helicoidal parameters, and other experimentally accessible properties, including hydration. Extensive comparisons are made between calculated data and probability distributions and average values from a survey of crystal structures in the Nucleic Acid Database (NDB).[30] Results are also presented from simulations on a DNA hexamer in 0 and 75% ethanol, and from a comparison of DNA simulations that differed only in the treatment of electrostatic interactions via Ewald or atom truncation methods. Details of the parameter optimization process are presented in the accompanying manuscript.[26]

Methods

Oligonucleotide atom names and torsional angles are defined as in Saenger.[31] The canonical A and B forms of the DNA are defined according to Arnott and Hukins,[32] and the sugar pseudorotation angle and amplitude have been determined following Altona and Sundaralingam, using the same reference state for $P = 0.0°$.[33] Helicoid parameters have been calculated using the FREEHELIX program.[34]

Empirical calculations were carried out with the CHARMM program[28, 29] using a dielectric constant of 1.0. The water model in all calculations was the CHARMM-modified TIP3P.[35, 36] Parameters for sodium are from Beglov and Roux,[37] and the magnesium parameters are based on reproduction of the experimental free energy of solvation (B. Roux, personal communication). MD simulations in solution were performed in the NPT ensemble[38] at 300 K with the Leap-Frog integrator. All calculations were performed using SHAKE[39] to constrain covalent bonds involving hydrogens, allowing for an integration time step of 0.002 ps, and images were generated using the CRYSTAL module[40] in CHARMM. Electrostatic interactions were treated via both atom truncation and the Ewald method[18] via the Particle Mesh Ewald (PME) approach.[41] Atom truncation was performed by using the force shift and force switch methods to smooth the electrostatic and Lennard–Jones (LJ) terms, respectively.[42] Nonbond pair lists were maintained out to 14 Å, and nonbond interactions were truncated at 12 Å. Nonbond lists were updated heuristically. PME calculations were performed using real space cutoffs of 8, 10, or 12 Å, as noted in Table I, with the LJ interactions truncated at the same distance, with lists maintained 2 Å beyond the real space cutoff and updated heuristically. The fast Fourier transform grid densities were set to approximately 1 $Å^{-1}$. The screening parameter was determined for each system by using a sixth-order smoothing spline and varying the screening parameter (kappa) from 0.20 to 0.50, and selecting the value at which the change in the energy as a function of kappa went to zero. In all cases the value of kappa was in the range of 0.28 to 0.34.

Preparation of the solvated systems was performed in several steps. Computations were initiated with the canonical A or B forms of DNA that were overlaid with a preequilibrated solvent box of the CHARMM TIP3P water model[35, 36] and sodium ions or, for the 75% ethanol simulations, a box of previously equilibrated 75% ethanol that included sodium counterions. Ethanol parameters

TABLE I.
DNA and RNA Duplex Structures Included as Target Data for the Parameter Optimization.

Sequence	Comment	Reference	Real Space[a]	Sampling[b]
d(CGCGAATTCGCG)	EcoRI recognition sequence	73, 95	8	0.5–3.0
d(CATTTGCATC)	NMR solution structure	53	8	0.5–2.0
d(CTCGAG)	A to B transition	56	12	0.5–2.0
d(GTACGTAC)	A form crystal, 2.2 Å resolution	65	NA	
d(CGATCGATCG)	B form crystal, 1.5 Å resolution	64	NA	
UAAGGAGGUGUA	RNA, 2 duplexes/asymmetric unit	96	10	0.5–2.0

[a] Distance of the real space cutoff in Å for the PME simulations.
[b] Time range in ns used for the analysis; the final value indicates the total duration of the simulation. NA indicates not applicable.

were extracted from the CHARMM22 protein force field.[43] All solvent molecules with a nonhydrogen atom within 1.8 Å of any DNA nonhydrogen atom were deleted. The number of sodiums was then checked to ensure electrostatic neutrality. If additional sodiums were required, they were added at random positions in the box or, if sodiums needed to be removed, those furthest from the oligonucleotide were deleted. All subsequent calculations were performed in the presence of periodic boundary conditions. The structures were minimized for 100 SD steps with all DNA nonhydrogen atoms fixed. The minimized system was then subjected to a 20 ps isothermal, constant volume MD simulation with all DNA atoms harmonically constrained with mass weighted force constants of 5 kcal/mol/Å to equilibrate the solvent around the DNA. In the case of the 75% ethanol simulations, the solvent equilibration simulation was extended to 1 ns. The final structure from the solvent equilibration simulation was subjected to a 100-step SD energy minimization, with the resulting structure used to initiate the production trajectory. For all simulations the initial portion of the production trajectories were considered equilibration, with analysis performed on the remainder of the simulations.

Analysis of the simulations was performed on the production trajectories following an initial 500 ps equilibration period, yielding a total analysis time of 2500 ps for EcoRI and 1500 ps for the other systems. Root-mean-square difference (RMSD) data and correlation analysis were performed on structures sampled every 5 ps, hydration numbers determined on structures sampled every 10 ps with the remaining analysis performed on structures sampled every 1 ps. RMSD values were obtained following least-squares fitting of all specified nonhydrogen atoms. Reported correlation coefficients were determined between the RMS differences with respect to the A and B forms of DNA based on the covariance of those values. All average values and standard errors were obtained by separating the simulations into five blocks (e.g., five 500-ps blocks for EcoRI and five 300-ps blocks for the other systems), obtaining the respective average values over the individual blocks with the final reported averages and standard errors obtained over those five block averages.[44]

Distributions of the dihedral angles and sugar pseudorotation angles in crystal structures of DNA duplexes were obtained from the NDB as of March 1998, while helicoidal parameter distributions were determined for structures obtained as of January 1999. Structures containing nonstandard DNA components, bound drugs, or proteins were excluded. The distributions are presented as probability distributions, and have been obtained separately for the A, B, and Z DNA families, by sorting the data into 2° bins, unless noted. For RNA dihedral and sugar puckering distributions, RNA duplexes, including those with mismatches and modifiers, and tRNA structures were used, while for the helicoidal parameters only the duplex structures were included.

Results and Discussion

Structures of DNA duplexes from X-ray crystallographic studies are known to contain local structural heterogeneites[31, 45] vs. the canonical structures of DNA obtained from fiber diffraction data.[32, 46] These differences are, in large part, related to base sequence and composition; however, interactions of the duplexes with the crystal environment can influence their structures.[47–50] Thus, during development of the new CHARMM27 all-atom nucleic acid parameters several DNA sequences were used

as target data for the optimization. Furthermore, calculations were performed both in solution and in the crystal environment. In the present manuscript results from the solution simulations are presented, while details of the crystal simulations are presented in the accompanying manuscript.[26] Data from the solution simulation of an RNA dodecamer not presented in the accompanying manuscript are included. A list of all the sequences studied is presented in Table I. In the remainder of the manuscript the following nomenclature will be used to identify the different systems: d(CGCGAATTCGCG), EcoRI; d(CATTTGCATC), CATTT; d(CTCGAG) initiated from canonical A DNA and performed in 75% ethanol, Hexamer A; d(CTCGAG) initiated from canonical B DNA and performed in aqueous solution, Hexamer B; d(GTACGTAC) in the crystal environment, A crystal; d(CGATCGATCG) in the crystal environment, B crystal; and the UAAGGAGGUGUA RNA simulation; RNA.

During optimization of the parameters, data from simulations of the A and B form crystals listed in Table I and results for a collection of small model compounds were compared with a variety target data.[26] To ensure that the parameters were not being dominated by the structural properties of the two crystal structures tests were also performed on three DNA duplexes in solution. These three systems, EcoRI, CATTT, and the hexamer, were selected because they vary in both size and sequence (see Table I). Furthermore, different types of experimental data are available for the structures, allowing for different comparisons between the calculation and experiment to be performed.

Results and discussion will be presented in sections based on the properties being analyzed rather than by the DNA sequences to facilitate comparison of the different sequences. Exceptions to this will occur with the analysis of structural changes associated with the influence of ethanol on the DNA hexamer and on the influence of truncation method on the MD simulations. In all cases, except the truncation section, the MD simulations were performed using the PME method to treat long-range electrostatic interactions.

