Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence of an Ideal B-DNA Structure to the Crystallographic Structure

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2006

Impact Factor: 3.3 · DOI: 10.1021/jp061421r · Source: PubMed

4 AUTHORS, INCLUDING:

Thomas Darden
OpenEye Scientific Software
146 PUBLICATIONS **6,167** CITATIONS
SEE PROFILE

Celeste Sagui
North Carolina State University
104 PUBLICATIONS **2,172** CITATIONS
SEE PROFILE

Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence of an Ideal B-DNA Structure to the Crystallographic Structure

Volodymyr Babin,[†] Jason Baucom,[†] Thomas A. Darden,[‡] and Celeste Sagui*,[†]

Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, and National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received: March 7, 2006; In Final Form: April 10, 2006

We have investigated to what extent molecular dynamics (MD) simulations can reproduce DNA sequence-specific features, given different electrostatic descriptions and different cell environments. For this purpose, we have carried out multiple unrestrained MD simulations of the DNA duplex $d(CCAACGTTGG)_2$. With respect to the electrostatic descriptions, two different force fields are studied: a traditional description based on atomic point charges and a polarizable force field. With respect to the cell environment, the difference between crystal and solution environments is emphasized, as well as the structural importance of divalent ions. By imposing the correct experimental unit cell environment, an initial configuration with two ideal B-DNA duplexes in the unit cell is shown to converge to the crystallographic structure. This convergence is measured by the appearance of sequence-dependent features that very closely resemble the crystallographic ones as well as by the decay of the all-atom root-mean-squared coordinates deviations (RMSD) with respect to the crystallographic structure. Given the appropriate crystallographic constraints, this is the first example of multiple nanosecond molecular dynamics trajectory that shows an ideal B-DNA model converging to an experimental structure, with a significant decay of RMSD.

I. Introduction

"Ideal" DNA originates from early X-ray diffraction studies of different forms of nucleic acid fibers, which produced several distinct diffraction patterns. When single crystal diffraction patterns became available, they identified several oligonucleotide models originally based on the fiber experiments: A- and B-DNA, A-RNA, and later Z-DNA. The ideal DNA structures differ on their relative size, length, handedness, and depth and width of grooves. These ideal structures are sequence-*independent*; i.e., the characteristic helicoidal parameters are always the same, independent of the base composition. However, when the first B-DNA crystal structure became available,[1] it was noticed that in addition to the general features corresponding to the ideal structure, there was a clear relationship between DNA sequence and structure, which is believed to be an important feature of essential cellular processes such as replication and transcription. This, in turn, has spawned a large number of experimental and theoretical investigations aimed at elucidating this relationship. Experimental work has primarily made use of X-ray crystallography and nuclear magnetic resonance (NMR), which probe the DNA structure in their crystal and solution environments, respectively. Complementing experimental work are theoretical investigations based on classical molecular dynamics (MD) simulations, which are important because they allow for a description of DNA structure and dynamics at the atomic level. It is now understood that a key feature of any MD simulation of DNA is the treatment of the delicate long-range electrostatic interactions. DNA is a highly charged and polar molecule, and correct electrostatics is absolutely essential for the stabilization of its three-dimensional structure.[2] Moreover, because a calculation of the Coulombic terms is expensive, the electrostatics typically represents the computational bottleneck of any long-time DNA simulation.[3] Large-scale simulations of DNA were enabled through the development of accurate and efficient algorithms for calculating electrostatics[4−6] just over a decade ago. Today, long-time stable simulations of DNA are commonplace, and several review articles discuss issues concerning MD simulations of nucleic acids.[7−11]

An important drawback of any MD simulation is the accuracy of the force fields, and considerable efforts have been devoted to improving the current models. In terms of biomolecules such as DNA, it has been recognized that there is a considerable loss of accuracy associated with the current description of the electrostatic interactions. This, in turn, has given impetus to the development of polarizable force fields, which potentially are considerably more accurate than the current descriptions based on point charges only. However, the capabilities and limitations of these new force fields remain largely unexplored, at least as far as DNA is concerned.

In this paper, we report on a large-scale simulation study aimed at elucidating the following: (i) the use of polarizable force fields in the description of nucleic acids; and (ii) the influence of the simulation environment, as given by the cell symmetry and divalent ions, in reproducing the experimental results. As a benchmark molecule, we have chosen the synthetic B-DNA decamer $d(CCAACGTTGG)_2$ whose experimental crystallographic structure is found in the PDB file with code 5DNB,[12] and for which both crystal and solution simulations[13−16] exist. However, there is no NMR data for this decamer, and therefore, we compared the root-mean-squared coordinates deviations (RMSD) with respect to the crystal structure (as previous simulations have done). Although our main purpose is to assess the simulations in crystal environment, to determine

* To whom correspondence should be addressed. E-mail: sagui@ncsu.edu.
[†] North Carolina State University.
[‡] National Institute of Environmental Health Sciences.

10.1021/jp061421r CCC: $33.50
Published on Web 05/19/2006

how the solution simulations compare with respect to the NMR experiments, the DNA dodecamer $d(CGCGAATTCGCG)_2$ has been chosen. Its high-resolution NMR structure is given in the PDB file with code 1NAJ.[17]

We elaborate on the aims annunciated above.

(i) Polarizable Force Fields. Recently, Baucom et al.[16] compared the performance of different force fields in preserving the structure of the $d(CCAACGTTGG)_2$ decamer over 25 ns unrestrained crystal simulations at constant pressure. The specific force fields investigated, based on the Cornell et al. force field,[18] were ff98, ff99, ff02, and ff02EP. These force fields characterize different electrostatic representations. In ff98 and ff99, the electronic charge density is simply represented by nonpolarizable atomic point charges. The ff02 force field is a polarizable variant, where the electronic charge density is represented by both atomic point charges and induced atomic point dipoles. The ff02EP force field is a polarizable force field with extra points on electron-donating atoms, which carry additional point charges to represent the electron lone pairs. The study found that all the force fields maintained the sequence-dependent features of the experimental structure to a reasonable degree. The polarizable ff02 force field, however, provided the most accurate representation of the crystal structure and sequence-dependent effects. Somewhat surprisingly, the use of ff02EP did not improve the accuracy of the results.

Although polarizable force fields tend to be considerably more costly to simulate, recent methodological advances for both permanent and induced dipoles,[19] as well as for higher order multipoles[20,21] based on the particle mesh Ewald (PME) method[5,6] and multigrid methods,[22] make such calculations feasible. In particular, the polarizable simulations in this work used a PME-based treatment of fixed and induced point dipoles implemented into the SANDER molecular dynamics module of AMBER 6, 7, and 8,[23−25] together with a Car−Parrinello scheme for the computation of the induced polarization. This method is 1.25−1.30 times more expensive than the PME method for charges alone, assuming a 1-fs time step for MD.

In this work, we carry out a more extensive comparison between the simple point-charge ff99 and the polarizable ff02 force fields both in crystal and in solution environments. Crystal simulations are started from the ideal B-DNA structure, instead of the crystal structure, to test whether MD simulations can converge to the correct experimental structure.

(ii) Simulation Environment: An important test for any new DNA force field is whether it can maintain the experimentally correct crystal structure in an MD simulation, without inducing further constraints than the experimental ones.[26] Whereas solution simulations tend to fluctuate more than experiments,[4] the packing density in crystal simulations may inhibit important atomic fluctuations. Constant pressure simulations, therefore, are recommended for crystal environments to avoid the problem of high pressure that may arise during a constant-volume simulation, which can lead to additional inhibitions of the motion. Results from successful crystal simulations[15,16,27−30] show that structural features characteristic of B-DNA are *maintained* over several nanoseconds. Whereas crystal simulations tend to preserve experimental features, solution simulations do not preserve all sequence-dependent features (e.g., helical twist). This indicates that crystal simulations may be used to facilitate the identification of force-field parameters that need to be adjusted to improve the agreement between experiment and simulation.

One important variable in the setup of the crystal environment is the presence of crystallographic Mg^{2+} ions. Although monovalent cations are needed to stabilize DNA conformation, they are *relatively* nonspecific in recognizing the surface of a regular DNA duplex (in contrast to the case of more complex nucleic acid structures, where they have specific stabilizing power). Their relative lack of hydration means that they interact with DNA through differences in the molecular electrostatic potential and hydrogen bonding in the major and minor grooves. This weak binding by monovalent cations does not affect the average conformation of duplex DNA under physiological conditions.[31−34] Divalent cations, on the other hand, are generally completely hydrated, and their interactions with duplex DNA are more sequence specific, mediated through their coordinated water ligands. In this sense, they play an important part in defining the structure of duplex DNA. As a byproduct of our crystal simulation studies, we find evidence for the very important role of Mg^{2+} ions in duplex DNA. For instance, crystal simulations with the particular symmetry and volume of the cell in the 5DNB PDB file could not be stabilized with Na^+ ions alone.

