Solution Structure of $[d(GGTATACC)]_2$: Wrinkled D Structure of the TATA Moiety†

Ning Zhou, Anna Maria Bianucci, Nagarajan Pattabiraman, and Thomas L. James*
Departments of Pharmaceutical Chemistry and Radiology, University of California, San Francisco, California 94143
Received March 16, 1987; Revised Manuscript Received June 22, 1987

ABSTRACT: Phase-sensitive two-dimensional nuclear Overhauser effect spectra of $[d(GGTATACC)]_2$ in aqueous deuterium oxide solution at four mixing times were quantified to give all nonoverlapping cross-peak intensities. A structural model for $[d(GGTATACC)]_2$ was built in which the GG- and -CC moieties were in the B-DNA form, while the middle -TATA- moiety was in the wrinkled-D form (BDB model). This model was subjected to energy refinement by molecular mechanics calculations with the program AMBER. Counterions (Na^+) were added to neutralize the charges, and water molecules were placed bridging across the minor groove. A complete relaxation matrix analysis was used to calculate two-dimensional nuclear Overhauser effect spectra of $[d(GGTATACC)]_2$ from the above models (before and after energy refinement) and from four other $[d(GGTATACC)]_2$ structural models: regular A, crystalline A, regular B, and energy-minimized B. Among them, the energy-minimized BDB model yielded a set of theoretical spectra that gave the best fit to the experimental spectra. It was also the energetically most stable. Therefore, it is a good representation of the ensemble- and time-averaged structure of the octamer in solution. This model has backbone torsion angles similar to those of B-form DNA in the GG- and -CC moieties and torsion angles similar to those of wrinkled D form DNA in the -TATA- moiety. The base stacking and base pairing are not interrupted at the junctions between the two structural moieties. Its minor groove is narrower than that of B DNA, and the solvent-accessible surface of the minor groove forms a closed hydration tunnel in the middle -TATA- segment.

Recent progress in DNA synthesis has enabled detailed structural studies of oligonucleotides with known sequences by single-crystal X-ray diffraction [e.g., Wang et al. (1979), Wing et al. (1980), and Dickerson and Drew (1981)] and by nuclear magnetic resonance [e.g., Feigon et al. (1983), Haasnoot et al. (1983), Wemmer et al. (1984), and Broido et al. (1985)]. It is obvious from these studies that the structures of oligonucleotides depend upon their base sequences and their environments. Torsion angles in oligonucleotide crystal structures are not uniform; they are sequence-dependent (Calladine, 1982; Dickerson, 1983). In solution, even more extensive structural variations can be expected. It is in the context of these sequence-dependent structures that we will have to understand the biological functions of DNA. Of particular interest in our laboratory are oligonucleotides containing alternating d-TA base pairs, as these commonly occur in the promoter region of genes, being recognized by RNA polymerase.

The crystal structure of the octamer $[d(GGTATACC)]_2$ was reported to be in A form (Shakked et al., 1981, 1983), while qualitative proton–proton one-dimensional nuclear Overhauser enhancement (NOE)[1] measurements by Reid et al. (1983) revealed a B-DNA type of structure for this molecule in aqueous solution. In a previous study from our group (Jamin et al., 1985), phase-sensitive two-dimensional (2D) proton–proton NOE spectra of $[d(GGTATACC)]_2$ were obtained, and nonoverlapping cross-peak intensities involving base aromatic protons and deoxyribose 1′-protons and 2′,2″-protons were quantified and used for detailed structure analysis with complete relaxation matrix analysis (CORMA) (Keepers & James, 1984). It was found that the GC base pairs were essentially in the regular B form but the -TATA- moiety could not be fit well by either regular B form or energy-minimized B form structures. The A-form structure can be excluded for this molecule from the qualitative pattern of the 2D NOE spectra, but more subtle structural differences between the $\overset{GG-}{CC-}$ parts and the $\overset{-TATA-}{-ATAT-}$ part do seem to exist. Another proton–proton 2D NOE study of this molecule in aqueous solution was reported shortly after (Patel et al., 1986). The authors concluded from relative 2D NOE peak intensities that the -TATA- part of this molecule adopts a B type of structure. It appears that only an approach using 2D NOE CORMA spectra has the power to detect detailed structural differences.

A recent study of $[d(AT)_5]_2$ from our group (Suzuki et al., 1986) found that a wrinkled-D (wD) structure (Arnott et al., 1983), a member of the B family, best accounted for the experimental 2D NOE spectra compared with seven well-known DNA structure models for this decamer duplex: regular B, regular A, alternating B, and left-handed B, C, D, and Z form structures. Molecular mechanics calculations also indicated that the wD structure was the best model for $[d(AT)_5]_2$ in moderate salt solution. Interesting structural features of wD DNA include (a) eight base pairs per helical turn; (b) a much narrower minor groove, which has the consequence of creating a hydration tunnel; (c) significant

†This work was supported by the National Science Foundation (Grant PCM 84-04198), the National Institutes of Health (Grant CA 27343), and the donors of the Petroleum Research Fund, administered by the American Chemical Society, as well as by instrumentation grants from the National Institutes of Health (RR 016688) and the National Science Foundation (DMB 84-06826). A.M.B. thanks the Italian National Council of Research for financial support.

* Address correspondence to this author at the Department of Pharmaceutical Chemistry.

[1] Abbreviations: 2D, two dimensional; ADI, average difference index; AMBER, assisted model building with energy refinement; BDB, a structural model for $[d(GGTATACC)]_2$ in which the GG- and -CC moieties are in regular B form while the -TATA- moiety is in wrinkled D form; CORMA, complete relaxation matrix analysis; DI, difference index; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; wD, wrinkled D.

0006-2960/87/0426-7905$01.50/0

cross-strand hydrophobic interactions, and (d) alternating torsion angle values at TA and AT steps (Suzuki et al., 1986).

Considering this apparent preference for wD structure for $[d(AT)_5]_2$, we built a model for $[d(GGTATACC)]_2$ in which the -TATA- moiety exists in the wD form and the $^{GG}_{CC}$ parts are in regular B form (referred to as BDB model hereafter). Seventeen Na^+ ions and nine water molecules were included in this BDB model for subsequent energy minimization on the basis of the following structural features Suzuki et al. (1986) found in the $[d(AT)_5]_2$ study: (A) counterions are important to reduce backbone phosphate group repulsions and to stabilize the comparatively narrow minor groove; (B) a hydration tunnel exists in the minor groove in which water molecules apparently can bridge between strands via hydrogen bonding.