RMS DIFFERENCES: ANTICORRELATED BEHAVIOR WITH RESPECT TO A AND B DNA

During optimization of the CHARMM27 force field, emphasis was placed on properly treating both the A and B forms of DNA. In solution, it is expected that the B form of DNA will dominate for both EcoRI and CATTT, as evidenced by NMR studies.[51–53] To investigate the solution structure of these two sequences the RMSD with respect to both canonical A and B DNA were determined. During the 500 to 2000-ps portions of the trajectories, the RMSDs for EcoRI were 4.50 ± 0.04 and 1.92 ± 0.05, and for CATTT they were 3.11 ± 0.05 and 2.10 ± 0.06 with respect to the A and B forms of DNA, respectively. Plots of the RMSD vs. time are presented in Fig. 1 of the supplementary material along with the RMSD of RNA with respect to the canonical A and B structures. From analysis of the RMSD vs. time, it is evident that the two DNA structures are sampling regions of conformational space between the A and B forms of DNA. To better investigate this phenomena, correlation coefficients for the RMSD vs. the A and B forms of DNA were determined from the 500 to 2000-ps regions of the two simulations. Values of -0.82 and -0.93 were obtained for EcoRI and CATTT, respectively, indicating a highly anticorrelated behavior of the RMSD with respect to the A and B forms of DNA. Thus, during the DNA simulations as a structure deviates from the canonical B form of DNA, it moves towards the canonical A form and vice versa, showing the force field to be sampling the conformational space between the A and B forms of DNA. As a control, correlation analysis was performed on the 500 to 2000-ps portion of the RNA simulation, yielding a correlation coefficient of 0.62. The RNA result shows a lower correlation coefficient of opposite sign, indicating that the anticorrelated results for EcoRI and CATTT are not an inherent property in the force field, but are specific to DNA. It should be noted that differences between DNA and RNA in the force field are only present in the parameters associated with the sugar moiety and uracil. Thus, the present force field for DNA yields solution properties that are sampling the region of conformational space between the A and B forms of DNA with that sampling occurring in a highly anticorrelated fashion. This property is suggested to allow for the sensitivity of the DNA force field to environmental conditions (see below) and for stable simulations of both the A and B crystals.

CTCGAG HEXAMER: INFLUENCE OF ETHANOL CONCENTRATION ON THE A AND B FORMS OF DNA

One of the major successes of recent force fields for nucleic acids is their ability to reproduce environmental influences on the equilibrium between the A and B forms of DNA.[8, 54, 55] Accordingly, it may be considered essential for new force fields to

retain this ability. To test the environmental sensitivity of the CHARMM27 force field, calculations were performed on the CTCGAG hexamer[56] in aqueous solution and in a 75% ethanol solution. This sequence was selected due to it having some A form characteristics while being primarily in the B form in the experimental crystal structure. Four simulations were undertaken—two in aqueous solution and two in 75% ethanol solution—starting from both the canonical A and B forms of DNA. Results from these simulations in the form of RMSD vs. time are presented in Fig. 1. RMSD vs. time for the two 75% ethanol simulations starting from A and B form DNA (Fig. 1A and B, respectively) show both forms to be stable. The A form being stable is consistent with experimental studies,[57] while the stability of the B form appears to be associated with kinetic barriers to the transition, consistent with previous MD studies.[8] Recently, it was reported that structural transitions between A and B DNA can take up to 5 ns to complete during simulations of DNA.[25] To test that the A form was indeed stable in 75% ethanol, the simulation was extended out to 5 ns. Average RMSD values over the last 1 ns (i.e., 4000 to 5000 ps) were 1.5 ± 0.2 and 3.1 ± 0.4 vs. the A and B forms of DNA, respectively, indicating that the CHARMM27 force field does yield the A form of DNA in a low-water activity environment.

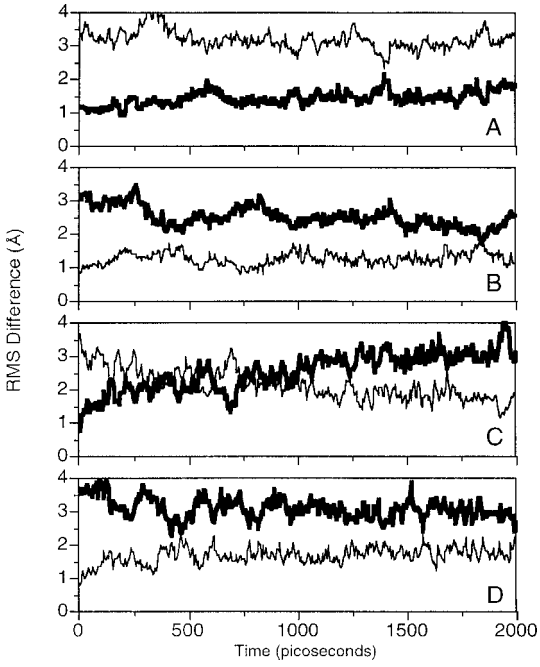

FIGURE 1. RMS differences vs. time from MD simulations of the DNA CTCGAG hexamer in 75% ethanol (A and B) and in water (C and D) starting from the canonical A (A and C) and B (B and D) forms of DNA. RMS differences are for all nonhydrogen atoms following least-squares fitting to the canonical A (bold lines) and B (thin lines) forms of the respective sequences.

Presented in Fig. 1C and D are the hexamer simulations in water started from the A and B forms of DNA, respectively. In the A initiated simulation, there is a gradual drift away from the A to a B form structure, with that transition taking over 1 ns. In the water simulation initiated from B form DNA (Hexamer B), the structure stays closer to the B form of DNA throughout the simulation. This simulation was also continued for 5 ns, with the RMSD over the final 1 ns being 3.5 ± 0.4 and 2.1 ± 0.4 Å with respect to the canonical A and B forms of DNA, respectively. The instability of A form DNA and the stability of B form DNA in high-water activity are again consistent with experimental data.[57] It should be noted that in a separate simulation of B DNA in water initiated with a different random number seed (not shown), base pairs 5 and 6 opened at the beginning of the simulation, leading to large RMSD values, although the RMSD with respect to B DNA was generally smaller than to A DNA.

The RMSD data in Fig. 1 indicate the present force field to be sensitive to environmental conditions. The A form structure is stabilized in low-water activity, and the B form in high-water activity, consistent with experiment.[31, 57] This has been verified in simulations using CHARMM27 with 4 M NaCl to create the low-water activity environment where the A form of DNA was observed to be stable.[27]

DIHEDRAL ANGLE PROBABILITY DISTRIBUTIONS

Optimization of the present force field was, in part, based on the reproduction of *ab initio* calculated dihedral angle potential energy surfaces for a collection of small model compounds designed to explicitly treat the backbone dihedrals, sugar puckering, and the glycosyi linkage, and on the reproduction of NDB survey probability distributions of these terms via MD simulations. As discussed in the accompanying manuscript, it was necessary to sacrifice the quality of the agreement between the *ab initio* and force field potential energy surfaces to adequately reproduce the survey data in simulations of the A and B crystals. To further investigate the quality of the dihedral distributions for DNA in solution, probability distributions were obtained from the EcoRI, CATTT, Hexamer A, and Hexamer B MD simulations.

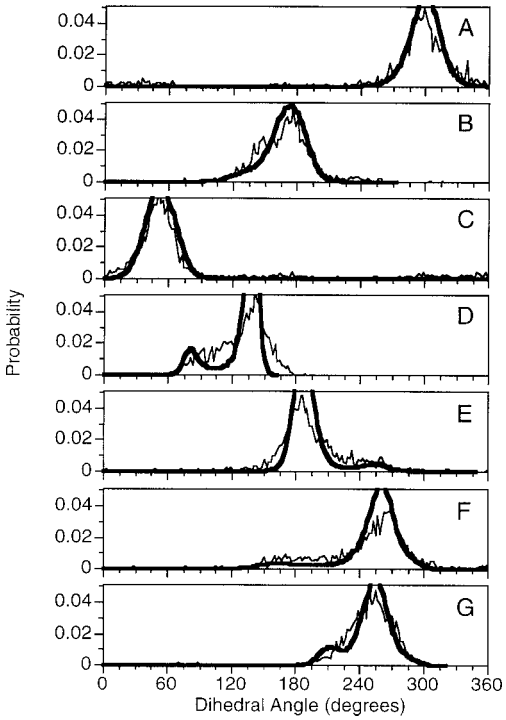

FIGURE 2. Probability distributions from an MD simulation of the EcoRI dodecamer (bold lines) and from a survey of the B structures in the NDB (thin lines) as a function of the dihedral angles α (A), β (B), γ (C), δ (D), ε (E), ζ (F), and χ (G).

