Haribabu Arthanari[1]
Kevin J. McConnell[2]
Richard Beger[3]
Matthew A. Young[4]
D. L. Beveridge[1]
Philip H. Bolton[1]

[1] *Department of Chemistry and Molecular Biophysics Program, Wesleyan University, Middletown, CT 06459*

[2] *Agilix Corporation, P. O. Box 8175, New Haven, CT 06530*

[3] *Division of Chemistry, National Center for Toxicological Research U.S. Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079-9502*

[4] *Department of Molecular and Cell Biology, 401 BARKER MC 3202, University of California, Berkeley, CA 94720-3202*

Assessment of the Molecular Dynamics Structure of DNA in Solution Based on Calculated and Observed NMR NOESY Volumes and Dihedral Angles from Scalar Coupling Constants

Received 25 April 2002; accepted 25 July 2002

Abstract: *To assess the accuracy of the molecular dynamics (MD) models of nucleic acids, a detailed comparison between MD-calculated and NMR-observed indices of the dynamical structure of DNA in solution has been carried out. The specific focus of our comparison is the oligonucleotide duplex, d(CGCGAATTCGCG)$_2$, for which considerable structural data have been obtained from crystallography and NMR spectroscopy. An MD model for the structure of d(CGCGAATTCGCG)$_2$ in solution, based on the AMBER force field, has been extended with a 14 ns trajectory. New NMR data for this sequence have been obtained in order to allow a detailed and critical comparison between the calculated and observed parameters. Observable two-dimensional (2D) nuclear Overhauser effect spectroscopy (NOESY) volumes and scalar coupling constants were back-calculated from the MD trajectory and compared with the corresponding NMR data. The comparison of these results indicate that the MD model is in generally good agreement with the NMR data, and shows closer accord with experiment than back-calculations based on the crystal structure of d(CGCGAATTCGCG)$_2$ or the canonical A or B forms of the sequence. The NMR parameters are not*

Correspondence to: Philip H. Bolton, email: pbolton@wesleyan.edu; or D. L. Beveridge, email: dbeveridge@wesleyan.edu

Contract grant sponsor: NIH
Contract grant number: 37909 and 65871
This article includes supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/0006-3525/suppmat/2002/68/v68.3.html

Biopolymers, Vol. 68, 3–15 (2003)

particularly sensitive to the known deficiency in the AMBER MD model, which is a tendency toward undertwisting of the double helix when the parm.94 force field is used. The MD results are also compared with a new determination of the solution structure of d(CGCGAATTCGCG)$_2$ using NMR dipolar coupling data. © 2002 Wiley Periodicals, Inc. Biopolymers 68: 3–15, 2003

Keywords: *molecular dynamics; nucleic acids; oligonucleotide duplex*

INTRODUCTION

Molecular dynamics (MD) simulations are a potentially valuable source of theoretical models for the structure and motions of DNA oligonucleotides in solution. The "dynamical structure" produced in MD is comprised of an ensemble of snapshots that approximates to a Boltzmann distribution as the length of the simulation increases. Recent improvements in nucleic acid force fields and the development of "second-generation" parameters designed for simulations including solvent have demonstrably advanced the ability of MD to predict DNA structure accurately.[1–3] However, assessments of the accuracy of MD models for DNA in solution have to date been based primarily on comparisons with structures obtained from x-ray crystallography. Crystal packing and other forces unique to the solid state effect the x-ray structures of DNA. Whether crystal structures of DNA provide accurate models for the dynamical structure of DNA in solution is not well established.[4–6] However this issue is resolved, it is obviously preferable to test MD models for solution structure against experimental data obtained directly on DNA in solution.

The method of choice for the study of DNA structures in solution is NMR spectroscopy. Two-dimensional (2D) nuclear Overhauser effect spectroscopy (NOESY) volumes and scalar coupling constants obtained from NMR provide a sensitive probes of structure[7] and the utilization of dipolar couplings in the determination of DNA structure from NMR has recently been introduced.[8] However, while certain aspects of secondary and tertiary structure of DNA can be deduced readily from NMR spectra, de novo determination of an all-atom DNA structure is generally underdetermined. Obtaining structures based on NMR data requires additional information, typically provided by empirical restraints and conformational energy calculations based on semiempirical force fields. As a consequence, details of NMR structures for uncomplexed DNA in solution tend to be sensitive to empirical energy refinement protocols, including the choice and parameterization of the energy functions per se, and therefore is not a suitable basis for unequivocal assessment of structures obtained from MD simulation.

In this study we adopt an alternative approach: direct "back-calculation" of the 2D NOESY volumes and dihedrals from the MD trajectories for DNA in solution. Corresponding NMR spectra were obtained at 500 MHz and analyzed with protocols for measuring the observables commensurate with those of the MD back-calculation. The comparison of the two allows the accuracy of the MD-calculated solution structure to be assessed without the issue of NMR structure refinement entering the picture. Specifically, the back-calculation of 2D NOESY volumes and dihedrals from a 14 ns room temperature MD trajectory on the prototype B-form duplex DNA duplex formed by the self-complementary sequence d(CGCGAATTCGCG)$_2$, obtained including counterions and water explicitly in the simulation, is described. The accuracy of the calculated MD model is assessed by a comparison of calculated and observed NOESY volumes and the corresponding dihedral angles obtainable by NMR experiments, both of which apply directly to the solution state. The sensitivity of the calculated NMR observables to the fine structure of the DNA is also discussed. In addition, a comparison of the MD structure for d(CGCGAATTCGCG)$_2$ a room temperature in solution with the independent "NMR-dipo" solution structure using dipolar coupling data in addition to NOESY data and scalar couplings is provided.

BACKGROUND

The current state of MD modeling of DNA structure has been reviewed.[1–3] Several recent studies[9–11] have documented that AMBER MD,[12] based on the Cornell et al. parm94 force field[13] with a particle mesh Ewald (PME) treatment of long-range interactions,[14] produce stable DNA MD trajectories on the nanosecond time scale and also provide a quite plausible description of the B-form double helix. The first applications of the parm94 force field to DNA and RNA sequences were reported by Kollman and co-workers.[15] Using AMBER and parm.94, Young et al.[16] reported an MD model of the Eco RI duplex, d(CGCGAATTCGCG)$_2$, in solution including counterions and water. This simulation produced an early theoretical prediction of fractional occupation of counterions in the grooves of DNA and particularly in the minor groove "spine of hydration," a result that has received considerable

subsequent attention from crystallography[17] and NMR spectroscopy studies[18] as well as more extensive MD analysis.[19] This MD protocol has been applied to the study of the dynamical structures of a number of DNA and RNA oligonucleotides.[1,2] The cases have included the dynamical structures of A-tract and B′ form DNAs,[10,20] as well as A_4T_4 and T_4A_4 motifs with periodic helix phasing,[21] A/B conformational preferences, and transitions.[22,23] Sprous et al.[24] characterized an AMBER MD model for an A-form structure of the Eco RI dodecamer, obtained from a simulation carried out in an 85% ethanol/water mixed solvent. Konderding et al.[25] described the use of the Cornell et al. force field with PME in restrained MD based on NMR-derived distance and torsional restraints for two previously determined DNA sequences,[26] and compared the resulting conformations to those obtained in free MD simulations. They found the agreement for some structural aspects of the NMR derived structure and the free MD to be quite good, but commented on significant differences in helical twist values.