A complete reanalysis of the phase-senstitive proton–proton 2D NOE spectra of $[d(GGTATACC)]_2$ from our previous study (Jamin et al., 1985) was carried out. We quantified all nonoverlapping cross-peak intensities, including intra- and interstrand NOE peaks between nonexchangeable base–base protons, base–sugar protons, and sugar–sugar protons; the numbers of nonoverlapping cross-peaks quantified and used in this analysis at mixing times of 400, 250, 175, and 100 ms were 152, 145, 130, and 115, respectively, with methyl protons of thymine excluded. Theoretical 2D NOE spectra were calculated from six models, namely, regular A, regular B, energy-minimized B, BDB, energy-minimized BDB models, and crystal structure A for $[d(GGTATACC)]_2$ with CORMA (Keepers & James, 1984); these theoretical spectra were then compared with experimental 2D NOE cross-peak intensities. The number of cross-peaks used in this study represents a large increase from the previous study (Jamin et al., 1985), where only 20 cross-peaks at each of the mixing times were quantified and used in the analysis.

Experimental Procedures

Quantification of Experimental 2D NOE Peaks. Experimental conditions for obtaining the phase-sensitive 2D NOE spectra have been described before (Jamin et al., 1985). Data were processed on a Vax 11/750 computer and peak intensities quantified with programs developed by Drs. S. Manogaran and R. Scheek at UC, San Francisco. Shifted 60° and 45° sine-bell apodizations were used in t_2 and t_1 dimensions before Fourier transformation to improve resolution of the spectra. One zero filling in the t_1 dimension was used, and the final data matrix size was 1K × 1K. The spectra were not symmetrized. The program for determining 2D NMR peak intensities allows one to examine contour plots of 2D NOE spectra on the screen and to select a square region enclosing one peak (or more peaks together in case of overlap) and then to add the intensities of all the data points (typically about 50 points defining one peak) inside this region to give the 2D NMR peak intensity. The intensities of all cross-peaks and diagonal peaks were quantified in this way. Then the total intensities of the peaks from each column (two or three columns, in case of overlap) of the contour plot were normalized to one (in case of overlap, two or three), and an average normalization factor was calculated from these. Finally, this average normalization factor was used to normalize each nonoverlapping cross-peak intensity of the 2D spectra. The intensities of the cross-peaks between protons i and j (I_{ij}) and between protons j and i (I_{ji}), which appear in the two triangles of the 2D NOE contour plot above and below the diagonal line (Jeener et al., 1979), were averaged to give the I_{ij} value actually used.

Model Building. The BDB model was built as follows: a wrinkled-D (wD) structure generated from the wD-DNA coordinates of Arnott et al. (1983) was assumed for the middle AT-rich segment $^{TATA}_{ATAT}$; a regular B structure generated from the coordinates of B DNA (Arnott & Hukins, 1973) for the $^{GG}_{CC}$ part was added to one end of the middle segment with the program MIDAS (Gallo et al., 1985) on an Evans & Sutherland color picture system (PS2) to adjoin the backbones of the two complementary chains forming a hexamer. Protons were added according to standard bond lengths and bond angles. In order to improve the atom contacts at the structural junction, the $^{GT}_{CA}$ part was then subjected to energy minimization (for details, see Molecular Mechanics) with the rest of the molecule fixed. Since the duplex $[d(GGTATACC)]_2$ has a $C2$ symmetry axis, the hexamer $^{GGTATA}_{CCATAT}$ model was rotated 180° about the symmetry axis to generate the other junction and thus the whole octamer.

Two water molecules were then added bridging interstrand N3 atoms of neighboring adenines by forming hydrogen bonds in the $^{TA}_{AT}$ region as suggested by the studies of Suzuki et al. (1986). Three water molecules were added bridging interstrand O2 atoms of thymine or cytosine by forming hydrogen bonds in neighboring base pairs in $^{AT}_{TA}$ and $^{GT}_{CA}$ regions. Four more water molecules were placed bridging the above five water molecules through hydrogen bonding. Sodium ions were placed by iteratively calculating and neutralizing the electrostatic potential around the DNA–water complex according to Devarajan and Pattabiraman (Fortran Program to Calculate Interactively Ionic Positions Around DNA, 1986, unpublished results). Instead of placing ions one by one as previously, we modified the program to place sodium ions simultaneously at the grid points that have electrostatic potential energies within 0.1 kcal and are more than 6 Å away from each other. The major distribution of ions is the same from the two methods except the positions of the last few ions added. Partial atomic charges from Singh and Kollman (1984) and a dielectric constant value of 80 were used in the calculation. The same procedures were utilized to add water molecules and sodium ions to the B model.

Molecular Mechanics. The coordinates for regular B form structure (Arnott & Hukins, 1973), as well as regular A (Arnott & Hukins, 1972) and crystal A form structures (Shakked et al., 1983), of $[d(GGTATACC)]_2$ were from X-ray diffraction studies and do not include proton coordinates. For molecular mechanics calculations and 2D NOE spectral calculations (see next section), protons were added by use of standard bond lengths and bond angles, as was done for the BDB model. Molecular mechanics calculations were carried out with complete B and BDB models (including sodium ions and water molecules) by use of the program AMBER, Assisted Model Building with Energy Refinement (Weiner & Kollman, 1981; Weiner et al., 1984a), with empirical potential energy functions. The energy functions include harmonic bond stretching and bond angle bending terms, a truncated Fourier series for torson angle terms, and standard Lennard–Jones and electrostatics terms for nonbonded interactions. The various constants used to evaluate the energy were from Weiner et al. (1984b, 1985). A distance-dependent dielectric constant (Hopfinger, 1973; Blaney et al., 1982) was utilized to mimic bulk solvent effects. In the calculations, an all-atom model, in which all hydrogen atoms are included explicitly, was used. The structures were refined until the root mean square derivative of the energy function relative to atomic coordinate changes was less than 0.1 kcal/(mol·Å).

Calculation of Theoretical 2D NOE Spectra. A complete relaxation matrix analysis developed by Keepers and James (1984) was used to calculate theoretical 2D NOE spectra from

each model. An isotropic motional model of the molecule was assumed. This assumption has been shown to introduce only small errors (Keepers & James, 1984). A correlation time (τ_c) of 4 ns was chosen for the overall isotropic motion of the molecule on the basis of the following: with either energy-minimized B, BDB, or energy-minimized BDB models and a τ_c of 4 ns, the decay with mixing time of the calculated diagonal peak intensity for nonoverlapping base protons and deoxyribose 1′-protons fits the experimental data well. A simplified treatment for the thymine methyl was used in which the three methyl protons were replaced by a three-proton centroid, and rotation around the C5–methyl axis was ignored. Since this was not an accurate treatment, the cross-peak intensities directly involving methyl protons were not used in subsequent comparison between different structural models, but the approximation is deemed sufficient to account for secondary effects entailing the methyl protons.

