# [5] Structure Determination and Analysis of Local Bending in an A-Tract DNA Duplex: Comparison of Results from Crystallography, Nuclear Magnetic Resonance, and Molecular Dynamics Simulation on d(CGCAAAAATGCG)

*By* MATTHEW A. YOUNG, JAYASHAREE SRINIVASAN, IGOR GOLJER, SURAT KUMAR, DAVID L. BEVERIDGE, and PHILIP H. BOLTON

## Introduction

The determination of the molecular structure of a DNA oligonucleotide can be undertaken by at least three different methodologies: X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and molecular dynamics (MD) simulation. X-ray crystallography can provide a detailed view of the structure, but the results are subject to crystal packing effects that may influence, to some degree, the results.[1] NMR experiments measure parameters for the molecule in the solution state, and the subsequent structure determination involves finding conformations consistent with the observed nuclear Overhauser effect (NOE) cross-peaks and with scalar coupling data. However, the limited distance range monitored by the NOEs mitigates against accurate determination of the global properties of a DNA helix,[2] including axis bending.[3] Theoretical structure determination via MD simulation is a third option, which for models of oligonucleotides in solution is an extensive undertaking on supercomputers. MD results are sensitive to assumptions about the underlying force field and in the simulation protocols employed.[4,5] In this chapter, we consider a system of considerable current interest that has been studied in detail by all three methods, describe the results, and then compare the resulting structures with respect to indices of helix morphology and fine structure. The objective is to determine the extent of similarities and differences in the various structure determinations and to assess critically the implications of the results.

The system under consideration is the duplex of sequence d(CGCAAAAATGCG), a sequence containing five consecutive A : T base

[1] R. E. Dickerson, D. S. Goodsell, M. L. Kopka, and P. E. Pjura, *J. Biolmol. Struct. Dyn.* **5,** 557 (1987).
[2] D. E. Wemmer, *Curr. Opin. Biol.* **1,** 452 (1991).
[3] K. Kaluarachi, R. P. Meadows, and D. G. Gorenstein, *Biochemistry* **30,** 8785 (1991).
[4] D. L. Beveridge and G. Ravishanker, *Curr. Opin. Struct. Biol.* **4,** 246 (1994).
[5] R. J. Loncharich and B. R. Brooks, *Proteins* **6,** 32 (1989).

pairs that form an "A-tract," which is a structural feature that has been implicated in DNA bending.[6] An issue of interest in A-tract sequences is the exact nature of local axis bending,[7] that is, whether it conforms to a "wedge model," [8] "junction model," [9] or other pattern.[10,11] Oligonucleotide duplexes with A-tracts exhibit a large retardation effect in gel migration studies, which is interpreted to be a consequence of axis bending.[6] These observations have important implications in various cases of the structural biochemistry of DNA, such as the solenoidal wrapping of DNA around histone proteins in the nucleosome.[12] Axis bending is an interesting feature of a number of recent structures of protein-DNA complexes[13] and is expected to be important to the understanding of molecular recognition events in gene regulation.

## The Crystal Structure(s) (and Analysis)

The 2.5 Å resolution crystal structure of the duplex formed by d(CGCAAAAATGCG) with its complementary strand has been reported by DiGabriele *et al.*[14] The duplex is observed to adopt two distinct orientations in the crystal referred to as the "up" and "down" forms. The two duplexes are related by a 180° rotation about the pseudo-dyad axis. Molecular structures for the "up" and "down" forms are shown in Fig. 1. Analysis of the results finds that both exhibit axis bending to the extent of ~20°. These results permit an assessment of whether the observed bend in the DNA is intrinsic to the sequence or arises as a consequence of crystal packing.

If bending is an intrinsic property of the DNA sequence, then the bends observed for the up and down helices should be the same when their sequences are identically aligned. If crystal packing forces are important, then the two structures should be the same when oriented the same way

[6] J. C. Marini, S. D. Levene, D. M. Crothers, and P. T. Englund, *Proc. Natl. Acad. Sci. U.S.A.* **79,** 7664 (1982).

[7] M. Sundaralingam and Y. C. Sekarudu, *in* "Sequence Directed DNA Bending and Curvature. An Overview" (W. K. Olson, M. H. Sarma, R. H. Sarma, and M. Sundaralingam, eds.). p. 9. Adenine Press, New York, 1988.

[8] L. E. Ulanovsky and E. N. Trifonov, *Nature* **326,** 720 (1987).

[9] H. S. Koo, J. Drak, J. A. Rice, and D. M. Crothers, *Biochemistry* **29,** 4227 (1990).

[10] M. A. Young, G. Raviahshanker, D. L. Beveridge, and H. M. Berman, *Biophys. J.* **68,** 2454 (1995).

[11] D. S. Goodsell and R. E. Dickerson, *Nucleic Acids Res.* **22,** 5497 (1994).

[12] W. Saenger, "Principles of Nucleic Acid Structure." Springer-Verlag, New York, 1984.

[13] T. A. Steitz, *Q. Rev. Biophys.* **23,** 205 (1990).

[14] A. D. DiGabriele, M. R. Sanderson, and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 1816 (1989).

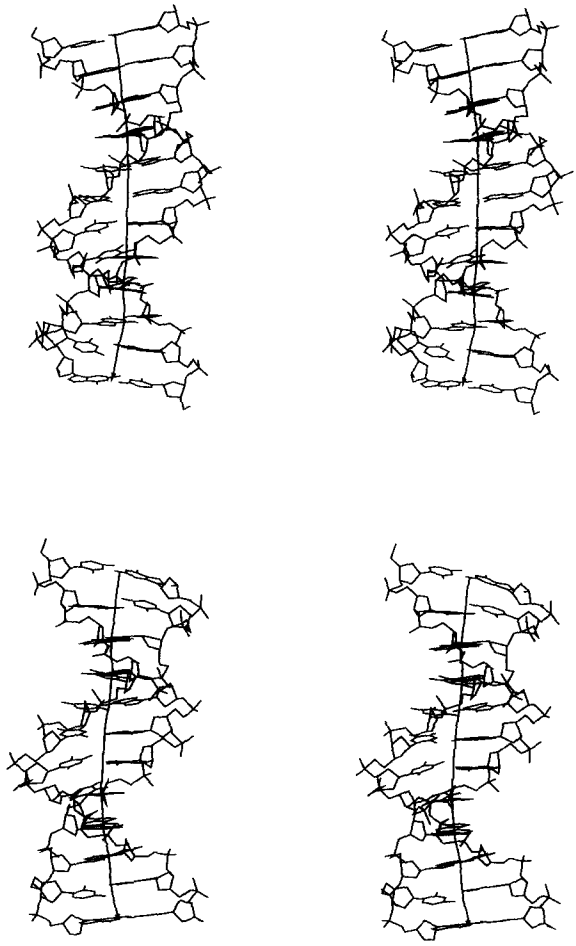


FIG. 1. Stereo views of the "up" (top) and "down" (lower) crystal structures. Superimposed on the stick diagrams is the global helical axis as calculated by CURVES.

in the crystal. When the duplexes are aligned, the axis bending of the "up" and "down" forms is opposite in sense: the up form bends toward the minor groove, whereas the down form bends toward the major groove. Thus, the orientation of the helix in the crystal and not its sequence dictates the direction of the bend. The sequence may, of course, confer a propensity to bend, but the manner in which this is realized obviously depends on the crystal context. The direction of the observed bend in the crystalline form is moreover 90° from the bend direction inferred from various solution

measurements by Crothers and co-workers who suggested that an A-tract bends toward the minor groove at the center of the A-tract.[15]

The A-tracts in both crystal structures are relatively straight compared to that of the flanking sequences and exhibit a high degree of propeller twist, averaging ~20° versus ~10° in the flanking sequence. It was noted that this brings the AN6 and TO4 into proximity so that bifurcated hydrogen bonds may be formed, conferring an extra rigidity that could explain the tendency of the A-tracts to be straight. However, recent theoretical analysis of this idea has indicated that the putative bifurcated hydrogen bonding does not necessarily confer any preferential stability.[16]

*Analysis of Fine Structure*

The "up" and "down" crystal structures of d(CGCAAAAATGCG) were analyzed in detail with respect to conformational and helicoidal parameters using the program "Dials and Windows," [17] based upon the "Curves" procedure devised by Lavery and Sklenar[18] using a global axis reference frame. Full details of this analysis are available by anonymous FTP,[19] and we provide here only an overview. The observed conformational parameters fall readily within sterically allowed regions for right-handed DNA structures.[12] The conformational parameters for the "up" and "down" forms are generally within a quadrant of one another, with the largest differences occurring in values for and in the regions at the junctions between A-tracts and flanking sequences. Note that the crystallographic refinements were performed with restraints on the sugar puckering set at C2′ endo values.