MD studies of a crystalline unit cell of four $d(CGCGAATTCGCG)_2$ under PME boundary conditions were recently carried out here and compared directly with corresponding crystal structure data.[27] The results show ~1 Å RMSD between the MD time-averaged and crystal structure. Comparing the MD results on a B-DNA sequence in the crystal with those for the same sequence in solution provides a basis for a purely theoretical study of crystal packing effects.[28] The results show the dynamic range of motions for the MD models of DNA is significantly less in the crystal than in solution, and the RMSD between the average structures found in the crystal and solution states is 2.1 Å. Differences between the crystal MD and solution MD structures show packing effects at the 3′ end of the sequence, at which there is an interpenetration of helices, and a G–G contact in the minor groove. These MD results indicated the influence of crystal packing on structure to be local and not global.

The field of NMR spectroscopy of nucleic acids has been treated in the monograph by Wüthrich,[7] in the reviews of van Deven and Hilbers[29] and Lane,[30] and in a series of topical articles.[31,32] In contrast with globular proteins, NMR studies of DNA are hampered by a low intrinsic nonexchangeable proton density and an absence of tertiary structure that gives rise to long-range NOEs between residues that are distant in the sequence, which can effectively serve as restraints in structural refinement. Low redundancy in the NOE information and a sensitivity of the NOEs to spin diffusion, internal rotation, and anisotropic rotational motions impedes accurate DNA structure determination still further. Addressing these problems with isotopic labeling is not as readily achieved as with proteins. The determination of torsion angles in DNA from scalar coupling constants has been advanced considerably by Altona and co-workers,[33] but addresses more local than global aspects of structure. The information from NMR data is insufficient to define a complete set of conformational and helicoidal parameters of DNA, but categorization of samples into A vs B forms and specification of certain structural parameters are well appreciated.[34,35]

The crystal structure of $d(CGCGAATTCGCG)_2$ has been the subject of numerous studies over the last 20 years.[36,37] The structure lacks palindromic symmetry (an obvious packing effect), shows axis bending of ~19° and sequence-dependent deviations from B-form DNA in some backbone torsion angles and sugar puckers. A number of NMR studies of $d(CGCGAATTCGCG)_2$ have been reported.[38–43] Following initial reports to the contrary,[41] Lane et al.[43] showed that NMR results are quite consistent with a regular palindromic B form of DNA in solution for the Eco RI sequence, and this result has generally stood the test of time. The crystal structure of $d(CGCGAATTCGCG)_2$ in complex with the restriction enzyme Eco RI endonuclease shows a local deformation, the Eco RI kink, in the vicinity of the ApT step, shown by MD to be a metastable form.[44,45] A statistical analysis of structural parameters for B-DNA duplexes in solution derived from NMR data was reported by Ulyanov and James,[46] and notably these NMR structures are slightly undertwisted, 34.6°, with respect to the 36°of canonical B DNA, and discrepancies with corresponding crystal structures were described. The sensitivity of NMR internuclear distances to B-DNA conformation has been discussed further by Lefebvre et al.[47]We note as well MD calculations of NMR relaxation parameters for a flexible polypeptide by Peter et al.[48]

The focus in this article is a detailed comparison of calculated and experimentally measured NMR observables for $d(CGCGAATTCGCG)_2$ in aqueous solution and the more general question of the accuracy of current MD models of DNA oligonucleotides in solution. The questions we specifically address are (a) how well does the back-calculation of 2D NOESY volumes from MD on DNA in solution compare with observed values; (b) how well do calculated and observed dihedrals calculated from scalar coupling constants agree, what are the major discrepancies between calculated and observed values and their implications; and (c) what is the sensitivity of NMR to small changes in DNA structure such as the differences between solution and crystalline forms. While this work was in progress, Bax and co-workers[28]

reported an NMR structure on d(CGCGAATTCGCG)$_2$ in a dilute aqueous liquid crystal phase that allowed a small degree of molecular alignment with the external field. Residual dipolar couplings are observable under these conditions that were incorporated with NOEs and scalar couplings to the structure determination. A comparison between our MD dynamical structure in solution and the results on this "NMR-dipo" structure of DNA in solution is included below.

METHODS

Experiments

The DNA sample was in a buffer of 20 m*M* sodium phosphate and 100 m*M* sodium chloride at pH 7.5. NMR data were obtained from 2D NOESY experiments on d(CGCGAATTCGCG)$_2$ performed on a Varian 500 MHz spectrometer using the States Haberkorn method with the sample in ^2H_2O at 25°C. The experiments were carried out with mixing times of both 100 and 250 ms, with an equilibration delay of 1 s. The spectral width in each dimension was 6000 Hz. In data acquisition, 512 t_1 increments were acquired with 64 transients each. The data was processed with a Gaussian weighting in both dimensions before a 4096 $\times$ 4096 Fourier transform. The 2D NOESY experiments with Watergate water suppression were run with the sample in 90% ^2H_2O and 10% ^2H_2O at 5°C with a mixing time of 100 ms and an equilibration time of 1 s. Spectral widths in both dimensions were 12,000 Hz and the data was processed with Gaussian apodization in both dimension before a 4096 $\times$ 4096 Fourier transform.

Band-selective total correlated spectroscopy (TOCSY) experiments were run on a Varian 400 MHz spectrometer equipped with a Nalorac IDG400-5 probe. The sample was in ^2H_2O at 30°C. The experiments were run with a mixing time of 70 ms and an equilibration delay of 1 s. Two Gaussian-shaped 180° pulses were applied to the H3′ region during the spin echo before the TOCSY spin–lock. Spectral width was 4000 Hz in the ^{31}P dimension and 12,000 Hz in the proton dimension. One hundred twenty-eight t_1 increments were acquired with 32 transients each. The *J* scale for the coupling to ^{31}P was set to 2 and 3 in separate experiments. The data was processed with a Gaussian apodization in both dimensions before a 4096 $\times$ 1024 Fourier transform. Purged Exclusive COSY (Correlation Spectroscopy) (PECOSY) spectra were run on a Varian 400 MHz spectrometer equipped with a Nalorac IDG400-5 probe with ^{31}P decoupling during the evolution time. A recycling delay of 1.4 s was used. 512 t_1 increments were collected with 48 transients each. The data was processed with a Gaussian apodization in both dimensions before a 2048 $\times$ 2048 Fourier transform. The data was quantified using the VNMR software.

Calculations

The 5 ns MD simulations of Young et al.[16] on the d(CGCGAATTCGCG)$_2$ dodecamer including water and counterions provides an ensemble of "snapshots" of DNA solution structures from which to calculate NMR observables. However, mobile cations are not fully stabilized even at this level of sampling.[2] The MD simulation was carried out using the AMBER 5.0 suite of programs,[49] parm94 force field,[13] and employed the PME treatment of long-range forces.[14] The simulation cell was comprised of dodecamer, 22 sodium counterions for electroneutrality, and TIP3P water molecules. In view of the possibility that the motions of solvent water and mobile ions are slower to stabilize than intrinsic DNA structure, the original MD was extended for this study from 5 to 14 ns.[50] The dynamical structure of the DNA did not change appreciably between 5 and 14 ns of trajectory, but the stabilization of the dynamical properties of the solvent molecules improved considerably.[50]

The MD trajectory was run for 14 ns to make sure that the DNA model was stable and to examine the convergence of the motions of the ions. The dynamical structure of the