Results and Discussion

Comparison of Experimental and Calculated 2D NOE Spectra. The normalized experimental 2D NOE cross-peak intensities at a mixing time of 250 ms are listed in Table I as examples, and normalized cross-peak intensities at 100, 175, and 400-ms mixing times are available as supplementary material (see paragraph at end of paper regarding supplementary material). These experimental 2D NOE cross-peak intensities at four different mixing times were compared with theoretical ones calculated from the six differential models for [d(GGTATACC)]$_2$. The difference between the experimental and the theoretical 2D NOE peak intensity was calculated for each detected cross-peak as shown in eq 1 and used as a difference index (DI) of this peak for comparison. The average DIs (ADIs) (see eq 2 for definition) were then calculated for each group of cross-peaks arising from interresidue and intraresidue interactions and for all of the nonoverlapping cross-peaks to evaluate these different models. $I_{i,j}$ denotes

$$\mathrm{DI}(\alpha,\beta,\tau_m) = |I^{\alpha}_{i,j} - I^{\beta}_{i,j}| \quad (1)$$

$$\mathrm{ADI}(\alpha,\beta,\tau_m) = \frac{\sum_{i,j}|I^{\alpha}_{i,j} - I^{\beta}_{i,j}|}{N} \quad (2)$$

intensity of detected cross-peak between proton i and j, $i \neq j$, superscripts α and β indicating experimental and theoretical spectra, respectively; τ_m is the mixing time of the 2D NOE experiment; N is the number of cross-peaks used in the calculation.

The ADI of a model can thus be used to estimate how well this model accounts for the experimental 2D NOE data. The ADI values of all nonoverlapping cross-peaks for A, crystalline A, B, energy-minimized B, BDB, and energy-minimized BDB models are plotted at four mixing times in Figure 1. From Figure 1, it is clear that both the regular A and crystalline A form structures give larger ADI values than the four B family structural models and, therefore, can be eliminated for the solution structure of the [d(GGTATACC)]$_2$. The ADI values of the other four models (B, BDB, and their energy-minimized forms), though, are relatively close, as would be expected since they all belong to the B family of DNA structures. For the lowest cross-peak intensity that was included in the analysis (0.001, i.e., 0.1% of the average intensity of a whole column), the signal/noise ratio was ~3.7; peaks with lower S/N ratio were not included though they were still easy to locate on contour plots. Therefore, the uncertainty in the cross-peak intensity is ≤0.001, and this unables distinction among these several models. When only cross-peaks

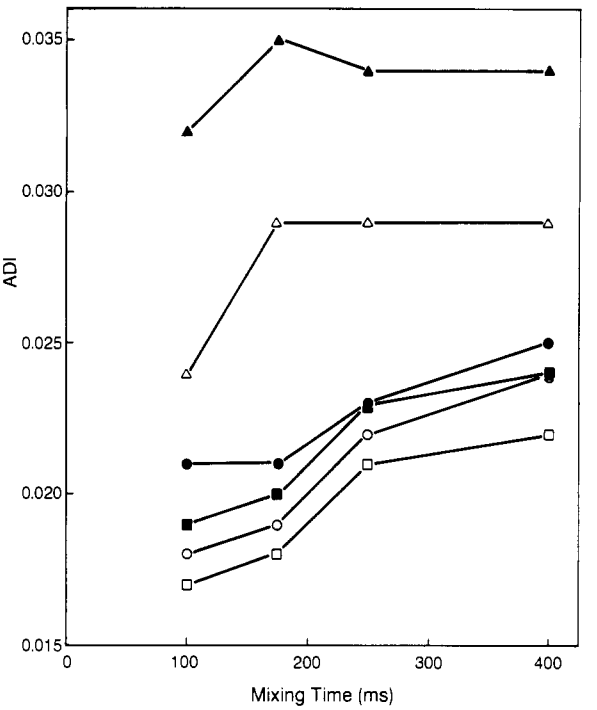

FIGURE 1: Average difference index (ADI) values (see text for definition) of the 2D NOE cross-peak intensities of [d(GGTATACC)]$_2$ for the six models at four mixing times: regular A (▲), crystalline A (△), regular B (●), energy-minimized B (○), BDB (■), and energy-minimized BDB (□).

involving base proton and involving sugar 1′-protons (92 cross-peaks at each mixing time) were used for comparison among models, the resulting differences between ADI values of the models increase, permitting greater distinction between models. For example, the energy-minimized BDB model gave ADI values of 0.011, 0.013, 0.024, and 0.025 at mixing times of 100, 175, 250, and 400 ms, respectively; the corresponding ADI values of the energy-minimized B model were 0.014, 0.016, 0.028, and 0.030. This can be understood since the distances involving base protons and sugar 1′-protons are generally more sensitive to the structural differences between these two models.

Our energy-minimized B model gives smaller ADIs than the standard B model. In the previous analysis by Jamin et al. (1985), the regular B model of [d(GGTATACC)]$_2$, without any counterions or hydrogen-bonded water molecule, was subjected to energy minimization. The resulting energy-minimized B model did not fit the experimental data better, indeed, for many cross-peaks compared, the fit was actually worse than the regular B model. This suggests that including counterions and hydrogen-bonded water molecules to simulate the immediate environment of the DNA molecule in the energy-minimization process is necessary to improve a solution structure model. Similarly, including counterions and hydrogen-bonded water molecules for energy minimization of the BDB model also resulted in an energy-minimized BDB model that fit the experimental 2D NOE data better than the starting BDB model. The approach we used to place counterions is a reasonable approximation to include Na^+–DNA interactions. Lacking good ways of including solvent water molecules, the effect of bulk solvent was simulated by using a distance-dependent dielectric constant (Hopfinger, 1973; Blaney et al., 1982), and this was shown to be a reasonable approach (Lybrand & Kollman, 1985). The nine water molecules that were placed across the minor groove did not have large effect on the resulting energy-refined structures:

Table I: Normalized 2D NOE Cross-Peak Intensities of $[d(GGTATACC)]_2$ at a Mixing Time of 250 ms[a,b]

