Structural Effects of the C^2-Methylhypoxanthine:Cytosine Base Pair in B-DNA: A Combined NMR and X-ray Diffraction Study of d(CGC[m^{2}I]AATTCGCG)†

Danzhou Yang,‡ Yi-gui Gao,‡ Howard Robinson,‡ Gijs A. van der Marel,§ Jacques H. van Boom,§ and Andrew H.-J. Wang*,‡

Division of Biophysics and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Gorlaeus Laboratories, Leiden State University, 2300RA Leiden, The Netherlands

Received March 2, 1993; Revised Manuscript Received May 6, 1993

ABSTRACT: C^2-Methylhypoxanthine (m^{2}I) is a synthetic analog of guanine with the N^2-amino group replaced by a methyl group. We have studied the structural consequence of the m^{2}I incorporation in DNA by a combination of X-ray crystallographic, NMR, and enzymatic analyses. The crystal structure of d(CGC-[m^{2}I]AATTCGCG) has been solved and refined to an R factor of 20.7% at 2.25-Å resolution. In the DNA duplex, the two independent m^{2}I:C base pairs maintain the Watson–Crick scheme. While the C^2-methyl group of m^{2}I is in van der Waals contact with the O^2 of the base-paired cytosine, it only causes the base pair to have slightly higher propeller twist and buckle angles. Its solution structure was analyzed by the NMR refinement procedure SPEDREF [Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533] using 2D nuclear Overhauser effect data. Two starting models, a relaxed fiber model and an X-ray model, were subjected to the NOE-constrained refinement using 1518 NOE cross-peak integrals to arrive at the final models with (NOE) R factors of 13.8% and 14.3%, respectively. The RMSD between the two refined models (all atoms included) is 1.23 Å, which presently seems to be near the limit of convergence of NOE-based refinement. The local structures of the two models are in better agreement as measured by the RMSD of the dinucleotide steps, falling in the range 0.54–0.98 Å. Both refined solution structures confirm that the m^{2}I dodecamer structure is of the B-DNA type with a narrow minor groove at the AT region, as observed in the crystal. However, significant differences exist between the crystal and solution structures in parameters such as pseudorotation angles, propeller twist angles, etc. The solution structure tends to have a more uniform backbone conformation, an observation consistent with that concluded from the laser Raman study of d(CGCAAATTTGCG) [Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., J. (1988) *Biochemistry 27*, 931–938]. Three related dodecamers, d(CGCGAATTCGCG), d(CGC[m^{2}I]AATTCGCG), and d(CGC[e^{6}G]AATTCGCG), were tested as substrates for the restriction endonuclease *Eco*RI. The m^{2}I dodecamer was active, but the e^{6}G dodecamer was not. Our results illustrate the complementarity in terms of the structural information provided by the two methods, X-ray diffraction and NMR.

Recognition between protein and DNA is a critical issue in biology and is being studied intensively. Recent exciting developments include the structural analysis of several complexes of repressor proteins and their DNA operators (Steitz, 1990; Ptashne, 1992). Another example is the crystal structure determination of the restriction endonuclease *Eco*RI complexed to a DNA fragment containing the recognition sequence 5′-GAATTC[1] (Kim et al., 1990). The DNA tridecamer, d(TCGCGAATTCGCG), in this enzyme–DNA complex has contacts with the enzyme mainly in the major groove, with specific hydrogen-bonding interactions between the enzyme and the O^6/N^7 of guanine and the N^6/N^7 of adenine in the 5′-GAATTC sequence. There is no apparent specific interaction involving the N^2-amino group of guanine in the minor groove of the DNA duplex of the restriction sequence. Nevertheless, replacement of the deoxyguanosine in 5′-GAATTC with a deoxyinosine (I) changes the enzymatic activity significantly. For example, the k_{cat}/K_m value of the modified DNA octamer d(GIAATTCC) is significantly reduced (to 10%) from that of native d(GGAATTCC) (Brennan et al., 1986). Similarly, the k_{cat}/K_m value of d(CTIAATTCAG) is only 17% of that of d(CTGAATTCAG) (McLaughlin et al., 1987).

It would be of interest to discern the molecular basis of this altered substrate–enzymatic activity due to the absence of an N^2-amino group in guanine. One possibility is that the DNA molecule adopts a somewhat different conformation resulting from the G→I replacement. The structures of several deoxyinosine-containing oligonucleotides, including d(CG-CIAATTCGCG) in the B-form (Xuan & Weber, 1992), d(CGCICICG) in the Z-form (Kumar et al., 1992), and d(GGIGCTCC) in the A-form (Cruse et al., 1989), have been studied crystallographically, and they were in general fairly similar to their respective canonical crystal structures, suggesting that dI, incorporated in DNA, disturbs the structure only minimally.

But what is the reason for the diminished *Eco*RI activity in the dI-containing substrate? Subtle changes in the physical and chemical properties of an I:C base pair in solution (e.g., the stability) may be responsible. We attempt to address this question by further modifying the deoxyinosine residue to

† This work was supported by NIH Grants GM-41612 and CA-52506 (A.H.-J.W.).

* Author to whom correspondence should be addressed.

‡ University of Illinois at Urbana–Champaign.

§ Leiden State University.

[1] Abbreviations: A, T, G, and C, adenine, thymine, guanine, and cytosine or their respective nucleotides in DNA; m^{2}I, C^2-methyl-2′-deoxyinosine or C^2-methylhypoxanthine; e^{6}G, O^6-ethylguanine; m^{6}A, N^6-methyladenine; m^{2}I-DODE, d(CGC[m^{2}I]AATTCGCG); e^{6}G-DODE, d(CGC[e^{6}G]AATTCGCG); NMR, nuclear magnetic resonance; 2D NOESY, two-dimensional nuclear Overhauser effect spectroscopy; pr tw, propeller twist; RMSD, root-mean-square deviation.

0006-2960/93/0432-8672$04.00/0

C^2-methyldeoxyinosine (m^2I). The extra methyl group at the C^2 position of m^2I may interfere with base-pair formation with the opposing cytosine, due to the possible van der Waals clash with the cytosine O^2. Additionally, is it possible that it causes the base pair to adopt uncommon pairing schemes which would further destabilize the helix? Hakoshima et al. (1981) have shown that several RNA fibers of poly(C^2-substituted A)·poly(U) produced X-ray diffraction patterns consistent with parallel-stranded RNA double helices with the contiguous reversed Hoogsteen C^2-modified A:U base pairs. This is not highly surprising, as the C^2 substitutions (including methyl and thiomethyl groups) provided extensive bulky hydrophobic groups (completely filling the minor groove) that would be likely to interfere with the uracil O^2 atom, if the C^2-substituted A and U were in the Watson–Crick geometry.

A fundamental, and often asked, question regarding the crystal structure determination is whether the crystal structure is the same as the solution structure. For example, many DNA sequences crystallized readily as A-DNA (Clark et al., 1990), yet when some of those sequences were examined in solution they turned out to be B-DNA. This paradox may be understood by the dynamic property of the DNA, causing them to coexist in many conformations (often in rapid equilibrium). For some DNA sequences they have a slightly higher propensity to adopt the A structure (even though the major population is B), and the crystal packing forces select out the A-DNA. The same question applies to the B-DNA crystals. For example, DNA dodecamers related to d(CGCGAATTCGCG) (DODE) sequences all crystallized with a characteristically narrow minor groove at the AT region (Drew & Dickerson, 1981; Coll et al., 1987; Teng et al., 1988; Sriram et al., 1992a,b). However, it is not entirely clear whether the solution structure retains this characteristic feature. This issue is a relevant one, and it is important to know what DNA conformation the enzyme, such as the *Eco*RI endonuclease, sees in solution.