The main results of this paper are as follows: First, not only base sequence but also crystal packing and divalent ions need to be taken into account if the simulations are to reproduce the experimental results. Second, polarizable force fields seem to perform better than simple point-charge force fields in reproducing the structural features of DNA. Third, by setting up the correct unit cell environment, it is possible to start from an ideal B-DNA structure and converge to the experimental structure. In our system, this convergence is measured by a significant decay of the RMSD with respect to the crystal structure as well as by the appearance of the experimental sequence-dependent features. There is, therefore, considerable hope that good predictive DNA simulations may be achieved with current and future polarizable force fields.

II. Methods

Our simulations used two all-atom force fields implemented in AMBER 7[24] and 8:[25]

(i) ff99, which is the 1999 version of the Cornell et al. force field,[18] with the same atom types, topologies (except for torsional parameters), and charges. Essentially, this is an all-atom nonpolarizable force field for proteins and nucleic acids. For DNA simulations, the parm98.dat and parm99.dat were found to produce the same results within statistical error.[16] These force fields differ from the previous parm94.dat in the torsion angle parameters for the glycosidic torsion in nucleic acids.[35]

(ii) The ff02 force field, which is a polarizable variant of ff99. It uses parm99.dat, and the charges are more "gas-phase" than those in the nonpolarizable force fields. Polarizable dipoles based on isotropic atomic polarizabilities are attached to the atoms, to represent the bond polarization in a condensed phase environment. The induced dipoles are therefore directly proportional to the total electrostatic field, and no damping is assumed. The ff02 force field uses interactive polarization, which includes the mutual effects of polarizable sites within a molecule on each other (intramolecular polarizability), in addition to the polarization induced by the external field. This generally requires smaller distributed atomic polarizabilities than in additive models, because the total molecular polarizability is increased by the intramolecular polarizability. However, in the AMBER parametrization, both permanent and induced electrostatic interactions between atoms that are 1−2 and 1−3 bonded are omitted, which results in molecular polarizabilities smaller than the real ones. As a consequence, the ff02 force field might be slightly underpolarized.[36] Dynamics is performed using a recent

implementation of a PME-based approach to fixed and induced dipolar interactions.[19] All parameters being equal, including the time step, this approach increases the CPU time of a system with both charges and induced dipoles by a factor of 1.3 with respect to a system with only charges. We note that we did not employ the ff02EP force field because previous results[16] indicate that this force field, at its very best, only performs as well as ff02.

The water models used in the simulations to solvate DNA were TIP3P[37] (nonpolarizable, no extra points) for ff99 and POL3[38] (polarizable, no extra points) for ff02. Some of the polarizable solution simulations were also run with the RPOL water model, and the results essentially agree with those obtained with POL3. We also tried a crystal simulation with nonpolarizable DNA and POL3 waters; we obtained results intermediate to those described by the ff99 and ff02 runs—an indication that the water polarization is also important.

The MD simulations were carried out using the SANDER module of AMBER 7 and 8. Van der Waals interactions were calculated using an 8 Å atom-based nonbond list, with a continuous correction for the long-range part. The long-range Coulomb energy was evaluated by the PME method,[5,6] with a heuristic pair list update (and a 1.0 Å nonbonded pair list buffer), an 8 Å cutoff, an Ewald coefficient of 0.34864, and B-splines of order 4 for ff99 and order 6 for ff02. The production (data gathering) stage of the simulations was carried out using the Berendsen algorithm[39] at constant temperature (300 K) with a time constant of 0.5 ps and constant pressure (which led to an average pressure of 1 atm with the huge pressure fluctuations typical of such simulations). The time step was 2 fs for ff99 and 1 fs for ff02 (as a consequence, the CPU for the ff02 calculations was approximately 2.6 times the CPU for the ff99 calculations). The SHAKE algorithm was applied to all bonds involving hydrogen atoms. Coordinates were saved for analysis every 1 ps during these simulations. Calculation of DNA structural parameters was done using 3DNA (v 1.5).[40]

The starting coordinates for the simulations were standard Arnott B-DNA,[41] taken from the crystallographic PDB file with code 5DNB,[12] or from the NMR file with code 1NAJ.[17] The crystal structure is in space group *C*2 with five base pairs per asymmetric unit, with decamer double helices stacked atop one another along the *c* axis in a manner that approximates a continuous B helix. The PDB file contains coordinates for only one strand of the duplex, so the second strand was generated through the appropriate symmetry transformation. Different equilibration procedures were followed for the solution and crystal simulations, as detailed below.

A. Solution Simulations. Three sets of solution simulations were carried out. (i) sol-99 and sol-02 systems and (ii) sol-nmr-99 and sol-nmr-02 systems. The initial coordinates for the sol-99 and sol-02 simulations were specified by the crystallographic 5DNB PDB file. The DNA was placed in a truncated octahedron box and 2355 water molecules from a pre-equilibrated water box were added (crystal axes equal to 46.28 Å, angles of 109.47°, Fourier mesh sizes for PME equal to 48). To achieve electroneutrality, eighteen water molecules were then replaced by Na^+ ions. Those replaced waters were at least 11 Å apart from DNA heavy atoms and newly introduced ions. The system has a final density of approximately 1 g/mL, with a total of 7661 atoms. The initial coordinates for the sol-nmr-*99* and sol-nmr-02 simulations were specified by the first conformer in the 1NAJ[17] PDB file whose coordinates correspond to the high-resolution structure of the DNA dodecamer d(CGCGAATTCGCG)$_2$ determined in aqueous dilute liquid crystalline phase (the five conformers specified in the file are very close to each other). Again, an octahedron box was chosen, with 3841 water molecules and 22 Na^+ ions with a total of 12 303 atoms (crystal axes equal to 54.24 Å, angles of 109.47°, Fourier mesh sizes for PME equal to 54).

For both sets of systems, equilibration was done as follows. The ff99 force field was used in the initial coordinates setup. Relaxation with harmonic restraints on heavy DNA atoms and Na^+ ions was carried out to adjust the positions of water molecules in too close contact with other atoms. Subsequently, MD simulations with ff99 were conducted with restrained DNA at constant volume, slowly increasing the temperature from 0 to 300 K in five steps, for a total of 50 ps. This configuration was saved and used for a 100 ps constant volume simulation at 300 K with DNA restrained under the ff02 force field. The last restrained configurations at 300 K for both ff99 and ff02 were then used for unrestrained constant volume simulations for 100 ps at $T = 300$ K. The last configurations from these runs became the time-zero configurations for ff99 and ff02 constant pressure runs at 300 K for 12 ns. (iii) mix-99 and mix-02 systems: The two DNA duplexes and the Mg^{2+} ions as given in the 5DNB PDB file were used to generate the starting coordinates and 2617 waters were added for a total of 9137 atoms in an orthorhombic cell (crystal axes of 48.27, 43.59, and 52.52 Å; mesh for Fourier transform in PME of sizes 48, 48, and 54). The initial equilibration procedure was done with not only DNA but also Mg^{2+} ions restrained, during the initial temperature ramp. In addition, 500 ps of constant volume MD at 300 K without any restraints was carried out using the two force fields. The resulting coordinates were then used as the initial configurations for the constant pressure runs, which were 25 ns long. Strictly speaking, these systems are not "solution" simulations because the two duplexes interact strongly with each other (in addition to their own periodic images); their density is approximately 1.1 g/mL. We refer to these simulations as the mix-99 and mix-02 systems.

B. Crystal Simulations. The crystal unit cell was given the geometry specified in the 5DNB PDB file (crystal axes equal to 34.38, 32.25, and 25.53 Å; for the PME simulations, mesh sizes of 48, 48, and 32 were used). The starting coordinates are those of the idealized Arnott B-DNA (ABDNA) structure, as generated by the program NUCGEN in AMBER. The unit cell in the 5DNB file contains two duplexes related by a symmetry operation specified in the file.[12] This same operation is applied to the structure generated by NUCGEN to obtain the second duplex in the unit cell. Special care is taken to rotate the initial ABDNA structure so that the RMSD between the *two* simulation duplexes and those in the crystal are minimized and equal to that of individual duplexes, 1.45 Å. In the crystal environment, the Mg^{2+} ions are very slow to diffuse.[16] Indeed, runs with randomly placed Mg^{2+} ions show that these move too slowly to reach their "crystallographic" positions during our available computational time. We therefore decided to keep the approximate crystallographic positions of the Mg^{2+} ions during the generation of the unit cell, with 14 Mg^{2+} ions being present in the system. The parameters for Mg^{2+} were taken from AMBER parm91.dat because these perform better than those currently implemented.[15] For ff02 the ion has a small polarizability of 0.12. To achieve electroneutrality, eight Na^+ ions were added.

