Effect of Periodic Box Size on Aqueous Molecular Dynamics Simulation of a DNA Dodecamer with Particle-Mesh Ewald Method

Osmar Norberto de Souza and Rick L. Ornstein
Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352 USA

ABSTRACT The particle-mesh Ewald (PME) method is considered to be both efficient and accurate for the evaluation of long-range electrostatic interactions in large macromolecular systems being studied by molecular dynamics simulations. This method assumes "infinite" periodic boundary conditions resembling the symmetry of a crystal environment. Can such a "solid-state" method accurately portray a macromolecular solute such as DNA in solution? To address this issue, we have performed three 1500-ps PME molecular dynamics (MD) simulations, each with a different box size, on the d(CGCGA$_6$CG)·(CGT$_6$CGCG) DNA dodecamer. The smallest box had the DNA solvated by a layer of water molecules of at least 5 Å along each orthogonal direction. The intermediate size box and the largest box had the DNA solvated by a layer of water molecules of at least 10 Å and 15 Å, respectively, along each orthogonal direction. The intermediate size box in the present study is similar to the box size currently chosen by most workers in the field. Based on a comparison of RMSDs and curvature for this single DNA dodecamer sequence, the larger two box sizes do not appear to afford any extra benefit over the smallest box. The implications of this finding are briefly discussed.

INTRODUCTION

Early molecular dynamics (MD) simulations of nucleic acids that used a cutoff distance to truncate the electrostatic interactions displayed extensive distortions in the DNA structure (Miaskiewicz et al., 1993), which, in some cases, were controlled by the introduction of artificial restraints (Swaminathan et al., 1991). However, substantial progress has been made recently in the quality of MD simulations for nucleic acids due to an efficient implementation of the particle-mesh Ewald (PME) method for biomolecular simulations (Essmann et al., 1995; Darden et al., 1993). Numerous well-converged nucleic acid simulations using the PME method have been reported (York et al., 1995; Cheatham et al., 1995; Zichi, 1995; Weerasinghe et al., 1995; Lee et al., 1995). To date, the structures of the simulated nucleic acids remained very close to their initial crystallographic or canonical conformations, clearly indicating that the previously noted long-range imbalance in the forces is avoided with the PME method. Does the PME method, however, artificially stabilize the initial structure by imposing "infinite" periodicity on the macromolecular system? Although a periodic system seems less disruptive than a vacuum boundary, its effect on the inherent dynamics of the macromolecular system is expected to be inversely related to the periodic box size. If the box is too small, one would expect a crystal-like behavior, which could significantly perturb the solution properties of the system. An unnecessarily large box may result in very little extra benefit, while being computationally impractical for routine use. In the present study, we compare MD simulations for a DNA dodecamer in three different box sizes to address this issue.

METHOD

MD simulations were performed with AMBER 4.1 (Pearlman et al., 1995), using the Cornell et al. (1995) force field. The starting conformation was taken from the x-ray crystallographic structure for the d(CGCGAAAAAACG)·d(CGTTTTTTCGCG) dodecamer and 42 crystallographic water molecules (DiGabriele and Steitz, 1993). The all-atom model of the DNA dodecamer contains 760 atoms. Each simulation system containing the DNA dodecamer and 22 sodium counterions, to neutralize the phosphate charge, was immersed in a rectangular box of TIP3P water molecules.

In the smallest simulation cell (B1), the DNA was solvated by a layer of water molecules, at least 5 Å wide in each orthogonal direction, resulting in a box of dimensions 56.0 × 41.0 × 41.0 Å^3. The intermediate size system (B2) is 70.0 × 45.0 × 45.0 Å^3 and has the DNA solvated by at least a 10-Å layer of water molecules. The largest simulation cell (B3) is 72.0 × 57.0 × 57.0 Å^3 and has the DNA solvated by a layer of water molecules of at least 15 Å in each orthogonal direction. The three simulation cells contained 8,162, 12,935, and 21,428 atoms.