Presented in Fig. 2 are the probability distributions for the backbone and glycosyl linkage dihedrals from the EcoRI simulation along with results from the crystal surveys; similar plots for CATTT are presented in Fig. 2 of the Supplementary Material. For both systems the agreement with the survey data is quite good. The largest differences between the MD and survey results occur with δ (Fig. 2D and Supplementary Material Fig. 2D), ε (Fig. 2E and Supplementary Material Fig. 2E), ζ (Fig. 2F and Supplementary Material Fig. 2F), and χ (Fig. 2G and Supplementary Material Fig. 2G). With δ, there are two distinct peaks in the MD results compared to the broader peak from the survey. The smaller peak, centered at ca. 80°, corresponds to sugars in the north pucker associated with A form DNA; a similar pattern occurs with the sugar-puckering distributions (see below). The presence of some north puckering is consistent with the B form survey data and with NMR data,[53, 58] indicating B form DNA to sample A form sugars. The narrowness of the MD distributions, however, is difficult to interpret. As discussed in the accompanying manuscript, it may be attributed to limited sampling in the MD simulations such that the full range of accessible dihedrals are not being sampled; however, limitations in the force field cannot be excluded. Alternatively, sampling of δ in the NDB could have contributions from crystal contacts in the survey data along with limitations associated with parameters used in the determination of the crystal structures; such a determination, however, is beyond the scope of the present work. With ε (Fig. 2E and Supplementary Material Fig. 2E) and ζ (Fig. 2F and Supplementary Material Fig. 2F), the extent of sampling in the regions of 255° and 180°, respectively, is decreased in the MD result compared to the survey. This sampling is associated with the B_{II} conformation, and is discussed below in more detail. For χ, the MD results show a small peak in the vicinity of 210°, which is due to the sampling of the A form of DNA, again consistent with NMR evidence indicating the sampling of A form sugars by B form DNA in solution.[53]

Analysis of the dihedral distribution from the hexamer simulations allows for determination of the quality of the force field in modeling A DNA in addition to the A DNA crystal simulations (see accompanying manuscript[26]), and an additional test of the B form of DNA. Dihedral distributions from the A form MD simulation in 75% ethanol, along with results from the NDB survey, are presented in Fig. 3; similar plots for the B form simulation in water are presented in Fig. 3 of the Supplementary Material. For Hexamer A, the results are generally in good agreement with the NDB survey data. The largest discrepancies occur with δ (Fig. 3D) and χ (Fig. 3G), both of which are due to sampling of south sugar puckers. At 75% ethanol the transition from the B to A form of DNA is not yet complete,[57] indicating the sampling of B conformations to be consistent with experiment. Significant deviations also occur with ε (Fig. 3E) and ζ (Fig. 3F), where tails in the MD distributions are present that are indicative of B form conformations. For Hexamer B, the probability distributions are in good agreement with the survey data, verifying the EcoRI and CATTT data, with the largest deviations again occurring with δ (Supplementary Material Fig. 3D) and χ (Supplementary Material Fig. 3G) due to sampling of north sugar puckers. This sampling is consistent with the hexamer experimental crystal structure where two of the six bases in the asymmetric unit have a north pucker.[56] As with the EcoRI and CATTT results, the B_{II} conformation is again being sampled, as evidenced in the ε (Supplementary Material Fig. 3E) and ζ (Supplementary Material Fig. 3F) surfaces (see below).

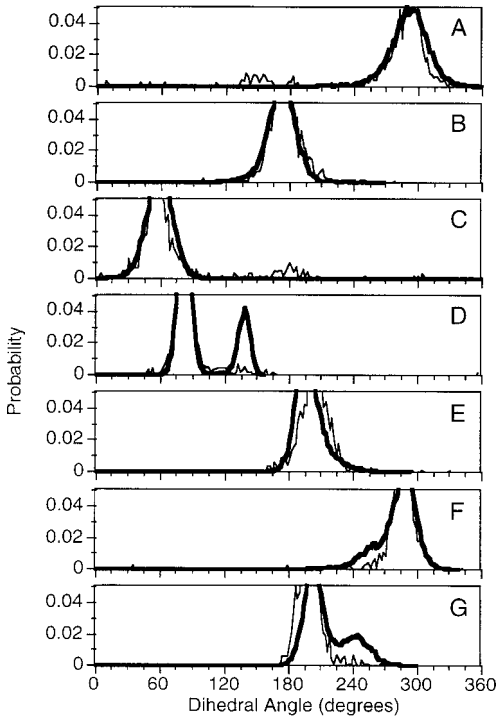

FIGURE 3. Probability distributions from an MD simulation of the CTCGAG hexamer (bold lines) in 75% ethanol starting from the canonical A form of DNA and from a survey of the A structures in the NDB (thin lines) as a function of the dihedral angles α (A), β (B), γ (C), δ (D), ε (E), ζ (F), and χ (G).

As discussed in the two previous paragraphs, experimental studies of B form DNA show the presence of two backbone conformations associated with the ε and ζ dihedrals. These substates are referred to as B_I and B_{II}, with the B_I being the most populated state in crystals and in solution.[45, 59, 60] In the present B form simulations, the percentages of B_I and B_{II} substates were 92 and 8%, respectively, for EcoRI, 94 and 6%, respectively, for CATTT, and 93 and 7%, respectively, for Hexamer B. Comparison of these values with those from the present survey of the NDB, 87 and 13% for the B_I and B_{II} substates, respectively, indicates the CHARMM27 force field to underestimate the sampling of the B_{II} conformation. Although the CHARMM27 underestimation may be due to limited sampling in the simulations, experimental studies indicate that the extent of sampling of the B_{II} substate in solution is decreased compared to the crystal,[45, 46] suggesting that the values from the MD simulations may be satisfactory. It should be noted that recent studies on hydrated DNA films indicate a relatively high population of the B_{II} substate,[62] indicating that further studies are required to clarify this point.

TABLE II.
Populations of the B_I and B_{II} Substates from the EcoRI MD Simulation.

Bases	B_I	B_{II}
C1-G2	0.60	0.40
G2-C3	0.98	0.02
C3-G4	0.87	0.13
G4-A5	0.94	0.06
A5-A6	0.99	0.01
A6-T7	0.99	0.01
T7-T8	1.00	0.00
T8-C9	0.99	0.01
C9-G10	0.94	0.06
G10-C11	0.94	0.06
C11-G12	0.87	0.13

The B_{II} substates are those with $\varepsilon > 210°$ and $\zeta < 210°$.

Analysis of the influence of sequence on the population of the B_I and B_{II} substates is presented in Table II for the EcoRI simulation. The data in Table II shows two significant results. First, the highest sampling of the B_{II} states occur at the duplex termini and, second, higher sampling of B_{II} generally occurs at internal basesteps where the second base is a purine (e.g., C3–G4), and the lowest sampling at steps where the second base is a pyrimidine (e.g., T7–T8). Experimental studies indicate higher sampling of the B_{II} substate in solution occurs at the duplex termini[60] and that B_{II} conformations in crystal structures occur to a greater extent at base steps where the second base is a purine.[59] The consistency of the MD simulation results in Table II with these experimental observations indicate that the present force field adequately models sequence-dependent contributions to the equilibrium between the B_I and B_{II} conformations, although the possibility that the overall sampling of the B_{II} state is underestimated cannot be excluded.

SUGAR PUCKERING PSEUDOROTATION ANGLE AND AMPLITUDE

Differences in sugar puckering between the A and B forms of DNA and the influence of the 2′ hydroxyl group on the sugar pucker of RNA are well documented.[31, 45] These properties, along with the central role of the sugar moiety in both the backbone of oligonucleotides and the linkage between the phosphodiester groups and the base, makes its proper treatment essential. To validate the quality of the treatment of sugar puckering by the CHARMM27 force field, probability distribu-

tions for both the pseudorotation angles and the amplitudes[33] were determined from the MD simulations and compared with the crystal survey data. In addition, pseudorotation angle time series are presented from the EcoRI simulation to investigate the nature of sampling of the sugar conformations by the present force field.

