ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/19740314

Refinement of the solution structure of the DNA dodecamer 5′d(CGCGPATTCGCG)2 containing a stable purine–thymine base pair: Combined use of nuclear magnetic resonance and restrained...

ARTICLE *in* BIOCHEMISTRY · JUNE 1988

Impact Factor: 3.02 · DOI: 10.1021/bi00411a042 · Source: PubMed

CITATIONS
48

READS
10

7 AUTHORS, INCLUDING:

G. Marius Clore
National Institutes of Health
550 PUBLICATIONS **59,952** CITATIONS
SEE PROFILE

Claudia Scalfi-Happ
Universität Ulm
36 PUBLICATIONS **330** CITATIONS
SEE PROFILE

Available from: G. Marius Clore
Retrieved on: 24 January 2016

Refinement of the Solution Structure of the DNA Dodecamer 5′d(CGCGPATTCGCG)$_2$ Containing a Stable Purine–Thymine Base Pair: Combined Use of Nuclear Magnetic Resonance and Restrained Molecular Dynamics†

G. Marius Clore,*,‡,⊥ Hartmut Oschkinat,‡ Larry W. McLaughlin,§ Fritz Benseler,§,∥ Claudia Scalfi Happ,‡ Erwin Happ,‡ and Angela M. Gronenborn*,‡,⊥

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG, and Department of Chemistry, Boston College, Chestnut Hill, Massachusetts 02167

Received November 3, 1987

ABSTRACT: The solution structure of the self-complementary dodecamer 5′d(CGCGPATTCGCG)$_2$, containing a purine–thymine base pair within the hexameric canonical recognition site GAATTC for the restriction endonuclease *Eco*RI, is investigated by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. Nonexchangeable and exchangeable protons are assigned in a sequential manner. A set of 228 approximate interproton distance restraints are derived from two-dimensional nuclear Overhauser enhancement spectra recorded at short mixing times. These distances are used as the basis for refinement using restrained molecular dynamics in which the interproton distance restraints are incorporated into the total energy function of the system in the form of effective potentials. Eight calculations are carried out, four starting from classical A-DNA and four from classical B-DNA. In all cases convergence to very similar B-type structures is achieved with an average atomic root mean square (rms) difference between the eight converged structures of 0.7 ± 0.2 Å, compared to a value of 6.5 Å for that between the two starting structures. It is shown that the introduction of the purine–thymine mismatch does not result in any significant distortion of the structure. The variations in the helical parameters display a clear sequence dependence. The variation in helix twist and propeller twist follows Calladine's rules and can be attributed to the relief of interstrand purine–purine clash at adjacent base pairs. Overall the structure is straight. Closer examination, however, reveals that the central 5 base pair steps describe a smooth bend directed toward the major groove with a radius of curvature of ~38 Å, which is compensated by two smaller kinks in the direction of the minor groove at base pair steps 3 and 9. These features can be explained in terms of the observed variation in roll and slide.

Sequence-specific "recognition" between proteins and nucleic acids can generally be understood in terms of functional group contacts between the amino acid side chains of the protein and functional groups available on the nucleobases. It is likely in these cases that the protein recognizes or interacts with a unique functional group pattern characteristic of a particular nucleic acid sequence. Oligodeoxynucleotides represent ideal substitutes for the longer DNA polymers in detailed studies of these phenomena. The incorporation of modified nucleobases into oligonucleotides provides a useful approach for dissecting recognition processes in that the functional group pattern available within a given sequence can be systematically altered. This approach has been employed in DNA binding studies involving the *lac* repressor (Yansura et al., 1977, 1979; Goeddel et al., 1977, 1978; Fischer & Caruthers, 1979; Caruthers, 1980), restriction endonucleases (Dwyer-Hallquist et al., 1982; Ono et al., 1984; Yolov et al., 1985; Seela & Driller, 1986; Jiricny et al., 1986; Fliess et al., 1986; Brennan et al., 1986a; McLaughlin et al., 1987), modification enzymes (Brennan et al., 1986b), and RNA polymerase (Dubendorff et al., 1987). The modifications that are often most useful in this respect are those that result in the deletion of a functional group from the sequence of interest. The substitution of 2′-deoxyuridine for 2′-deoxythymidine is one of the simplest such functional group deletions in which the thymine methyl group is removed. In a similar fashion, the substitution of 2′-deoxynebularine for 2′-deoxyadenosine replaces the adenine base with the purine base, resulting in the deletion of the exocyclic amino group of adenine. The purine base has been incorporated into oligonucleotides for the study of base pairing (Eritja et al., 1986) as well as interactions with the *Hin*II, *Sal*I, *Taq*I (Jiricny et al., 1986), and *Eco*RI (McLaughlin et al., 1987) restriction endonucleases. More recently, a preliminary nuclear magnetic resonance (NMR)[1] analysis has indicated that purine will base pair with thymidine in a Watson–Crick manner (Ikuta et al., 1987).

† This work was supported by the Max-Planck Gesellschaft and Grant 321/4003/0318909A from the Bundesministerium für Forschung und Technologie (G.M.C. and A.M.G.) and by Grant NSF-DMB-8519840 from the National Science Foundation (L.W.M.).

‡ Max-Planck-Institut für Biochemie.

⊥ Present address: Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.

§ Boston College.

∥ Permanent address: Max-Planck-Institut für experimentelle Medizin, D-3400 Göttingen, FRG.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; HOHAHA, two-dimensional homonuclear Hartmann–Hahn spectroscopy; ROESY, two-dimensional spin-locked NOE spectroscopy; E-COSY, two-dimensional homonuclear exclusive correlated spectroscopy; rms, root mean square; RD, restrained dynamics; TLC, thin-layer chromatography; EDTA, ethylenediaminetetraacetic acid.

0006-2960/88/0427-4185$01.50/0 © 1988 American Chemical Society

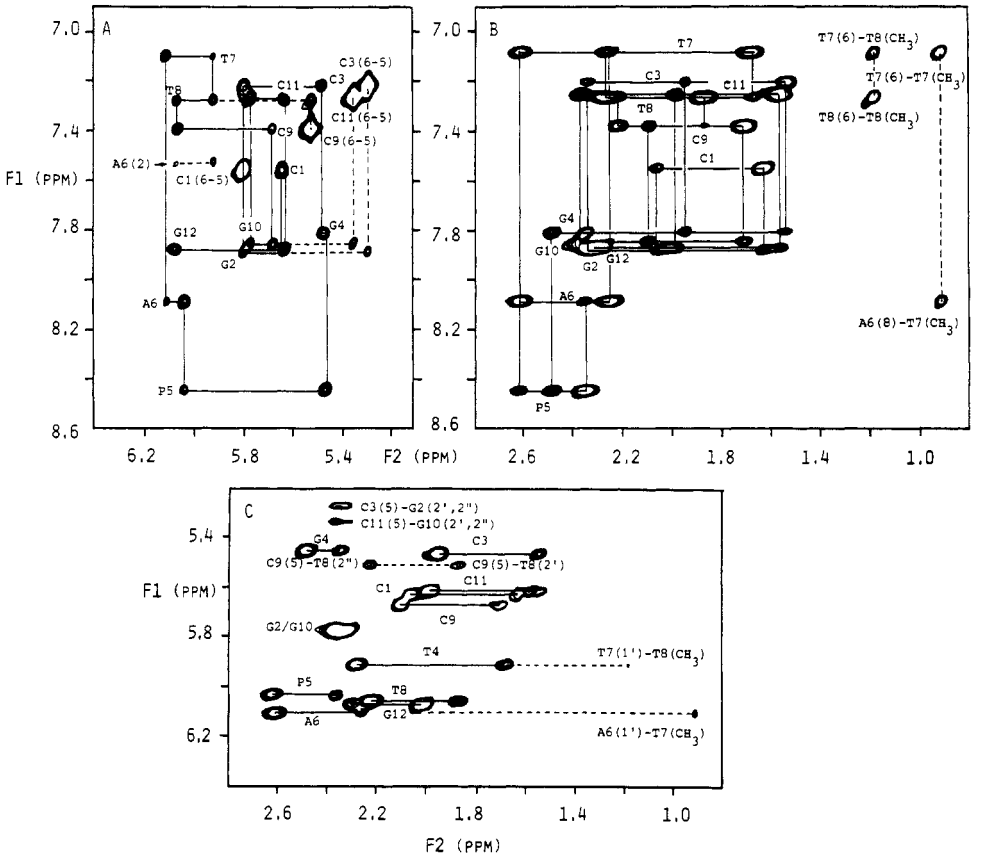

FIGURE 1: (A) H8/H6($F1$ axis)–H1′/H5($F2$ axis), (B) H8/H6($F1$ axis)–H2′/H2″/CH_3($F2$ axis), and (C) H1′/H5($F1$ axis)–H2′/H2″/CH_3 ($F2$ axis) of the 40-ms NOESY spectrum of the dodecamer in D_2O at 5 °C. Sequential H1′(i – 1) ↔ H8/H6(i) ↔ H1′(i) and H2′/H2″(i – 1) ↔ H8/H6(i) ↔ H2′/H2″(i) connectivities are shown in (A) and (B), respectively, as solid lines (—); H8/H6(i – 1) ↔ H5(i) and H8/H6(i – 1)–CH_3(i + 1) connectivities are shown in (A) and (B), respectively, as dashed lines (- - -).

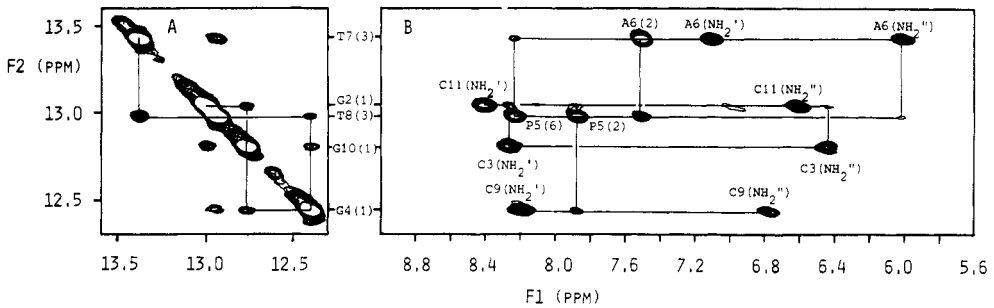

FIGURE 2: (A) NH($F2$ axis)–NH($F1$ axis) and (B) NH($F2$ axis)–aromatic/NH2($F1$ axis) of the 150-ms NOESY spectrum of the dodecamer in 90% H_2O/10% D_2O at 5 °C. Sequential connectivities are indicated by the solid lines.