In the helicoidal parameter set, the axis base pair parameter "X-displacement" (XDP) for both up and down forms is extremely close to canonical B values. There is a slight nonzero value in base pair inclination (INC) with respect to the helix axis into a region between values for canonical A and B forms of DNA. Within the A-tract the values are closer to canonical B than in the flanking sequences. The intra–base pair parameter set shows some particularly large values of base pair buckle (BKL) and, as discussed in the crystal structure paper, propeller (PRP). Values of PRP within the A-tract are uniformly large and negative;

[15] S. S. Zinkel and D. M. Crothers, *J. Mol. Biol.* **219,** 201 (1991).
[16] V. Fritsch and E. Westhof, *J. Am. Chem. Soc.* **113,** 8271 (1991).
[17] G. Ravishanker, S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar, *J. Biomol. Struct. Dyn.* **6,** 669 (1989).
[18] R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6,** 63 (1988).
[19] ftp://prophet.chem.wesleyan.edu/

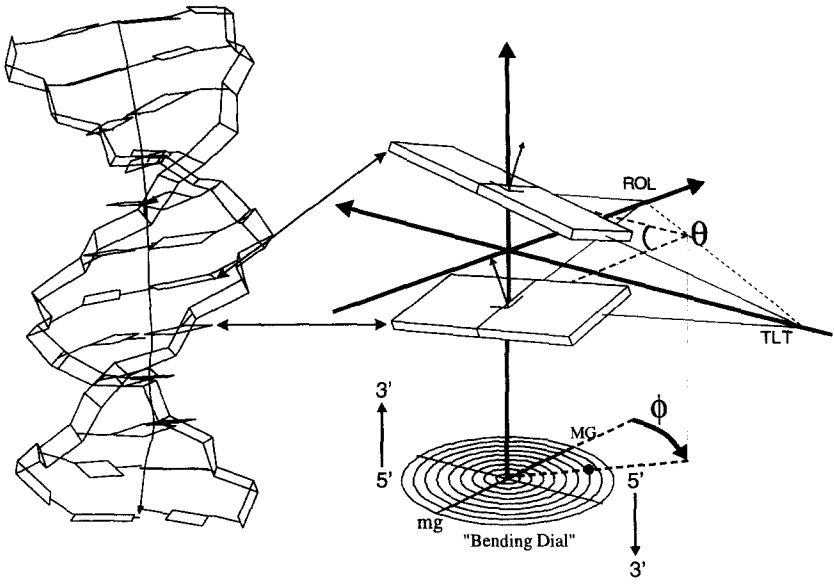



FIG. 2. Schematic definition of the axis bending direction $\phi$ and bending magnitude $\theta$ in terms of the helicoidal parameters roll ($\rho$) and tilt ($\tau$), defined with respect to the major groove (MG) and minor groove (mg) of the DNA double helix. Positive roll (bending compressing the major groove) is plotted in the northern hemisphere, and negative roll (bending into the minor groove) is plotted on the southern hemisphere of the dial.

however, some values of PRP are in the same range, whereas others are smaller.

### *Analysis of Helix Morphology*

Of the interbase pair parameters, interesting variations are seen in base pair roll and tilt that form a natural basis for the discussion of axis bending. In a previous series of papers,[10,20] we introduced the idea of the analysis of local axis bending using a graphical construct in which the bending magnitude and direction are projected onto a polar plot called a "bending dial," defined in Fig. 2. The local bending analysis of the "up" and the "down" crystal forms of d(CGCAAAAATGCG) duplex are shown in Fig.

[20] M. A. Young, R. Nirmala, J. Srinivasan, K. J. McConnell, G. Ravishanker, D. L. Beveridge, and H. M. Berman, *in* "Structural Biology: The State of the Art. Proc. Eighth Conversation in Biomolecular Structure and Dynamics." (R. H. Sarma, ed.) Vol. 2, p. 197. Adenine Press, Albany, NY, 1994.

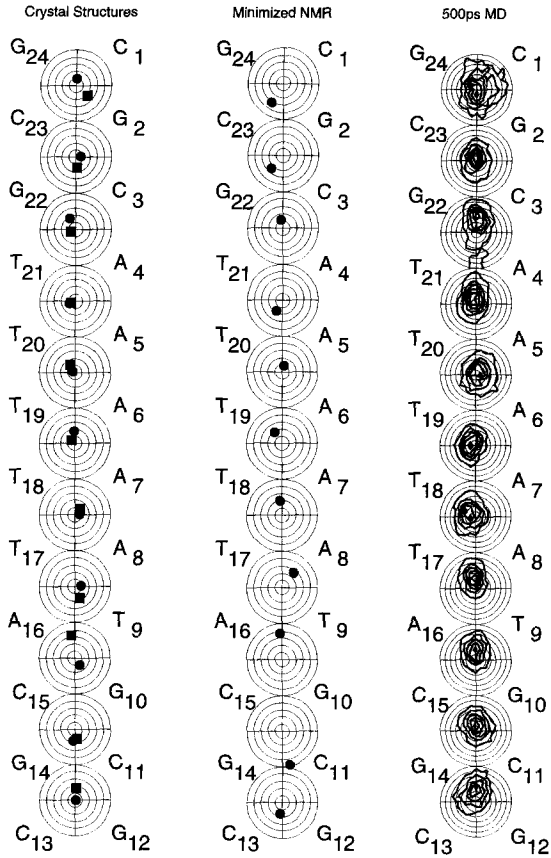



FIG. 3. Bending calculated at each step in the three structures determinations for the sequence d(CGCAAAAATGCG) presented as bending dials. Each concentric ring indicates a 5° deflection of the helical axis. (*Left*) The "up" and "down" structures determined by crystallography, presented as circles and squares, respectively. (*Middle*) Bending calculated from the NMR structure. (*Right*) Bending calculated in the range of structures visited during the 500 ps unrestrained MD trajectory. The sum of all MD structures contributes to the statistical contour levels plotted in this last column.

3. The results indicate that within the A-tract the helix axis is relatively straight, consistent with the other A-tract structures.[10,21,22] Bending at base pair steps within the flanking sequences is larger in magnitude. By this analysis the steps in the "up" form are relatively straight, and, thus, the

[21] A. DiGabriele and T. A. Steitz, *J. Mol. Biol.* **231,** 1024 (1993).
[22] D. S. Goodsell, M. K. Greskowiak, and R. E. Dickerson, *J. Mol. Biol.* **239,** 79 (1994).

overall bend results from the additive effect of concerted helical deflections. The "down" form shows a particular axis displacement toward the major groove at the T9-G10 step, a junction between the A-tract and the flanking sequence.

The "up" form shows a smaller displacement toward the minor groove at this point. The morphology of an oligonucleotide duplex can be investigated further by examination of the major and minor groove widths as a function of sequence. Various techniques have been proposed for the definition of groove widths, such as distance measures between opposing phosphates or spline fits to the sugar phosphate backbone. We find that a display of the "unrolled helix" (UH) is informative for most purposes. The UH representation of the two crystal structure of d(CGCAAAAATGCG) as compared with the canonical B-form DNA is shown in Fig. 4 (see color insert). Within the A-tract there is obviously a pronounced narrowing of the minor groove. This is accompanied by a slight widening of the major groove. In the flanking sequences, the minor groove is wider than the canonical, due presumably to crystal packing effects.[23]

## The NMR Structure

The NMR experiments were carried out on DNA samples of the sequence 5′-d($C_1G_2C_3A_4A_5A_6A_7A_8T_9G_{10}C_{11}G_{12}$)-3′ paired with the complementary strand 5′-d($C_{13}G_{14}C_{15}A_{16}T_{17}T_{18}T_{19}T_{20}T_{21}G_{22}C_{23}G_{24}$)-3′. The DNA single strands were obtained from Pharmacia and the purity of the samples checked both by NMR and HPLC. The stoichiometry of the two strands was adjusted to 1 : 1 by monitoring the one-dimensional proton NMR spectrum of the sample. The DNA duplex was annealed by heating to 90° and then allowed to cool to room temperature. The samples were at a concentration of 300 $OD_{260}$, and the buffer contained 100 m*M* NaCl, 50 m*M* phosphate buffer, and 0.5 m*M* EDTA at pH 7.0. All of the NMR data was processed with Varian VNMR software except as noted. Paramagnetic ions were rigorously excluded from the samples as the presence of such can lead to incorrect structures. The NOESY, ROESY, TOCSY, DQCOSY, PECOSY, and other NMR experiments were carried out in the usual fashion,[24–27] in both $H_2O$ and $^2H_2O$ solvents.

[23] N. Narayana, S. L. Ginell, I. M. Russu, and H. M. Berman, *Biochemistry* **30,** 4449 (1991).

