Solution Structure of the *Eco*RI DNA Octamer Containing 5-Fluorouracil via Restrained Molecular Dynamics Using Distance and Torsion Angle Constraints Extracted from NMR Spectral Simulations†

Ryszard Stolarski,‡,§ William Egan,‖ and Thomas L. James*,‡

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, and Biophysics Laboratory, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892

Received January 10, 1992; Revised Manuscript Received May 13, 1992

ABSTRACT: The self-complementary DNA octamer [d(GGAATU^{F}CC)]$_2$, containing the *Eco*RI recognition sequence with one of the thymines replaced by 5-fluorouracil (U^F), was synthesized. Proton homonuclear two-dimensional nuclear Overhauser effect (2D NOE) and double-quantum-filtered correlation (2QF-COSY) spectra, as well as one-dimensional spectra at different temperatures, were recorded for the octamer. Consequently, all proton resonances were assigned. The thermally induced transition from the duplex to single strands has been followed, demonstrating the stability of the duplex containing 5-fluorouracil. Simulations of the 2QF-COSY cross-peaks by means of the programs SPHINX and LINSHA were compared with experimental data, establishing scalar coupling constants for the sugar ring protons and hence sugar pucker parameters. The deoxyribose rings exhibit a dynamic equilibrium of N- and S-type conformers with 75–95% populations of the latter. Two programs used for complete relaxation matrix analysis 2D NOE spectra, CORMA and MARDIGRAS, were modified to account for the influence of the fluorines on dipolar interactions in the proton system. Quantitative assessment of the 2D NOE cross-peak intensities for different mixing times, in conjunction with the program MARDIGRAS, gave a set of interproton distances for each mixing time. The largest and smallest values of each of the interproton distances were chosen as the upper and lower bounds for each distance constraint. The distance bounds define the size of a flat-well potential function term, incorporated into the AMBER force field, which was employed for restrained molecular dynamics calculations. Torsion angle constraints in the form of a flat-well potential were also constructed from the analysis of the sugar pucker data. Several restrained molecular dynamics runs of 35 ps were performed, utilizing 284 experimental distance and torsion angle constraints and two different starting structures, energy-minimized A- and B-DNA. Convergence to similar structures with a root-mean-square deviation of 1.2 Å was achieved for the central hexamer of the octamer, starting from A- and B-DNA. The average structure from six different molecular dynamics runs was subjected to final restrained energy minimization. The resulting final structure was in good agreement with the structures derived from different molecular dynamics runs and showed a substantial improvement of the 2D NOE sixth-root residual index in comparison with classical and energy-minimized B-DNA. A detailed analysis of the conformation of the final structure and comparison with structures of similar sequences, obtained by different methods, were performed.

5-Fluorouracil (U^F)[1] is a thymine analog that is widely used for the treatment of gastrointestinal and other tumors; this application has been reviewed (Danenberg, 1977; Pinedo & Peters, 1988). In vivo, U^F is converted to 5-fluoro-2′-deoxyuridine 5′-monophosphate (F-dUMP), an inhibitor of thymidylate synthetase; in the absence of thymidine, DNA synthesis, and consequently cellular replication, is blocked. Additional cellular pathways allow F-dUMP to enter the nucleotide triphosphate pool and to be subsequently incorporated into DNA (Cheng & Nakayama, 1983; Major et al., 1982). The incorporation, in conjunction with the removal of U^F by uracil–DNA glycosylase (Ingraham et al., 1980), may prevent further elongation of DNA as well as produce strand breaks and DNA fragmentation (Lönn & Lönn, 1984; Schuetz & Diasio, 1985). Within DNA, U^F could serve as a site for point mutations by base pairing with guanine (Habener et al., 1988; Sowers et al., 1988), or it could disturb DNA transcriptional regulation through an alteration of DNA secondary structure. In addition to its effects on Watson–Crick hydrogen bonding, fluorine may also influence base stacking within a strand (for example, by changing the dipole moment of the base). This aspect of the substitution of H by F has recently been addressed in an X-ray crystallographic study (Coll et al., 1989): in a Z-DNA crystal, the stacking

† This work was supported by National Institutes of Health Grants GM 39247, CA 27343, and RR 01695. R.S. acknowledges the support of Polish Cancer Research Program CPBR-11.5 during the course of part of this work.

* Author to whom correspondence should be addressed. Telephone: (415) 476-1569. Fax: (415) 476-0688.

‡ University of California, San Francisco.

§ Permanent address: Institute of Experimental Physics, Department of Biophysics, University of Warsaw, 02-089 Warsaw, Poland.

‖ Center for Biologics Evaluation and Research.

[1] Abbreviations: U^F, 5-fluorouracil; F-dUMP, 5-fluoro-2′-deoxyuridine 5′-monophosphate; 1D NMR, one-dimensional nuclear magnetic resonance; 2D NMR, two-dimensional NMR; 2D NOE, two-dimensional nuclear Overhauser effect; COSY, correlation spectroscopy; 2QF-COSY, double-quantum-filtered COSY; rMD, restrained molecular dynamics; TLC, thin-layer chromatography; TEAB, triethylammonium bicarbonate; EGTA, [ethylenebis(oxyethylenenitrilo)]tetraacetic acid; FID, free induction decay; rMIN, restrained energy minimization; rms, root mean square; TSP, sodium 2,2,3,3-tetradeuterio-3-(trimethylsilyl)propionate; MD-A, structure resulting from rMD calculations using A-DNA as the starting model; MD-B, structure resulting from rMD using B-DNA as the starting model; MD$_{AV}$, final structure, resulting from rMD calculations after averaging different runs; avd, averaged deviation from the median of the flat well.

0006-2960/92/0431-7027$03.00/0

interaction of U^F with C was modified relative to that of T with C. Additionally, these authors saw no difference between U^F·G and T·G base pairing.

Nuclear magnetic resonance techniques, especially two-dimensional NMR (2D NMR) methods such as two-dimensional nuclear Overhauser effect (2D NOE) and correlation spectroscopy (COSY), are among the most powerful tools for investigation of the structure and dynamics of oligodeoxynucleotides in solution (Wüthrich, 1986), as well as their interactions with drugs (Zhou et al., 1989) and proteins (Lamerichs et al., 1989). This was made possible with the development of an array of computational methods for analyzing NMR data, namely, distance geometry (Havel & Wüthrich, 1985), complete relaxation matrix analysis (Borgias & James, 1988, 1989; Keepers & James, 1984), iterative relaxation matrix analysis (Boelens et al., 1988, 1989; Borgias & James, 1989, 1990; Madrid et al., 1991; Post et al., 1990), correlation spectroscopy cross-peak simulations (Widmer & Wüthrich, 1987), and restrained molecular dynamics simulations with constraints derived from NMR (Baleja et al., 1990; de Vlieg et al., 1986; Gochin & James, 1990; Nilsson et al., 1986). In order to better understand the role of U^F when incorporated into cellular DNA or to evaluate its potential role in other constructs, such as triplexes or antisense oligonucleotides, a detailed NMR study of the U^F-containing DNA octamer of the *Eco*RI recognition sequence (subsequently referred to as octamer) was performed:

		1	2	3	4	5	6	7	8		
5'	-	G1	G2	A3	A4	T5	U^F6	C7	C8	-	3'
3'	-	C8	C7	U^F6	T5	A4	A3	G2	G1	-	5'
		16	15	14	13	12	11	10	9		

This sequence was chosen for study because the parent sequence has been characterized by NMR spectroscopy (Broido et al., 1985; Nerdal et al., 1989), the *Eco*RI recognition sequence incorporated in dodecamers has been determined via X-ray diffraction (Larsen et al., 1991; Narayana et al., 1991), and the interaction of U^F with C, which was noted to be unusual in an X-ray crystallographic study, is preserved. Simple analysis of NMR data from the octamer can provide some general information about its structure in solution (helix type), and additional effort yields insights about conformational features of some parts of the molecule, e.g., sugar pucker. However, most of the subtle structural features of interest are not deduced in this way. A complete relaxation matrix analysis of 2D NOE spectra, employed in the MARDIGRAS iterative algorithm (Borgias & James, 1989, 1990), yields a large number of accurate interproton distances because, contrary to the commonly-employed isolated spin-pair approximation, spin diffusion is taken into account and longer mixing times can be utilized. However, a set of interproton distances, even with supplemental information from the analysis of sugar conformations, does not alone define a structure. Rather, these NMR-derived distances and sugar puckers constitute a set of structural constraints. Restrained molecular dynamics (rMD) calculations allow conformational space to be searched for structures which simultaneously satisfy the NMR-derived structural constraints and correspond to a reasonable potential energy, as given by an analytical potential or force field. A major drawback of this approach and of almost all refinement strategies currently employed is the interpretation of the derived structure in terms of only one rigid conformer, which is contradictory to the dynamic nature of DNA molecules. Nevertheless, our approach is justified in defining a time-averaged structure which is usually dominated by one conformer in solution.

MATERIALS AND METHODS

Chemical Syntheses

5'-O-(Dimethoxytrityl)-5-fluorodeoxyuridine. 5-Fluorodeoxyuridine was purchased from Sigma. Other reagents, unless noted otherwise, were purchased from Aldrich. 5-Fluorodeoxyuridine (1.39 g; 6 mmol) was dried by coevaporation with dry pyridine and then dissolved in pyridine (10 mL). Recrystallized dimethoxytrityl chloride (2.1 g; 6.2 mmol) was added at room temperature, and the progress of the reaction was monitored by TLC (benzene–CH_2Cl_2–MeOH = 1:1:0.1). After 2 h of reaction, 1 mL of MeOH was added and the reaction allowed to proceed for an additional 30 min. Solvents were evaporated to dryness, and the oily residue was dissolved in CH_2Cl_2, extracted with water (2 × 2 mL), dried with $MgSO_4$, filtered, evaporated, and purified by column chromatography. Column chromatography conditions were as follows: a 2.5 × 17 cm column of silica gel (Merck 230–400 mesh) eluting with a 1:1 benzene–CH_2Cl_2 mixture followed by a 1:1:0.2 mixture of benzene–CH_2Cl_2–MeOH (chromatography solvent mixtures should contain ca. 0.2% pyridine to protect the dimethoxytrityl group). The desired fraction (test fractions placed on a TLC plate developed an orange color when exposed to HCl vapors) was evaporated and dried under high vacuum to yield a heavy oil (2.4 g; 75% yield).

5'-O-(Dimethoxytrityl)-5-fluorodeoxyuridine 3'-O-(2-Cyanoethyl N,N-diisopropylphosphoramidite). 5'-(Dimethoxytrityl)-5-fluorodeoxyuridine (500 mg; 0.9 mmol) was added to 2 mL of CH_2Cl_2 under anhydrous conditions; diisopropylethylamine (472 mg; 636 mL; 3.6 mmol) was added to the mixture followed by 2-cyanoethyl *N,N*-diisopropylphosphoramidochloridite (230 mg; 0.9 mmol) with vigorous stirring at room temperature. The reaction mixture was allowed to stand at room temperature for 2 h. The reaction mixture was then added to a short silica gel column (1.2 × 5 cm) prepared with a benzene–CH_2Cl_2–Et_3N mixture (5:5:1); the phosphoramidite was eluted from the column with a CH_2Cl_2–Et_3N mixture (10:1). The desired fraction was collected, evaporated to a heavy oil, and then placed under high vacuum overnight to remove residual solvent, yielding 600 mg of phosphoramidite (86% yield). The ^{31}P NMR spectrum of the phosphoramidite consisted of two signals, of approximately equal intensity, at 111.4 and 111.7 ppm (relative to external trimethyl phosphate); no other signals were observed.

