Hashimoto, Y., Iijima, H., Nozaki, Y., & Shudo, K. (1986) *Biochemistry 25*, 5103–5110.
Hecht, S. M. (1986) *Acc. Chem. Res. 19*, 383–391.
Hertzberg, R. P., & Dervan, P. B. (1984) *Biochemistry 23*, 3934–3945.
Kessel, D. (1984) *Biochem. Pharmacol. 33*, 1389–1393.
Le Doan, T., Perrouault, L., Chassignol, M., Thuong, N. T., & Hélène, C. (1987) *Nucleic Acids Res. 15*, 8643–8659.
Lown, J. W., Sondhi, S. M., Ong, C. W., Skorobogaty, A., Kishihawa, H., & Dabrowiak, J. C. (1986) *Biochemistry 25*, 5111–5117.
McMurry, T. J., & Groves, J. T. (1986) in *Cytochrome P-450, Structure, Mechanism and Biochemistry* (Ortiz de Montellano, P. R., Ed.) Chapter I, pp 1–28, Plenum Press, New York.
Megnin, F., Faustino, P. J., Lyon, R. C., Lelkes, P. I., & Cohen, J. S. (1987) *Biochim. Biophys. Acta 929*, 173–181, and references cited therein.
Meunier, B. (1986) *Bull. Soc. Chim. Fr.*, 578–594.
Moan, J. (1986) *Photochem. Photobiol. 43*, 681–690.
Pasternack, R. F., Huber, P. R., Boyd, P., Engasser, G., Francesconi, L., Gibbs, E., Fasella, P., Cerio Venturo, G., & de C. Hinds, L. (1972) *J. Am. Chem. Soc. 94*, 4511–4517.
Pasternack, R. F., Gibbs, E. J., & Villafranca, J. J. (1983a) *Biochemistry 22*, 5409–5417.
Pasternack, R. F., Gibbs, E. J., & Villafranca, J. J. (1983b) *Biochemistry 22*, 2406–2414.
Pope, L. E., & Sigman, D. S. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3–7.
Praseuth, D., Gaudemer, A., Verlhac, J. B., Kraljic, I., Sissoëff, I., & Guillé, E. (1986) *Photochem. Photobiol. 44*, 717–724.
Pratviel, G., Bernadou, J., & Meunier, B. (1986) *Biochem. Biophys. Res. Commun. 136*, 1013–1020.
Pratviel, G., Bernadou, J., & Meunier, B. (1989) *Biochem. Pharmacol. 38*, 133–140.
Robert, A., & Meunier, B. (1988) *New J. Chem. 12*, 885–896.
Stubbe, J., & Kozarich, J. W. (1987) *Chem. Rev. 87*, 1107–1136.
Tabushi, I. (1988) *Coord. Chem. Rev. 86*, 1–42.
Tullius, T. D., Dombroski, B. A., Churchill, M. E. A., & Kam, L. (1987) *Methods Enzymol. 155*, 537–558.
Van Touw, J. H., Verberne, J. B., Retel, J., & Loman, H. (1985) *Int. J. Radiat. Biol. 48*, 567–578.
Ward, B., Skorobogaty, A., & Dabrowiak, J. C. (1986a) *Biochemistry 25*, 7827–7833.
Ward, B., Skorobogaty, A., & Dabrowiak, J. C. (1986b) *Biochemistry 25*, 6875–6883.

Solution Conformation of d-CTCGAGCTCGAG by Two-Dimensional NMR: Conformational Heterogeneity at *Xho*I Cleavage Site

Anu Sheth, R. V. Hosur,* and Girjesh Govil
Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India
M. V. Hosur and K. K. Kannan
Neutron Physics Division, Bhabha Atomic Research Centre, Bombay 400 085, India
Zu-kun Tan and H. Todd Miles
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892
Received January 31, 1989; Revised Manuscript Received May 12, 1989

ABSTRACT: Nonexchangeable proton resonances in the 500-MHz NMR spectrum of d-CTCGAGCTCGAG have been assigned by using two-dimensional correlated spectroscopy (COSY) and nuclear Overhauser enhancement spectroscopy (NOESY). ^1H–^1H coupling constants (J) in the deoxyribose rings have been measured by analyzing intensity and multiplet patterns in the phase-sensitive ω_1-scaled COSY spectra. A modification of the J-resolved technique, called amplitude-modulated J-resolved spectroscopy, has been described and used to increase the accuracy of J measurements. Absorption mode ω_1-scaled NOESY spectra at mixing times in the range 50–200 ms have been analyzed to monitor spin diffusion. A 50-ms spectrum has been used to estimate several interproton distances. The coupling constant and distance data have been used to arrive at sequence-specific sugar geometries and glycosidic torsion angles. The backbone structure has been refined by model building using the FRODO program, employing the sugar geometries and glycosidic torsion angles discussed above. The molecule shows interesting sequence-dependent variations in the structure. The cleavage site of the restriction enzyme *Xho*I exhibits unique differences in the sugar geometry and backbone torsion angles.