G1G1	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8
H1′								
H2″	0.085							
H2′	0.074							
H3′	0.032	0.068	0.097					
H4′	0.058			0.121				
H5″								
H5′								
H8	0.034	0.089	0.132	0.040	0.019			

G1G2	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8
H1′								0.047
H2″								0.052
H2′								
H3′								0.025
H4′								
H5″								
H5′								
H8								0.037

G2G2	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8
H1′								
H2″	0.104							
H2′		0.289						
H3′	0.043	0.081	0.109					
H4′	0.042			0.106				
H5″	0.044			0.169				
H5′	0.057							
H8	0.035	0.059		0.036	0.014			

G2T3	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6
H1′				0.012				0.031
H2″						0.034		
H2′								
H3′								
H4′								0.011
H5″								
H5′								
H8								0.016

T3T3	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6
H1′								
H2″	0.124							
H2′	0.073	0.125						
H3′	0.046	0.065	0.082					
H4′				0.032				
H5″	0.036							
H5′		0.046						
H6		0.063	0.084					

T3A4	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8	H2
H1′								0.039	
H2″								0.047	
H2′									
H3′								0.012	
H4′									
H5″									
H5′									
H6									

A4A4	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8	H2
H1′									
H2″	0.125								
H2′	0.074	0.138							
H3′	0.043	0.085							
H4′	0.045	0.022		0.042					
H5″									
H5′	0.032								
H8	0.044	0.059	0.084	0.035	0.013				
H2							0.039	0.017	

A4T5	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6
H1′								0.038
H2″								0.041
H2′								0.033
H3′								0.020
H4′								
H5″								
H5′								
H8						0.059		0.020
H2	0.011							

T5T5	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6
H1′								
H2″	0.108							
H2′	0.066							
H3′	0.048		0.078					
H4′								
H5″								
H5′								
H6	0.036	0.056	0.084	0.052				

T5A6	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8	H2
H1′								0.034	
H2″								0.040	
H2′								0.026	
H3′								0.016	
H4′									
H5″									
H5′									
H6								0.013	

A6A6	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H8
H1′								
H2″	0.110							
H2′	0.076							
H3′	0.041	0.113						
H4′	0.043							
H5″								
H5′	0.019							
H8	0.041	0.070	0.106	0.050	0.015			
H2								

A6C7	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6	H5
H1′									0.023
H2″									0.026
H2′									0.018
H3′								0.019	0.001
H4′									
H5″									
H8								0.027	0.032
H2	0.013								

C7C7	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6	H5
H1									
H2″	0.128								
H2′	0.070	0.268							
H3′	0.042	0.062	0.076						
H4′	0.012			0.033					
H5″									
H5′				0.130					
H6	0.029	0.039	0.069	0.036					
H5	0.007		0.025	0.013			0.151		

C7C8	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6	H5
H1′								0.027	0.011
H2″								0.044	0.023
H2′								0.033	0.021
H3′								0.019	
H4′									
H5″						0.017			
H5′						0.033		0.023	
H6								0.011	0.015
H5									0.016

C8C8	H1′	H2″	H2′	H3′	H4′	H5″	H5′	H6	H5
H1′									
H2″									
H2′									
H3′	0.043								
H4′	0.080								
H5″	0.063								
H5′									
H6	0.041			0.052		0.020			
H5	0.011			0.010			0.151		

[a] The nomenclature used for $[d(GGTATACC)]_2$ in this table is G1G2T3A4T5A6C7C8. [b] Only nonoverlapping cross-peak intensities are listed.

(a) when they were not included in the structure, the ion distribution calculation gave the same ion positions as when they were included, for both B and BDB models; (b) when the B DNA plus ions structure and the BDB DNA plus ions structure were subjected to energy refinement, the resulting refined structures were similar to the refined structures including nine water molecules, with total interaction energy for energy-refined B at −1481 kcal/mol and energy-refined BDB at −1552 kcal/mol (see next section and Table III for results including nine water molecules). There is a larger interaction energy difference between these two models (71 kcal/mol) favoring the energy-minimized BDB structure compared with calculations including nine water molecules (14 kcal/mol, Table III). To provide some check on the effect of maintaining hydration in the minor groove, we used the calculation including water molecules for the interaction energy comparison listed in Table III. More sophisticated approaches for solvent effects, for example, including phosphate group hydration, base hydration, and ion hydration and including a larger number of water molecules, should be considered for further improvement. Before energy refinement, the BDB model has a smaller ADI at each mixing time than the B model. The energy-minimized B model not only gave a smaller ADI value than did the B model, but it was smaller than the initial BDB model. Among all, the energy-minimized BDB model has the smallest ADI at each mixing time. As noted before (Keepers & James, 1984; Young & James, 1984), 2D NOE cross-peak intensity is very sensitive to interproton distances; ADI values derived from a large number of cross-peaks can very well be used to distinguish more subtle structural differences among different models even when their qualitative NOE patterns are the same, as in this case.

The ADI of various intra- and interresidue cross-peaks of the six models examined are listed in Table II. They can give some indication about how well the different models fit the experimental data in various moieties of the DNA fragment. Another feature can be seen from Figure 1; i.e., the ADI values are more dispersed at short mixing times, while they start to approach each other for the 250- and 400-ms mixing time data. This may be because that spin–diffusion dominates at these longer mixing times. The complete relaxation matrix analysis we used does not rely on NOE cross-peak buildup rates, so the analysis can, in principle, be done with a single mixing time. However, for the present study we used the data for all four mixing times.

Energy-Minimized BDB Model. An analysis of the energy-minimized BDB model is given here in comparison with other B-type models. The results of our molecular mechanics calculation on the energy-minimized BDB and B models are summarized as group interaction energies in Table III, where each structural model is divided into three groups for analysis—DNA molecule, sodium ions, and water molecules. From Table III, the total energy of the energy-minimized BDB model is clearly lower than that of the energy-minimized B model. The intramolecular energy of DNA favors the energy-minimized B model by about 39 kcal/mol; however, the sodium–DNA interaction energy stabilizes the energy-minimized BDB model over the energy-minimized B model by 141 kcal/mol. The repulsion, not considering diminution from nearby phosphate groups, between the sodium ions is higher in the energy-minimized BDB model than in the energy-minimized B model. The interaction energy among water molecules and their interactions with DNA and sodium ions favor the energy-minimized B model over the energy-minimized BDB model by about 31 kcal/mol. The unfavorable