5'-GGAATU^{F}CC. Solid-phase oligonucleotide synthesis was carried out on an Applied Biosystems Inc. Model 380B DNA synthesizer using standard β-cyanoethyl chemistry according to the manufacturer's protocol. All reagents used in the synthesis, with the exception of U^F phosphoramidite, were purchased from Applied Biosystems. The base-deprotected oligonucleotide bearing the dimethoxytrityl group was purified by reverse-phase HPLC using a Hamilton PRP-1 column (10 mm × 270 mm) under the following conditions: a linear gradient of 5–27.5% acetonitrile in 0.1 M triethylammonium bicarbonate (TEAB), pH 7.4, for 20 min followed by a 27.5–50% gradient of acetonitrile in TEAB for 5 min (flow rate of 2.5 mL/min). The purified oligomer was detritylated using 80% acetic acid (30 min), dried, and rehydrated in water. The solution was extracted three times with an equal volume of ethyl acetate. The aqueous layer was evaporated to dryness under vacuum (Savant) and repurified by reverse-phase HPLC using a Hamilton PRP-1 column under the following conditions: a 0–17.5% gradient of acetonitrile in 0.1 M TEAB for 20 min followed by a 17.5–50% gradient of acetonitrile in 0.1 M TEAB for 5 min (flow rate

of 2.5 mL/min). The oligonucleotide counterion, the triethylammonium ion, was exchanged for Na^+ as described previously for phosphoric acid diester-linked polysaccharides (Egan et al., 1982).

NMR Spectroscopy

Sample Preparation. For NMR measurements in H_2O, an amount corresponding to 60 absorbance units (ca. 3 mg) of the octamer was dissolved in 0.3 mL of 0.1 M NaCl, 10 mM phosphate buffer, pH 7.0, 0.1 mM EGTA with a trace amount of sodium azide, and 10% ^2H_2O for NMR spectrometer locking; the DNA concentration was 1.7 mM. For measurements in ^2H_2O, the sample was lyophilized three times from 99.95% ^2H_2O and finally dissolved in 99.995% ^2H_2O (Stohler). All spectra, except for those establishing the temperature dependence in H_2O, were run at a temperature of 15 °C.

NMR Measurements. All ^1H NMR experiments were performed at 500 MHz on a General Electric GN500 equipped with an Oxford Instruments magnet and Nicolet 1280 computer. One-dimensional NMR spectra in H_2O were obtained using a hard 1 3 $\bar{3}$ $\bar{1}$ pulse for water suppression (Hore, 1989).

Pure absorption 2D NOE spectra in ^2H_2O were carried out at three different mixing times, 100, 150, and 200 ms, employing the standard three-pulse sequence and the States' method of phase cycling (States et al., 1982). With alternating block acquisition, 400 free induction decays (FID), 16 scans each, were collected in the t_1 dimension with 4K data points in the t_2 dimension and a spectral width of 5400 Hz. A delay of 10 s between each scan was used to permit relaxation; this is essential for quantitative use of 2D NOE peak intensities. All data were transferred to a Sparc workstation for processing with locally written software. The spectra were apodized in both dimensions using Gaussian filter functions. The final data matrix consisted of 2K × 2K real points after zero-filling in the t_1 dimension. Base-line correction in both dimensions was applied after the second Fourier transformation. NOE cross-peaks were integrated using a modified version of the program CONTOUR (Gochin et al., 1990).

The 2D NOE spectra in H_2O were recorded using the three-pulse sequence with the third pulse replaced by the 1 3 $\bar{3}$ $\bar{1}$ water suppression pulse and a 10-ms homogeneity spoil gradient pulse at the beginning of the 120-ms mixing period (Boelens et al., 1985). Other experimental conditions were the same as for 2D NOE spectra in ^2H_2O (vide supra).

Measurements of proton spin–lattice relaxation times T_1 were performed by the inversion–recovery method, and measurements of spin–spin relaxation times T_2 were made using the Hahn spin-echo method.

Pure absorption double-quantum-filtered COSY (2QF-COSY) spectra were recorded by using time-proportional phase incrementation (Marion & Wüthrich, 1983). A spectral width of 5400 Hz in both dimensions was used, with 800 increments in the t_1 dimension, 32 scans each, a delay time between scans of 6 s, and 4K data points in the t_2 dimension. A square sine-bell filter function was used for data processing in both dimensions, shifted by 30° in t_2. The final data matrix consisted of 1K × 4 K data points, with a digital resolution of 1.3 Hz in ω_2 and 5.3 Hz in ω_1. Similar to 2D NOE spectra, base-line corrections were applied.

Model Building

The coordinates (and molecular topology) of the initial model octamer, used by NMR-analyzing programs, by restrained molecular dynamics, and by restrained molecular mechanics, were obtained from the standard suite of the AMBER program package (Pearlman et al., 1991), starting from canonical A-DNA (Arnott & Hukins, 1972) and B-DNA (Arnott & Hukins, 1973). The electrostatic charges for the new residue U^F were taken from the ab initio (STO-3G) optimized structure (Singh & Kollman, 1984) of 1-methyl-5-fluorouracil. The electrostatic point charges were fit to this structure at the 6–31G* level and were then adjusted to produce the AMBER residue consistent with the DNA model under consideration. Bond lengths and angles were taken from the quantum-optimized structure and were close to those from crystallography (Harris & Macintyre, 1964). Force constants were taken from existing values in the AMBER database when applicable, and the new terms needed for the fluorine-centered interactions were fit from existing MM2 parameters available in the literature (Burkert & Allinger, 1982). To test the validity of the new parameters, normal mode analysis was performed using the AMBER module NMODE. To mimic counterion effects, large sodium ions (hexahydrated, radius 5 Å) were added (Singh et al., 1985). Sodium ions were placed along the PO_2^- bisection, 6 Å away from the phosphorus, and were free to move during minimization and molecular dynamics procedures. In fact, they moved very little during the course of the trajectories. Energy minimizations of the starting structures were performed in vacuo by means of the MINMD module of AMBER via combined steepest descent and conjugate gradient methods.

NMR Data Analysis

Simulation of 2QF-COSY Cross-Peaks. SPHINX and LINSHA programs (Widmer & Wüthrich, 1987) were used to simulate 2QF-COSY cross-peaks for the deoxyribose spin system H1′, H2′, H2″, H3′, H4′ and the ^{31}P spin-coupled to H3′. SPHINX calculates stick spectra, treating all proton spins as strongly coupled nuclei. Line shapes, incorporating effects from digital resolution, apodization functions, truncations of FIDs, and natural line widths, were added to the stick spectra using LINSHA. Experimental cross-peaks were compared with simulated cross-peaks for many different sets of coupling constants and line widths. Conformations of the deoxyribose rings in the octamer were determined from the compilation of proton–proton coupling constants (Rinkel & Altona, 1987).

Extraction of Distance Constraints from 2D NOE Spectra. Interproton distances were calculated from 2D NOE cross-peak intensities using the program MARDIGRAS (Borgias & James, 1989, 1990); the version of the program utilized enabled more accurate distances involving methyl protons to be calculated (Liu et al., 1992). The program was modified to account for fluorine atoms. MARDIGRAS gave a set of interproton distances for each 2D NOE spectrum (three different mixing times: 100, 150, and 200 ms), using energy-minimized A- and B-DNA as starting models. Protons were added and their coordinates established with these starting models using the program NEWHYD, which was originally written here by Dr. N. Pattabiraman. An isotropic correlation time $\tau_c = 2$ ns was used for all protons. This value was derived from the equation relating the experimentally measured T_1 and T_2 relaxation times for different protons and was then averaged (Suzuki et al., 1986). All correlation times measured were in the range of 2.0 ± 0.5 ns. MARDIGRAS estimates lower and upper bounds to the calculated distances, on the basis of the agreement between experimental and converged MARDIGRAS cross-peak intensities and on the experimental signal-to-noise ratio (Borgias & James, 1990). For subsequent rMD

calculations, the largest value from the six distance sets (three mixing times and two starting structures) established the upper bound, and the smallest value from the six distance sets established the lower bound distance for any particular proton pair. The upper and lower bound values were used to determine the size of the flat well for the 204 distance constraints in the subsequent rMD calculations.

Simulation of 2D NOE Spectral Intensities. Cross-Peak intensities were simulated for 2D NOE spectra corresponding to various $[d(GGAATU^FCC)]_2$ structures during the process of structure refinement using the program CORMA (Borgias & James, 1988). As in the case of MARDIGRAS calculations, an overall tumbling correlation time of 2 ns was employed. CORMA also enabled a comparison of the simulated cross-peak intensities with experimental intensities and calculation of a residual index (vide infra).

Restrained Molecular Dynamics

The restrained molecular dynamics calculations were performed in vacuo using the SANDER module of AMBER Version 4.0 on a Sparc workstation (Pearlman et al., 1991). The pseudo-energy terms in the AMBER force field for the NOE-derived interproton distances and the 2QF-COSY-derived torsion angles have the form of flat wells with parabolic sides within 2 Å or 5°, respectively, from the flat part of the potential well and continue linearly beyond these margins. The width of the flat part for each proton pair arises from the accuracy of MARDIGRAS and SPHINX/LINSHA simulations. Four categories of interproton distance flat-well widths (0.00–0.25, 0.25–0.50, 0.50–1.00, and >1.00 Å) were respectively assigned force constants 100%, 70%, 40%, and 25% of the maximum during each stage of the rMD protocol (vide infra). The torsion angle flat wells were assigned 100% force constants. There were 204 distance constraints and 80 torsion angle constraints.

In addition, 40 constraints (20 distances between proton donor and acceptor and 20 flat angles determined by the three atoms forming a hydrogen bond) were included to guarantee maintenance of Watson–Crick hydrogen bonds. For a G·C base pair, lower and upper margins of a flat well were set to 2.81–3.01 Å (G-O6 to C-N4), 2.85–3.05 Å (G-N1 to C-N3), and 2.76–2.96 Å (G-N2 to C-O2); for A·T and A·U^F base pairs, values were 2.72–2.92 Å (A-N1 to T-N3 or U^F-N3) and 2.85–3.05 Å (A-N6 to T-O4 to U^F-O4) (Saenger, 1984). The force constants for the H-bond restraint were always maintained at 100% of maximum, but the force constants for the hydrogen bond angle constraints were 50% of maximum. This, together with a softer shape of the flat well, allowed more flexibility for orientation of complementary bases in the octamer.

The rMD runs consisted of 35 000 steps in 0.001-ps increments. All atoms within a 35-Å radius were included in nonbonded interactions. SHAKE was used to constrain all bonds (Ryckaert et al., 1977), and translational and rotational motions were removed every 100 steps. The energy-minimized coordinates of A-DNA and B-DNA were chosen as starting points. Initial velocities were taken from a Maxwellian distribution at 0.2 K. The system was heated gradually from 1 to 600 K between steps 500 and 5000 and maintained at 600 K for 12 000 steps, to avoid getting trapped in a local energy minimum. From step 17 000 to step 20 000, the temperature was decreased to the final value of 300 K. The weights of the NMR and hydrogen bond constraints were modulated by multiplying the force constants by a scaling factor. Up to step 1000, constraints were turned off. From step 1000 to step 4000 (before the system reached the highest temperature), the maximal (100%) restraint force constants were increased gradually to 150 kcal/mol·Å^2 and 150 kcal/mol·rad^2 and kept constant to step 22 000. During the subsequent 4000 steps, the force constants were reduced to a maximum of 20 kcal/mol·Å^2 and 20 kcal/mol·rad^2 as final values. For each rMD run of 35 ps, coordinate sets were recorded each 0.2 ps. The last 20 coordinate sets, arising from the last 4 ps of each run, were averaged and subjected to a restrained energy minimization (rMIN) using the SANDER module, with maximal force constants of 20 kcal/mol·Å^2 and 20 kcal/mol·rad^2. The final structure of the octamer was obtained by averaging coordinates of six structures obtained from rMD runs with slightly different protocols and A- and B-DNA as starting models. The averaged structure was subsequently subjected to a restrained energy minimization.

