Toward the Accurate Modeling of DNA: The Importance of Long-Range Electrostatics

Darrin M. York,*,†,‡ Weitao Yang,† Hsing Lee,§
Tom Darden,§ and Lee G. Pedersen§,⊥

Department of Chemistry, Duke University
Durham, North Carolina 27706
North Carolina Supercomputing Center
Research Triangle Park, North Carolina 27709-12889
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina 27709
Department of Chemistry, University of North Carolina
Chapel Hill, North Carolina 27599-3290

Received December 20, 1994

The relationship between DNA structure and function is fundamental to the understanding of biological processes. Currently, the most reliable source of biomolecular structural information comes from X-ray crystallographic data. Recent advances in theoretical modeling techniques have allowed molecular simulations to approach the accuracy obtainable from X-ray crystallography for proteins. We report the results of a 2.2 ns simulation of the B-DNA dodecamer d[CGCGAATTCGCG]$_2$ in a crystal unit cell, and demonstrate that, with rigorous accommodation of long-range forces, molecular simulation may be extended to provide atomic level accuracy of polynucleotide structures.

Most of what is known about DNA structure has been derived from crystal structure analyses. A concern that has been expressed brings into question the degree to which the intrinsic structure of DNA is disrupted by the crystal lattice. It has been recently pointed out by Dickerson[1] that lattice forces in DNA crystals are relatively weak, and that in fact these forces can be exploited to obtain useful information about DNA deformability. This notion strikes at the heart of a fundamental problem in structural biology: to understand the nature of DNA structure in the presence of environmental forces *in vivo*.

A powerful theoretical technique for studying biomolecular structure is molecular simulation. The reliability of molecular simulation to the study of polynucleotide structure, however, has lagged behind that of proteins. This is largely because of the difficulty associated with accommodation of long-range electrostatic forces. In conventional macromolecular simulations, pairwise Coulomb interactions are evaluated using a finite cutoff to decrease computational effort. Ultimately, this truncation of the Coulomb potential leads to unrealistic behavior.[2] The adverse effects of the cutoff methods are amplified in highly ionic systems such as DNA. In an attempt to circumvent these problems, artificial constructs are frequently introduced such as reduced phosphate charges on the nucleotide backbone,[3] modified Coulomb potentials in the form of distance dependent dielectric functions[4,5] or switching functions,[3,6] and distance restraints to enforce Watson−Crick hydrogen bonding.[6,7]

Perhaps the most studied DNA sequence both experimentally and theoretically is the dodecamer sequence d[CGCGAATTCGCG]$_2$.[8,9] Recent attempts to simulate the dodecamer in solution using full charges on the phosphates and explicit counterions to balance the charge led to structures that were significantly unwound,[10] with disrupted base stacking and base pair hydrogen bonding.[7,10] The results of a 1 ns molecular dynamics (MD) simulation of the same sequence that employed reduced phosphate charges on the DNA in conjunction with a switching function in the range 7.5−11.5 Å resulted in structures that were approximately 4.5 Å from that of canonical B-DNA[3] (for a review of the MD literature on DNA, see ref 11). Although these simulations represent progress, they bring into question the overall reliability of molecular simulation using cutoff methods for studying DNA structure.

The development of methods for evaluating long-range Coulomb forces is an ongoing area of active research (for a recent review, see ref 12). Recently, we have introduced two accurate methods for the rapid evaluation of Coulomb forces for macromolecular crystals.[13,14] Comparison of molecular simulations of ionic protein crystals indicates that rigorous treatment of electrostatic interactions can provide accuracy comparable to that observed between different crystal forms of the same protein,[15] and hence represents a major advance in the accuracy obtainable by molecular simulation. Recent crystal simulations of high-resolution Z-DNA hexamers show comparable results (Lee, H.; Darden, T.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *102*, 3830−3834); however, the methodology has not previously been tested on a significantly long segment of DNA where interstrand phosphate−phosphate repulsions become disruptive if not rigorously accommodated.

We have performed a 2.2 ns simulation of the dodecamer sequence d[CGCGAATTCGCG]$_2$ in a crystal unit cell[8,9] using full charges and explicit water and counterions. The trajectory is observed to be stable over the last nanosecond of simulation (Figure 1). The simulation cell-average structure over the last nanosecond has an overall root-mean-square positional deviation (rmsPD) from the crystal structure of 1.16 Å for all heavy atoms, and 0.90 Å for all base heavy atoms. Base stacking and Watson−Crick hydrogen bonding are rigorously maintained. Comparison of the simulation structure with the crystal structure shows the average deviation of corresponding base normal vectors is 11 ± 5°, and that of Watson−Crick hydrogen bond heavy atom−heavy atom distances is 0.1 ± 0.12 Å. The root-mean-square positional fluctuations (rmsPF) calculated from the simulation correlate well with fluctuations estimated from the crystallographic thermal factors (B values) using the relation $\langle \Delta r_i^2 \rangle^{1/2} = (3B_i/8\pi^2)^{1/2}$. In particular, the relative magnitudes of base, sugar, and phosphate heavy atom fluctuations in the simulation (0.77, 0.86, and 1.0 Å, respectively) are consistent with the corresponding experimentally derived values (1.01,

* To whom correspondence should be addressed.
† Duke University.
‡ North Carolina Supercomputing Center.
§ National Institute of Environmental Health Sciences.
⊥ University of North Carolina.