In this article we address the following questions: What are the respective structures of DNA in crystal in solution? What is the influence of a G→m^2I substitution on the DNA structure, if any? Does *Eco*RI still recognize the modified DNA? We have solved the 3D structure of d(CGC[m^2I]-AATTCGCG) (m^2I-DODE) in the solid state by X-ray crystallography and in solution by NMR and compared the two structures critically. We also compare the enzymatic reactivities of *Eco*RI toward three related dodecamers, DODE, m^2I-DODE, and d(CGC[e^6G]AATTCGCG) (e^6G-DODE).

MATERIALS AND METHODS

The procedure of Roelen et al. (unpublished results) was followed for the synthesis of C^2-methyl-2′-deoxyinosine and O^6-ethyl-2′-deoxyguanosine. The latter nucleosides were converted into the phosphoramidite precursor and incorporated into the oligonucleotides on a Pharmacia DNA synthesis. The synthesized DNA fragment was purified by Sephadex G-50 column chromatography. The purity of the final product was judged to be greater than 95% by HPLC analysis.

X-ray Diffraction Analysis. The sequence m^2I-DODE was selected as it contains a modified *Eco*RI site, so that it may be used as a substrate for the enzyme. Crystallization experiments using the procedure of Wang and Gao (1990) were carried out. We were able to obtain crystals of m^2I-DODE alone as well as its complexes with Hoechst 33258, Hoechst 33342, and netropsin. The crystallization solution of the free DNA contained 1.0 mM dodecamer (single-strand concentration), 50 mM cacodylate buffer at pH 6.5, 13 mM $MgCl_2$, 2 mM spermine, and 2% 2-methyl-2,4-pentanediol (2-MPD), and it was equilibrated against 40% 2-MPD by vapor diffusion at room temperature. Only the free DNA crystal structure is discussed in this article.

The m^2I-DODE crystal was mounted in a thin-walled glass capillary and sealed with a droplet of the crystallization mother liquor for data collection. It was isomorphous to the DODE crystal (Drew & Dickerson, 1981), with the orthorhombic space group $P2_12_12_1$ and unit cell dimensions of a = 25.44 Å, b = 40.91 Å, and c = 66.77 Å. The diffraction data set was collected to a resolution of 2.25 Å at room temperature on a Rigaku AFC-5R rotating-anode diffractometer at a power of 50 kV and 180 mA with graphite-monochromated Cu Kα radiation (1.5418 Å) by the ω-scan mode. Lorentz polarization, absorption, and decay corrections were applied to the data. Data beyond 2.5-Å resolution were weak.

Molecular replacement starting with the DODE crystal structure (Teng et al., 1988) was used to solve the m^2I-DODE structure. The model was refined using the Konnert–Hendrickson constrained refinement procedure (Hendrickson & Konnert, 1979; Westhof et al., 1985). After many cycles of refinement with all available data, the R factor reached ~30% at 2.25-Å resolution. Well-ordered solvent molecules were located from the Fourier ($2|F_o| - |F_c|$) maps. Initially, no hydrogen bond distance constraints were imposed on the two m^2I:C base pairs during the refinement to avoid any bias in the base-pairing scheme. At later stages of the refinement, Watson–Crick geometry was judged to exist and the constraints were restored. The final crystallographic R factor is 20.7% using 1459 reflections ($>2\sigma(F)$) with the inclusion of 51 waters. The RMSD for the bond lengths from the ideal values is 0.011 Å. A summary of the crystal data and refinement parameters is listed in Table 1S of the supplementary material. A sample electron density map is shown in Figure 1. The final atomic coordinates of the structure have been deposited at Brookhaven Protein Databank. The torsional angles of the refined crystal structure are listed in Table 2S of the supplementary material.

NMR Analysis. The DNA solution for NMR studies was prepared by dissolving the ammonium salt of m^2I-DODE in 0.5 mL of phosphate buffer solution (50 mM sodium phosphate, pH 7.0, and 0.15 M NaCl in 99.8% D_2O) to produce a final single-strand concentration of 2 mM. The solution was lyophilized twice with 99.8% D_2O and then dried in an NMR tube with a stream of argon gas, and finally 0.5 mL of 99.96% D_2O was added to produce the sample.

Both 1D and 2D NMR spectra were recorded on a GE GN500 500-MHz spectrometer. The chemical shifts (in ppm) are referenced to the HDO peak, which is calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate at different temperatures. Phase-sensitive NOESY spectra were recorded as 2×512 t_1 blocks of 2048 complex points each (in the t_2 dimension) and averaged for 16 scans per block. The total recycle delay was 3.5 s, and the mixing time was 200 ms for the NOESY experiments. The 2D data sets were processed with the program FELIX v1.1 (Hare Research, Woodinville, WA) using Silicon Graphics workstations. For the 2D NOE data sets, the 2048 complex points in the t_2 dimension were apodized with a sine-bell-squared function for the last quarter of the data to reduce truncation artifacts from the end of the FID and then exponentially multipled by a line broadening of 4 Hz. The 512 complex points in the t_1 dimension were apodized similarly and zero-filled to 2048 points prior to the Fourier transform.

The structure refinement of the complex has been carried out by an improved version of the procedure SPEDREF

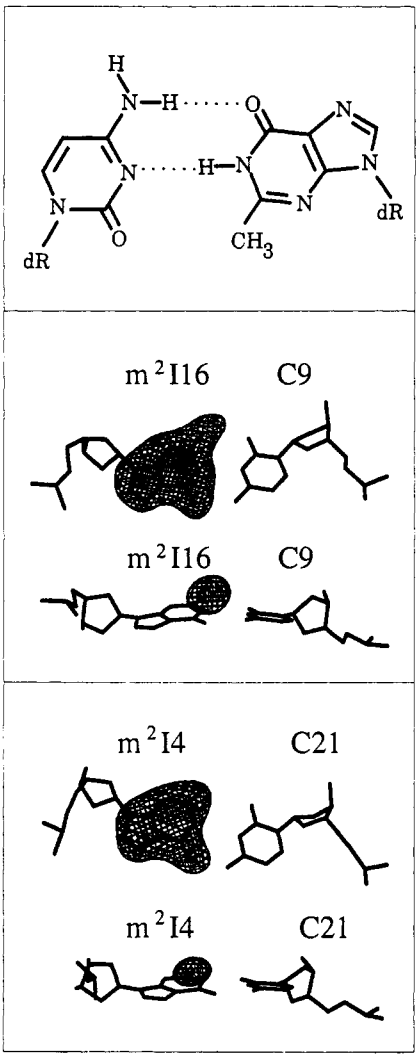

FIGURE 1: (Top) Watson–Crick m²I:C base pair. (Middle and Bottom) Stereoscopic ($|F_o| - |F_c|$) difference Fourier maps displaying the m²I16:C9 and m²I4:C21 base pairs of m²I-DODE, respectively. Either the entire m²I base or the methyl group of m²I was removed from the phase contribution for map calculation.

(Robinson & Wang, 1992). Our approach makes use of *every* measurable NOE cross-peak intensity in the 2D NOE spectrum. The mixing time of 200 ms was selected as it gave an optimum number of NOE observables without severe spin diffusion. Inversion recovery was used to determine the T_1 relaxation time for every spin, with an average T_1 of 2.25 s (the longest T_1 is 4.35 s for the A5H² proton). The recycle time of 3.5 s is about 1.5 times the average T_1, amounting to about 80% of a complete recovery. The correlation time τ_c was selected to be 7.7 ns using the procedure described before (Robinson & Wang, 1992). The selection of τ_c included the optimization of the agreement for the target distance for the H⁵ and H⁶ vector of the four cytosines (ideal distance 2.458 Å). The definition of the R factor has been described (Robinson & Wang, 1992), and it has been used similarly by others (Krisna et al., 1978).