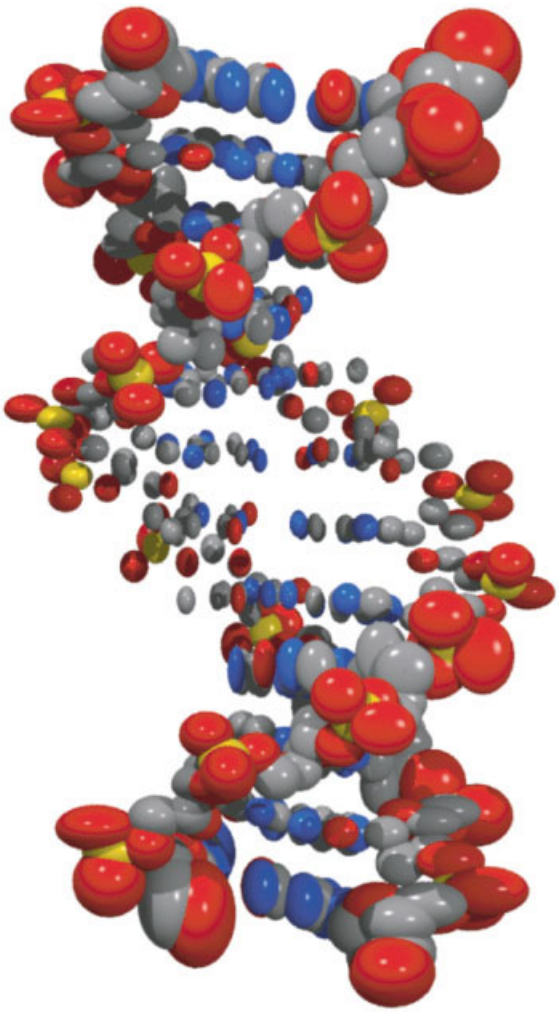

FIGURE 1 MD calculated solution structure of the d(CGCGAATTCGCG)$_2$ duplex based on 14 ns trajectory using AMBER 5.0 and the parm94 force field[50]; the dynamic range of the structures is presented in terms of thermal ellipsoids.

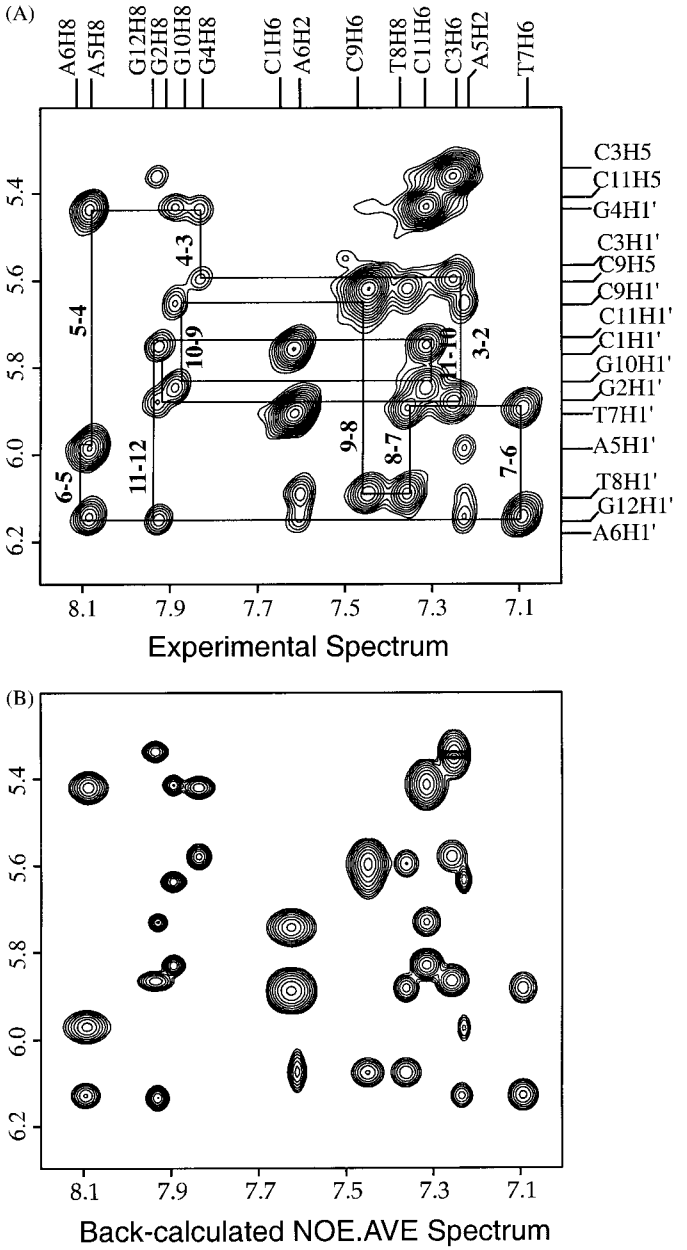

FIGURE 2 (A) Back-calculated 2D NOESY spectra (aromatic to H1′ region) based on the MD dynamical structure compared to (B) observed region 2D NOESY spectra for the d(CGCGAATTCGCG)$_2$ duplex from NMR spectroscopy.

DNA did not change significantly in the longer trajectory. The motions of the ions, however, did not fully converge in the longer trajectory but did show incipient palindromic symmetry.

The MD-calculated solution structure of d(CGCGAATTCGCG)$_2$ described using anisotropic thermal ellipsoids for the various atoms is shown in Figure 1. Young et al.[16] have previously reported the distribution of key helicoidal parameters from the MD compared with the distribution of corresponding quantities for all crystal structures of A- and B-form oligonucleotides. The MD structure was found to lie well within the expected behavior for that of B-form DNA, and sequence effects on structure and axis bending from crystal structures were generally well reproduced. The main deficiency of the AMBER MD model is a tendency toward undertwisting of the double helix, as discussed further by Beveridge and McConnell[1] and Cheathem and Young.[2]

MD on DNA produces a series of snapshots of the structure as a function of time at 2 fs intervals. For the back-calculation of NOESY volumes, MD snapshots were extracted at 5 ps intervals, ~3000 structures, in all. NOE volumes of all nonexchangeable protons were calculated via the complete relaxation matrix method[51] using XPLOR