One of the pitfalls in using DNA fragments containing such nucleobases in biochemical reactions, particular those involving macromolecular recognition, is the possibility that the modifications may induce noticeable structural alterations of the DNA helix. Indeed, particular structural features may themselves play a significant role in recognition processes, and changes in these may thus complicate subsequent assessment of the importance of a specific functional group. This is particularly a concern when the modification results in the loss of one or more hydrogen bonds. Although in aqueous solutions base stacking is thought to be the major source of DNA duplex stability (Petruska et al., 1986), hydrogen bonds clearly play a pivotal role in stabilizing double-stranded DNA as indicated from the recent thermodynamic studies of Turner et al. (1987). Deletion of hydrogen bonds generally results in reduced helix stability and possible alternate hydrogen-bonding patterns.

The restriction endonuclease *Eco*RI recognizes the hexameric double-stranded sequence 5′d(GAATTC)$_2$ and catalyzes the hydrolysis of both phosphodiester bonds between the dG and dA residues. Substitution of the purine base (2′-deoxynebularine) for the "outer" adenine residue in the *Eco*RI recognition sequence results in the loss of an exocyclic amino group from the 6-position of the purine base and deletion of the corresponding hydrogen bond to the O4 of thymine located in the major groove of the DNA helix. This functional group has been implicated, on the basis of recent X-ray diffraction data, in the binding interaction between the endonuclease and the canonical 6 base pair recognition site (Frederick et al., 1984; McClarin et al., 1986). We have also observed that the deletion of this functional group drastically affects catalytic efficiency by the enzyme (McLaughlin et al., 1987).

In order to better assess the structural implications of the dP-dT versus dA-dT base pairs, we prepared the purine-substituted *Eco*RI site GPATTC flanked by CGC on the 5′-side and GCG and the 3′-side, resulting in the dodecamer 5′d-(CGCGPATTCGCG)$_2$. This sequence thus corresponds to

Table I: Proton Resonance Assignments of the DNA Dodecamer at 5 °C

	chemical shift (ppm)											
residue	H8/H6	H5/CH_3	H2	H6	H1/H3	NH_2[a]	H1′	H2′	H2″	H3′	H4′	H5′, H5″
C1	7.56	5.80					5.66	1.89	2.33	4.61	3.98	3.64, 3.64
G2	7.88				13.04		5.80	2.58	2.63	4.90	4.26	4.00
C3	7.20	5.30				8.28, 6.45	5.48	1.81	2.22	4.75	4.06	4.05, 3.92
G4	7.81				12.45		5.47	2.61	2.74	4.94	4.27	3.93
P5	8.45		7.87	8.23			6.04	2.62	2.87	5.00	4.41	4.17, 4.15
A6	8.09		7.51			7.11, 6.01	6.11	2.50	2.86	4.97	4.43	4.19
T7	7.09	1.31			13.43		5.92	1.94	2.53	4.78	4.14	4.12, 4.31
T8	7.27	1.58			12.98		6.07	2.14	2.48	4.83	4.21	4.13, 4.05
C9	7.38	5.53				8.19, 6.78	5.69	1.97	2.35	4.82	4.07	4.11
G10	7.85				12.81		5.77	2.57	2.63	4.90	4.30	4.12
C11	7.25	5.37				8.42, 6.62	5.63	1.83	2.23	4.75	4.12	4.04
G12	7.87				13.10		6.08	2.55	2.27	4.63	4.11	4.08

[a] In the case of the 4- and 6-amino groups of cytosine andd adenine, respectively, the chemical shift of the hydrogen-bonded proton is listed first.

Table II: $(\langle r^{-6}\rangle)^{-1/6}$ Mean Interproton Distances Derived from the NOESY Spectra at 5 °C[a]

(A) Intranucleotide

	r_{ij} (Å)											
proton	C1	G2	C3	G4	P5	A6	T7	T8	C9	G10	C11	G12
sugar–sugar												
H1′–H2′	3.0		2.8	2.6	2.6		2.7	2.7	2.9		2.7	
H1′–H2″	2.2		2.2	2.2	2.2	2.2	2.2		2.2		2.1	2.1
H2′–H3′	2.3		2.2					2.3	2.2		2.2	2.2
H2″–H3′			2.6				2.6	2.6	2.7		2.6	
sugar–base												
H1′–H6/H8	3.1	3.7	3.5		3.5	3.5	4.0	3.7	3.7	3.7	3.7	3.4
H2′–H6/H8	2.2	2.1	2.0	2.1	2.1	2.2	2.1	2.1	2.0	2.1	2.0	2.1

(B) Internucleotide (Intrastrand)

	r_{ij} (Å)										
$i-(i+1)$	C_1pG_2	G_2pC_3	C_3pG_4	G_4pP_5	P_5pA_6	A_6pT_7	T_7pT_8	T_8pC_9	C_9pG_{10}	$G_{10}pC_{11}$	$C_{11}pG_{12}$
H1′–H6/H8		3.3		3.4	3.4	3.6	3.7	3.5	3.6	3.4	3.5
H2′–H6/H8	2.8		2.9		2.8	2.7	2.7	2.7	2.8		2.8
H2″–H6/H8	2.4	2.3	2.8	2.6		2.3	2.3	2.3	2.4	2.3	2.4
H2′–H5/CH_3						2.7	2.4	3.4			
H2″–H5/CH_3						2.8	2.7	3.0			
H2′/H2″–H5[b]		2.7								2.7	
H1′–CH_3						4.2	5.1				
H3′–CH_3						4.5	3.8				
H6/H8–H5/CH_3		3.8				3.2	3.0	3.6		3.8	
H1′–H5″			4.0	3.5		3.3					
H2–H1′						4.0					
H2–H2					3.5						
H1–H2				4.1							
H3–H3							3.6				
H1–NH_2'		4.5									
H1–NH_2''		4.9									

(C) Interresidue (Interstrand)

proton pair	r_{ij} (Å)	proton pair	r_{ij} (Å)	proton pair	r_{ij} (Å)
		Intra Base Pair			
A6(H2)–T19(H3)	2.9	P5(H2)–T20(H3)	3.1	P5(H6)–T20(H3)	3.1
		Inter Base Pair			
G2(H1)–G22(H1)	4.0	P5(H2)–T19(H3)	4.4	A6(H2)–T20(H1′)	4.0
G22(H1)–G4(H1)	4.1	A6(H2)–T20(H3)	3.5	A6(NH_2'')–T20(H3)	5.0
G4(H1)–T20(H3)	4.2				

[a] The estimated errors in the distances are as follows: $-0.2/+0.3$ Å for $r_{ij} < 3$ Å and $-0.3/+0.4$ Å for 3 Å $\leq r_{ij} \leq 5$ Å (see text). [b] The H2′ and H2″ resonances of residues G2 and G10 are superimposed. Consequently, the G2(H2′)–C3(H5) and G2(H2″)–C3(H5) cross-peaks are superimposed and similarly for the G10(H2′)–G11(H5) and G10(H2″)–C11(H5) cross-peaks. The intensities of these cross-peaks correspond to distances of 2.7 Å. In the restraint list, we set this value equal to the $(\langle r^{-6}\rangle)^{-1/6}$ average of the two corresponding distances.

the well-studied dodecamer 5′d(CGCGAATTCGCG)$_2$ crystallized in the B-helical form (Wing et al., 1980; Dickerson & Drew, 1981) and subsequently partially analyzed by ^{1}H NMR (Sarma et al., 1982; Patel et al., 1983; Hare et al., 1983). We then proceeded to determine its solution conformation by a combination of NMR and restrained molecular dynamics. First, the exchangeable and nonexchangeable resonances were assigned in a sequential manner by means of NOESY and HOHAHA spectroscopy. A set of interproton distance restraints was then derived from the NOESY spectra at short mixing times and used as the basis of structural refinement by restrained molecular dynamics (Kaptein et al., 1985; Clore et al., 1985, 1986a,b; Brünger et al., 1986; Nilsson et al., 1986). As in previous studies (Nilsson et al., 1986; Nilges et al., 1987a,b; Scalfi Happ et al., 1988), convergence was achieved by starting from classical A- and B-DNA structures. The converged structures are of the B type, with an average atomic rms difference between them of 0.7 ± 0.2 Å as compared to an atomic rms difference of 6.5 Å between the two starting structures. Clear sequence-dependent vari-

Table III: Atomic rms Differences between Initial and Final Structures[a]

	atomic rms difference (Å)		atomic rms difference (Å)
initial structures		rms distributions	
IniA vs IniB	6.45	⟨RDA⟩ vs ⟨RDA⟩	0.70 ± 0.20
rms shifts		⟨RDB⟩ vs ⟨RDB⟩	0.53 ± 0.11
IniA vs ⟨RDA⟩	5.24 ± 0.22	⟨RDA⟩ vs ⟨RDB⟩	0.68 ± 0.17
IniB vs ⟨RDB⟩	2.21 ± 0.14	⟨RD⟩ vs ⟨RD⟩	0.65 ± 0.17
$\overline{RD}$ vs ($\overline{RD}$)m	0.18	⟨RD⟩ vs $\overline{RD}$	0.43 ± 0.13
		⟨RD⟩ vs ($\overline{RD}$)m	0.47 ± 0.11

[a] The notation of the structures is as follows: IniA and IniB are the initial structures with classical A- and B-DNA geometries, respectively; ⟨RDA⟩ are the four final restrained dynamics structures derived from IniA; ⟨RDB⟩ are the four final restrained dynamics structures derived from IniB; ⟨RD⟩ refers to all eight final restrained dynamics structures (i.e., to ⟨RDA⟩ and ⟨RDB⟩ collectively); $\overline{RD}$ is the mean structure obtained by averaging the coordinate of the eight ⟨RD⟩ structures best fitted to each other; and ($\overline{RD}$)m is the restrained energy minimized average structure obtained by restrained energy minimization of $\overline{RD}$. The atomic standard rms error in the coordinates of the average structure $\overline{RD}$ is given by rmsd/$\sqrt{8}$ ~ 0.15 Å, where rmsd is the average atomic rms difference between the eight ⟨RD⟩ structures and the mean structure $\overline{RD}$.

ations in the helical parameters are observed, with the variations of helix twist and propeller twist closely following those observed in the crystal structure of the B-DNA dodecamer (Dickerson & Drew, 1982) and predicted by Dickerson's (1983) sum functions. Although the overall dodecamer is straight, the central 6 base pairs are smoothly bent in the direction of the major groove with a radius of curvature of approximately 38 Å. This bend, however, is compensated by two smaller kinks at the CpG base pair steps 3 and 9.