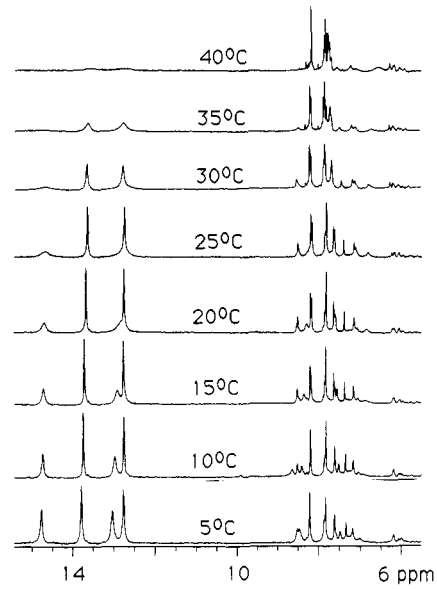


FIGURE 1: Temperature dependence of the imino and aromatic proton resonances of $[d(GGAATU^FCC)]_2$. For assignments, see Table I.

Structure Analysis and Display

Sugar pucker parameters and torsion angles of all structures were determined using the ANAL module of AMBER, and the helical parameters were calculated with the program CURVES (Lavery & Sklenar, 1988, 1990). All structures were displayed using MIDASplus on an Iris workstation (Gallo et al., 1989).

RESULTS AND DISCUSSION

Resonance Assignments

Assignment of exchangeable protons was made by analysis of cross-peaks in the 2D NOE spectrum in H_2O as previously described (Boelens et al., 1985) and confirmed by running 1D spectra in H_2O at different temperatures from 5 °C to the melting point (Figure 1). Both 1D and 2D spectra exhibit four imino protons of U^F6, T5, G1, and G2. The amino groups of C7 and C8 gave two signals each, the downfield ones arising from the protons involved in hydrogen bonding with the complementary guanines. The signals of guanine amino protons are very broad and those of adenines not visible, probably due to rotation of amino groups with rates intermediate on the NMR time scale (Boelens et al., 1985; Stolarski et al., 1987).

Assignment of all nonexchangeable protons was made in a sequential manner from 2D NOE and 2QF-COSY spectra in ^2H_2O following the strategy thoroughly described in earlier

Table I: Chemical Shifts[a] (in ppm versus Internal TSP) of All Protons in [d(GGAATU^{F}CC)]$_2$ at 15 °C

	G1	G2	A3	A4	T5	U^{F}6	C7	C8
H2			7.380	7.640				
H5							5.785	5.745
H6					7.185	7.875	7.645	7.640
H8	7.835	7.835	8.210	8.185				
CH_3					1.295			
H1′	5.615	5.440	6.080	6.200	5.975	6.185	6.065	6.250
H2′	2.470	2.665	2.770	2.585	2.140	2.185	2.225	2.280
H2″	2.665	2.750	2.995	2.985	2.570	2.680	2.515	2.280
H3′	4.815	5.015	5.110	5.040	4.895	4.925	4.900	4.580
H4′	4.170	4.345	4.500	4.510	4.280	4.290	4.205	4.050
H5′	3.650	4.050	4.175	4.320	4.365	4.240[b]	4.230[b]	4.080
H5″	3.650	4.130	4.240	4.280	4.200	4.175[b]	4.150[b]	4.190
NH	12.920	12.760			13.710	14.720		
NH_2	c	6.850	c	c	c	c	8.510	8.370
							7.140	7.050

[a] Accuracy ±0.005 ppm. [b] Reverse assignment for H5′ and H5″ possible. [c] Signal not detected.

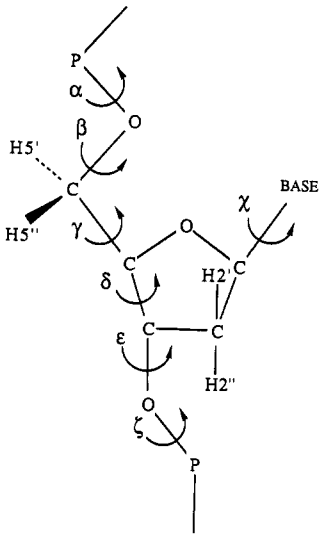


FIGURE 2: Definitions of the protons in the geminal parts H2′, H2″ and H5′, H5″ and of the torsion angles in the nucleotide units.

work (Broido et al., 1984; Feigon et al., 1983; Hare et al., 1983). Adenine H2 resonances were assigned from the strong cross-peaks to the imino protons of complementary thymines and relatively weak cross-peaks to H1′ of the adjacent 3′-nucleotides on the same strand. H5′ and H5″ were unambiguously assigned from the cross-peaks with H1′ of the adjacent 5′-neighboring nucleotide and/or intensity differences of the cross-peaks with H3′ of the same unit, except for U^{F}6 and C7.

The chemical shifts of all assigned protons at a temperature of 15 °C are reported in Table I. Each resonance represents two protons because of the symmetry of the helix. The unambiguous definitions of the protons in H2′, H2″ and H5′, H5″ geminal pairs are given in Figure 2.

General Conformational Features

Qualitative analysis of the 1D and 2D NMR data gives information about the overall structure of [d(GGAATU^{F}CC)]$_2$ in solution. The characteristic regular pattern of connectivities between H6, H8 and H2′, H2″ of the same and 5′-neighboring nucleotide as well as the signals of all imino protons points to the domination of a B-DNA-type double helix in aqueous solution. In particular, the modified nucleotide U^F behaves similarly to thymine, forming hydrogen bonds with the complementary adenine, with U^{F}6-N3H situated most downfield of all imino signals (Figure 1). Similar base pairing with the fluorouracil base has been already observed (Kremer et al., 1987; Sowers et al., 1987). The circular dichroism spectrum (not presented in detail) of [d(GGAATU^{F}CC)]$_2$ is also consistent with a B-DNA geometry.

The fluorouracil imino signal is broader than other imino resonances except for the first guanine, engaged in the fast exchange due to fraying. Intensities of exchange cross-peaks with water provide information on the relative exchange rates of imino protons (Dobson et al., 1986). It follows from the 2D NOE spectrum (see supplementary material) that the exchange rate of the U^{F}6-N3H is close to that of G1-N1H and almost two times greater than G2-N1H. This in turn is more than two times greater than the exchange rate of T5-N3H in the middle of the helix. Faster exchange of U^{F}6-N3H in comparison to other imino protons inside the helix does not arise from lower stability of the helix in this region. With increasing temperature the signal, although broad, remains intact almost up to the melting point of about 40 °C. Imino proton exchange with water is a two-step process (Teitelbaum & Englander, 1975): base pair opening with subsequent base-catalyzed proton transfer. The transfer rate depends on the difference between the pK values of the proton donor and the base catalyst (Kochoyan et al., 1987), phosphate buffer in our case. The pK of F-dUMP is 8.15 (Danenberg, 1977), in comparison with 9.9 for T. This makes the transfer rate of U^{F}6-NH3 greater than those of other imino protons.

Although one helical conformation of B-type prevails in solution for [d(GGAATU^{F}CC)]$_2$, several observations suggest the presence of at least one other conformer in dynamic equilibrium with the dominant one. In addition to a fast exchange between N and S forms of the deoxyribose rings of all nucleotides (vide infra), a slower exchange with a contribution from at least one minor global conformer is observed. Detailed inspection of 2D NOE spectra obtained in ^2H_2O reveals weak redundant cross-peaks. Their location as symmetrical pairs on both sides of the diagonal in the 2D NOE spectra, recorded with different mixing times, requires us to treat them as real cross-peaks and not artifacts. The characteristic weak connectivity between H2′, H2″ of C8 and H1′ of G1 on the opposite strand (too far apart from one another to be observed in the helical structure) can be attributed to fraying effects and more pronounced unwinding at the ends of the helix. Alternatively, such a cross-peak could arise from end-to-end stacking of the oligomer as well.

Some weak cross-peaks, typical for exchange, are clearly seen near the diagonal in regions which are not obscured by strong dipolar cross-peaks, e.g., aromatic and H1′ regions. The exchange connectivities are manifest in signals from the

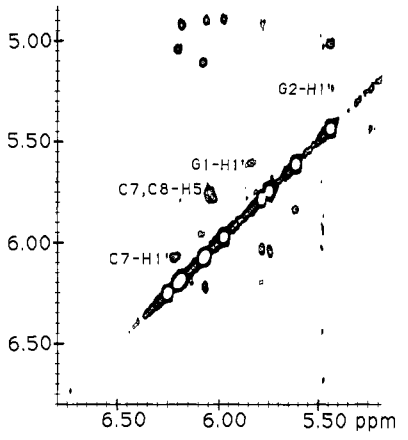


FIGURE 3: Part of the 2D NOE spectrum (mixing time 200 ms) showing some of the redundant cross-peaks due to exchange, marked with labels. Some of the dipolar H1′/H3′ cross-peaks are shown for comparison in the upper part.

nucleotides close to the ends of the helix: H1′ of G1, G2, and C7 and H5 of C7 and C8 (Figure 3). The resonances of the minor species are obscured by the strong signals of the dominating species, either on the diagonal of 2D spectra or in 1D spectra. The exchange behavior and salt dependence of signals emanating from the ends of the octanucleotides suggest that the minor conformer, with population always less than 20%, is probably some kind of helix maintained by two to three base pairs in the middle and with unwound ends; the population of the minor conformer decreases with salt concentration.

It is difficult to establish whether the presence of fluorines, especially with the ionized forms of the nucleotide, gives rise to helix destabilization and formation of the minor conformer. Double-helical structure prevents ionization of fluorouracil not only in the case of a limited number of U^F·A base pairs but also in oligomers like poly(U^F)·poly(A) (Szer & Shugar, 1963). In our case, stability of the octamer does not seem to be changed by replacement of a T·A base pair by U^F·A. On the other hand, thermodynamic data indicate that U^F·A and U^F·G base pairs are more stable than T·A and T·G, respectively (Habener et al., 1988), contrary to the observations of Kremer et al. (1987), who postulated lower stability of base pairs containing fluorouracil.

Determination of Deoxyribose Pucker

The well-established relationship between proton–proton vicinal coupling constants and dihedral angles, known as the modified Karplus relationship (Rinkel & Altona, 1987), permits fairly accurate determination of deoxyribose conformation. With accurate values of five vicinal coupling constants, $J(1',2')$, $J(1',2'')$, $J(2',3')$, $J(2'',3')$, and $J(3',4')$, it is possible to determine either the rigid conformation described by two parameters, the pseudorotation phase angle P and the pucker amplitude ϕ_m (Altona & Sundaralingam, 1972), or P and ϕ_m for the two N and S conformers in dynamic equilibrium with their percentage populations (Rinkel & Altona, 1987).

It is generally difficult to extract numerical values directly from J-couplings in oligonucleotides using COSY antiphase cross-peaks due to large inherent line widths—usually larger than the coupling constants. Additionally, digital resolution of COSY spectra generally exceeds 1 Hz in ω_2 and 5 Hz in ω_1. Several procedures, both experimental, e.g., E-COSY (Griesinger et al., 1985), and analytical, e.g., the DISCO method (Kessler et al., 1985), have been developed to extract coupling constants of macromolecules from COSY spectra. The best method at present, which gives coupling constants with reasonable accuracy (ca. 0.5 Hz), relies on comparison of experimentally detected COSY cross-peaks with simulations by means of programs such as SPHINX/LINSHA (Widmer & Wüthrich, 1987). This approach has been successfully applied to determine sugar pucker in oligodeoxynucleotides even when the line width exceeds the value of the coupling constants (Celda et al., 1989; Gochin et al., 1990; Schmitz et al., 1990).

The 2QF-COSY cross-peaks used for establishing conformations of the deoxyribose ring in the octamer are presented in Figure 4. H2′/H2″ and H3′/H4′ cross-peaks were not used since they change negligibly with conformational changes. Only H2″-ω_1/H1′-ω_2 cross-peaks exhibit 16 lines, as expected for one active coupling, $J(1',2'')$, and three passive couplings, $J(1',2')$, $J(2',2'')$, and $J(2'',3')$. In other cross-peaks, mutual cancellation of peaks with different phases yields a smaller number of peaks. Lack of H3′-ω_1/H2″-ω_2 cross-peaks, except for a weak one in C7, points to a strong preference for the S-type conformation typical of B-DNA (active coupling constants <3.5 Hz).