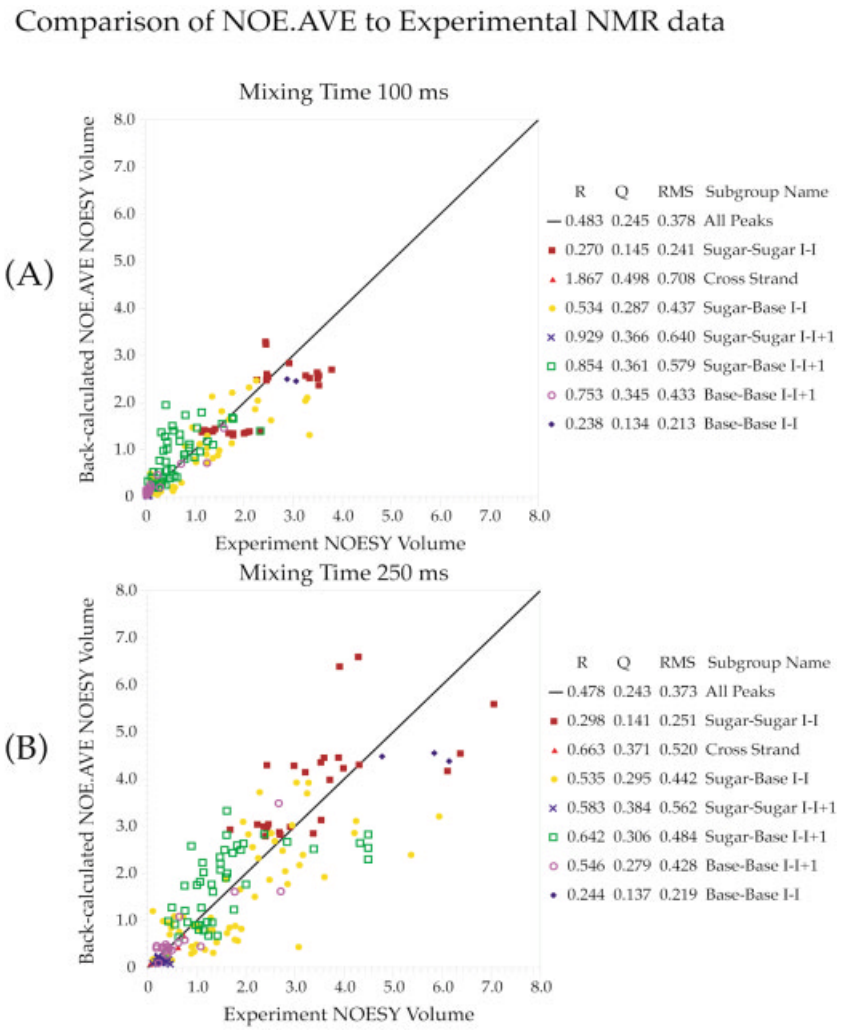

FIGURE 3 MD calculated (NOE.AVE) vs experimentally observed NOE volumes for the d(CGCGAATTCGCG)$_2$ duplex: (A) mixing time 100 ms; (B) mixing time 250 ms.

version 3.8.[52,53] Parameters used in calculating volumes were chosen to match the conditions under which the experimental data was obtained to the fullest extent possible. NOESY volumes were calculated for 100 and 250 ms mixing times assuming a magnetization leakage rate of 0.3 s^{-1}. A cutoff distance of 7.5 Å was applied in the relaxation matrix calculation. This cutoff distance was chosen after finding that several observed volumes were not present or had a greatly reduced values with a 5 Å cutoff. NOESY volumes for several structures were calculated with incremental increases in the cutoff, and the results showed essentially no change in volumes beyond a 7.5 Å cutoff. The longer distance cutoff presumably accounts better for the spin diffusion in the relaxation matrix calculation. The cutoff was observed to have a larger effect at the longer mixing time of 250 ms. A grid search to find the optimal value for anisotropic correlation time for DNA internal motions, comparing the calculated NOE volumes to experiment for several structures, resulted in a value of 5 ns, in reasonable accord with the 2–4.5 ns for B-form DNA dodecamers at 20°C reported by Eimer et al.[54] The thymine methyl protons were treated as a single spin. Calculated 2D NOESY spectra were constructed based on the well-established chemical shift assignments,[38–43] which were confirmed on the basis of the analysis of the spectra carried out in this study.

In all comparisons between MD-calculated and NMR-observed parameters, we consider two alternative procedures. One is based on calculating NMR parameters from each MD snapshot and then averaging the calculated values, and the other takes the single structure obtained as the average of the optimally aligned MD snapshots as the basis for parameter calculation. The former is subsequently referred to as the "NOE-averaged" (NOE.AVE) results, and the latter as the MD-averaged "MD.AVE" result. Actually, neither NOE.AVE nor MD.AVE provides an exact theoretical correspondence to what is measured in the NMR experiment. The NOE.AVE would be more rigorous if the MD trajectory were extended into the millisecond time frame, and the correlation functions

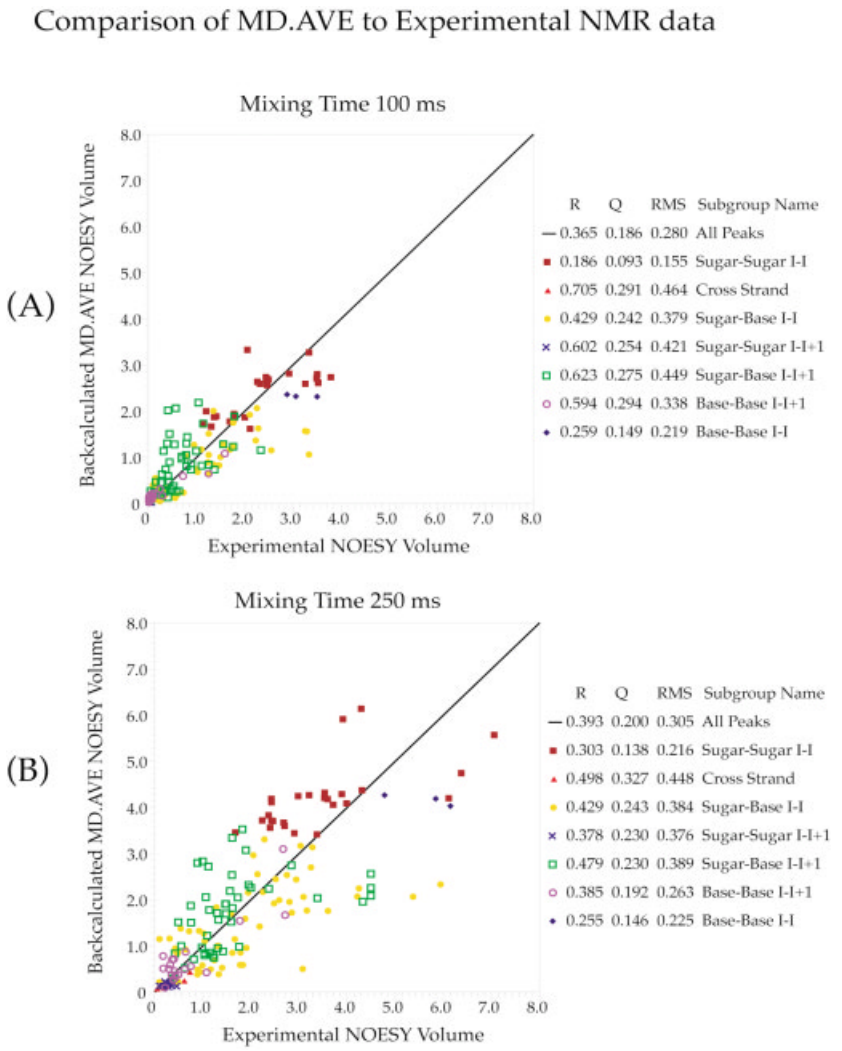

FIGURE 4 MD calculated (MD.AVE) vs experimentally observed NOE volumes for the d(CGC-GAATTCGCG)$_2$ duplex: (A) mixing time 100 ms; (B) mixing time 250 ms.

calculated from the trajectory, but current technology is, in fact, limited in this case to the nanosecond time frame for the MD trajectory. The MD.AVE assumes that the ensemble average of structure over the trajectory is representative of the ensemble per se, but this cannot be assured, and would not be the case if the MD trajectory visits structurally distinct, thermally accessible substates as opposed to oscillating within a single bound state. Since the MD average is the only well-defined single structure that can be derived from the overall trajectory, in the interest of sensitivity analysis it is of interest to report results from this alternative and how they compare with NOE.AVE results. The calculated NOE volume for each proton pair was normalized with respect to the total NOE volumes of the experimental data set.