Experimental Procedures

Synthesis of 5′-O-(9-Phenylxanthen-9-yl)-2′-deoxynebularine. To 0.7 g (2.96 mmol) of 2′-deoxynebularine (Nair & Chamberlain, 1984) suspended in 50 mL of dry pyridine was added dropwise 1.0 g (3.85 mmol) of 9-chloro-9-phenylxanthene dissolved in 10 mL of dry pyridine. After 3 h at ambient temperature, TLC analysis (CH_2Cl_2/CH_3OH in a ratio of 9:1) indicated that the reaction was complete. The reaction was stopped with 5 mL of methanol and partitioned between dichloromethane (75 mL) and 4% aqueous sodium bicarbonate (120 mL). The organic phase was dried (Mg_2SO_4) and the solvent removed by rotary evaporation. The residue was dissolved in toluene and evaporated to dryness 2 times in order to remove all traces of pyridine. The residue was chromatographed on 20 g of silica gel and eluted with dichloromethane/0.1% triethylamine and a gradient from 1% to 4% ethanol. The fractions containing the product were evaporated to dryness. The yield was 74% (1.11 g, 2.23 mmol). The ^{1}H NMR characteristics in DMSO-d_6 with a trace of D_2O were as follows: 2.3–3.1 (m, H2′, H2″, DMSO), 3.5 (m, H5′, H5″, HOD), 4.0 (m, 1 H, H4′), 4.5 (m, 1 H, H3′), 6.4 (t, 1 H, H1′, J = 6.4 Hz), 6.9–7.5 (m, 13 H, Ar H), 8.6 (s, 1 H, H8), 8.8 (s, 1 H, H2), 9.1 ppm (s, 1 H, H6).

Synthesis of 3′-O-[(N,N-Diisopropylamino)(2-cyanoethoxy)phosphinyl]-5′-O-(9-phenylxanthen-9-yl)-2′-deoxynebularine. To 1.1 g of the protected nebularine derivative

Table IV: rms Interproton Distance Deviations, Deviations from Ideality, and Restraint and Nonbonding Energies for the Initial and Final Structures[a]

	rms interproton distance deviations[b] (Å)			
structure	all (258)	intraresidue (102)	interresidue (126)	base pairing[c] (30)
IniA	0.85 (114)	0.86 (26)	0.93 (88)	0.16 (0)
IniB	0.45 (60)	0.21 (4)	0.62 (58)	0.15 (0)
⟨RDA⟩	0.17 ± 0.001 (0)	0.15 ± 0.001 (0)	0.20 ± 0.002 (0)	0.068 ± 0.003 (0)
⟨RDB⟩	0.17 ± 0.002 (0)	0.15 ± 0.001 (0)	0.20 ± 0.001 (0)	0.063 ± 0.001 (0)
$\overline{RD}$	0.19 (4)	0.15 (0)	0.24 (4)	0.069 (0)
($\overline{RD}$)m	0.17 (0)	0.15 (0)	0.20 (0)	0.052 (0)

	deviations from ideality		
structure	bond (Å) (812)	angles (deg) (1470)	impropers[d] (deg) (342)
IniA	0.008	3.17	0.32
IniB	0.011	3.08	0.32
⟨RDA⟩	0.009 ± 0	3.32 ± 0.03	0.33 ± 0.003
⟨RDB⟩	0.008 ± 0.001	3.35 ± 0.03	0.33 ± 0.013
$\overline{RD}$	0.113	13.33	12.57
($\overline{RD}$)m	0.009	5.56	0.30

	restraint energy (kcal/mol)		nonbonding energy (kcal/mol)		
structure	E_{NOE}[e] (258)	E_δ[f] (24)	van der Waals	electrostatic	H-bond
IniA	4212	50	−64	−347	−65
IniB	1106	0	−1157	−385	−50
⟨RDA⟩	129 ± 1	0.20 ± 0.17	−376 ± 2	−478 ± 4	−90 ± 1
⟨RDB⟩	132 ± 3	0.20 ± 0.03	−374 ± 2	−479 ± 3	−92 ± 1
$\overline{RD}$	151	0.12	−371	−469	−55
($\overline{RD}$)m	132	0.14	−375	−482	−81

[a] The number of distance restraints, bonds, angles, and impropers and the number of terms used in the NOE and δ torsion angle restraint potentials are given in parentheses below the respective headings. Note that the total number of interproton distance restraints used in the calculations is double that in Table II as the oligonucleotide is self-complementary and hence symmetric. [b] The number of interproton distance violations greater than 0.5 Å is given in parentheses next to the rms interproton distance deviations. [c] In addition to the experimental interproton distance restraints, a set of 30 base pairing restraints corresponding to the base pair hydrogen bonds was added to the NOE restraint energy function. These are as follows: for A-T base pairs $r_{A(N)-T(O4)}$ = 2.95 Å and $r_{A(N1)-T(H3)}$ = 2.82 Å; for P-T base pairs $r_{P(N1)-C(N3)}$ = 2.82 Å; for G–C base pairs $r_{G(O6)-C(N4)}$ = 2.91 Å, $r_{G(N1)-C(N3)}$ = 2.95 Å, and $r_{G(N2)-C(O2)}$ = 2.86 Å. The values were taken from the X-ray structure analyses of ApU (Seeman et al., 1976) and GpC (Rosenberg et al., 1976), and the error estimates for these values used in the calculations were ±0.2 Å, with the exception of the one for the P-T base pairs, which was set to ±0.4 Å. [d] The improper torsion angle restraints are the restraints used to maintain planarity. [e] The scale factor S (cf. eq 2) used in calculating the NOE restraints energy was 8. This corresponds to force constants of 59.6, 26.5, and 14.9 kcal mol^{-1} Å^{-2} for distance errors of 0.2, 0.3, and 0.4 Å, respectively. [f] The force constant for the δ torsion angle restraint was 40 kcal mol^{-1} rad^{-2}.

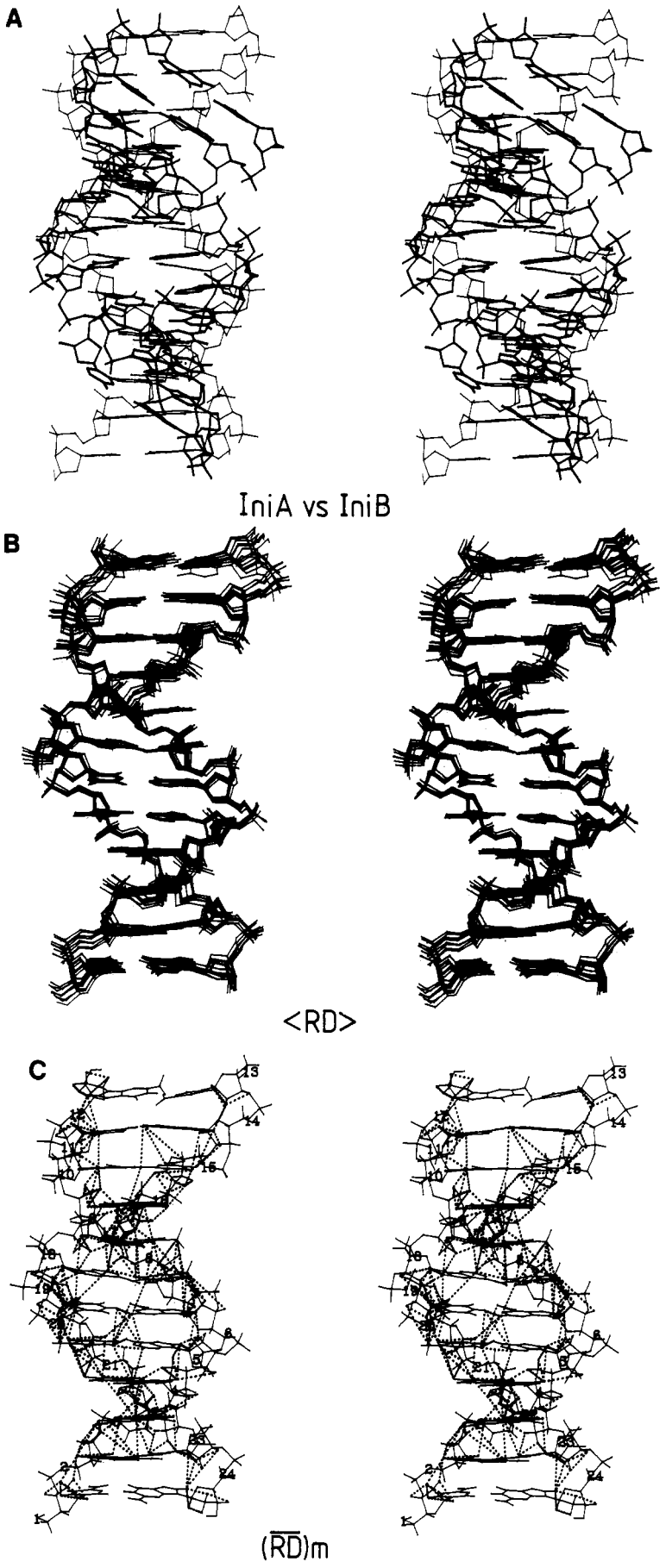

FIGURE 3: (A) Superposition of the two initial structures, IniA (thick lines) and IniB (thin lines), (B) superposition of the eight converged structures (⟨RD⟩), and (C) superposition of the interproton distance restraints (---) on the framework of the restrained energy minimized structure (RD)m.

