Abstract

BAUCOM, JASON BUTLER. Towards predictive molecular dynamics simulations of DNA: Role of electrostatics and of the cell environment. (Under the direction of Celeste Sagui.)

Molecular dynamics simulations of the DNA duplex $d(CCAACGTTGG)_2$ were used to study the relationship between DNA sequence and structure in a crystal environment. Three different force fields were used: a traditional description based on atomic point charges, a polarizable force field, and an “extra-point” force field (with additional charges on extranuclear sites). It is found that all the force fields reproduce fairly well the sequence-dependent features of the experimental structure. The polarizable force field, however, provides the most accurate representation of the crystal structure and the sequence-dependent effects observed in the experiment. These results point out to the need of the inclusion of polarization for accurate descriptions of DNA. This work has also 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 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. At later times, the polarizable force field is able to maintain this lower RMSD while the nonpolarizable force field starts to drift away.
Towards predictive molecular dynamics simulations of DNA: Role of electrostatics and of the cell environment

by

Jason Butler Baucom

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

BIOINFORMATICS

Raleigh

2006

APPROVED BY:

_________________________ _________________________
Christopher Roland   Steffen Heber

__________________________ _________________________
Maria Celeste Sagui   Carla Mattos
Chair of Advisory Committee
Dedication

To my loving wife and best friend: Adela.
Biography

Jason Baucom was born in Charlotte, NC in 1975. He attended school in Charlotte and spent most of his formative years in Matthews. In 1985 he moved to Wingate, NC, a small country town not too far from Charlotte. He attended Governors School in 1992. This summer school was a stark contrast to the small private school environment and further opened his eyes to the opportunities and intrigue of biology and chemistry. He started attending NCSU in 1993, working periodically in the computer science department. He was active in the Society of Physics Students, acting as vice-president his junior year and president his senior year. He graduated from NCSU in 1998 with a BS in physics with minors in computer science and mathematics, taking many electives in chemistry and biochemistry, further whetting his appetite for biology. He then worked for AT&T and Bank of America on Y2K computer issues in their human resources department. After the termination of those projects he returned to Raleigh to pursue his interest in merging his academic loves: biochemistry, physics and computer science. He found his home in the bioinformatics program studying under Dr Celeste Sagui of the Physics department. He was an IGERT fellow from 2002-2005. In December 2004 he was married to Adela Tordai Baucom in Lynchburg, Virginia.
# Table of Contents

List of Tables ..............................................................................................................vi  
List of Figures .............................................................................................................vii  

1: Introduction ............................................................................................................... 1  
   I. Crystallography ................................................................................................... 2  
   II. Molecular dynamics ....................................................................................... 3  
   III. Electrostatic treatments ....................................................................................... 6  
      A. Early attempts at electrostatics ............................................................... 7  
      B. Ewald base methods ........................................................................... 8  
      C. Fast Multipole Method ........................................................................... 9  
      D. Implicit water methods ......................................................................... 10  
      E. Polarizable force fields ......................................................................... 11  
      F. Extra-points ......................................................................... 12  
      G. Local environment ......................................................................... 12  
   IV. 5DNB and 1NAJ ..................................................................................... 13  

2: Molecular dynamics simulations of the d(CCAACGTTGG)₂ decamer in crystal  
   environment: Comparison of atomic point-charge, extra-point, and polarizable force  
   fields ......................................................................................................................... 15  
   I. Abstract ............................................................................................................. 15  
   II. Introduction ................................................................................................. 16  
   III. Methods for Crystal Simulations ...................................................................... 21  
   IV. Results ............................................................................................................. 25  
   V. Discussion and Conclusions ......................................................................... 32  

3: Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence  
   of an Ideal B-DNA Structure to the Crystallographic Structure  ......................... 57  
   I. Abstract ............................................................................................................. 57  
   II. Introduction ...................................................................................................... 58  
   III. Methods ............................................................................................................. 64  
      A. Solution Simulations ............................................................................... 66  
      B. Crystal Simulations ............................................................................... 68  
   IV. Results ............................................................................................................. 70
List of Tables

2.1 RMS deviations of structures relative to experimental crystal structure .......... 54
2.2 Helicoidal parameters from simulations ............................................................. 55
2.3 Standard deviations of time averaged, base parameters with respect to crystal ............. 56
3.1 RMSD of simulations relative to experimental crystal or solution structure .......... 94
3.2 Helicoidal and angle parameters for d(CCAACGTTGG)₂ simulations ................. 95
3.3 Helicoidal and angle parameters for d(CGCGAATTCGCG)₂ simulations ................. 96
4.1 RMS deviations of structures relative to experimental crystal structure .......... 121
4.2 Helicoidal and angle parameters for crystal simulations ....................................... 122
4.3 Error of averaged sequence-dependent features relative to crystal structure ............. 123
List of Figures

2.1 RMSD of two duplexes for ff99 simulation with respect to crystal structure .......... 37
2.2 RMSD of two duplexes for ff02EP simulation with respect to crystal structure .......... 38
2.3 RMSD of two duplexes for ff02 simulation with respect to crystal structure .......... 39
2.4 Torsional angles $\alpha$ and $\gamma$ as functions of time ............................................................. 40
2.5 Torsional angles $\alpha$ and $\gamma$ as function of sequence ............................................................. 41
2.6 Difference in torsional angles ($\epsilon - \zeta$) as function of sequence ......................... 42
2.7 Difference in torsional angles ($\epsilon - \zeta$) as functions of time ................................. 43
2.8 Phase angle of pseudorotation of sugar ring as function of base ......................... 44
2.9 Propeller twist as a function of base pair ................................................................. 45
2.10 Opening as a function of base pair ............................................................................. 46
2.11 Roll as a function of base-pair step ............................................................................. 47
2.12 Helical twist as a function of base-pair step ............................................................. 48
2.13 Helical twist as a function of time in the ff02 simulation ......................................... 49
2.14 Minor groove width as function of base-pair step ..................................................... 50
2.15 Distance between Mg$^{2+}$-45 and phosphorus DNA atoms ........................................ 51
2.16 RMS deviations of Mg$^{2+}$ atoms ..................................................................................... 52
2.17 B factors from simulation and crystal structure ......................................................... 53
3.1 RMSD of sol-99 and sol-02 simulations with respect to crystal structure .......... 84
3.2 RMSD of mix-99 and mix-2 simulations with respect to crystal structure .......... 85
3.3 RMSD of crys-Mg-99 and crys-Mg-02 simulations with respect to crystal structure .......................................................................................................................... 86
3.4 RMSD of crys-Mg-O-99 and crys-Mg-O-02 simulations with respect to crystal structure ................................................................. 87

3.5 RMSD of crys-random-99 and crys-random-02 simulations with respect to crystal structure ................................................................. 88

3.6 RMSD of sol-nmr-99 and sol-nmr-02 simulations with respect to NMR structure . 89

3.7 Phase angle of pseudorotation of the sugar ring as a function of base ............................... 90

3.8 Roll as a function of base-pair step ........................................................................... 91

3.9 Helical twist as a function of base-pair step ............................................................. 92

3.10 B factors as a function of heavy atom number ......................................................... 93

4.1 RMSD of crys-MG-O and crys-random simulations with respect to crystal structure .................................................................................................113

4.2 Phase angle of pseudorotation of the sugar ring as a function of base ................. 114

4.3 Buckle as a function of base-pair .......................................................................... 115

4.4 Propeller as function of base-pair .......................................................................... 116

4.5 Opening as function of base-pair .......................................................................... 117

4.6 Tilt as function of base-pair step .......................................................................... 118

4.7 Roll as function of base-pair step .......................................................................... 119

4.8 Helical twist as function of base-pair step ............................................................... 120
1: Introduction

Since the early days of molecular dynamics (MD), scientists have long sought to accurately model and understand the dynamic properties of biomolecules with atomic detail. MD provides a full atom dynamic picture of a molecule, providing resolution of detail that no current physical method can achieve. MD seeks to translate the static (or mostly static) detail of crystallography, nuclear magnetic resonance (NMR) and/or electron paramagnetic resonance (EPR) to a more kinetic picture of the system. This picture is developed by applying numerical solutions to Newton’s equations of motion to systems of molecules and allowing the system to move accordingly. MD explores the dynamic nature of biomolecules, and since biological processes are inherently dynamic, it is a perfect tool to analyze certain biological phenomena. Most current MD simulations are in the nanosecond time scale, with a few approaching microseconds (e.g. Duan, 1997). This time frame encompasses a large segment of biochemically interesting phenomena, but several complex activities, such as protein folding, which can extend into the hour range (Huang, 1995, Bryan, 1992), are still out of reach. Currently the largest computational time bottleneck of classical simulations and the largest source of error is the calculation of forces due to electrostatics (Sagui, 1999). Early MD simulations have demonstrated sensitivity to the proper treatment of electrostatics (York, 1993, Steinbach, 1994, Schreiber, 1992), and since electrostatics terms are long-range and their effects decay very slowly, it is computationally expensive to correctly treat and a bottleneck in current simulations.
I. Crystallography

For MD researchers it is important to have biologically meaningful and realistic molecules to study. The starting points for MD simulations are normally full atom pictures of biomolecules characterized by crystallography, NMR, EPR or some other physical technique that produces atomic coordinates. Pure DNA structures have been crystallized since 1981 when the Dickerson group crystallized a B-DNA dodecamer and a short section of Z-DNA (Drew, 1981, Drew, 1981-2). Long strands of normal helical A-, B-type DNA are notoriously difficult to crystallize (Neidle, 1998), with the longest strand crystallized currently being 12 base pairs in length (Drew, 1981 being one). Since Dickerson’s first two DNA structures submissions to the Protein Data Bank (PDB) (Berman, 2000) over a 1000 DNA structures have been submitted, and in recent years at a pace of slightly under a hundred a year. Compared to proteins, in excess of 30,000 at a pace currently near 5000 new submissions yearly, DNA submissions are quite far behind. RNA structures are being generated at an even slower pace, with fewer than 450 pure RNA molecules currently submitted. Many structures have been characterized that include both protein and nucleic acid with nearly 1500 nucleic acid/protein systems currently catalogued. Several non-standard form DNA systems have been crystallized or solved via NMR, including the telomeric region quadruplex (e.g. Wang, 1993 with 22 bases) and triplex form DNA (e.g. Radhakrishnan, 1994 with 31 bases). The general difficulty of crystallizing DNA presents a bottleneck for generating structural coordinates required for MD researchers to work with. While programs exist to generate ideal canonical DNA (or RNA) from first principles, there has been no evidence that such molecules are biologically reasonable or that simulations can converge to biologically accurate structures or the criteria required for such a convergence. Showing convergence of
canonical B-DNA to a more biologically accurate structure would open the world of MD to a plethora of possible structures. The consequences of the practical and physical limitations of crystallography, EPR and NMR can be diminished and physical methods that do not generate atomic-scale data, such as fluorescence, CD or other experimental methods could possibly be used to test hypothesis generated from MD simulations. While modern force fields may not be sophisticated enough to ensure proper RNA/protein folding, a good first step would be to demonstrate the predictive power for DNA systems.

Crystallography gives coordinates for heavy solute atoms, relatively stable heavy ions and occasionally crystallographic water oxygens. Water molecules still have a high degree of mobility in crystal structures, but crystallographic waters can be observed when water molecules become trapped or statistically are stable in a particular space region. This can occur with a single trapped water molecule or through water exchange (Halle, 1998). MD researchers have determined water exchange to occur in the time scale of tens of picoseconds to several hundred picoseconds (Bonvin, 1998), well within the time frame of modern simulations. They are generally close to the biomolecules surface, hydrogen bonding with the solute, forming a shell of hydration with ions or in hydrophilic regions of the solute or space filling (Kopka, 1983). In DNA, this is often observed as a spine of hydration along the minor groove (Denisov, 1997).

II. Molecular Dynamics

The evaluation of Newton’s equations requires a proper characterization of all of the atoms and forces involved. Solved structures and analogous chemical groups have been used
to calculate physical parameters required for the evaluation of Newton’s equations and these parameters are incorporated in the MD packages through force fields and atomic parameters. These parameters can be obtained experimentally by analogous groups or through quantum calculations. Biologically reasonable parameters for all aspects of the molecular system are required and the simulations can be quite sensitive to the accuracy of these parameters (Bevan, 2000). Improper models of ions have been shown to lead to artifacts in the simulation. As the fundamental understanding of the nature of biomolecules has developed and novel computational techniques have matured, force fields have improved. The proper treatment of electrostatics has always been a difficulty, and various techniques for the efficacious and proper treatment of electrostatics have developed with varying degrees of success. The second chapter of this thesis seeks to quantify the improvements of recent AMBER force fields over previous force fields and to show the potential improvements that incorporating polarizability or extra-points may prompt in crystal simulations of B-form DNA. These improvements are generally recognized by the reduction of the all-atom root mean square deviation (RMSD) in comparison to previous incarnations of AMBER force fields and by the reproduction of sequence specific features in relation to the crystal structure.

There are two general types of environments in which to perform MD on a system: periodic boundary conditions (PBC) and non-periodic boundary conditions (NPBC). Periodic boundary conditions can replicate the cell environment of X-ray crystallography structures, mimicking the structure composed of, for all practical purposes, an infinite number of identical cells. By mimicking the cell environment and cell conditions, molecular dynamics
can attempt to closely reproduce crystallographic results. It has been argued that structures obtained through crystallography may not be biologically accurate, since the biomolecules are artificially constrained (Koo, 1990?), but most comparisons between structures obtained through different methods (NMR, EPR) demonstrate a fairly close correlation (Wagner, 1992). Dickerson has argued that crystal packing effects play a minimal role in structure determination and the observation of sequence specific features (Dickerson, 1994). In fact, some crystallized enzymes have actually been shown to metabolize substrates while crystallized (e.g. Doscher, 1963). While DNA is difficult to crystallize, methods that determine structure in solution, such as NMR and EPR, currently have limitations on the size of structures that can be determined. The relative infrequency of hydrogen in DNA also makes NMR difficult, although recently there have been improvements to overcome this difficulty. PBC simulations do have certain potential problems. Some have theorized that the biomolecule will be influenced by artifacts created by the solute “feeling” each periodic image and the effects of crystal packing (See Kastenholz, 2004 for a review). Dickerson observed that crystal packing effects tend to play a minimal role in sequence specific characteristics of DNA (Dickerson, 1994). PBC simulations can also perform solvated simulations by expanding the cell dimensions and adding appropriate amounts of water. This also lessens the effects of feeling periodic solute images and the argument regarding lattice restrictions.

Non-periodic boundary conditions (NPBC) isolate a single image of the solved structure in a box. This method resolves the potential self-energy artifact, but it has its own unique dilemmas. In order for the solvent to remain in a box a rigid barrier is imposed on the
system. If a molecule touches the barrier, it is rebounded back. When this rebounding occurs, the improper handling of the rebounding forces can cause unrealistic energies in some simulations. Different treatments for barrier dynamics have been attempted, varying the manner in which the colliding molecule is rebounded, trying to make the transition smooth and energetically realistic for the system.

III. Electrostatic Treatments

Since the incept of the application of molecular dynamics to biological systems, electrostatics has posed a serious computational problem. Electrostatics is the longest reaching term in the potential energy equation and the slowest to decay. Several studies have shown that accurate treatment of electrostatic potentials is critical for accurate simulations of biological systems (York, 1993, Scheiber, 1992, Steinbach, 1994). The partial charge representations of atoms in current popular molecular dynamic software packages, such as AMBER (Pearlman, 1995), CHARMM (Brooks, 1993) and GROMOS (Gunsteren, 1988), insure that biological models are replete with charges. Electrostatic interactions are exceptionally important biologically, with effects ranging from salt bridges, hydrogen bonds, electrostatic steering, protein-protein interaction and DNA recognition (Sinha, 2002). Any interactions through space are generally considered electrostatic in nature. Characteristically DNA and RNA are more strongly charged than protein, having a negatively charged phosphate backbone, evidencing the critical need for the correct treatment of electrostatics in nucleic acid simulations. While most proteins may not be as highly charged, the partial charge representation of all atoms, the polar nature of many biologically relevant interactions and bonds and the typical presence of several charged residues require a proper treatment of
electrostatics (Sagui, 1999). The effect may not be as dramatic as in DNA simulations, but even a marginal increase in accuracy may provide a more biologically meaningful result. Many biological systems have been characterized that include both protein and nucleic acids, so the correct treatment of electrostatics will most definitely improve the simulations of these systems and help explain and provide insight into the binding strength, specificity and dynamic nature of these relationships.

