Journal of Biomolecular NMR, **13:** 119–131, 1999.
KLUWER/ESCOM

Restrained molecular dynamics of solvated duplex DNA using the particle mesh Ewald method

David E. Konerding[a], Thomas E. Cheatham III[b], Peter A. Kollman[c] & Thomas L. James[c,*]
[a]*Graduate Group in Biophysics, Box 0446, University of California, San Francisco, CA 94143, U.S.A.;* [b]*Laboratory of Biophysical Chemistry 12A-2041, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892-5626, U.S.A.;* [c]*Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, U.S.A.*

Received 7 July 1998; Accepted 14 October 1998

Key words: DNA structure, molecular dynamics, particle mesh Ewald, solvated refinement

Abstract

Restrained and unrestrained aqueous solution molecular dynamics simulations applying the particle mesh Ewald (PME) method to DNA duplex structures previously determined via in vacuo restrained molecular dynamics with NMR-derived restraints are reported. Without experimental restraints, the DNA decamer, d(CATTTGCATC)·d(GATGCAAATG) and trisdecamer, d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT), structures are stable on the nanosecond time scale and adopt conformations in the B-DNA family. These free DNA simulations exhibit behavior characteristic of PME simulations previously performed on DNA sequences, including a low helical twist, frequent sugar pucker transitions, B_I-B_{II}($\varepsilon - \zeta$) transitions and coupled crankshaft ($\alpha - \gamma$) motion. Refinement protocols similar to the original in vacuo restrained molecular dynamics (RMD) refinements but in aqueous solution using the Cornell et al. force field [Cornell et al. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197] and a particle mesh Ewald treatment produce structures which fit the restraints very well and are very similar to the original in vacuo NMR structure, except for a significant difference in the average helical twist. Figures of merit for the average structure found in the RMD PME decamer simulations in solution are equivalent to the original in vacuo NMR structure while the figures of merit for the free MD simulations are significantly higher. The free MD simulations with the PME method, however, lead to some sequence-dependent structural features in common with the NMR structures, unlike free MD calculations with earlier force fields and protocols. There is some suggestion that the improved handling of electrostatics by PME improves long-range structural aspects which are not well defined by the short-range nature of NMR restraints.

Abbreviations: NOE, nuclear Overhauser effect; MD, molecular dynamics; RMD, restrained MD; PME, Particle Mesh Ewald; rmsd, root-mean-square deviation.

Introduction

Improving the resolution of the structure of DNA in solution is a major challenge which requires both experimental data and theoretical modeling. It has long been recognized that solvent conditions profoundly influence the structure of DNA (Franklin et al., 1953) and that the specific sequence can also play a role in the structure and deformability of nucleic acids in solution (Wang et al., 1979). Static structure and dynamic deformability properties of DNA have important implications for recognition by proteins in biological processes such as transcription, as well as packaging and damage repair and may have applications in drug design and gene therapy. Therefore, correctly representing the structure of DNA in solution is paramount if understanding of these processes at a molecular level is to be achieved.

*To whom correspondence should be addressed.

Defining the atomic level structure of DNA in solution to adequate resolution has been difficult. NMR is the principal method for determining high resolution solution structures of DNA. However, NOE intensities can only be observed for interproton pair distances up to 5–6 Å, limiting our ability to define global DNA structure accurately beyond the base pair or base step level. Additionally, overlap of protons in the NOESY spectrum and poor signal-to-noise ratio can lead to the absence of data. Of the approximately 2000 interproton pairs with distances less than 6 Å in a canonical B-DNA decamer, we can expect to observe only about 400–500, with a precision of about 0.25–1 Å. In addition, due to molecular motion in solvated DNA, NOESY intensities are subject to averaging which complicates the process of determining accurate distances (Ulyanov et al., 1995).

For these reasons, structures based on NMR data must include explicit a priori knowledge of chemical structure. Normally, this information is present in the form of a force field, including explicit parameters for the bond lengths, bond angles and dihedral angles, as well as atomic charges and van der Waals parameters for the system being studied. Typically, a published DNA structure results from refinement of a starting model by in vacuo restrained molecular dynamics (RMD) using a particular force field. In this case, the chemical structure is maintained by the force field, while the specific tertiary conformation of the molecule is achieved by means of interproton distance and torsion angle restraints (derived from NOE intensities and COSY coupling constants, respectively) added to the force field (Schmitz et al., 1995).

Because DNA is a highly charged biomolecule, accurate unrestrained molecular simulations of duplexes have been very hard to generate. Consistent improvement of methodology (Beveridge et al., 1994) has improved the quality of simulations, but obtaining stable trajectories for long time periods (> 1 ns) has been very difficult. Recently (Darden et al., 1993; York et al., 1993), a promising technique, particle mesh Ewald, was developed. The PME method computes a full representation of the electrostatic interactions for a periodic lattice using screened real space sums and Fourier transforms to evaluate the reciprocal space interactions, providing an alternate method for computing long-range electrostatics in nucleic acid simulations. Its application has led to stable nucleic acid dynamics trajectories of 1 ns and longer (Cheatham et al., 1995; Louise-May et al., 1996), without need for artificial restraints. Use of PME with the Cornell et al. (1995) force field has been validated by impressive results reproducing many of the known conformational features of solvated DNA. Simulations using unrestrained PME starting from canonical B and canonical A form DNA converge to structures very similar to experimental crystal structures, reproducing sequence-specific properties such as roll and tilt (Cheatham et al., 1996).

To determine whether these advances in force field and simulation methodologies could increase the quality of DNA structures determined by NMR, we examined the effect of PME molecular dynamics simulations on a DNA decamer, d(CATTTGCATC)·(GATGCAAAATG), and a trisdecamer, d(ACGTTGCCTTGAG)·d(CTCAAGGCAAC GT), both free and restrained by NMR data, and compared the results with the in vacuo RMD simulations used in the originally refined structures (Mujeeb et al., 1993; Weisz et al., 1992, 1994). These sequences were chosen since they represent some of the highest resolution NMR structures refined in our laboratory. Unrestrained and restrained simulations have been carried out, and the resulting ensembles obtained from both types of simulations have been analyzed to whether more accurate structures are found using the improved methodology.

Methods

Decamer setup and equilibration

A total of five different MD simulations using the Cornell et al. (1995) force field with PME have been performed on the DNA decamer using the AMBER 4.1 and AMBER 5.0 suite of programs (Pearlman et al., 1995). One long unrestrained MD run was carried out for 2 ns. Following this, restrained PME simulations starting from four different starting structures were carried out for 100 ps each, using the distance and torsion angle restraints of the original in vacuo NMR refinements.