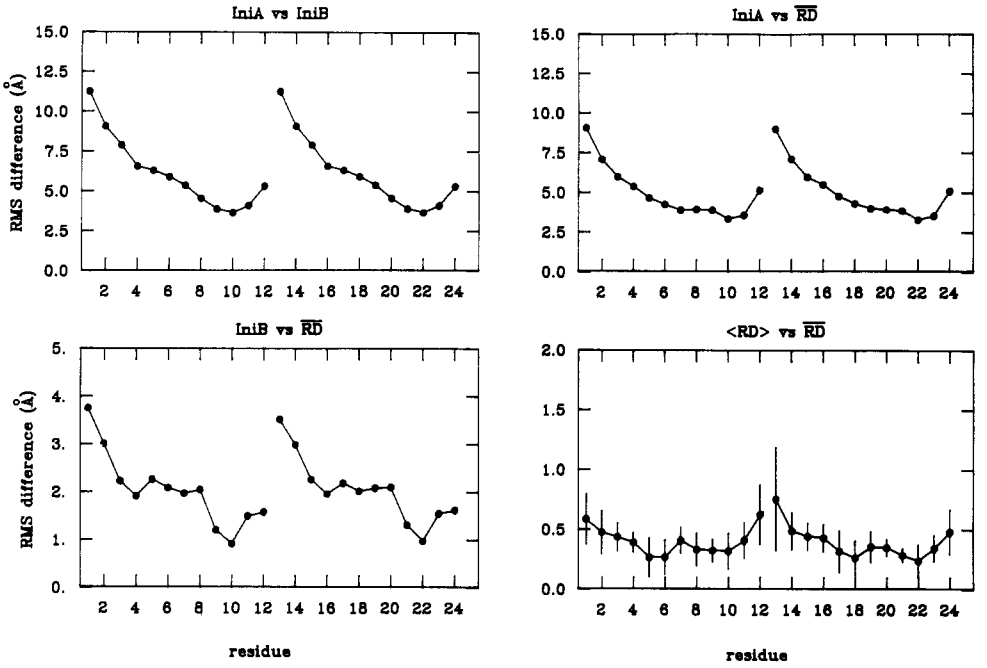

FIGURE 4: rms differences (Å) for all atoms between various structures. In the plot of ⟨RD⟩ vs $\overline{RD}$ the solid circles represent the average atomic rms difference and the bars the standard deviations in these values.

prepared above in 20 mL of dry dichloromethane were added 1.0 g (3.3 mmol) of bis(*N*,*N*-diisopropylamino)(2-cyanoethoxy)phosphine [prepared essentially as described by Barone et al. (1984)] and 0.28 g of diisopropyltetrazolide (Barone et al., 1984). After the resultant mixture was stirred for 3 h at ambient temperature, TLC analysis (ethyl acetate/petroleum ether/triethylamine, 49.5:49.5:1) indicated complete conversion of the starting material to two new products (diastereoisomers). The reaction was stopped with 3 mL of methanol and the mixture evaporated to dryness. The residue was dissolved in a small amount of dichloromethane/triethylamine (99:1), and a small quantity of activated charcoal was added. The mixture was introduced onto a column of silica gel (20 g) packed in petroleum ether/triethylamine (99:1). The products were eluted with dichloromethane/petroleum ether/triethylamine (49.5:49.5:1). Fractions containing the product were evaporated to dryness, dissolved in 6 mL of toluene, and precipitated by dropwise addition to 150 mL of petroleum ether stirred at −70 °C. The precipitate was filtered and dried under vacuum. The yield was 81% (1.23 g, 1.77 mmol). The ^{31}P NMR spectrum in $CDCl_3$ exhibited a signal at 146.94 ppm.

Oligonucleotide Synthesis. The oligodeoxynucleotide was synthesized on a controlled-pore glass support containing 10 μmol of starting nucleosides with an Applied Biosystem 380A synthesizer. The synthesis employed a glass column and a modified 10-μmol synthesis program, which will be described elsewhere. After deprotection (with concentrated ammonia for 6 h at 50 °C), the oligonucleotide containing the 5′-terminal protecting group was isolated by reversed-phase chromatography (McLaughlin & Piel, 1984). With the removal of the 5′-terminal protecting group, 360 A_{260} units (~18 mg) of purified oligodeoxynucleotide was obtained. To exchange the triethylammonium counterions for Na^+ ions, the dodecamer was then bound to DEAE-Sephadex A-25 and eluted with 1.0 M NaCl, desalted by solid-phase extraction using a Baker 10SPE C_{18} column and 60% aqueous ethanol as eluant, and finally passed over DNA-grade Sephadex G-25. Fractions containing pure product (334 A_{260} units) were pooled and lyophilized.

Sample Preparation. After extensive freeze-drying, 167 A_{260} units (~8 mg) of dodecamer was dissolved in either 99.995% D_2O or 90% H_2O/10% D_2O, containing 10 mM potassium phosphate, 100 mM KCl, and 0.01 mM EDTA, pH 6.5, to give a final concentration of ~4 mM in single strand. NMR experiments were carried out at 5 and 25 °C.

NMR Spectroscopy. All NMR spectra were recorded on a Bruker AM500 spectrometer. Two-dimensional NOESY (Jeener et al., 1979; Macura et al., 1982), ROESY (Bothner-By et al., 1984; Bax & Davis, 1985a), MLEV17 HOHAHA (Davis & Bax, 1985; Bax & Davis, 1985b), and E-COSY (Griesinger et al., 1982) spectra were recorded in the pure-phase absorption mode by the time-proportional incrementation method (Redfield & Kuntz, 1975; Bodenhausen et al., 1980; Marion & Wüthrich, 1983). In the case of the NOESY spectra recorded in H_2O, the last 90° pulse in the sequence was replaced by the $90°_x$–τ–$90°_{-x}$ jump return sequence (Plateau & Gueron, 1982) with the carrier placed at the position of the water resonance and a value of 50 μs for the delay τ. NOESY spectra in D_2O were recorded with mixing times of 40 and 70 ms at 5 °C and with mixing times of 75 and 150 ms at 25 °C. The NOESY spectrum in H_2O was recorded with a mixing time of 150 ms. Spectra in D_2O were recorded with a sweep width of 6042 Hz, an acquisition time of 0.17 s, and an incremental time step of 83 μs for t_1, while the corresponding values for the spectra in H_2O were 10000 Hz, 0.102 s, and 50 μs, respectively. Typically, 512–600 t_1 increments were recorded for each spectrum. To reduce t_1 noise, the first time domain data points were multiplied by a factor of 0.5 (Otting et al., 1986). Linear base line correction was carried out after the first and second transforms. In the case of the NOESY spectra that were used for quantification of cross-peak intensities, a relaxation delay between scans of 3–6 s was used and appropriate zero-filling was carried out in both dimensions to give a digital resolution of 1.47 and 2.44 Hz/point for the D_2O and H_2O spectra, respectively. Quantification of the NOESY cross-peak intensities was carried out on an ASPECT 1000 data station by determining the volume of each cross-peak by two-dimensional integration.

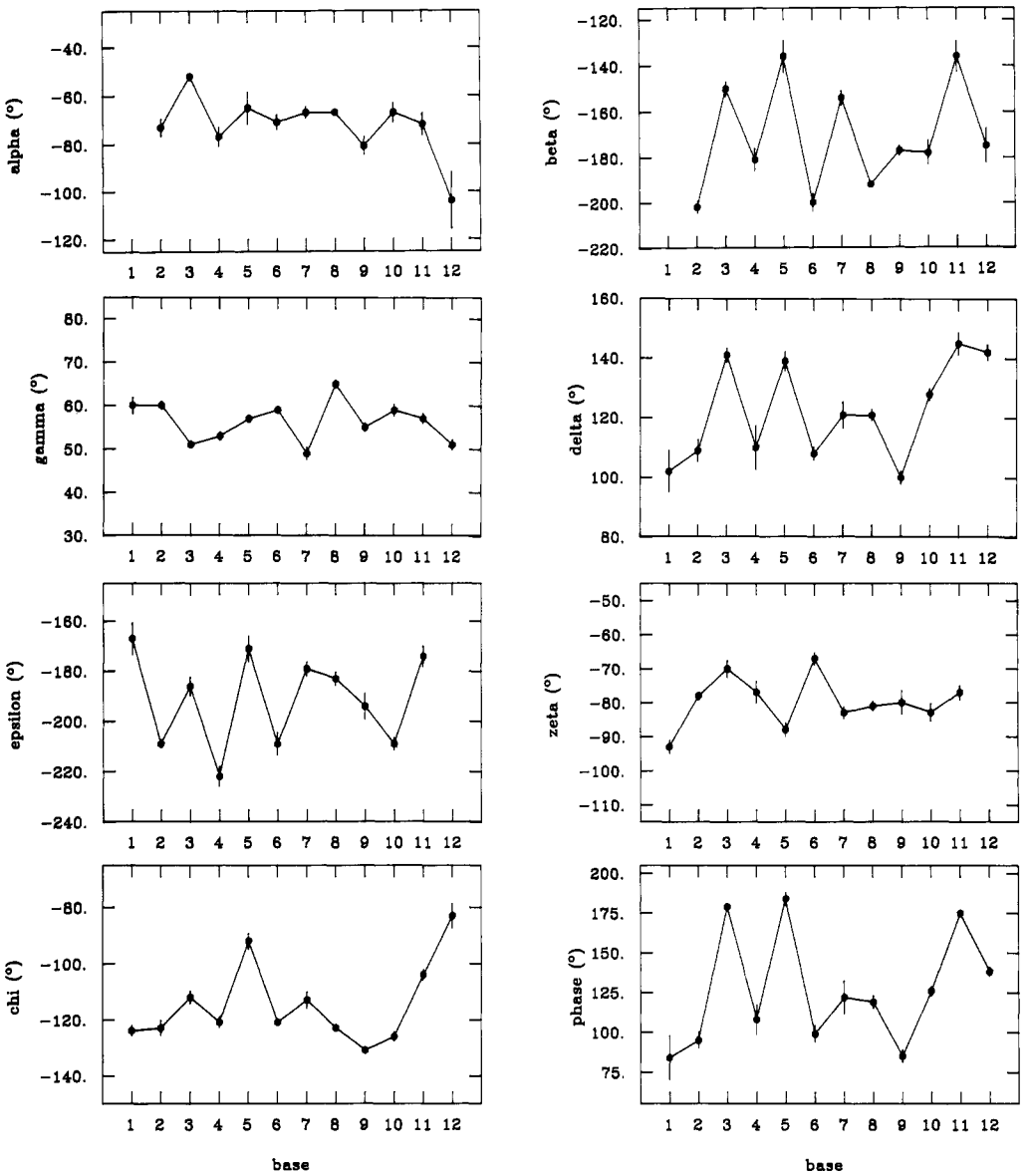

FIGURE 5: Variations in backbone and glycosidic bond torsion angles as well as in the phase angle describing the sugar pucker for the eight ⟨RD⟩ converged structures. The solid circles represent the averages of the values for the eight ⟨RD⟩ structures and the bars the rms deviations in these values. The sugar pucker phase angle is calculated as described by Cremer and Pople (1975) with the apex at atom 3 and C4′ = atom 0, C1′ = atom 1, and so on.