The agreement of calculated and observed NOE volumes for each proton pair as well as the overall spectrum were calculated based on *R* factors, *Q* factors, and RMSD. The *R* factor,

$$R = \frac{\sum |I_{\text{Exp}} - I_{\text{Obs}}|}{\sum I_{\text{Exp}}} \tag{1}$$

was proposed by Gonzalez et al.[55] and others.[56–58] The use of the *R* factor for comparison contains a bias in that different values of *R* can obtained when the same absolute difference between volumes is obtained. For example an experimental value of 3 units is compared to a theoretical intensity of 1 the *R* would equal 2/3. However, if the experimental intensity is 1 unit and the theoretical intensity is 3 units the *R* equals 2. Additional measures of comparison have been proposed which take account of this problem.[59] The *Q* factor,

$$Q = \frac{\sum |I_{\text{Exp}} - I_{\text{Obs}}|}{\sum I_{\text{Exp}} + \sum I_{\text{Obs}}} \tag{2}$$

removes the dependence on whether the experimental or theoretical is larger.

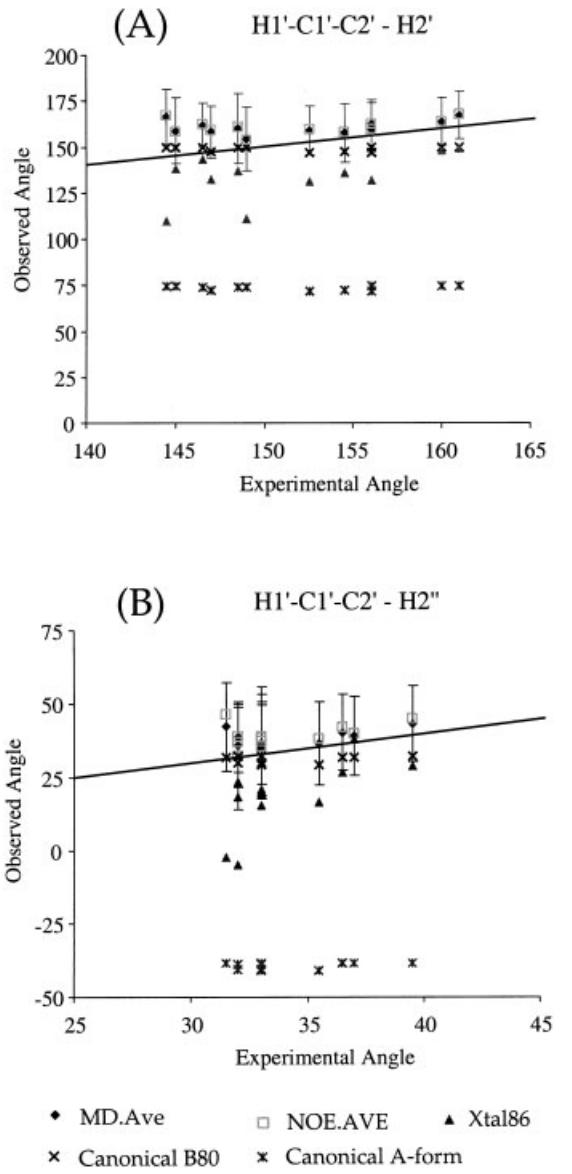

FIGURE 5 MD calculated and model values vs experimentally derived dihedral angles : (A) H1′–H2′ protons and (B) H1′–H2″.

A more statistically rigorous measure of comparison of calculated and observed volumes is the RMSD,

$$\mathrm{RMSD} = \left(\frac{\sum \{I_{\mathrm{Exp}} - I_{\mathrm{Obs}}\}^2}{\sum {I_{\mathrm{Exp}}}^2 + \sum {I_{\mathrm{Obs}}}^2} \right)^{1/2} \qquad (3)$$

Here, as well, similar measures of fit can be obtained for a model where several volumes may not be observed vs a model in which all volumes are observed but each have a slightly poorer fit. For the purposes of this study, the *R*, *Q*, and RMSD measures are all presented.

The dihedral angles (H2′—C2′—C1′— H1′), (H2″—C2′—C1′—H1′) and (P—O3′—C3′—H3′) were calculated for each snapshot and the average values were obtained. Experimental observed *J* coupling values were obtained from band-selective TOSCY and PECOSY[60] experiments. The corresponding dihedral angles were determined by use of a modified version of the Karplus equation by Altona et al.[61] as implemented in the PSEUROT program.[62] The helicoidal parameters were calculated using Curves 5.[63]

RESULTS

NOE Volumes

The region of the 2D NOESY spectrum of DNA that is most appropriate for structural comparisons and resolution of structural features corresponds to signals between aromatic and sugar H1′ protons which are observed from about 7 to 8.5 ppm and 5 to 7 ppm, respectively. The NOESY volumes back-calculated using the NOE.AVE protocol for the aromatic/sugar region for d(CGCGAATTCGCG)$_2$ are shown in Figure 2B. Here the volumes are placed on the spectrum based on experimentally observed chemical shifts assigned to the various protons, and the operational quantity is the peak volumes, not positions. The corresponding experimentally observed NOESY spectrum is shown in Figure 2A. All NOESY volumes assigned in the observed spectrum have a counterpart in the MD back-calculated spectrum. Comparing the two, one finds an overall close agreement with relatively minor differences in the calculated volumes and shapes of peaks. The MD back-calculated peak volumes tend to be more symmetric. The, presumed, undertwisting in MD model is not significantly effecting the calculated spectra.

A comparison of the calculated NOE.AVE and observed NOESY volumes as a scatter plot is shown in Figure 3, which permits a more quantitative comparison of calculated and observed values. The scatter plot for 100 ms mixing time (Figure 3B) shows *R*, *Q*, and RMS values of 0.483, 0.245, and 0.378 respectively, with an expected cone-shaped spreading of the distribution with respect to the size of NOEs. The results for 250 ms mixing time (Figure 3B) show an increased dispersion but overall *R*, *Q*, and RMS values are only marginally reduced. The data points are fairly symmetrical around the mean, indicating random and not systematic errors. Data points are labeled with respect to type of NOE: sugar–sugar, sugar–base, base–base, and cross-strand protons are differentiated with symbols. *R* and *Q* values are listed each case are listed for each subclassification. Longer range sugar–base volumes look to be on the high side but short-range values show the opposite trend. Overall, the results do not reveal any glaring discrepancies and support a conclusion that the MD model gives