To determine precisely the deoxyribose conformation, SPHINX/LINSHA was employed to simulate 2QF-COSY cross-peaks (Widmer & Wüthrich, 1987). Component separations within cross-peaks, detailed inspection of cross-peak patterns, appearance of additional small signals on both sides of cross-peaks in the ω_1 dimension due to FID truncation effects (Widmer & Wüthrich, 1987), and cross-peak intensities were considered simultaneously for different sets of coupling constants and line widths (5–11 Hz) in calculating the theoretical spectra. A rigid, single sugar conformation would not account for the experimental cross-peak patterns of any nucleotide in the octamer. However, as shown in Figure 5, the experimental peaks can be matched with simulations using a mixture of two conformers, one from the S and one from the N region of the pseudorotation circle (Table II). The pseudorotation phase angles of the dominant S conformation P_S vary from 171° for T5 to 126° for U^F and C7, and the S populations range from 95% for T5 to 75% for C7. The sugar pucker amplitude ϕ_m^S is about 37°. For the minor conformer, the parameters are P_N = 9° and ϕ_m^N = 37°. The sugar conformation of C8 is determined only tentatively because of a coalescence of H2′ and H2″ resonances. The accuracy of the sugar pucker parameters can be estimated to ±5° for P_S and ±5% for conformer populations. The sugar pucker amplitude ϕ_m is within the range 35°–40°. The pseudorotation parameters P_N and ϕ_m^N of the minor conformer are only approximations.

Determination of Interproton Distances via Analysis of 2D NOE Data Using MARDIGRAS

The presence of two fluorine atoms in the octamer posed some difficulty in using CORMA and MARDIGRAS because of dipolar interactions between the fluorines and the proton system. The CORMA (complete relaxation matrix analysis) program calculates theoretical 2D NOE cross-peak intensities and compares them with experimental values (Borgias & James, 1988, 1989; Keepers & James, 1984). MARDIGRAS uses the complete relaxation matrix to efficiently calculate interproton distances for the "real" structure by an iteration procedure which minimizes deviations between the calculated and experimental 2D NOE intensities (Borgias & James, 1989, 1990). We note that the version of the program used for the

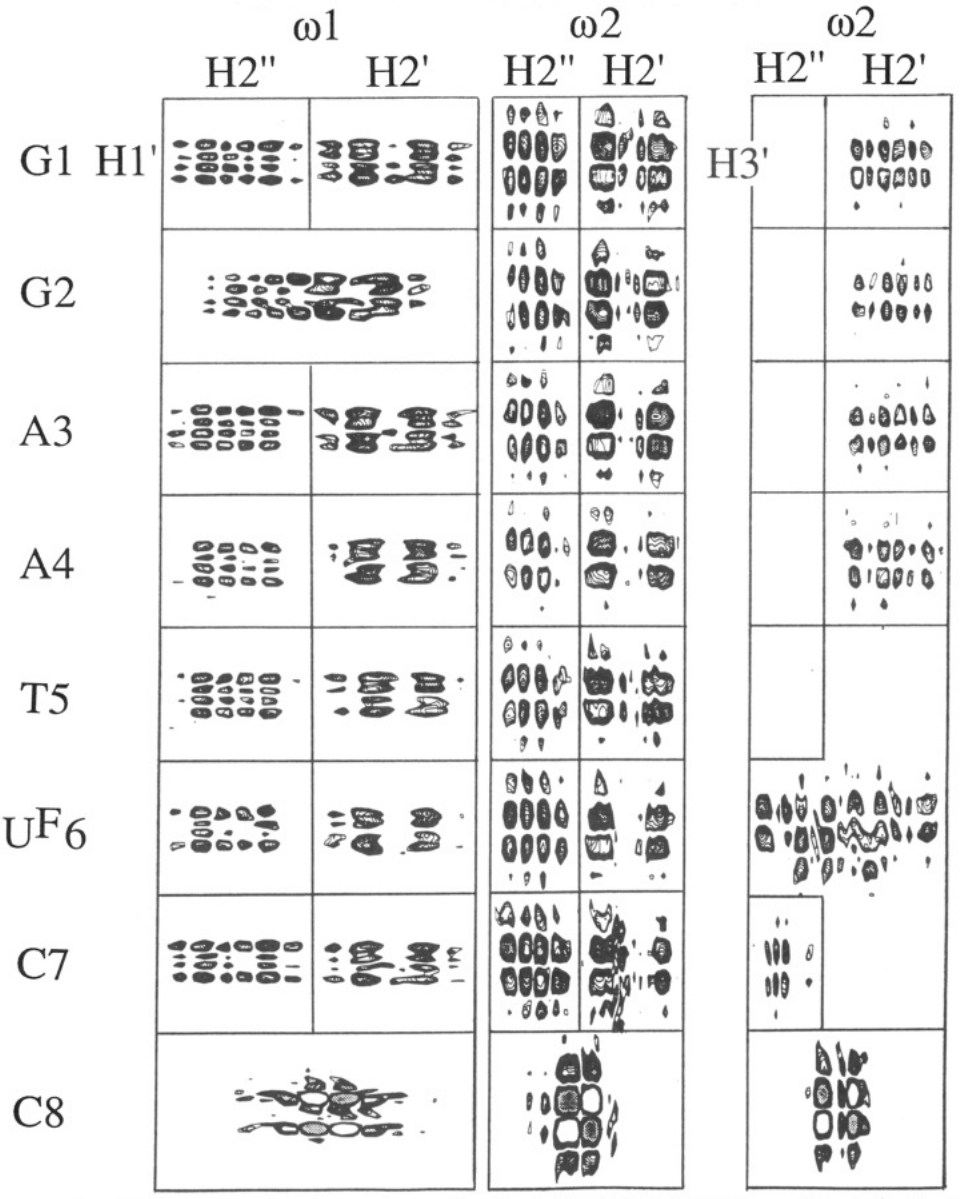


FIGURE 4: Experimental 2QF-COSY cross-peaks of successive nucleotides in [d(GGAATU^{F}CC)]$_2$ used for determination of sugar pucker: H1′-ω_2/H2″-ω_1, H1′-ω_2/H2′-ω_1, H1′-ω_1/H2″-ω_2, H1′-ω_1/H2′-ω_2, H3′-ω_1/H2″-ω_2, and H3′-ω_1/H2′-ω_2. Negative peaks are shaded.

studies reported also enables us to calculate distances involving rapidly rotating methyl groups with much greater accuracy than was previously possible (Liu et al., 1992). In oligodeoxynucleotides without modifications, dipolar interactions entailing heteroatoms, namely, ^{31}P of the phosphoryl groups, with protons are negligible (Broido et al., 1985). But ^{19}F can interact strongly with the proton system, changing intensities of 2D NOE cross-peaks for those protons which are spatially close to the ^{19}F, e.g., H6 of U^F and H1′, H2′, H2″ of the same and the 5′-neighboring residue. The conformation in this region is especially important because of possible changes introduced by the modified nucleotide. This prompted us to modify both programs so that corrections for the dipolar proton–fluorine interactions were included.

The quality of interproton distances in the octamer, derived from MARDIGRAS, is made evident by running the program twice for the same intensity data set but starting from different model structures. A comparison of interproton distances between the two starting models, A- and B-DNA, and the distances after two MARDIGRAS runs with A- and B-DNA as starting structures is presented in Figure 6.

Determination of [d(GGAATU^{F}CC)]$_2$ Solution Structure

Generation of Constraints from NMR Data. The construction of constraints for rMD and rMIN studies was based on both 2D NOE and 2QF-COSY data. The three categories of constraints are as follows: (a) proton–proton distance constraints derived from MARDIGRAS analysis of 2D NOE intensities; (b) sugar torsion angle constraints from vicinal proton–proton coupling constants determined by SPHINX and LINSHA simulations of 2QF-COSY cross-peaks; and (c) hydrogen bond distance and flat angle constraints, the existence of which can be inferred from the imino proton signals but with values taken from crystallographic data. All constraints had the form of a flat-well energy potential with the flat part, parabolic part, and force constants related to the accuracy of fitting the experimental data.

The three 2D NOE intensity sets analyzed with MARDIGRAS, employing the two starting structures, energy-minimized A- and B-DNA, yielded 204 distance constraints with associated upper and lower bounds. Of the 204 constraints, only about one-fourth are interresidue. Overlap of 2D NOE cross-peaks made it impossible to determine some interstrand

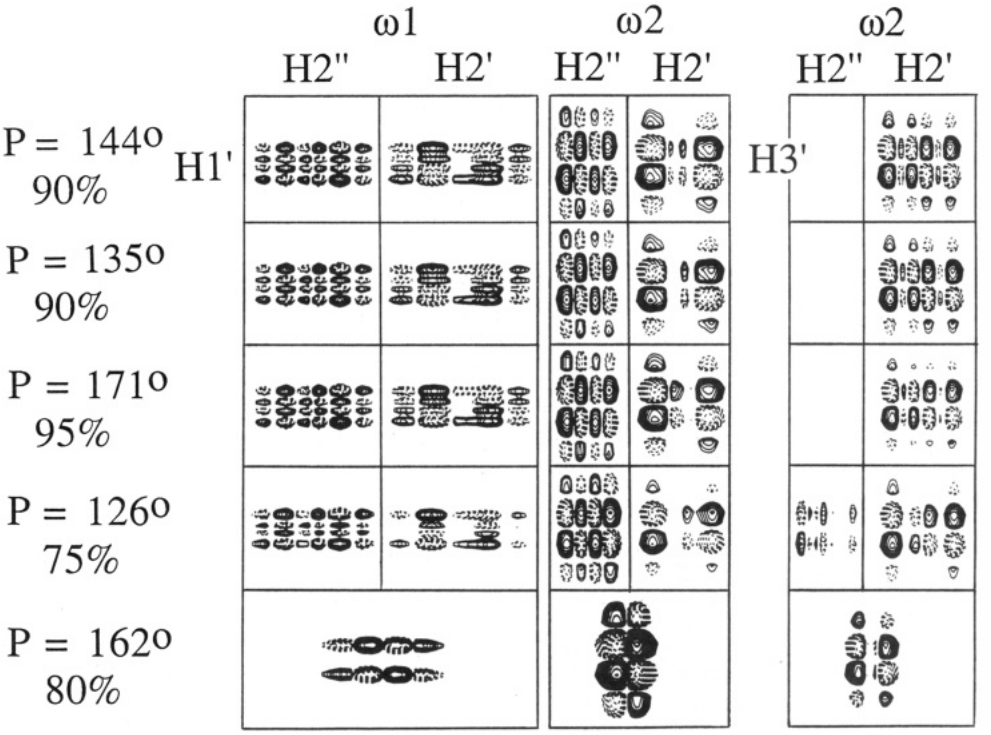


FIGURE 5: Best fit simulations of the 2QF-COSY cross-peaks. Negative peaks are indicated by broken lines. For values of coupling constants and pseudorotational parameters of the individual nucleotides, see Table II.

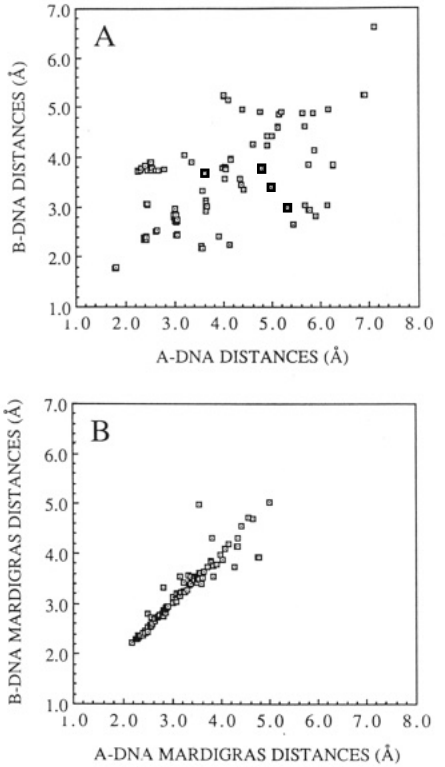


FIGURE 6: Comparison of interproton distances for [d(GGAATU^F-CC)]$_2$ in A-DNA and B-DNA conformations (A) and for corresponding MARDIGRAS-derived distances using A- and B-DNA starting models (B). The 200-ms 2D NOE data set was used.