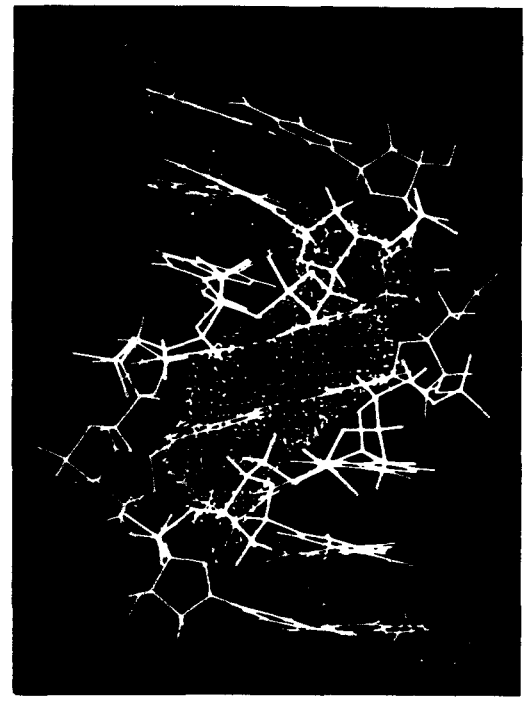

FIGURE 2: Structure of the energy-minimized BDB model. Note that the solvent-accessible surface of the minor groove for the -TATA- segment forms a closed hydration tunnel.

interactions in the energy-minimized BDB model are more than compensated by sodium ion–DNA interactions. When water molecules were not included in the models, the difference in interaction energy for energy-minimized BDB and energy-minimized B structures is larger, 71 kcal/mol, favoring the energy-minimized BDB structure (see also Comparison of Experimental and Calculated 2D NOE Spectra). We assumed a wD-DNA form for the middle TATA part in the BDB model. In the energy-minimized BDB model, the TATA part still maintains wD structural characteristics (Arnott et al., 1983; Suzuki et al., 1986): a very narrow and deep minor groove effectively creating a hydration tunnel. The $\frac{GG}{CC}$ moieties have a narrower minor groove than standard B DNA, but the surface remains open, so the hydration tunnel is not continued into the $\frac{GG}{CC}$ parts (Figure 2). The energy-minimized BDB model is shown as a stereoview pair, with all the Na^+ locatons, in Figure 3. More favorable electrostatic interactions between sodium ions and phosphate groups throughout the sequence and between interchain sugar rings and phosphate groups in the -TATA- moiety of the energy-minimized BDB model contribute to the stability of this model.

In Figure 4, the B, energy-minimized B, and energy-minimized BDB models are shown together for comparison. The view is perpendicular to the helix axes; the minor groove is at the right side, and the major groove is at the left side of each model. Energy refinement of the B model produces structural changes in the direction of the wD structure. It can be seen that the energy-minimized B model retains recognizable B structural features but has a narrower minor groove than standard B form, but is still wider than that of the energy-minimized BDB model. In the energy-minimized BDB model, the minor groove further narrows to form a closed channel in the -TATA- segment, as mentioned above. In addition, the major groove becomes more shallow than in the B form. We also note that including counterions in the energy refinement of the BDB model was necessary to stabilize the wD-type structure of the -TATA- segment; otherwise, the narrower minor groove would widen and the structure become

Table II: 2D NOE Cross-Peak ADI Values for Intra- and Interresidue Interactions of Six $[d(GGTATACC)]_2$ Structural Models[a]

mixing time (ms)	group interactions	structural models for $[d(GGTATACC)]_2$					
		A	crystalline A	B	energy-minimized B	BDB	energy-minimized BDB
100	G1	0.020	0.021	0.010	0.014	0.010	0.013
	G1G2	0.034	0.028	0.027	0.038	0.026	0.025
	G2	0.043	0.034	0.033	0.028	0.030	0.033
	G2T3	0.004	0.005	0.007	0.005	0.013	0.008
	T3	0.026	0.022	0.023	0.018	0.020	0.025
	T3A4	0.031	0.022	0.030	0.033	0.009	0.014
	A4	0.030	0.021	0.022	0.016	0.015	0.015
	A4T5	0.093	0.055	0.012	0.016	0.021	0.006
	T5	0.021	0.021	0.022	0.009	0.009	0.012
	T5A6	0.071	0.048	0.018	0.020	0.004	0.010
	A6	0.022	0.022	0.011	0.014	0.009	0.012
	A6C7	0.009	0.007	0.007	0.009	0.020	0.005
	C7	0.027	0.021	0.037	0.021	0.039	0.022
	C7C8	0.045	0.013	0.012	0.009	0.015	0.014
	C8	0.020	0.020	0.016	0.016	0.023	0.014
175	G1	0.024	0.026	0.015	0.013	0.014	0.012
	G1G2	0.034	0.021	0.019	0.040	0.017	0.015
	G2	0.054	0.045	0.043	0.035	0.035	0.038
	G2T3	0.037	0.010	0.006	0.018	0.016	0.012
	T3	0.028	0.033	0.018	0.018	0.018	0.027
	T3A4	0.042	0.068	0.022	0.027	0.007	0.012
	A4	0.026	0.020	0.019	0.013	0.013	0.011
	A4T5	0.041	0.034	0.005	0.009	0.011	0.005
	T5	0.026	0.040	0.020	0.012	0.009	0.012
	T5A6	0.069	0.058	0.024	0.029	0.004	0.014
	A6	0.025	0.025	0.010	0.015	0.008	0.011
	A6C7	0.015	0.009	0.010	0.016	0.020	0.007
	C7	0.040	0.032	0.046	0.027	0.040	0.028
	C7C8	0.055	0.017	0.017	0.014	0.015	0.021
	C8	0.020	0.017	0.014	0.018	0.022	0.014
250	G1	0.035	0.041	0.017	0.019	0.017	0.022
	G1G2	0.030	0.040	0.020	0.043	0.020	0.024
	G2	0.052	0.043	0.042	0.033	0.032	0.035
	G2T3	0.021	0.014	0.007	0.011	0.010	0.014
	T3	0.029	0.026	0.030	0.025	0.027	0.030
	T3A4	0.038	0.042	0.015	0.031	0.009	0.019
	A4	0.028	0.024	0.024	0.021	0.018	0.020
	A4T5	0.036	0.039	0.034	0.022	0.019	0.014
	T5	0.026	0.027	0.025	0.016	0.018	0.018
	T5A6	0.058	0.058	0.020	0.030	0.008	0.022
	A6	0.032	0.030	0.019	0.022	0.013	0.017
	A6C7	0.022	0.015	0.013	0.018	0.021	0.010
	C7	0.028	0.024	0.037	0.026	0.041	0.023
	C7C8	0.037	0.014	0.014	0.011	0.016	0.017
	C8	0.030	0.026	0.022	0.015	0.041	0.017
400	G1	0.033	0.037	0.020	0.016	0.020	0.015
	G1G2	0.026	0.031	0.007	0.044	0.010	0.017
	G2	0.064	0.051	0.052	0.046	0.042	0.041
	G2T3	0.015	0.013	0.012	0.015	0.020	0.017
	T3	0.028	0.026	0.024	0.022	0.029	0.030
	T3A4	0.027	0.034	0.015	0.025	0.013	0.021
	A4	0.029	0.027	0.026	0.022	0.019	0.022
	A4T5	0.023	0.030	0.011	0.018	0.014	0.012
	T5	0.027	0.026	0.023	0.019	0.023	0.021
	T5A6	0.042	0.052	0.051	0.033	0.010	0.027
	A6	0.027	0.027	0.020	0.021	0.020	0.020
	A6C7	0.028	0.016	0.022	0.026	0.015	0.010
	C7	0.031	0.026	0.034	0.028	0.038	0.022
	C7C8	0.038	0.019	0.024	0.017	0.026	0.021
	C8	0.040	0.029	0.028	0.015	0.050	0.021