(1) Dickerson, R. E.; Goodsell, D. S.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3579−3583.
(2) York, D.; Darden, T.; Pedersen, L. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(3) McConnell, K. J.; Nirmala, R.; Young, M. A.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 4461−4462.
(4) Srinivasan, J.; Withka, J. W.; Beveridge, D. L. *Biophys. J.* **1990**, *58*, 533−548.
(5) Rao, S. N.; Kollman, P. *Biopolymers* **1990**, *29*, 517−532.
(6) Swaminathan, S.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5027−5040.
(7) Withka, J. M.; Swaminathan, S.; Srinivasan, J.; Beveridge, D. L.; Bolton, P. H. *Science* **1992**, *255*, 597−599.
(8) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature (London)* **1980**, *287*, 755−758.
(9) Drew, H.; Wing, R.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179−2183.
(10) Miaskiewicz, K.; Osman, R.; Weinstein, H. *J. Am. Chem. Soc.* **1993**, *115*, 1526−1537.
(11) Beveridge, D. L.; Ravishanker, G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 246−255.
(12) Greengard, L. *Science* **1994**, *265*, 909−914.
(13) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(14) York, D.; Yang, W. *J. Chem. Phys.* **1994**, *101*, 3298−3300.
(15) York, D.; Wlodawer, A.; Pedersen, L.; Darden, T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8715−8718.
(16) Weiner, S.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 230−252.
(17) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(18) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(19) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.

0002-7863/95/1517-5001$09.00/0

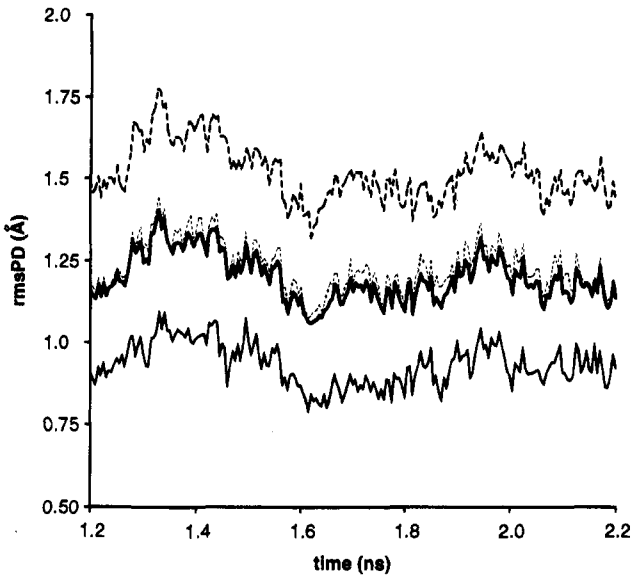

Figure 1. Time evolution of the rmsPD of the *instantaneous* cell-average structure from the crystal structure[8,9] for all heavy atoms (thick solid line), and separately for heavy atoms of the phosphates (medium dotted line at the top), sugars (thin dotted line in the middle), and bases (medium solid line at the bottom). The instantaneous cell-average structure was constructed at each time step by reverse symmetry transforming each of the four duplexes in the unit cell to a common local origin and averaging. Simulation was performed using a modified version of AMBER 3.0, implemented to perform Ewald sums using the particle-mesh Ewald method,[13] and based on the standard all-atom force field.[16] The unit cell (space group $P2_12_12_1$, cell parameters a = 24.87 Å, b = 40.39 Å, and c = 66.20 Å) contained 4 DNA duplexes, 1606 TIP3P[17] water molecules, and 88 sodium counterions (corresponding to the experimental crystal density of 1.5 g cm^{-1}). The starting coordinates for heavy atoms were those of the 1.9 Å resolution crystal structure.[8,9] Preparation and equilibration of the unit cell were analogous to those of previous crystal simulations.[15] Covalent bonds involving hydrogen were constrained using a modified version of the SHAKE algorithm.[18] Simulations were performed in the NVT ensemble (298 K) with a 1 fs integration time step, and carried out to 2.2 ns.

1.25, and 1.38 Å, respectively). Hence, the trend in thermal mobility observed by Drew et al.,[9] phosphates > sugars > bases, is also observed in the simulation.