In addition to the runs with the randomly positioned Mg^{2+} ions, two sets of simulations were carried out: in one set, the positions of not only the Mg^{2+} ions but also of the crystallographic waters were preserved, (plus extra, bulk waters

obtained from a preequilibrated water bath); in the other set all waters were bulk. The final system had a total of 2882 atoms, and a density of about 1.4 g/mL. Equilibration started with the ff99 force field, with the main equilibration procedure as in ref 16. First, all hydrogens were minimized to avoid overlaps and bad orientations. This was followed by minimization of bulk waters and Na^+ ions, (in the set that contained crystallographic waters, the corresponding oxygens were restrained). Subsequently, MD simulations were conducted at constant volume, slowly increasing the temperature from 0 to 300 K in five steps of 10 ps each for a total of 50 ps. In these runs, all hydrogens, bulk water, and Na^+ ions were allowed to move, but all the atoms in DNA (except hydrogens) and Mg^{2+} ions (and crystallographic water oxygens when present) were kept near their crystallographic positions through harmonic restraints. Then, waters were allowed to equilibrate during an additional 500 ps constant volume MD at 300 K, with DNA and Mg^{2+} ions still restrained. The final coordinates of this last step were then used as initial coordinates for 500 ps constant volume runs at 300 K *without restraints* under both the ff99 and the ff02 force fields. These finally became the initial configurations for constant pressure runs for a period of 12 ns. We refer to the simulations without initial crystallographic water oxygens as crys-Mg-99 and crys-Mg-02 systems and to the simulations with crystallographic water oxygens as crys-Mg-O-99 and crys-Mg-O-02 systems. In addition, the simulations that were equilibrated in a similar fashion but with random Mg^{2+} ions and waters are referred to as crys-random-99 and crys-random-02 systems (strictly speaking, the ions are not totally random; water molecules far away from ABDNA and other ions are "converted" into Mg^{2+} ions).

III. Results

The choice of the DNA decamer $d(CCAACGTTGG)_2$ as a benchmark molecule was made on the basis of the fact that both crystal and solution simulations exist, against which to compare. However, there is no NMR data for this decamer, and therefore, comparisons are carried out against the crystal structure. To answer the question of how the solution simulations compare with respect to the NMR experiments, the DNA dodecamer $d(CGCGAATTCGCG)_2$ was chosen.

Figures 1−5 show the instantaneous RMSD (iRMSD) of all DNA atoms (except hydrogens) relative to the 5DNB crystal structure for the solution simulations sol-99 and sol-02; the "mixed" simulations mix-99 and mix-02; and the crystal simulations crys-Mg-99, crys-Mg-02, crys-Mg-O-99, crys-Mg-O-02, crys-random-99, and crys-random-02. Figure 6 shows the iRMSD of all DNA atoms (except hydrogens) relative to the 1NAJ NMR structure for the solution simulations sol-nmr-99 and sol-nmr-02. Table 1 gives the averaged iRMSD after the initial transient time and the RMSD between the *average structures* computed from the trajectories (adjusted for overall translations and rotations) and the experimental structure (aRMSD). For comparison, values obtained from a crystal simulation[16] that relaxes the original crystallographic structure have also been computed. From these figures and table, we draw the following conclusions:

(i) The results involving the decamer $d(CCAACGTTGG)_2$ in the solution simulations sol-99 and sol-02 (Figure 1) and the "mixed" simulations mix-99 and mix-02 (Figure 2) are qualitatively similar, despite the fact that the systems are very different, with different boxes, number of atoms, counterions, and densities. Moreover, the mix-99 and mix-02 systems are not strictly "solution" simulations, because the two duplexes in the system interact strongly. After an initial transient regime, the iRMSD of the sol-99 and sol-02 systems oscillate around 2.9 Å, with an aRMSD of about 2.6 Å. The corresponding numbers are, in general, slightly larger for the mix-99 and mix-02 systems. In other words, "solution-like" systems, tend to give an all-atom RMSD around or less than 3 Å with respect to the

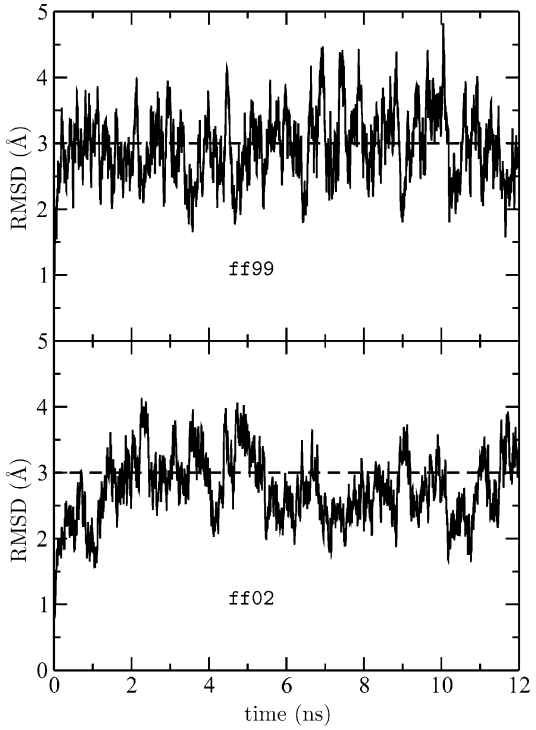

Figure 1. RMSD of all DNA atoms in $d(CCAACGTTGG)_2$ with respect to the crystal structure for the sol-99 (top) and sol-02 (bottom) solution systems. The dashed line is a guide for the eye.

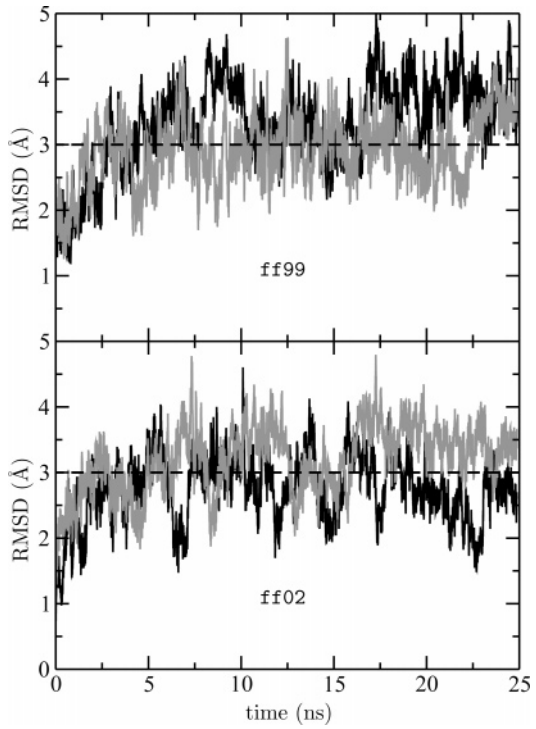

Figure 2. RMSD of all DNA atoms in $d(CCAACGTTGG)_2$ with respect to the crystal structure for the mix-99 (top) and mix-02 (bottom) "mixed" systems. Deviations for the two duplexes in the system are shown. The dashed line is a guide for the eye.

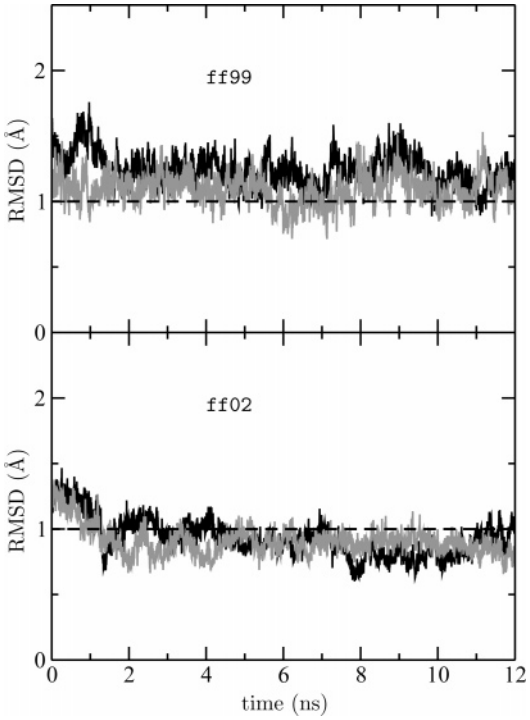

Figure 3. RMSD of all DNA atoms in d(CCAACGTTGG)$_2$ with respect to the crystal structure for the crys-Mg-99 (top) and crys-Mg-02 (bottom) crystal systems. Deviations for the two duplexes in the system are shown. The dashed line is a guide for the eye.