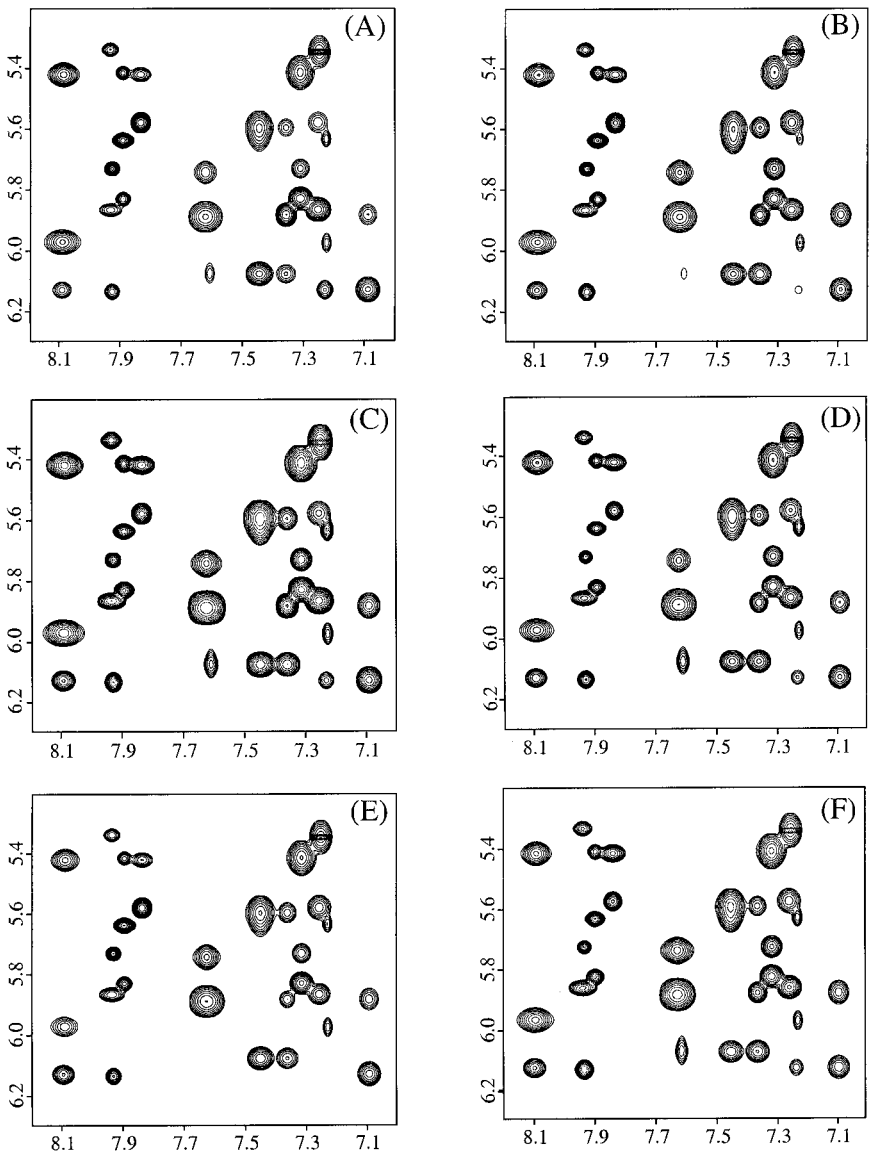

FIGURE 6 Back-calculated NOESY spectra for d(CGCGAATTCGCG)$_2$ duplex controls: (A) canonical B DNA, (B) canonical A DNA, (C) Williams et al. crystal structure (Xtal-86), (D) NMR structure using dipolar restrains (Bax), (E) Eco RI bound form of d(CGCGAATTCGCG)$_2$, and (F) Average MD structure.

back-calculated NOEs that agree well with the observed values.

A scatter plot of the NOESY volumes calculated from the ensemble average MD structure according to the MD.AVE protocol compared with the calculated ensemble average NOESY volumes is shown in Figure 4. The interpretation is similar to that of NOE.AVE, and the breakdown by type is parallel to that described above. Comparing the results calculated based on the NOE.AVE and MD.AVE protocols, Figures 3 and 4, the results are similar for the two methods of analysis, with the major discrepancies observed for the sugar–sugar protons and the cross-strand NOEs at 100 ms mixing time. Note that the difference between the prediction based on the calculated NOE.AVE and MD.AVE protocols are clearly less than the differences between calculated and observed volumes overall, indicating that the two approaches to the back-calculation of NOESY volumes show a high degree of internal correlation.

Scalar Coupling and Dihedrals

A comparison of calculated and observed dihedral angle for the sugar protons H1′–H2′ and H1′–H2″ are

Table I Comparison of Experimental NOE Volumes to the Models Listed for 100 and 250 ms Mixing Times

	100 ms			250 ms		
	R	*Q*	RMSD	*R*	*Q*	RMSD
NOE.AVE	0.483	0.245	0.378	0.478	0.243	0.373
MD.AVE	0.365	0.186	0.280	0.393	0.2	0.305
B-DNA	0.544	0.272	0.438	0.402	0.201	0.330
A-DNA	0.979	0.512	0.722	0.705	0.366	0.538
Xtal-86	0.537	0.268	0.382	0.514	0.257	0.372
EcoR I	0.598	0.297	0.445	0.603	0.301	0.431
NMR-Dipo	0.366	0.186	0.280	0.336	0.171	0.260

shown as scatter plots in Figure 5. The solid line is a regression line for the experimentally observed data. As in the case of the NOESY volumes described above, the results of calculated dihedral angles based on the NOE.AVE and MD.AVE protocols are quite similar. In this case the calculated points tend to be on the high side of the experimental data, indicating that the MD calculated results have a slight but systematic overestimation of the NOESY volumes.

DISCUSSION

The MD-calculated and NMR-observed results described in the preceding section exhibit a considerable (if not surprising) degree of agreement, no glaring discrepancies, and provisionally indicate that the MD model for the solution structure to be generally consistent with experimental NMR data. To investigate further the quality of the results and the significance of any discrepancies, back-calculation of the NOESY volumes was performed for the "control" structures, i.e., canonical A and B forms of d(CGCGAATTCGCG)$_2$,[64] the high resolution by Williams and co-workers,[37] the NMR structure solved with residual dipolar coupling by Bax and co-workers, the Eco RI protein bound formed featuring the Eco RI kink, and the average MD structure obtained by averaging the MD is snap shots in Cartesian space (MD.AVE). Comparison of the calculated NOE volumes for the controls in Figure 6A–E with the those from the MD model and experiment. Figure 2 shows that the MD gives a closer agreement with observed spectrum than any of the controls, providing further support for the accuracy of the MD model of d(CGCGAATTCGCG)$_2$ in solution.

MD-calculated dihedral angles and those obtained from the control structures of canonical A-form DNA, canonical B-form DNA, and crystal form of d(CGCGAATTCGCG)$_2$ is compared with the experimental dihedral angles derived from scalar couplings in Figure 5. Here the MD back-calculated dihedral angles also show better agreement with experiment than any of the controls. In contrast with the NOE results, the experimental MD-calculated dihedral angles are clearly closer to those from the canonical B form than those from the crystal structures. The control results for the canonical A form are well differentiated from the MD calculated and NMR observed values.

A summary of the comparisons with respect to the *R* factor is provided in Table I. The MD-based NOE.AVE shows agreement with the observed NOEs commensurate with that of the NMR Dipo structure. The crystal structures give a marginally better agreement with experiment than the canonical B80 form despite their lack of palindromic symmetry. The Eco RI and of course the canonical A form exhibit *R* values differentiated from those of the MD models and crystal B-form results. The NMR Dipo data was used by Bax et al.[8] to build an NMR structure for DNA in solution. Back-calculation of 2D NOESY spectra for this structure is shown in Figure 6D. The results correspond closely to those presented in Figure 2 for the MD model, and indicates the NMR Dipo and MD models to be in close accord on this measure.

Another means of assessing the quality of the fit is by scatter plots of the predicted data versus the experimental. The contour maps offer a qualitative comparison and the scatter plots a quantitative comparison. Figures 3 and 4 show the comparisons of the predicted NOE data with the experimental. The supplementary material contains scatter plots of the back-calculated vs the experimental data.