A comparison of the average helical and torsion angle parameters is given in Table 1. Overall, the agreement between the simulation and crystallographic structures is quite good. The average helical rotation and displacement of the simulation structure (35.2° and 3.21 Å, respectively) are close to those of the crystallographic structure (35.8° and 3.37 Å, respectively). The characteristic roll–bend and narrowing of the minor groove in the AATT tract observed in the crystal structure are also preserved in the simulation structure (Figure 2). Large fluctuations in the α, γ, ϵ, and ζ torsion angles present in the simulation reflect correlated (α, γ) "crankshaft" motions of the DNA backbone,[20] and transient sampling between B_I and B_{II} conformers involving concerted (ϵ, ζ) torsion angle rotations.[21]

These results demonstrate that, with rigorous modeling of long-range ionic forces, theoretical techniques can begin to approach the accuracy required to provide detailed information about polynucleotide structure at the atomic level. The impact of these findings on the development of future generation force fields is evident, and provides optimism that forthcoming experimental and theoretical methods will have a synergistic effect in the unraveling of macromolecular structure–function relationships.

(20) Olson, W. In *Topics in Nucleic Acid Structure*; Neidle, S., Ed.; MacMillan, Ltd.; London, 1982; Part 2, pp 1–76.
(21) Dickerson, R. E. In *Biological Macromolecules and Assemblies*; Jurnak, F. A., McPherson, A., Eds.; Wiley-Interscience: New York, 1985; Vol. 2.

Table 1. Average Torsion Angle and Helical Parameters for the Crystallographic (X-ray) and Simulated Average (MD Av) Structures[a]

Torsion Angle Parameters

	α	β	γ	δ	ϵ	ζ	χ
X-ray	297(8)	171(14)	60(26)	123(21)	191(25)	252(35)	243(14)
MD av	278(26)	163(23)	87(39)	115(13)	211(26)	246(42)	227(9)

Global Base-Base Parameters

	shear	stretch	stagger	buckle	propeller	opening
X-ray	0.0(0.3)	−0.1(0.1)	0.0(0.2)	0.3(5.9)	−13.5(6.8)	0.7(4.3)
MD av	0.0(0.1)	−0.1(0.1)	0.3(0.2)	−2.4(5.9)	−11.9(13.0)	−3.4(3.4)

Global Interbase Pair Parameters

	shift	slide	rise	tilt	roll	twist
X-ray	0.0(0.5)	0.0(0.4)	3.4(0.2)	−0.1(2.8)	0.4(5.5)	35.8(3.8)
MD av	0.0(0.3)	0.0(0.2)	3.2(0.2)	0.1(1.7)	1.2(7.2)	35.2(3.7)

[a] Helical parameters were determined using the CURVES algorithm generously provided by Richard Lavery.[19] Torsion angles and angular helical parameters are in degrees, and distance helical parameters are in angstroms. Standard deviations are shown in parentheses.

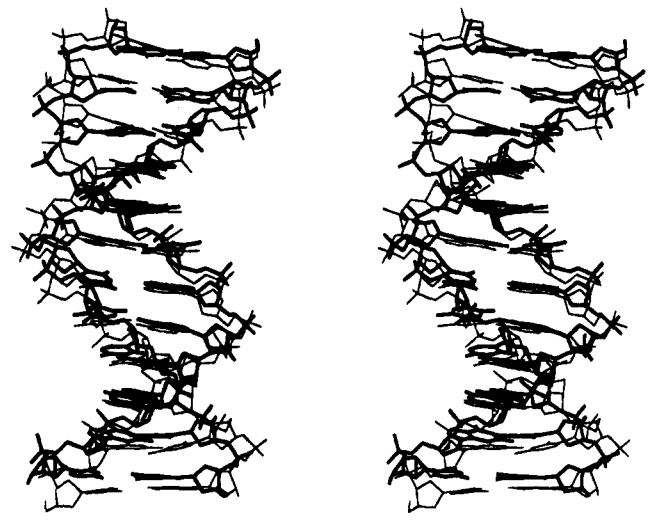

Figure 2. Stereo superposition of the heavy atoms of the crystallographic structure (thin lines) and the simulated (1.2–2.2 ns) cell-average structure (thick lines). The overall rmsPD for the fit was 1.16 Å for all heavy atoms, 1.33 Å for sugar–phosphate backbone heavy atoms, and 0.90 Å for purine/pyrimidine base heavy atoms.

Acknowledgment. We thank the North Carolina Supercomputing Center (NCSC), RTP, NC, the Biomedical Supercomputing Center, National Cancer Institute, FCRDC, and the Pittsburgh Supercomputing Center, Pittsburgh, PA, for allocations of computer time. DY was supported by an NSF postdoctoral fellowship jointly funded by the NCSC. WY is an Alfred P. Sloan Research Fellow. This work was supported by grants from NIEHS, RTP, NC (HL,TD,LP), NIH HL-27995 (LP), and NSF, NCSC, and MCNC through a subcontract agreement as part of the NIH Research Resource Program (DY,WY). DY thanks Professor Richard Lavery for providing the CURVES program.

JA944097Z