[a] The nomenclature used for $[d(GGTATACC)]_2$ in this table is G1G2T3A4T5A6C7C8.

Table III: Group Interaction Energies of $[d(GGTATACC)]_2$ for Energy-Minimized B and Energy-Minimized BDB Models

	energy-minimized B (kcal/mol)	energy-minimized BDB (kcal/mol)
intramolecular energy of DNA	−506	−467
DNA–sodium interaction energy	−1222	−1363
sodium–sodium interaction energy	249	306
intra- and interaction energy of water molecules	−208	−177
total energy	−1687	−1701

more "B-like" during minimization, similar to that observed by Suzuki et al. (1986) in their study of the $[d(AT)_5]$ duplex. On the other hand, interstingly, including counterions in the energy minimization of the B model produced a more "wD-like" structure for the middle part of the duplx. These indicate that the wD structure of the -TATA- moiety is more stable in aqueous NaCl solution than the B structure.

Table IV lists average backbone and glycosidic torsion angles of B, wD, and energy-minimized B models and torsion angles for the $\overset{\text{-TATA-}}{\text{-ATAT-}}$ and the $\overset{\text{GG-}}{\text{CC-}}$ regions of the BDB and

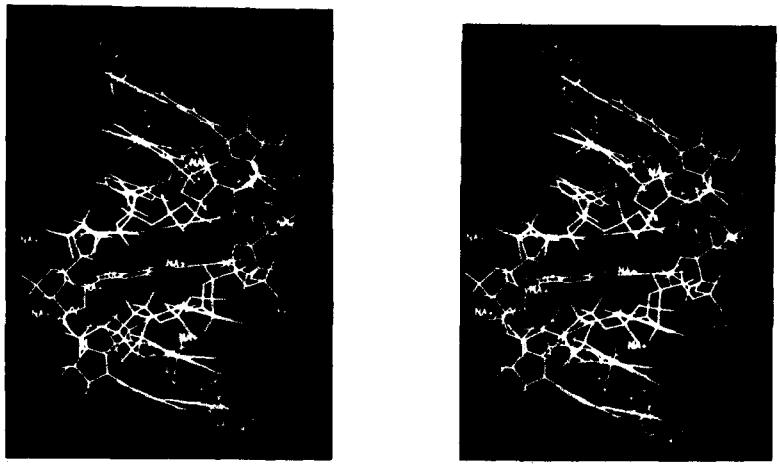

FIGURE 3: Stereo diagrams for the energy-minimized BDB model of [d(GGTATACC)]$_2$, shown with counterions (Na^+). The minor groove is oriented toward the reader. Water molecules are omitted for clarity.

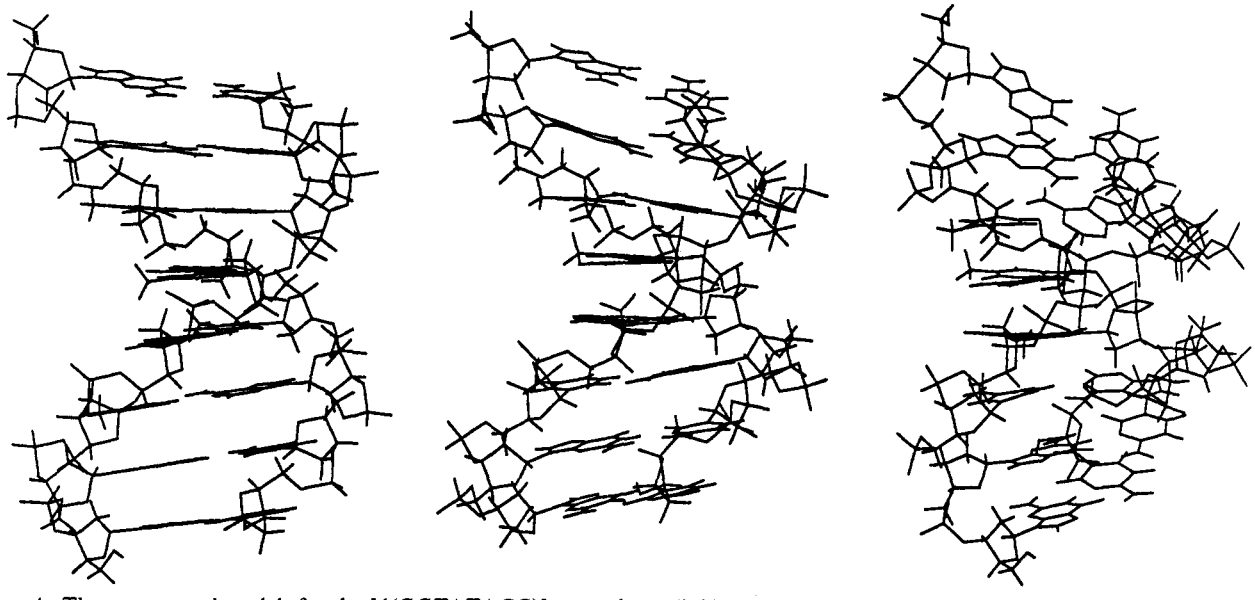

FIGURE 4: Three structural models for the [d(GGTATACC)]$_2$: regular B (left), energy-minimized B (middle), and energy-minimized BDB (right).