[24] G. M. Clore and A. M. Gronenborn, *Crit. Rev. Biochem.* **24,** 479 (1989).

[25] R. R. Ernst, G. Bodenhausen, and A. Wokaun, "Principles of NMR in One- and Two-Dimensions." Oxford University Press, London, 1987.

[26] F. J. M. V. d. Ven and C. W. Hilbers, *Eur. J. Biochem.* **178,** 1 (1988).

[27] K. Wüthrich, "NMR of Proteins and Nucleic Acids," p. 292. John Wiley & Sons, New York, 1986.

A 200 ms NOESY in 90% $H_2O$–10% $^2H_2O$ was obtained using a Varian VXR 400 spectrometer at 30° using the States–Haberkorn method. The final pulse of the experiment was a 1-1 sequence for suppression of the water signals. The relaxation delay was 1.5 s and the acquisition time 0.34 s. The spectra were obtained with a spectral width of 10,000 Hz, 1344 × 2 complex points in $t_2$, and 256 × 2 FIDs in $t_1$, with 128 transients obtained for each increment. Linear prediction was used to calculate the first three points in $t_2$ and the first five points in $t_1$ and to extend the number of points from 281 to 1024 points in $t_1$ by iterative linear prediction. A phase shifted sine weighting was used in $t_2$ and a Gaussian function in $t_1$. The spectra were zero-filled to 2048 × 2048 real points.

A PECOSY spectrum was obtained using a Varian VXR 400 spectrometer with the sample at 30° in $^2H_2O$ using the States–Haberkorn method. The relaxation delay was 1.6 s and the acquisition time was 0.34 s. The data were collected into 1024 × 2 complex points in $t_2$ and 512 × 2 points in $t_1$, with spectral widths of 3265 Hz in each dimension, and 64 transients were acquired for each increment of the evolution time. A Lorentzian weighting was used in both dimensions in both cases, and the spectra were zero-filled to 4096 × 4096 real points.

NOESY spectrum were also obtained using a Varian Unityplus 400 spectrometer with the sample at 30° in $^2H_2O$ using the States–Haberkorn method with $^{31}P$ decoupling during the evolution time. The relaxation delay was 1.5 s and the acquisition time was 0.34 s. The data were collected into 1024 × 2 complex points in $t_2$ and 512 × 2 points in $t_1$, with spectral widths of 4000 Hz in each dimension, and 64 transients were acquired for each increment of the evolution time. A Gaussian weighting was used in both dimensions, and the spectra were zero-filled to 4096 by 4096 real points. The heteronuclear $J_{PH}$ coupling constants were estimated by comparison of the proton linewidths along $F_1$ and $F_2$.

NOESY spectra (100 and 200 ms; 600 MHz) were obtained on the sample in $^2H_2O$ at 30° using a Bruker AM spectrometer with the TPPI method. The spectral width was 6024 Hz in each dimension, with 2048 complex points in $F_1$, and 600 increments of $t_1$ and 64 transients were acquired for each increment of the evolution time. The relaxation delay was 1.5 s and the acquisition time 0.34 s. The data was processed using FELIX 2.1 and the time domain data were processed with Gaussian weighting in both dimensions and zero-filled to 2048 points in both dimensions before Fourier transformation. The 600 MHz data were used in the structure determination.

One-dimensional $^{31}P$ NMR spectra were obtained at 161.9 MHz with proton decoupling. The spectral width was 2000 Hz with 3264 complex

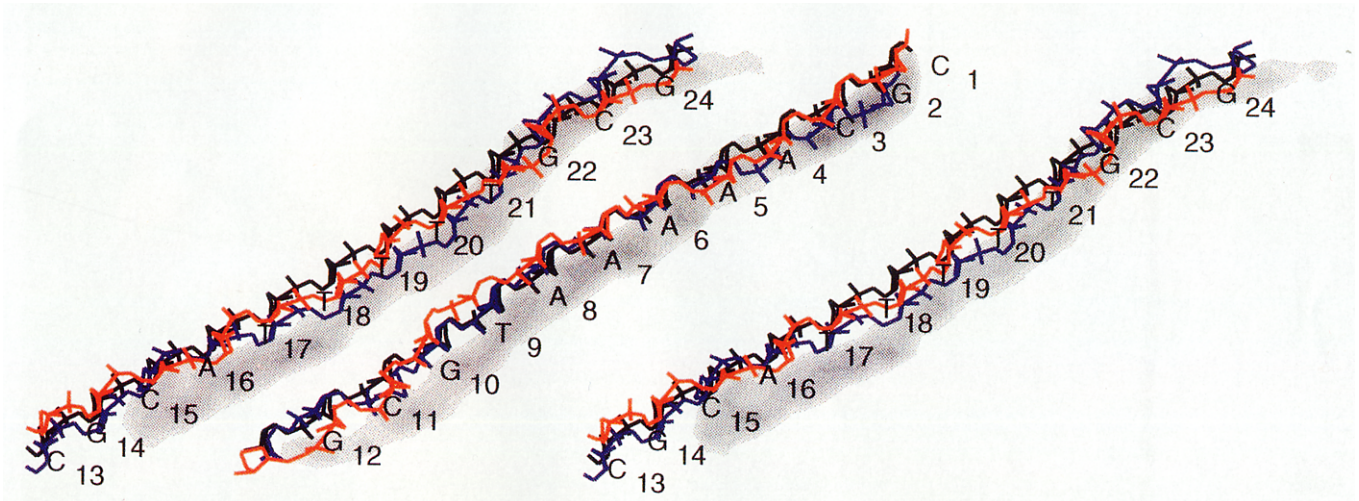



FIG. 4. "Unrolled helix" representation of the locations of the DNA backbone atoms. The plot is generated by projecting the location of the backbone atoms onto a cylinder of uniform 8.5 Å radius. Shown superimposed on this plot are the two crystallographic solutions (in red and blue), a canonical B-80 form of the DNA double helix (in black), and a statistical representation of the structures visited throughout the MD trajectory (grey shading).

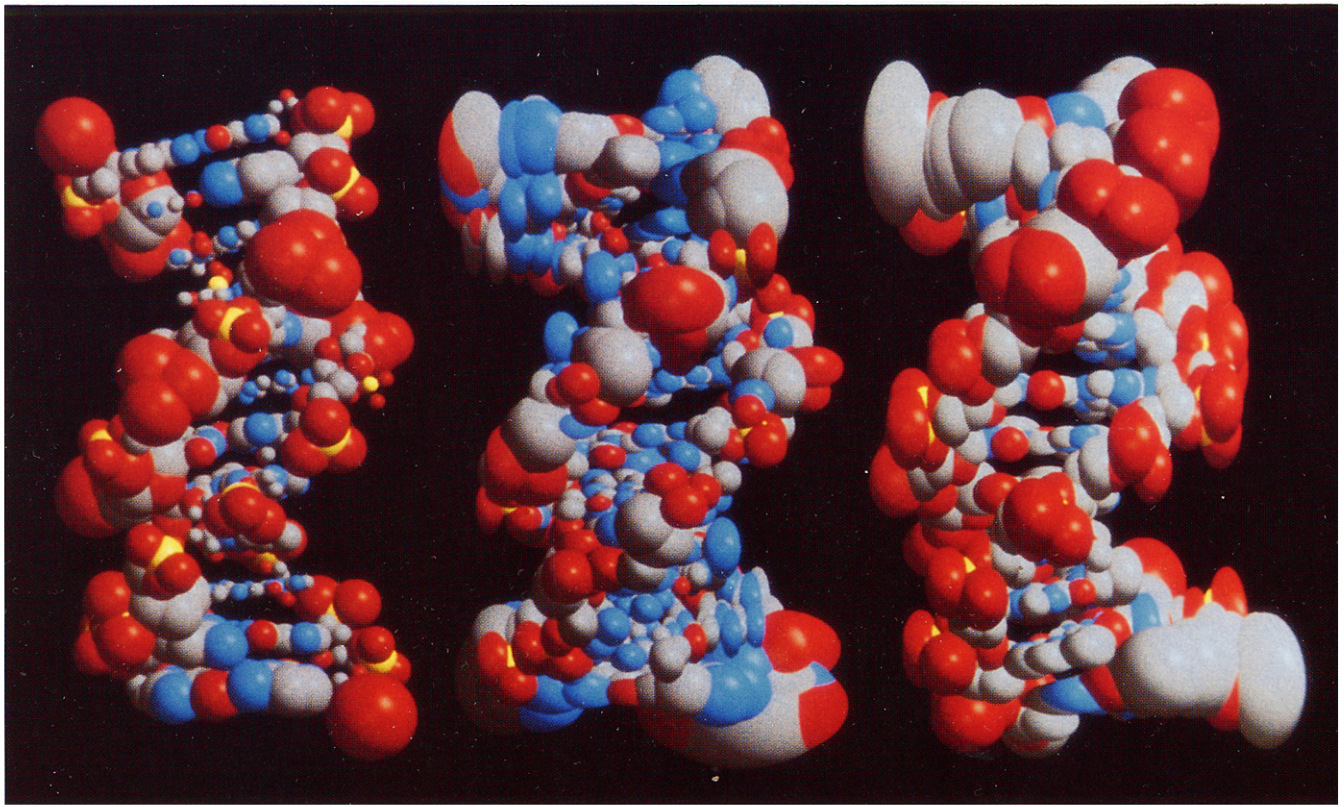


FIG. 8. The structures determined by the three methods as shown as ellipsoids. The size of the ellipsoid is one standard deviation about the average position. The "up" crystal structure is shown on the left, the NMR structure in the middle, and the MD structure on the right.

points and 2048 scans. One-dimensional spectra of the imino protons were obtained at 400 MHz using a 1-1 pulse for water suppression.