A. Early attempts at electrostatics:

When researchers applied molecular dynamics simulation techniques to biological systems computer speed and memory was minimal. A direct treatment of the electrostatic term scales as \(O(N^2)\) with \(N\) representing the number of atoms in the system, which is computationally intractable as system sizes get large. Biochemical systems vary in size from 10’s to millions of atoms in size, and as systems get much larger than 1000, direct evaluation of electrostatics becomes computationally costly. The effects of long-range electrostatics were not well understood or characterized at the time these simulations began, so in an effort to save time and resources, electrostatic terms were dealt with by invoking a strict cut-off. As simulations progressed, DNA molecules would act in a biologically uncharacteristic manner and disassociate (for a brief review, see Cheatham, 2000). In order to obtain more stable trajectories, artificial constraints or the modification of the highly charged phosphate residue were added. Further investigation pointed to correct treatment of electrostatics as the prime culprit for simulation problems and instabilities (e.g., see York, 1993). Efforts were made to smooth the cut-off transition by invoking shifting (Brooks, 1983, Brooks, 1985) or switching functions (Brooks, 1983). Shifting functions multiply the \(1/r\) electrostatic term by a function
that slowly decays to zero at a predetermined cut-off. Shifting functions tend to distort short range interactions of charged groups and in general tend to overestimate forces. Switching functions treat the electrostatic potential exactly to a predetermined $R_{on}$ region and apply a function that decays to zero from a region $R_{on}$ to $R_{off}$. This tends to result in better represented water molecules in simulations, but forces are still overestimated (Feller, 1996). These functions are useful in that they keep a strict cutoff and thus minimal computation time, but the simulations are still inaccurate. Different techniques needed to emerge to more accurately treat electrostatics.

**B. Ewald based methods**

In 1993 Darden et al applied a method previously developed by Ewald to effectively deal with electrostatics in these systems (Darden, 1993). Ewald based methods had originally been developed to deal with the accurate treatment of long-range interactions between particles and all of their periodic images. (Ewald, 1921). Ewald based methods decompose the electrostatic $1/r$ term into error functions, specifically $1/r = \text{erfc}(r)/r + \text{erf}(r)/r$ (the complementary error function is given by $\text{erfc}(x) = 1 - \text{erf}(x)$).

The benefit of such analysis is that $\text{erfc}(r)/r$ decays very fast, and therefore a cutoff can be used to compute this term directly in real space. Therefore the calculation of the “direct sum” with a fixed cutoff becomes $O(N)$. The term $\text{erf}(r)/r$, on the other hand, does not decay fast but varies very smoothly and can be computed via Fourier transforms. The insight of the Particle Mesh Ewald Method was to compute the reciprocal part very efficiently by interpolating the particles onto a grid and using Fast Fourier Transforms (FFT) to calculate
the electrostatic potential. Forces are computed in reciprocal space and then interpolated back to real space via B-Splines. The Ewald based methods have proven a marked improvement over previous electrostatics work with comparable or decreased computation time as compared to cut-off techniques that were employed at the time of implementation (York, 1993). PME simulations scale as \( O(N \ln N) \), which is a marked improvement over the \( O(N^2) \) of the Ewald exact treatment. PME provides stable DNA simulations that extend well into the microsecond range and has become the method of choice for most explicit solvent biomolecular simulations. Part of the speed accomplished by the PME technique depends on the rapid calculations afforded by FFT’s. Unfortunately, the numerical calculation of FFT requires global communication, and this creates problems in parallelization. As computing systems start heading in a massively parallel direction, other techniques that parallelize well and still maintain the same degree of accuracy may be favored in the future.

C. Fast Multipole Method

The Fast Multipole Method (FMM) was developed as an alternative to the Ewald based approach to electrostatics by Greengard and Rokhlin (Greengard, 1987, Greengard, 1987-2, White, 1994). FMM recursively segments the system into smaller and smaller octal sub-cells. Coulombic interactions are treated exactly for particles inside of neighboring sub-cell, which are short range in nature. Electrostatic interactions outside of the neighboring sub-cell are handled via multipole expansions. The splitting of Coulombic interactions into local and distant interactions is not a smooth function of \( R \), and may be a potential source of inaccuracies. The method is \( O(N) \), in theory a great savings over Ewald based methods. In practice, it is not much of a savings over PME’s \( O(N \ln N) \) scaling on single processors.
(Pollock, 1996), since there is a large overhead for splitting cells and very high order
multipoles are needed to achieve energy conservation (Bishop, 1997). With newer and
extremely efficient implementations of FFT, FMM codes to not scale better than PMEMD
(the newest, highly parallelizable version of PME implemented in AMBER 8 and 9).

D. Implicit water methods

Computationally it is very inefficient to deal explicitly with waters. Running heavily
solvated simulations can greatly increase the number of atoms in a system, which in turn
slows a simulation down dramatically. As an alternative for explicit representations of water,
Generalized Born (GB) fields were developed (Constanciel, 1984, Still, 1990). GB fields
remove most explicit waters and apply a background dielectric, allowing the system to
effectively “feel” the presence of water. This technique may be valuable for saving time if
there is minimal concern about how specific waters affect the system and interact with the
biomolecule. If explicit waters are suspected to play a minimal role in the system’s dynamics,
the GB approach may be appropriate. Hybrid procedures have been developed that create
water shells, perhaps one or two shells of explicit water around the biomolecule. In this
treatment, electrostatics outside of the water shells is treated using a GB approach while
explicit waters are dealt with explicitly. This gets around the problem of biologically
important water interactions while minimizing computational time. Removing waters may
not be biologically accurate, but it does speed up simulations. Ewald based simulations tend
to scale like O(NlogN), so anything that can be done to reduce the number of atoms will
dramatically reduce computational time. Since removing waters will reduce viscosity,
conformational space can be sampled more rapidly than biologically expected or though
explicit solvent techniques (Ponder, 2002). Another approach is taken by the reaction field procedure. Rather than ignoring waters or including a small water shell, reaction field (RF) techniques seek a compromise by maintaining a sufficiently hydrated system but only explicitly treating waters within a certain radius (Onsager, 1936). Waters outside of this radius are ignored and a Generalized Born type treatment is applied. By doing this, some computational time is saved while the accuracy dealing with local water interactions is maintained. Close waters can be dealt with by the explicit evaluation of modern approaches already discussed. While implicit water methods may save computational time and have some merit, they will lack the delicate electrostatic interactions of individual water molecules and ions that inevitably play an important role in electrostatics.

**E. Polarizable force fields**

Chemically speaking, atoms do not exist in space as simple partial charges. In fact, electron clouds are always distorted by the local electronic environment (Griffiths, 1989). The maturation of current force fields has recently encompassed the polarizable nature of biomolecules. It has been hoped that this progression will provide biologically more reasonable simulations. The first attempt at explicitly calculating dipoles was implemented as a recursive process (Saboungi, 1988, Sprik, 1988, Wilson, 1993). Polarizable forces are calculated and every dipole is modified according to the electrostatic environment. However, once a single dipole is modified, it affects every other dipole in the system. At this point, recalculation of all other dipoles is required. This process is repeated until convergence is reached within a predetermined tolerance. As with most recursive processes the analysis is very computationally intense. As one alternative, the original Carr-Parrinello technique
(Carr, 1985) was modified and applied to the situation (Toukmaji, 2000). Using this technique, dipoles are treated as dynamics variables, adding additional degrees of freedom to the system. These additional equations are solved using modified LaGrangian techniques. The procedure does not require strict convergence and is computationally much more tractable. Improvements to force fields are detailed in the second and third chapters.

**F. Extra-Points**

Some force fields in recent years have attempted to represent lone pairs in simulations, but the proper characterization of extra points in AMBER has proven difficult, in particular calculating the van Der Waals surface (Case, 2005). At present, charges representing extra points in AMBER do not have a van der Waals surface, which can lead to an electrostatic “catstrophy” when the extra points become too close to other charges. In AMBER lone pairs of electrons in DNA simulations can be present in base pairs and water oxygen. Nucleic acids carry extra points exclusively on base groups and optionally water, so the effects on the overall dynamics may not be dramatically affected. Factoring in additional biologic information does not necessarily ensure more accurate simulations, particularly if the additions are difficult to characterize. The second chapter characterizes the accuracy of current descriptions of extra point representations in DNA simulations.

**G. Local environment**

The local environment surrounding DNA plays an important role. Crystal simulations maintain the cell symmetry of the crystallographic cell. Simulations are often started from crystallographic coordinates, which generally include heavy ion coordinates. The effects of
the local environment and ions in these simulations are very important to understand the resulting DNA structure. Divalent ions have been shown to have more effect on standard duplex form DNA conformation (Yuan, 1992) and diffuse slower than monovalent ions (Bevan, 2000). Solvated simulations extend the simulation beyond the cell environment and allow a greater deal of translational and rotational motion for the solute. This is normally associated with greater deviations of all atom root mean square deviations (RMSD). Crystal packing effects have been of concern, but B-factors for most simulations demonstrate that the system is typically not more constrained than crystal structures and other groups have demonstrated the crystal packing effects have been minimal (for a review, see Dickerson, 1994). The role of the cell environment and of divalent ion placement in the lattice is explored in the third chapter.

IV. 5DNB and 1NAJ

As a baseline for studying the effects of electrostatics we have chosen to study the DNA segment detailed in the 5DNB pdb file. This file contains coordinates and information necessary to generate a duplex, with two identical duplexes of the double stranded decamer d(CCAACGTTGG)₂. The sequence was crystallized and the structure determined by the Dickerson group and published in 1991 (Prive, 1991). The DNA was synthesized in the lab, and not characteristic of any particular biologically interesting DNA sequence. Biologically speaking, the sequence is not uniquely interesting, but the existing work with previous simulations that employed older force fields provides a good baseline for comparison for this work. Cheatham and Kollman used MD on the 5DNB system to demonstrate the transition of A-DNA to B-DNA in aqueous solution (Cheatham, 1996). Bevan et al used this structure to
detect the effects of crystal environment, demonstrating the preservation of sequence specific features (Bevan, 2000). These studies used the Cornell force field (ff94) and analyzed MD trajectories for both crystal and solvated simulations. Since that time, several updates for AMBER force fields have occurred and the treatment of electrostatics has not been well characterized with DNA molecules or compared with previous simulations. The accuracy of polarizable force fields or extra-points in DNA simulations with relation to older force fields were unknown, and chapters two and three take an effective step in that direction.

Currently there is no solved NMR structure for the d(CCAACGTTGG)\textsubscript{2} DNA sequence. There may be a practical limit of accuracy when comparing solvated simulations against crystal structures, so the Dickerson dodecamer d(CGCGAATTCCGCG)\textsubscript{2} (Drew, 1981) was used to test the limit of how close, gauged by RMSD, current force fields can replicate an NMR structure in a solvated simulation. 1NAJ is a NMR structure of the Dickerson dodecamer published by the Wu group in 2003 (Wu, 2003). Like 5DNB, 1NAJ is a synthetic construct and not of great biological importance, though it does contain a recognition sequence for the EcoRI restriction enzyme. It was the first large crystallized B-form DNA segment (Drew, 1981) and has been used in many theoretical studies (e.g. Ravishanker, 1989, Swaminathan, 1991, Duan, 1997).
Molecular dynamics simulations of the d(CCAACGTTGG)$_2$ decamer in crystal environment: Comparison of atomic point-charge, extra-point, and polarizable force fields

Jason Baucom$^1$, Thomas Transue$^2$, Miguel Fuentes-Cabrera$^1$, J. M. Krahn$^2$, Thomas A. Darden$^2$, Celeste Sagui$^1$

$^1$ Department of Physics, North Carolina State University, Raleigh, North Carolina 27695
$^2$ National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received 7 April 2004; accepted 9 July 2004

Abstract

Molecular dynamics simulations of the DNA duplex d(CCAACGTTGG)$_2$ were used to study the relationship between DNA sequence and structure in a crystal environment. Three different force fields were used: a traditional description based on atomic point charges, a polarizable force field, and an “extra-point” force field (with additional charges on extranuclear sites). It is found that all the force fields reproduce fairly well the sequence-dependent features of the experimental structure. The polarizable force field, however, provides the most accurate representation of the crystal structure and the sequence-dependent effects observed in the experiment. These results point out to the need of the inclusion of polarization for accurate descriptions of DNA.
I. INTRODUCTION

The relationship between DNA sequence and structure is a subject of considerable interest, and its understanding is believed important for essential cellular processes such as replication and transcription. The relationship between DNA sequence and structure is currently probed by experimental methods such as x-ray crystallography and nuclear magnetic resonance (NMR), and theoretical methods such as molecular dynamics (MD) simulations. Currently, x-ray crystallography offers the most precise picture of DNA sequence dependent structure. However, due to difficulties in crystallization, this technique has been limited to small oligomers (e.g., no more than dodecamers for A- or B-form DNA, in the absence of protein). In addition, due to its flexibility, DNA may be sensitive to crystal packing artifacts which may complicate the effects of sequence dependence on structure. NMR offers a complementary solution-phase experimental picture which is free of crystal packing artifacts. Sequence-dependent dynamics as well as structure can be studied using NMR. However, due to the generally extended structure of DNA as well its relative scarcity of protons, NMR structures tend to be of lower precision than those provided by x-ray crystallography (long-range information from residual dipolar couplings improves the precision considerably).

At present, a large number of molecular dynamics theoretical approaches are being used to study DNA structure. MD studies are important because they allow for a description of DNA structure and dynamics at the atomic level, and thus provide a complementary tool to the experimental methods. Until fairly recently (i.e., up to ten years ago) MD simulations of DNA were known to be unstable, and usually produced unrealistic outcomes; e.g.,
partially or fully melted duplex DNA, within a few hundred picoseconds of simulation. This instability was traced to numerical artifacts due to the nonbond cutoff schemes in use at the time (York, 1993, York, 1995), which were particularly severe in the case of DNA with its high net charge. With the introduction of methods to efficiently calculate Coulombic terms while reducing (Steinbach, 1994) or largely eliminating (Darden, 1993, Essmann, 1995) artifacts due to long-ranged electrostatics, long stable simulations of DNA became commonplace. Several review articles discuss issues concerning MD simulations of nucleic acids (Beveridge, 2000, Norberg, 2002, Cheatham, 2000) as well as the importance of the correct treatment of electrostatics (Sagui, 1999). With the recent better force fields coupled to accurate methods, an abundance of information has been obtained regarding the dynamics, structure, and energetics of nucleic acids, as well as the role of water and ions and their specific interactions with DNA.