Energy minimization and molecular dynamics (equilibration and production phase) were carried out under periodic boundary conditions. Equilibration was performed as described in reference 17 of Cheatham et al. (1995). Each simulation was computed in the NPT ensemble at 300 K with Berendsen temperature coupling and constant pressure (1 atm) with isotropic scaling. SHAKE was applied to fix all covalent bonds, allowing the use of a time step of 2.0 fs in the integration of the equations of motion. No extra restraints were placed on the DNA after the equilibration phase. The electrostatic forces and energies were computed with the PME method (Essmann et al., 1995; Darden et al., 1993), and the Lennard-Jones interactions were evaluated with a 9.0-Å residue-based cutoff. The PME charge grid spacing was ~1.0 Å, and the charge grid was interpolated on a cubic grid with the direct sum tolerance set to 4.0 × 10^{-6}. The nonbond pair list was updated every 10 steps. Data were collected for 1500 ps for each of the three systems. Structures for analysis were saved every 0.5 ps.

Received for publication 12 December 1996 and in final form 4 March 1997.

Address reprint requests to Dr. Rick L. Ornstein, Environmental Molecular Sciences Laboratory, Mailstop k1-83, Pacific Northwest National Laboratory, Richland, WA 99352. Tel.: 509-375-2132; Fax: 509-375-6904; E-mail: rl_ornstein@pnl.gov.

0006-3495/97/06/2395/03 $2.00

RESULTS AND DISCUSSION

Fig. 1 shows the time course of the instantaneous root mean square deviation (RMSD) with respect to the canonical A form (*top*) and the initial B-form crystallographic (*bottom*) structure. The RMSD parameter measures the overall change in conformation from the initial or any other reference structure. The MD trajectories appear to be stable in the second half of the simulations, whereas the RMSD converges to within 1.22 Å, 1.13 Å, and 1.35 Å of their average structure in the 750- to 1500-ps interval for B1, B2, and B3, respectively (Table 1). The relative large RMSD with respect to the A-DNA form (Fig. 1 and Table 1) indicates that structures in the trajectories are significantly different (~4.5 Å) from a canonical A-DNA. The observed RMSD (~3.3 Å) with respect to the B-DNA crystal structure is due mainly to curvature of the double-helix axis (see below). These values are similar to those of a recent solution MD investigation of the A-DNA to B-DNA transition (Cheatham and Kollman, 1996), but greater than the 1.16 Å RMSD observed for the Drew-Dickerson dodecamer (York et al., 1995). In the latter study, however, the authors attemped to reproduce the crystal environment by using constant volume and crystal symmetry in their simulations.

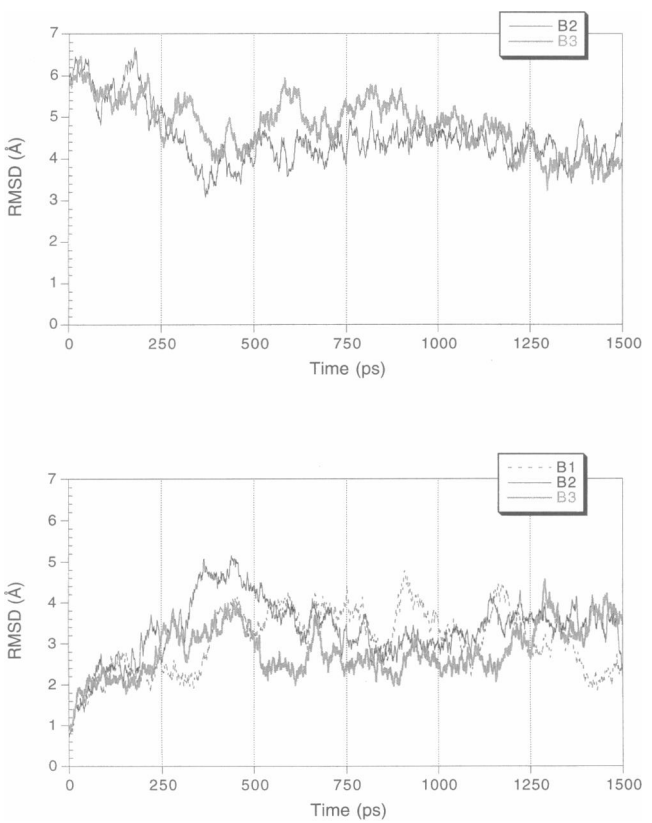

FIGURE 1 (*Top*) Time course of the root mean square deviation (RMSD) for MD simulations B2 (*dark line*) and B3 (*gray line*) with respect to the canonical A-DNA conformation. The RMSD of B1 is very similar to that of B2 and B3 and, therefore, is omitted for clarity. (*Bottom*) Time course of the RMSD with respect to the initial, crystallographic conformation for MD simulations B1, B2, and B3.