Restrained Molecular Dynamics. All energy minimization and molecular dynamics calculations were carried out on a CONVEX C1-XP computer with the program XPLOR (Brünger et al., 1986, 1987a,b; Clore et al., 1986a; A. T. Brünger, unpublished data), which is derived from the program CHARMM (Brooks et al., 1983) and has been specially adapted for restrained molecular dynamics. The energy function used comprises an all-hydrogen empirical energy function ($E_{\text{empirical}}$) developed for nucleic acids (Nilsson & Karplus, 1985) and effective interproton distance (E_{NOE}) and torsion angle (E_ϕ) restraint energy functions. The effective NOE restraint potential, E_{NOE}, has the form of a skewed biharmonic potential (Clore et al., 1985) given by

$$\begin{aligned} E_{\text{NOE}}(\mathbf{r}_{ij}) &= c_1(\mathbf{r}_{ij} - \mathbf{r}_{ij}^\circ)^2 \quad \text{if } \mathbf{r}_{ij} > \mathbf{r}_{ij}^\circ \\ &= c_2(\mathbf{r}_{ij} - \mathbf{r}_{ij}^\circ)^2 \quad \text{if } \mathbf{r}_{ij} < \mathbf{r}_{ij}^\circ \end{aligned} \tag{1}$$

where $\mathbf{r}_{ij}$ and $\mathbf{r}_{ij}^\circ$ are the calculated and experimental distances, respectively, and c_1 and c_2 are force constants given by

$$c_1 = k_B TS/2(\Delta_{ij}^+)^2 \qquad c_2 = k_B TS/2(\Delta_{ij}^-)^2 \tag{2}$$

where k_B is the Boltzmann constant, T is the absolute temperature, S is a scale factor, and Δ_{ij}^+ and Δ_{ij}^- are the positive and negative error estimates on the value of $\mathbf{r}_{ij}^\circ$. The effective torsion angle restraint potential has the form of a square well given by (Clore et al., 1986b)

$$\begin{aligned} E_\phi &= c(\phi_{ij} - \phi_{ij}^{\text{u}})^2 \quad \text{if } \phi_{ij} > \phi_{ij}^{\text{u}} \\ &= c(\phi_{ij} - \phi_{ij}^{\text{l}})^2 \quad \text{if } \phi_{ij} < \phi_{ij}^{\text{l}} \end{aligned} \tag{3}$$

where c is a force constant, ϕ_{ij} is the calculated value of the torsion angle, and ϕ_{ij}^{u} and ϕ_{ij}^{l} are the upper and lower experimental limits of the torsion angle. With respect to the electrostatic component of the empirical energy function, the effect of solvent was approximated by a $1/r$ screening function (Gelin & Karplus, 1977; Brooks et al., 1983) and by reducing the net charge on the phosphate group to −0.32e (Tidor et al., 1982). The nonbonded interactions were switched off, by using a cubic switching function, between 9.5 and 10.5 Å, with pairs up to 11.5 Å included in the nonbonded list. Integration of the classical equations of motion was performed with a Verlet integration algorithm (Verlet, 1967) with initial velocities

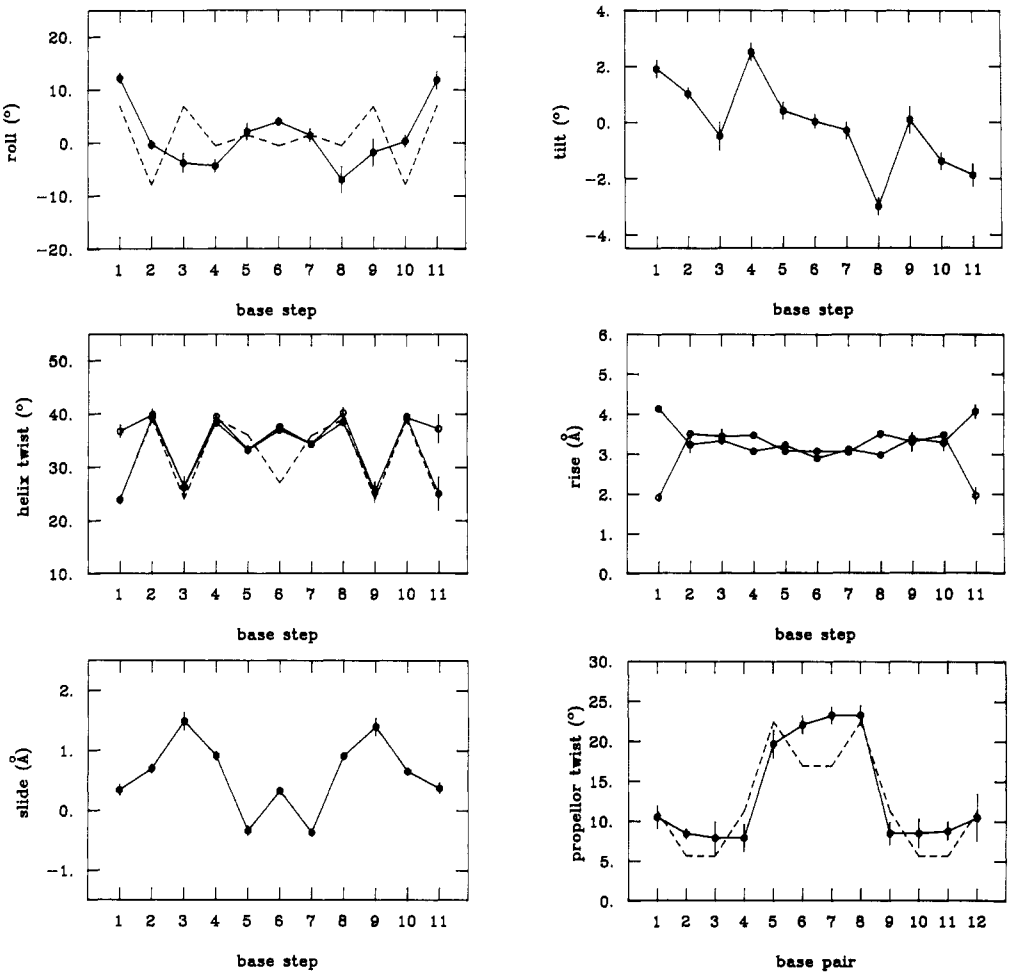

FIGURE 6: Variations in helical parameters for the eight ⟨RD⟩ converged structures. The solid and open circles represent the averages of the values for the eight ⟨RD⟩ structures and the bars the rms deviations in these values. In the case of helix twist and rise, the solid circles (●) represent the global helix twist and rise and the open circles (○) the local helix twist and rise. The exact definitions of the various helical parameters are given by Dickerson (1983). The variation in helix twist, roll, and propeller twist predicted by Dickerson's sum function $\sum_1$, $\sum_2$, and $\sum_4$, respectively, is shown as dashed lines (- - -).

assigned from a Maxwellian distribution at 400 K. The temperature of the system was maintained constant by rescaling the velocities of the atoms every 0.1 ps. The time step of the integrator was 0.001 ps, and the nonbonded interaction lists were updated every 0.02 ps. Bond lengths involving hydrogen atoms were kept fixed with the SHAKE algorithm (Ryckaert et al., 1977).

Structural Analysis. Displaying of trajectories was carried out on an Evans & Sutherland PS390 color graphics system using a modified version of the function network of FRODO (Jones, 1978) interfaced with XPLOR. Analysis of helical parameters was carried out with the program HETRAN (M. Nilges and G. M. Clore, unpublished data), which is a modified version of the programs AHELIX (written by J. Rosenberg), BROLL, and CYLIN (written by R. E. Dickerson) adapted to deal with dynamics trajectories (Nilges et al., 1987a).

Results and Discussion

Sequential Resonance Assignment. The assignment of the nonexchangeable protons was accomplished in a sequential manner (Reid et al., 1983; Scheek et al., 1983; Hare et al., 1983; Feigon et al., 1983; Weiss et al., 1984a; Clore & Gronenborn, 1983, 1985a) by using (a) Hartmann–Hahn spectroscopy to demonstrate direct and relayed through-bond connectivities along the H1′ ↔ H2′ ↔ H3′ ↔ H4′ ↔ H5′/H5″ pathway within each sugar unit and (b) NOESY spectroscopy to demonstrate through-space (<5 Å) connectivities along the H1′/H2′(i − 1) ↔ H8/H6(i) ↔ H1′/H2′(i) and the H8/H6(i)–H5/CH$_3$(i + 1) pathways. Some examples of NOESY spectra in D_2O are shown in Figure 1. Assignment of the exchangeable imino and amino protons, together with the A(H2), P(H2), and P(H6) resonances, was also carried out sequentially (Chou et al., 1983; Ulrich et al., 1983; Clore & Gronenborn, 1983; Gronenborn et al., 1984a; Weiss et al., 1984b; Boelens et al., 1985) as illustrated by the NOESY spectrum in H_2O shown in Figure 2. Note that the presence of an imino proton for the T-P base pair, together with the observation of approximately equal NOEs between the T8(H3) proton on the one hand and the P(H2) and P(H6) protons on the other, indicates unambiguously that the P-T base pair is hydrogen bonded in a Watson–Crick manner. The complete list of assignments is given in Table I.