Pseudorotation probability distributions for the EcoRI, CATTT, RNA, Hexamer A, and Hexamer B solution simulations are presented in Fig. 4A through E, respectively. For the three B-form DNA simulations (Fig. 4A, B, and E) the MD distributions are in good agreement with the survey data. This includes sampling of the region between 60 and 120°, which is dominated by the pyrimidines, consistent with crystal[45] and Raman[63] studies and some sampling of the north pucker. Contributions associated with sequence, composition, and/or size effects are evident from the differences in the three DNA distributions, especially upon going from the two longer EcoRI and CATTT sequences to Hexamer B (Fig. 4E). With the RNA dodecamer (Fig. 4C), the range of pseudorotation angles sampled is in good agreement with the survey data, although the shape of the distribution in the north region is significantly different. During optimization of the force field, it was deemed desirable to reproduce the range of values rather than the shape of the distribution, considering the non-Gaussian form of the crystal survey distribution. Although the distribution of sugar pseudorotation angles in RNA may represent the experimental regimen, it may be suggested that force fields used in the optimization of RNA crystal structures may have an influence on the distribution. Of note is the sampling of the south conformation in RNA by the force field; while only a small amount of sampling occurs, it is consistent with the survey data. With the A form of DNA in 75% ethanol (Fig. 4D), the MD distribution is in satisfactory agreement with the survey; the overall range of pseudorotation angles in the north region is good; however, the total overlap of the peaks could be improved. Considering the overall quality of agreement of both the A and B sugar puckers additional optimization of the parameters was not deemed necessary. For both the A- and B-form DNA simulations alternate sugar pseudorotation angles were sampled (e.g., in the Hexamer A simulation south sugar puckering occurs), consistent with a variety of experimental studies.[53, 56, 64, 65] In the CATTT simulation (Fig. 4B) approximately 15% of the sugars occupy the north conformation. With B-form DNA it has been indicated that sampling of the north sugar occurs at levels less the 25% of the total population,[53] indicating the extent of sampling of the alternate conformation to be satisfactory. For Hexamer A at 75% ethanol the system is not expected to be completely in the A form such that sampling of B form sugars may be expected.

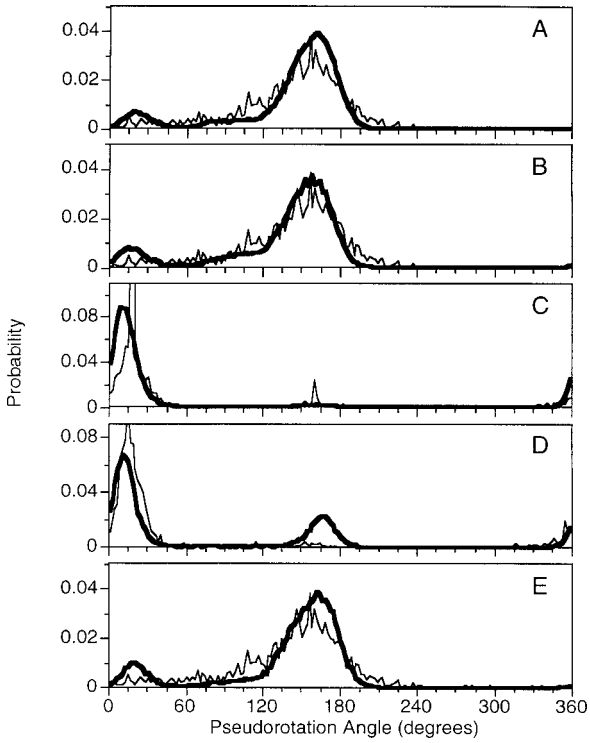

FIGURE 4. Probability distributions of the sugar pseudorotation angles from MD simulations (bold lines) of the (A) EcoRI dodecamer, (B) CATTTGCATC decamer, (C) RNA UAAGGAGGUGUA dodecamer, (D) CTCGAG hexamer in 75% ethanol (initiated from A DNA), and (E) the CTCGAG hexamer in aqueous solution (initiated from B DNA). Thin lines represent the sugar pseudorotation angle probability distributions from a survey of the NDB for B (A, B, and E) and A (D) form DNA structures and for RNA structures (C).

Analysis was also performed on the distributions of the sugar pucker amplitudes to obtain a better picture of the overall conformational properties of the furanose ring. Shown in Fig. 4A through E of the Supplementary Material are the amplitude distributions from the five MD simulations and the NDB survey data. In all cases the extent of agreement is satisfactory, with small shifts occurring in the MD distributions with respect to the survey data. The total ranges of amplitude sampled are in good agreement with the survey data. Of note are the changes in the MD distribution upon going from B DNA (Supplementary Material Fig. 4A, B, and E) to A DNA (Supplementary Material Fig. 4D)

to RNA (Supplementary Material Fig. 4C), which mimic the differences seen in the survey. Thus, the present force field adequately reproduces both pseudorotation angle and amplitude distributions from surveys of the NDB.

To investigate the nature of the dynamic properties of sugar puckering in CHARMM27 the individual time series of the pseudorotation angles for all 24 sugars from the EcoRI MD simulation are presented in Fig. 5. In the majority of cases sampling occurs primarily in the south region (ca. 120 to 180°, B form), with selected residues sampling the north region (ca. 0 to 60°, A form) for significant periods of time (e.g., strand 1, nucleotides 6, 9, 11, and 12; and strand 2, nucleotides 9 and 11). For the majority of nucleotides the sugars are primarily sampling one pucker, but make short-lived excursions towards the alternate pucker. For example, in nucleotide 4 of strand 1 there is an excursion towards the north pucker at ca. 700 ps. Similar excursions occur from the north region towards the south as in nucleotide 9 of strand 1 in the region of 1600 to 1900 ps. Concerning purines (nucleotides 2, 4, 5, 6, 10, and 12) vs. pyrimidines (nucleotides 1, 3, 7, 8, 9, and 11), the purines typically sample, in a discreet fashion, either the north or south regions with excursions between the two (i.e., over the region of 60 to 120°) occurring quickly. With the pyrimidines, there is more sampling of the region between 60 to 120°, consistent with the experiment.[45, 53] Of the nucleotides in Fig. 5 the cytosines (nucleotides 1, 3, 9, and 11) spend the most time in the north conformation, consistent with *ab initio* calculations on nucleosides, indicating cytosine to favor the A form of DNA.[66] Overall, the sugar-puckering time series further indicate the quality of the force field for treating sugar puckering, although rigorous verification of the time scales of the transitions is not currently possible due to limitations in available experimental data.

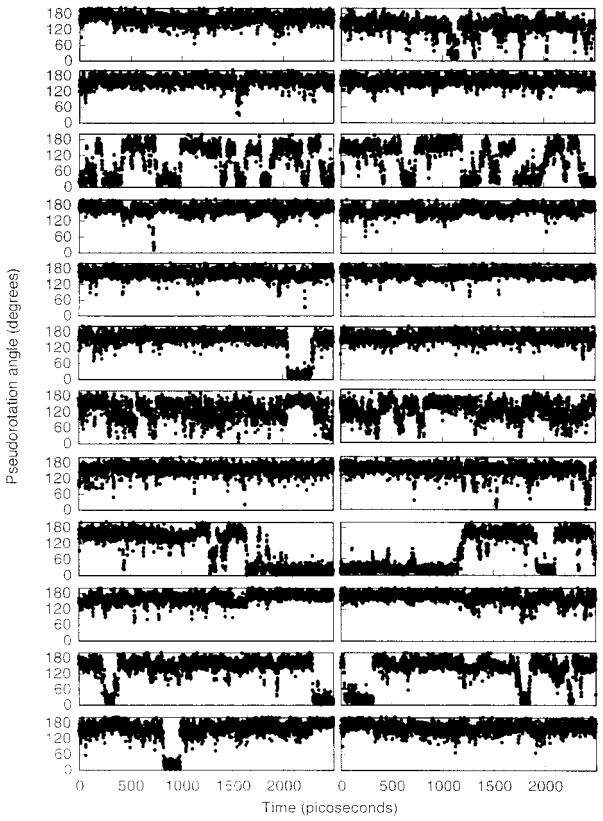

FIGURE 5. Sugar pucker pseudorotation angles as a function of time from the MD simulation of the EcoRI dodecamer in solution. The left and right columns are for strands one and two, respectively, with the individual nucleotides numbered from top (1) to bottom (12). Zero time represents the beginning of the production portion of the trajectory (500 ps of the 3-ns trajectory).