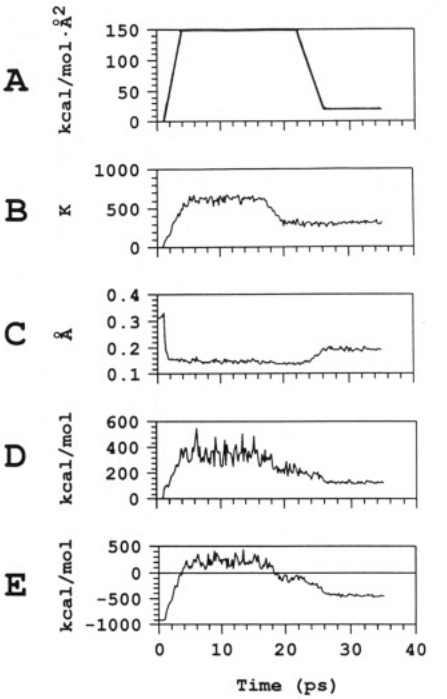


FIGURE 7: Parameter evolution during a rMD run, starting from B-DNA: (A) maximal (100%) constraints' force constants k_c; (B) temperature; (C) distance average deviation from the median of the constraint's flat well, avd; (D) constraint violation energy E_{con}; and (E) potential energy E_p.

constraints which we have previously found to be quite valuable (Schmitz et al., 1991).

Sugar ring torsion angles for the dominant S conformer were utilized (Table II). Widths of the flat well varied from 9° to 14°, and the parabolic part of the potential extended 5° on each side. Force constants were set to 100% of maximum for the entire rMD run to ensure constraint forces similar to those of distances determined with the greatest accuracy.

Molecular Dynamics Simulations. Although there have been many high-resolution structures (predominantly of proteins and peptides) determined by molecular dynamics simulations [for review see Karplus and Petsko (1990) and van Gunsteren and Berendsen (1990)], there is no consensus on a standard procedure as yet. Restrictions are imposed by the compute-intensive nature of the calculations and the interplay of kinetic energy at a given temperature with both conventional potential energy terms and applied experimental

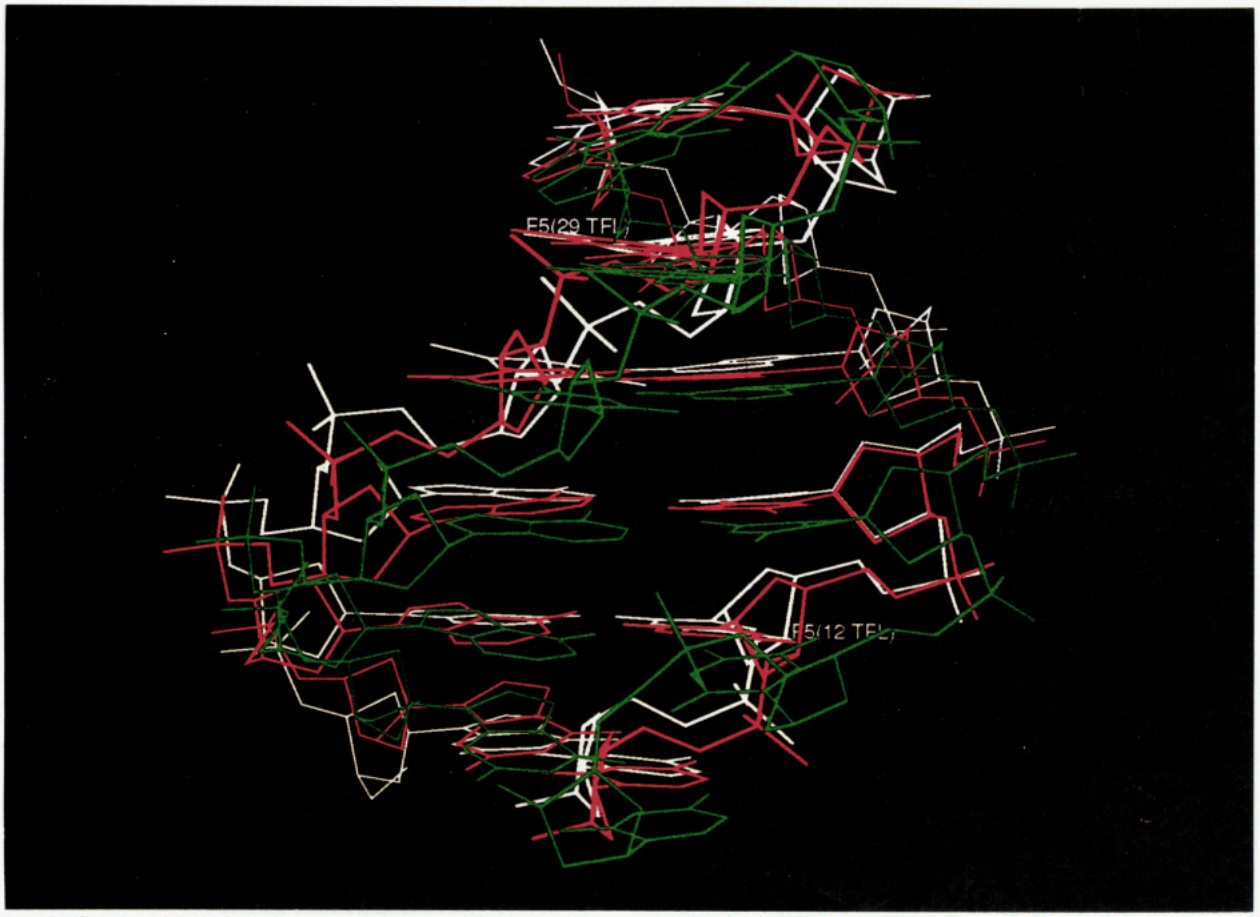


FIGURE 8: Superposition of the final structure MD_{AV} (white) and two structures after the rMD simulations, MD-A (green) and MD-B (red).

Table II: Sugar Ring Coupling Constants (in Hz) and Sugar Pucker Parameters (in deg) for [d(GGAATU^{F}CC)]$_2$

	G1	G2	A3	A4	T5	U^{F}6	C7	C8
J(1′,2′)	9.5	9.5	9.4	9.4	9.6	8.5	8.1	8.1
J(1′,2″)	5.6	5.6	5.7	5.7	5.6	6.1	6.2	6.0
J(2′,2″)	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
J(2′,3′)	6.5	6.5	7.0	7.0	5.4	7.5	7.4	5.9
J(2″,3′)	2.1	2.1	2.1	2.1	1.8	3.0	3.5	3.5
J(3′,4′)	2.8	2.8	3.4	3.4	1.5	4.6	4.8	3.0
J(3′,P)	5.8	5.8	5.8	5.8	5.8	5.8	5.8	5.8
P_S	144	144	135	135	171	126	126	162
ϕ_m^S	37	37	37	37	37	37	37	37
% S	90	90	90	90	95	80	85	80
P_N	9	9	9	9	9	9	9	9
ϕ_m^N	37	37	37	37	37	37	37	37
% N	10	10	10	10	5	20	25	20

constraints. Therefore, several protocols were evaluated with different weighting of distance and torsion angle restraints and with different simulation temperatures. Various starting models should ultimately converge to the same final structure if the constraint set characterizes the conformational properties unambiguously. Monitoring potential energy (E_p), constraint violation energy (E_{con}), and the average deviation from the median of the constraint's flat well (avd) enables evaluation of the performance of rMD procedures.

Protocols, based on previous rMD simulations of DNA octamers (Kerwood et al., 1991; Schmitz et al., 1991), were modified until successful convergence for two different starting structures, A- and B-DNA, was achieved, together with reasonably low values of E_p, E_{con}, and a distance avd less than 0.23 Å, the average width of a flat well. Typical evolution of E_p, E_{con}, and distance avd during a successful rMD run starting from B-DNA, together with the course of restraint force constants and temperature, is shown in Figure 7. For the final 8 ps, no drifting in total energy or average constraints' deviation was observed, indicating that the conformation is fluctuating in an energy minimum well. Slight modifications of that protocol in conjunction with various starting structures (energy-minimized in different ways) and different sets of initial velocities still yielded a set of convergent structures.

Assuming that structural excursions during the last part of a run fluctuate around the energy minimum, coordinate sets covering the last 4 ps of a run were averaged and subjected to a final restrained energy minimization, which removed averaging artifacts. Force fields used during the minimization were the same as those at the end of the rMD run, i.e., 20 kcal/mol·Å^2 and 20 kcal/mol·rad^2. The averaged final structures from different rMD runs were compared by calculating their rms deviations and visually inspecting via computer graphics.

Modifications of the above rMD protocol did not influence the structures (rms deviation less than 1 Å) when starting either from A-DNA or from B-DNA. On the other hand, rms deviations for the final structures derived from rMD runs with A- or B-DNA as the starting model, MD-A and MD-B, respectively, exceeded 2.3 Å. This value, although much less than the rmsd value of 4.5 Å between the starting models themselves, is far from satisfactory. A detailed inspection of rms deviations per residue (see supplementary material) reveals a very good structural consistency for the inner six base pairs. The rms deviation for this part of the molecule is 1.2 Å. Large structural discrepancies, however, are observed for the two terminal base pairs. In MD-B, one of the terminal base pairs is actually destacked and oriented almost perpendicular to the plane of the next pair. In MD-A both terminal base pairs

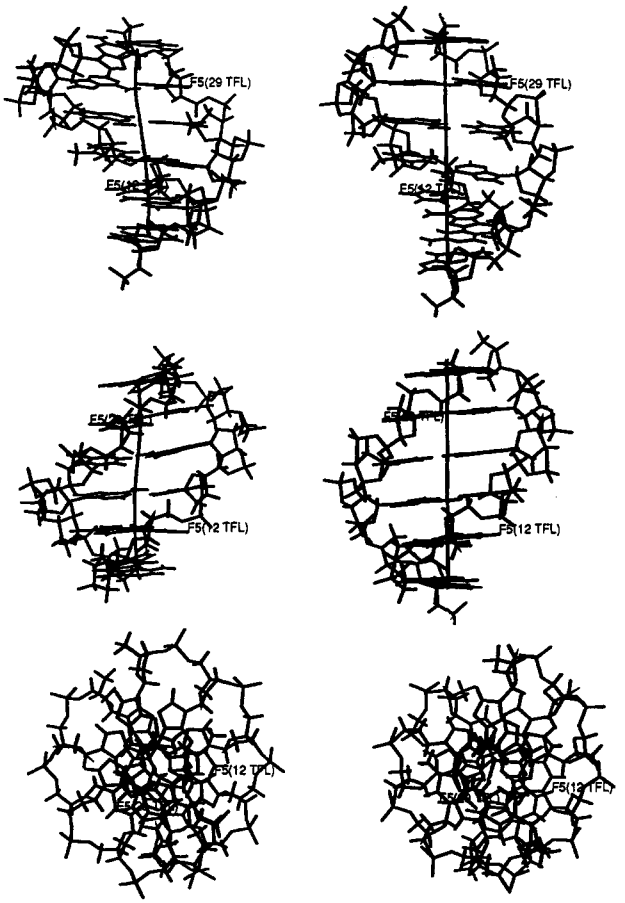


FIGURE 9: Comparison of MD_{AV} and classical B-DNA hexamers (inner six base pairs of $[d(GGAATU^{F}CC)]_2$, including their helix axes). Views are into the major groove (top), minor groove (middle), and down the helix axis (bottom). Fluorines are marked with labels.

have irregular structures, far from B-DNA. It is obvious that constraints at the ends of the octamer are not sufficient to enable A- and B-DNA starting models to converge at the helix termini. One reason is that distance constraints at the ends of the octamer are inaccurate due to exchange phenomena (vide supra), which influence the observed 2D NOE intensities. Additionally, there are fewer interresidue distance constraints between the terminal and penultimate nucleotides than in the rest of the octamer.