Restriction Enzymatic Analysis. The enzymatic reactivities of *Eco*RI toward three related dodecamer substrates, DODE, m²I-DODE, and e⁶G-DODE, were analyzed as follows. A 1.3 μM DNA sample was incubated with 2000 units of *Eco*RI (Bethesda Research Lab., catalog no. 5202SH) in 0.5 mL of reaction buffer at room temperature (~22 °C). Aliquots (50 μL) of the reaction mixture were taken out every 2 h, and the reaction was stopped by adding the loading dye. Six samples were taken for the 12-h time course. The reaction in the remaining solution was finally stopped after 31 h. The reaction at every time point was analyzed by gel electrophoresis on a 20% nondenaturing polyacrylamide gel (20:1 cross-linked), run at 130 V for 2 or 4 h in Tris–borate–EDTA buffer at room temperature. The gel was stained in a 0.5 μg/mL ethidium bromide solution for 20 min and destained for 10 min in water. The gel was then photographed with a Polaroid camera under UV light. A second set of experiments was carried out in a similar manner, but with 0.51 μM DNA and 20 units of *Eco*RI (BRL, catalog no. 52025A) in a 20-μL final volume at 37 °C, run for 4 h in gel.

RESULTS AND DISCUSSION

Crystal Structure. Figure 2 displays the overall structure of the m²I-DODE double helix, which is similar to other related dodecamers (Drew & Dickerson, 1981; Coll et al., 1987; Teng et al., 1988; Sriram et al., 1992a,b). They all belong to the B-type double helix. Several other gross features such as the 18° bend of the helix axis and the characteristically narrow minor groove at the AATT region are conserved. The similarity is reflected in the relatively small RMSD (0.68 Å) between the m²I-DODE and the native DODE structure (Drew & Dickerson, 1981). In view of this, we focus our attention on only a few salient features.

In most of the DODE-related structures, the base pairs in the central AT region have either high propeller twist (ω) or buckle (κ) angles. In the present structure, the four AT base pairs have similar properties. For example, A6–T19, T7–A18, and T8–A17 maintain high pr tw angles of –27°, –27°, and –25° respectively (Table I). The A5–T20 base pair has a low pr tw angle (–4°) but a slightly higher buckle angle (–13°). The distances between N⁶ of A5 and O⁴ of T20 (2.86 Å) or O⁴ of T19 (3.30 Å) and N⁶ of A17 and O⁴ of T8 (2.77 Å) or O⁴ of T7 (2.97 Å), respectively, satisfy the condition of the interbase bifurcated hydrogen bond (Coll et al., 1987; Nelson et al., 1987). The two G:C base pairs at both ends of the helix are involved in the minor groove interhelix lattice interactions using the G14:G24# and G12:G2# (# indicates a symmetry-related duplex) base pairings. Therefore, the conformations of the terminal and the penultimate base-paired (C1:G24, G2–C23, C11–G14, and G12:C13) nucleotides are strongly influenced by those crystal packing interactions. This should be kept in mind when the crystal structure is compared to the solution structure.

Both m²I4:C21 and m²I16:C9 base pairs in m²I-DODE adopt normal Watson–Crick configurations (Figure 1). In a Watson–Crick G–C base pair, the hydrogen bond distances between the acceptors and the donors are as follows: CN⁴–GO⁶, 2.85 Å; CN³–GN¹, 2.93 Å; and CO²–GN², 2.89 Å (Gessner et al., 1989). Replacement of the N^2-amino group by a methyl group will create a close contact between the methyl group and the cytosine O² oxygen, since the sum of their van der Waals radii is ~3.2 Å. The actual distance of the methyl of m²I and the O² of C is 3.04 Å (average of two). This causes the m²I:C base pair to be distorted from planarity with a somewhat large buckle or propeller twist angle. The m²I4:C21 base pair has a –11° buckle and 1° pr tw angle, while the m²I16:C9 base pair has a 6° buckle and a –19° pr tw angle (Table I).

NMR Refinement. The complete experimental 2D NOESY spectrum is shown in Figure 3. All NOE cross peaks, except some associated with the H⁵′/H⁵″ protons, have been unambiguously assigned. The ambiguity due to those overlapping H⁵′/H⁵″ resonances did not seriously affect the refinement outcome, since the cross-peak NOE intensities associated with H⁵′ and H⁵″ do not vary much, regardless of whether a

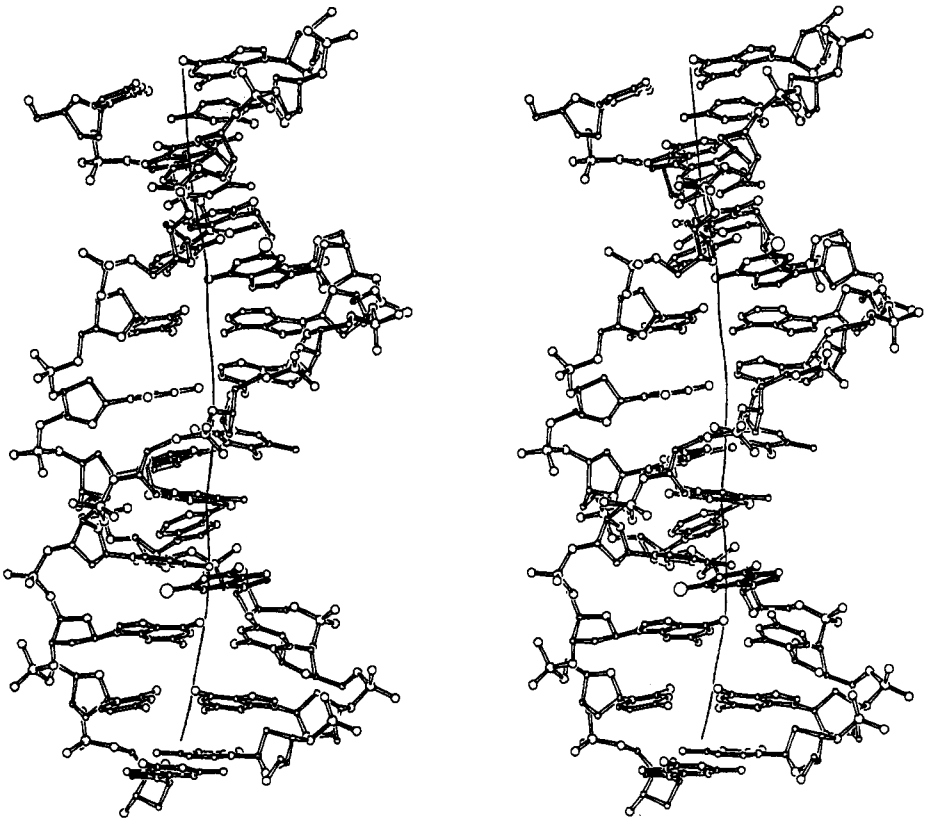

FIGURE 2: Stereoscopic skeletal drawing of the structure of the d(CGC[m²I]AATTCGCG) duplex. The two m²I's are drawn with filled bonds.

Table I: Selected Helical Parameters[a] for Different Models

	P			Ω			ω			κ		
residue	F-R	X-R	X-ray[b]	F-R	X-R	X-ray	F-R	X-R	X-ray	F-R	X-R	X-ray
C1	137	106	143 (167)	38	43	45	−5	−13	−15	22	4	−7
G2	161	185	228 (185)	38	35	36	−4	−32	−11	−6	2	1
C3	100	96	35 (66)	37	29	30	−9	−3	−1	2	9	13
m²I4	173	160	147 (178)	36	39	34	−1	−6	1	−20	−12	−11
A5	185	179	218 (31)	38	39	39	−13	−28	−4	−11	−1	−13
A6	146	147	148 (47)	33	27	29	−5	−21	−27	−9	1	−5
T7	98	108	139 (179)	38	39	37	−5	−21	−27	8	−1	−5
T8	147	122	201 (217)	36	39	40	−13	−28	−25	11	1	−6
C9	121	103	207 (185)	37	29	29	−1	−6	−19	20	12	6
G10	165	151	168 (171)	38	35	38	−9	−3	1	−2	−9	−19
C11	143	119	175 (169)	38	43	46	−4	−32	−27	6	−2	−11
G12	165	164	4 (327)				−5	−13	7	−22	−4	−11

[a] Nomenclature of the helical parameters follows that of Dickerson et al. (1989). *P* is the pseudorotation angle, Ω is the helical twist angle, κ is the base-pair buckle, and ω is the propeller twist angle, all in degrees. [b] The numbers in parentheses are for the second strand of the X-ray structure. F-R is the FIBREF model and X-R is the X-RAYREF model.

resonance was assigned as $H^{5'}$ or $H^{5''}$. There are 111 unique resonances of nonexchangeable protons and therefore 6105 possible cross peaks. We have measured 1518 NOEs which were considered to be above the noise level, and they are used in the refinement.