HELICOIDAL PARAMETERS

The structure of oligonucleotides can be defined solely based on the dihedral angles comprising the phosphodiester backbone, the sugar, and the glycosyl linkage. However, to facilitate the description of the orientation of the bases and base pairs, a collection of parameters have been developed. Helical or helicoidal parameters describe the relative orientation of bases with respect to either a global helical axis or locally, relative to adjacent base pairs.[45, 67] During optimization of the CHARMM27 nucleic acid force field the helicoidal parameters were not considered as target data. It can be expected, however, that a successful force field for nucleic acids should be able to accurately reproduce these terms. Presented in Table III are selected helicoidal parameters commonly used to describe duplex DNA and RNA. Included are reported values for canonical forms of DNA, average values, and standard deviations determined from the present survey of the NDB for A DNA, B DNA, and RNA, calculated using the FREEHELIX program,[34] and average values from the nonterminal bases for the EcoRI, CATTT, Hexamer A, Hexamer B, and RNA MD simulations.

Generally, the results from the MD simulations are in good agreement with the experimental data (Table III). With rise and slide the MD results mirror the experimental data. This is also true for twist, with the exception of Hexamer A and Hexamer B, where the calculated values are larger than those

TABLE III.
Selected Helicoidal Parameters from the MD Simulations, Canonical Structures, and a Survey of Crystal Structures in the Nucleic Acids Database.

Source	Rise	Inclination	Roll	Twist	X-disp	Prop. Twist	Slide
Canonical, fiber diffraction							
A-DNA	2.56	20.7	0.0	32.7	−5.28	−7.5	0.0
B-DNA	3.38	1.5	0.0	36.0	0.0	−13.3	0.0
NDB survey							
A-DNA	2.8 ± 0.4	14.0 ± 4.9	6.9 ± 4.7	32.3 ± 3.8	−4.1 ± 0.8	−9.2 ± 6.3	−1.4 ± 0.4
B-DNA	3.4 ± 0.5	2.8 ± 7.7	1.9 ± 6.0	35.7 ± 5.5	0.0 ± 1.4	−11.3 ± 8.2	0.4 ± 0.7
RNA	2.5 ± 0.8	14.6 ± 13.0	8.1 ± 17.7	29.9 ± 12.9	−3.6 ± 3.6	−3.2 ± 27.1	−2.6 ± 1.7
MD simulations							
EcoRI	3.4 ± 0.4	8.7 ± 3.0	5.6 ± 3.0	34.9 ± 2.5	−0.9 ± 0.4	−11.4 ± 5.4	0.1 ± 0.3
CATTT	3.4 ± 0.2	7.3 ± 1.9	5.2 ± 4.6	33.4 ± 2.3	−1.2 ± 0.1	−11.1 ± 5.4	−0.2 ± 0.5
Hexamer, A	2.5 ± 0.4	25.3 ± 7.7	10.8 ± 13.1	35.8 ± 3.1	−3.5 ± 0.4	−12.8 ± 7.9	−1.0 ± 0.2
Hexamer, B	3.4 ± 0.5	7.8 ± 2.1	5.3 ± 6.3	36.3 ± 3.2	−0.3 ± 0.0	−13.3 ± 0.6	0.3 ± 0.1
RNA	2.6 ± 0.2	14.5 ± 3.8	2.8 ± 6.5	31.3 ± 3.4	−2.6 ± 2.7	−5.4 ± 5.3	−2.3 ± 2.0

Terms calculated with FREEHELIX are as follows: rise and twist are associated with strand 1 and inclination, roll, twist, X-displacement, propeller, and slide are calculated for the base pairs. The use of only strand 1 is due to a large number of crystal structures having a single strand as the asymmetric unit. See original references for more details. Averages and standard deviations from the MD simulations were over all nonterminal bases or basepairs in the duplexes with the values for the individual bases or base pairs determined from averages of five individual block averages.

from experiment. Because the value of twist for the other two B form simulations is close to 34°, in excellent agreement with experimental data for DNA in solution,[68–71] and the ordering of the Hexamer twist values is correct (i.e., Hexamer A twist is smaller than Hexamer B), the larger twist values are attributed to either the size of the duplex or the base sequence. Inclination and roll values are all in satisfactory agreement with experiment, with the B form simulations having values intermediate to the A and B DNA survey values. This is consistent with the RMSD for these simulations indicating some A character (see Fig. 1 and Supplementary Material Fig. 1), and indicates that only selected aspects of DNA have A-like characteristics while the overall structure is predominately in the B form. Examination of the X-displacement values also shows the B form simulations to have slightly negative values, although those values are closer to the survey and canonical B form data than to that of the A form. Finally, the values of propeller twist from the simulations are in satisfactory agreement with the experimental data, although the error estimates for both the survey and simulation results limit rigorous comparison. Although limited by the magnitude of the errors, it is worth noting that the decreased propeller twist upon going from the DNA simulations to the RNA simulation mimics that seen in the survey. Thus, the MD simulations show the present force field to satisfactorily reproduce a variety of helicoidal parameters. The largest deviations from the canonical and survey data occur in the hexamer simulations. This difference may be attributable to the total size of the duplex, being able to only average over the four central bases, and to possible sequence dependent conformational properties in those four bases (TCGA).

Sequence dependent properties of DNA are essential for its biological function, and have been observed in crystal structures.[59] To analyze the ability of the present force field to represent sequence-dependent properties, the six local (vector based) basestep parameters calculated by the FREEHELIX program were determined for the EcoRI simulation and compared with five experimental crystal structures of the EcoRI dodecamer. Experimental structures included bd0005,[72] bd1001,[73] bd1002,[74] bd1005,[75] and bd1020 (◯),[76] where the identifiers are from the NDB. Results, presented in Fig. 5 of the Supplementary Material, show the simulation data to generally reproduce the trends in the helicoidal parameters as a function of residue present in the crystal structures. Of note is the increased Vrise (Supplementary Fig. 5A) and decreased Vtwist (Supplementary Fig. 5F) at steps 2 and 9, the decreased Vslide values (Supplementary Fig. 5C) in

the central region of the dodecamer, and the decrease in Vroll (Supplementary Fig. 5B) going from nucleotides 2 to 6 followed by an increase from 6 to 9. Although limited to only a single sequence, these results indicate that the presented force field adequately reproduces sequence-dependent behavior with respect to the helicoidal parameters.

HYDRATION PROPERTIES

Environmental influences on the structure of oligonucleotides, particularly DNA, are well documented, with one of the best studied being the relationship between water activity and the equilibrium between the A and B forms of DNA.[31] The majority of structural information on the hydration of DNA and RNA has been obtained via crystallography and NMR studies,[77–79] with theoretical approaches also making significant contributions.[1, 80–82] Although a detailed analysis of the interaction of water with DNA and RNA in the present simulations is beyond the scope of this study, examination of the hydration numbers of selected moieties is appropriate to gauge the quality of the force field with respect to hydration. Presented in Table IV are hydration numbers associated with the first hydration shell from the EcoRI, CATTT, and RNA solution simulations. Results are presented for all nonhydrogen atoms along with selected moieties. In all cases, the terminal bases were excluded from the calculations. For the two DNA simulations the hydration numbers per base are slightly greater than 10, in good agreement with experimental analysis of crystal structures indicating eigth to nine well-defined waters per base[83] and 5 to 12 tightly bound waters, as indicated by apparent molar volume and adiabatic compressibility experiments.[84] It should be noted that the method used presently to estimate the hydration numbers (see legend of Table IV) will typically overestimate the hydration numbers compared to the static structures from crystallography.[22] Concerning the individual moieties, analysis of crystal structures indicates the presence of at least two and three waters for each pyrimidine and purine base, respectively,[85] in good agreement with the values in Table IV. That same study indicates two and three

TABLE IV.
Hydration Numbers for Selected Atoms from the EcoRI, CATTTGCATC, and RNA UAAGGAGGUGUA Solution Simulations.