When new force fields are tried, one of the major tests to verify their correctness involves crystal simulations that study whether the experimentally determined structure can be maintained without seriously constraining the molecule (Cheatham, 2000). These simulations start from the crystal structure and maintain the crystal environment throughout the simulation. Crystal simulations can be challenging because the high packing density may inhibit atomic fluctuations. In contrast, solution simulations fluctuate perhaps too much with respect to experiments (Steinbach, 1994). Also, constant pressure simulations are recommended for crystal environments in order to avoid the problem of high pressures that may develop during a constant volume simulation, which can lead to additional motion inhibition. Examples of early successful crystal simulations were carried out by Darden and
co-workers. An MD simulation of the dodecamer \(d(CGCGAATTCCCG)\)_2 in the crystal unit cell yielded a stable trajectory and an average structure with a root mean square (RMS) deviation for all heavy atoms of 1.2 Å from the crystal structure (York, 1995). Similarly, an MD simulation of Z-DNA\(d(CGCGCG)\) in its crystal environment produced an average structure with an RMS deviation of 1.16 Å from the crystal structure (Lee, 1995). MD simulations of the RNA dinucleotides ApU and GpC (Lee, 1995) were found to be within 0.4 Å of the crystal structures, and the calculated and experimental temperature factors were comparable. Other crystal simulations that go back to the same time are those performed by MacKerell and co-workers (MacKerrell, 1995), who carried out several short-time simulations. The structure that we consider in this work, the \(d(CCAACGTTGG)\)_2 decamer, has been studied before. Cheatham and Kollman (Cheatham, 1996) performed solution simulations of this structure. Out of four MD simulations, two were started from canonical B-DNA and two from canonical A-DNA. All simulations converged to B-DNA structures in less than 1 ns. These structures were within 0.8–1.6 Å RMS deviation of one another and 3.1–3.6 Å RMS deviation of the published crystal structure (Prive, 1991). A more recent study is that of Bevan et al. (Bevan, 2000), where they carried two crystal simulations—one consisting of one unit cell containing two duplexes, and the other of two unit cells containing four duplexes—and two solution simulations. The authors studied the relation between DNA sequence and structure and found that for many parameters crystal and solution simulations were the same, but for other parameters such as some backbone torsion angles, and phase and helical twist, there were quantitative differences, with the solution simulations diverging from the experimental results.
A general picture that emerges from these and other simulations is that structural features characteristic of B-DNA (for experimental structures that can be characterized as such) are maintained over several nanoseconds. In solution simulations not all sequence-dependent features are preserved (e.g., helical twist), but crystal simulations tend to preserve the experimental features. This indicates that crystal simulations may be used to facilitate the identification of force-field parameters that need to be adjusted to improve agreement between experiment and simulation. In the present study we compare the performance of different force fields in reproducing the structure of \( d(\text{CCAACGTTGG})_2 \) decamer in relatively long (25 ns) crystal simulations. These force fields, based on the Cornell et al. force field (Cornell, 1995), are the \( \text{ff98, ff99, ff02, and ff02EP} \) force fields. The \( \text{ff98} \) and \( \text{ff99} \) force fields use the atom types, topologies (except for torsional parameters), and charges form the Cornell et al. force field to create an all-atom nonpolarizable force field for proteins and nucleic acids. The \( \text{ff98} \) force field differs from the \( \text{parm94.dat} \) force field (the one used by Bevan et al.) in the torsion angle parameters involving the glycosidic torsion in nucleic acids, while \( \text{ff99} \) is similar to \( \text{ff98} \) except to additional changes in the torsion angle parameters. In this work we are not so much interested in the bonded-atom parametrizations as in the electrostatics characterization. The chosen force fields characterize different electrostatic representations. In \( \text{ff98} \) and \( \text{ff99} \) the electronic charge density is simply represented by atomic point charges. The \( \text{ff02} \) force field is a polarizable variant of the Cornell et al. force field, where the electronic charge density is represented by both atomic point charges and induced atomic point dipoles. The \( \text{ff02EP} \) force field is a polarizable force field with extra points on electron-donating atoms. These atoms carry additional point charges which represent electron lone pairs.
It has long been known that accurate electrostatics is a key issue for improving current force fields for large-scale biomolecular simulations. Typically, this calls for an improved and more accurate description of the molecular electrostatic potential, which eliminates the artifacts associated with current point charge-based descriptions. In this sense the polarizable \textit{ff02} and extra-point \textit{ff02EP} force fields represent attempts to move forward in accuracy. The crystal studies described below generally indicate that the polarizable \textit{ff02} indeed performs better in reproducing sequence specific features than the simpler \textit{ff98} and \textit{ff99}; however \textit{ff02EP} is not as good as \textit{ff02}. The results for the polarizable \textit{ff02} are slightly surprising if one considers 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 (Ren, 2002) is assumed. However, after initial equilibration, the simulations run smoothly and the results speak for themselves. On a physical basis, these results are less surprising: nucleic acids are highly charged polymers where polarization effects are bound to play a very important role. The results presented here are very encouraging for future parameterizations that include higher order multipoles and polarization (Sagui, 2004). Indeed, these calculations are feasible from a computational point of view. Recently, Sagui \textit{et al.} presented an efficient simulation scheme for such a description using the Cartesian tensor formalism (Sagui, 2004). Essentially, the long-range electrostatic interactions are divided in two sums according to the usual Ewald scheme: the \textit{direct} sum, which evaluates the rapidly varying, particle-particle interactions with a McMurchie-Davidson formalism (McMurchie, 1978), considered up to a given cutoff in real space; and the “reciprocal” sum, which evaluates the smoothly varying, long-range part of the interaction. The reciprocal part of the problem was treated in three different ways: using
the standard Ewald scheme, a particle mesh Ewald (PME) (Darden, 1993, Essmann, 1995), based formalism, and a multigrid-based approach (Sagui, 2001). It was found (Sagui, 2004) that a highly accurate calculation of the interactions up to hexadecapoles costs only a factor of 8.5 times more a regular AMBER simulation using monopoles (charges) only. For comparison, a straightforward implementation which includes hexadecapoles (i.e., 35 independent multipolar components in a cartesian representation) may be expected to be $O(35^2=1225)$ times more expensive than a treatment based on monopoles only. An earlier version of this work (Toukmaji, 2000) included an implementation of classical Ewald (Smith, 1982) and PME based treatments of fixed and induced point dipoles into the SANDER molecular dynamics module of AMBER6 and AMBER 7, (Case, 2002, Pearlman, 1995) together with a Car-Parrinello scheme for the computation of the induced polarization. This implementation is the one that has being used for the computation of the induced dipoles and their interactions. Higher order multipoles and polarization have already been successfully implemented for protein force fields (Ren, 2002, Ren, 2003, Ponder, 2003), the results presented here clearly indicate that this is the route for the accurate representation of DNA.

II. METHODS FOR CRYSTAL SIMULATIONS

Four different MD simulations were carried out in this project, using four force fields implemented in AMBER 7 (Case, 2002). The four all-atom force fields are as follows:

(i) The Cornell et al. force field, in its 1998 and 1999 version. These are a modification of the force field developed by the Kollman group in the 1990s (Cornell, 1995). The simulations by Bevan et al. (Bevan, 2000) were carried out using the parm94.dat force
field; the *parm98.dat* and *parm99.dat* force fields differ from *parm94.dat* in the torsion angle parameters for the glycosidic torsion in nucleic acids (Cheatham, 1999). We refer to these fields as *ff98* and *ff99*.

(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. Dynamics are performed using a recent implementation of a PME based approach to fixed and induced dipolar interactions (Toukmaji, 2000). During MD the induced dipoles can be propagated along with the atomic positions either by iteration to self-consistency at each time step or by a Car-Parrinello technique using an extended Lagrangian formalism.

(iii) The *ff02EP* force field is a polarizable force field with extra points on electron-donating atoms that carry additional point charges (Dixon, 1997) (which therefore would represent electron lone pairs). In particular, the nucleic acids carry extra points only on the bases and not on the sugars or phosphate groups.

The water models used for each field were (i) TIP3P (Jorgensen, 1983) (nonpolarizable, no extra points) for *ff98* and *ff99*, (ii) POL3 (Caldwell, 1995) (polarizable, no extra points) for *ff02*, and (iii) TIP5P (Mahoney, 2000) (nonpolarizable, extra points) for *ff02EP*. The parameters for Mg$^{2+}$ were taken from AMBER *parm91.dat*, because they have
been tested to perform better than those currently implemented (Bevan, 2000). For *ff02* and *ff02EP* the ion has a small polarizability of 0.12.

MD simulations were carried out using the SANDER module of AMBER7. Simulations were conducted at 300 K using a 2 fs time step for *ff98* and *ff99*, and 1 fs time step for *ff02* and *ff02EP*. The SHAKE algorithm was applied to all bonds involving hydrogen atoms. 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.

The starting coordinates for the crystal simulations were taken from the PDB file with code *5dnb* (Prive, 1991). The PDB file contains coordinates for only one strand of the duplex, so the remaining strand was generated through the appropriate symmetry transformation. A second duplex was added to the unit cell by translating the initial duplex according to guidelines in the file (Prive, 1991). The crystallographic positions of the Mg$^{2+}$ ions were maintained during the generation of the unit cell, with 14 Mg$^{2+}$ ions being present in the system. Crystallographic waters were added first by applying the appropriate symmetry transformations as was done to generate the duplex structures and then by adding hydrogens using the GWH program within AMBER. We started equilibration with the *ff98* force field. The LEAP program in AMBER was used to generate the topology and coordinate files containing the positions of DNA atoms, Mg$^{2+}$ ions, and crystallographic waters. Bulk water molecules were then added to the system and minimization was done to adjust the positions of water molecules that were in close contact with other atoms in the system. To
achieve electroneutrality, eight Na$^+$ ions were added. The final system had a total of 2882 atoms and a density of 1.41 g/ml. The equilibration process was done in steps: First, we minimized all hydrogens, to avoid overlaps and bad orientations, and followed with minimization of bulk waters and Na$^+$ ions. Subsequently, MD simulations were conducted at constant volume, slowly increasing the temperature from 0 to 300 K in five steps of 8 ps each, for a total of 40 ps. In these runs, all hydrogens, bulk water, and Na$^+$ ions were allowed to move, but all the atoms in DNA (except hydrogens), Mg$^{2+}$ ions and crystallographic water oxygens, were fixed in their crystallographic positions. Then we ran MD at constant volume and 300 K for 50 ps more, in the same conditions but with the crystallographic oxygens not fixed but constrained. Test unconstrained constant pressure runs indicated that the system had the right number of waters and the right density—a fact further corroborated by the very long constant pressure simulations described below. After this, we went back to the last configuration of the constrained constant volume simulation and carried out a 500 ps constant volume equilibration without restraints under each of the four force fields. This became the initial configuration for constant pressure runs for a period of 25 ns. Coordinates were saved for analysis every 1 ps during these simulations.

Calculation of DNA structural parameters was done using 3DNA (v1.4.1) (Lu, 2000, Lu, 2003). This is a comprehensive, user-friendly software package for the analysis, reconstruction, and visualization of three dimensional nucleic acid structures by Lu and Olson.
III. RESULTS

We found that results for \textit{ff98} and \textit{ff99} are—perhaps not surprisingly—very similar and not much meaning can be attributed to their differences which fall within statistical errors. Bearing this in mind, we show results for either \textit{ff98} or \textit{ff99} (mainly \textit{ff99}).

The atomic RMS deviations, relative to the starting structure, for the two duplexes in the crystal unit cell simulations for each of the force fields are shown in Fig. 2.1 (\textit{ff99}), Fig. 2.2 (\textit{ff02EP}), and Fig. 2.3 (\textit{ff02}). The long simulations allow for a careful analysis of the stability of the trajectory. For each of the two duplexes simulated under each force field, Table 2.1 gives the averaged RMS deviations for the last 20 ns (third column) and the last 10 ns (fourth column) as well as the maximum RMS deviation (last column). For comparison, values obtained for a \textit{parm94.dat} simulation (Bevan, 2000) have also been computed. A comparison of the figures and the results in Table 2.1 show that the RMS deviations from the crystal structures, whether averaged or maximum, are better for \textit{ff02} and definitely worse for the older \textit{parm94.dat} force field. Notably, \textit{ff02EP} does not provide a remarkable increase of accuracy with respect to \textit{ff99}.

Much of the data pertaining to DNA structure from the unrestrained simulations is summarized in Table 2.2. Analysis of the helicoidal parameters of each duplex calculated over the last 5 ns of the simulation shows only small differences between the duplexes, as might be expected based on their similar RMS deviations (Table 2.1). For each force field, the standard deviations in each helicoidal parameter are comparable for each duplex, and therefore it is possible to express the helicoidal parameters as a simple average over the two
duplexes. We also checked that the values reported in the 20–25 ns time interval do not differ substantially from the values obtained in a 5–25 ns time interval; the quantities that experience the largest variation with the time interval chosen for averaging are the backbone torsional parameters. It should also be noticed that the total values for each parameter sometimes convey only reduced information, since most parameters are strongly dependent on the sequence as will be shown below. Both experimental and simulation data clearly indicate a B-DNA structure.

First, we consider some backbone torsional parameters. In Fig. 2.4, we show results for strand 1 for the torsional angles $\alpha$ and $\gamma$ as a function of time. There is an apparent correlation between these angles: $\alpha$ decreases in time by approximately the same amount that $\gamma$ increases in time. An example of this approximate anticorrelation is shown in Fig. 2.5 which shows $\gamma$ as function of sequence for one of the strands in duplex 2 for $ff02$ (data obtained as an average over the 20–25 ns time interval). Figures 2.4 and 2.5 show discrete changes in these angles that can be attributed to changes among different structural forms. In particular, after 5 ns, $ff98$, $ff99$, and $ff02$ do not show steady “drift” in the values of the angles, only some discrete changes. This is not the case for $ff02EP$, where a steady drift seems to be superposed to the discrete changes. It has been pointed out, however, that these angles can be trapped in long-time metastable states and therefore present noncanonical backbone geometries (Varnai, 2002). Another backbone torsional parameter of interest is the glycosidic angle $\chi$. During the entire time regime this angle oscillates around a value of approximately 2100 (or 260), 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). This agrees well with the experimental data. Some interesting quantitative features can also be obtained for the backbone torsional angles \(\epsilon\) and \(\zeta\). The difference \((\epsilon - \zeta)\) is often used, because it assumes values that are characteristic (Lavery, 1994) of the \(BI\) backbone conformation \((-90^\circ)\) and the \(BII\) backbone conformation \((+90^\circ)\). In the crystal structure, the values of \((\epsilon - \zeta)\) for C-2 and T-8 are indicative of the \(BII\) conformation while the others are characteristic of the more common \(BI\) conformation. The sequence-dependent nature of this parameter is clearly shown in Fig. 2.6, where \((\epsilon - \zeta)\) averaged over the 25 ns time regime is shown as a function of the base for strand 1. The polarizable force field \(ff02\) agrees extremely well with the crystal data, while \(ff02EP\) gives an incorrect maximum for A-4. The standard deviation of the time-averaged, base-dependent values of \((\epsilon - \zeta)\) with respect to the crystal values are reported in Table 2.3. Figure 2.7 shows values of \((\epsilon - \zeta)\) as a function of time for the \(ff02\) simulation. It is observed that transitions between \(BI\) and \(BII\) occur at several bases. For this reason the time averages for Fig. 2.6 are carried out over the entire time regime. These transitions show the flexibility of the molecules, which are not unduly constrained in the crystal lattice during the simulation.

Among the sugar conformational parameters, we examined the phase angle of pseudorotation of the sugar ring as function of the base. The results averaged over the 20–25 ns time interval are presented in Fig. 2.8. This figure corresponds to strand 1 of duplex 1. From the figure, it is clear that sequence-dependent features are well preserved. Depending on the base, the puckering observed in the simulations is \(C18\)-exo, \(C28\)-endo, or \(C38\)-exo. The three conformations correspond to a phase angle that varies between approximately \(108^\circ\) and \(216^\circ\). These are also the three types of puckering reported in the
crystal structure. In these figures, the minima are slightly higher in the simulations than in the crystal, but the maxima agree very well. We calculated the standard deviation of the time-averaged, base-dependent values of the phase with respect to the crystal values per strand. For strand 1 and strand 2 these standard deviations are 17.36 and 18.78 for ff02; 20.60 and 19.57 for ff98; 20.24 and 19.98 for ff99; and 19.38 and 22.26 for ff02EP, indicating that the polarizable force field ff02 reproduces the crystal data the best, followed by ff99 (or ff98) and then ff02EP.

Among the local base-pair parameters we examined the propeller twist and opening, shown in Figs. 2.9 and 2.10 (data averaged over the 20–25 ns time interval). The propeller twist is not too well reproduced, and there are large variations in the 25 ns time interval. However, sequence dependent features seem to correlate with those observed in the crystal structure. In the crystal structure, the maximum value of the propeller is reached in the “outer” A-T base pairs, and the minimum in the middle C-G and G-C pairs. Opening is much better reproduced by ff02 and ff99 (or ff98) than by ff02EP. As it has been observed before, positive opening corresponds to A-T base pairs while negative opening corresponds to C-G base pairs.

Among the local base-pair step parameters, we examined roll and local helical twist. The results averaged over the 20–25 ns time interval are presented in Figs. 2.11 and 2.12. Both parameters show marked sequence-dependent features that are well reproduced by the simulations. 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. 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 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 (Cheatham, 1996, Cheamtha, 1997, Cieplak, 1997). However, crystal simulations 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 (Bevan, 2000). We have examined the time dependence of individual base steps. We found that there are fast oscillations (of amplitude ~ 10° in the original data taken every 1 ps) in short period of times. For longer periods of time, the helical twist for each base step either stays constant for the entire time regime, or undergoes some change which is compensated by an opposite change in a different base step. These changes seem to occur as slow oscillations over long time intervals. As an example, Fig. 2.13 shows the helical twist as a function of time in the *ff*02 simulations for the three base steps that have the highest values of helical twist in the crystal structure: 2 CA/TG, 5 CG/CG, and 8 TG/CA. It is observed that the twist does not increase or decrease in time but seems to undergo time oscillations with very long periods. The long-time oscillations cancel each other when the three base steps are averaged for each duplex.

Finally, the minor groove width (El Hassan, 1998) is shown in Fig. 2.14. These data were obtained with 3DNA (Lu, 2000, Lu, 2003). To compare with other DNA analysis packages and to take into account the van der Waals radii of the phosphate groups, one needs to subtract 5.8 Å from the values presented. Figure 2.14 shows data averaged over the 20–25 ns time interval. In this particular interval, *ff*02 reproduces the crystal minor groove width.
better than \textit{ff99} or \textit{ff98}. However we looked at the entire time regime and found that there are considerable oscillations in the groove width for all force fields, and that perhaps longer time intervals are needed to get a good average of this quantity. In the 25 ns time interval, both \textit{ff02} and \textit{ff99} (or \textit{ff98}) seem to perform comparably, and better than \textit{ff02EP}.