The crystal structure described above showed that the molecule adopted a B-type double helix. The solution structure was presumed to be in the B-DNA family. How closely was the crystal structure preserved in solution? This may be tested by using the X-ray structure as a starting model for the SPEDREF refinement procedure (Robinson & Wang, 1992). However, the crystal structure of the m²I-DODE duplex does not have an exact 2-fold symmetry due to the crystal packing constraint (the RMSD between the two halves of the duplex is 0.478 Å), so symmetry is added in the refinement procedure. The *R* factor between the experimental NOEs and the simulated NOEs, calculated on the basis of the X-ray model, is 49.8%, suggesting that this model is different from the solution structure. We proceeded to refine the model by carrying out 120 SPEDREF refinement cycles to converge to an *R* factor of 14.3%.

We also carried out a parallel refinement study using the energy-minimized fiber model as the starting model. Interestingly, the initial *R* factor was 33.2%, which is significantly lower than that from the crystal model. By inspecting the cross-peak pattern of a key region (aromatic to aromatic) of the 2D NOE spectrum (Figure 4), it is clear that there are changes in the relative positions (e.g., helical twist angles) between adjacent bases of the two models. For example, some of the simulated NOE cross peaks (Figure 4, bottom panel, above diagonal), including $C1H^6$–$G2H^8$, $C3H^6$–$I4H^8$, and $A6H^8$–$T7H^6$, appear to be too strong as compared to the experimental NOEs (Figure 4, top panel). After the refinement, the agreement is substantially better (*R* = 13.8%) (Figure 4, bottom panel, below diagonal) and the refined model is denoted FIBREF. Portions (aromatic to $H^{1',5}$ and aromatic to $H^{2',2''}$) of the simulated 2D NOESY based on FIBREF are shown in Figure 5, and they are similar to their corresponding

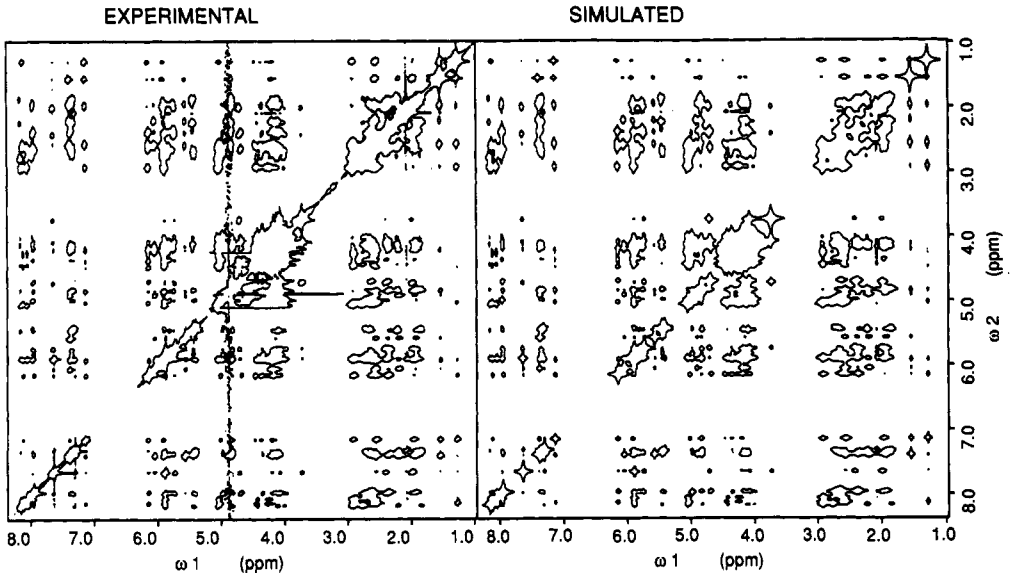

FIGURE 3: Complete experimental and simulated two-dimensional NOESY spectra of d(CGC[m²I]AATTCGCG) at a mixing time of 200 ms. The simulated spectrum was calculated with the refined model of FIBREF as described in the text.

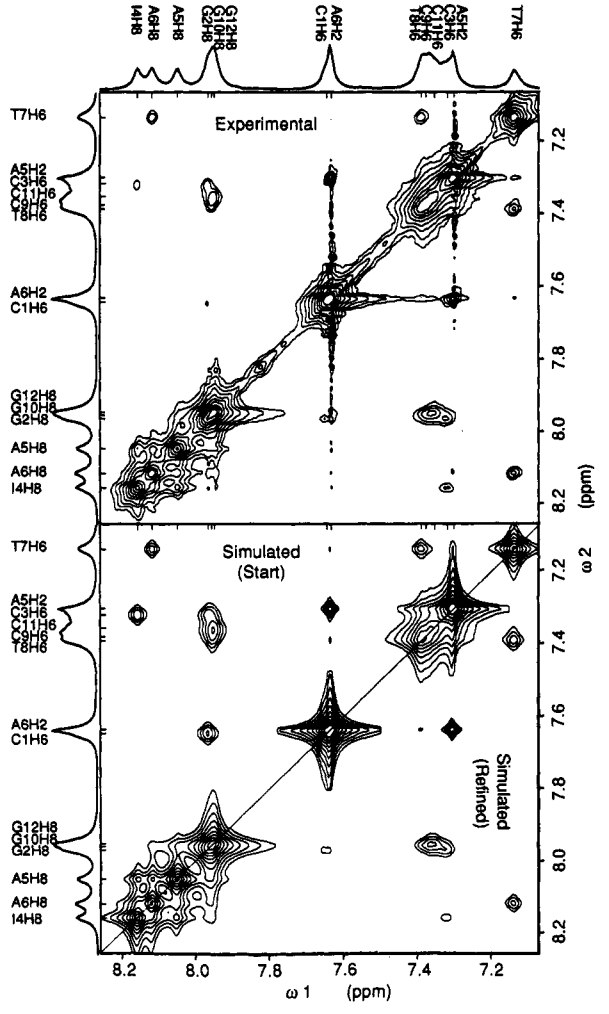

FIGURE 4: Expanded aromatic to aromatic NOE cross-peak region. (Top) Experimental. (Bottom) Upper left half is the simulated spectrum based on the fiber starting model, and the lower-right half is the simulated spectrum based on the FIBREF refined model.