With recent advances in methods for synthesis of DNA segments and the physical tools for structure determination, the understanding of DNA structure has undergone radical changes. It is no longer believed that DNA is a regular double-helical structure. New evidence suggests that DNA exhibits wide structural diversity depending upon the base sequence and experimental conditions. It is envisaged that the information content in DNA with respect to its function may be coded and modulated not only by the base sequence but also by the associated structural diversities.

Recent 2D NMR studies have focused on DNA segments recognized by proteins in a highly specific manner (Ravikumar

0006-2960/89/0428-7275$01.50/0

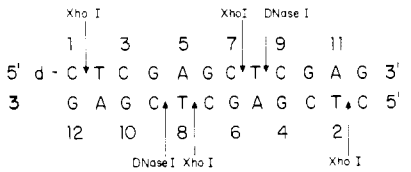

FIGURE 1: Preferential cleavage sites for the enzymes *Xho*I and DNase I in the dodecanucleotide d-CTCGAGCTCGAG.

et al., 1985; Nilges et al., 1987; Sheth et al., 1987a; Chary et al., 1987, 1988; Hosur et al., 1985a, 1986a,b; 1988a,b; R. V. Hosur, K. V. R. Chary, A. Saran, G. Govil, and H. T. Miles, unpublished results). We have investigated solution structures of I, d-GGATCCGGATCC (Hosur et al., 1985a; Ravikumar et al., 1985), II, d-GAATTCGAATTC (Chary et al., 1987; Hosur et al., 1986a), III, d-GAATTCCCGAATTC (Hosur et al., 1986b), and IV, d-GGTACGCGTACC (Chary et al., 1988). These sequences are more than 10 units long, and in each case, it is observed that the molecule adopts an overall structure belonging approximately to the B-DNA family. The sugar geometries show wide variations ranging from C2′-endo to O4′-endo in contrast to the C2′-endo geometry implicated in regular B structure. This is in contrast to a regular structure observed in d-CGCGCGCGCGCG (Sheth et al., 1987b) under low-salt conditions. In addition, important structural variations at the cleavage sites of restriction enzymes *Bam*H1 (molecule I) and *Eco*RI (molecules II and III) have been observed. On the other hand, the cleavage sites of *Rsa*I and *Fnu*dII (molecule IV) do not show such variations. It is interesting to note that the latter two enzymes recognize 4-unit-long sequences while the former two endonucleases recognize 6-unit-long sequences of DNA. It is important to study a large number of such recognition sequences and subsequently their interactions with proteins so that general conclusions about the mechanism of recognition and cleavage by such enzymes can be derived.

In this paper we report the solution structure of d-CTCGAGCTCGAG (Figure 1) (V), which is recognized by the restriction enzyme *Xho*I, using 2D NMR methods. A preliminary paper on this molecule describing the assignments and qualitative structural features (Sheth et al., 1987a) has been published earlier.

Materials and Methods

Synthesis. d-CTCGAGCTCGAG was prepared by a manual solid-phase (aminomethylpolystyrene with 1% divinylbenzene) phosphotriester method as described by Tan et al. (1982). The 3′-linked dG resin was prepared and coupled to the dA monomer. The protected dimers dCG, dCT, dAG, dCG, and dCT were coupled in turn to the 5′ end of the growing chain. After detritylation, cleavage from the resin, and deblocking, the oligomer was purified on DEAE-cellulose (7 M urea, 0.15–0.4 M NaCl gradient) [cf. Tan et al. (1982)]. A center cut of the main peak was homogeneous on gel electrophoresis. The sequence was confirmed by the method of Maxam and Gilbert (1980).

Sample Preparation. For the ^{1}H NMR experiments on d-CTCGAGCTCGAG, 6 mg of the sample was dissolved in 400 μL of 50 mM sodium phosphate buffer (pH 7.3) and 1 mM ethylenediaminetetraacetic acid (EDTA). The solution was repeatedly lyophilized from ^2H_2O solution. Finally the sample was made up to 0.4 mL with 99.9% ^2H_2O for the NMR experiments.

Nuclear Magnetic Resonance Spectroscopy. All proton NMR spectra were recorded at 25 °C. The melting temperature of the molecule has been determined to be 60 °C in

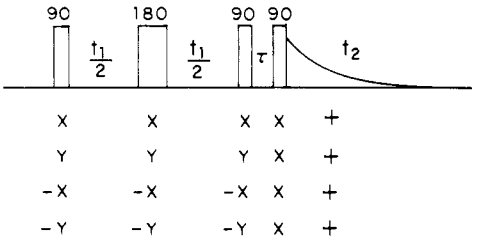

FIGURE 2: Experimental scheme of amplitude-modulated 2D *J*-resolved spectroscopy. τ is a small fixed period for phase switching. The phase cycling of pulses is indicated by x, y, etc. The plus sign implies that the data are added in the memory. t_1 and t_2 are the usual evolution and detection periods.