*Determination of Assignments and Coupling Constants*

The nonexchangeable proton resonances of the DNA duplex were assigned using the standard intra- and interresidue connectivities.[26,28] Both strands exhibited sequential H8/H6 to H1′, H2′, H2″, and H3′ NOE connectivities. NOESY spectra containing these NOEs are shown in Fig. 5. The imino protons were assigned via the imino–imino NOE connectivities. The AH2 protons were assigned on the basis of intra-base pair AH2-imino connectivities. The assignments of the protons are listed in Table I. The $J_{H1',H2''}$ couplings were determined from PECOSY results and the couplings are listed in Table II. All of the couplings were found to be in the general range expected for B-DNA.[29] The heteronuclear $J_{PH}$ couplings were estimated as described earlier and are listed in Table II.

*Quantitation of NOE Cross-Peak Volumes*

The volumes of the NOE cross-peaks of the 600 MHz data obtained with 100 and 200 ms mixing times were quantified using FELIX 2.1 software. For each assigned and resolved cross-peak, the volume in a standard area was determined. The spectral baseplane was flattened before the volumes were determined. Two additional NOESY data sets were obtained at 400 MHz with 1.5 s and 7 s relaxation delays and 0.34 s acquisition times. The comparison of the cross-peak volumes from the two experiments showed less than 20% variation. This showed that the volumes obtained using the shorter equilibration time on the 600 MHz spectrometer, on which only limited time was available, could be used as constraints.

*Structure Determination Procedures*

Structure refinement was performed by restrained molecular dynamics using X-PLOR 3.1.[30] Restrained molecular dynamics have been successfully

[28] I. Goljer and P. H. Bolton, "Two-Dimensional NMR Spectroscopy. Applications for Chemicals and Biochemists," p. 699. VCH, New York, 1994.

[29] M. M. W. Mooren, S. W. Wijmenga, G. A. v. d. Marel, J. H. v. Boom, and C. W. Hilbers, *Nucleic Acids Res.* **22,** 2658 (1994).

[30] A. T. Brünger, "X-PLOR (Version 3.0) Manual." The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1992.

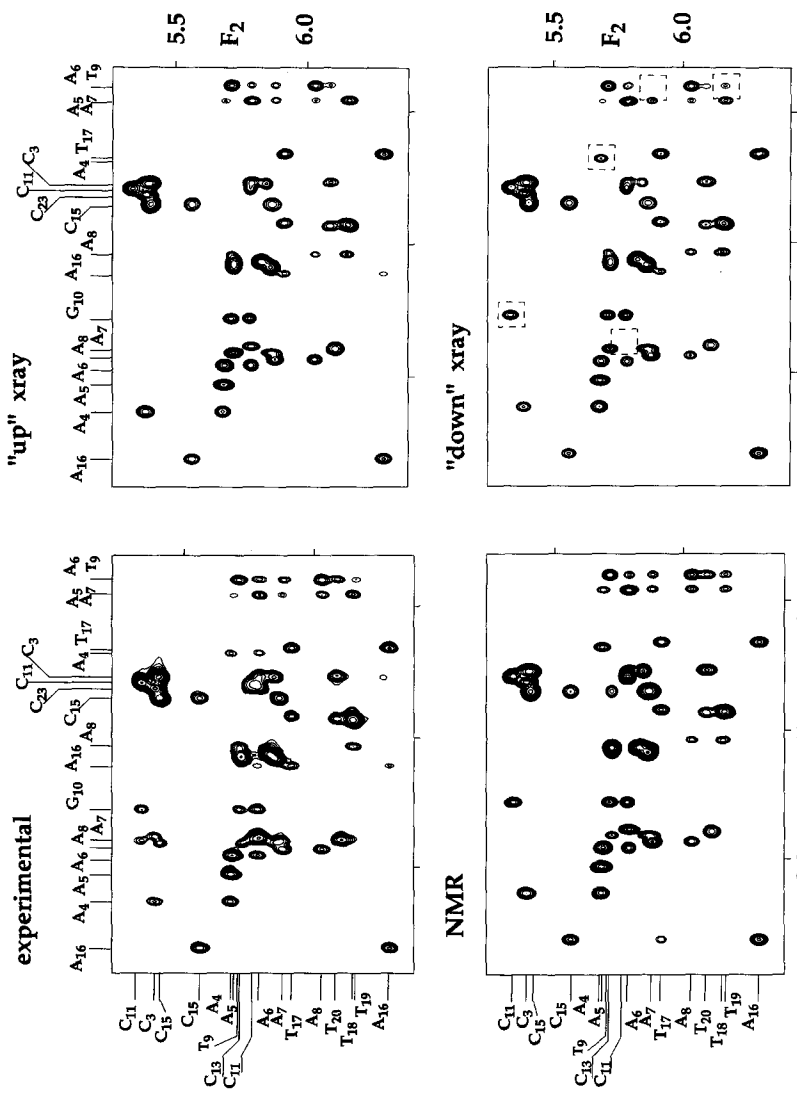



FIG. 5. The top left spectrum is the experimental two-dimensional NOESY spectrum of the DNA duplex obtained at 600 MHz with a mixing time of 200 ms. The assignments of the signals are indicated, and this region contains aromatic to H1′ and H6-H5 cross-peaks. The top right spectrum is that predicted from the "up" crystal structure, and the bottom right spectrum is that predicted from the "down" crystal structure. The bottom left spectrum is that calculated from the NMR structure.

TABLE I
CHEMICAL SHIFTS ASSIGNMENTS

| | C6H/C8H | C5H/5-Me | H1′ | H2′ | H2″ | H3′ | H4′ | C2H | Imino |
|---|---|---|---|---|---|---|---|---|---|
| C1 | 7.64 | 5.92 | 5.76 | 1.98 | 2.40 | 4.69 | 4.03 | | |
| G2 | 7.94 | | 5.87 | 2.65 | 2.72 | 4.93 | N/A | | 12.9 |
| C3 | 7.29 | 5.43 | 5.40 | 1.82 | 2.18 | 4.77 | N/A | | |
| A4 | 8.16 | | 5.71 | 2.68 | 2.77 | 5.00 | 4.31 | 7.17 | |
| A5 | 8.06 | | 5.71 | 2.56 | 2.78 | 5.00 | 4.37 | 6.95 | |
| A6 | 7.99 | | 5.82 | 2.51 | 2.88 | 5.01 | 4.38 | 6.89 | |
| A7 | 7.92 | | 5.92 | 2.51 | 2.92 | 5.01 | 4.40 | 6.96 | |
| A8 | 7.97 | | 6.05 | 2.43 | 2.89 | 4.97 | 4.41 | 7.55 | |
| T9 | 6.92 | 1.19 | 5.74 | 1.89 | 2.35 | 4.82 | 4.25 | | 13.45 |
| G10 | 7.81 | | 5.80 | 2.58 | 2.66 | 4.96 | 4.33 | | 12.60 |
| C11 | 7.34 | 5.37 | 5.77 | 1.89 | 2.33 | 4.77 | N/A | | |
| G12 | 7.94 | | 6.16 | 2.60 | 2.35 | 4.66 | 4.31 | | 12.80 |
| C13 | 7.61 | 5.89 | 5.76 | 1.95 | 2.39 | 4.69 | 4.03 | | |
| G14 | 7.94 | | 5.89 | 2.66 | 2.72 | 4.96 | N/A | | 12.90 |
| C15 | 7.39 | 5.43 | 5.58 | 2.08 | 2.40 | 4.81 | 4.32 | | |
| A16 | 8.34 | | 6.32 | 2.73 | 2.98 | 5.04 | 4.44 | | |
| T17 | 7.19 | 1.41 | 5.94 | 2.04 | 2.61 | 4.81 | 4.31 | | 13.75 |
| T18 | 7.46 | 1.55 | 6.19 | 2.25 | 2.67 | 4.89 | N/A | | 14.13 |
| T19 | 7.48 | 1.60 | 6.19 | 2.25 | 2.67 | 4.89 | N/A | | 13.99 |
| T20 | 7.47 | 1.62 | 6.12 | 2.16 | 2.65 | 4.92 | N/A | | 13.75 |
| T21 | 7.29 | 1.68 | 5.83 | 2.06 | 2.46 | 4.91 | N/A | | 13.65 |
| G22 | 7.92 | | 5.81 | 2.66 | 2.72 | 4.96 | N/A | | |
| C23 | 7.36 | 5.42 | 5.76 | 1.89 | 2.33 | 4.77 | N/A | | |
| G24 | 7.94 | | 6.16 | 2.60 | 2.35 | 4.66 | 4.31 | | 13.03 |

used to determine the structures of nucleic acids,[31-36] and the next section describes the details of the approach we used that vary in some respects from the methods used by others. The inclusion of the effects of anisotropic tumbling and local motion into the model for the DNA can effect the fit of the predicted and experimental NOE data.[37] These factors were not included here since the development of the methods to incorporate these into restrained molecular dynamics has not been completed. We are cur-