The standard deviations of the time-averaged, base dependent values of \((\varepsilon - \zeta)\), opening, propeller twist, roll, and helical twist with respect to the crystal values are presented in Table 2.3. In all cases, as the previous figures and tables indicate, the polarizable force field \textit{ff02} performs better than, or at worst equal to \textit{ff99} (or \textit{ff98}), the force field described only in terms of charges at the nuclear positions. For the propeller twist, \textit{ff99} gives a slightly better value than \textit{ff02} in the 20–25 ns interval, but this quantity varies greatly in the 25 ns regime, and perhaps longer times should be needed for a more credible average. Both \textit{ff02} and \textit{ff99} perform considerably better than the \textit{ff02EP}, the polarizable force field with charges present also at extranuclear sites.

We also examined the behavior of the Mg\textsuperscript{2+} ions during the crystal simulations. The water molecules that constitute the hydration sphere around each Mg\textsuperscript{2+} ion were examined for the crystal structure at various times and no exchange of waters was observed within the primary coordination shell. Mg\textsuperscript{2+} ions maintain the hydration shell, with the molecules in the inner sphere being 2.0–2.1 Å from the Mg\textsuperscript{2+}. Analysis of the trajectories revealed that some of the Mg\textsuperscript{2+} ions (with their associated sphere of hydration) moved relative to the DNA molecules. The movement of the individual Mg\textsuperscript{2+} atoms also depends on the force field. As an example, Fig. 2.15 shows the position of a Mg\textsuperscript{2+} ion relative to nearby DNA phosphorus
atoms. For both ff98 and ff02, the ion oscillates around the initial position, although the amplitude of the oscillations is larger for ff98. For ff02EP the oscillations are combined with motions of larger magnitude and at the end of the 25 ns the ion is approximately 3 Å away from the initial position. In general, for all the force fields the Mg$^{2+}$ ions tend to drift away from the DNA. This is reflected in Fig. 2.16 where the RMS deviations of the seven Mg$^{2+}$ ions associated with each duplex are shown as a function of time. These RMS deviations are expressed with respect to the initial crystallographic positions; yet, the relatively large magnitude of the RMS deviations as compared to those of DNA in Figs. 2.1–2.3 indicates a relative displacement of the Mg$^{2+}$ ion relative to nearby DNA atoms. Only for duplex 1 in the ff02 simulation that displacement is very small.

Finally, we examined the simulation $B$ factors (defined by the atomic position fluctuations multiplied by $8\pi^2/3$). The data presented are averaged over the 20–25 ns time regime. The $B$ factors from the simulation tend to be slightly higher than those from experiment. Those from ff02 approach the experimental data the best while those from ff99 (or ff98) and ff02EP have higher values on average. In general, there is good agreement between the $B$ factors calculated from the crystal simulations and the experimental ones, indicating once more that crystal simulations can represent the dynamics of the duplex in the crystal environment. Examples of $B$ factors are given in Fig. 2.17. In these figures it is observed that the $B$ factors have not completely converged, in the sense that the $B$ factors of the two strands of the duplex are not symmetric as would be expected, a common situation in simulations since individual B factors are subject to large fluctuations.
IV. DISCUSSION AND CONCLUSIONS

The results presented in this work show that crystal simulations can in fact provide a means of performing MD simulations under conditions in which many structural features of nucleic acids are preserved without overly constraining the molecules within the crystal lattice. Since a direct comparison between the experimental crystal structure and the crystal simulation provides a way to evaluate the quality of the simulations rigorously, it is possible to compare the performance of different force fields.

Under constant-pressure MD, the density of the system experiences small oscillations around a constant value of 1.41 g/ml and the volume standard deviation (of the order of ≈120 Å3 for each force field) puts the average volume for each force field in close agreement with the experimental value of 25 980 Å3. Experimentally, crystal packing effects have been cited as a limitation of crystallography in analysis of DNA structure. However, although crystal packing may constrain the DNA conformation in the crystal, the DNA can still be expected to adopt some of the structures that are favored by its sequence (Dickerson, 1994). In this sense, base sequence dependence can be regarded as conferring an inherent bendability on that region of the helix, and different crystallographic environments allow to visualize directly the nature and some extent of the sequence-directed polymorphism. On the other hand, it is an open debate whether the analysis of DNA structure in solution, either by NMR or MD, does represent a realistic environment for the DNA inside the cell. The density in the nucleus is 1.3–1.4 g/ml, which means that DNA is tightly packed and constrained in a manner that may resemble more its crystalline than its solution state.
The crystal simulations in this work preserve sequence dependent features extremely well, better than those reported with older force fields (Bevan, 2000, Cheatham, 1997). With respect to the backbone torsion angles, the (anti)correlated motion involving $\alpha$ and $\gamma$ and shown in Figs. 2.4 and 2.5 has been observed in previous MD simulations (Bevan, 2000) and NMR studies (Xu, 1994). In general, an increase in one of the angles correlates with a decrease in the other such that the structural changes are balanced. The values reported for these angles in Table 2.2, however, do not agree well with those obtained experimentally, which is indicative of a deficiency in the force field (and not an artifact of crystal packing) since the same trend has been observed in solution simulations (Bevan, 2000).

The torsion angles ($\varepsilon - \zeta$) in Fig. 2.6 present strong sequence-dependent features that reproduce very well the experimental structure (except for $ff02EP$). The averaged values of ($\varepsilon - \zeta$) for C-2 and T-8 are indicative of the BII conformation while the others are characteristic of the more common BI conformation. However, a time representation of ($\varepsilon - \zeta$) as shown in Fig. 2.7 reveals that there are transitions between BI and BII at several bases. These transitions show the flexibility of the molecules, which are not unduly constrained in the crystal lattice during the simulation. However, a Boltzmann weighting for the BI/BII distribution would require a few hundred of nanoseconds, which at present is beyond the scope of this work. The time-averaged phase angle of pseudorotation of the sugar ring in Fig. 2.8 also shows sequence-dependent features that are well preserved. Depending on the base, the time-averaged puckering observed in the simulations is C18-exo, C28-endo, or C38-exo. However, sugar repuckering is observed in the simulations at different times.
Among the local base-pair parameters there is considerable variation in the propeller twist for the comparison in Fig. 2.9 to be totally valid. Opening, however, seems to be reproduced better, with positive opening corresponding to A-T base pairs and negative opening corresponding to C-G base pairs. Among the local base-pair step parameters, both roll and helical twist reproduce sequence-dependent features. Positive and negative values of roll alternate. The helical twist reaches its lowest values in the CC/GG and AA/TT steps; its highest values being in the CA/TG and CG/CG steps, and intermediate values in the AC/GT steps. Crystal simulations do not suffer from the “undertwisting” characteristic of the solution simulations, a result that is believed to derive from the constraints placed upon the DNA duplexes within the unit cell (Bevan, 2000). In addition to fast oscillations the twist also seems to experience slow oscillations over long time intervals that may correspond to collective motions. Finally, the minor groove width obtained in these crystal simulations tends to change with time, but on average ff02 and ff99 (or ff98) reproduce the experimental minor groove width fairly well (less so does ff02EP).

The polarizable ff02 and extra-point ff02EP force fields represent attempts to achieve a better representation of electrostatics. The crystal simulations indicate that the polarizable ff02 field in combination with a polarizable water model (POL3 in this case) generally reproduces sequence specific features better than the simpler ff99 and ff98 on the nanosecond time scale; whether this is to hold for longer time scales is an open question. The results for the polarizable ff02 are slightly surprising if one considers 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 the current implementation of induced dipoles in AMBER careful initial equilibration of the system is important. Once the system was equilibrated, the simulation ran smoothly for the 25 ns (some researchers have reported ‘crashes’ due to the ‘polarization catastrophe,’ which can occur in absence of a damping term, but this was not the case in these well-equilibrated systems with relatively homogeneous density). We note that the formalism that we have implemented in AMBER can be used for both permanent and induced dipoles, and at present we are extending the Car-Parrinello scheme to include a more realistic description of polarization. We did find on the other hand that there are problems with the \textit{ff02EP} representation in combination with an extra-point water model (in this case, TIP5P): the simulations tended to crash often in spite of careful equilibration. By restarting the runs, it is possible to gather data for the entire 25 ns regime. Yet, this representation seems less accurate than the other two. We believe that both the crashes and the loss in accuracy are due to the lack of van der Waals radius for the extranuclear charges. We also tried two 5 ns test runs for \textit{ff02EP} combined with POL3. The simulations are more stable but the results are at best equal to those obtained with \textit{ff02} (but then one has the additional cost of the extra points). Again, these conclusions could change for longer time scales.

Nucleic acids are highly charged polymers where polarization effects are bound to play a very important role. The results presented here are very encouraging for future parametrizations that include higher order multipoles and polarization. Indeed, for the first time these calculations are feasible from a computational point of view (Sagui, 2004). Higher order multipoles and polarization have already been successfully implemented for protein
force fields (Ren, 2002, Ren, 2003, Ponder, 2003), the results presented here would indicate
that this is the route for the accurate representation of DNA. This, in turn, will open the door
to more detailed investigation of the interaction of nucleic acids with solvent and ions,
particularly divalent ions, as well as allow better characterization of modified nucleic acids
and subtle drug–nucleic acid or protein–nucleic acid interactions.

ACKNOWLEDGMENT

This research was supported by NSF under Grant Nos.

ITR-0121361 and DGE-9987555.
FIG. 2.1 Atomic RMS deviations of the two duplexes in the crystal for the ff99 simulation. Data are shown every 5ps.
FIG. 2.2 Atomic RMS deviations of the two duplexes in the crystal for the ff02EP simulation. Data are shown for every 5 ps.
FIG. 2.3 Atomic RMS deviations of the two duplexes in the crystal for the ff02 simulation. Data are shown for every 5 ps.
FIG. 2.4 Torsional angles $\alpha$ (top curve) and $\alpha$ (bottom curve) as functions of time for the different force fields. ff99 shows similar results to ff98.
FIG. 2.5 Torsional angles $\alpha$ (solid line) and $\alpha$ (dashed line) as functions of sequence for one of the strands of duplex 2 in the ff02 simulation. The numbers 1, …, 10 in the horizontal axis correspond to CCAACGTTGG. The representation of $\alpha$ in this figure differs from that in Fig. 4 and in Table II by the addition of 360°.
FIG. 2.6 The difference in torsional angles ($\varepsilon - \zeta$) is shown as a function of the base for strand 1. The numbers 1, ..., 9 in the horizontal axis correspond to CCAACGTTG. Solid line – crystal, dashed line – simulations.
FIG. 2.7 The difference in torsional angles ($\varepsilon - \zeta$) is shown as a function of time for strand 1 of duplex 1 in the ff02 simulation. Data are presented for every 5 ps.
FIG. 2.8 Phase angle of pseudorotation of the sugar ring for strand 1 of duplex 1 as function of the base. The numbers 1,…,10 in the horizontal axis correspond to CCAACGTGG. Solid line – crystal, dashed line – simulation.
FIG. 2.13 Helical twist as a function of time in the f002 simulations. The left column contains data for duplex 1 and the right column data for duplex 2. From top to bottom the four rows contain data for the here base steps that have the highest values of helical twist in the crystal structure: 2 CA/TG, 5 CG/CG, 8 TG/CA base steps, and the average of these three steps. For clarity, data were smoothed by performing a 50-point running average in time, for 5 ps spaced data points. The very smooth lines are polynomial fittings to serve as a guide for the eye.
FIG. 2.14 Minor groove width as function of base-pair step. The solid black line represents the crystal; the dashed line and the solid gray line represent the two duplexes in the simulation.
FIG. 2.15 Distance between Mg$^{2+}$-45 and phosphorous DNA atoms as function of time. Distances are between Mg$^{2+}$-45 and T7P (Black line) and A14P (gray line). Data are shown for every 5ps.
FIG. 2.16 RMS deviations of the 7 MG$^{2+}$ ions associated with duplex 1 (black line) and the 7 MG$^{2+}$ ions associated with duplex 2 (gray line). Data are shown for every 5 ps.
FIG. 2.17 B factors calculated from the simulations compared with B factors from the crystal structure. In each panel, the black line represents the B factors from the crystal structure and the gray line those calculated from the simulation. Results for each strand of duplex one are shown for ff02 (top two panels) and for ff99 (bottom two panels).
Table 2.1 Atomic RMS deviations of the structures from the molecular dynamics simulations relative to the crystal structure. The definition of the force fields is given in the text. The RMS for parm94.dat has been calculated from the original data in Bevan et al (Bevan, 2000). The first RMS column contains averages of the data for the last 20 ns; the second RMS column contains averages for the last 10 ns of the simulations. The last RMS column contains the maximum RMS in the entire time regime.