	Atoms					
	EcoR1		CATTTGCATC		RNA	
DNA	#/Atom	#/Base	#/Atom	#/Base	#/Atom	#/Base
DNA	0.50 ± 0.00	10.22	0.51 ± 0.00	10.38	0.54 ± 0.00	11.40
Bases						
Pyrmidines	0.39 ± 0.00	3.30	0.39 ± 0.00	3.43	0.45 ± 0.00	3.64
Purines	0.37 ± 0.00	3.87	0.35 ± 0.00	3.59	0.40 ± 0.00	4.22
Minor	0.83 ± 0.00	2.49	0.79 ± 0.00	2.38	0.82 ± 0.01	2.47
Major	0.79 ± 0.01	2.38	0.79 ± 0.01	2.36	1.12 ± 0.01	3.37
C8	0.69 ± 0.01		0.58 ± 0.01		0.68 ± 0.01	
C6	0.49 ± 0.01		0.39 ± 0.01		0.52 ± 0.00	
Backbone						
Sugar O4′	1.26 ± 0.01		1.23 ± 0.01		0.96 ± 0.01	
Ester O	0.92 ± 0.01	1.83	0.94 ± 0.01	1.88	0.88 ± 0.01	1.75
Anionic O	3.02 ± 0.03	6.04	3.04 ± 0.01	6.08	2.80 ± 0.01	5.60
Sugar O2′					2.51 ± 0.01	

Values represent the hydration numbers and standard errors from five 500 ps blocks from the EcoRI simulation and five 300 ps blocks from the CATTTGCATC and RNA simulations. Hydration numbers per atom (#/atom) were determined based on all waters within 3.5 Å of the selected atoms and averaged over values from windows every 10 ps in each block. Values are normalized with respect to the time frame and the the number of DNA atoms. Hydration number per base (#/base) was obtained by multiplying the #/atom values by the total number of atoms and dividing by the associated number of bases, except in the case of the minor and major grooves where #/base represents the total number of waters hydrating the basepairs comprising the grooves. DNA represents all nonhydrogen atoms, Base represents all base nonhydrogen atoms, Minor represents the purine N2, N3, or pyrimidine O2 atoms in the minor groove, and Major represents the purine O6, N6, N7, or pyrimidine N4, O4 in the major groove.

waters in the minor and major grooves, respectively, with the calculations in satisfactory agreement for the minor groove and underestimated for the major groove. It should be noted that the larger hydration numbers for the pyrimidines and purines compared to the minor and major grooves are due to all nonhydrogen atoms in the bases being included in the hydration determination, while only heteroatoms in the minor and major grooves are included (see legend of Table IV). The survey analysis also indicates the pyrimidine C6 and purine C8 atoms to be hydrated, as is observed in the simulations. For the backbone, the total hydration of the anionic oxygens per base in the simulations is close to six, again consistent with the crystal survey.[83] The calculated hydration number of the ester oxygens, however, is close to one, in contrast to the crystal analysis, which indicated hydration of the ester oxygens to be minimal. Visual analysis of the final structure from the EcoRI MD simulation showed the waters hydrating the ester oxygens (i.e., water oxygens within 3.5 Å of the ester oxygens) to typically be closer to the anionic oxygens, indicating that the calculated hydration numbers may be influenced by the selected methodology. The hydration number of the sugar O4′ oxygen being larger than the ester oxygen is also consistent with the crystal analysis.

A more detailed examination of minor groove hydration in DNA is shown in Fig. 6 of the Supplementary Material, where the final structure from the EcoRI MD simulation is presented. Included in the figure are the oxygens of water molecules in the minor groove. In Supplementary Material Fig. 6A the single-file spine of hydration in the central portion of the minor groove is evident. Close analysis of the figure reveals the number of waters in the AATT region of the minor groove to be seven, in good agreement with an estimate from NMR experiments.[86] Beyond the central AATT portion of the dodecamer (see Supplementary Material Fig. 6B) the minor groove hydration is no longer single file, with water molecules occurring double file in the groove. These results are in good agreement with both crystal structures[31] and NMR analysis.[86]

For the RNA MD simulation, the data are similar to the DNA, with some notable exceptions (see Table IV). The overall hydration of the RNA is larger than in DNA by approximately one water molecule, consistent with crystallographic studies.[87] This is due to the significant hydration of the O2′ atom, which is 2.5 waters per base in the present study. This value is in good agreement with the $\{r(C4G4)\}_2$ crystal structure, where between two and three waters per O2′ were identified.[87] The decreased hydration number of the sugar O4′ atoms is suggested to be due to competition for water molecules by the O2′ atom based on visual inspection of the final structure from the RNA MD simulation. The O2′ atom, which protrudes into the minor groove of the RNA, also contributes to the larger hydration number of the minor groove, as well as the increased pyrimidine and purine hydration numbers. The larger hydration number of the major groove is consistent with the estimate of five waters from the $\{r(C4G4)\}_2$ crystal structure and with NMR studies.[88] This increased hydration appears to also contribute to the increased hydration number of the purines in the RNA simulation. Data from crystal studies on the hydration of the minor groove of RNA is limited by crystal contacts typical in RNA structures.[87] Thus, the present force field yields hydration properties consistent with experiment for both DNA and RNA. This consistency, along with the ability of the force field to correctly treat the equilibrium between the A and B forms of DNA associated with changes in environment, indicate it to be appropriate for the study of environmental influences on the properties of oligonucleotides.

PARTICLE MESH EWALD VS. ATOM TRUNCATION

As discussed in the introduction, initial successes in performing simulations of duplex DNA both in solution and crystal environments were attributed to the use of Ewald sums to treat the long-range electrostatic interactions.[2, 19, 89, 90] Subsequently, it has been shown that stable simulations of DNA can be performed with atom-truncation methods that include smoothing functions.[5, 7, 20] To test the sensitivity of the CHARMM27 force field to the treatment of the long-range electrostatic interactions the EcoRI simulation was repeated using atom-based truncation at 12 Å, combined with the force shift method to smooth the energy and forces as they approached the truncation distance. Results from this simulation, which was extended to 2 ns, are presented in Figs. 7 and 8 of the Supplementary Material. RMSD vs. time in Supplementary Material Fig. 7 may be compared with the PME results in Supplementary Material Fig. 1A. In both cases, the structures stay close to the B form of DNA, but are fluctuating between the A and B canonical forms. RMS differences averaged over the 0.5 to 2.0-ns portion of the simulations were 4.5 ± 0.4 and 1.9 ± 0.3 Å for the PME simulation vs. and A and B forms, respectively, and 4.8 ± 0.3 and 1.8 ± 0.3 Å for the atom-truncation simulation, respectively, where the

errors are the standard deviations. Thus, the two methods yield similar structural differences with respect to the canonical forms of DNA. Furthermore, the atom truncation simulation shows a similar anticorrelated behavior with respect to the canonical forms of DNA, with a correlation coefficient of −0.74. Similarities of the two systems also extend to probability distributions for the backbone dihedrals and sugar puckering, as shown in Supplementary Material Fig. 8. Comparison of the atom truncation backbone and χ distributions (Supplementary Material Fig. 8A to G) with those from the PME simulation (Fig. 2A to G) show the results to be quite similar. The most significant difference occurs in sampling of the B_{II} conformation as may be seen in the ε and ζ distributions (see Fig. 2E and F and Supplementary Material Fig. 8E and F). In the atom truncation simulation, 10% of the phosphodiester linkages are in the B_{II} conformation compared with 6% in the PME simulation, although the statistical significance of this difference has not been validated. Sugar-puckering distributions are also similar for the atom-truncation (Supplementary Material Fig. 8H) and PME (Fig. 4A) MD simulations. Thus, the present force field yields DNA structural properties that are similar using either PME and atom-truncation schemes for the treatment of the electrostatic interactions, although some differences are evident. Although the use of Ewald approaches may be considered preferable for periodic boundary simulation, especially now that PME based methods are typically faster than atom truncation, alternate simulation approaches including stochastic boundary and other reaction field methods,[37, 91] where Ewald approaches cannot be applied, must use atom truncation or extended electrostatic methods for the treatment of electrostatic interactions. CHARMM27 should be suitable for such studies. It should be noted that the present results cannot be extrapolated to the use of shorter truncation distances, atom truncation without proper smoothing functions or high-salt simulations where long-range electrostatic effects may have a larger effect.

Conclusion

In the present manuscript, results from simulations of five different nucleic acid duplexes (Table I) have been performed as a test of the new CHARMM27 all-atom force field for nucleic acids.[26] It should noted that MD simulations on these systems were performed at different stages of the optimization procedure to ensure that the force field was not being biased by the A and B crystal simulations used as the primary macromolecular target data. Accordingly, the presented results may not be considered independent tests of the force field, but do act as a means to verify that the force field reproduces a variety of experimental observables. Independent tests of earlier versions of the CHARMM27 force field support the quality of the results presented in this work (L. Nilsson, personal communication; B. M. Pettit and M. Feig, personal communication; D. Langley, personal communication; D. Strahs and T. Schlick, personal communication; T. Cheatham, personal communication).