[31] K. D. Bishop, F. J. H. Blocker, W. Egan, and T. L. James, *Biochemistry* **33,** 427 (1994).
[32] O. Y. Federoff, M. Salazar, and B. R. Reid, *J. Mol. Biol.* **233,** 509 (1993).
[33] J. Santa-Lucia and D. H. Turner, *Biochemistry* **32,** 12612 (1993).
[34] P. Schultze, R. F. Macaya, and J. Feigon, *J. Mol. Biol.* **235,** 1532 (1994).
[35] B. I. Schweitzer, T. Mikita, G. W. Kellogg, K. H. Gardner, and G. P. Beardsley, *Biochemistry* **33,** 11460 (1994).
[36] K. Weisz, R. H. Shafer, W. Egan, and T. L. James, *Biochemistry* **33,** 354 (1994).
[37] J. M. Withka, S. Swaminathan, J. Srinivasan, D. L. Beveridge, and P. H. Bolton, *Science* **255,** 597 (1992).

TABLE II
HOMONUCLEAR AND HETERONUCLEAR COUPLING CONSTANTS

| Base | $J_{1'2'}$ | $J_{1'2''}$ | $J_{PH3'}$ |
|---|---|---|---|
| C1 | n/d | n/d | — |
| G2 | n/d | n/d | — |
| C3 | 9.06 | 5.44 | — |
| A4 | 9.16 | 5.06 | 1.3 |
| A5 | 9.16 | 5.08 | 0.8 |
| A6 | 8.96 | n/d | — |
| A7 | 8.04 | 6.1 | — |
| A8 | 8.12 | 6.24 | 3.6 |
| T9 | 10.2 | 5.96 | 0.5 |
| G10 | — | — | 1.5 |
| C11 | — | — | — |
| G12 | — | — | — |
| C13 | — | — | — |
| G14 | — | — | — |
| C15 | 7.4 | 4.9 | — |
| A16 | 9.1 | 6.2 | — |
| T17 | 7.1 | 5.6 | 0.5 |
| T18 | 8.4 | 5.3 | — |
| T19 | 8.4 | 5.3 | 0.5 |
| T20 | 6.0 | 5.8 | 0.5 |
| T21 | — | 6.1 | 1.1 |
| G22 | — | — | — |
| C23 | — | 7.8 | — |
| G24 | — | — | — |

rently investigating how to combine restrained molecular dynamics with free molecular dynamics as a route to incorporating anisotropic tumbling and local motion.

The experimental NOE cross-peak volumes were used as constraints. The experimental homonuclear and heteronuclear couplings were used to generate dihedral constraints defining sugar pucker and the conformation of the backbone. In addition, dihedral constraints were used to keep the purine rings planar. The nonbonded interaction cut-off was set to 11.5 Å. The distance over which the switching functions for nonbonded interactions was switched from on to off was 9.5 to 10.5 Å. The distance cut-off for hydrogen bonding interactions was set to 7.5 Å, and the switching function was applied from 4.0 to 6.5 Å.

The total energy of the system during the simulation is given by

$$E(\text{total}) = E(\text{bond}) + E(\text{angl}) + E(\text{impr}) + E(\text{Vdw}) + E(\text{elec}) + E(\text{Hbon}) + E(\text{dih}) + E(\text{bp}) + E(\text{plan}) + E(\text{NOE}).$$

In this force field there are terms corresponding to the energy of the molecule, as well as to the degree of agreement between the NMR parameters predicted by the structure and those observed experimentally. The first six terms are the standard CHARMM force field.[38] The constraints on dihedral angles gives rise to the term $E$(dih). The agreement between the predicted and observed NOEs gives rise to the $E$(NOE) term. The energy penalty for keeping the purine bases planar is the $E$(plan) term. The energy term for constraining the hydrogen bonding of the A-T and G-C Watson–Crick base pairs is referred to as $E$(bp). Explicit descriptions of these nonstandard energy terms are given in the next section.

*Analytical Expression for the Gradient in Intensities Minima Search.* The derivative of the calculated NOE volumes, $I^{c}_{ij}$, with respect to coordinate $\mu$ of each proton is needed for the minimization procedure. The gradient field for each mixing time, $\tau_m$, is generated using:

$$\sum_{i=1}^{n} \nabla\mu(I^{c}_{ij}) = \nabla\mu[\exp(-\mathbf{R}\{\tau_m\})]^{ij} = \text{Trace}[(\nabla\mu\mathbf{R})\mathbf{L}\mathbf{F}^{(ij)}\mathbf{L}^{T}],$$

where $\mathbf{F}^{(ij)}$ is defined as

$$-\mathbf{L}_{ir}\mathbf{L}_{uj}{}^{T}[\exp(-\lambda_r{}^{t}) - \exp(-\lambda_u{}^{t})]/[t(\lambda_r - \lambda_u)] \text{ if } r \neq u;$$
$$F^{(ij)}{}_{ru} = \mathbf{L}_{ir}\mathbf{L}_{rj}{}^{T} \exp(-\lambda_r{}^{t}),$$

if $r = u$. The $\mathbf{L}_{ir}$ and $\mathbf{L}_{uj}$ are the matrix of eigenvectors and eigenvalues of the relaxation matrix $\mathbf{R}_{ir}$ and $\mathbf{R}_{uj}$, respectively. $\lambda_r$ is the $r$th eigenvalue of the relaxation matrix. The superscript $T$ indicates the trace.

*E(NOE).* $E$(NOE) is the function of difference between the observed and calculated NOE cross-peak volumes

$$E_{(\text{NOE})} = W_N \sum \sum \text{well}(I^{c}_{i}, k_s I^{O}_{i}, \Delta_i, \text{n})^{m},$$

with $W_N$ an energy constant of 20 kcal/mol. The first sum runs over the mixing times of 100 and 200 ms. The second sum runs over all of the cross-peak volumes. The calibration factor for each mixing time, $k_s$, normalizes the experimental and calculated volumes. The well function is given by:

$$\text{well}(a, b, \Delta, n) = (b - \Delta)^n - a^n \quad \text{if } a^n < (b - \Delta)^n$$
$$\text{well}(a, b, \Delta, n) = 0 \quad \text{if } (b - \Delta)^n < a^n < (b + \Delta)^n$$
$$\text{well}(a, b, \Delta, n) = a^n - (b + \Delta)^n \quad \text{if } a^n > (b + \Delta)^n,$$

with $a = I^{c}_{i}$, $b = k_s I^{O}_{i}$. The parameter $\Delta$, 10%, is the estimate of the error

[38] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4,** 187 (1983).

in the determination of the experimental cross-peak volumes. The $I^c_{ij}$ are the cross-peak volumes determined by a complete relaxation matrix calculation. The values of $n$ and $m$ were chosen following the procedures of James and co-workers[39] as $m = 2$ and $n = 1/6$.

*E(bp).* A biharmonic potential was applied to the heavy atoms involved in the hydrogen bonding of A-T and G-C base pairs since preliminary calculations on the Dickerson dodecamer showed that the standard X-PLOR potential did not describe these base pairs well.

$$E(\text{bp}) = \min(\text{ceil}, (S^*K_b^*T/2c_{ij}^2)^*(R - d)^2).$$

The ceil parameter is 30 kcal/mol; $S$ is a scale factor of 20 kcal/mol; the temperature, $T$, is 300°K; Boltzmann's constant is $K_b$; $c_{ij} = d_o - \Delta$ if $R < d$ and $c_{ij} = d_o + \Delta$ if $R > d$; and $\Delta = 0.2$ Å. The distances are $d_o = 1.86$ Å for the carbonyl oxygen to amino proton distances and $d_o = 1.849$ Å for the imino nitrogen proton of $G$ to ring nitrogen of $C$ distance. The $d_o$ for the carbonyl oxygen of $T$ to amino proton of $A$ is 1.964 Å and for the 1.882 Å for the imino nitrogen proton of $T$ to ring nitrogen of $A$ distance.