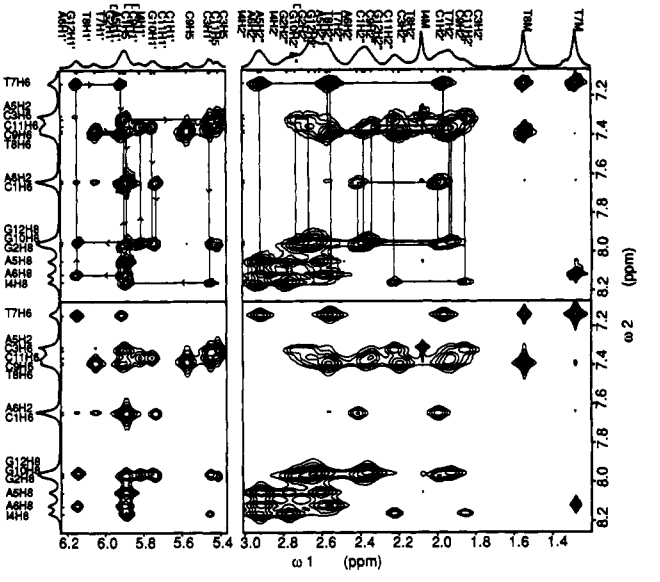

FIGURE 5: Portions (cross peaks between the $H^{1',5,2',2''}$ protons and the aromatic protons) of the 2D NOESY spectra showing the agreement between the experimental (top two panels) and the simulated spectra based on the FIBREF model (lower two panels).

experimental spectra. The stereoscopic drawing of the FIBREF structure is displayed in Figure 6A. The torsional angles of the FIBREF model are listed in Table 3S of the supplementary material.

NMR-Refined Structures. An important question to ask is the following: which solution structure is "correct"? This is difficult to answer. A number of approaches have been taken by different investigators (Suzuki et al., 1986; Gupta et al., 1988; Sarma et al., 1988; Clore & Gronenborn, 1989; Metzler et al., 1990; James, 1991; Koning et al., 1991; Kim & Reid, 1992), and different conclusions have been reached. The FIBREF structure (R = 13.8%) has a (marginally) lower R factor than the X-RAYREF structure (R = 14.3%), and both have good chemical energies. When these two structures are compared globally, they look somewhat different with an RMSD of 1.23 Å (Figure 6B).

This apparent nonconvergence of the global structures is not surprising as the NOE intensity is principally a short-range (<6 Å) phenomenon. In addition, the paucity of protons on a DNA molecule and the linear nature of the helix make it likely to have multiple conformers that agree with the same NOE data equally well. While it is expected that the local structures between the FIBREF and X-RAYREF models

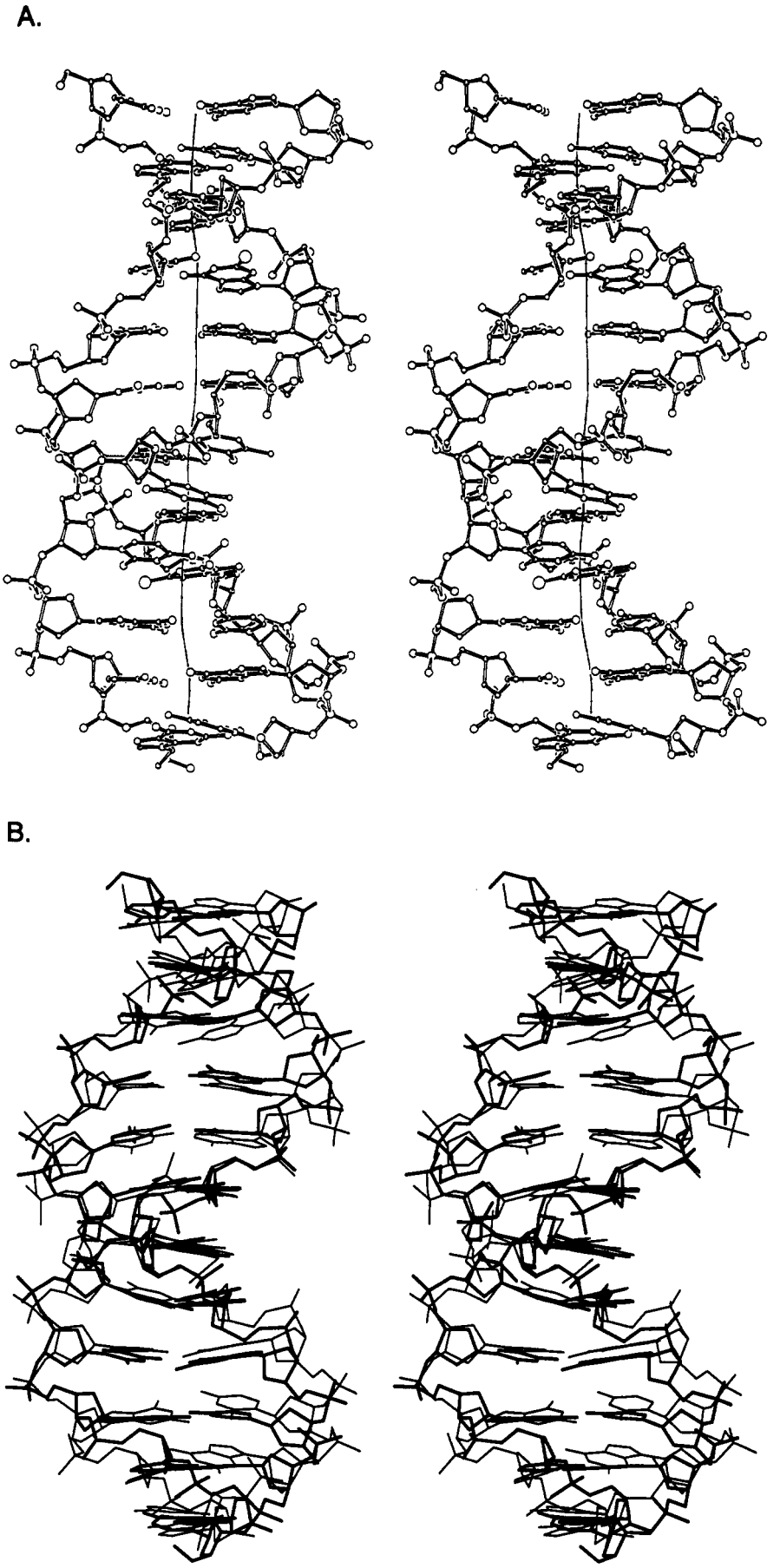

FIGURE 6: (A) Stereoscopic skeletal drawing of the FIBREF refined structure. (B) Least-squares superposition of the X-RAYREF and the FIBREF (thick lines) refined structures (RMSD = 1.23 Å).

should agree more closely, the refined global structures may deviate somewhat depending on the starting models.

We earlier showed that NOE refinement is capable of driving different starting models to reasonable convergence

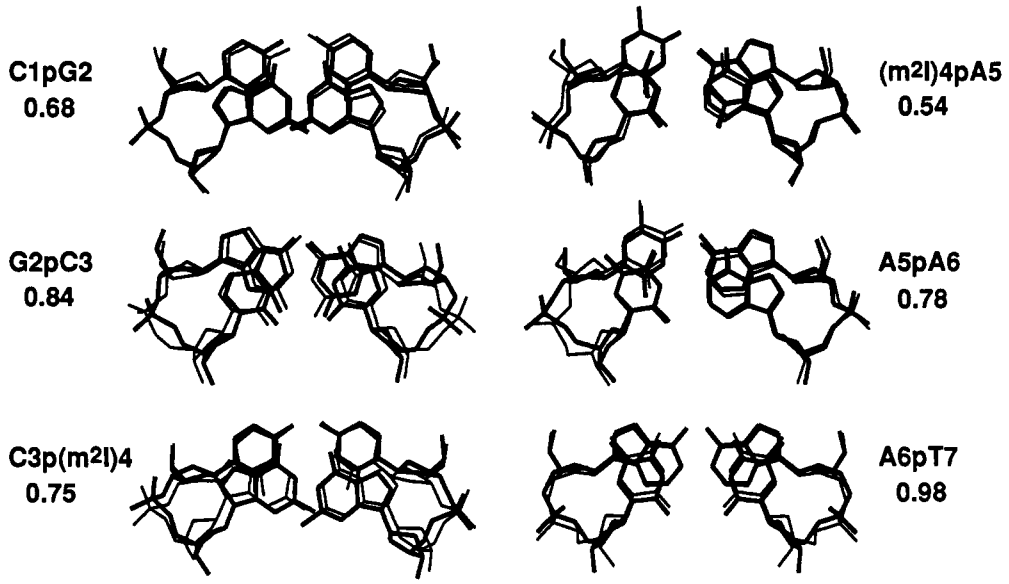

FIGURE 7: Detailed comparison of the six dinucleotide steps by a least-squares fit between the two refined models, FIBREF (thin lines) and X-RAYREF (thick lines). The nucleotide sequences are shown on the left, and the RMSD is shown on the right.