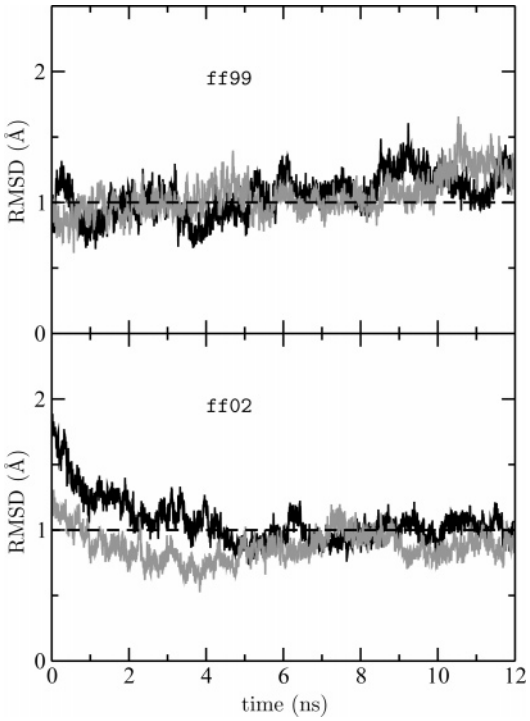

Figure 4. RMSD of all DNA atoms in d(CCAACGTTGG)$_2$ with respect to the crystal structure for the crys-Mg-O-99 (top) and crys-Mg-O-02 (bottom) crystal systems. Deviations for the two duplexes in the system are shown. The dashed line is a guide for the eye.

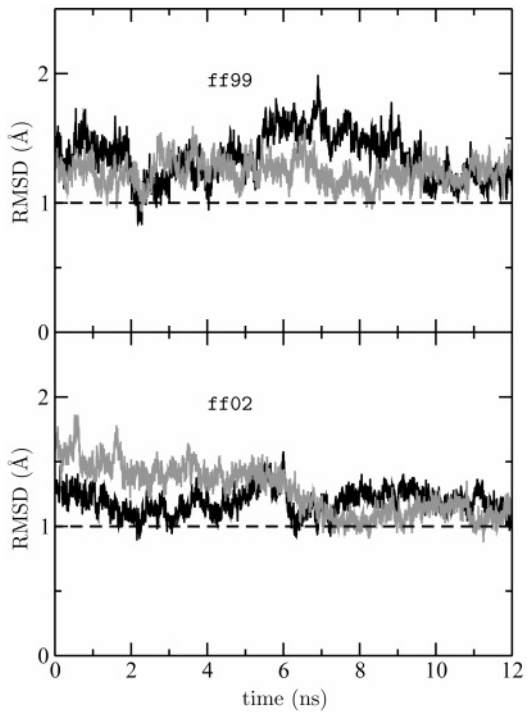

Figure 5. RMSD of all DNA atoms in d(CCAACGTTGG)$_2$ with respect to the crystal structure for the crys-random-99 (top) and crys-random-02 (bottom) crystal systems. Deviations for the two duplexes in the system are shown. The dashed line is a guide for the eye.

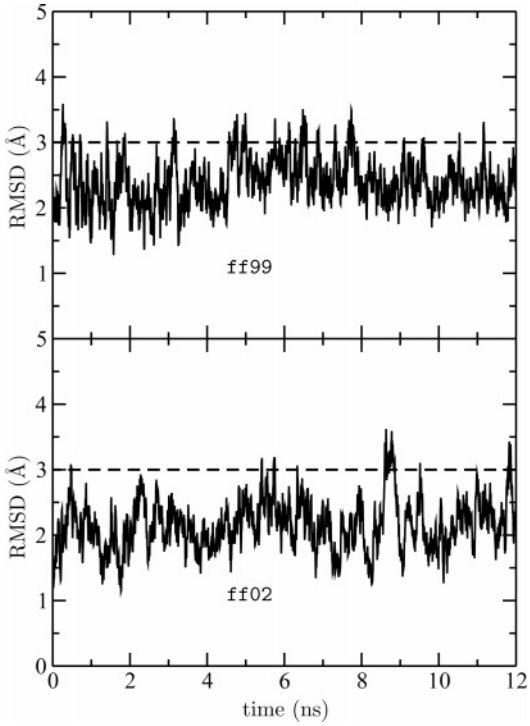

Figure 6. RMSD of all DNA atoms in d(CGCGAATTCGCG)$_2$ with respect to the NMR structure for the sol-nmr-99 (top) and sol-nmr-02 (bottom) solution systems. The dashed line is a guide for the eye.

crystal structure, even when the initial coordinates for both DNA duplexes and divalent ions are the same as those of the crystal.

(ii) The results involving the dodecamer d(CGCGAATTCGCG)$_2$ in the solution simulations sol-nmr-99 and sol-nmr-02 (Figure 6) show smaller RMSD with respect to the NMR structures than the ones reported in (i) for sol-99 and sol-02 (with respect to the crystal structure): the average of the iRMSD improves approximately by 25%, and the aRMSD improves approximately by 50%. The low value of the aRMSD (1.62 Å) for ff02 is also very encouraging for this type of simulation.

(iii) The RMSD obtained by taking initial crystallographic water oxygens in the crys-Mg-O-99 and crys-Mg-O-02 systems are comparable to the RMSD obtained with all waters placed

TABLE 1: All-Atom RMSD of the Molecular Dynamics Simulations Relative to the Experimental Structure (Either 5DNB or 1NAJ); Both the Average of the Instantaneous RMSD (iRMSD) as Shown in Figures 1−6 and the RMSD of the Average Structure (aRMSD) Are Listed[a]

simulation	time for average (ns)	iRMSD (Å)[b]	aRMSD (Å)[b]
sol-99	5−12	3.04	2.77
sol-02	5−12	2.69	2.40
mix-99	5−25	3.54/2.98	2.98/3.02
mix-02	5−25	2.82/3.33	2.48/3.26
crys-Mg-99	3−12	1.22/1.08	1.06/0.86
crys-Mg-02	3−12	0.88/0.89	0.63/0.72
crys-Mg-O-99	3−12	1.09/1.09	0.80/0.79
crys-Mg-O-02	3−12	0.99/0.86	0.81/0.64
crys-random-99	3−12	1.38/1.24	1.06/0.95
crys-random-02	3−12	1.20/1.21	1.01/1.01
ref 16, ff99	15−25	1.22/1.06	0.88/1.04
ref 16, ff02	15−25	0.87/0.80	0.70/0.58
sol-nmr-99	0−12	2.39	1.84
sol-nmr-02	0−12	2.15	1.62

[a] The definition of the various runs and force fields are given in the text. [b] When two RMSD values are present, they are the corresponding values for each duplex.

randomly in the crys-Mg-99 and crys-Mg-02 systems. The fact that crys-Mg-O-02 performs slightly worse than crys-Mg-02 is attributed to the initial positioning of crystal oxygens, which naturally do not match the ABDNA structure as well as the crystal structure. These can, therefore, be partially "trapped", taking a longer time to relax. Even though the average values in Table 1 would indicate that crys-Mg-O-99 has a smaller RMSD than crys-Mg-99, Figure 4 shows that at later times, crys-Mg-O-99 deviates more from the crystal structure and becomes comparable to crys-Mg-99.

(iv) The RMSDs for the crys-random-99 and crys-random-02 systems in Figure 5 are larger than those in Figures 3 and 4. This is just a reflection of the importance of the positions of the Mg^{2+} ions in the generation of the DNA crystal lattice.[42−44] The Mg^{2+} ions diffuse very slowly in the crystal simulations; therefore, the results in Figure 5 have not converged.

(v) With respect to the role of the polarizable force field ff02, it is noticed that the differences with respect to ff99 are small but consistent: the RMSD for ff02 are equal or smaller than those for ff99 in all of the systems investigated.

(vi) With both force fields, occasional fraying at the ends is observed. The fraying is negligible for the crystal structures and is more pronounced in the solution simulations. The average structure, computed without excluding any configuration, is well-behaved.

(vii) The most remarkable fact about the $d(CCAACGTTGG)_2$ crystal simulations with initial fixed Mg^{2+} ions is that the RMSD actually decreased from its initial value: unlike the $d(CCAACGTTGG)_2$ solution simulations that started from the crystallographic structure, the crystal simulations started from idealized ABDNA structures that were 1.45 Å away from the crystal structure. After an initial equilibration period, the structures converged to the crystal structure with an aRMSD for each duplex of 1.06 Å/0.86 Å for crys-Mg-99 and 0.63 Å/0.72 Å for crys-Mg-02.