*E(dih.).* PECOSY data were used to determine the proton H1′,H2′ couplings, and $^{31}$P decoupled NOESY spectra were used to determine the H3′, $^{31}$P couplings as described previously. The H1′,H2″ dihedral angles of residues $C_3$, $A_4$, $A_5$, $A_7$, $A_8$, $T_9$, $C_{15}$, $A_{16}$, $T_{17}$, $T_{18}$, $T_{19}$, $T_{20}$, $T_{21}$, and $C_{23}$ were constrained during the relaxation matrix simulation. The $M_{(n)}$H3′ – $M_{(n+1)}$P dihedral angles were constrained for residues $A_4$, $A_5$, $A_6$, $A_8$, $T_9$, $G_{10}$, $T_{17}$, $T_{19}$, $T_{20}$, and $T_{21}$. The coupling constants were used to calculate the dihedral angles using standard Karplus equations.[29] The energy term is given by:

$$E(\text{dih}) = \sum_{i=1}^{m} k_{\phi i}(1 + \cos(n\phi_i + \delta_i)) \qquad \text{if } n_i > 0$$

$$E(\text{dih}) = \sum_{i=1}^{m} k\phi_i(\phi_i - \delta_i)^2 \qquad \text{if } n_i = 0,$$

where $k_{\phi i}$ is the energy constant, 8 kcal/mol; $n_i$ is the periodicity, 2; $\delta_i$ is the dihedral angle calculated from the relevant Karplus equations; $\phi_i$ is the dihedral angle of the trial structure; and $m$ is the number of dihedral angles that are considered. The energy constraints on the dihedral angles were chosen to be sufficiently loose that inaccuracies in the determination of the scalar couplings due to scalar relaxation and linewidth effects could be neglected.

[39] U. Schmitz, A. Kumar, and T. L. James, *J. Am. Chem. Soc.* **114,** 10654 (1992).

*E(plan).* Preliminary runs showed that the purine bases became puckered using the standard X-PLOR force field. Therefore, an additional term,

$$E(\text{plan}) = W_{\text{plan}} \sum_{i}^{gi},$$

was used to keep these residues planar. The $W_{\text{plan}}$ constraint is 100 kcal-mol$^{-1}A^{-2}$, and $g_i$ is the orthogonal distance of the particular atom form, the least squares plane defined by all of the atoms of the purine.[40] This method was found to be more computationally efficient than the use of dihedral constraints to keep the purines planar.

*Restrained Dynamics*

The starting structure was canonical B-form DNA. The energy of this structure was minimized by 50 steps of conjugate gradient minimization using the standard CHARMM force field. This was followed by a 15-step minimization using the force field with the restraints. After minimization, a restrained molecular dynamics trajectory of 150 ps was run, which was followed by an additional 250 steps of conjugate gradient energy minimization. This structure was subjected to an additional 100 ps trajectory run that was followed by a 250-step conjugate gradient energy minimization. The structures obtained after the 150 and 250 ps time points were found to be essentially identical indicating that there was little to be gained by running the simulation longer than 150 ps. The values for the energy terms for the beginning structure and at various times during the structure refinement procedure are given in Table I, and the structure is shown in Fig. 6.

*Calculation of Spectra*

The volumes of the cross-peaks in the two-dimensional spectra were calculated using a complete matrix relaxation calculation using the proton coordinates of the structure and numerical integration of the Bloch equations. These calculations were performed using the MODEL program of FELIX 2.1, using an overall correlation time of 5 ns, a leakage rate of 0.33 $s^{-1}$, and a distance cut-off of 5.5 Å. The volumes were combined with the assignments to generate the simulations of the two-dimensional spectra. Calculated spectra using the NMR and crystal structures are shown in Fig. 5.

[40] V. Shomaker, J. Waser, R. E. Marsh, and G. Bergman, *Acta. Crystallogr.* **12,** 600 (1959).

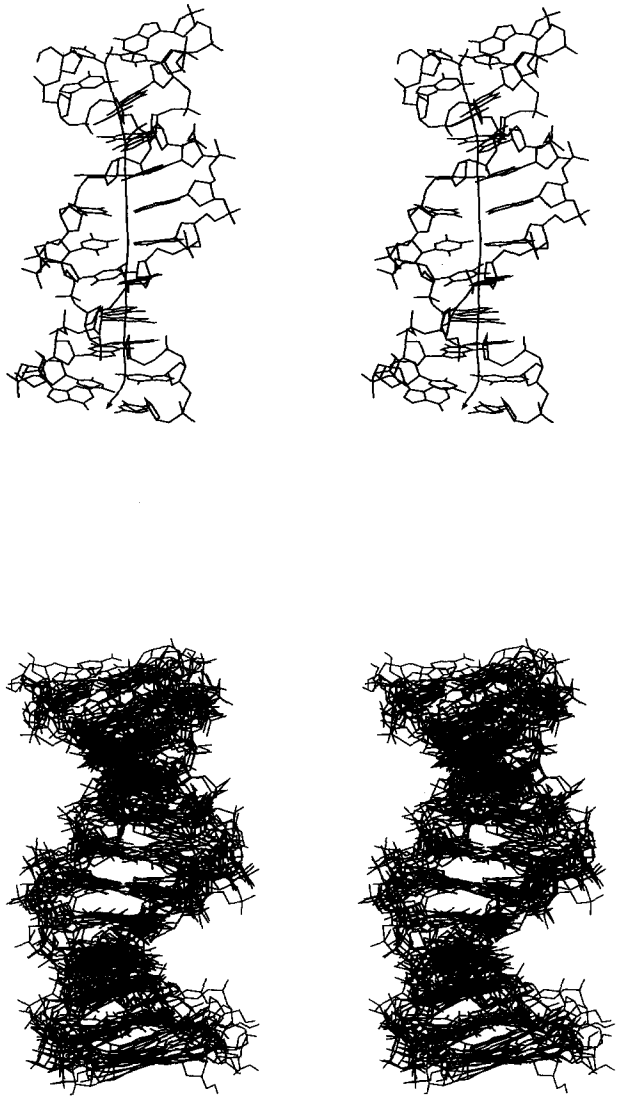


FIG. 6. The top of this figure shows stereo stick representations of the NMR structure. The bottom shows an overlay of 10 structures from the 500 ps MD trajectory taken at equal 50 ps intervals. The global helical axis calculated by Curves is presented within the top figures.

*Analysis of Convergence*

The parameters from Curves, Dials, and Windows for the structure were determined.[17,18] The quality of the fit of the experimental and predicted NOEs was determined by calculation of the NMR *R*-factor as well as the *Q* and RMS as a function of the time of the trajectory.[42] It was found that the *R*, *Q*, and RMS dropped from their initial values associated with the starting structure during the first 1–3 ps of the trajectory. The values of *R*, *Q*, and RMS for the final structure are 0.227, 0.125, and 0.111, respectively. The energy of the DNA at various points in the trajectory are given in Table I. Analysis via Curves, Dials, and Windows indicates that the structures in the trajectory are quite stable.

The analysis of the helicoidal parameters and the energy of the DNA indicated that the trajectory was stable after about 10 ps. However, the agreement between the calculated and experimental NOESY spectra improves as the length of the trajectory increases. The results in Fig. 5 show that the intensities of the AH2-H1′ cross-peaks change significantly from 10 to 20 ps, and further change occurs between 20 and 150 ps. It appears to be the case that the restrained molecular dynamics trajectory first fits the high intensity NOE cross-peaks, which correspond to the shorter distances, and requires more time to fit the low intensity cross-peak volumes.

*Analysis of NMR Structure*

The structure of the DNA as determined by NMR is shown in Fig. 6. This is a structure in the B-DNA family with modest deviations from the canonical values of roll, tilt, and the other helicoidal parameters. The terminal base pairs are not well defined in the NMR structure. The roll and tilt are shown in Fig. 7. This structure is in good agreement with the experimental NOE data collected at two mixing times, as well as the homonuclear and heteronuclear couplings. This structure bends in the major groove.

This structure does show the characteristic narrowing of the minor groove, which has been observed in the "up" crystal structure of this DNA and in other crystal structures. The minor groove widths for the NMR structure and for the "up" X-ray structure of this DNA are 10.36/14.36 for A5-G24, 10.49/13.67 for A6-C23, 8.31/9.88 for A7-G22, 8.31/9.19 for A8-T21, 8.55/9.00 for T9-T20, 9.25/8.31 for G10-T19, 10.61/10.86 for C11-T18, and 10.82/9.97 Å for G12-T17. The minor groove width for canonical B-form

[41] Deleted in proof.