A comparison of the helicoidal parameters for the NOE.AVE and the NMR Dipo structure is shown in Figure 7. Statistical uncertainties are indicated on the MD plot; corresponding values for the NMR results are not available. Both the MD and NMR dipo models show evidence of 5′ and 3′ end effects, so analysis is focused on the inner 10 base pairs of the sequence. For tilt, roll, shift, rise, buckle, propeller, opening angle, and base-pair stagger, the two models agree,

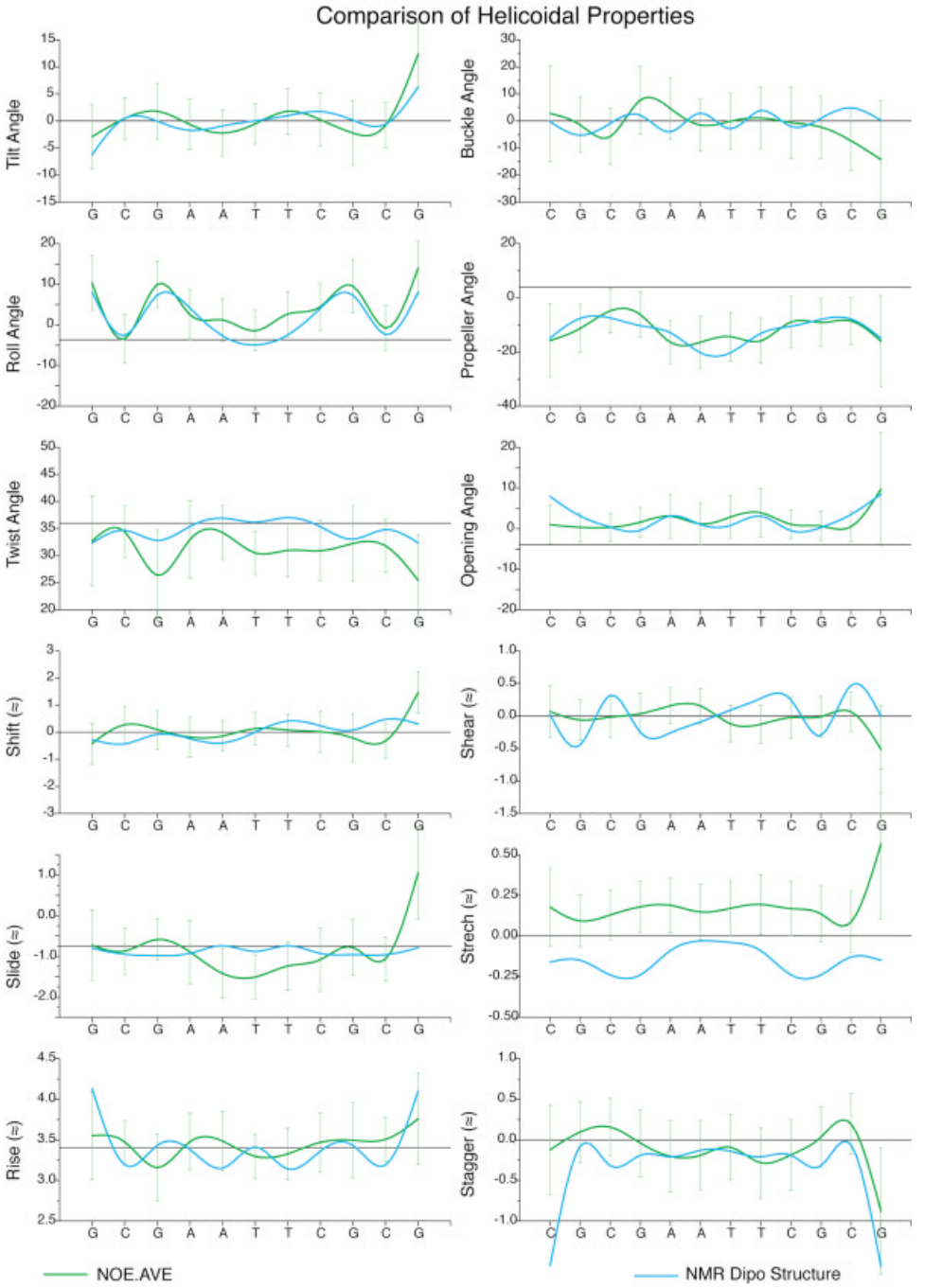

FIGURE 7 A comparison of the helicoidal parameters between NOE.AVE structure and Bax et al. NMR-dipo structure.

within the statistical uncertainty, with the MD. Base-pair slide and shear are close but show slight significant differences, with the behavior of shear being contravariant in the two cases. The magnitude of these effects as with stretch, which shows the largest significant difference, is not large enough to cause substantial differences in the structures. The tendency toward undertwisting in the MD model, mentioned earlier, is clear in this comparison as well. This point obviously needs further attention, but it should be noted that the average behavior of certain base-pair steps in the crystal structure data base show undertwisted values as well.

The determination of DNA structure in solution based on NMR spectra has remained problematic. In particular, the energy refinement of the structure on

the basis of NMR parameters has been criticized due to the use of oversimplified DNA potential functions and a lack of consideration of solvent, leading to results of uncertain accuracy. For the back-calculated NOESY volume from MD trajectory, the ensemble of structures giving rise to the spectrum is known. This suggests using this is a test of existing NMR refinement protocols and as a basis for developing improved refinement protocols, using the theoretical spectrum to provide for a test case in which the "answer" is known. This project will be the subject of a subsequent article.

SUMMARY AND CONCLUSIONS

A detailed comparison between MD-calculated and NMR-observed indices of the dynamical structure of DNA in solution has been carried out. The B-form DNA sequence $d(CGCGAATTCGCG)_2$, for which extensive crystal structure data has been reported and an MD model for the structure in solution has been derived from a 14 ns simulation based on the parm94 force field. New measurements of the NMR spectrum for this DNA were obtained in order to make the most detailed possible comparison between calculated and observed parameters. Observable 2D NOESY volumes and dihedral angles were back-calculated from the MD trajectory and compared with corresponding NMR measurements. The results indicate that the MD model is generally in good agreement with the NMR data, indicating the MD model for the solution structure is largely accurate. Further assessment using a sensitivity analysis based on control structures from crystallography and fiber diffraction shows that the MD model exhibits closer accord with experiment than calculations based on the $d(CGCGAATTCGCG)_2$ crystal structure or canonical A and B forms of the sequence. The NMR parameters are not particularly sensitive to one supposed deficiency in the MD model, a tendency toward undertwisting of the double helix. The agreement between the back-calculated NOESY volumes and experiment and from the recent structure obtained from a refinement using dipolar couplings turn out to be commensurate.

Our implementation and analysis of the MD of DNA over the last 15 years has been significantly enhanced by the scientific accomplishments, personal generosities, and many helpful discussions with our friend and colleague Peter A. Kollman. We are pleased to acknowledge the use of the AMBER suite of programs and the extent to which interactions with Peter on his many visits to our program at Wesleyan has motivated and enhanced of our own studies and scientific thought about DNA dynamics. This research was supported by NIH grant 37909 to DLB and NIH grant 65871 to PHB. KJM and MAY were recipients of NIH graduate traineeships under the Molecular Biophysics Training Program, Grant 08271.