(viii) It is interesting to compare these results to those of ref 16. There, unrestrained crystal simulations with the ff99 and ff02 force fields were shown to *preserve* sequence-dependent features extremely well in comparison to those reported with slightly older force fields.[14,15] These simulations started from the crystal structure, which was then relaxed and equilibrated for approximately 600 ps before constant pressure runs 25 ns long. The structural features were preserved without overly constraining the molecules within the crystal lattice, as evidenced by the B factors. For ff02, the aRMSD was relatively the same for the entire time regime: during the 0−15 ns time regime it was 0.602/0.671 Å and during the 15−25 ns time regime it was 0.698/0.578 Å. Instead, the aRMSD increased for ff99 with the values of 0.715/0.876 Å (0−15 ns) and 0.879/1.044 Å (15−25 ns). Because these ff99 results worsened with time, the results for crys-Mg-99 and crys-Mg-O-99 are expected to follow a similar trend. As a further proof of convergence, we can compute the aRMSD between the structures from this previous work and those obtained in the present simulations. The aRMSD between the ff99 crystal structures of ref 16 (15−25 ns) and crys-Mg-99 are 0.769/0.791 Å and the analogous comparison for the ff02 structures and crys-Mg-02 gives 0.725/0.600 Å.

Much of the data characterizing the DNA decamer $d(CCAACGTTGG)_2$ structure from the unrestrained simulations is summarized in Table 2, which gives standard angles and helicoidal parameters, averaged over all the residues, base pairs, or base-pair steps (where appropriate) for the various systems. The average values have been calculated in two ways: either by determining the values for each structure at every time step and then averaging the results over time, or by computing the average duplex structures and then obtaining the corresponding parameter. Both ways give essentially similar values. Of course, because most parameters are strongly dependent on the sequence, the average values for each parameter convey only reduced information.[53] In addition, the standard deviations for the individual pairs/steps are—at the very worst—10 times larger than the standard deviations obtained in Tables 2 and 3. The duplexes in all of the systems analyzed correspond to a B-DNA type of structure. However, certain base pairs in the solution simulations can exhibit instant A and TA forms.

Now we consider specific features pertaining to the decamer $d(CCAACGTTGG)_2$. With respect to the backbone torsional parameters, the behavior obtained from these trajectories is very similar to that reported previously, either in solution[15] or in crystal environment.[16] The α and γ torsions are in the g^-/g^+ ($-60/+60°$) canonical conformation. The glycosidic angle χ oscillates around $-100°$ (or 260°) in the crystal and $-115°$ (or 245°) in solution, which corresponds to the anti conformation of the right-handed double-stranded helices (although the magnitude of the angle tends to be more characteristic of A-DNA than of B-DNA). The torsion angles ($\epsilon-\zeta$) display strong sequence-dependent features, with some bases assuming the BI backbone conformation ($-90°$) and others the BII backbone conformation ($+90°$).[45] A time representation of ($\epsilon-\zeta$) reveals that there are transitions between BI and BII at several bases. All in all, the behavior of these parameters is statistically similar to that reported in previous work, and no qualitative difference is observed between ff99 and ff02.

Figures 7−9 show several sequence-dependent parameters.[54] In these figures, the solid line represents the MD results; the dashed line, the 5DNB crystal structure; and the dotted line, the ideal ABDNA structure, which is sequence-independent except for the small variations in the helical twist, due to the definition of local helical axis in 3DNA.[55] All crystal simulations were started from the dotted, featureless line and evolved into the sequence-dependent solid line.

Among the sugar conformational parameters, we examined the phase angle of pseudorotation of the sugar ring as a function of the base for strand I. The results are presented in Figure 7. All figures show clear sequence-dependent features with the crystal runs showing—as expected—better correlation with the

TABLE 2: Standard Angle and Helicoidal Parameters, Together with Standard Deviations, Averaged over All the Residues, Base Pairs, or Base-Pair Steps (Where Appropriate) for the d(CCAACGTTGG)$_2$ Decamer[a]

parameter[b]	5DNB	ABDNA	sol-99	sol-02	crys-Mg-99	crys-Mg-02	crys-Mg-O-99	crys-Mg-O-02
α (deg)	296.5	313.2	290.1 ± 2.94	289.7 ± 3.56	289.2 ± 3.19	286.7 ± 2.83	290.4 ± 3.64	287.7 ± 3.05
β (deg)	166.4	214.0	170.6 ± 3.17	171.8 ± 3.32	170.0 ± 3.29	167.9 ± 2.43	172.0 ± 3.87	170.0 ± 2.97
γ (deg)	49.59	36.35	53.61 ± 3.25	60.61 ± 3.74	53.65 ± 4.45	53.71 ± 2.71	57.16 ± 5.06	54.35 ± 4.59
δ (deg)	128.8	156.4	124.7 ± 4.18	128.2 ± 3.60	138.6 ± 2.35	136.1 ± 2.58	137.1 ± 3.27	136.5 ± 2.35
ϵ (deg)	203.2	155.0	197.1 ± 5.58	196.5 ± 5.06	214.7 ± 4.94	211.2 ± 3.99	215.8 ± 7.92	210.8 ± 3.72
ζ (deg)	244.4	264.9	259.8 ± 5.79	264.4 ± 5.35	238.2 ± 5.18	243.2 ± 4.22	244.5 ± 7.13	242.5 ± 4.07
χ (deg)	257.8	262.1	243.0 ± 4.77	248.0 ± 4.62	260.5 ± 3.48	262.3 ± 2.92	257.8 ± 3.35	261.7 ± 3.44
phase (deg)	145.4	191.8	135.9 ± 6.33	141.0 ± 5.58	158.2 ± 4.19	154.3 ± 4.26	156.1 ± 4.76	155.0 ± 3.98
amplitude (deg)	38.82	35.65	38.85 ± 1.37	38.86 ± 1.36	38.39 ± 1.28	38.53 ± 1.21	38.41 ± 1.25	38.61 ± 1.17
x-disp (Å)	0.45	−0.01	−2.05 ± 0.63	−1.62 ± 0.56	0.21 ± 0.27	0.22 ± 0.26	0.15 ± 0.27	0.17 ± 0.26
y-disp (Å)	0.0	0.0	0.03 ± 0.31	0.03 ± 0.32	0.07 ± 0.26	−0.09 ± 0.26	−0.03 ± 0.29	−0.04 ± 0.25
inclin. (deg)	5.11	−5.94	9.17 ± 3.48	10.17 ± 3.61	3.92 ± 2.19	5.58 ± 1.93	4.53 ± 2.64	4.77 ± 2.06
tip (deg)	0.0	0.0	0.0 ± 0.00	−1.78 ± 2.40	−0.19 ± 1.91	−0.19 ± 2.00	−0.35 ± 2.09	0.38 ± 2.21
shear (Å)	0.0	0.0	0.0 ± 0.10	−0.09 ± 0.14	−0.01 ± 0.09	0.04 ± 0.10	0.02 ± 0.09	0.01 ± 0.10
stretch (Å)	−0.13	−0.13	−0.03 ± 0.04	−0.04 ± 0.07	−0.06 ± 0.04	−0.08 ± 0.04	−0.05 ± 0.04	−0.08 ± 0.04
stagger (Å)	0.08	0.01	−0.05 ± 0.16	−0.09 ± 0.17	0.23 ± 0.16	0.22 ± 0.15	0.18 ± 0.17	0.25 ± 0.16
buckle (deg)	0.0	0.0	0.04 ± 4.35	−3.17 ± 4.50	−1.48 ± 4.74	0.37 ± 4.29	0.86 ± 5.25	−0.75 ± 4.38
propeller (deg)	−10.25	4.20	−10.42 ± 3.15	−13.65 ± 3.18	−7.52 ± 2.34	−9.21 ± 2.41	−9.96 ± 2.79	−9.03 ± 2.33
opening (deg)	0.75	−5.55	0.08 ± 1.42	1.09 ± 2.03	0.19 ± 1.24	0.27 ± 1.22	−0.09 ± 1.27	0.22 ± 1.34
shift (Å)	0.0	0.0	−0.02 ± 0.17	0.10 ± 0.18	−0.02 ± 0.12	0.06 ± 0.13	0.03 ± 0.14	0.0 ± 0.14
slide (Å)	0.78	−0.34	−0.57 ± 0.27	−0.30 ± 0.25	0.52 ± 0.15	0.65 ± 0.14	0.52 ± 0.19	0.59 ± 0.14
rise (deg)	3.31	3.36	3.37 ± 0.07	3.39 ± 0.08	3.37 ± 0.04	3.34 ± 0.04	3.36 ± 0.04	3.34 ± 0.04
tilt (deg)	0.0	0.0	0.03 ± 1.17	0.85 ± 1.22	0.07 ± 0.98	0.12 ± 1.01	0.42 ± 1.07	−0.03 ± 1.08
roll (deg)	2.23	−3.64	4.65 ± 1.87	5.51 ± 1.90	1.56 ± 1.20	2.24 ± 1.06	1.79 ± 1.48	1.79 ± 1.21
twist (deg)	35.24	35.61	31.72 ± 1.28	31.93 ± 1.30	34.99 ± 0.58	34.68 ± 0.44	34.75 ± 0.55	34.84 ± 0.40

[a] The average values have been calculated by determining the values for each structure at every time step and then averaging over time. The numbers thus obtained are in good agreement with the values calculated from each average structure. For reference, values for experimental (5DNB) and ideal Arnott B-DNA (ABDNA) structures are also given. [b] The different parameters are given either in angstroms (Å) or degrees (deg).