DNA decamer solvation and equilibration

A rectangular periodic box containing TIP3P water and 18 sodium counterions was constructed using the AMBER EDIT module to solvate and neutralize the originally refined NMR DNA structure. The water box extended approximately 10 Å away from any solute atom, yielding approximately 3000 water molecules with a box size of 52 Å by 44 Å by 45 Å, giving an approximate concentration of 16 mM DNA and 270 mM

sodium. PME simulations were run with SHAKE on hydrogens (tolerance = 0.0005 Å), a 1 fs time step, a temperature of 300 K with Berendsen temperature coupling (Berendsen et al., 1984) with solvent and solute coupled to a common bath, a 9 Å cutoff applied to Lennard-Jones interactions in PME, and constant pressure with pressure scaling ($tau_T = 0.2$ ps, $tau_P = 0.2$ ps). The nonbonded list was updated every 10 steps. The PME charge grid spacing was approximately 1.0 Å, and the charge grid was interpolated with a cubic B-spline (Essman et al., 1995).

Water and solute equilibration was performed by minimizing the water and counterions for 2500 steps while holding the DNA fixed to its initial atomic coordinates (the original in vacuo NMR structure). Next, 25 ps of non-PME dynamics were run, raising the temperature of the system from 100 K to 300 K while holding the DNA fixed to its atomic coordinates. Then, 1000 steps of minimization were applied, allowing the water and counterions to move freely while the DNA was restrained to its atomic coordinates using a harmonic potential of 25 kcal/mol. Following this, 10 ps of non-PME dynamics were run, allowing the water and counterions to move freely while restraining the DNA with a 1000 kcal/mol harmonic potential. Next, five consecutive 2000-step minimizations were performed, decreasing the harmonic potential from 20 to 0 kcal/mol in 5 kcal/mol steps. As a final equilibration step, a 3 ps PME dynamics run with no restraints on DNA, counterions or water, warmed the system from 100 K to 300 K. At this point the system was considered to be equilibrated, and production runs at 300K were initiated. As a control for the PME force field, a long (2 ns) free PME simulation was run. The free PME simulation, referred to as fPME in this paper, remained stable and exhibited structural behavior similar to previously published DNA free PME simulations (Cheatham et al., 1996, 1998; Young et al., 1997). The rmsd (all atom, mass-weighted) of the average structure to the initial in vacuo refined NMR structure is 3.23 Å (Table 1), the average twist of the molecule is 29°, roll is positive, inclination is slightly negative, and has an average pseudorotation value near 120° (Table 2).

Original DNA decamer refinement in vacuo

The original in vacuo NMR refinement is described in detail in Weisz et al. (1994). Following 20 ps of RMD refinement with a 1 fs timestep at 300 K, the restraint-minimized structure was submitted to 100 ps of in vacuo RMD, generating the trajectory referred to as ivRMD in this paper.

DNA decamer refinement using PME

For the DNA decamer, four independent 100 ps restrained PME molecular dynamics runs were performed. The restraint force field utilized was k_{dist} = 20 kcal/(mol Å^2) for distance restraints and k_{tors} = 60 kcal/(mol rad^2) for torsion angle restraints. The restraint set used was identical to the original NMR restraint set: 100 torsion angle restraints, 398 distance restraints, and 48 additional Watson–Crick hydrogen bond restraints to maintain base pairing. The hydrogen bond flat angle restraints were 10 kcal/(mol rad^2) and the distance restraints were 18 kcal/(mol A^2). These hydrogen bond restraints were maintained for consistency with the initial refinement. The four starting structures for the runs corresponded to the initial and final frames of the unrestrained PME simulation, canonical A-DNA and canonical B-DNA. These simulations are referred to as RMDI, RMDF, RMDA and RMDB. The starting structures from the free PME simulation were already equilibrated in solvent, while the canonical A and B forms were solvated and equilibrated using the method described above for equilibrating the original NMR structure.

Velocities were assigned from a Maxwellian distribution to give a temperature of 300 K. The restraints were identical to those used in the original in vacuo NMR refinement. The restraint protocol is as follows: after an initial unrestrained PME dynamics equilibration period of 10 ps, the distance restraint forces were ramped from k_{dist} = 1 kcal/(mol Å^2) to 20 kcal/(mol Å^2) and torsion angle restraint forces were ramped from k_{tors} = 3 kcal/(mol rad^2) to 60 kcal/(mol rad^2). The target temperature remained constant at 300 K throughout the restrained portion of the simulation. This protocol is similar to the original NMR refinements but had a duration of 100 ps instead of 20 ps.

Analysis of decamer simulation trajectories

For each simulation, a representative ensemble comprising the final 50 ps at 1 ps intervals of the simulation, and the corresponding average structure over this range, were chosen to represent the structure and allow direct comparison between the simulations. These subsets were used to eliminate bias from the early sections of the trajectories, which were close to the starting structure, and to select the same number of samples from each simulation.

Table 1. Rms deviations between pairs of average structures from the free PME, restrained PME, and in vacuo decamer simulations. Rms deviations between pairs of coordinate averaged structures of the decamer. Lower left is all-atom mass-weighted rms, upper right is the inner octamer mass-weighted rms

	fPME[a]	RMDI	RMDF	RMDA	RMDB	ivRMD	Adna	Bdna
fPME		1.78	1.75	1.89	2.01	2.89	2.56	3.65
RMDI	2.04		0.24	0.61	0.78	1.23	2.75	2.79
RMDF	2.02	0.25		0.60	0.76	1.28	2.73	2.82
RMDA	2.19	0.62	0.60		0.86	1.41	2.73	2.86
RMDB	2.27	0.88	0.83	0.95		1.39	2.91	2.59
ivRMD	3.23	1.35	1.39	1.46	1.58		3.18	2.68
Adna	2.87	3.23	3.22	3.25	3.40	3.73		4.73
Bdna	4.23	3.14	3.16	3.24	2.96	2.96	5.59	

[a]Acronyms used in Tables 1, 2 and 5: fPME: 2 nsec free PME simulation of decamer DNA; RMDI: RMD PME starting from the initial free PME conformation; RMDF: RMD PME starting from the final free PME conformation; RMDA: RMD PME starting from canonical A form; RMDB: RMD PME starting from canonical B form; IvRMD: RMD in vacuo (original NMR simulation); Adna: A-DNA; Bdna: B-DNA.

To describe the conformational space sampled in the simulations, the backbone torsion angles and helical parameters of the DNA structures were calculated using the Dials and Windows (Ravishanker et al., 1989) interface to Curves (Lavery et al., 1989). The parameters were computed for each structure in an ensemble, then arithmetically averaged. The results of the average of the parameters is more informative than parameters calculated from the average structure, because coordinate averaging is subject to motion artifacts and, with the ensemble, a statistical distribution of values is obtained rather than a single value.