<table>
<thead>
<tr>
<th>Force field</th>
<th>Structure</th>
<th>RMS (Å) 5–25 ns</th>
<th>RMS (Å) 15–25 ns</th>
<th>Maximum RMS (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>parm94.dat</td>
<td>Duplex 1</td>
<td>1.41</td>
<td>1.31</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>Duplex 2</td>
<td>1.23</td>
<td>1.27</td>
<td>1.71</td>
</tr>
<tr>
<td>ff99</td>
<td>Duplex 1</td>
<td>1.17</td>
<td>1.22</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>Duplex 2</td>
<td>1.03</td>
<td>1.06</td>
<td>1.52</td>
</tr>
<tr>
<td>ff02</td>
<td>Duplex 1</td>
<td>0.85</td>
<td>0.87</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>Duplex 2</td>
<td>0.84</td>
<td>0.80</td>
<td>1.35</td>
</tr>
<tr>
<td>ff02EP</td>
<td>Duplex 1</td>
<td>1.08</td>
<td>1.12</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>Duplex 2</td>
<td>1.02</td>
<td>1.04</td>
<td>1.42</td>
</tr>
</tbody>
</table>
Table 2.2 Helicoidal parameters for the crystal structure (5dnb) and for the duplexes from the MD simulations. The different parameters are given either in angstroms (Å) or degrees (deg).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5dnb</th>
<th>ff02EP Duplex 1</th>
<th>ff02EP Duplex 2</th>
<th>ff99 Duplex 1</th>
<th>ff99 Duplex 2</th>
<th>ff02 Duplex 1</th>
<th>ff02 Duplex 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ (deg)</td>
<td>296.6</td>
<td>312.78±2.98</td>
<td>311.00±6.16</td>
<td>258.75±6.40</td>
<td>265.48±3.91</td>
<td>286.67±8.08</td>
<td>289.07±5.00</td>
</tr>
<tr>
<td>$\beta$ (deg)</td>
<td>166.3</td>
<td>170.32±3.47</td>
<td>172.62±3.05</td>
<td>175.25±2.89</td>
<td>169.69±4.99</td>
<td>154.92±3.63</td>
<td>163.40±2.67</td>
</tr>
<tr>
<td>$\gamma$ (deg)</td>
<td>49.6</td>
<td>95.75±3.05</td>
<td>92.16±4.51</td>
<td>80.65±4.59</td>
<td>79.15±6.52</td>
<td>85.55±4.54</td>
<td>72.05±4.60</td>
</tr>
<tr>
<td>$\delta$ (deg)</td>
<td>128.7</td>
<td>135.69±2.79</td>
<td>139.63±2.50</td>
<td>140.08±2.78</td>
<td>138.85±2.87</td>
<td>138.07±2.19</td>
<td>135.40±2.34</td>
</tr>
<tr>
<td>$\epsilon$ (deg)</td>
<td>204.9</td>
<td>205.54±3.61</td>
<td>219.68±4.08</td>
<td>216.97±4.58</td>
<td>215.05±4.72</td>
<td>211.24±2.96</td>
<td>208.22±3.52</td>
</tr>
<tr>
<td>$\zeta$ (deg)</td>
<td>239.7</td>
<td>245.53±3.28</td>
<td>231.29±4.23</td>
<td>229.33±4.66</td>
<td>233.67±4.73</td>
<td>247.68±2.93</td>
<td>240.96±3.62</td>
</tr>
<tr>
<td>$\chi$ (deg)</td>
<td>257.7</td>
<td>252.62±3.28</td>
<td>261.89±2.67</td>
<td>260.32±2.85</td>
<td>259.46±3.00</td>
<td>259.08±2.99</td>
<td>260.39±2.80</td>
</tr>
<tr>
<td>Phase (deg)</td>
<td>145.5</td>
<td>154.71±4.68</td>
<td>161.29±4.67</td>
<td>162.22±4.17</td>
<td>159.67±4.23</td>
<td>157.73±3.71</td>
<td>154.51±3.81</td>
</tr>
<tr>
<td>Amplitude (deg)</td>
<td>38.8</td>
<td>37.32±1.21</td>
<td>38.32±1.26</td>
<td>38.62±1.19</td>
<td>38.13±1.15</td>
<td>38.10±1.18</td>
<td>38.33±1.20</td>
</tr>
<tr>
<td>$x$-disp (Å)</td>
<td>0.45</td>
<td>0.25±0.28</td>
<td>0.31±0.26</td>
<td>0.09±0.30</td>
<td>0.03±0.30</td>
<td>0.29±0.29</td>
<td>0.28±0.26</td>
</tr>
<tr>
<td>$y$-disp (Å)</td>
<td>0.0</td>
<td>-0.01±0.27</td>
<td>0.12±0.26</td>
<td>0.10±0.28</td>
<td>0.13±0.32</td>
<td>0.04±0.29</td>
<td>0.14±0.26</td>
</tr>
<tr>
<td>Incln. (deg)</td>
<td>5.10</td>
<td>4.20±2.36</td>
<td>6.55±1.98</td>
<td>5.99±2.06</td>
<td>5.12±2.01</td>
<td>4.52±2.34</td>
<td>6.63±1.95</td>
</tr>
<tr>
<td>Tip (deg)</td>
<td>0.0</td>
<td>0.63±1.92</td>
<td>0.03±2.22</td>
<td>0.54±2.00</td>
<td>-0.70±2.04</td>
<td>0.52±1.91</td>
<td>-0.31±2.25</td>
</tr>
<tr>
<td>Shear (Å)</td>
<td>0.0</td>
<td>0.01±0.10</td>
<td>-0.02±0.11</td>
<td>0.01±0.09</td>
<td>-0.03±0.09</td>
<td>-0.03±0.09</td>
<td>-0.04±0.09</td>
</tr>
<tr>
<td>Stretch (Å)</td>
<td>-0.13</td>
<td>-0.01±0.04</td>
<td>0.02±0.04</td>
<td>-0.06±0.04</td>
<td>-0.05±0.04</td>
<td>-0.06±0.04</td>
<td>-0.07±0.07</td>
</tr>
<tr>
<td>Stagger (Å)</td>
<td>0.08</td>
<td>0.31±0.15</td>
<td>0.17±0.14</td>
<td>0.17±0.15</td>
<td>0.11±0.14</td>
<td>0.24±0.15</td>
<td>0.21±0.14</td>
</tr>
<tr>
<td>Buckle (deg)</td>
<td>0.0</td>
<td>0.84±4.59</td>
<td>-1.84±4.29</td>
<td>-0.19±4.51</td>
<td>2.82±4.27</td>
<td>-2.35±4.00</td>
<td>0.40±4.71</td>
</tr>
<tr>
<td>Propeller (deg)</td>
<td>-10.24</td>
<td>-12.00±2.11</td>
<td>-13.58±2.55</td>
<td>-8.52±2.14</td>
<td>-10.64±2.32</td>
<td>-11.17±2.42</td>
<td>-10.42±2.24</td>
</tr>
<tr>
<td>Opening (deg)</td>
<td>0.75</td>
<td>-1.42±1.31</td>
<td>-1.13±1.34</td>
<td>0.21±1.18</td>
<td>0.04±1.28</td>
<td>1.31±1.25</td>
<td>0.42±1.27</td>
</tr>
<tr>
<td>Shift (Å)</td>
<td>0.0</td>
<td>-0.05±0.13</td>
<td>-0.08±0.17</td>
<td>-0.08±0.12</td>
<td>-0.08±0.13</td>
<td>-0.07±0.13</td>
<td>-0.06±0.16</td>
</tr>
<tr>
<td>Slide (Å)</td>
<td>0.78</td>
<td>0.55±0.20</td>
<td>0.79±0.18</td>
<td>0.49±0.13</td>
<td>0.44±0.14</td>
<td>0.59±0.21</td>
<td>0.77±0.18</td>
</tr>
<tr>
<td>Rise (deg)</td>
<td>3.31</td>
<td>3.30±0.08</td>
<td>3.27±0.08</td>
<td>3.29±0.08</td>
<td>3.31±0.07</td>
<td>3.29±0.08</td>
<td>3.26±0.09</td>
</tr>
<tr>
<td>Tilt (deg)</td>
<td>0.0</td>
<td>-0.33±1.01</td>
<td>0.02±1.15</td>
<td>-0.47±0.94</td>
<td>0.20±1.03</td>
<td>-0.29±1.01</td>
<td>0.18±1.16</td>
</tr>
<tr>
<td>Roll (deg)</td>
<td>2.23</td>
<td>1.78±1.44</td>
<td>2.85±1.08</td>
<td>2.39±1.11</td>
<td>2.46±1.09</td>
<td>2.06±1.47</td>
<td>2.96±1.08</td>
</tr>
<tr>
<td>Twist (deg)</td>
<td>35.27</td>
<td>34.74±0.41</td>
<td>34.65±0.40</td>
<td>34.42±0.38</td>
<td>34.44±0.38</td>
<td>34.73±0.40</td>
<td>34.68±0.40</td>
</tr>
</tbody>
</table>
Table 2.3 Standard deviation of the time-averaged, base dependant values of \((\varepsilon - \zeta)\), opening propeller twist, roll and helical twist with respect to the crystal values. The base-dependent values of \((\varepsilon - \zeta)\) are obtained as an average over the complete 25 ns time interval, while the base-dependent values of all the other parameters are obtained as an average over the 20 – 25 ns time interval.

<table>
<thead>
<tr>
<th></th>
<th>((\varepsilon - \zeta))</th>
<th>Opening</th>
<th>Propeller</th>
<th>Roll</th>
<th>Twist</th>
</tr>
</thead>
<tbody>
<tr>
<td>ff02</td>
<td>16.02</td>
<td>1.38</td>
<td>2.36</td>
<td>2.24</td>
<td>2.34</td>
</tr>
<tr>
<td>ff99</td>
<td>34.14</td>
<td>1.30</td>
<td>1.58</td>
<td>2.98</td>
<td>3.62</td>
</tr>
<tr>
<td>ff02EP</td>
<td>103.94</td>
<td>2.88</td>
<td>4.21</td>
<td>3.30</td>
<td>3.81</td>
</tr>
</tbody>
</table>
Abstract

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 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, independently of the base composition. However, when the first B-DNA crystal structure became available (Wing, 1980), 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 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 (York, 1993). Moreover, since a calculation of the Coulombic terms is expensive, the electrostatics typically represents the computational bottleneck of any long-time DNA simulation (Sagui, 1999). Large-scale simulations of DNA were enabled through the development of accurate and efficient algorithms for calculating electrostatics (Steinbach, 1994, Darden, 1993, Essmann, 1995), 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 (Beveridge, 2000, Cheatham, 2000, Norberg, 2002, Giudice, 2002, Orozco, 2003).

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 (Prive, 1991), and for which both crystal and solution simulations (Chuprina, 1991, Cheatham, 1996, Bevan, 2000, Baucom, 2004) exist. However, there is no NMR data for this decamer and therefore we compared the RSMD with respect to the crystal structure (as previous simulations have done). Although our main purpose is to assess the simulations in crystal environment, in order to determine 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 (Wu, 2003).

We elaborate on the aims annunciated above.

(i) Polarizable force fields: Recently, Baucom et al. (Baucom, 2004) compared the performance of different force fields in preserving the structure of the d(CCAACGTTGG)₂ decamer over 25ns unrestrained crystal simulations at constant pressure. The specific force fields investigated, based on the Cornell et al. force field, (Cornell, 1995) 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 (Toukmaji, 2000), as well as for higher order multipoles (Sagui, 2004, Sagui, 2004) based on the Particle Mesh Ewald (PME) method (Darden, 1993, Essmann, 1995) and multigrid methods (Sagui, 2001) 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 (Pearlman, 1995, Case, 2002, Case, 2004), together with a Car-Parrinello scheme for the computation of the induced polarization. This method 1.25-1.30 times more expensive than the PME method for charges alone, assuming a 1fs 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 environment. Crystal simulations are started from the ideal B-DNA structure, instead of the crystal structure, in order 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 (Cheatham, 2000). Whereas
solution simulations tend to fluctuate more than experiments, (Steinbach, 1994) the packing density in crystal simulations may inhibit important atomic fluctuations. Constant pressure simulations, therefore, are recommended for crystal environments in order 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 (Bevan, 2000, Baucom, 2004, York, 1995, Lee, 1995, Lee, 1995, MacKerell, 1995) show that structural features characteristic of B-DNA are maintained over several nanoseconds. While 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 in order 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 non-specific 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 (Yuan, 1992, Young, 1996, Varnai, 2004, Ponomarev, 2004). 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 (Case, 2002) and 8 (Case, 2004):

(i) ff99, which is the 1999 version of the Cornell et al. force field (Cornell, 1995), 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 (Baucom, 2004). These force fields differ from the previous parm94.dat in the torsion angle parameters for the glycosidic torsion in nucleic acids (Cheatham, 1999).

(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 in 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 (Ponder, 2003). Dynamics is performed using a recent implementation of a PME based approach to fixed and induced dipolar interactions (Toukmaji, 2000). 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 (Baucom, 2004) 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 (nonpolarizable, no extra points) (Jorgensen, 1983) for ff99 and POL3 (polarizable, no extra points) (Caldwell, 1995) 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 non-polarizable 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 8Å atom-based nonbond list, with a continuous correction for the long-range part. The long-range Coulomb energy was evaluated by the PME method (Darden, 1993, Essmann, 1995), with a heuristic pair list update (and a 1.0Å nonbonded pair list buffer), an 8Å cutoff, Ewald coefficient of 0.34864, 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 (Berendsen, 1984) at constant temperature (300 K) with a time constant of 0.5ps and constant pressure (which led to an average pressure of 1atm 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
1ps during these simulations. Calculation of DNA structural parameters was done using 3DNA (v 1.5) (Lu, 2003).

The starting coordinates for the simulations were standard Arnott B-DNA (Arnott, 1972), taken from the crystallographic PDB file with code 5DNB (Prive, 1991), or from the NMR file with code 1NAJ (Wu, 2003). The crystal structure is in space group C2 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 46.28Å, 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 11Å 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 (Wu, 2003) 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 12303 atoms (crystal axes equal to 54.24Å, 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 in order to adjust the positions of water molecules in too close a contact with other atoms. Subsequently, MD simulations with ff99 were conducted with restrained DNA at constant volume, slowly increasing the temperature from 0K to 300K in five steps, for a total of 50 ps. This configuration was saved and used for a 100ps constant volume simulation at 300K with DNA restrained under the ff02 force field. The last restrained configurations at 300K for both ff99 and ff02 were then used for unrestrained constant volume simulations for 100ps at T=300K. The last configurations from these runs became the time zero configurations for ff99 and ff02 constant pressure runs at 300K for 12ns. (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 48.27Å, 43.59Å and 52.52Å; 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, 500ps of constant volume MD at 300K 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 25ns long. Strictly speaking, these systems are not “solution" simulations, since 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 34.38Å, 32.25Åand 25.53Å; 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 (Prive, 1991). This same operation is applied to the structure generated by NUCGEN in order 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, 1.45Å. In the crystal environment, the Mg$^{2+}$ ions are very slow to diffuse (Baucom, 2004). 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, since these perform better than those currently implemented (Bevan, 2000). 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 pre-equilibrated 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 Baucom et al. (Baucom, 2004). 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 0K to 300K in five steps of 10ps each, for a total of 50ps. 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 500ps constant volume MD at 300K, with DNA and Mg\(^{2+}\) ions still restrained. The final coordinates of this last step were then used as initial coordinates for 500ps constant volume runs at 300K 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 12ns. 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 based on 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. In order to answer the question of how the solution simulations compare with respect to the NMR experiments, the DNA dodecamer d(CGCGAATTCCGC)$_2$ was chosen.

Figures 3.1-3.5 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 3.6 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 3.1 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 (Baucom, 2004) 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 2.1) and the “mixed" simulations mix-99 and mix-02 (Figure 3.2) 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, since 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 2.9Å, with an aRMSD of about 2.6Å. 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 3Å with respect to the crystal structure, even when the initial coordinates for both DNA duplexes and divalent ions are the same as the crystal ones.

(ii) The results involving the dodecamer d(CGCGAATTCGCG)$_2$ in the solution simulations sol-nmr-99 and sol-nmr-02 (Figure 3.6) 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% while the aRMSD improves approximately by 50%. The low value of the aRMSD (1.62Å) 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, crys-Mg-O-02 systems are comparable to the RMSD obtained with all waters placed randomly in the crys-Mg-99 , 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 3.1 would indicate that crys-Mg-O-99 has smaller RMSD than crys-Mg-99, Figure 3.4 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 3.5 are larger than those in Figures 3.3 and 3.4. This is just a reflection of the importance of the positions of the Mg$^{2+}$ ions in the generation of the DNA crystal lattice (Lapanov, 1993, Minasov, 1998, Chiu, 2000). The Mg$^{2+}$ ions diffuse very slowly in the crystal simulations; and therefore the results in Figure 3.5 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 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. Naturally, 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 1.45Å away from the crystal structure. After an initial equilibration period, the structures converged to the crystal structure with an aRMSD for each duplex of 1.06Å/0.86Å for crys-Mg-99 and 0.63Å/0.72Å for crys-Mg-02.

(viii) It is interesting to compare these results to those of Baucom et al, (Baucom, 2004). 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 (Cheatham, 1996, Bevan, 2000). These simulations started from the crystal structure, which was then relaxed and equilibrated for approximately 600ps before constant pressure runs 25ns 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 0-15ns time regime it was 0.602Å/0.671Å and during the 15-25ns time regime it was 0.698Å/0.578Å. Instead, the aRMSD increased for ff99 with the values of 0.715Å/0.876Å (0-15ns) and 0.879Å/1.044Å (15-25ns). Since 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 Baucom et al (Baucom, 2004) (15-25ns) and crys-Mg-99 are 0.769Å/0.791Å and the analogous comparison for the ff02 structures and crys-Mg-02 gives 0.725Å/0.600Å.
Much of the data characterizing the DNA decamer d(CCAACGTTGG)$_2$ structure from the unrestrained simulations is summarized in Table 3.2, 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, since most parameters are strongly dependent on the sequence, the average values for each parameter convey only reduced information$^1$. In addition, the standard deviations for the individual pairs/steps are -at the very worst- 10 times larger than the standard deviations obtained in Table 3.2 and 3.3. The duplexes in all 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

$^1$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 1ps were independent, the error of the mean value during the last 10ns of the simulations could be obtained by dividing the reported SD by $\sqrt{10000} = 100$. 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 3.5ns, the error estimate is approximately only 5 times worse than the one computed in 1ps windows. Even if one assumes that the real error is 10 times worse than that for the 1ps
windows, one would still have to divide the SD in Table 2.2 by approximately 10, which gives very small values for the error of measurement of the mean value.

The trajectories is very similar to that reported previously, either in solution (Bevan, 2000) or in crystal environment (Baucom, 2004). The $\alpha$ and $\gamma$ torsions are in the $g^\prime/g^+$ ($-60^\circ/+60^\circ$) canonical conformation. The glycosidic angle $\chi$ oscillates around $-100^\circ$ (or $260^\circ$) in the crystal and $-115^\circ$ (or $245^\circ$) 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 ($\varepsilon - \zeta$) display strong sequence-dependent features, with some bases assuming the BI backbone conformation ($-90^\circ$) and others the BII backbone conformation ($+90^\circ$) (Lavery, 1994). A time representation of ($\varepsilon - \zeta$) 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 3.7 to 3.9 show several sequence-dependent parameters. 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. All crystal simulations were started from the dotted, feature-less 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 function of the base for strand I. The results are presented
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.

In Figure 3.7. All figures show clear sequence-dependent features with the crystal runs showing -as expected- better correlation with the 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 3.8 and 3.9 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 3.8a, but it is remarkably reproduced by the crystal ff02 runs shown in Figures 3.8b,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 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

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°.

al. force field typically result in an undertwisting of the duplex, such that values for twist are lower than those observed in crystal structures (Cheatham, 1996, Cieplak, 1997). This is indeed observed in Figure 3.9a and Table 3.2. However, crystal simulations, as those shown in Figure 3.9b,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 (Bevan, 2000).

The simulation B factors (defined by the atomic position fluctuations multiplied by $8\pi^2/3$) are presented in Figure 3.10. The good agreement between the B factors calculated from the crystal simulations and the experimental ones indicates that crystal simulations can represent well the dynamics of the duplex 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 3.1, the agreement with experiment is not bad but it 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 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 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, since 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 since 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°, 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 (Beveridge, 1995); 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 3.3.