(Robinson & Wang, 1992; Jaishree et al., 1993). To illustrate this point, we have generated an overwound dodecamer model (5 deg/step) and subjected it to refinement. The starting overwound model had an RMSD of 1.9 Å with the fiber B-DNA model (Figure 1S of the supplementary material). After 20 cycles of refinement, the RMSD between the refined model (starting overwound) and FIBREF is ~0.8 Å.

This local/global issue was further examined by comparing the six unique dinucleotide steps (Figure 7). They show a good agreement between the FIBREF and X-RAYREF models, with the RMSDs ranging from 0.54 (m^2I4pA5 step) to 0.98 Å (A6pT7 step). Overall, the helical conformations (e.g., helical twist angles) do not vary greatly (Table I). This indicates that in a dinucleotide step the intranucleotide and the internucleotide NOEs are sufficient to pull the local structures to convergence, as reflected in the pseudorotation angles. Whereas most of the sugars are refined to an average pseudorotation angle (P) of ~160° ($C^{2'}$-endo), three pyrimidine residues (C3, T7, and C9) are refined to an average value of ~100° ($C^{3'}$-exo). In fact, the pyrimidines seem to have a consistently lower P value (av 117°) than the purines (av 165°). Since each nucleotide has about 126 NOEs for fixing the conformation, this variation in P is likely to be reliable.

It has been suggested that J-coupling constants may be used as additional input constraints in the refinement, although there is still no agreement about how best to use this information. Some have suggested that a three-state sugar pucker model is needed to generate a simulated COSY pattern that matches the observed COSY pattern (Schmitz et al., 1990), while others argue to the contrary (Salazar et al., 1993). We measured, from the PE-COSY spectrum (Figure 2S of the supplementary material), the J-coupling constants (e.g., $H^{1'}$–$H^{2''}$ and $H^{1'}$–$H^{2'}$) to estimate the sugar pseudorotation angles. Some COSY patterns are well-defined (e.g., C1 and C3), while others are weak and diffuse (e.g., m^2I4 and A4). The measured J values agree with the calculated values derived from a general class of the pucker range, i.e., of the S type (Table 4S of the supplementary material).

The agreement between the experimental and calculated NOEs of selected cross peaks is shown in Table II. Most of those NOEs are from the nearest-neighbor intrastrand cross peaks (<6 Å). However, there are several cross peaks that are from the interstrand proton pairs (e.g., $A6H^2$–$T8H^{1'}$). Those NOEs help define the global structure more definitively. For example, the $A6H^2$–$T8H^{1'}$ cross peak can only occur to a significant extent when the minor groove of the double helix is narrow. This has been demonstrated in the NMR study of a number of oligonucleotides with the A_nT_n sequence (Behling et al., 1987; Sarma et al., 1988; Celda et al., 1989; Katahira et al., 1990). Naturally, we ask whether the SPEDREF refinement resulted in structures with a narrow minor groove in the dodecamers.

Table II: Comparison of NOEs for Selected Spin Pairs between the Experimental and the Simulated Spectra of Starting and Refined Models

		simulated NOE[b]	
proton pairs	observed NOE[a]	start[c]	refined[c]
	Aromatic to Aromatic		
C1H6–G2H8	76	329 (4.86 Å)	85 (5.59 Å)
G2H8–C3H5	232	252 (3.69 Å)	180 (3.91 Å)
G2H8–C3H6	129	210 (4.90 Å)	152 (5.01 Å)
C3H6–m^2I4H8	84	241 (5.08 Å)	84 (5.30 Å)
m^2I4H8–A5H8	121	160 (4.96 Å)	151 (4.79 Å)
A5H2–A6H2	194	230 (3.83 Å)	233 (3.83 Å)
A5H8–A6H8	100	215 (4.96 Å)	135 (5.29 Å)
A6H8–T7H6	154	190 (4.96 Å)	195 (5.12 Å)
T7H6–T8H6	147	206 (4.97 Å)	223 (4.93 Å)
T8H6–C9H5	244	223 (4.00 Å)	242 (3.50 Å)
T8H6–C9H6	255	209 (5.07 Å)	213 (4.73 Å)
C9H6–G10H8	110	263 (5.07 Å)	116 (5.76 Å)
G10H8–C11H5	217	238 (3.70 Å)	212 (3.85 Å)
G10H8–C11H6	130	202 (4.91 Å)	146 (5.17 Å)
C11H6–G12H8	149	258 (5.14 Å)	169 (5.15 Å)
	Interstrand		
A5H2–T20H1′	24	24 (5.28 Å)	29 (5.24 Å)
A5H2–C21H1′	21	39 (4.86 Å)	25 (5.31 Å)
A6H2–A18H2	79	166 (4.06 Å)	107 (4.37 Å)
A6H2–T19H1′	23	22 (5.36 Å)	25 (5.20 Å)
A6H2–T20H1′	94	30 (5.06 Å)	101 (4.18 Å)
A6H2–T20H4′	21	12 (6.21 Å)	21 (5.77 Å)
T7H1′–A18H2	23	23 (5.33 Å)	25 (5.20 Å)
T8H4′–T20H5′	22	11 (5.75 Å)	20 (5.69 Å)
T8H4′–T20H4′	87	16 (5.51 Å)	54 (4.30 Å)
m^2I4Me–C21H1′	33	40 (4.85 Å)	30 (5.12 Å)
m^2I4Me–G22H1′	39	64 (4.39 Å)	77 (4.30 Å)

[a] NOE intensity is calculated on an arbitrary scale. [b] Interproton distance in the models. [c] Starting model is fiber B-DNA and the refined model is F-REF.

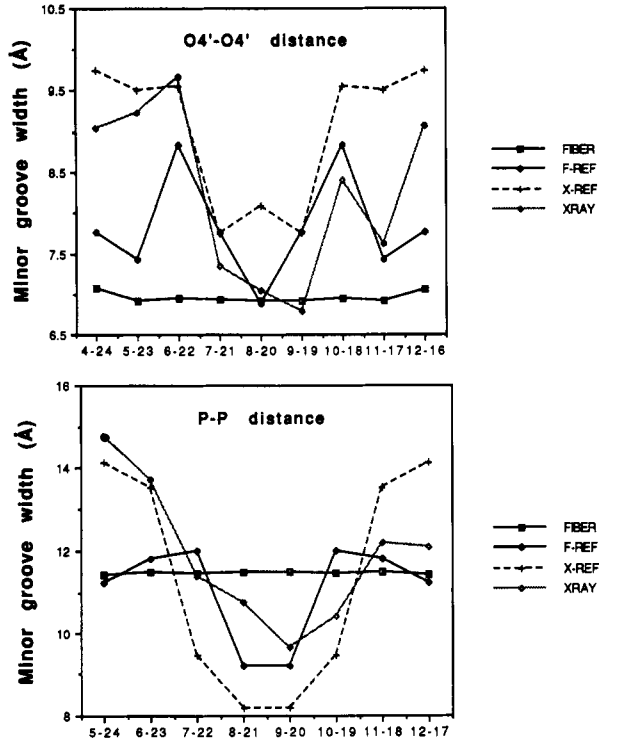

FIGURE 8: Comparison of the minor groove widths as measured by the O4'–O4' or P–P distances across the groove among different models: (1) starting fiber model (FIBER); (2) NMR-refined fiber model (F-REF); (3) NMR-refined X-ray model (X-REF); and (4) crystal structure X-ray model. All models show a smaller minor groove width near the center of the helix, except for the starting idealized fiber model.