The quality of fit of structures compared to the experimental data was calculated using a sixth-root-weighted R-factor, R_x (James, 1991). The calculation of the R_x factor used the experimental NOESY intensities with a mixing time of 140 ms and a correlation time of 2 ns. The experimental NOESY intensities were available for the decamer but not the trisdecamer, so R_x has not been calculated for the trisdecamer simulations.

For the calculation of structural energy and figures of merit, the sampling ensembles were used. Energies of structures from the simulations were computed using AMBER 5.0 with the Cornell et al. force field applied to water- and counterion-stripped sample frames. No electrostatic cutoff was applied, and a distance-dependent dielectric with a dielectric constant of 4 was used to represent crudely the dielectric screening by bulk solvent. Energies of the individual members of the sampling ensembles were computed, then arithmetically averaged. The coordinate-average structures were computed by averaging the trajectory subsets described above. Following this averaging, the rmsd of all atoms with mass weighting was computed for pairs of structures. The CARNAL module of AMBER 4.1 was used for coordinate averaging and computing rmsd.

Trisdecamer setup and equilibration

A series of different MD simulations using the Cornell et al. force field and PME (Cornell et al., 1995) have been performed on the DNA trisdecamer using the AMBER 4.1 and AMBER 5.0 suite of programs. One long unrestrained MD simulation using the PME force field was carried out for 1 ns, and will be referred to as fPME. As with the decamer fPME simulation, the trisdecamer simulation exhibited the same structural properties as previously published free PME simulations of duplex DNA. rmsd (all-atom, mass-weighted) to the initial in vacuo refined NMR structure is 3.17 Å (Table 3), the average twist of the molecule is 30°, roll is positive, the molecule has a slightly negative inclination and an average pseudorotation value near 120° (Table 4). Following this, restrained PME simulations starting from four different starting structures (two canonical A-DNA and two canonical B-DNA form DNAs), each with a different random number seed, were carried out for 250 ps each using the distance and torsion angle restraints of the original in vacuo NMR refinements. These simulations will be referred to as RMDA1, RMDA2, RMDB1 and RMDB2. The PME equilibration was essentially the same as the decamer

Table 2. Average helical parameters and backbone angles of decamer simulation ensembles. Standard angle and helical values averaged over residues, base pairs, or base pair steps (where appropriate) for the decamer structures specified. Average values were calculated by arithmetically averaging the values calculated for the individual structures within each sampling ensemble. Standard deviations are parenthesized

	fPME[a]	RMDI	RMDF	RMDA	RMDB	ivRMD
Shear	0.0(0.2)	0.1(0.1)	0.0(0.1)	0.0(0.1)	0.0(0.1)	0.1(0.1)
Stretch	0.2(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.0)	−0.3(0.1)
Stagger	−0.2(0.2)	−0.1(0.1)	−0.1(0.1)	−0.1(0.2)	−0.1(0.1)	0.0(0.1)
Buckle	−0.8(4.5)	−2.1(3.3)	−2.2(2.4)	−1.7(2.9)	−2.1(2.7)	−5.0(3.0)
Propeller	−11.8(3.4)	−12.4(3.2)	−11.9(3.1)	−12.5(2.7)	−11.4(2.8)	−18.4(3.2)
Opening	3.8(2.3)	−0.5(1.7)	−0.7(1.5)	−0.2(1.6)	−2.1(1.5)	−5.9(1.7)
Shift	−0.0(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.0)
Slide	−0.2(0.1)	−0.2(0.1)	−0.2(0.1)	−0.2(0.1)	−0.3(0.0)	−0.2(0.0)
Rise	3.3(0.2)	3.2(0.1)	3.2(0.1)	3.1(0.1)	3.2(0.1)	3.1(0.0)
Tilt	0.5(1.9)	1.3(1.2)	0.8(1.2)	1.2(1.1)	0.8(1.0)	1.0(0.9)
Roll	7.4(2.7)	1.9(1.5)	2.0(1.6)	1.0(1.5)	3.8(1.3)	1.0(1.4)
Twist	29.0(1.2)	33.1(0.5)	32.8(0.6)	32.9(0.6)	33.0(0.5)	36.3(0.6)
X Disp.	−1.9(0.8)	−1.9(0.3)	−1.8(0.3)	−2.0(0.3)	−1.3(0.2)	−1.6(0.2)
Y Disp.	−0.0(0.5)	−0.4(0.2)	−0.4(0.3)	−0.4(0.2)	−0.5(0.2)	−0.4(0.2)
Inclination	−4.8(6.2)	4.1(2.6)	4.0(2.4)	5.7(2.6)	−0.0(2.2)	6.3(2.3)
Tip	−0.6(3.8)	1.4(2.1)	1.5(2.4)	1.5(2.0)	0.5(1.9)	1.2(1.7)
Axis X Disp.	−0.0(0.1)	−0.0(0.0)	−0.0(0.0)	−0.0(0.0)	0.0(0.0)	−0.0(0.0)
Axis Y Disp.	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)
Axis inc.	0.5(1.3)	0.8(0.8)	0.7(0.9)	0.9(0.8)	−0.2(0.7)	0.4(0.7)
Axis tip	6.4(2.2)	1.9(1.0)	2.1(1.2)	1.5(1.1)	3.4(0.8)	1.5(1.0)
delta	110.7(3.3)	114.1(1.4)	113.7(1.5)	113.3(1.6)	113.2(1.9)	117.6(1.6)
epsilon	178.8(3.0)	172.3(1.6)	172.8(1.7)	172.0(1.5)	172.4(1.8)	171.3(1.5)
zeta	254.9(3.7)	256.1(1.7)	255.9(1.6)	256.4(1.6)	256.4(1.9)	255.4(1.9)
alpha	264.9(1.8)	267.0(2.5)	267.5(2.2)	277.9(1.7)	277.3(2.1)	262.8(2.9)
beta	162.5(2.7)	164.3(2.0)	164.0(1.8)	163.0(1.7)	163.2(1.5)	165.1(1.5)
gamma	61.2(2.1)	57.0(2.0)	56.8(1.7)	48.5(1.7)	48.4(1.5)	61.9(2.4)
chi	229.9(2.7)	236.8(1.8)	236.8(1.7)	237.3(2.0)	237.0(1.6)	237.3(1.6)
Pucker	114.8(9.5)	129.6(1.4)	129.3(1.4)	129.1(1.9)	128.8(1.7)	131.9(1.6)
Amplitude	40.6(1.4)	32.2(0.7)	32.2(0.9)	32.3(0.9)	32.4(0.6)	32.3(0.7)

[a] Acronyms are explained in the footnote to Table 1.

simulation, except in the case of the restrained simulations, the first 25 ps involved non-PME dynamics and were followed by 25 ps of PME dynamics with the DNA held fixed in both cases before PME production dynamics. In the production dynamics, the restraint force constants were $k_{dist} = 20.0$ kcal/(mol Å^2) and $k_{tors} = 60.0$ kcal/(mol rad^2). The restraints were ramped up smoothly to 1/4 strength during the first 2–10 ps, followed by ramping to full strength over the next 10 ps. The time step was 2 fs. Similarly, in the absence of artificial hydrogen bond restraints, it is necessary to ramp up the restraint force constants slowly since otherwise the structure may rapidly distort (e.g., by breaking base pairs) in order to instantaneously satisfy the restraints.