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 2.9Å with an aRMSD of about 2.6Å. 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 3.55Å, for approximately 1ns simulation (Cheatham, 1996) and aRMSD equal to 4.22Å for a 5ns simulation (Bevan, 2000)). 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 1.84Å for sol-nmr-99 and 1.62Å 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 (Baucom, 2004) 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 re-create 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 1.45Å for ABDNA to an average crystal structure with an aRMSD for each duplex of 0.63Å/0.72Å for crys-Mg-02. In addition, the aRMSD between the ff02 crystal duplexes
in Ref. 16 (15-25ns) and crys-Mg-02 are 0.725Å/0.600Å. 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 counter-examples. 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, 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 3Å 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; 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 (Chiu, 2000). Experimentally it has been found that divalent ions are key in stabilizing the crystal unit cells (Yuan, 1992). Even the type of divalent ion plays a big role in the resulting crystal. Thus, Mg\textsuperscript{2+} and Ca\textsuperscript{2+} ions generate different DNA crystal lattices and stabilize different end-to-end overlaps and lateral contacts between duplexes (Liapnov, 1993, Minasov, 1998, Chiu, 2000). Moreover, monovalent cations not revealed by the crystallographic structures do not crucially affect the conformation of duplex DNA, since they do not directly coordinate to DNA atoms (Minasov, 1998). Interestingly, the simulations are sensitive enough to echo these experimental results. Thus, while hydrated Mg\textsuperscript{2+} ions stabilize the crystal by bridging the two helical duplexes within the unit cell, the 18 Na\textsuperscript{+} ions have a higher mobility and are seen to sample more freely the available space. These considerations imply that a good parameterization of the Mg\textsuperscript{2+} ion is crucial for accurate simulations of DNA where these ions are present. Mg\textsuperscript{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 (Deerfield, 1995, Munoz, 2001). Indeed, a classical calculation that compared simulation to experimental results concluded that deficiencies in the results concerning the Mg\textsuperscript{2+} ions were in part due to the lack of explicit
polarization/charge transfer terms for all molecules, including water, in the force field (York, 1992).

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 (Ponder, 2003). The success of even this simple representation is understood from a physical point of view, since 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 (Ponder, 2003, Ren, 2003, Ren, 2003) to include nucleic acids, with a relatively small increase in computational cost.
Acknowledgments

This research was supported by NSF under grants ITR-0121361 and CAREER DMR-0348039.
Figure 3.1 RMSD of all DNA atoms in the $d($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 3.2 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 3.3 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 3.4 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 3.5 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 3.6 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.
Figure 3.7 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, cry-Mg-O-99, and crys-Mg-O-02.
Figure 3.8 Roll as a function of the 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, cry-Mg-O-99, and crys-Mg-O-02.
Figure 3.9 Helical twist as a function of the 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, cry-Mg-O-99, and crys-Mg-O-02.
Figure 3.10 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.
Table 3.1: 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 3.1-3.6 and the RMSD of the average structure (aRMSD) are listed\textsuperscript{a}

<table>
<thead>
<tr>
<th>simulation</th>
<th>time for average (ns)</th>
<th>iRMSD (Å)\textsuperscript{b}</th>
<th>aRMSD (Å)\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>sol-99</td>
<td>5—12</td>
<td>3.04</td>
<td>2.77</td>
</tr>
<tr>
<td>sol-02</td>
<td>5—12</td>
<td>2.69</td>
<td>2.49</td>
</tr>
<tr>
<td>mix-99</td>
<td>5—25</td>
<td>3.54/2.98</td>
<td>2.98/3.02</td>
</tr>
<tr>
<td>mix-02</td>
<td>5—25</td>
<td>2.82/3.33</td>
<td>2.48/3.26</td>
</tr>
<tr>
<td>crys-Mg-99</td>
<td>3—12</td>
<td>1.22/1.08</td>
<td>1.06/0.86</td>
</tr>
<tr>
<td>crys-Mg-02</td>
<td>3—12</td>
<td>0.88/0.89</td>
<td>0.63/0.72</td>
</tr>
<tr>
<td>crys-NH-O-99</td>
<td>3—12</td>
<td>1.09/1.09</td>
<td>0.80/0.79</td>
</tr>
<tr>
<td>crys-NH-O-02</td>
<td>3—12</td>
<td>0.99/0.86</td>
<td>0.81/0.64</td>
</tr>
<tr>
<td>crys-random-99</td>
<td>3—12</td>
<td>1.38/1.24</td>
<td>1.06/0.95</td>
</tr>
<tr>
<td>crys-random-02</td>
<td>3—12</td>
<td>1.20/1.21</td>
<td>1.01/1.01</td>
</tr>
<tr>
<td>ref 16, hp99</td>
<td>15—25</td>
<td>1.22/1.06</td>
<td>0.88/1.04</td>
</tr>
<tr>
<td>ref 16, hO2</td>
<td>15—25</td>
<td>0.87/0.80</td>
<td>0.70/0.58</td>
</tr>
<tr>
<td>sol-nmr-99</td>
<td>0—12</td>
<td>2.39</td>
<td>1.84</td>
</tr>
<tr>
<td>sol-nmr-02</td>
<td>0—12</td>
<td>2.15</td>
<td>1.62</td>
</tr>
</tbody>
</table>

\textsuperscript{a} The definition of the various runs and force fields are given in the text. \textsuperscript{b} When two RMSD values are present, they are the corresponding values for each duplex.
Table 3.2 Standard angle and helicoidal parameters, together with standard deviations, averages over all the residues, base pairs, or base-pair steps (where appropriate) for the d(CCAACGTGTTG)₂ decamer.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5DNB</th>
<th>ABDNA</th>
<th>sol-99</th>
<th>sol-02</th>
<th>crys-Mg-99</th>
<th>crys-Mg-02</th>
<th>crys-Mg-O-99</th>
<th>crys-Mg-O-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>α (deg)</td>
<td>290.5</td>
<td>313.2</td>
<td>290.5 ± 294</td>
<td>280.7 ± 3.56</td>
<td>285.2 ± 3.19</td>
<td>286.7 ± 2.83</td>
<td>290.4 ± 3.64</td>
<td>287.7 ± 1.05</td>
</tr>
<tr>
<td>β (deg)</td>
<td>166.4</td>
<td>214.0</td>
<td>176.8 ± 31.7</td>
<td>171.8 ± 3.22</td>
<td>170.0 ± 3.29</td>
<td>167.9 ± 2.43</td>
<td>172.0 ± 3.87</td>
<td>170.0 ± 2.97</td>
</tr>
<tr>
<td>γ (deg)</td>
<td>49.59</td>
<td>36.35</td>
<td>53.61 ± 3.25</td>
<td>60.61 ± 3.74</td>
<td>53.65 ± 4.45</td>
<td>53.71 ± 2.71</td>
<td>57.16 ± 5.06</td>
<td>54.35 ± 4.59</td>
</tr>
<tr>
<td>δ (deg)</td>
<td>128.8</td>
<td>156.4</td>
<td>124.7 ± 4.18</td>
<td>128.2 ± 3.60</td>
<td>138.6 ± 2.35</td>
<td>136.1 ± 2.58</td>
<td>137.1 ± 3.27</td>
<td>136.3 ± 2.35</td>
</tr>
<tr>
<td>ε (deg)</td>
<td>202.2</td>
<td>155.0</td>
<td>197.1 ± 5.58</td>
<td>196.5 ± 5.06</td>
<td>214.7 ± 4.04</td>
<td>211.2 ± 3.99</td>
<td>215.8 ± 7.92</td>
<td>210.8 ± 3.72</td>
</tr>
<tr>
<td>ζ (deg)</td>
<td>244.4</td>
<td>264.9</td>
<td>259.8 ± 5.79</td>
<td>264.4 ± 3.35</td>
<td>238.2 ± 5.18</td>
<td>243.2 ± 4.22</td>
<td>244.3 ± 7.13</td>
<td>242.5 ± 4.07</td>
</tr>
<tr>
<td>χ (deg)</td>
<td>257.8</td>
<td>262.1</td>
<td>243.0 ± 4.77</td>
<td>248.0 ± 4.62</td>
<td>260.5 ± 3.48</td>
<td>262.3 ± 2.92</td>
<td>257.8 ± 3.35</td>
<td>261.7 ± 3.44</td>
</tr>
<tr>
<td>Phase (deg)</td>
<td>145.4</td>
<td>191.9</td>
<td>135.0 ± 6.33</td>
<td>141.0 ± 5.58</td>
<td>158.2 ± 4.10</td>
<td>154.2 ± 4.26</td>
<td>156.1 ± 4.76</td>
<td>155.0 ± 3.98</td>
</tr>
<tr>
<td>Amplitude</td>
<td>38.82</td>
<td>35.65</td>
<td>38.85 ± 1.37</td>
<td>38.86 ± 1.36</td>
<td>38.39 ± 1.28</td>
<td>38.33 ± 1.21</td>
<td>38.41 ± 1.25</td>
<td>38.61 ± 1.17</td>
</tr>
<tr>
<td>α-disp (Å)</td>
<td>0.45</td>
<td>0.01</td>
<td>-2.05 ± 0.63</td>
<td>-1.62 ± 0.56</td>
<td>0.21 ± 0.27</td>
<td>0.22 ± 0.26</td>
<td>0.15 ± 0.27</td>
<td>0.17 ± 0.26</td>
</tr>
<tr>
<td>β-disp (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.03 ± 0.31</td>
<td>0.03 ± 0.32</td>
<td>0.07 ± 0.26</td>
<td>-0.09 ± 0.26</td>
<td>-0.03 ± 0.20</td>
<td>-0.04 ± 0.25</td>
</tr>
<tr>
<td>γ-disp (Å)</td>
<td>5.11</td>
<td>-5.94</td>
<td>9.17 ± 3.48</td>
<td>10.17 ± 3.61</td>
<td>3.92 ± 2.19</td>
<td>5.58 ± 1.93</td>
<td>4.53 ± 2.64</td>
<td>4.77 ± 2.06</td>
</tr>
<tr>
<td>δ-disp (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0 ± 0.00</td>
<td>0.0 ± 0.00</td>
<td>-1.78 ± 2.40</td>
<td>-0.19 ± 1.01</td>
<td>-0.19 ± 2.00</td>
<td>-0.35 ± 2.09</td>
</tr>
<tr>
<td>ε-disp (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0 ± 0.10</td>
<td>0.0 ± 0.14</td>
<td>-0.01 ± 0.09</td>
<td>0.04 ± 0.10</td>
<td>0.02 ± 0.09</td>
<td>0.01 ± 0.10</td>
</tr>
<tr>
<td>ζ-disp (Å)</td>
<td>-0.13</td>
<td>-0.13</td>
<td>-0.03 ± 0.04</td>
<td>-0.04 ± 0.07</td>
<td>-0.06 ± 0.04</td>
<td>-0.08 ± 0.04</td>
<td>-0.05 ± 0.04</td>
<td>-0.08 ± 0.04</td>
</tr>
<tr>
<td>stretch (Å)</td>
<td>0.08</td>
<td>0.01</td>
<td>-0.05 ± 0.16</td>
<td>-0.09 ± 0.17</td>
<td>0.23 ± 0.16</td>
<td>0.22 ± 0.15</td>
<td>0.18 ± 0.17</td>
<td>0.25 ± 0.16</td>
</tr>
<tr>
<td>hull (Å)</td>
<td>-0.0</td>
<td>0.0</td>
<td>0.04 ± 0.25</td>
<td>-3.17 ± 4.50</td>
<td>-1.48 ± 4.74</td>
<td>0.37 ± 4.29</td>
<td>0.86 ± 5.25</td>
<td>-0.75 ± 4.38</td>
</tr>
<tr>
<td>propeller (deg)</td>
<td>-10.25</td>
<td>4.20</td>
<td>-10.42 ± 3.15</td>
<td>-13.65 ± 3.18</td>
<td>-7.52 ± 2.34</td>
<td>-9.21 ± 2.41</td>
<td>-9.96 ± 2.79</td>
<td>-9.03 ± 2.33</td>
</tr>
<tr>
<td>opening (deg)</td>
<td>0.75</td>
<td>-5.55</td>
<td>0.08 ± 1.42</td>
<td>1.09 ± 2.03</td>
<td>0.16 ± 1.24</td>
<td>0.27 ± 1.22</td>
<td>-0.09 ± 1.27</td>
<td>0.22 ± 1.34</td>
</tr>
<tr>
<td>shift (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.02 ± 0.17</td>
<td>0.10 ± 0.18</td>
<td>-0.02 ± 0.12</td>
<td>0.06 ± 0.13</td>
<td>0.03 ± 0.14</td>
<td>0.0 ± 0.14</td>
</tr>
<tr>
<td>slide (Å)</td>
<td>0.78</td>
<td>0.34</td>
<td>-0.57 ± 0.27</td>
<td>-0.39 ± 0.25</td>
<td>0.52 ± 0.15</td>
<td>0.65 ± 0.14</td>
<td>0.52 ± 0.19</td>
<td>0.59 ± 0.14</td>
</tr>
<tr>
<td>rise (deg)</td>
<td>3.31</td>
<td>3.36</td>
<td>3.37 ± 0.07</td>
<td>3.39 ± 0.08</td>
<td>3.33 ± 0.04</td>
<td>3.34 ± 0.04</td>
<td>3.35 ± 0.04</td>
<td>3.34 ± 0.04</td>
</tr>
<tr>
<td>tilt (deg)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.03 ± 1.17</td>
<td>0.85 ± 1.22</td>
<td>0.07 ± 0.98</td>
<td>0.12 ± 1.01</td>
<td>0.42 ± 1.67</td>
<td>-0.03 ± 1.08</td>
</tr>
<tr>
<td>roll (deg)</td>
<td>2.23</td>
<td>-3.34</td>
<td>4.65 ± 1.87</td>
<td>5.51 ± 1.90</td>
<td>1.54 ± 1.20</td>
<td>2.28 ± 1.06</td>
<td>1.79 ± 1.48</td>
<td>1.79 ± 1.21</td>
</tr>
<tr>
<td>twist (deg)</td>
<td>35.24</td>
<td>35.61</td>
<td>31.72 ± 1.28</td>
<td>31.93 ± 1.30</td>
<td>34.99 ± 0.58</td>
<td>34.68 ± 0.44</td>
<td>34.75 ± 0.55</td>
<td>34.84 ± 0.40</td>
</tr>
</tbody>
</table>

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 give. b The different parameters are given either in angstroms (Å) or degrees (deg).
Table 3.3 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>INAJ</th>
<th>sol-nmr-99</th>
<th>sol-nmr-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ (deg)</td>
<td>208.7 ± 1.40</td>
<td>200.6 ± 3.36</td>
<td>200.1 ± 3.24</td>
</tr>
<tr>
<td>$\beta$ (deg)</td>
<td>171.3 ± 1.13</td>
<td>170.1 ± 2.90</td>
<td>170.9 ± 3.37</td>
</tr>
<tr>
<td>$\gamma$ (deg)</td>
<td>50.58 ± 1.01</td>
<td>54.51 ± 3.09</td>
<td>59.43 ± 4.11</td>
</tr>
<tr>
<td>$\delta$ (deg)</td>
<td>124.1 ± 0.65</td>
<td>129.7 ± 3.41</td>
<td>127.8 ± 3.26</td>
</tr>
<tr>
<td>$\epsilon$ (deg)</td>
<td>190.4 ± 0.25</td>
<td>207.8 ± 6.79</td>
<td>195.1 ± 4.87</td>
</tr>
<tr>
<td>$\zeta$ (deg)</td>
<td>257.5 ± 0.18</td>
<td>257.1 ± 5.62</td>
<td>264.7 ± 5.29</td>
</tr>
<tr>
<td>$\chi$ (deg)</td>
<td>247.8 ± 0.52</td>
<td>253.3 ± 5.84</td>
<td>248.6 ± 4.50</td>
</tr>
<tr>
<td>Phase (deg)</td>
<td>132.0 ± 0.66</td>
<td>142.3 ± 4.09</td>
<td>140.9 ± 4.95</td>
</tr>
<tr>
<td>Amplitude (deg)</td>
<td>32.65 ± 0.21</td>
<td>38.87 ± 1.22</td>
<td>36.58 ± 1.23</td>
</tr>
<tr>
<td>x-disp (Å)</td>
<td>-0.86 ± 0.90</td>
<td>-1.37 ± 3.89</td>
<td>0.0 ± 0.00</td>
</tr>
<tr>
<td>y-disp (Å)</td>
<td>0.0 ± 0.00</td>
<td>0.11 ± 0.56</td>
<td>0.0 ± 0.25</td>
</tr>
<tr>
<td>inclin. (deg)</td>
<td>6.11 ± 0.67</td>
<td>6.16 ± 5.46</td>
<td>8.51 ± 2.98</td>
</tr>
<tr>
<td>tip (deg)</td>
<td>-0.01 ± 0.94</td>
<td>-1.62 ± 4.32</td>
<td>1.47 ± 1.95</td>
</tr>
<tr>
<td>shear (Å)</td>
<td>0.0 ± 0.00</td>
<td>-0.12 ± 0.15</td>
<td>0.03 ± 0.10</td>
</tr>
<tr>
<td>stretch (Å)</td>
<td>-0.37 ± 0.40</td>
<td>0.02 ± 0.21</td>
<td>-0.07 ± 0.04</td>
</tr>
<tr>
<td>stagger (Å)</td>
<td>0.08 ± 0.00</td>
<td>0.0 ± 0.20</td>
<td>0.0 ± 0.14</td>
</tr>
<tr>
<td>buckle (deg)</td>
<td>0.01 ± 0.20</td>
<td>-1.31 ± 4.26</td>
<td>1.00 ± 4.11</td>
</tr>
<tr>
<td>propeller (deg)</td>
<td>-17.09 ± 0.64</td>
<td>-8.32 ± 3.37</td>
<td>-12.61 ± 2.69</td>
</tr>
<tr>
<td>opening (deg)</td>
<td>-0.35 ± 0.58</td>
<td>2.69 ± 3.00</td>
<td>0.13 ± 1.50</td>
</tr>
<tr>
<td>shift (Å)</td>
<td>0.0 ± 0.00</td>
<td>0.07 ± 0.19</td>
<td>-0.09 ± 0.14</td>
</tr>
<tr>
<td>slide (Å)</td>
<td>-0.18 ± 0.90</td>
<td>0.06 ± 0.37</td>
<td>-0.15 ± 0.24</td>
</tr>
<tr>
<td>rise (deg)</td>
<td>3.23 ± 0.03</td>
<td>3.21 ± 0.26</td>
<td>3.35 ± 0.07</td>
</tr>
<tr>
<td>tilt (deg)</td>
<td>0.01 ± 0.03</td>
<td>-1.17 ± 3.43</td>
<td>-0.87 ± 1.08</td>
</tr>
<tr>
<td>roll (deg)</td>
<td>3.57 ± 0.36</td>
<td>4.84 ± 3.31</td>
<td>4.75 ± 1.64</td>
</tr>
<tr>
<td>twist (deg)</td>
<td>35.45 ± 0.08</td>
<td>31.06 ± 2.98</td>
<td>32.48 ± 1.10</td>
</tr>
</tbody>
</table>