RMSD analysis of the force field shows the B form to be stable in aqueous solution (high-water activity, see Fig. 1 and Supplementary Material Fig. 1). During the DNA B form simulations, the duplexes sample regions of conformational space between the A and B forms of DNA, with that sampling occurring in a highly anticorrelated fashion. A low-water activity environment, via the addition of 75% ethanol, stabilizes the A form of DNA, consistent with experiment.[31, 57] In the present calculations as well as previously published work,[8] the B form of DNA is stable in 75% ethanol, in disagreement with experiment. This is suggested to be associated with kinetic barriers to the B to A transition. The ability of the CHARMM27 force field to stabilize the A and B forms of DNA in low- and high-water activity, respectively, the anticorrelated nature of the RMSD with respect to canonical A and B DNA, and the quality of the force field in reproducing experimental hydration numbers and the minor groove spine of hydration in EcoRI (see Table IV and Supplementary Material Fig. 6) indicates the present force field to properly balance interactions between the oligonucleotides and the environment with the intrinsic conformational properties of the oligonucleotides. Such a balance has previously been emphasized to be an essential property of force fields designed for condensed phase simulation studies.[21, 43]

Detailed analysis of the structural properties, including dihedral and sugar puckering probability distributions and helicoidal parameters shows good agreement with experiment (see Figs. 2 through 5, Supplementary Material Figs. 2 through 5 and Table III). Both the probability distributions and helicoidal parameters indicate sampling of A form conformations, even when the structures are primarily in the B form, and vice versa. This is exemplified by the sugar pseudorotation angle profiles where the B form simulations show some north sugar puckers (see Fig. 4A, B, and E), while both the DNA

Hexamer A (Fig. 4D) and RNA (Fig. 4C) simulations include sampling of south sugars. Of note are the values of twist in the EcoRI and CATTT simulations being close to the experimentally determined value of 34°. The ability of the force field to properly reproduce both local geometric properties (i.e., the dihedral probability distributions) while simultaneously reproducing more global properties (i.e., RMSD and average twist) indicates that the force field properly balances local structural contributions to yield the correct macromolecular properties. Particularly satisfying is the adequate reproduction of the helicoidal parameters, including the differences between A DNA, B DNA, and RNA (Table III), as well as the change in selected helicoidal parameters as a function of sequence (Supplementary Material Fig. 5). These terms were not included in the target data during the optimization of the force field; their correct treatment validates the applied parametrization methodology.

Testing of the present force field included the influence of the treatment of electrostatic interactions on a simulation of duplex DNA (see Supplementary Material Figs. 7 and 8). Although limited to only one test, it does indicate that the force field may be used with both the Ewald and atom-truncation methods for the treatment of electrostatic interactions. The latter is important in that it indicates that the force field can be used with reaction field methods[37, 91] designed to minimize the size of simulation systems.

Comparison of the present duplex DNA solution simulation results with those reported for the AMBER96,[21] revised AMBER (AMBER98),[25] and BMS force fields is informative. The BMS simulations were performed using atom truncation for the treatment of long-range electrostatics while AMBER96, AMBER98, and CHARMM27 primarily used PME. All force fields yield stable simulations of B DNA in solution. Based on RMSD, the BMS force field yields structures in a high-water activity environment that are closer to the canonical B form of DNA than CHARMM27, with both AMBER force fields being even less B like than CHARMM27. Both AMBER98 and BMS appear to have greater sampling of the B_{II} conformation than CHARMM27. The authors of Cheatham et al.[25] cite work indicating that this is consistent with DNA in solution,[60] however, other works indicate the percent B_{II} to be low in solution.[45, 61] Concerning the influence of water activity on the A to B equilibrium, AMBER96, BMS, and CHARMM27 adequately reproduce the experiment, while AMBER98 fails to stabilize the A form of DNA in low-water activity. Of note is that the BMS force field does allow for a B to A transition in 75% ethanol for the d(GGGCCC) hexamer, which did not occur with the d(CTCGAG) hexamer in the present work or with both AMBER force fields. This difference may be sequence dependent, based on the greater tendency for GC dominated duplexes to assume the A form conformation[92] and on the extensive equilibration of the 75% ethanol environment in this study, which has been suggested to effect transition rates between A and B DNA in MD simulations.[27] Two other properties distinguishing the A and B forms of DNA are the helicoidal parameters twist and X-displacement. With CHARMM27 the EcoRl (dodecamer) and CATTT simulations (decamer) had twist values in good agreement with expected value of 34° for DNA in solution, while the hexamer yielded a larger value of 36.3°. This value is consistent with that reported for BMS with the d(CGTACG) hexamer. AMBER96 yielded twist values in the range of 30°, while AMBER98 yields twist values lower than 34°, except in the case of dA_{10}-dT_{10}, where a value of 34.7° is reported. X-displacement values for B form structures for all force fields are intermediate to that expected for the B and A forms of DNA, with BMS closest to B form, followed by CHARMM27 and then AMBER98. Overall, all four force fields satisfactorily model the B form of DNA in solution, with AMBER96, CHARMM27, and BMS also treating the equilibrium between and A and B forms of DNA as a function of water activity. Although details of the merits of the four force fields for simulations of DNA and RNA in solution requires additional study, the most significant advantage of CHARMM27 is the extensive use of model compound target data during the parameter optimization to ensure that the proper balance of local contributions are combining to yield the desired macromolecular properties.

In summary, the CHARMM27 all-atom force field, when applied towards simulations of DNA and RNA duplexes in solution, yields satisfactory agreement with a variety of experimental observables. This ability, along with its compatibility with the CHARMM force fields for proteins,[43] and lipids,[93, 94] will allow for modeling and MD studies of nucleic acids in a variety of environments, including complexes with proteins and lipids.

Acknowledgments

We thank the DOD ASC Major Shared Resource Computing and High Performance Computing and

the NSF PACI Program for providing computational resources. Appreciation to Drs. T. Cheatham, M. Feig, N. Foloppe, D. Langely, L. Nilsson, M. Pettitt, T. Schlick, and D. Strahs for tests of the force field and helpful discussions.

Supplementary Material

Included in the Supplementary Material are eight figures. Color representations of Supplementary Material Figure 6 may be found on the web page of ADM Jr. at www.pharmacy.umaryland.edu/~alex.