Interproton Distance Restraints. Interproton distance restraints were obtained from cross-peak intensities in the NOESY spectra using the H2′–H2″ (1.8 Å), C(H6)–C(H5) (2.5 Å), and T(H6)–T(CH$_3$) (~2.7 Å) distances as internal references (Gronenborn et al., 1984b) from the equation (Wagner & Wüthrich, 1979; Dobson et al., 1982; Clore & Gronenborn, 1985b)

$$(\langle r_{ij}^{-6} \rangle)^{-1/6} = [a_{ref}(\tau_m)/a_{ij}(\tau_m)]^{-1/6} r_{ref} \qquad (4)$$

where r_{ij} and $a_{ij}(\tau_m)$ are the distance and NOE cross-peak intensity at a mixing time τ_m, respectively, between protons i and j and r_{ref} and $a_{ref}(\tau_m)$ have the corresponding meanings

Table V: Average Values of the Glycosyl and Backbone Torsion Angles, Helix Twist, Helix Rise, and Propeller Twist for the Initial Structures, the Final Converged Restrained Dynamics Structures, and the Crystal Structure of the B-DNA Dodecamer 5'd(CGCGAATTCGCG)$_2$

structure	glycosyl χ (deg)	backbone torsion angles (deg)						helix twist (deg)		local helix rise (Å)	propeller twist (deg)
		α	β	γ	δ	ϵ	ζ	global	local		
IniA	−154	−86	−151	47	83	178	−46	32.7	32.7	2.56	11.7
IniB	−98	−47	−146	36	156	155	−95	36.0	36.0	3.33	4.2
B-DNA$_C$[a]	−117 ± 14	−63 ± 8	171 ± 14	54 ± 8	123 ± 21	−169 ± 25	−108 ± 34	35.8 ± 4.1	37.3 ± 3.8	3.33 ± 0.13	13.4 ± 4.9
⟨RD⟩	−114 ± 15	−72 ± 13	180 ± 43	56 ± 5	122 ± 17	169 ± 19	−80 ± 7	34.6 ± 5.5[b]	35.0 ± 5.7[b]	3.33 ± 0.20[b]	13.0 ± 6.4

[a] From Dickerson and Drew (1981). [b] Excludes the terminal base pair steps 1 and 11.

for the reference vector. The validity of eq 4 rests on the assumption that the effective correlation times of the i–j and reference vectors are about the same and that the initial rate condition is approximately valid.

For distances involving only nonexchangeable protons we used the 40-ms NOESY spectrum at 5 °C and the 75-ms NOESY spectrum at 25 °C, while for distances involving exchangeable protons we used the 150-ms NOESY spectrum recorded in H_2O at 5 °C. The validity of the initial rate approximation under these conditions was verified by selective one-dimensional experiments using the NOESY pulse sequence with the first nonselective 90° pulse replaced by a selective 90° Gaussian-shaped pulse (Kessler et al., 1986). In this respect it should be noted that the reason that a much longer mixing time could be used for the distances involving the exchangeable protons as well as the A(H2), P(H2), and P(H6) protons is simply that they are relatively isolated in terms of nearest proton neighbors (with the obvious exception of the two amino protons of a given amino group).

Because the H2′ and H2″ protons are so close together (1.8 Å), it is often the case that a small amount of spin diffusion resulting in the transmission of a direct NOE involving one of the two methylene protons to the other is invariably observed even at the shortest mixing times. An example of this is the small indirect intraresidue NOE between the base H8/H6 and H2″ protons (see Figure 1B). As spin-diffusion effects are generally not observed in ROESY spectra (Bax et al., 1986), a 100-ms mixing time ROESY spectrum was used as a check to ensure that all the cross-peaks integrated in the NOESY spectra and used for the subsequent refinement calculations arose from first-order direct NOEs.

When eq 4 is used, the effects of variations in effective correlation times also have to be considered. No base-to-base variation in effective correlation times for the fixed distance intraresidue H2′–H2″, C(H5)–C(H6), and T(H6)–T(CH$_3$) vectors could be detected. The apparent effective correlation time of the intraresidue H2′–H2″ sugar vectors, however, was slightly shorter than that of the C(H6)–C(H5) and T(H6)–T(CH$_3$) vectors, in agreement with previous findings on other DNA oligonucleotides (Clore & Gronenborn, 1984; Nilges et al., 1987a,b). For this reason, we used the H2′–H2″ vector as a reference for all sugar–sugar and sugar–base distances (with the exception of the H1′–base distances) and the C(H6)–C(H5) vector as a reference for all base–base and sugar H1′–base distances. The justification for this approach which is easily verified empirically by calculating distances, which, although not fixed, have a very limited range (±0.2 Å), has been discussed in detail previously (Gronenborn et al., 1984; Gronenborn & Clore, 1985; Nilges et al., 1987a).

A summary of the calculated interproton distances is given in Table II. Taking into account both the considerations discussed above as well as the errors involved in determining cross-peak intensities by volume integration, we estimate that the errors are −0.2 Å/+0.3 Å for $r_{ij} < 3$ Å and −0.3 Å/+0.4 Å for 3 Å $\leq r_{ij} < 5$ Å. Within the experimental errors, no differences could be detected in the distances measured at 5 and 25 °C. For consistency, only the 5 °C data were used in the subsequent calculations, as the spectra in H_2O were only recorded at 5 °C.

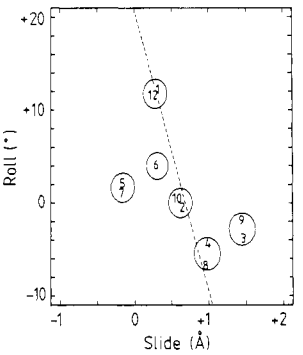

FIGURE 7: Roll–slide diagram for the restrained energy minimized average structure ($\overline{RD}$)m. The dashed line from roll, slide = −10°, 1 Å to +20°, 0.2 Å represents the break between A- and B-type DNA geometries, which lie to the right and left, respectively, of the line (Calladine & Drew, 1984).

C4′–C3′ Bond Torsion Angle Restraints. Information on the C4′–C3′ (δ) bond torsion angle was deduced from $^3J_{1'2'}$ coupling constants measured from an E-COSY spectrum. These could be estimated to be larger than 7 Hz for all sugar units, indicative of a value of $\delta > 110°$ (Altona & Sundaralingam, 1972; Davies, 1985). Thus, in the calculations we restrained the δ torsion angles in the range 110–170°. Because we were unable to measure the $^3J_{1'2'}$ coupling constants with a precision of better than 1 Hz and because evaluation of the other sugar coupling constants was precluded due to resonance overlap, we did not feel justified in imposing any more precise restraints on the δ bond torsion angles.

Structure Refinement. In order to obtain the structure of the dodecamer in solution, we proceeded to carry out restrained molecular dynamics calculations incorporating the experimental interproton distances and δ torsion angle data into the total energy of the system in the form of effective potentials (cf. eq 1 and 3). Two initial structures were used, namely, classical A- and B-DNA (Arnott & Hukins, 1972). The atomic rms difference between the two initial structures is 6.5 Å. The protocol of restrained dynamics used was the same as that previously used on an RNA hexamer (Scalfi Happ et al., 1988) and involved three stages: (i) 8 ps of quenched restrained dynamics at 400 K in which the velocities were rescaled to 400 K every 0.1 ps and the NOE restraint scale factor S (cf. eq 1 and 2) was increased from 0.32 to a maximum value of 8.0 and the δ torsion angle restraint force constant from 0.63 kcal mol^{-1} rad^{-2} to a maximum value of 40 kcal mol^{-1} rad^{-2} by multiplying their respective values by $10^{0.2}$ every 0.1 ps (the values of the NOE and δ restraint force constants reached at the end of this stage were maintained for the rest of the calculation); (ii) 12 ps of quenched restrained dynamics at 300 K in which the velocities were rescaled every