Analysis of trisdecamer simulation trajectories

In each case, a representative ensemble comprising the final 50 ps at 1 ps intervals of the simulation and the corresponding average structure over this range was chosen to represent the structure and allow direct comparison between the simulations. The conformational properties of the trisdecamer ensembles were calculated using Dials and Windows, and the energetics and quality of fit factors were computed using AMBER 5.0, in the same manner as for the decamer ensembles.

Table 3. Rms deviations between pairs of average structures from the free PME, restrained PME, and in vacuo trisdecamer simulations. Lower left is all-atom mass-weighted, upper right is the inner decamer

	fPME	ivRMD	RMDA1	RMDA2	RMDB1	RMDB2
fPME		2.83	3.28	3.03	2.10	1.78
ivRMD	3.17		2.90	2.42	1.48	1.85
RMDA1	3.43	3.30		0.79	2.79	3.15
RMDA2	3.15	2.76	0.92		2.37	2.79
RMDB1	2.21	1.70	2.94	2.43		0.68
RMDB2	1.89	2.10	3.26	2.84	0.72	

[a] Acronyms used in Tables 3, 4 and 6: fPME: 1 nsec free PME simulation of trisdecamer DNA; ivRMD: Original NMR structure; RMDA1: RMD PME starting from A-DNA conformation; RMDA2: RMD PME starting from A-DNA conformation; RMDB1: RMD PME starting from B-DNA conformation; RMDB2: RMD PME starting from B-DNA conformation.

Simulation details

The decamer and trisdecamer simulations were run using the Sander module of AMBER 4.1 and 5.0 on the Cray T3D and T3E at the San Diego Supercomputer Center, Pittsburgh Supercomputer Center and local computers.

Results

Unrestrained and restrained PME simulations were run for both the DNA decamer, d(CATTTGCATC)·d(GATGCAAATG), and trisdecamer, d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT). These sequences have been well defined by NMR data, with a high number of restraints per residue. The decamer has nearly 20 distance restraints per residue, which are derived from the cross-relaxation matrix MARDIGRAS analysis of the NOE intensities. The MARDIGRAS bounds are extremely tight; the average flat-well width over all distance restraints was 0.25 Å. The tightness of the bounds is due to the MARDIGRAS technique. A newer method for determining distance bounds from NOE intensities, RANDMARDI (Liu et al., 1995), would produce wider bounds which better represented the inherent imprecision of the intensities. However, the original distance bounds were used to maintain consistency with the original refinement. The trisdecamer has approximately 10 distance restraints per residue. The results of both the decamer and trisdecamer simulations demonstrate similar general trends between the free and restrained simulations; for brevity, we will focus first on the results of the decamer simulations in depth, then compare these with the trisdecamer simulations.

Decamer simulations

Comparison of energies and goodness of fit between unrestrained PME, restrained PME, and restrained in vacuo structures

As described in the Methods, the structural energy and goodness of fit were computed for sampling periods from the MD simulations. Instead of computing the structural energy and fit of the average structures, we have computed the average structural energy and fit from the sample structures making up the ensemble. There is a distinct advantage in computing the average energy of these samples from an ensemble in that the conformational energy is not artificially high due to the coordinate averaging process. Dynamic processes in the simulation, such as backbone torsion angle fluctuations, helix axis flexibility, and methyl rotation lead to anomalously high bond and angle energies upon straight coordinate averaging. Minimization is necessary to eliminate these artifacts but will only move the structure to a nearby local energy minimum. The local energy minimum may not correspond to the global minimum sought by RMD simulation. Therefore it is not very informative to compare the structural energies of minimized average structures generated from different ensembles. Even in the best case, where the dynamics represent fluctuations around a single mean rather than transitions between conformationally accessible substates, these energies are not particularly informative.

Average energies computed from the simulations are illuminating. The free PME simulation has a

Table 4. Average helical parameters and backbone angles of trisdecamer simulation ensembles. Standard angle and helical values averaged over residues, base pairs, or base pair steps (where appropriate) for the trisdecamer structures specified. Average values were calculated by arithmetically averaging the values calculated for the individual structures within each sampling ensemble. Standard deviations are parenthesized. The ivRMD standard deviations are zero because it is only a single structure