Figure 8 shows the comparison of the minor groove width as measured by the O4'–O4' distance or the P–P distance across the groove. It is clear that the general trend of a narrow minor groove in the central AATT segment of the helix is present in all models. The results suggest that, when all NOEs are used in the refinement, important global structural features can be obtained. This has also been shown in the structures of two other related dodecamers, e⁶G-DODE and d(CGCAAATTTGCG), refined by SPEDREF (D. Yang and A. H.-J. Wang, unpublished results).

NMR vs X-ray Structures. In the crystal structure there is a 18° bending of the helix axis, most likely due to the crystal packing forces (Drew & Dickerson, 1981; Coll et al., 1987; Teng et al., 1988; Sriram et al., 1992a,b). Does the NMR refinement detect this as well? The helix axes for the two refined structures, defined by the program CURVE (Lavery & Skelner, 1989), are nearly straight. Clearly, the X-ray crystal and NMR solution structures are different. This was demonstrated and supported by a laser Raman study of d(CGCAAATTTGCG) in the crystalline state and in solution (Benevides et al., 1988). Benevides et al. (1988) showed that many of the fingerprint Raman bands are significantly broader in the crystal Raman spectrum than those recorded from solution. This was attributed to the fact that the DNA conformation is more heterogeneous in the crystal. Our study here reaffirms this conclusion, as supported by the more uniform distribution of the helical parameters for the FIBREF or X-RAYREF NMR structures (Figure 9).

What does an NMR structure mean? It has been suggested that DNA in solution is dynamic. The deoxyribose ring is constantly equilibrated between the C2'-endo (S-type) and the C3'-endo (N-type) pucker conformations. In B-DNA the sugars are mostly in the S-type conformation (Drew & Dickerson, 1981). Our refined structure (e.g., F-REF) is

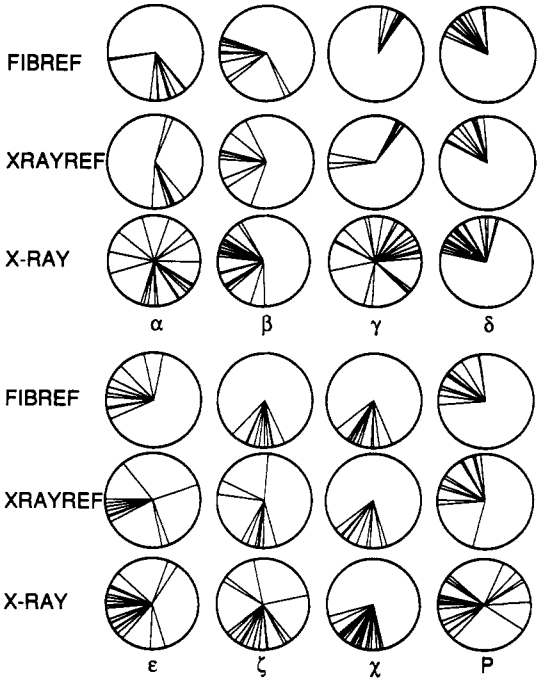

FIGURE 9: Comparison of the backbone conformational angles (α–ζ, the glycosyl angle χ, and the pseudorotation angle P) among the three models, NMR FIBREF model, NMR X-RAYREF model, and the crystal structure X-ray model. The two NMR models have narrower ranges in their angular distributions than those of the X-ray model.

represented by a static picture with a fixed set of sugar pseudorotation angles (Table I), but it should not be viewed as the only conformer that exists in solution. Instead, it is a model that best accounts for all of the NOE data, while still making sense chemically. The model should be viewed as a representative structure of an ensemble of many conformations. Several investigators have demonstrated that the sugar puckers of the DNA molecules are best interpreted as having varying S/N ratios by using the simulation of *J*-coupling COSY patterns (Altona, 1982; Schmitz et al., 1990).

Which structure is more relevant, the crystal structure of the solution structure? This cannot be answered easily. These two techniques, X-ray diffraction and NMR, have their respective strengths and limitations. Our comparative study provides a good illustration of those points. The X-ray structure analysis of the m²I-DODE at ~2.25 Å provides an unambiguous picture of the global features (e.g., bend in the helix axis) and many of the helical parameters (e.g., propeller twist, base-pair type), but it cannot define other structural features, such as sugar pucker, nearly as conclusively. The influence of crystal lattice forces on the DNA conformation is often a concern as well. However, this may be viewed as a reflection of the desirable property of DNA, which is highly plastic and is capable of adapting to environmental factors including proteins. Indeed, the conformation of the DNA tridecamer is distorted (with "kinks") in the complex of *Eco*RI and DNA (Kim et al., 1990).

Finally, we used the *Eco*RI restriction enzyme to probe the effect of a chemical modification in the guanine of the 5'-GAATTC restriction sequence of the three related dodecamers, DODE, m²I-DODE, and e⁶G-DODE. The cutting patterns in Figure 10 showed that DODE was almost completely digested after 4 h under the reaction conditions, whereas the e⁶G-DODE was nearly totally inactive (the intensity of the ethidium stain was only slightly diminished after 31 h). This was to be expected as the modification of guanine on the O⁶ (in the major groove) position, which caused

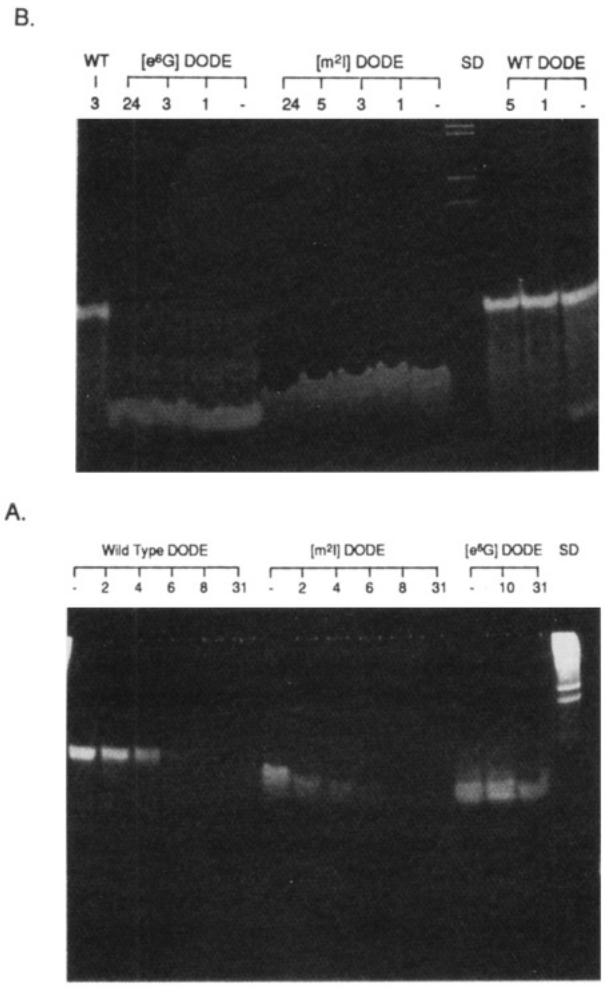

FIGURE 10: Digestion of three related DNA dodecamers, DODE, m^2I-DODE, and e^6G-DODE, by the restriction endonuclease *Eco*RI. (A) Lanes are numbered from left to right. The gel was run for 2 h under a higher *Eco*RI/DNA (1584 units/μM) ratio. Lanes 1 and 7 are the controls (no enzyme added), while lanes 2–6 and 8–12 are the reaction time course at 2, 4, 6, 8, and 31 h for DODE and m^2I-DODE, respectively. Lanes 13–15 are the control and the reaction product at 10 and 31 h for e^6G-DODE, respectively. Only the intact duplex molecule was stained, while the digested fragments (e.g., CGCG and AATTCGCG) did not stain well with ethidium, presumably due to the unstable duplex structure under the gel running conditions. The outer left lane was standards (SD) (New England Biolabs catalog no. 303-2S). Notice that among the three related DNA dodecamers, DODE, m^2I-DODE, and e^6G-DODE, the native DODE has the lowest mobility whereas e^6G-DODE has the fastest mobility. This observation may be related to their respective stabilities of the duplex molecule under the gel running conditions (B). The gel was run for 4 h under a lower *Eco*RI/DNA (40 units/μM) ratio at 37 °C. Lanes are numbered from right to left. Lanes 1, 5, and 10 are the controls (no enzyme added), while lanes 2, 3, and 14, 6–9, and 11–13 are the reaction products at indicated time points for DODE, m^2I-DODE, and e^6G-DODE, respectively. The difference in mobility is seen more clearly here.