Table IV: Average Torsion Angles, Deoxyribose Ring Pseudorotational Parameters, and Helical Parameters of the Different Models for [d(GGTATACC)]$_2$

structure models	base sequence	α (P–O5′) (deg)	β (O5′–C5′) (deg)	γ (C5–C4′) (deg)	δ (C4′–C3′) (deg)	ϵ (C3′–C3′) (deg)	ξ (O3′–P) (deg)	χ^b (deg)	P (deg)	θ_m (deg)	base pairs per turn	propeller base twist (deg)
B	–	313	190	36	156	155	265	142	191	36	10.0	–4
wD	T-A	288	143	69	154	227	207	138	176	43		
											8.0	–21
	A-T	315	142	55	142	203	215	128	148	45		
BDB	G-G											
		314	214	37	153	159	265	138	180	35	10.2	–6
	C-C											
	Py-Py	288	142	71	139	224	207	138	172	45		
											8.3	–23
	Pu-Py	315	141	56	141	203	215	128	153	44		
energy-minimized BDB	G-G											
		292	171	56	124	185	269	124	140	40	11.1	–17
	C-C											
	Py-Pu	208	146	79	144	233	218	148	176	36		
											8.5	–25
	Pu-Py	286	149	63	108	209	224	119	128	43		
energy-minimized B	–	291	169	61	127	186	256	118	138	40	9.6	–12

[a] Torsion angles α, β, γ, ϵ, and ζ belong to the dinucleotide link indicated; all other parameters belong to the second nucleotide in the link. [b] Glycosidic torsion angles χ are defined by C2′–C1′–N1–C2 for pyrimidines and C2′–C1′–N9–C4 for purines.

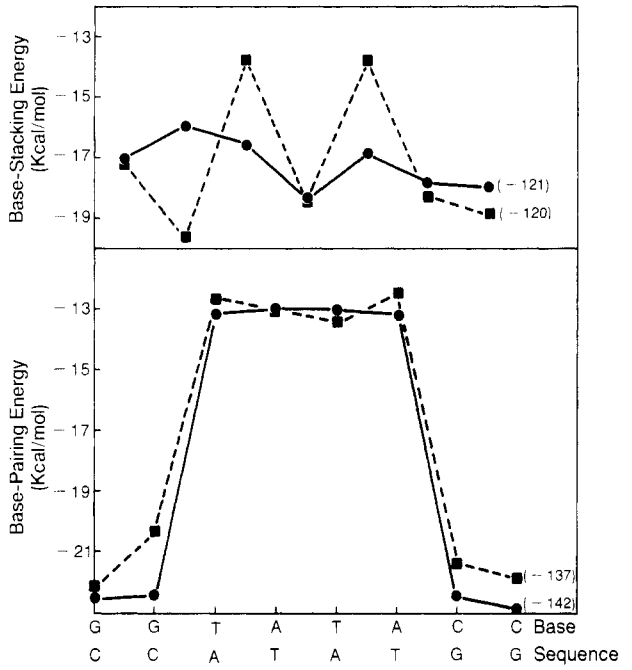

FIGURE 5: Base-stacking (top) and base-pairing (bottom) energies for energy-minimized B (●) and energy-minimized BDB (■) models. Shown on the abscissa are base pairs along the sequence. The numbers in the parentheses are total base-stacking and base-pairing energies for the two models, respectively.

energy-minimized BDB models. For the BDB and the energy-minimized BDB models, torsion angles from the B–wD junctions were included in the $\frac{\text{-TATA-}}{\text{-ATAT-}}$ part. This is because when building the BDB model, wD backbone structure was used for the junctions (see Model Building). Angles for pyrimdine–purine (T-A) and purine–pyrimidine (A-T, G-T, A-C) links were averaged separately and listed in Table IV to compare with those of T-A and A-T links of the wD structure. In energy-minimized BDB, each torsion angle varies in a different range along the sequence; but it is clear from Table IV that the average backbone torsion angles for the -TATA- segment in the energy-minimized BDB model stayed very close to those of wD form. The average backbone torsional angles for the $\frac{\text{GG-}}{\text{CC-}}$ parts were changed from those of the B structure, but were close to those of energy-minimized B, which is essentially still in B form. The energy-minimized B model has more uniform backbone torsion angles along the sequence. The average values of all the backbone angles were slightly changed from the B model values toward values more characteristic of wD. The transition from wD to B structures is apparently possible by changing only local backbone torsion angles. This was consistent with a helical parameter of 8.5 bp/turn for the -TATA- moiety, which is close to 8 bp/turn in wD-form structure, indicating a more tightly wound helical structure compared with B structure (10 bp/turn).

Also shown in Table IV are average propeller twist angles for the different models. In the energy-minimized BDB model, an average propeller twist of −25° was seen for the -TATA- segment, which is slightly larger than the value for wD form (−21°); the GC parts had an average value of −17°, slightly larger than the −12° twist in the energy-minimized B model. However, as plotted in Figure 5, the base pairing and base stacking were uninterrupted at the B–wD junctons. The total base-stacking energy between intrastrand neighboring bases in energy-minimized BDB (−120 kcal/mol) was almost the same as in energy-minimized B (−121 kcal/mol) but alternated for purine–pyrimidine and pyrimidine–purine sequences (Figure 5); this alteration is similar to the situation with the initial wD form. The total base-pair energy of the energy-minimized BDB model (−137 kcal/mol) was a little higher than that in the energy-minimized B model (−142 kcal/mol) (Figure 5), probably due to larger propeller twist angles.

Average deoxyribose ring pseudorotational phase angles P and maximum ring puckering θ_m [defined according to Altona and Sundaralingam (1972)] were calculated for the GC parts and the AT part of the BDB and energy-minimized BDB models, and for the whole sequence of the energy-minimized B model. These are listed together with B and wD structure pseudorotational parameters in Table IV. In the energy-minimized BDB model, the average phase angle P of the GC parts (140°) was very close to the P value of the energy-minimized B model (138°); the phase angles of T and A residues in the energy-minimized BDB model average to 176° and 128°, comparable to the respective P values in the wD model: 176° and 148°. The energy-minimized BDB model had average θ_m values of 40° for the GC parts, 43° for the T residues, and 36° for the A residues. These are between the maximum ring puckering values of the B and wD structures. All the deoxyribose sugar rings were still in the C2′-endo conformation region.