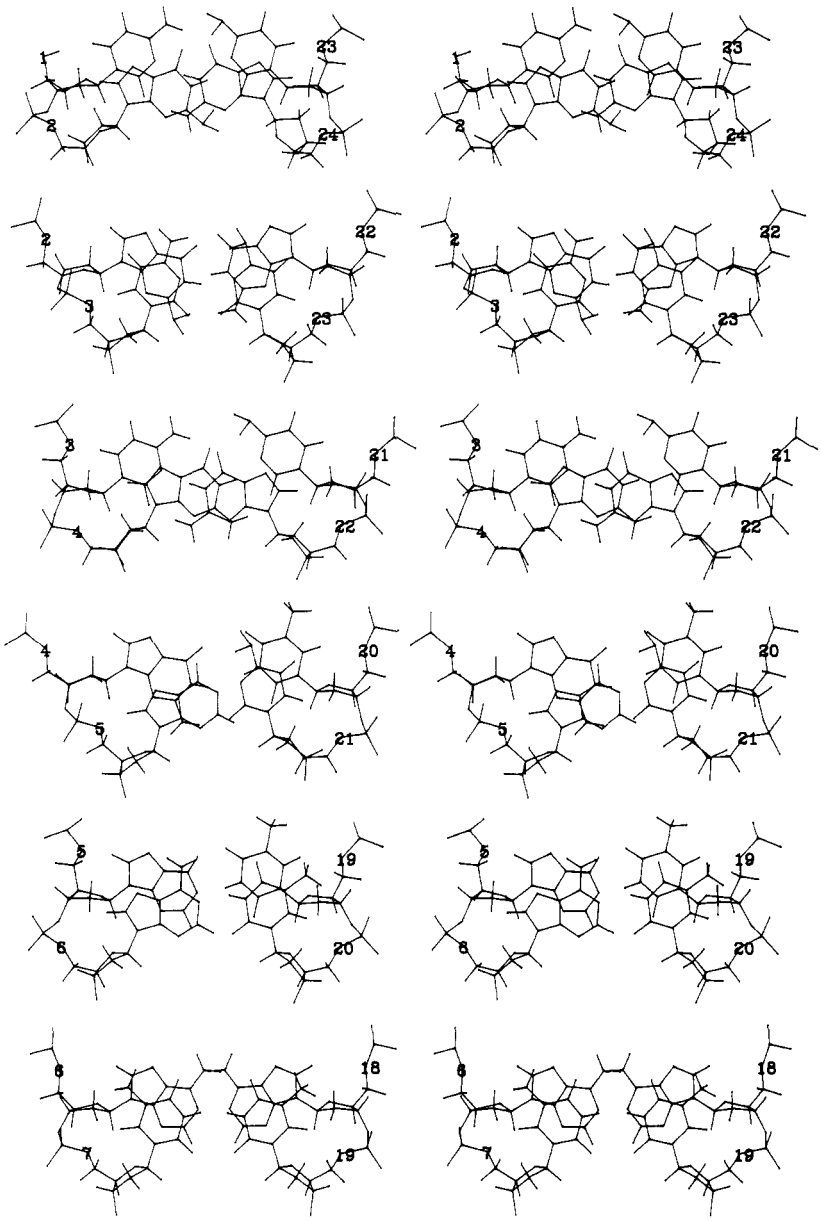

FIGURE 8: Stereoviews of base pair steps 1–6 of the restrained energy minimized average structure (RD)m viewed down the global helix axis. (Note that because of symmetry base pair steps 7–11 are virtually identical with base pair steps 1–5 and hence are not shown.)

0.1 ps; and (iii) 400 cycles of restrained energy minimization of the coordinates obtained by averaging the coordinate trajectories over the last 8 ps of the second stage. Four calculations were carried out from each initial structure by using different random number seeds for the assignments of the initial velocities. Each calculation took approximately 10 h on the Convex C1XP computer. The final four structures obtained by starting from IniA are referred to as ⟨RDA⟩ and the final four starting from IniB as ⟨RDB⟩; ⟨RDA⟩ and ⟨RDB⟩ are also referred to collectively as ⟨RD⟩. The coordinates of the eight final structures were also averaged to yield the average structure $\overline{RD}$, which was subjected to 400 cycles of restrained energy minimization to produce the restrained energy minimized average structure $(\overline{RD})$m.

The atomic rms differences between the structures are given in Table III, and the rms differences between the calculated and experimental interproton distances, the deviations from ideality for bonds, angles, and planes, the NOE and δ torsion angle restraint energies, and the nonbonding energies are given in Table IV. Superpositions of the two initial structures and of the final structures are shown in Figure 3 and plots of atomic rms difference between various structures as a function of residue number in Figure 4. Also shown in Figure 3 is a superposition of the interproton distances on the restrained energy minimized average structure $(\overline{RD})$m.

It is clear from the data in Tables II and III and Figures 3 and 4 that convergence to essentially the same structures, both globally and locally, has been achieved by starting from both initial structures. Further, the atomic rms differences between the final structures are independent of the starting structures. Thus, the difference between the final structures arises from the different random number seeds used to assign the initial velocities. The average atomic rms difference between the final structures is ~0.7 Å, and that between the final structures and the mean structure $\overline{RD}$ is ~0.4 Å, which is comparable to the atomic rms fluctuations of the atoms about their average positions. The rms difference between the calculated and target interproton distances (~0.2 Å) is within the experimental errors specified, and the δ torsion angles lie within their target range. In addition, the extent of conver-

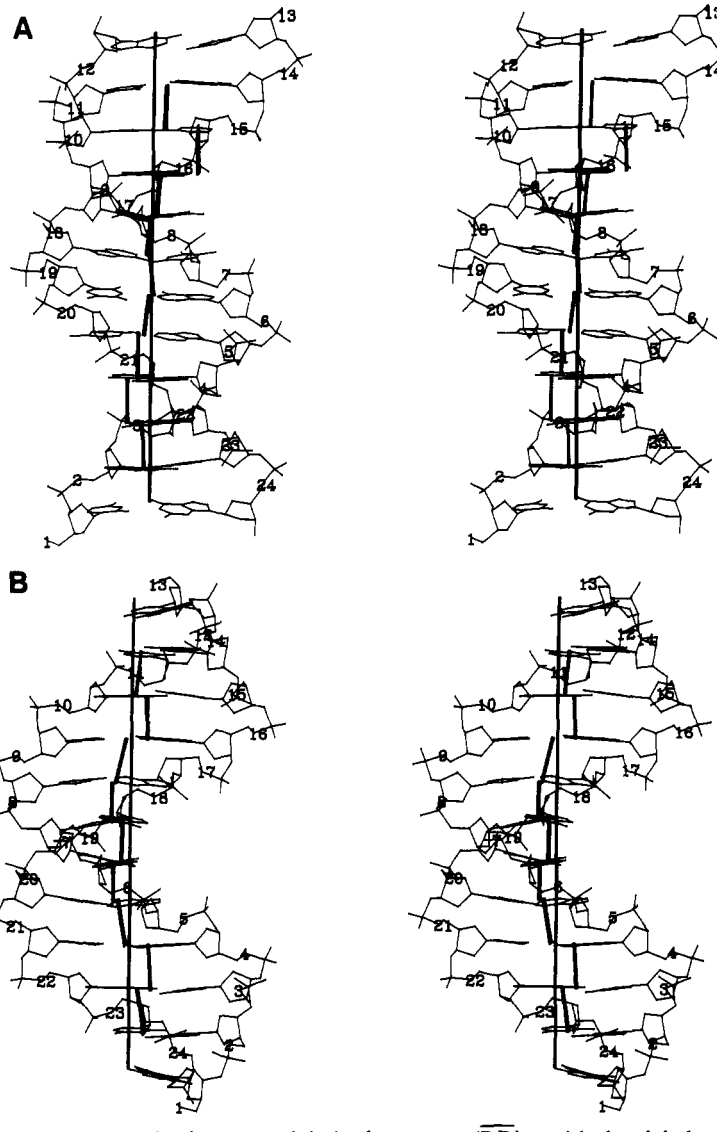

FIGURE 9: Two stereoviews of the average restrained energy minimized structure $(\overline{RD})m$ with the global and local helix axes superimposed.

gence can be assessed from the plots of backbone torsion angles (Figure 5) and helical parameters (Figure 6) as a function of residue number.

It should be noted that in the absence of experimental restraints (i.e., using the same dynamics protocol but with the force constants for the NOE and δ torsion angle restraints set to zero) convergence from the two starting structures does *not* occur. The structure starting from IniA remains A type and that from IniB B type, with an atomic rms difference between them of 6.3 Å. Thus, as in the previous cases (Nilsson et al., 1986; Nilges et al., 1987a,b; Scalfi Happ et al., 1988), convergence is entirely due to the incorporation of the experimental restraints into the total energy function of the system in the form of effective potentials, and the structural features that emerge are not in any way artifacts arising from the empirical energy function. The role of the latter is solely to ensure that the local stereochemistry and nonbonded interactions are approximately correct.

Structural Features. The overall structure of the dodecamer is of the B type (Figure 3), and the substitution of a purine for an adenine at position 5 clearly does not lead to any major perturbation in the structure of the dodecamer.

Examination of the backbone torsion angles indicates that they all lie in the range characteristic of B-DNA. Indeed, the average values are very close to those found in the crystal structure of the B-DNA dodecamer 5′d-$(CGCGAATTCGCG)_2$ (Dickerson & Drew, 1981) (Table V). As expected, the C4′–C3′ bond torsion angles δ are correlated with the sugar pucker phase angles and anticorrelated with the glycosidic bond torsion angles χ (Figure 5).

Considering the helical parameters, there are clear sequence-dependent variations in base pair roll, slide, tilt, helix twist, and propeller twist (Figure 6). The variations of helical twist and propeller twist are similar to those found in the crystal structure of the B-DNA dodecamer and follow the trends predicted by Dickerson's (1983) sum functions (Figure 6), suggesting that these variations are driven by the need to relieve interstrand steric hindrance between purines at Pyr-Pur and Pur-Pyr steps (Calladine, 1982). The variation in roll, on the other hand, is not so well correlated with either the predicted or observed crystal structure variation, the principal deviation occurring at base pair steps 3 and 9.

Roll and slide are correlated (Figure 7). Base pair steps 1/12, 2/10, 4/8, and 6 lie along the line roll, slide = −10°, 1 Å to +20°, 0.2 Å that separates A- and B-type geometries (Calladine & Drew, 1984). Base pair steps 3/9 and 5/7, on the other hand, lie clearly in the A- and B-type regions, respectively. These features are manifested in the pattern of base pair stacking (Figure 8). The stacking of the CpG base pair steps 3 and 9 are of the A-type with interstrand stacking of

the six-membered ring of the purine residues of opposite strands. Associated with this is a decrease in the helical twist (Figure 6). The PpA base pair steps 5 and 7, however, are of the B type with good intrastrand overlap of the six-membered purine rings. The base pair stacking of base pair steps 2 and 10 is intermediate between A and B type, whereas all the other base pair steps are stacked in a B-type manner. Note, though, that the intrastrand stacking of the purine residues for the GpA steps 4 and 8 is not as optimal as that of the PpA steps 5 and 7, although both involve Pur-Pur steps.

These local variations in slide, roll, and base pair stacking patterns give rise to kinking and bending of the DNA. This is illustrated in Figure 9, which shows a superposition of the global and local helix axes on the restrained energy minimized average structure $(\overline{RD})$m. Overall, the dodecamer is clearly straight. The central 5 base pair steps, however, are smoothly bent in the direction of the major groove with a radius of curvature of ~38 Å and an angle between the local helix axes of base pair steps 4 and 8 of 25°. This is compensated by two smaller kinks in the direction of the minor groove at the CpG steps 3 and 9 This can be understood in terms of the roll–slide diagram and can be attributed to the relative changes in roll and slide of base pair steps 3/9 and 5/7 with respect to those of base pair steps 4/8: in particular, the decrease in roll and increase in slide of base pair steps 3 and 9, on the one hand, and the increase in roll and decrease in slide of base pair steps 5 and 7, on the other, relative to those of base pair steps 4 and 8.