References

1. Young, M. A.; Ravishanker, G.; Beveridge, D. L. Biophys J 1997, 73, 2313.
2. Yang, L.; Pettitt, B. M. J Phys Chem 1996, 100, 2550.
3. Cheatham, T. E., III; Kollman, P. A. J Am Chem Soc 1997, 119, 4805.
4. Auffinger, P.; Westhof, E. Biophys J 1996, 71, 940.
5. Norberg, J.; Nilsson, L. J Chem Phys 1996, 104, 6052.
6. Cheatham, T. E., III; Kollman, P. A. Structure 1997, 5, 1297.
7. MacKerell, A. D., Jr. J Phys Chem B 1997, 101, 646.
8. Cheatham, T. E., III; Crowley, M. F.; Fox, T.; Kollman, P. A. Proc Natl Acad Sci USA 1997, 94, 9626.
9. Sprous, D.; Young, M. A.; Beveridge, D. L. J Phys Chem B 1998, 102, 4658.
10. Miaskiewicz, K.; Miller, J.; Cooney, M.; Osman, R. J Am Chem Soc 1996, 118, 9156.
11. Cieplak, P.; Cheatham, T. E., III; Kollman, P. A. J Am Chem Soc 1997, 119, 6722.
12. Sen, S.; Nilsson, L. J Am Chem Soc 1998, 120, 619.
13. Zichi, D. A. J Am Chem Soc 1995, 117, 2957.
14. Luo, J.; Bruice, T. C. J Am Chem Soc 1998, 120, 1115.
15. Shields, G. C.; Laughton, C. A.; Orozco, M. J Am Chem Soc 1997, 119, 7463.
16. Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y.-K.; Pettitt, B. M. J Am Chem Soc 1995, 117, 2147.
17. Pardo, L.; Pastor, N.; Weinstein, H. Biophys J 1998, 75, 2411.
18. Ewald, P. P. Ann Phys 1921, 64, 253.
19. Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. J Am Chem Soc 1995, 117, 4193.
20. Norberg, J.; Nilsson, L. J Biomol NMR 1996, 7, 305.
21. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
22. MacKerell, A. D., Jr.; Wiórkiewicz–Kuczera, J.; Karplus, M. J Am Chem Soc 1995, 117, 11946.
23. MacKerell, A. D., Jr. In Molecular Modeling of Nucleic Acids; Leontis, N. B.; SantaLucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1998; p. 304, vol. 682.
24. Feig, M.; Pettitt, B. M. Biophys J 1998, 75, 134.
25. Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. J Biomol Struct Dynam 1999, 16, 845.
26. Foloppe, N.; MacKerell, A. D., Jr. J Comp Chem 2000, to appear.
27. Langley, D. R. J Biomol Struct Dynam 1998, 16, 487.
28. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
29. MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R.; N. L. A.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; p. 271, vol. 1.
30. Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. Biophys J 1992, 63, 751.
31. Saenger, W. Principles of Nucleic Acid Structure; Springer Verlag: New York, 1984.
32. Arnott, S.; Hukins, D. W. L. J Mol Biol 1973, 81, 93.
33. Altona, C.; Sundaralingam, M. J Am Chem Soc 1972, 94, 8205.
34. Dickerson, R. E. Nucleic Acids Res 1998, 26, 1906.
35. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926.
36. Reiher, W. E., III Theoretical Studies of Hydrogen Bonding. Ph.D., Harvard University (1985).
37. Beglov, D.; Roux, B. J Chem Phys 1994, 100, 9050.
38. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, R. W. J Chem Phys 1995, 103, 4613.
39. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comp Phys 1977, 23, 327.
40. Field, M. J.; Karplus, M. CRYSTAL: Program for Crystal Calculations in CHARMM; Harvard University: Cambridge, MA, 1992.
41. Darden, T. A.; York, D.; Pedersen, L. G. J Chem Phys 1993, 98, 10089.
42. Steinbach, P. J.; Brooks, B. R. J Comp Chem 1994, 15, 667.
43. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
44. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Biopolymers 1992, 32, 523.
45. Hartmann, B.; Lavery, R. Q Rev Biophys 1996, 29, 309.
46. Arnott, S.; Hukins, D. W. L.; Dover, S. D.; Fuller, W.; Hodgson, A. R. J Mol Biol 1973, 81, 102.
47. Dickerson, R. E.; Goodsell, D. S.; Neidle, S. Proc Natl Acad Sci USA 1994, 91, 3579.
48. Jain, S.; Sundaralingam, M. J Biol Chem 1989, 264, 12780.
49. Shakked, Z.; Guerstein–Guzikevitch, G.; Eisenstein, M.; Frolow, F.; Rabinovitch, D. Nature 1989, 342, 456.
50. Lipanov, A.; Kopka, M. L.; Kaczor–Grzeskowiak, M.; Quintana, J.; Dickerson, R. E. Biochemistry 1993, 32, 1373.
51. Gronenbom, A. M.; Clore, G. M. Biochemistry 1989, 28, 5978.
52. Lane, A. N.; Jenkins, T. C.; Brown, T.; Neidle, S. Biochemistry 1991, 30, 1372.
53. Weisz, K.; Shafer, R. H.; Egan, W.; James, T. L. Biochemistry 1992, 31, 7477.
54. Cheatham, T. E., III; Kollman, P. A. Structure 1997, 5, 1297.

55. Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. J Am Chem Soc 1998, 120, 10629.
56. Wahl, M. C.; Rao, S. T.; Sundaralingam, M. Biophys J 1996, 70, 2857.
57. Ivanov, V. I.; Krylov, D. Y. Methods Enzymol 1992, 211, 111.
58. Weisz, K.; Shafer, R.; Egan, W.; James, T. L. Biochemistry 1994, 33, 354.
59. Dickerson, R. E. Methods Enzymol 1992, 211, 67.
60. Gorenstein, D. G. Chem Rev 1994, 94, 1315.
61. Szyperski, T.; Ono, A.; Fernandez, C.; Iwai, H.; Tate, S.-I.; Wüthrich, K.; Kainosho, M. J Am Chem Soc 1997, 119, 9901.
62. Pichler, A.; Rudisser, S.; Mitterbock, M.; Huber, C. G.; Winger, R. H.; Liedl, K. R.; Hallbrucker, A.; Mayer, E. Biophys J 1999, 77, 398.
63. Peticolas, W. L.; Ghomi, M.; Spassky, A.; Evertsz, E. M.; Rush, T. S., III In Molecular Modeling of Nucleic Acids; Leontis, N. B., SantaLucia, J., Jr., Eds.; Americal Chemical Society: Washington, DC, 1998; p. 150.
64. Grzeskowiak, K.; Yanagi, K.; Prive, G. G.; Dickerson, R. E. J Biol Chem 1991, 266, 8861.
65. Langlois D'Estaintot, B.; Dautant, A.; Courseille, C.; Precigoux, G. Eur J Biochem 1993, 213, 673.
66. Foloppe, N.; MacKerell, A. D., Jr. Biophys J 1999, 76, 3206.
67. Dickerson, R. E. J Biomol Struct Dynam 1989, 6, 627.
68. Rhodes, D.; Klug, A. Nature 1980, 286, 573.
69. Peck, L. J. Nature 1981, 292, 375.
70. Rhodes, D.; Klug, A. Nature 1981, 292, 378.
71. Ulyanov, N. B.; James, T. L. Methods Enzymol 1995, 261, 90.
72. Shui, X.; Sines, C. C.; McFail–Isom, L.; VanDerveer, D.; Williams, L. D. Biochemistry 1998, 37, 16877.
73. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. S. Proc Natl Acad Sci USA 1981, 78, 2179.
74. Drew, H. R.; Samson, S.; Dickerson, R. E. Proc Natl Acad Sci USA 1982, 79, 4040.
75. Holbrook, S. R.; Dickerson, R. E.; Kim, S.-H. Acta Crystallogr B 1985, 41, 255.
76. Westhof, E. J Biomol Struct Dynam 1987, 8, 581.
77. Westhof, E. Annu Rev Biophys Biophys Chem 1988, 17, 125.
78. Saenger, W. Annu Rev Biophys Biophys Chem 1987, 16, 93.
79. Wütrich, K. Cold Spring Harbor Symp Quant Biol 1993, 149.
80. Hummer, G.; Garcia, A. E.; Soumpasis, D. M. Biophys J 1995, 68, 1639.
81. De Winter, H.; Lescrinier, E.; Van Aerschot, A.; Herdewijn, P. J Am Chem Soc 1998, 120, 5381.
82. Auffinger, P.; Westhof, E. J Mol Biol 1997, 269, 326.
83. Schneider, B.; Patel, K.; Berman, H. M. Biophys J 1998, 75, 2422.
84. Chalikian, T. V.; Sarvazyan, A. P.; Plum, G. E.; Breslaur, K. J. Biochemistry 1994, 33, 2394.
85. Schneider, B.; Berman, H. M. Biophys J 1995, 69, 2661.
86. Jóhannesson, H.; Halle, B. J Am Chem Soc 1998, 120, 6859.
87. Egli, M.; Portmann, S.; Usman, N. Biochemistry 1996, 35, 8489.
88. Conte, M. R.; Conn, G. L.; Brown, T.; Lane, A. N. Nucleic Acids Res 1996, 24, 3693.
89. York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. J Am Chem Soc 1995, 117, 5001.
90. Lee, H.; Darden, T.; Pedersen, L. G. J Chem Phys 1995, 102, 3830.
91. Brooks, C. L.; Karplus, M. J Chem Phys 1983, 79, 6312.
92. Peticolas, W. L.; Wang, Y.; Thomas, G. A. Proc Natl Acad Sci USA 1988, 85, 2579.
93. Schlenkrich, M.; Brickmann, J.; MacKerell, A. D., Jr.; Karplus, M. In Biological Membranes: A Molecular Perspective from Computation and Experiment; Merz, K. M.; Roux, B., Eds.; Birkhaüser: Boston, 1996; p. 31.
94. Feller, S. E.; Yin, D.; Pastor, R. W.; MacKerell, J., A. D. Biophys J 1997, 73, 2269.
95. Drew, H. R.; Dickerson, R. E. J Mol Biol 1981, 151, 535.
96. Schindelin, H.; Zhang, M.; Bald, R.; Fuerste, J.-P.; Erdmann, V. A.; Heinemann, U. J Mol Biol 1995, 249, 595.
97. Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. J Mol Graph 1988, 6, 13.