0.1 M Na^+ from UV measurements. The melting curves indicate that temperature of our measurements is in the range where the molecule is in an ordered conformation. The chemical shifts (δ) were measured with respect to sodium 3-(trimethylsilyl)[2,2,3,3-^{2}H]propionate (TSP).

The spectra were recorded on a Bruker AM 500 FT-NMR spectrometer equipped with an ASPECT 3000 computer. COSY data were recorded by using a (RD–90–t_1–Δ–90–Δ–t_2)$_n$ pulse sequence, where Δ is the fixed delay (5 ms) used for the enhancements of cross peaks (Bax & Freeman, 1981; Anil Kumar et al., 1984; Wynants & Van Binst, 1984). The carrier was placed at one end of the spectrum. The time domain data set consisted of 2048 and 512 data points along the t_2 and t_1 axes, respectively. For each t_1 value, 64 transients were collected. The data set was zero filled to 1024 along the t_1 axis and was multiplied by unshifted sine square bell and unshifted sine bell window functions along the t_2 and t_1 axes, respectively, prior to Fourier transformation. The digital resolution along both axes was 7.5 Hz/point. Phase-sensitive ω_1-scaled COSY (Hosur et al., 1985c, 1988a) spectra have been recorded by using the pulse sequence 90–$(\gamma + \alpha)t_1/2$–180–$(\gamma - \alpha)t_1/2$–ϕ–t_2, where α and γ are the shift and *J*-scaling factors, respectively. The scaling factors selected were $\alpha = 0.5$ and $\gamma = 1.3$. The flip angle (ϕ) of the mixing pulse was $\pi/4$. The time domain data consisted of 2048 t_2 and 350 t_1 points, and the number of transients for each t_1 value was 96. The data set was zero filled to 4096 along t_2 and 2048 along t_1 prior to two-dimensional Fourier transformation. The digital resolution along both axes was 1.8 Hz/point.

ω_1-scaled absorption mode NOESY (Jeener et al., 1979; Anil Kumar et al., 1980; Macura et al., 1980) spectra were recorded with scaling factors $\alpha = 0.5$ and $\gamma = 0.6$ and with mixing times of 50, 70, 130, 170, and 200 ms. The time domain data consisted of 2048 t_2 and 256 t_1 points. These data were zero filled only along t_1 to 1024 points and multiplied by sine bell functions shifted by $\pi/16$ along both axes before Fourier transformation. The final digital resolution in these spectra was 3.5 Hz/point along both axes.

An amplitude-modulated *J*-resolved spectrum of the oligonucleotide was recorded by using the pulse sequence shown in Figure 2. The phase of the 90° pulses is cycled so as to retain *z* magnetization at the end of the second 90° pulse. The phase of the 180° pulse was alternated between 0° and 180° so as to minimize effects of imperfections in the 180° pulse. The *z* magnetization is converted into observable magnetization by a third 90° pulse. In this experiment the t_1 domain has only *J* information, and the observed signal in t_2 is amplitude modulated by evolution in t_1. Consequently, the peaks have pure phases and good line shapes. Tilting of the data in frequency domain is not necessary. The method, however, does not remove *J* information along the ω_2 axis, and hence the resolution along ω_2 is essentially the same as in the 1D spec-

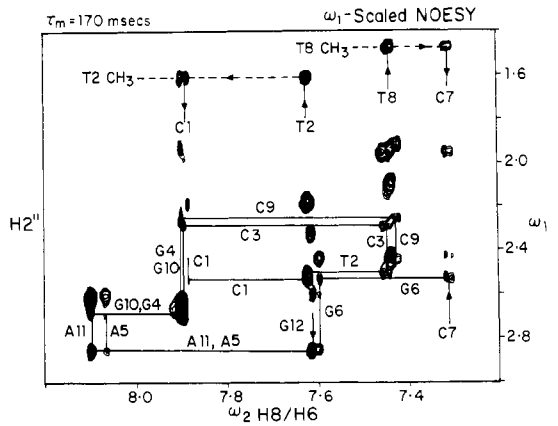

FIGURE 3: Portion of the ω_1-scaled absorption mode NOESY spectrum of d-CTCGAGCTCGAG, recorded with a mixing time of 170 ms. Sequential d3 connectivities via base (H8/H6) and H2″ protons are indicated by solid lines. The H2″ protons are identified on horizontal lines, and the base protons are identified near the vertical lines. Two d1 connectivities involving TCH_3 and CH6 are also observed in this region and are indicated by dashed lines.

trum. The method can be used to obtain pure absorptive line shapes as opposed to the phase-twisted line shapes prevalent in normal *J*-resolved spectroscopy.