Concluding Remarks. This dodecamer represents the third DNA oligonucleotide whose structure we have solved by a combination of NMR and restrained molecular dynamics and analyzed in detail, the previous two being the hexamer 5′d-$(GCATGC)_2$ (Nilges et al., 1987a) and the decamer 5′d-$(CTGGATCCAG)_2$ (Nilges et al., 1987b). It is therefore of interest to see whether any common sequence-dependent features emerge from these studies. The most striking feature common to all three oligonucleotides is the behavior of the nonterminal CpG steps. They all have values of slide and roll in the A-type region of the slide–roll diagram, A-type interstrand stacking of the purines, and a decrease in helical twist. These features result in a kinking of the DNA at the CpG steps. It must be emphasized, however, that the glycosidic bond and C4′–C3′ bond torsion angles of these residues all lie in the range characteristic of B-DNA. The second most obvious common feature is the variation in helical twist, which appears to follow reasonably well, although by no means perfectly, the trend predicted by Dickerson's (1983) sum functions.

Other features are more difficult to predict as they depend on the actual sequence of bases rather than the sequence written purely in terms of purines and pyrimidines. Thus, for example, the central 6 base pairs of the decamer (GGATCC) and the dodecamer (GPATTC) have the same sequence of purine and pyrimidines: Pur-Pur-Pur-Pyr-Pyr-Pyr. They differ, however, by the substitution of a G for a P at position 2 and of a C for a T at position 5. The sequence dependence of the helical parameters is similar but not identical, and these differences are sufficient to ensure that the central 6 base pairs of the decamer are straight, while those of the dodecamer are smoothly bent.

Registry No. d(CGCGPATTCGCG), 114155-95-2; purine, 120-73-0; thymine, 65-71-4; 5′-*O*-(9-phenylxanthen-9-yl)-2′-deoxynebularine, 114155-96-3; 2′-deoxynebularine, 4546-68-3; 9-chloro-9-phenylxanthene, 42506-03-6; 3′-*O*-[(*N*,*N*-diisopropylamino)(2-cyanoethoxy)phosphinyl]-5′-*O*-(9-phenylxanthen-9-yl)-2′-deoxynebularine, 114155-97-4; bis[*N*,*N*-diisopropylamino)(2-cyanoethoxy)phosphine, 102691-36-1.

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.

Barone, A. D., Tang, J.-Y., & Caruthers, M. H. (1984) *Nucleic Acids Res. 12*, 4051–4061.

Bax, A., & Davis, D. G. (1985a) *J. Magn. Reson. 65*, 355–360.

Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 65*, 207–213.

Bax, A., Sklenar, V., & Summers, M. F. (1986) *J. Magn. Reson. 70*, 327–331.

Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.

Boelens, R., Scheek, R. M., Dijkstra, K., & Kaptein, R. (1985) *J. Magn. Reson. 62*, 378–386.

Bothner-By, A. A., Stephens, R. L., Lee, J. T., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811–813.

Brennan, C. A., Van Cleve, M. D., & Gumport, R. I. (1986a) *J. Biol. Chem. 261*, 7270–7278.

Brennan, C. A., Van Cleve, M. D., & Gumport, R. I. (1986b) *J. Biol. Chem. 261*, 7278–7286.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 74*, 187–217.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987a) *Science (Washington, D.C.) 235*, 458–460.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1987b) *Protein Eng. 1*, 399–406.

Calladine, C. R., & Drew, H. R. (1984) *J. Mol. Biol. 178*, 773–782.

Caruthers, M. H. (1980) *Acc. Chem. Res. 13*, 155–160.

Chou, S. H., Hare, D. R., Wemmer, D. E., & Reid, B. R. (1983) *Biochemistry 22*, 3037–3041.

Clore, G. M., & Gronenborn, A. M. (1983) *EMBO J. 2*, 2109–2115.

Clore, G. M., & Gronenborn, A. M. (1984) *FEBS Lett. 172*, 219–225.

Clore, G. M., & Gronenborn, A. M. (1985a) *FEBS Lett. 179*, 187–198.

Clore, G. M., & Gronenborn, A. M. (1985b) *J. Magn. Reson. 61*, 158–164.

Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.

Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986a) *J. Mol. Biol. 191*, 523–551.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986b) *EMBO J. 5*, 2729–2735.

Cremer, D., & Pople, J. A. (1975) *J. Am. Chem. Soc. 97*, 1358–1367.

Davies, D. B. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 12*, 135–225.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2821–2822.

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

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

Dobson, C. M., Olejniczak, E. T., Poulsen, F. M., & Ratcliffe, R. G. (1982) *J. Magn. Reson. 48*, 87–110.

Dubendorff, J. W., DeHaseth, P. L., Rosendahl, M. S., & Caruthers, M. H. (1987) *J. Biol. Chem. 262*, 892–898.

Dwyer-Hallquist, P., Kezdy, F. J., & Agarwal, K. L. (1982) *Biochemistry 21*, 4693–4700.

Eritja, R., Horowitz, D. M., Walker, P. A., Ziehler-Martin, J. P., Boosalis, M. S., Goodman, M. F., Itakura, K., & Kaplan, G. E. (1986) *Nucleic Acids Res. 14*, 8135–8153.
Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983) *Biochemistry 22*, 5930–5942.
Fisher, E. F., & Caruthers, M. H. (1979) *Nucleic Acids Res. 7*, 401–416.
Fliess, A., Wolfes, H., Rosenthal, A., Schwellnus, K., Blocker, H., Frank, R., & Pingoud, A. (1986) *Nucleic Acids Res. 14*, 3463–3474.
Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L. J., Wang, B.-C., Greene, P., Boyer, H. W., & Rosenberg, J. M. (1984) *Nature (London) 309*, 327–331.
Gelin, G. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002–2006.
Goeddel, D. V., Yansura, D. G., & Caruthers, M. H. (1977) *Nucleic Acids Res. 4*, 3039–3054.
Goeddel, D. V., Yansura, D. G., Winston, C., & Caruthers, M. H. (1978) *J. Mol. Biol. 123*, 661–687.
Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6802.
Gronenborn, A. M., & Clore, G. M. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 17*, 1–33.
Gronenborn, A. M., Clore, G. M., Jones, M. B., & Jiricny, J. (1984a) *FEBS Lett. 165*, 216–222.
Gronenborn, A. M., Clore, G. M., & Kimber, B. J. (1984b) *Biochem. J. 221*, 723–736.
Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.
Ikuta, S., Eritja, R., Kaplan, B. E., & Itakura, K. (1987) *Biochemistry 26*, 5646–5650.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jiricny, J., Wood S. G., Martin, D., & Ubasawa, A. (1986) *Nucleic Acids Res. 14*, 6579–6590.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.
Kessler, H., Oschkinat, H., Griesinger, C., & Bermerl, W. (1986) *J. Magn. Reson. 70*, 106–133.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
McLaren, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science (Washington, D.C.) 234*, 1526–1541.
McLaughlin, L. W., & Piel, N. (1984) in *Oligonucleotide Synthesis: A Practical Approach* (Gait, M. J., Ed.) pp 199–218, IRL, Oxford.
McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N., & Scholtissek, S. (1987) *Biochemistry 26*, 7238–7245.
Nair, V., & Chamberlain, S. (1984) *Synthesis* 401–403.
Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson, L. (1987a) *Biochemistry 26*, 3718–3733.
Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987b) *Biochemistry 26*, 3734–3744.
Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 691–716.
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455–475.
Ono, A., Masayuki, S., Oktani, Y., & Ueda, T. (1984) *Nucleic Acids Res. 12*, 8939–8949.
Otting, G., Widmer, W., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Patel, D. J., Kozowski, S. A., & Bhatt, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3908–3912.
Petruska, J., Sowers, L. C., & Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1559–1562.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.
Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., & Williams, D. H. (1983) *Biochemistry 22*, 2019–2025.
Rosenberg, J. M., Seeman, N. C., Day, R. O., & Rich, A. (1976) *J. Mol. Biol. 104*, 145–167.
Rychaert, J. P., Cicotto, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–337.
Saenger, W. (1984) *Principles of Nucleic Acid Structures*, Springer-Verlag, New York.
Sarma, R. H., Warner, B. G., & Mitra, C. K. (1982) in *Biomolecular Stereodynamics* (Sarma, R. H., Ed.) Vol. 1, pp 89–98, Adenine, New York.
Scalfi Happ, C., Happ, E., Nilges, M., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 1735–1743.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.
Seela, F., & Driller, H. (1986) *Nucleic Acids Res. 14*, 2319–2332.
Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P., & Rich, A. (1976) *J. Mol. Biol. 104*, 109–144.
Tidor, B., Irikura, K., Brooks, B. R., & Karplus, M. (1983) *J. Biomol. Struct. Dyn. 1*, 231–252.
Turner, D. H., Sugimoto, N., Kierzek, R., & Dreiker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3738–3785.
Ulrich, E. L., Jogn, E. M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., & Markley, J. L. (1983) *Biochemistry 22*, 4362–4365.
Verlet, L. (1967) *Phys. Rev. 159*, 98–105.
Wagner, G., & Wüthrich, K. (1979) *J. Magn. Reson. 33*, 675–680.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984a) *Proc. Natl. Acad. Sci. U.S.A. 81*, 130–134.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984b) *Nucleic Acids Res. 12*, 4035–4047.
Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tamaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.
Yansura, D. G., Goeddel, D. V., Cribbs, D. L., & Caruthers, M. H. (1977) *Nucleic Acids Res. 4*, 723–737.
Yansura, D. G., Goeddel, D. V., Kundu, A., & Caruthers, M. H. (1979) *J. Mol. Biol. 133*, 117–135.
Yolov, A. A., Vinogradova, M. N., Gromova, E. S., Rosenthal, A., Cech, D., Veiko, V. P., Metelev, V. G., Kosykyh, V. G., Buryanov, Y. I., Bauev, A. A., & Shaborova, Z. A. (1985) *Nucleic Acids Res. 13*, 8983–8998.