Poor characterization of terminal base pairs but good convergence found for the inner six base pairs forced us to examine only the inner six base pairs (further referred to as the hexamer) when global helical parameters were being calculated and to consider as meaningful only conformational parameters of the hexamer. The final structure from molecular dynamics simulations, MD_{AV}, was derived by averaging the coordinations of six different MD-A and MD-B structures, with mutual rms deviations less than 1.4 Å, and running a restrained energy minimization. The restrained force field was the same as that used during the last 4 ps of the rMD runs. Superposition of the final structure, MD_{AV}, and two typical structures, MA-A and MD-B, for the inner six base pairs of the octamer is shown in Figure 8.

It is pertinent to consider some methodological problems related to constraints. They reflect a time-averaged structure of a molecule and are used in molecular dynamics simulations which give a single final structure. However, it is known from experimental data, e.g., sugar pucker analysis, that the structure is a mixture of conformations in dynamic equilibrium. So the final structure, although close to the dominant conformer, is not identical with it. Structural parameters describing the refined, time-averaged structure may differ from those reflected in experimental data, e.g., sugar pucker phase P and amplitude ϕ_m. Additionally, an averaged structure arising from averaging distances is not necessarily identical with that arising from averaging torsion angles; i.e., constraints of one type may contradict constraints of another type. To check this problem, a restrained energy minimization of B-DNA using only torsion angle constraints, with $k_c = 150$ kcal/mol·rad^2, was performed. The resulting values of pseudorotational phase P were close to those of MD_{AV} within 20°, except for the terminal C8 residues. The consistency of different types of constraints in this particular study is further stressed by the fact that it was possible to achieve a convergence to similar structures, starting from A-DNA and B-DNA models, only if both distance and torsion angle constraints were applied.

Features of the Final Structure

Energy Considerations. The total energy of the final structure MD_{AV} is −766.2 kcal/mol, 120–140 kcal/mol higher than the energies of starting energy-minimized A- and B-DNA models. The constraint energy of MD_{AV}, 78.8 kcal/mol, which represents violations of the NMR data, is considerably lower than for the B-DNA starting model, 577.0 kcal/mol, or for the A-DNA starting model, 4396.3 kcal/mol. Distance constraints seem to be consistent with torsion angle constraints, except for the terminal cytosine residues (see supplementary material). This confirms previous conclusions derived from analysis of the rMD runs. Consistency of MD_{AV} with experimental results is also reflected in small average deviations from the median of the constraint's flat well, 0.175 Å and 9.7° for distances and torsion angles, respectively. The lowest possible values, corresponding to complete consistency of a structure with experimental data, within experimental error, are 0.115 Å for distance avd and 6° for torsion angle avd (each value is half of the average width of the flat well).

Residual Index. The complete relaxation matrix program CORMA enables comparison of experimental 2D NOE cross-peak intensities with values calculated by assuming any particular structure. Experience has shown us that 2D NOE peak intensities are quite sensitive to interproton distances; i.e., it is easy to generate plausible model structures which yield theoretical spectra at strong variance with experimental spectra. Various numerical indices could be used as a figure of merit to evaluate the overall fit of experimental and calculated intensities (Gochin & James, 1990; Suzuki et al., 1986; Thomas et al., 1991) and, hence, the consistency of that structure with the experimental data. Most commonly, a residual index R, analogous to a crystallographic R factor, has been used. However, this R factor is dominated by evaluation of the fit of cross-peaks corresponding to very short (<2.5 Å) distances. As the 2D NOE cross-peak intensities depend on distances essentially with a sixth-root dependence, we have found that a more sensitive monitor of fitting is the sixth-root residual index (James, 1991; Kerwood et al., 1991; Thomas et al., 1991):

$$R_1^x = \frac{\sum_i |I_o^{1/6}(i) - I_c^{1/6}(i)|}{\sum_i I_o^{1/6}(i)}$$

where $I_o(i)$ is the experimental intensity of cross-peak i and

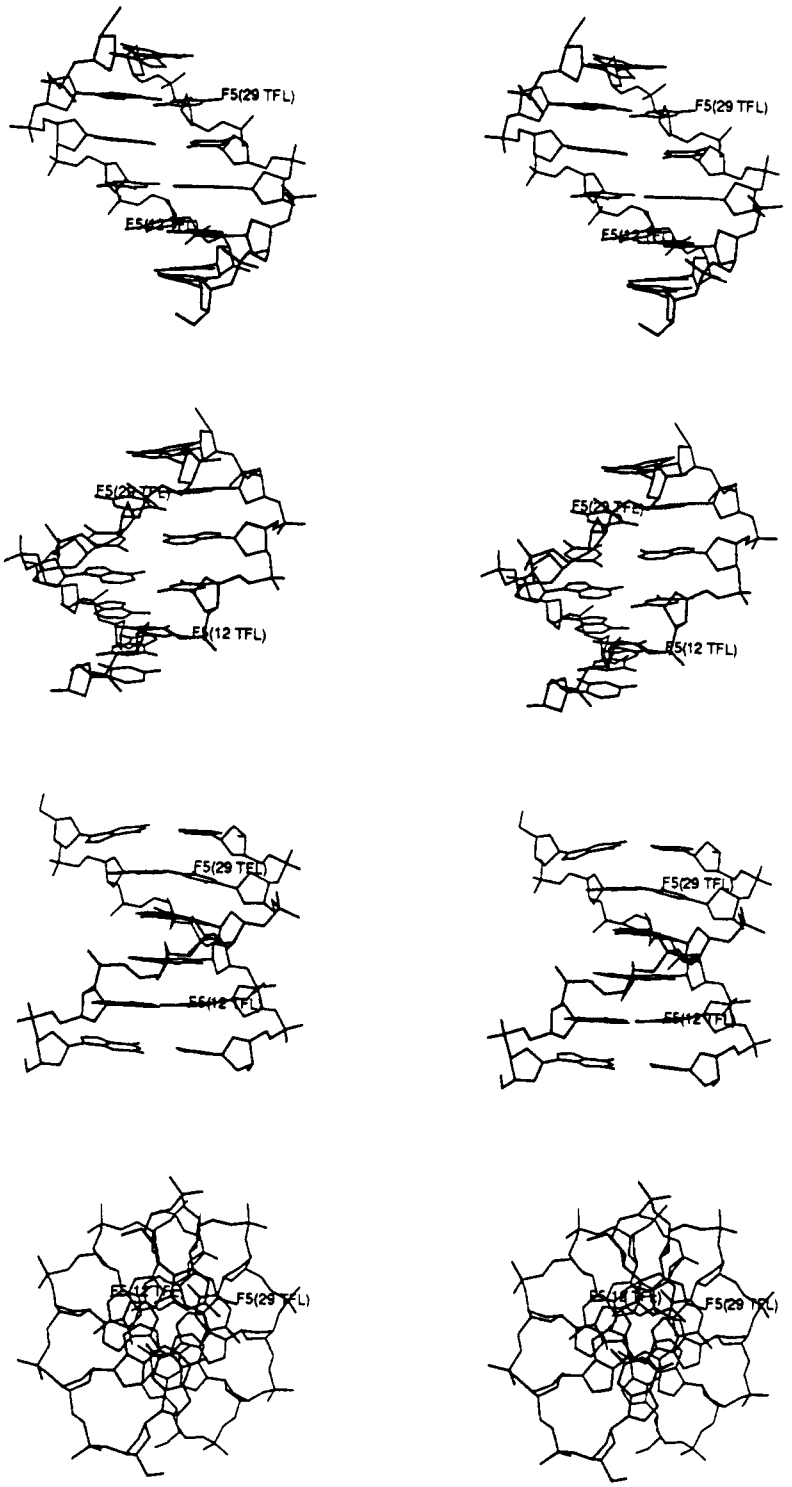
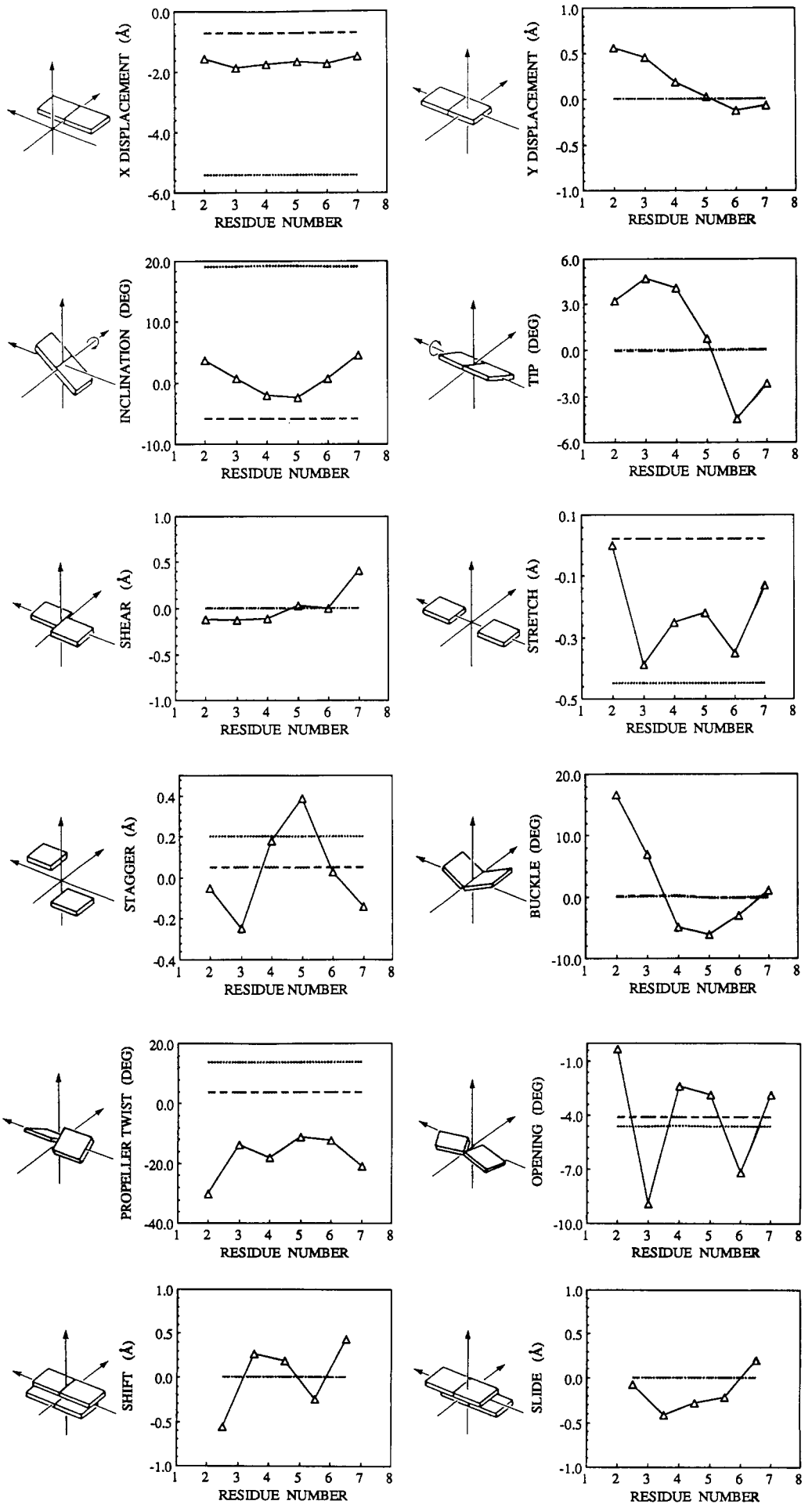


FIGURE 10: Stereoviews of the final structure MD_{AV} showing the inner six base pairs of [d(GGAATU^{F}CC)]$_2$ from different perspectives. Fluorines are marked with labels.