Results and Discussion

Resonance Assignments. Sequence-specific resonance assignments have been obtained by following well-established procedures (Clore & Gronenborn, 1985; Hare et al., 1983, Hosur et al., 1985a; Reid et al., 1983; Scheek et al., 1983, 1984). In the first step, a COSY spectrum is used to identify the sugar protons H1′, H2′, H2″, H3′, and H4′ belonging to the individual networks of coupled spins. The cytidine and thymidine base protons are identified by using *J*(H5, H6) and *J*(CH_3–H6) correlations. In the next step, the intranucleotide and internucleotide NOEs between the base protons (H6, H8, and CH_3) and the sugar protons (H1′, H2′, and H2″) in the NOESY spectrum are used to assign the spin networks to individual nucleotide units. Four different types of sequential correlations have been observed:

d1 $(\text{base})_i \rightarrow (\text{base})_{i\pm1}$

d2 $(\text{base})_i \rightarrow (\text{H1}')_{i-1}$

d3 $(\text{base})_i \rightarrow (\text{H2}', \text{H2}'')_{i-1}$

d4 $(\text{base})_i \rightarrow (\text{H3}')_{i-1}$

As an illustration, Figure 3 shows the d3 connectivities in a NOESY spectrum with τ_m of 170 ms, involving the H2″ protons of the $(i-1)$th residues and the (H8/H6) protons of the ith residues. Two d1 connectivities involving thymine methyl protons are also seen in this portion of the spectrum and are indicated by dashed lines. The figure also contains sequential connectivities via H2′ protons; however, they are not drawn for the sake of clarity.

The sequence-specific assignments thus obtained were confirmed from the H1′–H4′ cross peaks in the NOESY spectrum (not shown). Stereospecific assignment of H2′ and H2″ protons of the individual nucleotide units has been obtained by monitoring the intensities of $(\text{H1}')_i$–$(\text{H2}',\text{H2}'')_i$ cross peaks in a ω_1-scaled NOESY spectrum recorded with τ_m = 70 ms (not shown). H1′–H2″ cross peaks are invariably stronger than H1′–H2′ cross peaks as discussed earlier (Chary et al., 1987, 1988; Hosur et al., 1986a). The complete resonance assignments are listed in Table I.

Table I: Chemical Shifts of Nonexchangeable Protons in d-CTCGAGCTCGAG (ppm from TSP)

base	H6/H8	H1′	H2′	H2″	H3′	H4′	H5/H2/CH_3
C1	7.87	5.85	2.24	2.59	4.66	4.07	5.92
T2	7.62	6.15	2.23	2.55	4.89	4.24	1.66
C3	7.44	5.55	2.02	2.34	4.85	4.07	5.62
G4	7.88	5.38	2.65	2.72	4.97	4.28	
A5	8.06	6.05	2.64	2.91	5.09	4.40	7.66
G6	7.58	6.65	2.47	2.59	4.93	4.32	
C7	7.30	5.81	2.00	2.46	4.63	4.17	5.14
T8	7.43	6.05	2.15	2.50	4.85	4.18	1.52
C9	7.41	5.49	1.97	2.31	4.83	4.05	5.64
G10	7.89	5.39	2.65	2.72	4.97	4.28	
A11	8.10	6.10	2.64	2.91	5.02	4.40	7.80
G12	7.60	5.96	2.38	2.25	4.60	4.14	

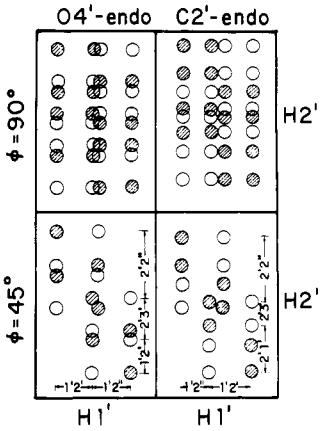

FIGURE 4: Calculated multiplet patterns of H1′–H2′ cross peak in COSY and COSY 45 spectra for the sugar geometries O4′-endo and C2′-endo. The peak separations useful for estimation of coupling constants *J*(H1′–H2′), *J*(H1′–H2″), *J*(H2′–H2″), and *J*(H2′–H3′) are indicated in COSY 45; solid and open circles indicate positive and negative peaks, respectively. For C3′-endo geometry, the active coupling *J*(H1′–H2′) = 0 Hz, and hence this cross peak will not be observed.

3D Structure of d-CTCGAGCTCGAG. Interstrand NOEs such as A5H2 → C9H5, A5H2 → T8H1′, A5H2 → C9H1′, and A11H2 → C3H1′ emerging from the halves of the oligonucleotide have been observed. These confirm that the molecule is in a double-stranded form. The directionality of the NOEs and the sequential NOEs in Figure 3 indicate that the molecule adopts a right-handed conformation. A more precise description of the structure has been obtained from the proton–proton coupling constants (*J*) in the deoxyribose rings and the proton–proton distances (*r*) in the molecule.