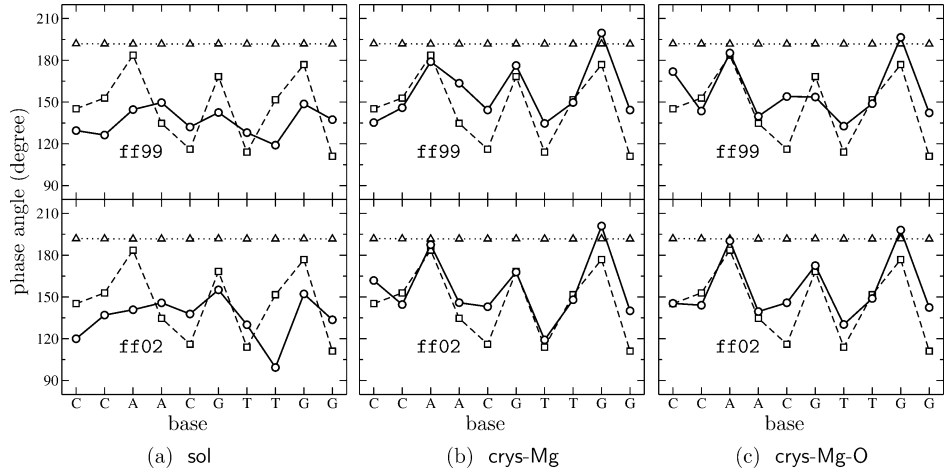

Figure 7. Phase angle of pseudorotation of the sugar ring for strand I as a function of the base. Solid line: simulation, dashed line: 5DNB, dotted line: Arnott B-DNA. Average is over last 7 ns for solution runs sol-99 and sol-02 and over last 9 ns for crystal runs crys-Mg-99, crys-Mg-02, crys-Mg-O-99, and crys-Mg-O-02.

experimental data. Depending on the base, the time-averaged puckering observed in the crystal simulations is C1′-exo, C2′-endo, or C3′-exo. However, sugar repuckering is observed at different times.

In general, the local base-pair parameters present too much variation for the averages to be meaningful. As an example of local base-pair *step* parameters, we focus on roll and helical twist, shown in Figures 8 and 9, respectively. Both present sequence-dependent features, even in the solution runs, where the two force fields are in relatively good agreement. Experimentally, roll is negative for the CA/TG and AC/GT steps and positive for the CC/GG, AA/TT, and CG/CG steps. Positive and negative values therefore alternate. This feature is not reproduced in the solution runs shown in Figure 8a, but it is remarkably reproduced by the crystal ff02 runs shown in Figure 8b, c. The helical twist reaches its lowest values in the CC/GG and AA/TT steps, its highest values in the CA/TG and CG/CG steps ("YR/YR" type of step), and its intermediate values in the AC/GT steps ("RY/RY" type of step). It has been pointed out before that solution simulations of DNA with the Cornell et al. force field typically result in an undertwisting of the duplex, such that values for twist are lower than those observed in crystal structures.[14,46] This is indeed observed in Figure 9a and Table 2. However, crystal simulations, as those shown in Figure 9b,c, reproduce the helical twist fairly well, a result that is believed to derive from the constraints placed upon the DNA duplexes within the unit cell.[15]

The simulation B factors (defined by the atomic position fluctuations multiplied by $8\pi^2/3$) are presented in Figure 10. The good agreement between the B factors calculated from the crystal simulations and the experimental ones indicates that crystal simulations can well represent the dynamics of the duplex

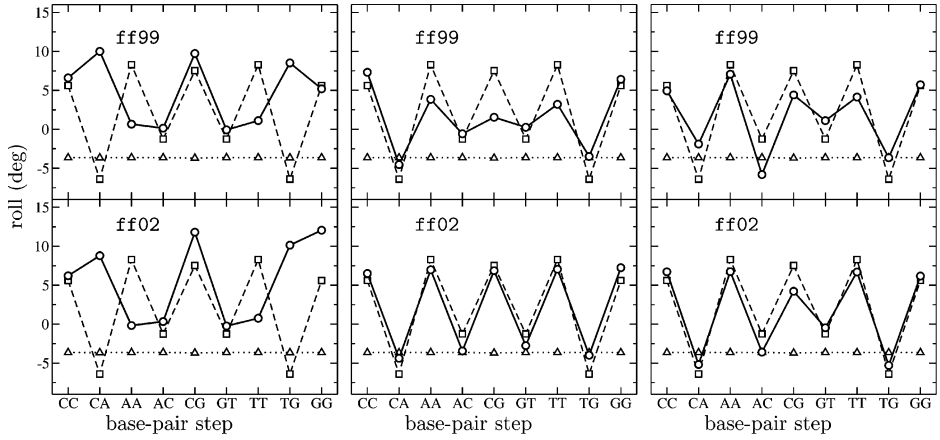

Figure 8. Roll as a function of base-pair step. Solid line: simulation, dashed line: 5DNB, dotted line: Arnott B-DNA. Average is over last 7 ns for solution runs sol-99 and sol-02 and over last 9 ns for crystal runs crys-Mg-99, crys-Mg-02, crys-Mg-O-99, and crys-Mg-O-02.

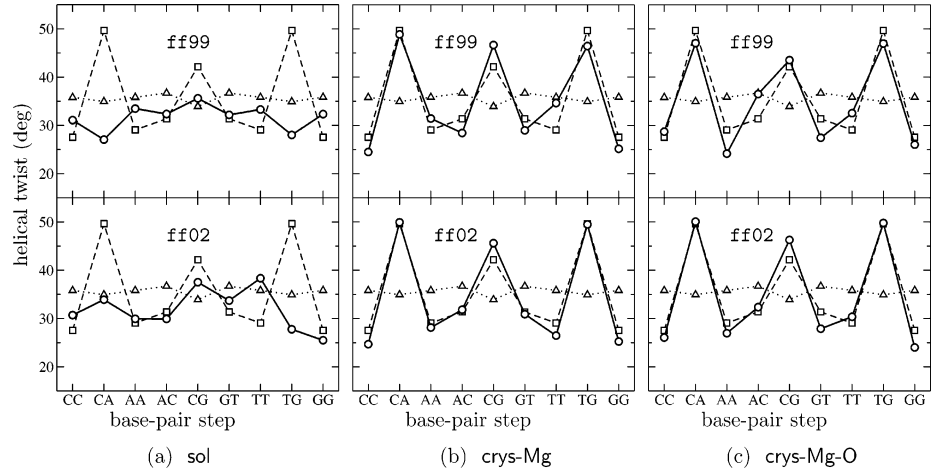

Figure 9. Helical twist as a function of base-pair step. Solid line: simulation, dashed line: 5DNB, dotted line: Arnott B-DNA. Average is over last 7 ns for solution runs sol-99 and sol-02 and over last 9 ns for crystal runs crys-Mg-99, crys-Mg-02, crys-Mg-O-99, and crys-Mg-O-02.

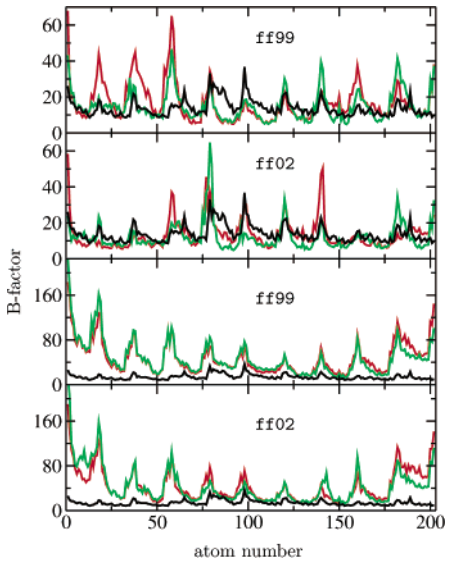

Figure 10. B factors as function of heavy atom number for the decamer d(CCAACGTTGG)$_2$. Red color represents strand 1; green, strand 2; and black, 5DNB crystallographic values. From top to bottom, the systems represented in each panel are crys-Mg-99, crys-Mg-02, sol-99, and sol-02.

in the crystal environment. As it is well-known, solution simulations provide considerably higher values than crystallographic B factors.