	fPME	ivRMD	RMDA1	RMDA2	RMDB1	RMDB2
Shear	−0.0(0.1)	−0.0(0.0)	−0.0(0.1)	−0.1(0.1)	−0.1(0.1)	−0.2(0.1)
Stretch	0.1(0.1)	−0.2(0.0)	−0.0(0.2)	0.1(0.1)	0.0(0.1)	0.0(0.1)
Stagger	−0.2(0.1)	0.1(0.0)	−0.1(0.1)	−0.1(0.2)	0.0(0.1)	−0.1(0.2)
Buckle	−0.1(3.8)	0.6(0.0)	−9.4(2.4)	−9.4(3.3)	−2.4(3.1)	−0.3(2.8)
Propeller	−5.8(3.8)	−18.3(0.0)	−3.2(3.8)	−5.8(2.7)	−10.6(3.2)	−10.7(3.5)
Opening	2.1(1.9)	−4.4(0.0)	−1.6(2.3)	−0.8(1.8)	−2.7(1.7)	−2.2(1.6)
Shift	−0.1(0.1)	−0.0(0.0)	0.2(0.1)	0.1(0.1)	0.0(0.1)	0.0(0.1)
Slide	−0.2(0.1)	0.0(0.0)	−0.1(0.1)	−0.1(0.1)	−0.1(0.1)	−0.1(0.1)
Rise	3.6(0.1)	3.1(0.0)	3.1(0.1)	3.1(0.1)	3.3(0.1)	3.3(0.1)
Tilt	0.6(1.8)	1.4(0.0)	−1.5(1.6)	0.2(1.2)	0.4(1.3)	0.4(1.3)
Roll	8.5(1.7)	3.0(0.0)	4.0(2.3)	2.7(2.0)	4.2(1.4)	4.4(1.4)
Twist	30.2(0.6)	34.3(0.0)	32.7(0.9)	33.6(0.7)	33.4(0.6)	32.8(0.6)
X Disp.	−1.2(0.4)	−2.4(0.0)	−3.1(0.4)	−3.2(0.4)	−1.9(0.3)	−1.8(0.3)
Y Disp.	−0.1(0.4)	−0.4(0.0)	0.5(0.5)	0.1(0.3)	−0.2(0.3)	−0.2(0.3)
Inclination	−9.3(3.1)	6.3(0.0)	20.1(4.0)	21.0(3.6)	3.9(2.7)	1.0(2.9)
Tip	1.8(2.7)	0.6(0.0)	−6.9(3.6)	−3.7(3.0)	−2.2(2.4)	−2.2(2.0)
Axis X Disp.	−0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	−0.0(0.0)
Axis Y Disp.	−0.1(0.0)	−0.0(0.0)	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)	−0.1(0.0)
Axis inc.	−0.6(1.1)	0.3(0.0)	−1.8(1.0)	−0.5(1.0)	−0.3(1.0)	−0.2(1.0)
Axis tip	7.2(1.3)	2.2(0.0)	5.2(1.9)	3.5(1.6)	3.9(1.0)	3.8(1.1)
delta	114.2(2.7)	117.4(0.0)	118.7(1.3)	119.2(1.5)	119.0(1.3)	118.6(1.7)
epsilon	198.4(2.5)	178.5(0.0)	198.5(1.9)	199.2(2.0)	186.5(1.6)	185.7(1.7)
zeta	259.1(3.4)	272.1(0.0)	254.9(2.1)	256.3(1.8)	267.1(1.7)	267.3(1.9)
alpha	280.8(1.7)	291.7(0.0)	286.7(7.7)	291.2(1.9)	292.0(1.8)	294.0(2.3)
beta	171.2(2.1)	180.8(0.0)	170.3(3.3)	172.0(1.6)	176.4(1.5)	177.3(1.6)
gamma	62.3(1.7)	54.0(0.0)	43.4(2.9)	44.0(2.8)	50.7(1.6)	49.2(2.0)
chi	231.9(2.9)	244.7(0.0)	257.2(1.9)	255.3(1.8)	245.1(2.0)	244.2(2.3)
Pucker	117.8(5.8)	138.5(0.0)	141.9(1.8)	142.4(2.0)	141.7(1.5)	141.4(1.5)
Amplitude	40.4(1.3)	29.1(0.0)	28.6(0.8)	28.5(0.9)	28.8(0.8)	28.9(0.8)

[a]Acronyms are explained in the footnote to Table 3.

Table 5. Energies and statistics of fit for decamer simulation structures

	ivRMD[a]	RMDI	RMDF	RMDA	RMDB	free PME
Eamber	582.90	543.79	554.59	551.68	544.93	439.76
Econst	284.87	283.03	299.69	284.66	283.87	2396.33
AVDB	0.15	0.14	0.15	0.14	0.14	0.39
Rx	0.06	0.06	0.06	0.06	0.06	0.13

[a]Acronyms are explained in the footnote to Table 1.

slightly lower conformational energy (439 kcal/mol vs. 540–580 kcal/mol, where conformation energy is the sum of bond length, angle, dihedral and non-bonded terms) (Table 5) than the restrained simulations, which makes sense because the restraints tend to move the structure simulations away from the idealized, lower conformational energy structure preferred by the force field in an effort to minimize the artificial

restraint energy. It is quite reasonable that the constraint energy of the free PME structure is significantly higher than the constraint energy of the restrained runs (2396 kcal/mol vs. about 280–290 kcal/mol), and the R_x, 0.13, is much higher than for the original in vacuo RMD structure (0.06). An R_x of 0.13 is approximately the same canonical A- (0.16) or B-DNA (approximately 0.11), essentially a non-fit to the data. Comparing the in vacuo RMD energies with the RMD PME energies, we see that virtually the same E_{amber} and E_{const} values, as well as AVDB and R_x, are found for both types of restrained simulations.

No matter which type of fit we use to compare the free MD with the RMD runs, a clear trend exists: the free MD structures do not fit the data nearly as well as the RMD runs. What is surprising is that all the RMD runs, in vacuo or solvent/PME, have the same quality of fit, independent of the force field and whether explicit solvent is included. Given the dissimilarity of the magnitude of twist (vide infra) between the in vacuo run and the restrained PME runs, this demonstrates that the quality of fit is not adequate to distinguish the absolute value of twist between the two structures. However, the restraints are still sufficient to determine the relative value of twists between base steps. This is rationalized by the fact that NOEs are short range, never giving direct information beyond the base pair or base step. Nevertheless, if short-range distances are determined with sufficient precision and combined with a sophisticated force field, which accurately represents the long-range features of the molecule, the global structure of the DNA should be accurately defined.

Restrained PME molecular dynamics compared to restrained in vacuo molecular dynamics: Comparison of time-averaged helical and pseudorotation parameters

The restrained PME simulations agree well with the restrained in vacuo simulation when helical parameters are compared (Figure 1). This demonstrates that the restraints act independently of the force field to determine sequence-specific variations in helical parameters. While all the different RMD PME runs converge to identical structures, ranging between 0.25 and 0.95 Å (all atoms, mass-weighted pairwise rmsd between average structures from the sampling ensembles) as shown in Table 1, the RMD PME runs do not find precisely the same structure as the original in vacuo RMD simulation structure (rmsd ranging between 1.3 and 1.6 Å).