^1H–^1H Coupling Constants (J). Phase-sensitive ω_1-scaled COSY spectra have proved very useful for obtaining *J* values (Hosur et al., 1988b, and references cited therein). Recently a version of PE COSY (Mueller, 1987) has also been used to estimate *J* values in oligonucleotides (Bax & Lerner, 1988). An invaluable feature of such techniques is the presence of positive and negative extrema corresponding to the "active" coupling constant which aids the unambiguous assignment of *J* values. The use of different flip angles can further help in resolution enhancement. The general appearance of such spectra depends on the relative magnitudes of the couplings involved, which in turn depend on the sugar puckers. Figures 4 and 5 show the calculated patterns of H1′–H2′ and H1′–H2″ cross peaks (in the upper triangle of the spectrum) for different sugar geometries with two different flip angles, $\pi/2$ and $\pi/4$, of the mixing pulse. In the H1′–H2″ cross peak, the ω_1 axis contains the multiplicity of the H2″ proton, while the ω_2 axis contains the multiplicity of the H1′ proton. In the case of

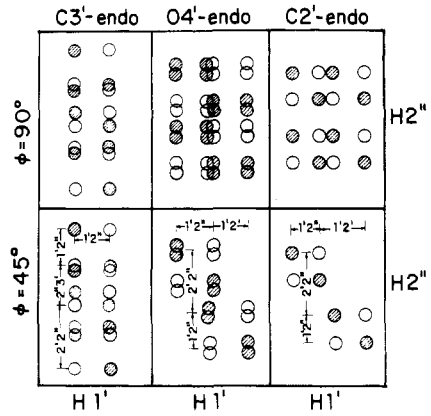

FIGURE 5: Calculated multiplet patterns in H1′–H2″ cross peak in COSY and COSY 45 spectra for three different sugar geometries, namely, C3′-endo, C2′-endo, and O4′-endo. Here again, the peak separations useful for estimation of coupling constants are indicated. Solid and open circles indicate positive and negative peaks, respectively.

C2′-endo geometry, all the multiplet components in the cross peak are well resolved (Widmer & Wüthrich, 1987; Hosur et al., 1988a,b; Chary et al., 1988) and measurement of J(H1′–H2′), J(H1′–H2″), and J(H2′–H2″) from the experimental spectra is straightforward. For O4′-endo geometry, a small but nonzero value of J(H2″–H3′) is buried in the widths of the component peaks along the ω_1 axis [see, for example, Figure 6, Chary et al. (1988)]. The ω_1-scaled COSY spectrum with $\phi = \pi/4$ has a particular advantage in the measurement of J values. In this spectrum, the number of components is reduced to about half the number in the spectrum with $\phi = \pi/2$. The spectrum enables direct measurement of J(H1′–H2′) and J(H1′–H2″) along the H1′ axis of the cross peak even when these coupling constants are similar in magnitude. Such an evaluation is not possible with $\phi = \pi/2$ since, here, the overlap between the positive and negative central components of the cross peaks results in cancellation of their intensities. Thus, better estimates of the two coupling constants J(H1′–H2′) and J(H1′–H2″) can be obtained from the COSY spectrum with $\phi = \pi/4$.

Accurate determination of J(H2″–H3′) from the H1′–H2″ cross peak and of J(H2′–H3′) from the H1′–H2′ cross peak is often difficult due to insufficient resolution of the components in the respective cross peaks. However, qualitative estimates of these coupling constants can be obtained by monitoring the relative intensities of all the cross peaks in a low-resolution COSY spectrum. Considering the digital resolution in our low-resolution COSY spectra, we feel that the absence of a particular cross peak indicates that the relevant active coupling constant is less than 3 Hz. In most DNA segments we have studied, H2″–H3′ cross peaks are found to be absent in the COSY spectrum. Estimates of J(H3′–H4′) can also be similarly obtained by comparing the intensities of the cross peaks in the low-resolution COSY spectrum. Recently a special J-scaling technique has been used to estimate the J(H3′–H4′) coupling constants (Bax & Lerner, 1988).

Variations in intensities of cross peaks in a COSY spectrum can occur either due to variations in the coupling constants or due to large variations in line widths for particular protons. Several studies have indicated that there can be significant variations in the values of the coupling constants in a given molecule. But large variations in line widths are generally not expected, since they would amount to significantly different motions in different parts of the duplex molecules. Therefore, it is logical to interpret intensity variations in COSY in terms of variations of J values rather than of line widths.

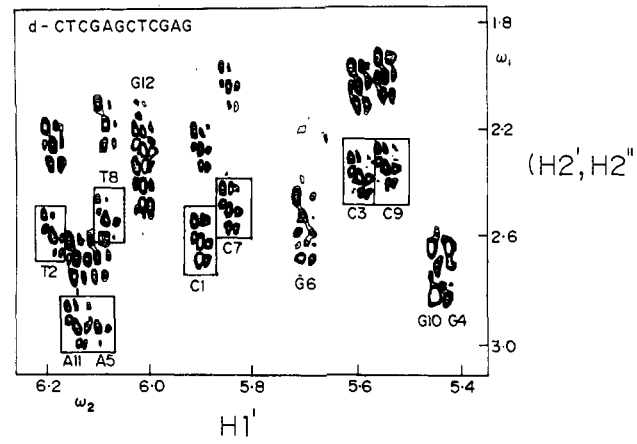

FIGURE 6: H1′–H2′,H2″ region of ω_1-scaled phase-sensitive COSY 45 spectrum of d-CTCGAGCTCGAG. The multiplets belonging to particular H1′–H2″ cross peaks are enclosed in boxes in all cases except G4, G10, and G12, where the separation between H1′–H2′ and H1′–H2″ cross peaks is not adequate. In all cases except G12, H2″ resonates downfield of H2′ proton (from NOESY). The patterns may be compared with the simulated patterns in Figure 5 for estimation of coupling constants.