[42] J. M. Withka, J. Srinivasan, and P. H. Bolton, *J. Magn. Reson.* **98,** 611 (1992).

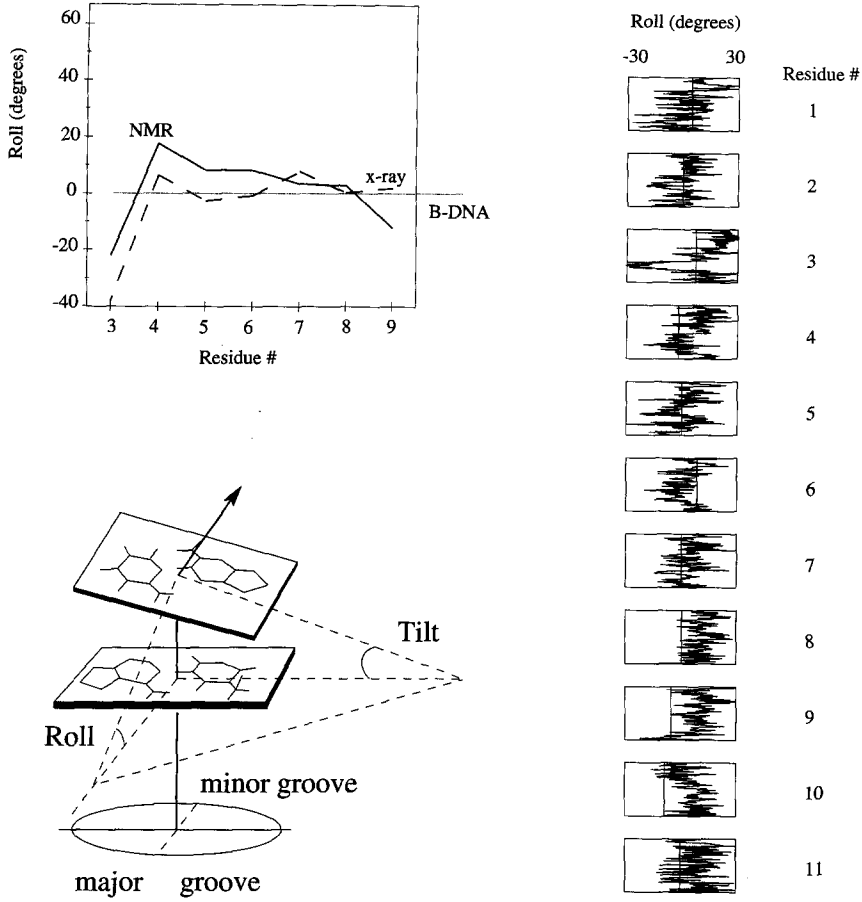



FIG. 7. The roll of the base pairs for the NMR structure and the "up" crystal structure are shown in the upper left plot. The roll of the base pairs during the MD trajectory is shown at the left. The roll of the base pairs during the trajectory are shown as windows, with the start of the trajectory at the top and the end of the trajectory at the bottom. Also shown is a depiction of roll and tilt with the angles as defined in Fig. 2.

DNA is 11.5 Å. The "up" crystal structure shows a progressive narrowing of the minor groove, from the 5′ to 3′ end, whereas the solution structure has a more uniformly narrow minor groove.

The solution and "up" crystal state structures are both shown in Figs. 1 and 8 (see color insert for Figure 8) for comparison purposes. The two structures have many similarities with bending in the same direction and at the same location. There are differences between the two structures in detail, but the overall nature of the bend is essentially the same in both cases. As shown

in Fig. 7, the propeller twist in the NMR and crystal structures is strikingly similar.

Thus, it appears that in the case of the duplex DNA formed of d(CGCAAAAATGCG) paired with the complementary strand, similar bending can occur in the "up" crystal and solution states. Fig. 5 contains the NOESY map predicted by the "up" and "down" crystal structures as well as the experimental data. This comparison shows that both crystal structures offer a reasonable description of the NMR data. However, neither crystal structure predicts particularly good agreement with the experimentally observed H2-H1′ interstrand NOEs, as well as in several other regions, which are also poorly predicted by the canonical B-form structure. The two crystal structures also predict significantly different spectra as highlighted in Fig. 5.

As a control we applied the same structure determination methodology to the Dickerson dodecamer.[43] The structure of the Dickerson DNA, determined by this approach, showed no significant deviation from a B-form DNA structure in agreement with prior solution state NMR studies.[37,44]

## Molecular Dynamics Simulations

Three MD simulations have been recently carried out on the d(CGCAAAAATGCG) duplex, employing as starting configurations the canonical B-form of the duplex and the crystallographic "up" and "down" forms. All elements of the MD other than the starting configuration of the DNA and surrounding water were conserved across the three MD trajectories. The simulations were carried out using the program WESDYN and treated with the DNA duplex immersed within a periodically replicated hexagonal prism cell composed of ~3000 water molecules, sufficient to extend at least 12 Å beyond all DNA atoms. The simulations employed the GROMOS86 force field[45] augmented with a

$$E(W.C.) = k\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{10}\right\},$$

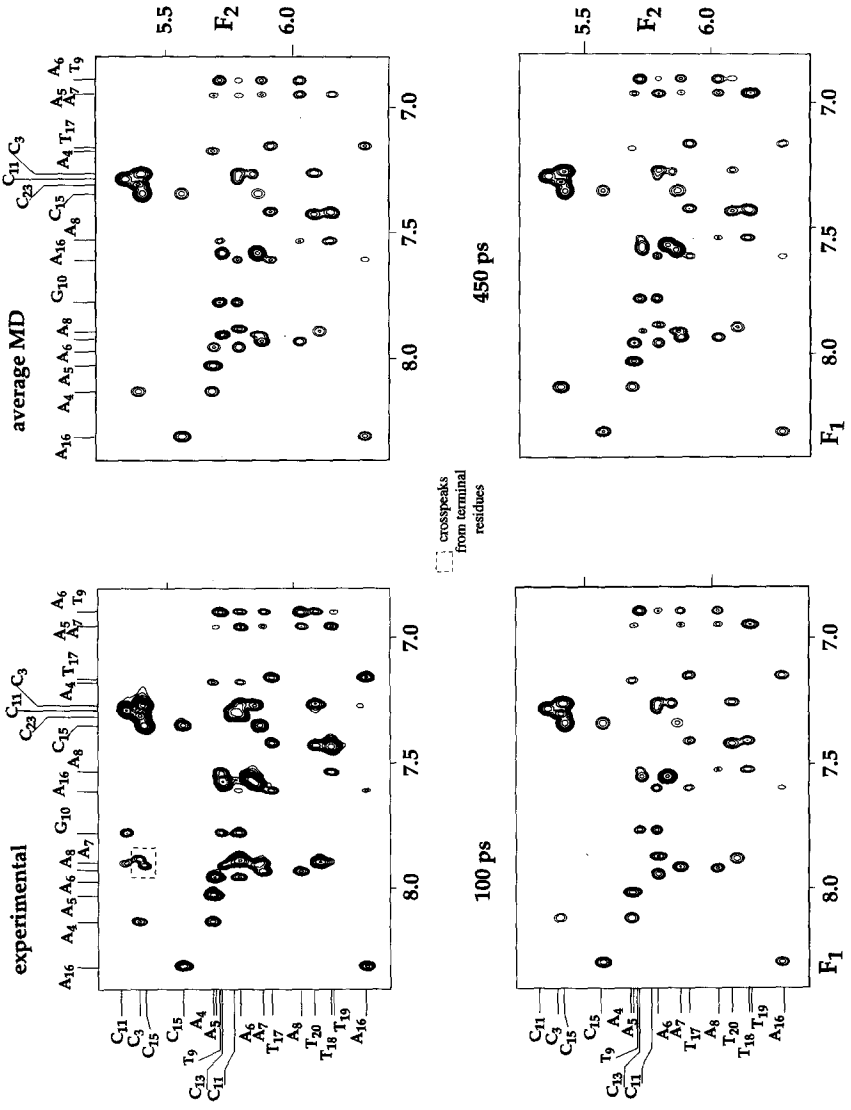
12–10 potential for Watson–Crick hydrogen bonds. Charges on the phosphate groups were set at −0.24 eu to model counterion condensation effects.[46] A long-range switching function, 4 Å in width (on at 7.5 Å, off at 11.5

[43] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itikura, and R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2179 (1981).

[44] A. Lane, T. C. Jenkins, T. Brown, and S. Neidle, *Biochemistry* **30,** 1372 (1991).

[45] W. F. van Gunsteren and H. J. C. Berendsen, "GROMOS-86." University of Groningen, Groningen, 1986.