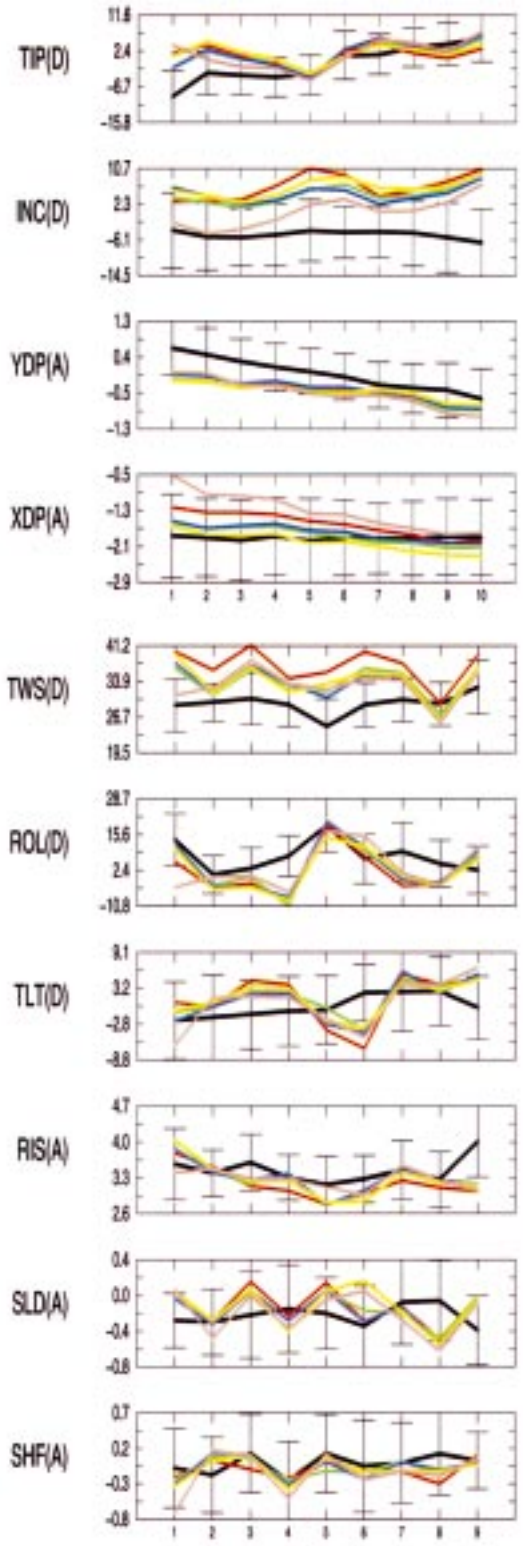

Figure 1. Average of helical parameters and backbone angles over the ensemble structures from the original in vacuo RMD, free PME and PME RMD simulations calculated with the Dials and Windows interface to Curves. Parameters were calculated for individual structures taken from the sampling ensembles of the trajectories, and then arithmetically averaged. The x-axis represents the base position in the sequence and the y-axis is the parameter value. Parameters in Å are marked (A) and parameters in degrees are marked (D). The lines are colored as follows: Free PME (black), ivRMD (red), RMDI (green), RMDF (blue), RMDA (yellow), RMDB (brown). Vertical bars represent the standard deviation for the free PME simulation.

Despite the agreement in helical parameters, we are not confident that 100 ps simulations using this protocol represent adequate sampling; for example, the unusual α-γ conformation at the T5pG6 step seen in the free PME simulation is reproduced only in the RMD simulations which started from the PME or in vacuo structures which had that conformation initially. The RMD simulations started from canonical A-DNA and B-DNA, which do not have that α-γ conformation, do not converge to the unusual α-γ conformation in the 100 ps refinement. Moreover, careful inspection of backbone angle and helical parameters shows RMD simulations starting from A and B tend to cluster together and simulations starting from the in vacuo or free PME cluster together. The lack of complete agreement suggests incomplete sampling. This clustering effect is most noticeable in the χ angle at the G6 base and may represent a correlation between the α-γ conformation and the χ angle at that step. Admittedly, the backbone of DNA is really not defined by NMR restraints, with neither the α nor γ torsions being measured by COSY. Moreover, the barriers to rotation around these dihedral angles are likely large enough to preclude observing transitions during short simulations, so this probably represents a sampling problem with MD which is exacerbated by restraints which further inhibit sampling and by the presence of explicit water which slows the dynamics. All of these observations suggest the need for longer restrained simulations or a methodology which increases conformational sampling by reducing energy barriers.

In spite of the lack of complete convergence, magnitude and sequence-specific variations of roll and tilt, as well as pseudorotation, are nearly identical between the PME and in vacuo RMD simulations. X-displacement, inclination and other helical parameters are in excellent agreement in both magnitude and sequence-specific variation as well (Figure 1), although some solvated simulations do not converge as well as others. This lack of convergence also suggests that 100 ps RMD simulations are not long enough to adequately allow the transition to the final RMD structure. In free PME, the A to B DNA transition takes approximately 250 to 500 ps; whereas in the restrained simulations there is no repuckering and limited sampling, yet an A to B transition is enforced in 50 ps. Even with restraints, the energy barrier between the starting structure and the correct solution structure may be high enough that longer restrained simulations will be necessary to find the global minimum energy.

While sequence-specific variations in twist are maintained between PME and in vacuo restrained structures, the magnitude differs. It appears to be 'stuck' at an average of 33° (Table 2), which is between the value favored by the free PME structure (average = 29°) and the restrained in vacuo structure (average = 36°). This difference in magnitude of twist is the primary contributor to the surprisingly high atomic rmsd (1.3–1.6 Å) between RMD PME and in vacuo RMD structures. Naturally the question is raised why tilt and roll converge to values so close to the original NMR structure while twist mirrors the sequence-specific variations but not the actual magnitude. If one considers the influences of NMR restraints and the force field on a DNA structure, the logical explanation for this derives from the combination of the lack of explicit twist-defining NMR data and the underestimated twist with the Cornell et al. force field. 'Local' (base pair step) twist is poorly defined by the NMR data, and the 'global' twist of the entire molecule is primarily determined by the force field. Roll and tilt components of base pair steps are very well defined by NMR data, primarily due to the spatial arrangement of NOESY distances between base pairs in a step (Ulyanov et al., 1992). This is an important aspect of DNA solution structure that has not been fully addressed in the literature to date.

The equilibrium twist in the restrained PME run is balanced between the value predicted by the free PME simulation and the value predicted by the original restrained in vacuo simulation. Thus, the force field used does play a significant role in defining the global helical parameters, even when experimental distance restraints are added. We note that the distance restraints, which never extend beyond a single base pair or base step, are not necessarily violated when twist, tilt, and roll are modulated. There are two reasons why this is the case. First, NOE restraints are imprecise, i.e., upper and lower bounds define acceptable values, and so small changes in twist may not cause the NOE restraints to be violated. Second, compensations for

large changes in twist by other parameters such as slide and displacement can avoid strong violations of the NOEs (Ulyanov et al., 1992). Thus, it is not unreasonable that we can expect the force field to exert a significant influence on the average twist magnitude. In spite of this, NMR data still have an important role in determining sequence-specific variations from the mean. It is hoped that with optimization of the Cornell et al. force field to mend the low-twist bias, a more reasonable representation of the twist in these structures can be determined.

It is important to note that the RMD PME simulations do have an average twist which is close to the helical periodicity (10.6 bases per turn, twist = 34°) measured in solution using an independent enzymatic method (Rhodes et al., 1980). Additionally, we note that re-inspection of the original decamer NMR spectra reveals the absence of a cross peak between methyl hydrogens of bases T5 and T13. In the original in vacuo RMD structure, the methyl hydrogens are close enough (<6 Å) that a small peak should be observable. In the free PME and RMD PME structures, the methyl hydrogen distances are increased, due to the decreased bending at the T5-A8 steps, in better agreement with the spectra. While absence of a peak does not prove there is no intermolecular contact, it suggests that the original model may not be as accurate as the PME RMD models. These two observations represent important validations of the Cornell et al. force field with the PME method for use in NMR structure refinement.