REFERENCES

1. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182–196.
2. Cheatham, T. E., 3rd; Young, M. A. Biopolymers 2001, 56, 232–256.
3. MacKerell, J.; A. D.; Banavali, N. Comput Chem 2000, 21, 105–120.
4. Tereshko, V.; Subirana, J. A. Acta Crystallogr, Sect. D: BiolCrystallogr 1999, D55, 810–819.
5. Mack, D. R.; Chiu, T. K.; Dickerson, R. E. J Mol Biol 2001, 312, 1037–1049.
6. Berman, H. M. Biopolymers 1997, 44, 23–44.
7. Wuthrich, K. Acta Crystal Sect D Biol Crystal 1995, 51, 249–270.
8. Tjandra, N.; Tate, S.-I.; Ono, A.; Kainosho, M.; Bax, A. J Am Chem Soc 2000, 122, 6190–6200.
9. Cheatham, T. E., III; Kollman, P. A. J Am Chem Soc 1997, 119, 4805–4825.
10. Young, M. A.; Beveridge, D. L. J Mol Biol 1998, 281, 675–687.
11. Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. J Am Chem Soc 1999, 121, 5981–5991.
12. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. University of California, San Francisco, 1997.
13. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179–5197.
14. Darden, T. A.; York, D. M.; Pedersen, L. G. J Chem Phys 1993 98, 10089–10092.
15. Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. J Am Chem Soc 1995, 117, 4193–4194.
16. Young, M. A.; Ravishanker, G.; Beveridge, D. L. Biophys J 1997, 73, 2313–2336.
17. Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. Biochemistry 1998, 37, 8341–8355.
18. Hud, N. V.; Feigon, J. J Am Chem Soc 1997, 119, 5756–5757.
19. McConnell, K. J.; Beveridge, D. L. J Mol Biol 2000, 304, 803–820.
20. McConnell, K. J.; Beveridge, D. L. J Mol Biol 2001, 314, 23–40.
21. Sprous, D.; Young, M. A.; Beveridge, D. L. J Mol Biol 1999, 285, 1623–1632.
22. Cheatham, T. E., III; Kollman, P. A. J Mol Biol 1996, 259, 434–444.

23. Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. J Am Chem Soc 1998, 120, 10629–10633.
24. Sprous, D.; Young, M. A.; Beveridge, D. L. J Phys Chem 1998, 102, 4658–4667.
25. Konerding, D. E.; Cheatham, T. E., III; Kollman, P. A.; James, T. L. J Biomol NMR 1999,13, 119–131.
26. Weisz, K.; Shafer, R. H.; Egan, W.; James, T. L. Biochemistry 1994, 33, 354–366.
27. McConnell, K. J. In Molecular Biology: The Molecular Dynamics of DNA and Free Energy Studies of Protein DNA Interpretations, Wesleyan University: Middletown, CT, 2002; pp 300, PhD thesis.
28. McConnell, K. J.; Young, M. A.; Beveridge, D. L. Submitted, 2002.
29. Van de Ven, J. M.; Hilbers, C. W. Eur J Biochem 1988, 178, 1–38.
30. Lane, A. N. Biochim Biophys Acta 1990, 1049, 189–204.
31. Han, X. G.; Gao, X. L. Curr Med Chem 2001, 8, 551–581.
32. Patel, D. J.; Suri, A. K.; Jiang, F.; Jiang, L. C.; Fan, P.; Kumar, R. A.; Nonin, S. J Mol Biol 1997, 272, 645–664.
33. Rinkel, L. J.; Altona, C. A. J Biomol Struct Dynam 1987, 4, 621–649.
34. Metzler, W. J.; Wang, C.; Kitchen, D. B.; Levy, R. M.; Pardi, A. J Mol Biol 214, 711–736.
35. Meadows, R. P.; Kaluarachchi, K.; Post, C. B.; Gorenstein, D. G. Bull Magn Reson 1992,13, 22–48.
36. Wing, R. M.; Drew, H. R.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, I.; Dickerson, R. E. Nature 1980, 287, 755–758.
37. Howerton, S. B.; Sines, C. C.; VanDerveer, D.; Williams, L. D. J Am Chem Soc 2001, in press.
38. Hare, D. R.; Wemmer, D. E.; Chou, S. H.; Drobny, G.; Ried, B. R. J Mol Biol 1983, 171, 319–336.
39. Broido, M. S.; James, T. L.; Zon, G.; Keepers, J. W. Eur J Biochem 1985, 150, 117–128; issn: 0014-2956.
40. Patel, D. J.; Shapiro, L.; Hare, D. J Biol Chem 1986, 261, 1223–1229.
41. Nerdal, W.; Hare, D. R.; Reid, B. R. Biochemistry 1989, 28, 10008–10021.
42. Denisov, A. Y.; Zamaratski, E. V.; Maltseva, T. V.; Sandstrom, A.; Bekiroglu, S.; Altmann, K. H.; Egli, M.; Chattopadhyaya, J. J Biomol Struct Dynam 1998, 16, 547–568.
43. Lane, A.; Jenkins, T. C.; Brown, T.; Neidle, S. Biochemistry 1991, 30, 1372–1385.
44. Kumar, S.; Duan, Y.; Kollman, P. A.; Rosenberg, J. M. J Biomol Struct Dynam 1994, 12, 487–525.
45. Young, M. A.; Nirmala, R.; Srinivasan, J.; McConnell, K. J.; Ravishanker, G.; Beveridge, D. L.; Berman, H. M. In Structural Biology: The State of the Art. Proceedings of the 8th Conversation; Sarma, R. H., Ed.; Adenine Press: Albany, NY, 1994; pp 197–214.
46. Ulyanov, N. B.; James, T. L. Appl Magn Reson 1994, 7, 21–42.
47. Lefebvre, A.; Fermandjian, S.; Hartmann, B. Nucleic Acids Res 1997, 25, 3855-3862.
48. Peter, C.; Daura, X.; van Gunsteren, W. F. J Biomol NMR 2001, 20, 297–310.
49. Srinivasan, J.; Case, D. Private communication, 1997.
50. Young, M. A.; Jayaram, B.; Beveridge, D. L. J Phys Chem B 1998, 102, 7666–7669.
51. Boelens, R.; Koning, T. M. G.; van der Marel, G. A.; Van Bloom, J. H.; Kaptein, R. J. Magn Res 1989, 82, 290.
52. Brunger, A. T. X-PLOR, Yale University, New Haven, CT, 1996.
53. White, S. A.; Nilges, M.; Huang, A.; Brunger, A. T.; Moore, P. B. Biochemistry 1992, 31, 1610–1621.
54. Eimer, W.; Williamson, J. R.; Boxer, S. G.; Pecora, R. Biochemistry 1990, 29, 799–811.
55. Gonzalez, C.; Rullmann, J. A. C.; Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. J Magn Res 1991, 91, 659–664.
56. Baleja, J. D.; Pon, R. T.; Sykes, B. D. Biochemistry 1990, 29, 4828–4839.
57. Gochin, M.; James, T. L. Biochemistry 1990, 29, 11172–11180.
58. Nikonowicz, E. P.; Meadows, R. P.; Fagan, P.; Gorenstein, D. G. Biochemistry 1991,30, 1323–1334.
59. Withka, J. M.; Srinivasan, J.; Bolton, P. H. J Mag Res 1992, 98, 611.
60. Bax, A.; Lerner, L. J Magn Res1988, 79, 429–438.
61. Haasnoot, C. A. G.; De Leeuw, F. A. A. M.; Altona, C. Tetrahedron 1980, 36, 2783–2792.
62. DeLeeuw, F. A. A. M.; Altona, C. J Comp Chem 1983, 4, 428–437.
63. Lavery, R.; Sklenar, H. J Biomo Struct Dynam 1988, 6, 63–91.
64. Arnott, S.; Hukins, D. W. L. Biochem Biophys Res Comm 1972, 47, 1504–1510.