the base pair to adopt a wobble or bifurcated configuration (Sriram et al., 1992a,b), should have a strong interfering effect for enzyme recognition and binding and therefore cause a substantial reduction in enzyme activity. A similar result was also seen in the enzyme activities toward m^6G-containing oligonucleotides (Voigt & Topal, 1990).

The m^2I-DODE was readily cut by *Eco*RI (Figure 10). This is evident by a comparison of lane 8 (m^2I-DODE at 2 h) and lane 2 (DODE at 2 h) of the bottom panel in which significant intensity loss of the ethidium stains in lane 8 is seen. McLaughlin et al. (1987) have shown that d(C-T*I*AATTCAG) is a slightly less active substrate than the native decamer. The introduction of a methyl group at the C^2 position of the I4 residue did not appear to significantly affect m^2I-DODE as a substrate for *Eco*RI. This is not surprising since the overall structure of m^2I-DODE in solution is nearly the same as that of DODE (RMSD = 0.4 Å) by NMR analysis (D. Yang and A. H.-J. Wang, unpublished data). The Watson–Crick m^2I4–C16 base pair, while slightly distorted, is recognized by the enzyme.

CONCLUSION

In this article, we performed structural analyses of the m^2I-containing DNA dodecamer by a combination of X-ray diffraction and NMR refinements. We showed that the structure in solution is neither the same as the crystal structure nor is it the same as the idealized B-DNA. One conspicuous difference is that the solution structures have a more uniform backbone conformation, an observation consistent with that concluded from the laser Raman study of d(CGCAAATTTGCG) (Benevides et al., 1989). Major common structural features between the crystal and the solution structures include the Watson–Crick m^2I:C base pairs (no Hoogsteen base pair) with minor structural perturbations in the double helix. This suggests that a single isolated methyl group is not bulky enough to provide sufficient hindrance to destabilize the base pair. We have shown previously that the true restriction methylation modification, m^6A, did not alter the structure significantly, as is evident from the crystal structure of d(CGCG[m^6A]ATTCGCG) (Frederick et al., 1988). The role of the methyl group on the N^6 of adenine base is, most likely, simply to provide physical hindrance between the enzyme amino acid side chains and the adenine base. In comparison, the effect of m^2I in replacing the G in the 5′-GAATTC sequence on the enzymatic activity appears to be negligible.

ACKNOWLEDGMENT

We thank Mr. Y. Zhuang for assistance in the gel electrophoresis experiments.

SUPPLEMENTARY MATERIAL AVAILABLE

Two figures showing the comparison of the NMR refined models using different starting models (Figure 1S) and representative PE-COSY spectra (Figure 2S) and four tables showing crystal data (Table 1S), conformational torsional angles of various models (Tables 2S and 3S), and *J*-coupling constants (Table 4S) (7 pages). Ordering information is given on any current masthead page.

REFERENCES

Aiken, C. R., McLaughlin, L. W., & Gumport, R. I. (1991) *J. Biol. Chem. 266*, 19070–19078.
Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas 101*, 413–433.
Behling, R. W., Rao, S. N., Kollman, P., & Kearns, D. R. (1987) *Biochemistry 26*, 4674–4681.
Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1988) *Biochemistry 27*, 931–938.
Brennan, C. A., Van Cleve, M. D., & Gumport, R. I. (1986) *J. Biol. Chem. 261*, 7270–7278.
Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., & Wuthrich, K. (1989) *Biochemistry 28*, 1462–1471.
Clark, G. R., Brown, D. G., Sanderson, M. R., Chwalinsky, T., Neidle, S., Veal, J. M., Jones, R. L., Wilson, W. D., Zon, G., Garman, E., & Stuart, D. I. (1990) *Nucleic Acids Res. 18*, 5521–5528.
Clore, C. M., & Gronenborn, A. M. (1989) *Crit. Rev. Biochem. Mol. Biol. 24*, 479–564.
Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Cruse, W. B. T., Aymami, J., Kennard, O., Brown, T., Jack, A. G. C., & Leonard, G. A. (1989) *Nucleic Acids Res. 17*, 55–72.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekman, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989) *Nucleic Acids Res. 17*, 1797–1803.
Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.
Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1988) *J. Biol. Chem. 263*, 17872–17879.
Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., & Wang, A. H.-J. (1989) *J. Biol. Chem. 264*, 7921–7935.
Gupta, G., Sarma, M. H., & Sarmar, R. H. (1988) *Biochemistry 27*, 7909–7919.
Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron 36*, 2783–2792.
Hakoshima, T., Fukui, T., Ikehara, M., & Tomita, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7309–7313.
Hendrickson, W. A., & Konnert, J. H. (1979) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinvasan, R., Ed.) pp 43–57, Pergamon, Oxford.
Jaishree, T. N., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1993) *Biochemistry 32*, 4903–4911.
James, T. L. (1991) *Curr. Opin. Struct. Biol. 1*, 1042–1053.
Katahira, M., Sugeta, H., Kyogoku, Y., & Fujii, S. (1990) *Biochemistry 29*, 7214–7222.
Kim, S.-G., & Reid, B. R. (1992) *Biochemistry 31*, 12103–12116.
Kim, Y., Grable, J. C., Love, R., Greene, P. J., & Rosenberg, J. M. (1990) *Science 249*, 1307–1309.
Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry 30*, 3787–3797.
Krisna, N. R., Agresti, D. G., Glickson, J. D., & Walter, R. (1978) *Biophys. J. 24*, 791–814.
Kumar, V. D., Harrison, R. W., Andrews, L. C., & Weber, I. T. (1992) *Biochemistry 31*, 1541–1550.
Lavery, R., & Skelnar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655–667.
McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N., & Scholtissek, S. (1987) *Biochemistry 31*, 7238–7245.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711–736.
Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.
Ptashne, M. (1992) *A Genetic Switch*, 2nd ed., Cell Press & Blackwell Scientific Publications, Cambridge, MA.
Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533.
Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., & Reid, B. R. (1993) *Biochemistry 32*, 4207–4215.
Sarma, M. H., Gupta, G., & Sarma, R. H. (1988) *Biochemistry 27*, 3423–3432.
Schmitz, U., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 2357–2368.
Sriram, M., Roelen, H. C. P. F., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1992a) *EMBO J. 11*, 225–232.
Sriram, M., van der Marel, G. A., Roelen, H. C. P. F., van Boom, J. H., & Wang, A. H.-J. (1992b) *Biochemistry 31*, 10510–10517.
Steitz, T. (1990) *Q. Rev. Biophys. 23*, 205–280.
Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–6865.
Teng, M.-K., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671–2690.
Wang, A. H.-J., & Gao, Y.-G. (1990) *Methods (San Diego) 1*, 91–99.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Xuan, J.-C., & Weber, I. T. (1992) *Nucleic Acids Res. 20*, 5457–5467.