$I_c(i)$ is the corresponding cross-peak intensity calculated for a particular structure. The summation is generally carried out for all observed cross-peaks. But indices, giving discrepancies between experimental and calculated intensities for smaller parts of the molecule or to highlight certain features of the molecule, can also be defined. In an intraresidue index, both protons of the ith pair must belong to the same residue (nucleotide) to be included in the summation. And in an interresidue index, only those intensities are summed up which arise from proton pairs belonging to different residues. A comparison of R_1^x values for different structures is shown in Table III. Energy-minimized B-DNA fits the experimental data better than classical B-DNA and much better than A-DNA. But R_1^x values are substantially reduced for the refined structure after refinement using restrained molecular dynamics (MD_{AV} structure). We note that our CORMA and MARDIGRAS programs calculate the standard R factor as well as R_1^x; the standard R factor improvement with refinement mirrored that of R_1^x.

Conformational Details. Comparison of the final structure MD_{AV} with the classical B-DNA form for the inner six base pairs of the octamer is presented in Figure 9. Stereoviews of the hexamer are shown in Figure 10. It should be realized that the experimental NMR parameters, being short-range in nature, largely establish local structure; consequently, many helical parameters are only secondarily established to be


X DISPLACEMENT (Å)
0.0
-2.0
-4.0
-6.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
Y DISPLACEMENT (Å)
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
INCLINATION (DEG)
20.0
10.0
0.0
-10.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
TIP (DEG)
6.0
3.0
0.0
-3.0
-6.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
SHEAR (Å)
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
STRETCH (Å)
0.1
-0.1
-0.3
-0.5
1 2 3 4 5 6 7 8
RESIDUE NUMBER
STAGGER (Å)
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7 8
RESIDUE NUMBER
BUCKLE (DEG)
20.0
10.0
0.0
-10.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
PROPELLER TWIST (DEG)
20.0
0.0
-20.0
-40.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
OPENING (DEG)
-1.0
-4.0
-7.0
-10.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
SHIFT (Å)
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER
SLIDE (Å)
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8
RESIDUE NUMBER

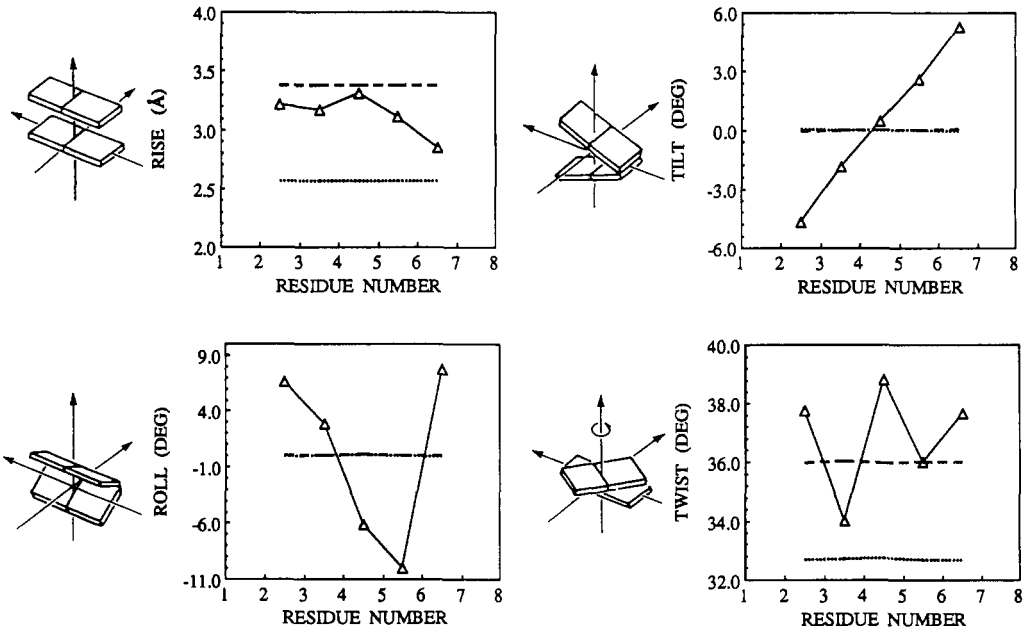


FIGURE 11: Helix parameters for MD$_{AV}$ (△), classical A-DNA (···), and classical B-DNA (- - -).

consistent with the local structure and with the AMBER force field. So these helical parameters will be subject to any bias or inaccuracies in that force field.

The resulting structure, MD$_{AV}$, is slightly shorter than the classical B-DNA hexamer and exhibits some irregularities in the helix axis. Although the axis is fairly smooth as a whole, with global curvature of 5.8°, more pronounced bends are observed near U^F·A base pairs. The width of the minor groove in MD$_{AV}$, measured by the phosphorus–phosphorus distance across the groove (minus the combined radii of the phosphorus atoms, 5.8 Å), varies from residue to residue but is about 1.5 Å smaller than that of standard B-DNA, which is 6 Å. Since the octamer has a 2-fold axis of symmetry, it might be expected that the structure resulting from rMD analysis would reflect this symmetry. To a certain extent, this symmetry is manifest (see Figure 12), but since the duplex is not completely constrained by NMR-determined distances and angles, there is also some variability in the conformational parameters between the two symmetric halves of the duplex, as each half has been able to follow its own rMD trajectory.

Three types of helix parameters, axis–base pair, intrabase pair, and interbase pair, according to the definitions of the EMBO workshop (Diekman, 1989), for MD$_{AV}$, classical A-DNA, and B-DNA are displayed in Figure 11. The global helix axis of the MD$_{AV}$ hexamer has been visualized in Figure 9 using MACB, the graphics module of CURVES, so we have refrained from representing the individual axis junction parameters. Among the axis–base pair parameters, MD$_{AV}$ reveals larger *X* displacements than in B-DNA and small, but nonzero *Y* displacements. Values for inclination in MD$_{AV}$ are smaller than in classical B form and tip angles nonzero. Both U^F·A base pairs have the smallest inclination and the greatest tip angle among all six base pairs of MD$_{AV}$. Shear, stretch, and stagger are minimal, <0.4 Å, and base pair opening is more pronounced only for the two U^F·A base pairs. Propeller twist and buckle are larger than in classical B-DNA, resulting in deviations from planarity of complementary base pairs. For the interbase pair parameters, regular B-DNA and A-DNA structures exhibit values of zero for slide, shift, tilt, and roll. In MD$_{AV}$, slide and shift are up to 0.6 Å and tilt and roll up to 10°, resulting in bends in the helix near U^F·A base pairs. The average twist is close to that of B-DNA, but the average rise is smaller by 0.25 Å, thus making the MD$_{AV}$ hexamer shorter than classical B-DNA.

Values of torsion angles α through ζ and χ (Figure 2) for MD$_{AV}$ and A- and B-DNA are presented in Figure 12. Most torsion angle values for different residues of MD$_{AV}$ differ by less than ca. 20°, except for χ and ζ which are more scattered, up to 40°. The ranges of torsion angles values are shifted in comparison with values typical for classical B-DNA, especially for β, χ, and δ. The angle δ reflects the deoxyribose ring conformation and is related to the pseudorotational phase angle *P*. Since rMD searches for a single conformation, the *P* values in Figure 12 are not identical to those derived from analysis of 2QF-COSY spectra (vide supra). Nevertheless, they reflect the major conformer of the equilibrium. The values of χ and α are between those characteristic of classical B-DNA and A-DNA. Values of ζ are close to that of classical B-DNA and values of ϵ to that of classical A-DNA. The values of γ are scattered in the +synclinal area around 60° and those of β in the antiperiplanar area around 175°, outside the ranges of classical A- and B-DNA. On average, values of torsion angles in MD$_{AV}$ are much closer to those in B-DNA after energy minimization even without constraints than to those in classical B-DNA. Nevertheless, individual values of torsion angles may differ by 20° when MD$_{AV}$ is compared with energy-minimized B-DNA.

The results of conformational analysis suggest a dependence of DNA structure on sequence. Replacement of T by U^F seems to introduce some local conformational changes, e.g., local bending of the molecule. The lack of a precise conformational analysis for the parent octamer makes it impossible to describe the changes explicitly (vide infra).

Comparison with Similar DNA Sequences. The parent octamer, i.e., without the replacement of T by U^F, was investigated previously by means of 2D NOE measurements and CORMA calculations (Broido et al., 1985). However, without any structural refinement, it was only possible to perform a qualitative analysis of the structure.

Several papers, both theoretical and experimental, have been published on the dodecamer [d(CGCGAATTCGCG)]$_2$, which contains the same *Eco*RI endonuclease recognition sequence GAATTC as our octamer. The structure of the dodecamer was analyzed in the crystal (Dickerson & Drew,

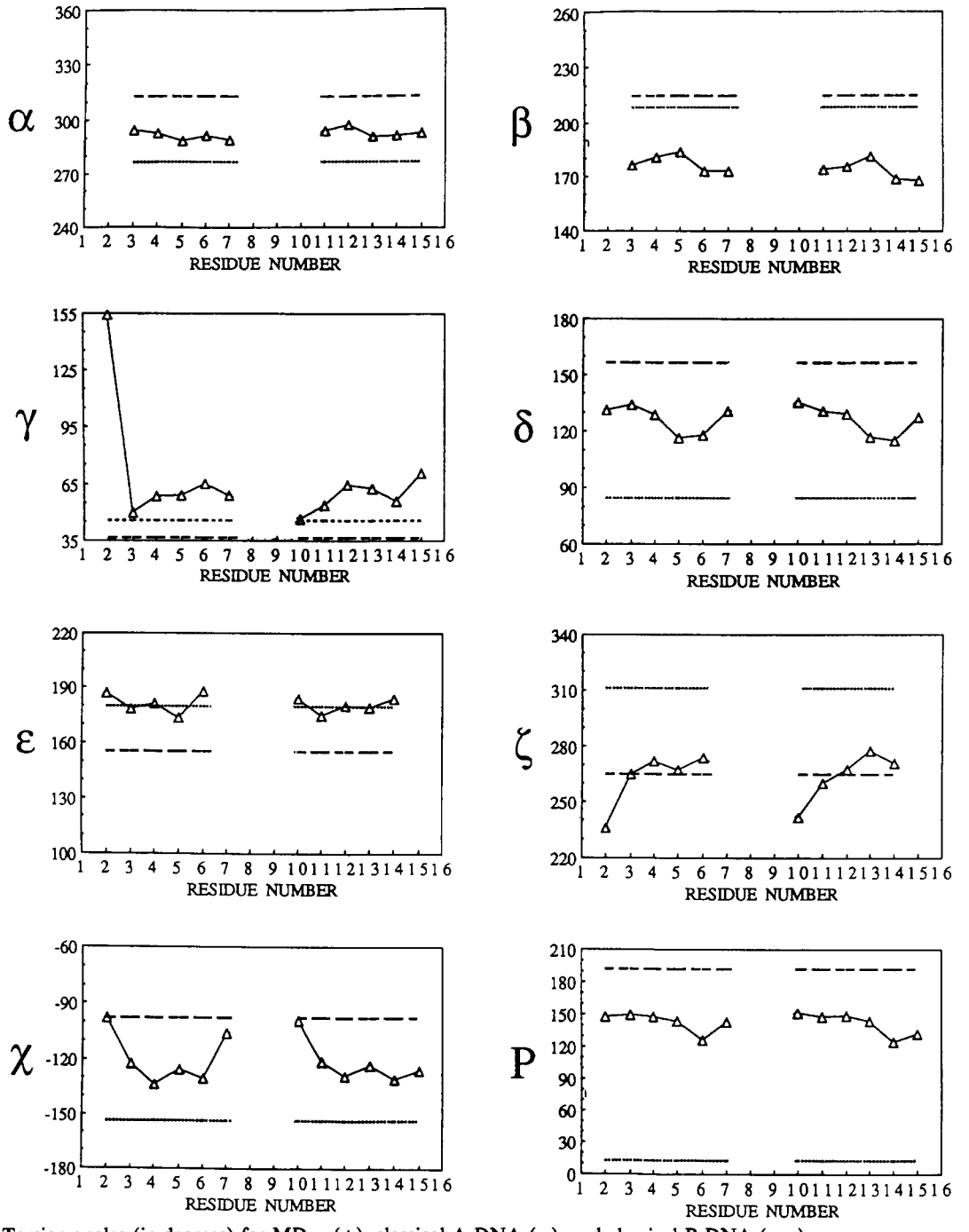


FIGURE 12: Torsion angles (in degrees) for MD_{AV} (Δ), classical A-DNA (···), and classical B-DNA (- - -).