Thus it is not surprising that we and others see larger RMSD in constant-pressure simulations, in an attempt to replicate the solution behavior of DNA oligonucleotides.

Global helix axis curvature is another marker for overall change in DNA conformation. The computed curvatures in our solution simulations of the dodecamer in all three boxes are nearly equivalent and significantly larger than those seen in the crystal structure (Table 1, Fig. 2).

The uniform cross RMSD values, between the average structures of the three different trajectories, and the similarity in helix axis curvature (Table 1) indicate that the DNA is not sensitive to the size of the box in the three simulations of the present work. We anticipated, however, that simulation B1 would result in significant motional damping due to the imposed periodicity of the PME method. Surprisingly, this was not observed after the first 350 ps, whereas the RMSD increased and oscillated in a fashion similar to that observed in the other two simulations.

The above results indicate that for MD simulations of a DNA dodecamer, using periodic boundary conditions and electrostatic interactions evaluated with the PME method, a box with a minimum layer of water of at least 5 Å around the DNA seems adequate to reproduce the global structure and dynamics in solution. A box with a minimum layer of water of 10 Å or 15 Å, although more costly, does not appear to afford significant extra benefit. The present results are consistent with the recent findings of Smith and Pettit (1996), that there is only a negligible affect, due to "infinite periodicity," when the PME method is used, on the rotational potential energy surface of a dipole, quadrupole, or small protein in a high dielectric solvent.

Whereas the difference in computational cost between B1 and B2 is small, B3 requires twice as much CPU time as B2. Thus, if CPU time and disk space are severely limited, the use of a 5-Å minimum layer of water for DNA simulations may be adequate. However, we suggest that additional simulations on a variety of DNA sequences and compositions are appropriate to test whether the current finding is generally applicable. In this regard, it is possible that some DNA sequences exist that have significantly greater inherent flexibility than the sequence employed in the present study; this could result in the need for a larger minimum layer of water. Although the need for a 10-Å minimum layer of water, for DNA simulations, has not been proved necessary in the present work, we nevertheless suggest the use of a 10-Å minimum layer of water, for the sake of caution, until an adequate range of additional sequences and compositions has been similarly tested.

We thank Dr. Yajun Zheng for helpful discussions.

This work was supported by the Laboratory Directed Research and Development Program (LDRD) at Pacific Northwest National Laboratory (RLO). Pacific Northwest National Laboratory is a multiprogram national laboratory operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC06-76RLO 1830. The postdoctoral fellowship of ONS is administered by Associated Western Universities, Northwest Division (AWU NW), under grant DE-FG06-89ER-75522 or DE-FG06-92RL-12451 with the U.S. Department of Energy.

TABLE 1 Comparison of RMSD, cross RMSD, and curvature for the B1, B2, and B3 simulations

RMSD*	$B1_{750-1500\ ps}$#	$B2_{750-1500\ ps}$	$B3_{750-1500\ ps}$	Crystal	A-DNA	B-DNA
B1	1.22 ± 0.27	1.45 ± 0.25	1.50 ± 0.38	3.22 ± 0.65	4.67 ± 0.41	3.51 ± 0.62
B2	1.37 ± 0.20	1.13 ± 0.19	1.48 ± 0.20	3.32 ± 0.32	4.37 ± 0.31	3.55 ± 0.37
B3	1.63 ± 0.35	1.65 ± 0.28	1.35 ± 0.41	2.93 ± 0.56	4.56 ± 0.60	3.27 ± 0.61
Crystal	—	—	—	—	5.96	1.70
Curvature§						
12 bp	33 ± 15	34 ± 14	34 ± 17	20	0	0
A_6-tract	23 ± 12	28 ± 11	27 ± 13	7	0	0

* The RMSD was calculated by optimizing the overlap between every snapshot in the dynamic trajectory taken at 0.5-ps intervals and the reference structure (the average structure in the 750–1500-ps time period, the crystal structure, the canonical A or B form). All 12 base pairs and all atoms are included in the calculation. Average conformations from the trajectories in the 750–1500-ps interval were calculated after superposition as described above using the MDANAL module of AMBER 4.1 (Pearlman et al., 1995). The crystal structure (1D89) was taken from DiGabriele and Steitz (1993). The models for the canonical structures (Arnott and Hukins, 1972) were generated using the NUCGEN module of AMBER 4.1.