An additional approach for obtaining J values is to use J-resolved spectroscopy. The J-resolved spectrum has exclusively coupling constant information along the ω_1 axis and consequently digital resolutions of the order of 0.2–0.3 Hz/point are achieved. The H1′ region of the spectrum that can often be analyzed yields precise values of J(H1′–H2′) and J(H1′–H2″). The J-resolved spectrum has another advantage. Estimation of coupling constants from COSY is hindered when the H2′ and H2″ protons have nearly identical chemical shifts. In such cases H1′–H2′ and H1′–H2″ cross peaks cannot be properly analyzed and the J-resolved spectrum can be helpful, although it is difficult to specifically assign the measured J values to J(H1′–H2′) or J(H1′–H2″). In principle, this can be achieved by the analysis of H2′, H2″, and H3′ regions of the spectrum, but these regions suffer from poor sensitivity. Therefore, it is essential to use J-resolved spectroscopy in conjunction with COSY to obtain accurate values of coupling constants.

Figure 6 shows the H1′–H2′ and H1′–H2″ cross-peak region of the phase-sensitive ω_1-scaled COSY spectrum, with $\phi = \pi/4$, of the present oligonucleotide. In Figure 7, blowups of H1′–H2″ cross peaks belonging to C1, C7, C3, and C9 units are shown with positive and negative signals identified. Such blowups are used for identifying the relative magnitudes of J(H1′–H2′) and J(H1′–H2″). An important feature of Figure 7 is that in the case of C1 and C7 J(H1′–H2″) is larger than J(H1′–H2′), whereas in C3 and C9 the opposite is the case. The pattern observed in C3 and C9 is the more commonly observed feature, and the observation of the opposite pattern in C1 and C7 is unique. Figure 8 shows the H1′ region of the amplitude-modulated J-resolved spectrum, from which fairly accurate values (±0.2 Hz) of the coupling constants J(H1′–H2′) and J(H1′–H2″) have been obtained. Rough estimates of J(H2″–H3′), J(H3′–H4′), and J(H2′–H3′) have been obtained by comparing the intensities of all the cross peaks in the low-resolution COSY spectrum. Figure 9 shows the (H2′, H2″)–H3′ and H3′–H4′ cross peaks of the low-resolution COSY spectrum used for this purpose. Both the H3′–H4′ and H2′–H3′ cross peaks show intensity variations and can be classified into three categories: strong (S), medium (M), and weak (W), which may be taken to indicate coupling constants in the ranges 6–10, 4–6, and 3–4 Hz, respectively. The coupling constants generally vary between 0 and 10 Hz, and cross

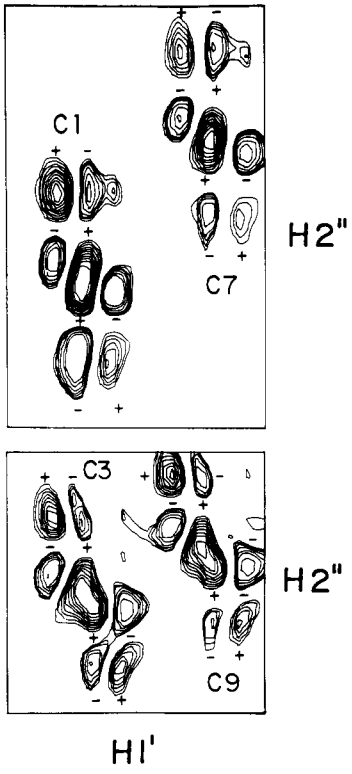

FIGURE 7: Blowups of H1′–H2″ cross peaks of C1, C7, C3, and C9 nucleotide units. The + − separations and the patterns of cross peaks in C1 and C7 are distinctly different from those of C3 and C9, pointing to unique sugar geometry differences.

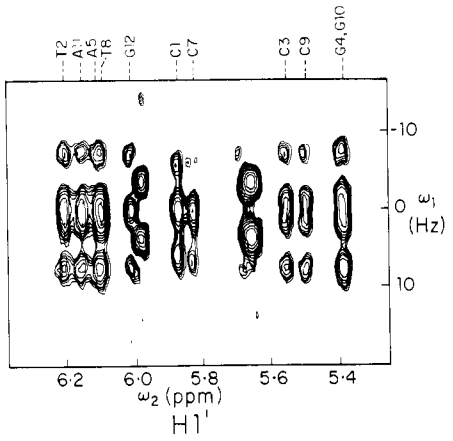

FIGURE 8: H1′ region of amplitude-modulated J-resolved spectrum of d-CTCGAGCTCGAG. The H1′ chemical shifts are identified at the top of the figure. A total of 32 t_1 increments were collected with 2048 t_2 points for each t_1 value. Final digital resolution after zero filling and Fourier transformation was 0.4 Hz/point along the ω_1 axis and 3.6 Hz/point along the ω_2 axis. Unlabeled peaks are H5 doublets of cytosines.