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 experimenta INAJ structure are also given. b The different parameters are given either in angstroms (Å) or degrees (deg).
ABSTRACT

We have investigated the role of the electrostatic description and cell environment in molecular dynamics simulations of DNA. Multiple unrestrained MD simulations of the DNA duplex d(CCAACGTTGG)$_2$ have been carried out using two different force fields: a traditional description based on atomic point charges and a polarizable force field. For the time scales probed, and given the “right” distribution of divalent ions, the latter performs better than the nonpolarizable force field. In particular, by imposing the experimental unit cell environment, an initial configuration with ideal B-DNA duplexes in the unit cell acquires sequence-dependent features that very closely resemble the crystallographic ones. Simultaneously, the all-atom root-mean-squared coordinates deviations (RMSD) with respect to the crystallographic structure is seen to decay. At later times, the polarizable force field is able to maintain this lower RMSD while the nonpolarizable force field starts to drift away.
I. INTRODUCTION

Early X-ray diffraction studies identified several distinct diffraction patterns corresponding to different forms of nucleic acid fibers, such as 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, which are constant for any composition of DNA, i.e., they are sequence-independent. The resolution of the first B-DNA crystal structure (Wing, 1980) showed that there is a clear relationship between DNA sequence and structure. This relationship has been explored by experimental work such as X-ray crystallography and Nuclear Magnetic Resonance (NMR), which probe DNA structure in their crystal and solution environments, respectively; and theoretical investigations mainly based on classical molecular dynamics (MD) simulations, which allow for an atomistic description of DNA structure and dynamics.

A key feature of any MD simulation of DNA is the treatment of the delicate long-range electrostatic interactions that are fundamental for the three-dimensional structure of this highly charged and polar molecule. Electrostatics not only represent a computational bottleneck of any long-time DNA simulation (Sagui, 1999), but they are also a bottleneck in accuracy, since they are the classical representation of the molecular electronic density and its associated quantum behavior. Today, long time stable simulations of DNA are commonplace thanks to accurate and efficient algorithms for calculating electrostatics (York, 1993, Steinbach, 1994, Darden, 1993, Essmann, 1995). However, the present force fields based on partial point charges for the description of electrostatics are inadequate to capture certain structural features. This, in turn, has given impetus to the development of polarizable
force fields, which potentially are considerably more accurate than the current descriptions because they can better describe the distortion of the electronic cloud in condensed phase.

In this paper, we present results of large-scale simulation studies 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 (Prive, 1991), and for which both crystal and solution simulations (Chuprina, 1991, Cheatham, 1996, Bevan, 2000, Baucom, 2004) exist. This work is essentially a continuation of previous work (Babin, 2006); here we have been able to significantly extend the length of the crystal simulations and to confirm the main conclusions reached before. 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, given the previous conditions, the polarizable force field performs better than the nonpolarizable one 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. These extended MD simulations show, however, that at late times the nonpolarizable force field starts to deviate from the crystallographic structure. These results give 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 (Case, 2002) and 8 (Case, 2004):

(i) The ff99 force field, which is the 1999 version of the Cornell et al. force field (Cornell, 1995), 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 (Baucom, 2004). These force fields differ from the previous parm94.dat in the torsion angle parameters for the glycosidic torsion in nucleic acids (Cheatham, 1999).

(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 non-polarizable 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\(^1\). Although

\(^1\) 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, since 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 (Ponder, 2003).
polarizable force fields tend to be considerably more costly to simulate, recent methodological advances for both permanent and induced dipoles (Toukmaji, 2000), as well as for higher order multipoles (Sagui, 2004, Sagui, 2004-B) based on the Particle Mesh Ewald (PME) method (Darden, 1993, Essmann, 1995) and multi-grid methods (Sagui, 2001) 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, 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 1fs time step for MD.

The water models used in the simulations to solvate DNA were TIP3P (Jorgensen, 1983) (non-polarizable, no extra points) for ff99 and POL3 (Caldwell, 1995) (polarizable, no extra points) for ff02. We also tried a crystal simulation with non-polarizable 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 parameters for Mg$^{2+}$ were taken from AMBER parm91.dat, since these perform better than those currently implemented (Bevan, 2000). For ff02 the ion has a small polarizability of 0.12.

The MD simulations were carried out using the SANDER module of AMBER 7 and 8. Van der Waals interactions were calculated using an 8Å atom-based nonbond list, with a continuous correction for the long-range part. The long-range Coulomb energy was evaluated by the PME method (Darden, 1993, Essmann, 1995), with a heuristic pair list update (and a
1.0Å nonbonded pair list buffer), an 8Å cutoff, Ewald coefficient of 0.34864, 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 (Berendsen, 1984) at constant temperature (300 K) with a time constant of 0.5 ps and constant pressure. 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) (Lu, 2003).

The starting coordinates for the simulations were either standard Arnott B-DNA (Arnott, 1972) or taken from the crystallographic PDB file with code 5DNB (Prive, 1991). The crystal structure is in space group C2 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, as detailed below.

(i) Solution simulations: 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 from a pre-equilibrated water box were added for a total of 9137 atoms in an orthorhombic cell (crystal axes of 48.27Å, 43.59Å and 52.52Å; mesh for Fourier transform in PME of sizes 48, 48 and 54). To achieve electroneutrality, eight water molecules were then replaced by Na$^+$ ions.
Those replaced waters were at least 11Å apart from DNA heavy atoms and newly introduced ions. 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 Mg$^{2+}$ ions was carried out in order to adjust the positions of water molecules in too close a contact with other atoms. Subsequently, MD simulations with ff99 were conducted with restrained DNA and Mg$^{2+}$ ions at constant volume, slowly increasing the temperature from 0K to 300K in five steps, for a total of 50 ps. This configuration was saved and used for a 100 ps constant volume simulation at 300K with DNA restrained under the ff02 force field. The last restrained configurations at 300K for both ff99 and ff02 were then used for unrestrained constant volume simulations for 500ps at T=300K. The last configurations from these runs became the time zero configurations for ff99 and ff02 constant pressure runs at 300K for 25ns. Strictly speaking, these systems are not “solution" simulations, since 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.

(ii) Crystal simulations: The crystal unit cell was given the geometry specified in the 5DNB PDB file (crystal axes equal to 34.38Å, 32.25Åand 25.53Å; 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 (Prive, 1991). This same operation is applied to the structure generated by NUCGEN in order 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, 1.45Å. In the crystal environment, the Mg$^{2+}$ ions are very slow to diffuse (Baucom, 2004). 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. To achieve electroneutrality, eight Na$^+$ ions were added. 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 Baucom et al (Baucom, 2004). First, all hydrogens were minimized to avoid overlaps and bad orientations. This was followed by minimization of bulk waters and Na$^+$ ions. Subsequently, MD simulations were conducted at constant volume, slowly increasing the temperature from 0K to 300K in five steps of 10ps each, for a total of 50ps. In these runs, all hydrogens, bulk water and Na$^+$ ions were allowed to move, but all the atoms in DNA (except hydrogens), Mg$^{2+}$ ions and crystallographic water oxygens were kept near their crystallographic positions through harmonic restraints. Then waters were allowed to equilibrate during an additional 500ps constant volume MD at 300K, with DNA and Mg$^{2+}$ ions still restrained. The final coordinates of this last step were then used as initial coordinates for 500ps constant volume runs at 300K 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 30ns. We refer to these simulations 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

Table 4.1 shows the average of the instantaneous RMSD (iRMSD) and the RMSD between the average structures (aRMSD) computed from the trajectories (adjusted for overall translations and rotations) of all DNA atoms (except hydrogens) relative to the 5DNB crystal structure. Results are shown for the “mixed" simulations mix-99 and mix-02 ; and the crystal simulations crys-Mg-O-99 , crys-Mg-O-02 , crys-random-99 and crys-random-02 . For comparison, values obtained from a crystal simulation (Baucom, 2004) that relaxes the original crystallographic structure have also been computed. The results of the “mixed" simulations mix-99 and mix-02 give substantial larger values than the crystal simulations. Given that the initial coordinates for both DNA duplexes and divalent ions are the same as the crystal ones, this is a reflection of role of the cell environment on the nucleotide structure.

Now consider the crystal structures, whose iRMSD are shown in Fig. 4.1. The most remarkable fact about the d(CCAACGTTGG)$_2$ crystal simulations with initial crystallographic 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 structures, the crystal simulations started from idealized ABDNA structures that were 1.45Å away from the crystal structure. After an initial equilibration period, the structures in the 3ns - 12 ns time regime have an aRMSD for each duplex of 0.80Å/0.79Å for crys-Mg-O-99 and 0.81Å/0.64Å for crys-Mg-O-02 , while the structures with randomly positioned Mg$^{2+}$ ions...
show higher RMSD (albeit, smaller than 1.45Å). However, as time evolves, crys-Mg-O-99 drifts away from the experimental structure and acquires RMSD that are comparable to those for crys-random-99, as evidenced by the 6ns - 30ns averages. The polarizable system crys-Mg-O-02, on the other hand, does a much better job of staying close to the experimental structure with an aRMSD for each duplex of 0.77Å/0.77Å. It is interesting to compare these results to those of Baucom et al (Baucom, 2004). 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 (Cheatham, 1996, Bevan, 2000). These simulations started from the crystal structure, that was then relaxed and equilibrated for approximately 600ps before constant pressure runs 25ns long. For ff02, the aRMSD was relatively the same for the entire time regime: during the 0-15ns time regime it was 0.60Å/0.67Å and during the 15-25ns time regime it was 0.70Å/0.58Å. Instead, the aRMSD increased for ff99 with the values of 0.72Å/0.88Å (0-15ns) and 0.88Å/1.04Å (15-25ns). This deterioration of the ff99 results is also observed in the crys-Mg-O-99 run. The aRMSD between the ff99 crystal structures of Baucom et al (Baucom, 2004) (15-25ns) and crys-Mg-O-02 (6-30ns) are 0.80Å/0.62Å, whose small value would indicate that the system has converged to a minimum of the force field that is close to the experimental structure.

Much of the data characterizing the DNA decamer d(CCAACGTTGG)₂ structure from the unrestrained simulations is summarized in Table 4.2, which gives standard angles and helicoidal parameters, averaged over all the residues, base pairs, or base pair steps (where appropriate) for the various systems. Since most parameters are strongly dependent
on the sequence, the average values for each parameter convey only reduced information. The duplexes in all systems analyzed correspond to a B-DNA type of structure.

Now we consider specific structural features. Figures 4.2 to 4.8 show several sequence-dependent parameters. 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. All crystal simulations were started from the dotted, feature-less line and evolved into the sequence-dependent solid line. These figures should be read in conjunction with Table 4.3, that shows the error of the averaged sequence-dependent function with respect to the experimental data.

Among the sugar conformational parameters, we examined the phase angle of pseudorotation of the sugar ring as function of the base for strand I. The results are presented in Fig. 4.2. All figures show clear sequence-dependent features (sugar repuckering was observed at different times). Best fits to experimental data are provided by crys-Mg-O-02 and

---

2 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.

3 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±.
crys-random-99, although this quantity fluctuates considerably with standard deviations around 50±. With respect to the local base-pair parameters we show those that do not oscillate too close to zero; mainly the three rotational degrees of freedom, buckle, propeller and opening, shown in Figs. 4.3 to 4.5. In all cases, the best fit is significantly provided by crys-Mg-O-02. We also show the rotational degrees of freedom for the local base-pair step parameters, tilt, roll and twist, shown in Figs.4.6 to 4.8 respectively. For the palindromic sequence CCAACGTTGG, buckle and tilt have inversion symmetry while the other base-pair and step rotational parameters are symmetric. 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. 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 intermediate values in the AC/GT steps ("RY/RY" type of step). Again, in all cases these features are best reproduced by the crys-Mg-O-02 system. Notice that twist is reproduced amazingly well, a result that is believed to derive from the constraints placed upon the DNA duplexes within the unit cell (Bevan, 2000).

IV. DISCUSSION AND CONCLUSIONS

We have investigated the role of different electrostatic descriptions and cell environments on DNA sequence-specific structure as obtained from MD simulations. 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.

Previous work (Baucom, 2004) 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 shows that by starting from the idealized
Arnott B-DNA structure, it is possible to re-create sequence-dependent features completely
absent in the ABDNA structure. The values of the RMSD for both force fields initially decay
considerably from an initial value of 1.45Å for ABDNA to an average crystal structure with
an aRMSD for each duplex of approximately 0.80Å for crys-Mg-O-99 and slightly less for
crys-Mg-O-02 in the 3ns - 12ns interval, while the structures with randomly positioned Mg\(^{2+}\)
ions show higher RMSD (albeit, smaller than 1.45Å). However, as time evolves, crys-Mg-O-
99 drifts away from the experimental structure and acquires RMSD that are comparable to
those for crys-random-99, as evidenced by the 6ns - 30ns averages. The polarizable system
crys-Mg-O-02, on the other hand, does a much better job of staying close to the experimental
structure with an aRMSD for each duplex of 0.77Å/0.77Å.

To identify what are the ingredients that contribute to this close convergence we used
counter-examples. In this way, we were able to identify three key ingredients:

(i) The cell environment: The crystal symmetry is carefully set up in the simulations,
so that the unit cell is exactly the same as used in experiments. As a counterexample, the
RMSD increases to approximately 3Å for simulations that do not have a crystal environment,
even though in this case the initial coordinates for the DNA and Mg\textsuperscript{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\textsuperscript{2+} ions are initially placed approximately in their crystallographic positions (although allowed to move during equilibration). For the polarizable force field, the reproduction of the experimental features seems to remain constant during the entire time of the simulations. For the nonpolarizable field ff99, the agreement is good during approximately the first 12ns, but later the system drifts away from experiment. If, on the other hand, the crystal environment is correctly set up in the simulation but the Mg\textsuperscript{2+} ions are placed at random, as in runs crys-random-99 and crys-random-02, the RMSD are always slightly larger and the sequence-dependent features agree less well with experiments. After 12ns, crys-Mg-99, crys-random-99 and crys-random-02 give all comparable accuracies. Unfortunately the Mg\textsuperscript{2+} ions move so slowly for both force fields that their final distribution is very close to the initial one, and these simulations cannot provide an answer to issues of ion convergence or distribution.

(iii) The electrostatic description: Given the right cell symmetry and Mg\textsuperscript{2+} ion distribution, the polarizable force field performs systematically better than the non-polarizable one.
Experimentally it has been found that divalent ions are key in stabilizing the crystal unit cells (Yuan, 1992, Lipanov, 1993, Minasov, 1998, Chiu, 2000). It has also been found that although monovalent cations are needed to stabilize DNA conformation, they are relatively non-specific 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 (Minasov, 1998). Interestingly, the simulations are sensitive enough to echo these experimental results. Thus, while hydrated Mg\(^{2+}\) ions stabilize the crystal by bridging the two helical duplexes within the unit cell, the 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 (York, 1992, Deerfield, 1995, Munoz, 2001).