The diagonal values in the first three columns depict the average RMSDs and their standard deviations for each trajectory against their average conformation in the 750–1500-ps interval. The off-diagonal values in the first three columns represent the cross RMSDs (for instance, column 1, row 2 shows the average RMSD and standard deviation between the average conformation of B1 and the dynamic trajectory of B2 in the 750–1500-ps interval). The cross RMSD is a simple way of comparing two independent dynamic trajectories of the same DNA system.

§ Global DNA curvature was calculated by fitting a circumference to the set of helix axis reference points determined using the program CURVES (Lavery and Sklenar, 1989). The angle of curvature (or, simply, curvature) is obtained by taking the scalar product of the two radii vectors encompassing the helical axis of the DNA fragment (Norberto de Souza, 1994).

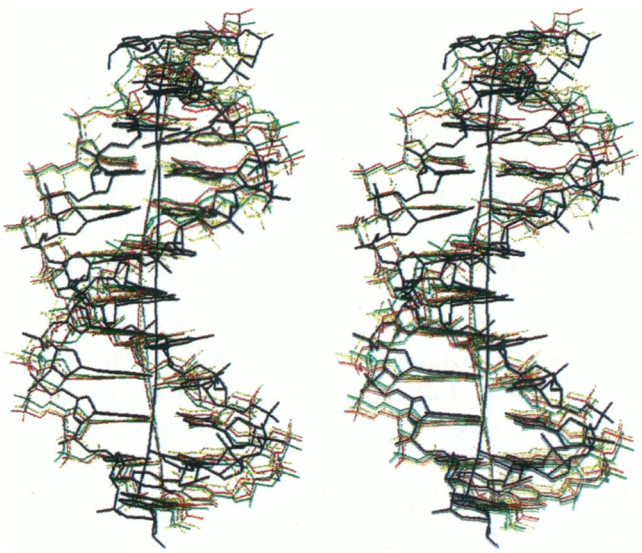

FIGURE 2 Stereoplot of the average structure and helix axis of the double-helix dodecamer d(CGCGA6CG)·d(CGT6CGCG) in simulations B1 (*green*), B2 (*red*), and B3 (*yellow*) superimposed on the crystal structure (*black*). Notice the increased curvature of the average structure in solution compared to the relatively straight helix axis of the structure in the crystal.

REFERENCES

Arnott, S., and D. W. L. Hukins. 1972. Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* 47:1504–1509.

Cheatham, T. E., III, and P. A. Kollman. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259:434–444.

Cheatham, T. E., III, J. J. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald methods leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* 117:4193–4194.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 103:8577–8593.

DiGabriele, A. D., and T. A. Steitz. 1993. A DNA dodecamer containing and adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* 231:1024–1039.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 98:10089–10092.

Lavery, R., and H. Sklenar. 1989. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Lee, H., T. A. Darden, and L. G. Pedersen. 1995. Molecular dynamics study of a high resolution Z DNA crystal. *J. Chem. Phys.* 102: 3830–3834.

Miaskiewicz, K., R. Osman, and H. Weinstein. 1993. Molecular dynamics simulation of the hydrated $d(CGCGAATTCGCG)_2$ dodecamer. *J. Am. Chem. Soc.* 115:1526–1537.

Norberto de Souza, O. 1994. Intrinsic curvature of DNA: a molecular dynamics study. Ph.D. thesis. University of London.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, III, D. M. Ferguson, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 1995. AMBER 4.1. University of California, San Francisco, CA.

Smith, P. E., and B. M. Pettit. 1996. Ewald artifacts in liquid state molecular dynamics simulations. *J. Chem. Phys.* 105:4289–4293.

Swaminathan, S., G. Ravishanker, and D. L. Beveridge. 1991. Molecular dynamics of B-DNA including water and counterions: a 140-ps trajectory for $d(CGCGAATTCGCG)_2$ based on the GROMOS force field. *J. Am. Chem. Soc.* 113:5027–5040.

Weerasinghe, S., P. E. Smith, V. Mohan, Y.-K. Cheng, and B. M. Pettit. 1995. Nanoseconds dynamics and structure of a model DNA triplex in saltwater solution. *J. Am. Chem. Soc.* 117:2147–2158.

York, D. M., W. Yang, H. Lee, T. Darden, and L. G. Pedersen. 1995. Toward accurate modeling of DNA: the importance of long-range electrostatics. *J. Am. Chem. Soc.* 117:5001–5002.

Zichi, D. A. 1995. Molecular dynamics of RNA with the OPLS force field. Aqueous simulations of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.* 117:2957–2969.