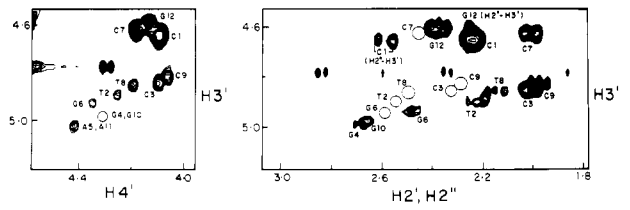

FIGURE 9: Portion of the absolute value COSY spectrum of d-CTCGAGCTCGAG showing H2′–H3′, H2″–H3′, and H3′–H4′ cross peaks. In the right panel, the open circles indicate the expected positions of H2″–H3′ cross peaks that are not observed.

peaks will not be seen if J is less than 3 Hz. The estimated coupling constants and intensity distributions are listed in Table II.

Table II: Measured Coupling Constants (J, Hz), H2′–H3′ and H3′–H4′ Cross-Peak Intensities, and Sugar Geometries in d-CTCGAGCTCGAG

	H1′–H2′	H1′–H2″	H2″–H3′	H2′–H3′[a]	H3′–H4′[a]	inferred[b] P value (deg)
C1	5.7	7.9	3–5	S	S	80
T2	9.0	5.8	<3	M	W	144
C3	8.8	6.0	<3	S	M	126
G4	9.5	5.8	<3	W	A	162
A5	9.0	5.8	<3	A	W	144
G6			<3	W	W	90–270
C7	5.6	7.9	<3	S	S	90
T8	9.0	5.8	<3	M	W	144
C9	8.8	6.0	<3	S	M	126
G10	9.5	5.8	<3	W	A	162
A11	9.0	5.8	<3	A	W	144
G12	7.2	7.2	<3	S	S	100

[a] S = strong; W = weak; M = medium, A = absent. [b] The values should be taken to correspond to the center of a narrow domain.

Conformation of Deoxyribose Rings. The coupling constant and intensity data (Table II) indicate that there is a wide variation in sugar geometries along the sequence of the molecule [for strategies see Hosur (1986), Hosur et al. (1986a,b, 1988a,b), and Chary et al. (1988)]. In the units C1 and C7, J(H1′–H2″) is larger than J(H1′–H2′); strong H3′–H4′ cross peaks are seen in the low-resolution COSY spectrum. For C1, the H2″–H3′ cross peak is seen, but this is absent for C7. These observations indicate that for the C7 unit the pseudorotation angle P is restricted to a narrow domain (around 90°), and C1 exhibits conformational mixture with the major conformer having P around 90°.

In the nucleotide units C3 and C9, J(H1′–H2′) is larger than J(H1′–H2″), and the H3′–H4′ cross peaks are of medium intensity. H2″–H3′ cross peaks are absent in the COSY spectrum. Therefore, the sugar rings in these two units can be fixed around P = 126°. Similarly, in the case of T2, T8, A5, and A11 units, J(H1′–H2′) is larger than J(H1′–H2″), but H3′–H4′ cross peaks in COSY are weak. Also no H2″–H3′ cross peaks are seen, indicating that the sugar pucker for these four units is around P = 144°. In the case of A5 and A11, the H2′–H3′ cross peaks also are not observed. This is unusual since J(H2′–H3′) is fairly large in the entire range of P (6–10 Hz), and generally this cross peak is not expected to be absent in the COSY spectra. Therefore, absence of these cross peaks for A5 and A11 must be attributed to unusually large line widths of H2′ protons in these two units.

Nucleotides G4 and G10 are at the centers of the halves of the molecule. For these two units, the chemical shifts of all the protons are nearly identical, which suggests that the sugar geometries are also similar. The COSY spectrum does not show separate H1′–H2′ and H1′–H2″ cross peaks. The J-resolved spectrum provides two coupling constant values (9.5, 5.8), but does not provide unique identification of J(H1′–H2′) and J(H1′–H2″). This poses a problem in the determination of the sugar geometries. However, the absence of an H3′–H4′ cross peak in the low-resolution COSY spectrum helps to restrict the sugar geometries to the domain P = 140°–280° (Hosur et al., 1986a). Considering that the two coupling constants J(H1′–H2′) and J(H1′–H2″) are widely different, one can further restrict the geometries to two narrower domains centered around P = 162° and P = 216°. From energetics and previous data, one may be inclined to assign the geometry to P around 162°, which is in the standard B-DNA region.