Trisdecamer results: Comparison of in vacuo RMD with free PME and RMD PME

In contrast to the straightforward RMD PME refinements of the decamer, restrained simulations of the trisdecamer from different starting structures did not converge as readily to a common structure (Figure 2, Tables 3 and 6). In part, this is due to the quality of the restraints: there are fewer restraints per residue for the trisdecamer, and bounds are not as precise. Additionally, we found it necessary to add Watson–Crick hydrogen bond restraints to maintain base pairing during the simulations starting from A-DNA, since without these, the base pairs broke to instantaneously satisfy the 'B-DNA' restraints while in an 'A-DNA' geometry. The simulations were run for a longer period (250 ps instead of 100 ps) and still had trouble converging: the A-DNA simulations still have high inclination (approximately 20°), although the x-displacement and pseudorotation angles are compatible with the B-DNA

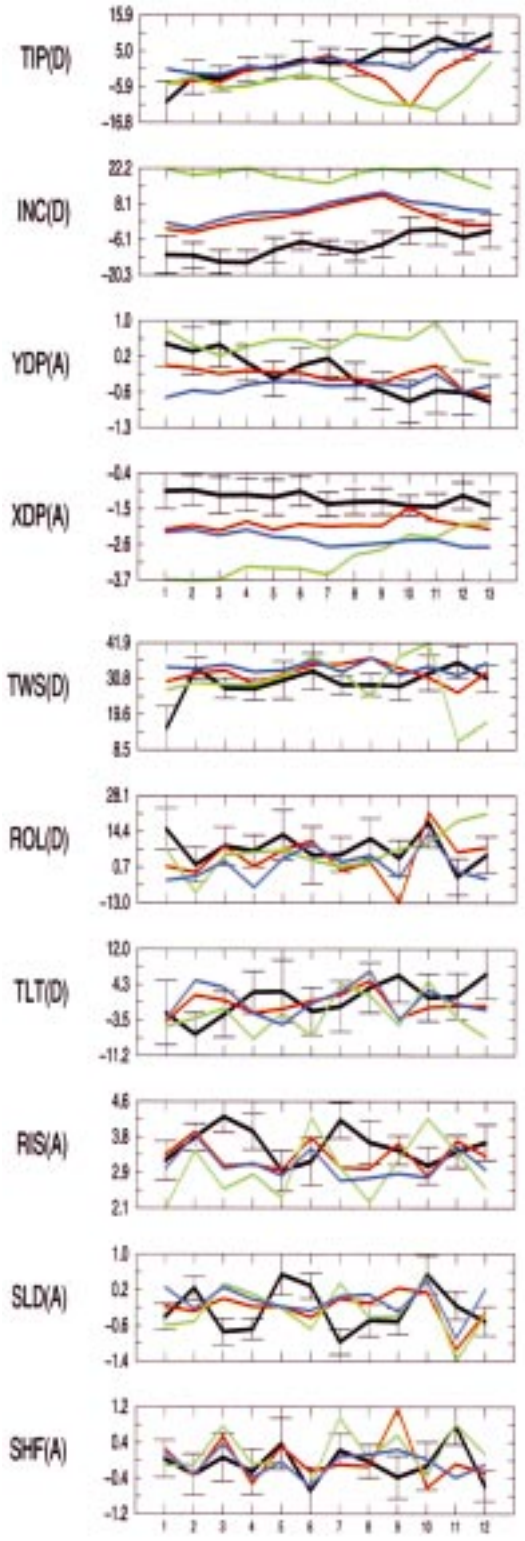

Figure 2. Average of helical parameters and backbone angles over the ensemble structures from the original in vacuo RMD structure, free PME simulation and PME RMD simulations calculated with the Dials and Windows interface to Curves. Parameters were calculated for individual structures taken from the sampling ensembles of the trajectories, and then arithmetically averaged. The x-axis represents the base position in the sequence and the y-axis is the parameter value. Parameters in Å are marked (A) and parameters in degrees are marked (D). The lines are colored as follows: Free PME (black), RMDB4 (red), RMDA4 (green), ivRMD (blue). Vertical bars represent the standard deviation for the free PME simulation.

simulations. The average structures from the simulations starting from A-DNA are 2.76 to 3.3 Å from the original NMR structure (Table 3), and the simulations starting from B-DNA are 1.7 to 2.1 Å from the original NMR structure, while the simulations starting from A-DNA are 2.4 to 3.2 Å from the simulations starting from B-DNA.

The difficulty of converging from A-DNA to B-DNA, despite the longer simulation, is likely due to difficulties in conformational sampling coupled with lower quality restraints. Adding explicit water slows conformational transitions, and spontaneous A-DNA to B-DNA transitions in free PME require ca. 500 ps (Cheatham et al., 1996). This time scale, coupled with the inhibited sampling observed with the restraints, suggests that longer simulation times are necessary or alternative methods need to be applied to enhance sampling.

Mirroring the trouble we had in obtaining converged structures (vide supra), the goodness of fit of the RMD PME structures starting from A-DNA does not fit the NMR data as well as the B-DNA start simulations or the originally determined in vacuo structure. The ensemble average E_{const} for the A-DNA starting structure simulations were above 700 kcal/mol, while the B-DNA starting structure simulations were around 230 kcal/mol and the original in vacuo simulation around 330 kcal/mol (Table 6).

Discussion and conclusions

We have shown that unrestrained MD simulations of DNA sequences, which have previously been characterized by NMR, are stable on the nanosecond time scale using the Cornell et al. force field with PME. However, the sampled structures are not fully consistent with the NMR data. In contrast, the restrained PME simulations converge to a common structure even when different starting structures are used, and the sequence-specific properties observed in the original in vacuo RMD simulations are reproduced fairly well, with the exception of helical twist. Unrestrained simulations using PME with the Cornell et al. force field have already been observed to show a lower twist than experimental data imply, so the lower twist in the RMD simulations is not surprising. Moreover, the PME RMD simulations, which converge to a structure with much lower average helical twist from the original in vacuo refinements, manifest nearly identical structural energies and figures of merit. This shows that NOE intensities, even in well-determined systems, cannot accurately define the magnitude of helical twist. Dependence on an accurate force field is therefore necessary in DNA structure refinement using NMR data.