[46] G. S. Manning, *Acct. Chem. Res.* **12,** 443 (1979).


experimental
average MD
100 ps
450 ps
crosspeaks from terminal residues
F1
F2
5.5
6.0
7.0
7.5
8.0

Å), was selected to eliminate simulation artifacts caused by the presence of a more abrupt truncation of electrostatic interactions. Prior to initiating the dynamics, the water in each system was equilibrated using the Metropolis Monte Carlo technique on all rotational and translational degrees of freedom. The systems were then iteratively heated to 300°K by introducing random velocities into the atoms for a period of 1.5 ps. The velocities were rescaled to a Gaussian distribution over the next 3.5 ps. For the duration of the trajectories, the temperature of the system was allowed to fluctuate within a temperature window of ±4°K. The simulations were executed at constant energy, as the temperature never drifted outside of this window.

The assessment of trajectory convergence was made using a combination of atomic RMS deviation and the analysis of DNA helicoidal parameters. All simulations were initially carried out to 250 ps, at which point the results from the canonical and the "down" crystal form MD showed close agreement when analyzed using bending dials and helicoidal analysis. At this point in the trajectory beginning with the "up" crystallographic form, the duplex was found to show a net bending into the minor groove centered at the C3-A4 step, whereas the others showed net bending toward the major groove at this position. Extending the "up" form simulation to 500 ps led to an essentially complete coincidence of local bending behavior in all three simulations, leading us to conclude that the MD results are independent of the choice of starting configuration. Full details of this analysis have been given by Young *et al.*[20]

*Analysis of Fine Structure*

The MD simulation was also subjected to conformational and helicoidal analysis using Curves,[18] Dials, and Windows.[17] The helicoidal dynamics show from the behavior of XDP that a good B-form DNA is maintained by the simulation with some tendency to A-form base-pair inclination within the A-tract. There is extensive BKL and PRP dynamics, the latter only slightly larger within the A-tract. The conformational analysis shows considerable activity in the parameters $\alpha$ and $\gamma$, executing correlated conformational transitions that preserve the intact helix, which are called crankshaft motions. These are seen within the A-tract, as well as at the junctions. The

FIG. 9. The top left spectrum is the experimental two-dimensional NOESY spectrum of the DNA duplex obtained at 600 MHz with a mixing time of 200 ms. The assignments of the signals are indicated, and this region contains aromatic to H1′ and H6-H5 cross-peaks. The top right spectrum is that predicted from the average MD structure. The bottom left spectrum is that predicted from the structure at 100 ps into the trajectory. The bottom right spectrum is that calculated from the structure at 450 ps into the trajectory.

dynamics of the sugar rings consists of excursion in many cases and a transient repuckering into A form C3′ endo pucker at the junction, correlated with local bending. An overlay of the structures obtained in the trajectory is shown in Fig. 9.

Analysis of ROL and TLT is best viewed in the presentation of bending dials for the MD simulation (Fig. 3C). Here the dials present the direction of bending calculated at each step over the course of the 500 ps trajectory as a probability contour. The bending dials indicate that the overall bending is centered close to the origin within the A-tract, and thus the calculated A-tract is seen to be relatively straight. Significant displacements in local axis bending occur at both junctions of the A-tract with the flanking sequences, consistent with a generalized junction bending model.

### *Analysis of Helix Morphology*

The UH analysis of Fig. 4 presents the ensemble of MD structures as a distribution superimposed over the crystallographically determined starting structure. When the entire trajectory is viewed in this manner we observe an overall narrowing of the minor groove in the region of the A-tract, but we also observe that a considerable range of backbone atom locations is visited during the course of the trajectory.

The MD trajectory has been compared with the experimental NOE data via comparison of the spectra calculated from the structures of the MD trajectory. This comparison has been made several ways. The structures taken at two time points have been used with the "best" and "worst" time points being taken as examples. In addition, the average of the results predicted by 10 structures taken at 50 ps intervals has also been calculated. The predicted and experimental spectra are shown in Fig. 9. This comparison shows that the average over the trajectory offers good agreement with the experimental results. In particular, the spectrum predicted by the average of the trajectory offers good agreement with the AH2-H1′ intrastrand and interstrand NOEs. The results predicted by the two snapshots of the trajectory indicate the range of agreement for the structures over the time course of the trajectory.

## Discussion

The structure determined for d(CGCAAAAATGCG) obtained from the crystallography, NMR refinement, and MD simulation are compared in Figs. 1, 8, and 9. The mode of presentation uses ellipsoids constructed around the mean positions of each particle, indicating the dynamic range of contributing structures. From Fig. 8 and the comparison of bending dials

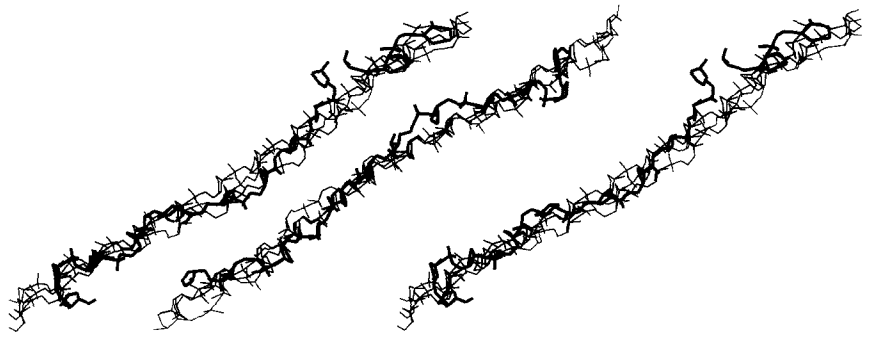


FIG. 10. "Unrolled helix" representation of the minimized structure generated using NMR restraints superimposed on the two crystallographic structure determinations.

in Fig. 3, we find that the A-tract region of the crystal structure and the MD structures is relatively straight. In the NMR structure, there is some concerted bending toward the major groove within the A-tract region. The results of all three structure determinations are consistent in localizing considerable axis bending at or near the junctions of the A-tract with the flanking sequences. Young *et al.*[10] have recently analyzed all of the crystal structure data on A-tract DNAs using bending dials. They found that the crystal structures of the A-tract DNAs are all relatively straight, with bends occurring at or near the junctions. The junction bending is not, however, identical to that originally proposed by Crothers *et al.* based on gel electrophoresis experiments.[15,47] In each of the structures obtained, there is an overall bending observed in the helix axis. The up and down crystal forms show bending into the major and minor grooves, respectively. The NMR structure exhibits axis bending toward the major groove. The MD results show an overall bend into the major groove, but an oscillation between major groove and minor groove bending within the simulation. This is consistent with the sensitivity of observed bending to lattice forces in the crystal.

The groove widths in the crystal structures and MD model are compared in Fig. 4 using the "unrolled helix" representation. There is distinct narrowing of the minor groove in the A-tract of the crystalline forms, a feature commented on in a number of other structures as well. In the MD model, the grooves seem more displaced than narrowed. This occurs as a consequence of a slight untwisting of the helix to an average twist angle 32.5 at the end of the trajectory. This effect can be expected to be sensitive to the

[47] H.-M. Wu and D. Crothers, *Nature* **308,** 509 (1984).

force field underlying the simulation and the choice of a reduced charge model for the phosphates in the simulation.

The comparison of the crystal structure and NMR results indicates that the structure of the "up" crystal form is most similar to that observed in solution. The comparison of the experimentally based structures and the MD structure indicates that the details of the bending in the dynamical model are in reasonable accord with axis bending in both the crystallographic and the NMR structure, and it also supports the idea that the "up" form is more like the structure for d(CGCAAAAATGCG) in solution.

## Summary and Conclusions

We have presented a detailed analysis for structure determinations for the DNA duplex d(CGCAAAAATGCG) obtained from X-ray crystallography, nuclear magnetic resonance, and molecular dynamics simulation. Each of the structures for the duplex deviates from the structure of the canonical form of B-DNA in a number of observable characteristics. Specifically, the three determinations all contain DNA axis deflections at the junctions of the A-tract with the flanking sequences. The analysis provided shows that the general characteristics of the structures obtained for d(CGCAAAAATGCG) from X-ray, NMR, and MD methods turn out to be quite similar. The extent to which this result can be generalized remains to be established by consideration of similar cross-comparisons on diverse oligonucleotide structures.

## Acknowledgments

This work was supported by grant NP-750B from the American Cancer Society to P. H. Bolton and grant GM 37909 to D. L. Beveridge from the NIH. The 600 MHz spectra were obtained with the assistance of Dr. Frits Abildgaard using the NMR facility at the University of Wisconsin, Madison, which is supported by NIH grant RR02301 with equipment purchased with support from the University of Wisconsin, NSF DMB-8415048, NIH RR02781, and the USDA.