This work shows that crystal simulations with the polarizable ff02 field 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. Slightly better performance of ff02 is also seen in solution simulations (Babin, 2006). There is indication that polarization effects need to be considered not only for DNA and water but also for divalent ions. Improvements in the structural description of DNA are measured in spite of the
very simple implementation of polarization in the AMBER ff02 force field; and the fact that it may suffer from an underestimation of polarization effects (Ponder, 2002). The success of even this simple representation is understood from a physical point of view, since nucleic acids are highly charged polymers with important polarization effects. Although the gains in accuracy of the polarizable ff02 field over the nonpolarizable ff99 field are somewhat small (but systematic) these results give hope that good predictive DNA simulations may be achieved with future polarizable force fields, such as extensions to the AMOEBA force field (Ponder, 2003, Ren, 2002, Ren, 2003) to include nucleic acids. The feasibility of these more accurate descriptions is guaranteed by accurate and highly efficient methods for computation of higher order multipole interactions and polarizability (Toukmaji, 2000, Sagui, 2004).

Acknowledgments

This research was supported by NSF under grants ITR-0121361 and CAREER DMR-0348039.
**FIG. 4.1** RMSD of all heavy DNA atoms in d(CCAACGTTGG)2 with respect to the crystal structure for the crys-Mg-O (left) and crys-random (right) crystal systems. Deviations for the two duplexes in the system are shown. The dashed line is a guide for the eye. Data shown every 50ps.
FIG. 4.2 Phase angle of pseudorotation of the sugar ring for strand I as a function of the base. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right - crys-random.
FIG. 4.3 Buckle as a function of base-pair. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
FIG. 4.4 Propeller as function of the base-pair. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
FIG. 4.5 Opening as function of the base-pair. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
FIG. 4.6 Tilt as function of the base-pair step. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
FIG. 4.7 Roll as function of the base-pair step. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
FIG. 4.8 Helical twist as function of the base-pair. Circles – simulation, squares – 5DNB, triangles – Arnott B-DNA. Average is over last 24ns and over both decamers. Left panel – crys-Mg-O, right – crys-random.
TABLE 4.1 All-atom (except for H) RMSD of the molecular dynamics simulations relative to the experimental structure 5DNB. The two RMSD values correspond to each duplex. The definition of the various runs and force fields is given in the text.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Time for average</th>
<th>iRMSD (Å)</th>
<th>aRMSD (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mix-99</td>
<td>5 ns – 25 ns</td>
<td>3.54/2.98</td>
<td>2.98/3.02</td>
</tr>
<tr>
<td>mix-02</td>
<td>5 ns – 25 ns</td>
<td>2.82/3.33</td>
<td>2.48/3.26</td>
</tr>
<tr>
<td>crys-Mg-O-99</td>
<td>3 ns – 12 ns</td>
<td>1.09/1.09</td>
<td>0.80/0.79</td>
</tr>
<tr>
<td>crys-Mg-O-99</td>
<td>6 ns – 30 ns</td>
<td>1.38/1.32</td>
<td>0.98/1.00</td>
</tr>
<tr>
<td>crys-Mg-O-02</td>
<td>3 ns – 12 ns</td>
<td>0.99/0.86</td>
<td>0.81/0.64</td>
</tr>
<tr>
<td>crys-Mg-O-02</td>
<td>6 ns – 30 ns</td>
<td>0.96/0.95</td>
<td>0.77/0.77</td>
</tr>
<tr>
<td>crys-random-99</td>
<td>3 ns – 12 ns</td>
<td>1.38/1.24</td>
<td>1.06/0.95</td>
</tr>
<tr>
<td>crys-random-99</td>
<td>6 ns – 30 ns</td>
<td>1.34/1.35</td>
<td>1.10/1.09</td>
</tr>
<tr>
<td>crys-random-02</td>
<td>3 ns – 12 ns</td>
<td>1.20/1.21</td>
<td>1.01/1.01</td>
</tr>
<tr>
<td>crys-random-02</td>
<td>6 ns – 30 ns</td>
<td>1.21/1.09</td>
<td>0.97/0.86</td>
</tr>
<tr>
<td>Ref. 11, ff99</td>
<td>15 ns – 25 ns</td>
<td>1.22/1.06</td>
<td>0.88/1.04</td>
</tr>
<tr>
<td>Ref. 11, ff02</td>
<td>15 ns – 25 ns</td>
<td>0.87/0.80</td>
<td>0.70/0.58</td>
</tr>
</tbody>
</table>
TABLE 4.2 Standard angle and helicoidal parameters, together with standard deviations, averaged over all the residues, base pairs, or base pair steps (where appropriate) for the specified duplex structures. 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. The different parameters are given either in Angstroms (Å) or degrees (d).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5DNB</th>
<th>ABDNA</th>
<th>crys-Mg-O-99</th>
<th>crys-Mg-O-02</th>
<th>crys-random-99</th>
<th>crys-random-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>α (d)</td>
<td>296.3</td>
<td>313.2</td>
<td>288.9±1.95</td>
<td>286.9±3.09</td>
<td>291.8±6.06</td>
<td>298.5±4.62</td>
</tr>
<tr>
<td>β (d)</td>
<td>166.4</td>
<td>214.0</td>
<td>170.3±4.40</td>
<td>169.5±2.94</td>
<td>170.4±5.50</td>
<td>169.2±3.30</td>
</tr>
<tr>
<td>γ (d)</td>
<td>49.59</td>
<td>36.35</td>
<td>61.27±7.26</td>
<td>54.92±3.89</td>
<td>58.84±7.14</td>
<td>61.77±3.59</td>
</tr>
<tr>
<td>δ (d)</td>
<td>128.8</td>
<td>156.4</td>
<td>136.5±2.96</td>
<td>137.8±2.41</td>
<td>132.6±3.55</td>
<td>130.4±4.20</td>
</tr>
<tr>
<td>ε (d)</td>
<td>203.2</td>
<td>155.0</td>
<td>217.3±7.73</td>
<td>210.8±3.42</td>
<td>209.9±8.22</td>
<td>198.0±6.80</td>
</tr>
<tr>
<td>ζ (d)</td>
<td>244.4</td>
<td>264.9</td>
<td>248.6±5.13</td>
<td>240.8±4.16</td>
<td>256.1±6.09</td>
<td>258.1±5.31</td>
</tr>
<tr>
<td>χ (d)</td>
<td>257.8</td>
<td>262.1</td>
<td>257.6±3.76</td>
<td>263.0±3.43</td>
<td>254.9±4.58</td>
<td>255.6±5.21</td>
</tr>
<tr>
<td>phase</td>
<td>145.4</td>
<td>191.8</td>
<td>153.8±1.59</td>
<td>156.8±4.14</td>
<td>156.6±5.41</td>
<td>146.9±6.52</td>
</tr>
<tr>
<td>amplitude</td>
<td>38.82</td>
<td>35.65</td>
<td>38.93±1.28</td>
<td>38.69±1.16</td>
<td>38.99±1.58</td>
<td>37.67±1.41</td>
</tr>
<tr>
<td>x-disp (Å)</td>
<td>0.45</td>
<td>-0.01</td>
<td>0.01±0.32</td>
<td>0.13±0.26</td>
<td>-0.08±0.34</td>
<td>-0.01±0.26</td>
</tr>
<tr>
<td>y-disp (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.18±0.34</td>
<td>-0.03±0.25</td>
<td>0.01±0.33</td>
<td>-0.03±0.27</td>
</tr>
<tr>
<td>inclin. (d)</td>
<td>5.11</td>
<td>-5.94</td>
<td>4.81±3.07</td>
<td>5.31±2.08</td>
<td>4.72±3.08</td>
<td>5.15±2.74</td>
</tr>
<tr>
<td>tip (d)</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.48±0.21</td>
<td>0.11±2.07</td>
<td>0.29±2.19</td>
<td>0.61±2.14</td>
</tr>
<tr>
<td>shear (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.02±0.09</td>
<td>0.0±0.10</td>
<td>0.01±0.09</td>
<td>-0.01±0.09</td>
</tr>
<tr>
<td>stretch (Å)</td>
<td>-0.13</td>
<td>-0.13</td>
<td>-0.04±0.04</td>
<td>-0.08±0.04</td>
<td>-0.05±0.04</td>
<td>-0.08±0.04</td>
</tr>
<tr>
<td>stagger (Å)</td>
<td>0.08</td>
<td>0.01</td>
<td>0.12±0.16</td>
<td>0.2±0.15</td>
<td>0.13±0.15</td>
<td>0.15±0.15</td>
</tr>
<tr>
<td>buckle (d)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.69±5.29</td>
<td>-0.35±4.45</td>
<td>-0.30±5.15</td>
<td>-0.11±4.48</td>
</tr>
<tr>
<td>propeller (d)</td>
<td>-10.25</td>
<td>4.20</td>
<td>-10.25±2.61</td>
<td>-8.80±2.22</td>
<td>-10.0±2.43</td>
<td>-12.0±2.45</td>
</tr>
<tr>
<td>opening (d)</td>
<td>0.75</td>
<td>-5.55</td>
<td>-0.01±1.35</td>
<td>0.18±1.24</td>
<td>-0.47±1.41</td>
<td>-0.47±1.36</td>
</tr>
<tr>
<td>shift (Å)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.11±0.14</td>
<td>0.0±0.12</td>
<td>-0.02±0.17</td>
<td>-0.02±0.17</td>
</tr>
<tr>
<td>slide (Å)</td>
<td>0.78</td>
<td>-0.34</td>
<td>0.44±0.21</td>
<td>0.59±0.15</td>
<td>0.37±0.22</td>
<td>0.40±0.19</td>
</tr>
<tr>
<td>rise (d)</td>
<td>3.31</td>
<td>3.36</td>
<td>3.36±0.05</td>
<td>3.34±0.04</td>
<td>3.34±0.04</td>
<td>3.33±0.06</td>
</tr>
<tr>
<td>tilt (d)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.45±1.08</td>
<td>0.0±1.03</td>
<td>-0.31±1.10</td>
<td>-0.22±1.12</td>
</tr>
<tr>
<td>roll (d)</td>
<td>2.22</td>
<td>-3.64</td>
<td>2.06±1.46</td>
<td>2.10±1.16</td>
<td>2.06±1.45</td>
<td>2.58±1.42</td>
</tr>
<tr>
<td>twist (d)</td>
<td>35.24</td>
<td>35.61</td>
<td>33.96±1.04</td>
<td>34.84±0.39</td>
<td>33.43±0.84</td>
<td>34.15±0.69</td>
</tr>
</tbody>
</table>
TABLE 4.3 Error for the *averaged* sequence-dependent feature relative to the corresponding 5DNB function. Phase, base-pair and base-pair step parameters are shown in degrees.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>crys-Mg-O-99</th>
<th>crys-Mg-O-02</th>
<th>crys-random-99</th>
<th>crys-random-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>phase</td>
<td>20.784</td>
<td>18.135</td>
<td>12.597</td>
<td>20.211</td>
</tr>
<tr>
<td>buckle</td>
<td>3.321</td>
<td>2.956</td>
<td>3.734</td>
<td>4.145</td>
</tr>
<tr>
<td>propeller</td>
<td>2.318</td>
<td>2.180</td>
<td>3.445</td>
<td>2.871</td>
</tr>
<tr>
<td>opening</td>
<td>1.905</td>
<td>1.227</td>
<td>2.112</td>
<td>2.313</td>
</tr>
<tr>
<td>tilt</td>
<td>2.072</td>
<td>0.834</td>
<td>1.699</td>
<td>1.088</td>
</tr>
<tr>
<td>roll</td>
<td>3.586</td>
<td>1.334</td>
<td>4.108</td>
<td>4.444</td>
</tr>
<tr>
<td>twist</td>
<td>3.924</td>
<td>1.952</td>
<td>3.517</td>
<td>4.296</td>
</tr>
</tbody>
</table>
5: Conclusion

In this work, we explored the use of molecular dynamics simulations of the DNA duplex d(CCAACGTTGG)$_2$ to study the relationship between DNA sequence and structure in a crystal environment. Three different force fields were used in the second chapter: a traditional description based on atomic point charges (ff99), a polarizable force field (ff02), and an “extra-point” force field (ff02EP). The initial coordinates were taken from the crystallographic file and then thoroughly relaxed. It was found that all the force fields maintain fairly well the sequence-dependent features of the experimental structure. The polarizable force field, however, provides the most accurate representation of the crystal structure and sequence-dependent effects observed in the experiment. The results point to the need of the inclusion of polarization for accurate descriptions of DNA.

Encouraged by these results, we decided to investigate to what extent molecular dynamics simulations can reproduce DNA sequence-specific features, given different electrostatic descriptions and different cell environments. With respect to the electrostatic descriptions we used the simple point charge ff99 and the polarizable ff02 force fields. 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. At later times, the polarizable force field is able
to maintain the low RMSD while the nonpolarizable force field starts to drift away.

In conclusion, modern force fields are mature enough to not only reproduce
crystallographic features but are capable of recovering sequence specific characteristics and
to converge to crystal from sequence independent DNA. Polarizable force fields without
extra points consistently outperformed both extra point and simple point charge force fields.
Proper treatment of electrostatic and cell environment is critical for accurate molecular
dynamics simulations.

Where will simulations go from here?

As force fields continue to develop, procedures such as demonstrated in this work
must be carried out. The accuracy of future force fields can be characterized based not only
on how well they reproduce experimental results but on how well they can recover sequence
specific features from computationally generated ideal structures.

A logical next step from this work is to expand the number of systems being
analyzed. Extending the work to include well established and characterized systems such as
the Dickerson dodecamer (CGCGAATTCGCG)\textsubscript{2}, poly(A) tract DNA (for a review, see
Cheatham-2, 2000) and other well characterized systems will provide ample sample space for
analysis of polarizable force field predictive power. Analyzing systems with unique dynamic features can add to our confidence. Simulations with different conformations of DNA, such as A- and Z-forms of DNA can demonstrate the general applicability of classical polarizable force fields. These, in turn, can be used for structural predictions of alternate conformations of DNA, such as triplex, quadraplex, cruciform and other conformations. Expanding to RNA simulations could open a library of various modified nucleic acids for the study of the characteristic secondary and tertiary structure of some RNA molecules.

In conclusion, this work provides a good first step towards predictive molecular dynamics of DNA, indicating the important role of electrostatics and incorporating polarizability. The correct placement of heavy ions and accurate periodic box can dramatically improve simulations of crystal DNA structures. The accuracy of solution simulations has improved over previous force fields. As new force fields are developed and electrostatic treatments get more accurate, the reproductive and predictive power of DNA simulations will continue to increase. As these techniques are applied to a wider variety of systems, including RNA, protein and nucleic acid/protein simulations, the power of the procedures and force fields detailed and currently being developed will improve the simulations of these systems.
Bibliography


Bevan, D, Li, L, Pedersen, L, Darden, T. Molecular dynamics simulations of the d(CCAACGTTGG)2 decamer: Influence of the crystal environment. Biophys. J. 78:668-682, 2000


Chiu, T, Dickerson, R. 1 Å crystal structure of B-DNA real sequence specific and groove specific bending of DNA by magnesium and calcium. J. Mol. Biol. 301:915-945, 2000


Chiu, T, Dickerson, R. 1Å crystal structures of B-DNA reveal sequence specific binding and groove-specific bending of DNA by Magnesium and Calcium. J. Mol. Biol. 301:915-945

128


Halle, B, Denisov, V P. Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. Biopolymers 48:210-233


Lipanov, A, Kopka, M, Kaczor-Grzeskowiak, M, Quintana, J, Dickerson, R. The structure of the B-DNA decamer CCACITTG in two different space groups: conformational flexibility of B-DNA. Biochemistry 32:1373-1389, 1993


Smith, W. Point multipoles in the Ewald summation. CCP5 Information Quarterly 4, 13, 1982


Swaminathan, S, Ravishanker, G, Beveridge, DL. Molecular Dynamics of B-DNA including water and counterions: a 140-ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field. J. Am. Chem. Soc. 113:5027-5040, 1991


Xu, X, Chiu, W, Au-Yeung, S. Chemical shift and structural relationship in nucleic acids: correlation of backbone torsion angles $\gamma$ and $\alpha$ with $^{13}$C chemical shifts. J. Am. Chem. Soc. 120:4230-4231, 1994


York, D M, Darden, T, Deerfield, D, Pedersen, L G. The interaction of NA(I), CA(II) and MG(II) metal-ions with duplex DNA – a theoretical modeling study. Intl. J. Quant. Chem. 19:145-166, 1992