Finally, the crys-random-99 and crys-random-02 systems also display sequence-dependent features. Similarly to the RMSD shown in Table 1, the agreement with experiment is not bad but is poorer than in the other two crystal runs. For instance, the standard deviation of the *average* sequence-dependent roll with respect to the crystal structure goes from 3.33° and 3.04° for crys-Mg-99 and crys-Mg-O-99, respectively, to 4.13° for crys-random-99; and it goes from 1.64° and 1.71° for crys-Mg-02 and crys-Mg-O-02, respectively, to 4.25° for crys-random-02. Similar behavior is observed in the helical twist. The slight worsening of the agreement with experiments is expected, because the divalent ion distribution is not converged in these runs, and therefore, the DNA structure—most likely—is not converged either. Unfortunately, these simulations cannot predict when the Mg^{2+} ions would eventually converge because the ions move so slowly (particularly in the crystal environment) that their final distribution is very close to the initial one.

Now we briefly turn to the dodecamer d(CGCGAATTCGCG)$_2$. In this case, the standard deviations for different sequence-dependent quantities are approximately the same magnitude than the ones obtained for the solution simulations sol-99 and sol-02. However, the sequence-dependent features of the NMR structure are less pronounced than those of the crystal structure; therefore, the experimental structure falls within the standard deviations of the simulation, and comparisons are less meaningful. For instance, for the helical twist, the maximum amplitude of the NMR helical twist oscillations is about 7°,

TABLE 3: Standard Angle and Helicoidal Parameters, Together with Standard Deviations, Averaged over All the Residues, Base Pairs, or Base-Pair Steps (Where Appropriate) for the d(CGCGAATTCGCG)$_2$ Dodecamer[a]

parameter[b]	1NAJ	sol-nmr-99	sol-nmr-02
α (deg)	298.7 ± 1.40	290.6 ± 3.36	290.1 ± 3.24
β (deg)	171.3 ± 1.13	170.1 ± 2.90	170.9 ± 3.37
γ (deg)	50.58 ± 1.01	54.51 ± 3.09	59.43 ± 4.11
δ (deg)	124.1 ± 0.65	129.7 ± 3.41	127.8 ± 3.26
ϵ (deg)	190.4 ± 0.25	207.8 ± 6.79	195.1 ± 4.87
ζ (deg)	257.5 ± 0.18	257.1 ± 5.62	264.7 ± 5.29
χ (deg)	247.8 ± 0.52	253.3 ± 5.84	248.6 ± 4.50
phase (deg)	132.0 ± 0.66	142.3 ± 4.99	140.9 ± 4.95
amplitude (deg)	32.65 ± 0.21	38.87 ± 1.22	38.58 ± 1.23
x-disp (Å)	−0.86 ± 0.00	−1.37 ± 0.89	0.0 ± 0.00
y-disp (Å)	0.0 ± 0.00	0.11 ± 0.56	0.0 ± 0.25
inclin. (deg)	6.11 ± 0.67	6.16 ± 5.46	8.51 ± 2.98
tip (deg)	−0.01 ± 0.04	−1.62 ± 4.32	1.47 ± 1.95
shear (Å)	0.0 ± 0.00	−0.12 ± 0.15	0.03 ± 0.10
stretch (Å)	−0.37 ± 0.00	0.02 ± 0.21	−0.07 ± 0.04
stagger (Å)	−0.08 ± 0.00	0.0 ± 0.20	0.0 ± 0.14
buckle (deg)	0.01 ± 0.20	−1.31 ± 4.26	1.00 ± 4.11
propeller (deg)	−17.09 ± 0.64	−8.32 ± 3.37	−12.61 ± 2.69
opening (deg)	−0.35 ± 0.68	2.09 ± 3.00	0.13 ± 1.50
shift (Å)	0.0 ± 0.00	0.07 ± 0.19	−0.09 ± 0.14
slide (Å)	−0.18 ± 0.00	0.06 ± 0.37	−0.15 ± 0.24
rise (deg)	3.23 ± 0.03	3.21 ± 0.26	3.35 ± 0.07
tilt (deg)	0.01 ± 0.03	−1.17 ± 3.43	−0.87 ± 1.08
roll (deg)	3.57 ± 0.36	4.84 ± 3.31	4.75 ± 1.64
twist (deg)	35.45 ± 0.08	31.06 ± 2.98	32.48 ± 1.10

[a] The average values have been calculated by determining the values for each structure at every time step and then averaging over time (0−12ns). The numbers thus obtained are in good agreement with the values calculated from each average structure. For reference the values for experimental 1NAJ structure are also given. [b] The different parameters are given either in angstroms (Å) or degrees (deg).

but the standard deviations for the individual steps can be as large as 15°. In addition, some quantities—specially in the ff99 force field—present big deviations in the base pairs or steps at the ends, due to the "fraying" that we alluded to previously. This problem has been reported in the literature;[47] a usual "fix" is to cap the ends of the sequence with another CG pair (not done in this work). Overall averages are given in Table 3.

IV. Discussion and Conclusions

We have investigated how MD simulations reproduce DNA sequence-specific features, given different electrostatic descriptions and cell environments. With respect to the electrostatic descriptions, two different force fields available in AMBER have been used: ff99 (nonpolarizable atomic point charges) and ff02 (atomic point charges and induced atomic point dipoles). With respect to the cell environment, this work stresses the importance of the crystal or the solution environments and the structural importance of divalent ions.

First, consider the solution simulations. For the DNA decamer d(CCAACGTTGG)$_2$, the RMSD with respect to the crystal structure for sol-99 and sol-02, as well as for the hybrid mix-99 and mix-02 simulations, are qualitatively similar. After an initial transient regime, the iRMSD of the sol-99 and sol-02 systems oscillate around 2.9 Å with an aRMSD of about 2.6 Å. The corresponding numbers are, in general, slightly larger for the mix-99 and mix-02 systems. The values for the aRMSD for sol-99 and sol-02 are considerably lower than the values reported before (aRMSD ranging from 3.14 to 3.55 Å, for approximately 1 ns simulatio,n[14] and aRMSD equal to 4.22 Å for a 5 ns simulation[15]). The angle and helicoidal parameters definitely show sequence-dependent features, although naturally these correlate poorly with the crystal structure. For the dodecamer d(CGCGAATTCGCG)$_2$, the aRMSD with respect to the NMR structure is 1.84 Å for sol-nmr-99 and 1.62 Å for sol-nmr-02. These numbers probably indicate the best that one can obtain with present force fields in solution studies.

Now consider the crystal simulations. Previous work[16] showed that unrestrained crystal simulations with the ff99 and ff02 force fields *preserve* sequence-dependent features extremely well, when they start from the crystal structure. The present work goes one step further in showing that by starting from the idealized Arnott B-DNA structure, it is possible to recreate sequence-dependent features completely absent in the ABDNA structure. This is also shown by the values of the RMSD that decay from an initial value of 1.45 Å for ABDNA to an average crystal structure with an aRMSD for each duplex of 0.63/0.72 Å for crys-Mg-02. In addition, the aRMSD between the ff02 crystal duplexes in ref 16 (15−25 ns) and crys-Mg-02 are 0.725/0.600 Å. These numbers indicate that the crys-Mg-02 system has indeed converged to a minimum of the force field which is very close to the experimental crystal structure.

To identify what are the ingredients that contribute to this close convergence, we used counterexamples. In this way, we were able to identify two key ingredients:

(i) The cell environment: The crystal symmetry is carefully set up in the simulations (runs crys-Mg-99, crys-Mg-02, crys-Mg-O-99, and crys-Mg-O-02), so that the unit cell is exactly the same as used in experiments. As a counterexample, the RMSD increases to approximately 3 Å for a simulation that does not have a crystal environment, even though in this case the initial coordinates for the DNA and Mg^{2+} ions are exactly those of the experimental crystal (runs mix-99 and mix-02). Clearly, initial conditions do not play a role after equilibration; the solution simulations starting from the crystal structure lose the features characteristic of the crystal while the crystal simulations that start from an ideal, sequence-independent structure acquire the crystal features.

(ii) The positioning of the divalent ions (when present): The Mg^{2+} ions are initially placed approximately in their crystallographic positions (although allowed to move during equilibration). If, on the other hand, the crystal environment is correctly set up in the simulation but the Mg^{2+} ions are placed at random, as in runs crys-random-99 and crys-random-02, the RMSD are slightly larger and the sequence-dependent features agree less well with experiments. Unfortunately, the Mg^{2+} ions move so slowly (particularly in the crystal environment) that their final distribution is very close to the initial one. As a consequence, these simulations are not able to provide an answer to issues of convergence or distribution; they just point to the fact that the ions are structurally important.