[d(GGTATACC)]$_2$ Solution Structure. Our energy-minimized BDB model yielded the closest fit of the experimental 2D NOE spectra of the octamer among the six structural models tested; it is also energetically most stable. This model can therefore be considered as a reasonable representation of the ensemble- and time-averaged structure of the octamer in solution. In this energy-minimized BDB model, backbone torsion angles of the residues are distributed over a certain range, but the average values in the -TATA- segment approximate those of wD DNA, and the average values in the $\frac{\text{GC-}}{\text{CC-}}$ ends are close to those of the energy-minimized B DNA.

The narrow minor groove of the TATA double-helical segment may be important for its interaction with TA-specific minor groove binding drugs, and the shallow major groove in addition to the minor groove may affect protein recognition in its functioning in the DNA promoter region of genes. The fact that the minor groove surface forms a closed tunnel in the middle part of the molecule suggests that minor groove drug binding involves opening of the minor groove first.

The junction between wD and B parts was built in a somewhat arbitrary way (see Model Building). Compared with the energy-minimized B model, the interresidue cross-peaks involving junctions (G-T and A-C) have only slightly higher ADI values in BDB but are lower with the energy-minimized BDB model (Table II). The base pairing and base stacking with the energy-minimized BDB are not interrupted at junctions as shown in Figure 5. This suggests that the junction structure in our energy-minimized BDB model is a feasible way for joining the $\frac{\text{-TATA-}}{\text{-ATAT-}}$ and $\frac{\text{GG-}}{\text{CC-}}$ parts in [d-(GGTATACC)]$_2$.

The energy-minimized B model gave an improved fit to the 2D NOE data compared with regular B DNA structure; this apparently results from change toward a more "BDB-like" structure during energy refinement and thus manifests a closer representation to the solution structure of [d(GGTATACC)]$_2$ than does regular B DNA. From our analysis, the stability of the wD structure in the middle alternating $\frac{\text{-TATA-}}{\text{-ATAT-}}$ moiety in aqueous NaCl solution is intrinsic to the sequence even when it is only four base pairs long.

A Raman spectroscopy study on [d(CGCGTATACGCG)]$_2$ in aqueous solution has been carried out by Nishimura et al. recently (personal communication, 1986). The authors observed two types of spectral lines in a frequency region sensitive

to backbone torsion angles at an intensity ratio of 2:1 and interpreted them as evidence for two types of B-form structures, which they named Ba and Bn′, in the molecule. The Ba structure is the same as that observed for poly[d(GC)], and Bn′ is the same as that observed for poly[d(AT)] solution structures by their Raman measurements of these samples. The central six base pairs of this molecule (-GTATAC-) are the same as in the octamer under present study; it is likely that the -TATA- segment of their oligonucleotide also exists in a wD type of structure. It is not clear to us now how the structure they assigned to the middle TA-rich region correlates with the wD structure we used in our study, but their observation of two types, and only two types, of Raman lines for [d(CGCGTATACGCG)]$_2$ is consistent with our conclusion that the transition between the two B-family structures (B and wD) in the backbone is abrupt.

Acknowledgments

We thank Dr. Nadege Jamin for obtaining the 2D NOE spectra that were reanalyzed in this work and Dr. Gerald Zon for providing [d(GGTATACC)]$_2$. The coordinates of the [d(GGTATACC)]$_2$ crystal structure were kindly provided by Dr. Olga Kennard and Dr. William Cruse. Dr. Ei-ichiro Suzuki is thanked for helpful discussion. We gratefully acknowledge our use of the University of California, San Francisco Computer Graphics Laboratory (Director, Dr. R. Langridge; supported by NIH Grant RR 01081).

Supplementary Material Available

One table of the normalized nonoverlapping cross-peak intensities at mixing times of 100, 175, and 400 ms (intensities were rounded to the 0.1% position for subsequent analysis) (11 pages). Ordering information is given on any current masthead page.

Registry No. [d(GGTATACC)]$_2$, 80407-93-8; TATA, 39862-54-7.

References

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.

Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93–105.

Arnott, S., Chandrasekaran, R., Puigjaner, L. C., Walker, J. K., Hall, I. H., Birdsall, D. L., & Ratliff, R. L. (1983) *Nucleic Acid Res. 11*, 1457–1474.

Blaney, J. M., Weiner, P. K., Dearing, A., Kollman, P. A., Jorgensen, E. C., Oatley, S. J., Burridge, J. M., & Blake, C. C. F. (1982) *J. Am. Chem. Soc. 104*, 6424–6434.

Broido, M. S., James, T. L., Zon, G., & Keepers, J. W. (1985) *Eur. J. Biochem. 150*, 117–128.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343–352.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983) *Biochemistry 22*, 5943–5951.

Gallo, L., Huang, C., Ferrin, T., & Langridge, R. (1985) *User's Manual, Molecular Interactive Display and Simulation (MIDAS)*, Computer Graphic Laboratory, University of California—San Francisco, San Francisco.

Haasnoot, C. A. G., Westerink, H. P., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 1*, 131–149.

Hopfinger, A. (1973) in *Conformational Properties of Macromolecules*, Academic, New York.

Jamin, N., James, T. L., & Zon, G. (1985) *Eur. J. Biochem. 152*, 157–166.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.

Lybrand, T. P., & Kollman, P. (1985) *Biopolymers 24*, 1863–1879.

Patel, D. J., Shapiro, L., & Hare, D. (1986) *Biopolymers 25*, 693–706.

Pattabiraman, N., Langridge, R., & Kollman, P. A. (1984) *J. Biomol. Struct. Dyn. 1*, 1525–1533.

Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., & Williams, D. H. (1983) *Biochemistry 22*, 2019–2023.

Shakked, Z., Rabinovich, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A., & Viswamitra, M. A. (1981) *Proc. R. Soc. London, B 213*, 479–487.

Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A., & Viswamitra, M. A. (1983) *J. Mol. Biol. 166*, 183–201.

Singh, U. C., & Kollman, P. A. (1984) *J. Comput. Chem. 5*, 129–149.

Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–6865.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature (London) 282*, 680–686.

Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.

Weiner, P. K., Singh, U. C., Kollman, P. A., Caldwell, J., & Case, D. A. (1984a) *A Molecular Mechanics and Dynamics Program*, University of California—San Francisco, San Francisco.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., & Weiner, P. K. (1984b) *J. Am. Chem. Soc. 106*, 765–784.

Weiner, S. J., Kollman, P. A., Case, D., & Nguyen, D. (1985) *J. Comput. Chem. 7*, 230–252.

Wemmer, D. E., Chou, S.-H., Hare, D. R., & Reid, B. R. (1984) *Biochemistry 23*, 2262–2268.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.

Young, G. B., & James, T. L. (1984) *J. Am. Chem. Soc. 106*, 7986–7988.