Commonly, NMR refinements will use restrained MD with several different starting structures. Convergence from several different starting structures to a single common structure (usually measured by atomic rmsd) is treated as a measure of precision, in that the NMR data is sufficient to guide an RMD run to a single structure which satisfies the restraints. In this work, we used canonical A- and B-DNA forms as starting structures and measured whether convergence was reached. Although excellent convergence was readily achieved using the decamer data set and A-DNA and B-DNA forms, the trisdecamer data set was not sufficient to drive the A-DNA and B-DNA forms to the same final structure under the simulation conditions employed. In particular, the trisdecamer RMD simulation that started from A-DNA still maintained significant A-DNA conformational features such as x-displacement and inclination. Additionally, the α and γ torsion angles at step T5pG6 of the decamer do not converge from the simulations started from the original NMR structure conformation or the final PME conformation to the more typical values seen for the rest of the sequence and those seen in the canonical A-DNA and B-DNA simulations.

Normally, RMD refinement utilizes a simulated annealing approach, where the temperature of the simulation is raised to high values along with the force constants of the experimental restraints, followed by a cooling period where temperature and restraints are dropped significantly. This should help guide initial structures over large energy barriers to a region near the correct solution structure. Simulated annealing is challenging to implement when full solvation and periodic boundary conditions are used, since the commonly used water models were not parameterized for

Table 6. Energies and statistics of fit for trisdecamer simulation structures

	ivRMD[a]	A1	A2	B1	B2	fPME
Eamber	169.80	914.06	895.66	669.73	655.69	565.77
Econst	337.78	820.28	735.69	230.12	232.11	3553.34
AVDB	0.16	0.28	0.26	0.14	0.14	0.57

[a]Acronyms are explained in the footnote to Table 3.

use at high temperature, and high temperature leads to lower water density (which could disrupt the structure) or higher pressure (which may inhibit sampling further). However, judicious modifications to the simulation protocol, such as constant volume instead of constant pressure as well as shorter time steps, should allow higher temperatures during RMD runs. In the meantime, other methods of passing over the energy barrier of A-DNA to B-DNA interconversion need to be used. In the current work, it was necessary to use H-bond restraints and longer (250 ps vs. 100 ps) simulations for the trisdecamer in canonical A-DNA and B-DNA conformation to approach the 'correct' final structure without distortion of terminal base pairs. Since we know that free PME simulations readily interconvert from A-DNA to B-DNA on a 500 ps–1 ns timescale, this suggests that longer simulations, at least 500 ps–1 ns, may be necessary when using PME RMD with explicit water. A simple way to overcome some of these problems is to continue with the standard, rapid and efficient in vacuo refinement to generate a set of structures compatible with the data followed by submitting these structures to 50–500 ps of RMD in explicit solvent with PME and a reasonable nucleic acid force field. Alternatively, enhanced sampling methodologies, such as locally enhanced sampling, may be applied to effectively reduce barriers to conformational transition (Roitberg et al., 1991; Simmerling et al., 1998). It is believed that with better nucleic acid force fields, modern simulation techniques and inclusion of explicit solvent, more reliable refinement of NMR models can be performed to produce more realistic structures.

Acknowledgements

D.E.K. would like to acknowledge the Computer Graphics Lab at the University of California, San Francisco, for their generous donation of computer time on the Cray T3D and T3E at the San Diego Supercomputer Center. We would also like to thank the Pittsburgh Supercomputer Center for significant computational resources (MCA93S017P and CHE880090P). P.A.K. would like to acknowledge the National Institutes of Health for research support (CA-25644). T.L.J. would like to acknowledge the National Institutes of Health for research support (GM-39247).

References

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Beveridge, D.L. and Ravishanker, G. (1994) *Curr. Opin. Struct. Biol.*, **4**, 246–255.

Cheatham, T.E. and Kollman, P.A. (1996) *J. Mol. Biol.*, **259**, 434–444.

Cheatham, T.E. and Kollman, P.A. (1998) *Structure, Motion, Interaction and Expression of Biological Macromolecules*, Proceedings of the Tenth Conversation, State University of New York, Albany, NY.

Cheatham, T.E., Miller, J.L., Fox, T., Darden, T.A. and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 4193–4194.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Kenneth M. Merz, J., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.

Darden, T.A., York, D. and Pedersen, L.G. (1993) *J. Chem. Phys.*, **98**, 10089–10092.

Essman, U., Perera, L., Berkowitz, M., Darden, T., Lee, H. and Pedersen, L. (1995) *J. Chem. Phys.*, **103**, 8577–8593.

Franklin, R.E. and Gosling, R.G. (1953) *Acta. Crystallogr.*, **6**, 673–677.

James, T.L. (1991) *Curr. Opin. Struct. Biol.*, **1**, 1042–1053.

Lavery, R. and Sklenar, H. (1989) *J. Biomol. Struct. Dyn.*, **6**, 655–667.

Liu, H., Spielmann, H., Ulyanov, N. and Wemmer, D.E. (1995) *J. Biomol. NMR*, **6**, 390–402.

Louise-May, S., Auffinger, P. and Westhof, E. (1996) *Curr. Opin. Struct. Biol.*, **6**, 289–298.

Mujeeb, A., Kerwin, S.M., Kenyon, G.L. and James, T.L. (1993) *Biochemistry*, **32**, 13419–13431.

Ravishanker, G., Swaminathan, S., Beveridge, D.L., Lavery, R. and Sklenar, H. (1989) *J. Biomol. Struct. Dyn.*, **6**, 669–699.

Rhodes, D. and Klug, A. (1980) *Nature*, **286**, 573–578.

Roitberg, A. and Elber, R. (1991) *J. Chem. Phys.*, **95**, 9277–9286.

Schmitz, U. and James, T.L. (1995) *Methods Enzymol.*, **261**, 3–44.

Simmerling, C., Miller, J. and Kollman, P.A. (1998) *J. Am. Chem. Soc.*, **120**, 7149–7155.

Ulyanov, N.B., Gorin, A.A., Zhurkin, V.B., Chen, B.-C., Sarma, M.H. and Sarma, R.H. (1992) *Biochemistry*, **31**, 3918–3930.

Ulyanov, N.B., Schmitz, U., Kumar, A. and James, T.L. (1995) *Biophys. J.*, **68**, 13–24.
Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) *Nature*, **283**, 680–686.
Weisz, K., Shafer, R.H., Egan, W. and James, T.L. (1992) *Biochemistry*, **31**, 7477–7487.
Weisz, K., Shafer, R.H., Egan, W. and James, T.L. (1994) *Biochemistry*, **33**, 354–356.
York, D.M., Darden, T.A. and Pedersen, L.G. (1993) *J. Chem. Phys.*, **99**, 8345–8348.
Young, M.A., Ravishanker, G. and Beveridge, D.L. (1997) *Biophys. J.*, **73**, 2313–2336.