The results for the crystal runs emphasize the crucial role of divalent ions, which are more important for the structure of duplex DNA than the monovalent ions.[44] Experimentally it has been found that divalent ions are key in stabilizing the crystal unit cells.[31] Even the type of divalent ion plays a big role in the resulting crystal. Thus, Mg^{2+} and Ca^{2+} ions generate different DNA crystal lattices and stabilize different end-to-end overlaps and lateral contacts between duplexes.[42−44] Moreover, monovalent cations not revealed by the crystallographic structures do not crucially affect the conformation of duplex DNA, because they do not directly coordinate to DNA atoms.[43] Interestingly, the simulations are sensitive enough to echo these experimental results. Thus, whereas hydrated Mg^{2+} ions stabilize the crystal by bridging the two helical duplexes within the unit cell, the 18 Na^+ ions have a higher mobility and are seen to

sample more freely the available space. These considerations imply that a good parametrization of the Mg^{2+} ion is crucial for accurate simulations of DNA where these ions are present. Mg^{2+} ions are particularly challenging because their interactions with nucleic acids involves polarization, charge transfer, and other molecular orbital effects, which generally require a quantum treatment.[48,49] Indeed, a classical calculation that compared simulation to experimental results concluded that deficiencies in the results concerning the Mg^{2+} ions were in part due to the lack of explicit polarization/charge-transfer terms for all molecules, including water, in the force field.[50]

Finally, with respect to the electrostatic description, the polarizable ff02 force field represents an attempt to achieve a better representation of the electrostatics. The crystal simulations indicate that the polarizable ff02 field in combination with a polarizable water model (POL3 or RPOL) have smaller RMSD and reproduce sequence-specific features better than the simpler ff99 on the nanosecond time scale, as long as divalent ion convergence is achieved. In addition, a good representation of divalent ions requires the inclusion of polarization effects, probably in a more sophisticated manner that the one currently implemented in AMBER. Improvements with the use of the polarizable ff02 are seen despite the simple nature of the present description: the dipoles are simply assumed to be proportional to the total electrostatic field, where the polarizability is the scalar constant of proportionality and no damping is assumed. In addition, the ff02 field seems to suffer from an underestimation of polarization effects.[36] The success of even this simple representation is understood from a physical point of view, because nucleic acids are highly charged polymers where polarization effects are bound to play a very important role. The ff02 field results give hope that good predictive DNA simulations may be achieved with future polarizable force fields, such as extensions to the AMOEBA force field[36,51,52] to include nucleic acids, with a relatively small increase in computational cost.

Acknowledgment. This research was supported by NSF under Grants ITR-0121361 and CAREER DMR-0348039. We also thank NCSA for extensive computer support.

References and Notes

(1) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. *Nature* **1980**, *287*, 755−758.
(2) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(3) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.
(4) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.
(5) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(6) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(7) Beveridge, D.; McConnell, K. *Curr. Opinion. Struct. Biol.* **2000**, *10*, 182−196.
(8) Cheatham, T. E., III; Young, M. *Biopolymers* **2000**, *56*, 232−256.
(9) Norberg, J.; Nilsson, L. *Acc. Chem. Res.* **2002**, *35*, 465−472.
(10) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350−357.
(11) Orozco, M.; Pérez, A.; Noy, A.; Luque, F. *Chem. Soc. Rev.* **2003**, *32*, 350−364.
(12) Prive, G.; Yanagi, K.; Dickerson, R. *J. Mol. Biol.* **1991**, *217*, 177−199.
(13) Chuprina, V.; Hienemann, U.; Nurislamov, A.; Zielenkiewicz, P.; Dickerson, R.; Saenger, W. *PNAS* **1991**, *88*, 593−597.
(14) Cheatham, T. E., III; Kollman, P. A. *J.Mol. Biol.* **1996**, *259*, 434−444.
(15) Bevan, D.; Li, L.; Pedersen, L.; Darden, T. *Biophys. J.* **2000**, *78*, 668−682.
(16) Baucom, J.; Transue, T.; Fuentes-Cabrera, M.; Krahn, J.; Darden, T.; Sagui, C. *J. Chem. Phys.* **2004**, *121*, 6998−7008.
(17) Wu, Z.; Delaglio, F.; Tjandra, N.; Zhurkin, V.; Bax, A. *J. Biomol. NMR* **2003**, *26*, 297−315.
(18) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(19) Toukmaji, A.; Sagui, C.; Board, J. A.; Darden, T. *J. Chem. Phys.* **2000**, *113*, 10913−10927.
(20) Sagui, C.; Pedersen, L.; Darden, T. A. *J. Chem. Phys.* **2004**, *120*, 73−87.
(21) Sagui, C.; Pomorski, P.; Darden, T. A.; Roland, C. *J. Chem. Phys.* **2004**, *120*, 4530−4544.
(22) Sagui, C.; Darden, T. A. *J. Chem. Phys.* **2001**, *114*, 6578−6591.
(23) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Comm.* **1995**, *91*, 1−41.
(24) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K.; Stanton, R.; Cheng, A.; Vincent, J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.; Singh, U.; Weiner, P.; Kollman, P. *AMBER 7*; University of California: San Francisco, 2002.
(25) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Wang, B.; Pearlman, D.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.; Ross, W.; Kollman, P. *AMBER 8*; University of California: San Francisco, 2004.
(26) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.
(27) York, D. M.; Yang, W.; Lee, H.; Darden, T. A.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.
(28) Lee, H.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *102*, 3830−3834.
(29) Lee, H.; Darden, T. A.; Pedersen, L. G. *Chem. Phys. Lett.* **1995**, *243*, 229−235.
(30) MacKerell, A.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(31) Yuan, H.; Quintana, J.; Dickerson, R. *Biochemistry* **1992**, *31*, 8009−8021.
(32) Young, M.; Jayaram, B.; Beveridge, D. *J. Am. Chem. Soc.* **1996**, *119*, 59−69.
(33) Várnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *320*, 4269−4280.
(34) Ponomarev, S.; Thayer, K.; Beveridge, D. *PNAS* **2004**, *101*, 14771−14775.
(35) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(36) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.
(37) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(38) Caldwell, J.; Kollman, P. *J. Phys. Chem.* **1995**, *99*, 6208−6219.
(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(40) Lu, X.; Olson, W. *Nucleic Acids Res.* **2003**, *31*, 5108−5121.
(41) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Comm.* **1972**, *47*, 1504−1509.
(42) Lipanov, A.; Kopka, M.; Kaczor-Grzeskowiak, M.; Quintana, J.; Dickerson, R. *Biochemistry* **1993**, *32*, 1373−1389.
(43) Minasov, G.; Tereshko, V.; Egli, M. *J. Mol. Biol.* **19998**, *291*, 83−99.
(44) Chiu, T.; Dickerson, R. *J. Mol. Biol.* **2000**, *301*, 915−945.
(45) Lavery, R. *Adv. Comput. Biol.* **1994**, *1*, 69−145.
(46) Cieplak, P.; Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6722−6730.
(47) Beveridge, D.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; 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. *Biophys. J.* **2004**, *87*, 3799−3813.
(48) Deerfield, D.; Pedersen, L. *J. Biomol. Struct. Dyn.* **1995**, *13*, 167−180.
(49) Muñoz, J.; Sponer, J.; Hobza, P.; Orozco, M.; Luque, F. *J. Phys. Chem. B* **2001**, *105*, 6051−6060.
(50) York, D. M.; Darden, T.; Deerfield, D.; Pedersen, L. G. *Int. J. Quantum Chem.* **1992**, *19*, 145−166.
(51) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497−1506.
(52) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(53) [The standard deviations (SD) are computed as the second moments of the distribution. To obtain an estimate of *error of measurement of the mean value*, one would have to compute the SD from a group of independent sampling points and then divide by the square root of the number of sampling events. For instance, if the configurations stored every 1 ps were independent, the *error of the mean value* during the last 10 ns of the simulations could be obtained by dividing the reported SD by $\sqrt{10\,000} = 100$. Of course, these configurations are not independent and one has to carry block averages, and the SD tend to increase with the size of the block.

We find that for block sizes as large as 3.5 ns, the error estimate is approximately only 5 times worse than the one computed in 1-ps windows. Even if one assumes that the real error is 10 times worse than that for the 1-ps windows, one would still have to divide the SD in Table 2 by approximately 10, which gives very small values for the error of measurement of the *mean value*.

(54) In the mixed systems, mix-99 and mix-02, the duplexes strongly interact in a complex way (whose study is beyond the scope of the present work) and are not analyzed here.

(55) Even in the ideal structure of B-DNA, the base pairs are not geometrically identical in terms of bond lengths, valence angles, complementary parameters. For example, the coordinate frame of a standard guanine is slightly different from a standard adenine and the base pairs are not coplanar nor perfectly aligned perpendicular to the double helical axis. The twist of consecutive base-pair frames can thus be slightly sensitive to sequence, due to the definition of the standard base (and base-pair) frame. If, instead of using a local helical axis at each base-pair step, a global linear helical axis is used, ABDNA produces the ideal value of 36°.