Table III: Comparison of Sixth-Root Residual Indices R_1^x (Mixing Time 200 ms) for the Initial and Resulting Structures

	energy-minimized A	classical B	energy-minimized B	MD_{AV}
intraresidue	0.160	0.070	0.061	0.038
interresidue	0.325	0.144	0.087	0.056
total R_1^x	0.195	0.086	0.066	0.042

1981), in solution by means of 2D NOE and distance geometry (Nerdal et al., 1989), and by unrestrained molecular dynamics simulations in water (Swaminathan et al., 1991). The crystal structure of $[d(CGTGAATTCACG)]_2$ has also been determined (Larsen et al., 1991; Narayana et al., 1991). The structure of $[d(TCGCGAATTCGCG)]_2$ in a complex with *Eco*RI endonuclease was investigated in the crystalline state (Frederick et al., 1984). All studies reveal structural irregularities of the B-DNA double helix, with some structural variations reported between the different studies. Characteristic kinks at the A-T, C4-G5 and C10-G11 base steps are observed in the oligomer–*Eco*RI endonuclease complex, resulting in a 12° bend of the helix and enlarging of the grooves. The existence of the A-T kink was further supported by 2D NOE and distance geometry in solution (Nerdal et al., 1989) but was observed neither in the crystalline dodecamers (Dickerson & Drew, 1981; Larsen et al., 1991; Narayana et al., 1991) nor in the solution structure suggested by molecular dynamics simulations (Swaminathan et al., 1991). On the other hand, the molecular dynamics solutions did not confirm a narrowing of the minor groove in the AT region as observed in the crystalline state and in the present NMR/rMD study. Comparing the common hexameric moiety, the global helical features of MD_{AV} resemble those of the crystalline dodecamers.

CONCLUSIONS

The present study demonstrates that incorporating MARDIGRAS-derived distances, calculated from 2D NOE cross-peak intensities, and SPHINX/LINSHA-derived torsion angles, from 2QF-COSY cross-peaks, in rMD calculations is an effective method to determine the high-resolution structure of DNA fragments in solution. The structure determined for [d(GGAATU^{F}CC)]$_2$ has been influenced both by the experimental NMR data and by the empirical force field used in the rMD calculations. It is most satisfying that convergence was achieved when two quite different DNA models, A-DNA and B-DNA, were used as starting structures for the rMD runs. Such convergence seems to be the most convincing argument for the validity of the resulting structure. This was further supported by the total energy, the constraints' violation energy, and the fit with experimental data (R_1^x) that the inner six base pairs of [d(GGAATU^{F}CC)]$_2$ exhibited in the final structure resulting from averaging the resultant coordinates from six rMD runs, followed by restrained energy minimization.

One of the biggest methodological limitations seems to be imposed by the paucity of NOE data. A reliable structure was found only for the inner six base pairs of the octamer. This probably results from the limited number of 2D NOE intensities involving the terminal base pairs and by exchange processes (from fraying) which make even those few distances inaccurate. MARDIGRAS calculations, yielding a consistent set of accurate interproton distances based on a complete relaxation matrix analysis, seem to be crucial for determination of the subtleties of sequence-dependent structure. It must also be noted that convergence to similar structures from different starting models with rMD was achieved with torsion angle constraints in addition to distance constraints. This demonstrates the value of using as many consistent constraints as possible, especially interresidue (and, whenever possible, interstrand) distance constraints.

Replacement of thymine by 5-fluorouracil seems to introduce some local, conformational changes in the DNA helix. Some minor changes were also found in the structure of an alternating d(AT) decamer with one of the thymines replaced by uracil (Kerwood et al., 1991).

ACKNOWLEDGMENT

We gratefully acknowledge the use of the University of California, San Francisco, Computer Graphics Laboratory (supported by NIH Grant RR 01081). We are indebted to Dr. D. Ferguson for preparing the AMBER database for U^F, to Dr. U. Schmitz for help with the molecular dynamics simulations, and to P. Orlewski for assistance in modifying CORMA and MARDIGRAS for the effects of ^{19}F. We also gratefully acknowledge the use of the Cray-YMP supercomputer which was supported by a grant from the Pittsburgh Supercomputing Center through the NIH Division of Research Resources Cooperative Agreement U41RR04154 and a grant from the National Science Foundation Cooperative Agreement ASC-8500650.

SUPPLEMENTARY MATERIAL AVAILABLE

Three figures showing exchange cross-peaks between the imino and water protons, rms deviations of atomic coordinates for each residue of the structures, and inter- and intraresidue constraint violation energies of MD_{AV} (3 pages). Ordering information is given on any current masthead page.

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.
Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry 29*, 4828–4839.
Boelens, R., Scheek, R. M., Dijkstra, K., & Kaptein, R. (1985) *J. Magn. Reson. 62*, 378–386.
Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173*, 299–311.
Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290–308.
Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493–512.
Borgias, B. A., & James, T. L. (1989) in *Methods in Enzymology, Nuclear Magnetic Resonance, Part A: Spectral Techniques and Dynamics* (Oppenheimer, N. J., & James, T. L., Eds.) Vol. 176, pp 169–183, Academic Press, New York.
Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475–487.
Broido, M. S., Zon, G., & James, T. L. (1984) *Biochem. Biophys. Res. Commun. 119*, 663–670.
Broido, M. S., James, T. L., Zon, G., & Keepers, J. W. (1985) *Eur. J. Biochem. 150*, 117–128.
Burkert, U., & Allinger, N. L. (1982) *Molecular Mechanics*, American Chemical Society Monograph 177, American Chemical Society, Washington, DC.
Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., & Wüthrich, K. (1989) *Biochemistry 28*, 1462–1470.
Cheng, Y.-C., & Nakayama, K. (1983) *Mol. Pharmacol. 23*, 171–174.
Coll, M., Saal, D., Frederick, C. A., Rich, A., & Wang, A.-H.-J. (1989) *Nucleic Acids Res. 17*, 911–923.
Danenberg, P. (1977) *Biochim. Biophys. Acta 473*, 73–92.
de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Isr. J. Chem. 27*, 181–188.
Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.
Diekman, F. (1989) *EMBO J. 8*, 1–4.
Dobson, C. M., Lian, L.-Y., Redfield, C., & Topping, K. D. (1986) *J. Magn. Reson. 69*, 201–209.
Egan, W. M., Schneerson, R., Werner, K. E., & Zon, G. (1982) *J. Am. Chem. Soc. 104*, 2898–2910.
Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983) *Biochemistry 22*, 5930–5942.
Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W., & Rosenberg, J. M. (1984) *Nature 209*, 327–331.
Gallo, K., Huang, C., Ferrin, T. E., & Langridge, R. (1989) *Molecular Interactive Display and Simulation (MIDASplus)*, University of California, San Francisco.
Gochin, M., & James, T. L. (1990) *Biochemistry 29*, 11172–11180.
Gochin, M., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 11161–11171.
Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1985) *J. Am. Chem. Soc. 107*, 6394–6396.
Habener, J. F., Vo, C. D., Le, D. B., Gryon, G. P., Ercolani, L., & Wang, A.-H.-J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1735–1739.
Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.
Harris, D. R., & Macintyre, W. M. (1964) *Biophys. J. 4*, 203–225.
Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.
Hore, P. J. (1989) in *Methods in Enzymology, Nuclear Magnetic Resonance, Part A: Spectral Techniques and Dynamics* (Oppenheimer, N. J., & James, T. L., Eds.) Vol. 176, pp 64–77, Academic Press, New York.

Ingraham, H. A., Tseng, B. Y., & Goulian, M. (1980) *Cancer Res. 40*, 998–1001.
James, T. L. (1991) *Curr. Opin. Struct. Biol. 1*, 1042–1053.
Karplus, M., & Petsko, G. A. (1990) *Nature 347*, 631–639.
Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.
Kerwood, D. J., Zon, G., & James, T. L. (1991) *Eur. J. Biochem. 197*, 583–595.
Kessler, H., Müller, A., & Oschkinat, H. (1985) *Magn. Reson. Chem. 23*, 844–852.
Kochoyan, M., Leroy, J. L., & Gueron, M. (1987) *J. Mol. Biol. 196*, 599–609.
Kremer, A. B., Mikita, T., & Beardsley, G. P. (1987) *Biochemistry 26*, 391–397.
Lamerichs, R. M. J. N., Boelens, R., van der Marel, G., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Ruterjans, H. (1989) *Biochemistry 28*, 2985–2991.
Larsen, T. A., Kopka, M. L., & Dickerson, R. E. (1991) *Biochemistry 30*, 4443–4449.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63–91.
Lavery, R., & Sklenar, H. (1990) *CURVES 3.0. Helical Analysis of Irregular Nucleic Acids*, Laboratory for Theoretical Biology, CNRS, Paris.
Liu, H., Thomas, P. D., & James, T. L. (1992) *J. Magn. Reson. 98*, 163–175.
Lönn, U., & Lönn, A. (1984) *Cancer Res. 44*, 3414–3418.
Madrid, M., Llinas, E., & Llinas, M. (1991) *J. Magn. Reson. 93*, 329–346.
Major, P. P., Egan, E., Herrick, D., & Kufe, D. W. (1982) *Cancer Res. 42*, 3005–3009.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Narayana, N., Ginell, S. L., Russu, I. M., & Berman, H. M. (1991) *Biochemistry 30*, 4449–4455.
Nerdal, W., Hare, D. R., & Reid, B. (1989) *Biochemistry 28*, 10008–10021.
Nilsson, L., Clore, G. M., Gronenborn, A., Brunger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455–475.
Pearlman, D. A., Case, D. A., Caldwell, J., Seibel, G. L., Singh, U. C., Weiner, P. K., & Kollman, P. A. (1991) *AMBER 4.0 (UCSF)*, University of California, San Francisco.
Pinedo, H. M., & Peters, G. F. (1988) *J. Clin. Oncol. 6*, 1653–1664.
Post, C. B., Meadows, R. P., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc. 112*, 6796–6803.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–649.
Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–341.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York.
Schmitz, U., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 2357–2368.
Schmitz, U., Pearlman, D. A., & James, T. L. (1991) *J. Mol. Biol. 221*, 271–292.
Schuetz, J. D., & Diasio, R. B. (1985) *Biochem. Biophys. Res. Commun. 133*, 361–367.
Singh, U. C., & Kollman, P. A. (1984) *J. Comput. Chem. 5*, 129–145.
Singh, U. C., Weiner, P. K., & Kollman, P. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 755–759.
Sowers, L. C., Eritja, R., Kaplan, B. E., Goodman, M. F., & Fazakerley, G. V. (1987) *J. Biol. Chem. 262*, 15436–15442.
Sowers, L. C., Eritja, R., Kaplan, B. E., Goodman, M. F., & Fazakerley, G. V. (1988) *J. Biol. Chem. 262*, 14794–14801.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Stolarski, R., Buck, F., Fera, B., & Ruterjans, H. (1987) *Eur. J. Biochem. 169*, 603–609.
Suzuki, E.-I., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–6865.
Swaminathan, S., Ravishanker, G., & Beveridge, D. L. (1991) *J. Am. Chem. Soc. 113*, 5027–5040.
Szer, W., & Shugar, D. (1963) *Acta Biochim. Pol. 10*, 219–231.
Teitelbaum, H., & Englander, S. W. (1975) *J. Mol. Biol. 92*, 55–78.
Thomas, P. D., Basus, V. J., & James, T. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1237–1241.
van Gunsteren, W. F., & Berendsen, H. J. C. (1990) *Angew. Chem., Int. Ed. Engl. 29*, 992–1023.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316–336.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zhou, N., James, T. L., & Shafer, R. H. (1989) *Biochemistry 28*, 5331–5339.