Units G6 and G12 deviate from the symmetrical behavior observed for other nucleotides. In G12, H2′ and H2″ are nonequivalent; they are equivalent in G6. Also in G12, the

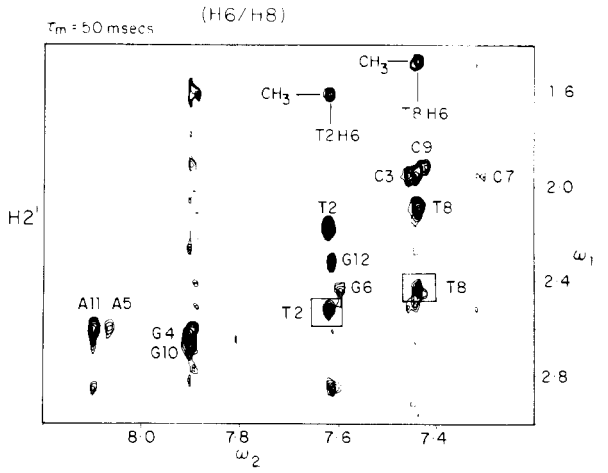

FIGURE 10: H6/H8–(H2′,H2″) region of a 500-MHz ω_1-scaled absorption mode NOESY spectrum recorded with a mixing time of 50 ms. All the self cross peaks to H2′ protons are labeled by specific nucleotide labels. Two peaks enclosed in square boxes are H8/H6–H2″ self cross peaks. The intensities of the peaks are monitored to get estimates of H8/H6–H2′ distances.

H2″ proton appears upfield of the H2′ proton, while the opposite is true in other units. Thus, the sugar geometries for G6 and G12 are not identical. This is also evident from the observation of a strong H3′–H4′ cross peak for G12, while G6 produces only a weak peak. Further, the fact that in G12 the coupling constants J(H1′–H2′) and J(H1′–H2″) are equal suggests that the geometry for this nucleotide unit is around $P = 100°$ (O4′-endo). For G6, the sugar geometry can be only qualitatively fixed in the S domain.

The sequence-specific sugar geometries discussed above are also listed in Table II.

A common practice in dealing with small and single-stranded segments (Govil & Hosur, 1982; Rinkel & Altona, 1987) has been to interpret the observed coupling constants in terms of mixtures of C3′-endo and C2′-endo sugar geometries used in models of A- and B-DNA or the related N and S families of conformations. In long double-helical oligonucleotides, rapid interconversion between widely different sugar geometries may be difficult due to higher energy barriers, and dynamism in sugar puckers is expected to be restricted to a narrow range of P values. Comparison of the experimental data in Table II with calculated values for an N–S mixture (Hosur et al., 1988b; Widmer & Wüthrich, 1987) suggests that in the present oligonucleotide, indeed, the dynamism is restricted to a narrow domain around the P value given for each nucleotide unit. In the case of terminal nucleotides, the experimental data are indicative of a more relaxed conformation.

Glycosidic Torsion Angles (χ). The relative intensities of intranucleotide H8/H6–H1′ and H8/H6–H2′,H2″ cross peaks in the NOESY spectrum, recorded with short mixing time (70 ms), indicate that the torsion angle χ is in the anti domain for all the nucleotide units. A more precise definition of χ has been obtained from the knowledge of interproton distances between the base protons and the sugar ring protons obtained from the NOESY spectrum recorded with a mixing time of 50 ms (Figure 10) (Wüthrich, 1986; Chary et al., 1988; Hosur et al., 1988a). Except for T8, T2, and possibly G12, no cross peak is seen from H8/H6 to H2″ proton, while all the H8/H6–H2′ cross peaks are observed. This spectral region also contains cross peaks from thymine methyl protons to their own H6. The intensity of this peak provides a marker for interproton distances, since this distance is around 3 Å. From a quantitative comparison of these intensities, we conclude that

Table III: Sequence-Specific[a] Values (deg) of Torsion Angles in the Central Portion of the Molecule Obtained from the Model in Figure 11

	χ	α	β	γ	δ	ϵ	ζ
A5	−120	−5	−180	53	142	−171	−107
G6	−121	−62	173	59	141	−157	−114
C7	−145	−54	153	51	104	120	−126
T8	−100	−137	−108	76	142	173	−50
A17	−122	−65	170	59	142	−161	−113
G18	−120	−50	168	44	143	−157	−130
C19	−152	−45	144	59	98	150	−135
T20	−101	−117	−111	60	143	164	−53
B-DNA	−102	−41	136	38	139	−133	−157

[a] The sequence is

1 12
C-T-C-G-A-G-C-T-C-G-A-G

G-A-G-C-T-C-G-A-G-C-T-C
24 13

the H8/H6–H2′ distances are in the range 2.5–3.0 Å in all nucleotides except C7. In C7, the H6–H2′ cross peak is extremely weak, suggesting that this distance is larger than 3.5 Å. Noting that the H8/H6–H2′ distance varies only mildly with sugar pucker in the anti domain of χ (Wüthrich, 1986), the above distance of 2.5–3.0 Å indicates that the χ value is in the range $-130° \pm 30°$. The values are close to −100° for T2 and T8 as evidenced by the observation of H6–H2″ cross peaks in these two cases. For C7, the distance of 3.5 Å suggests that χ is around −160°.

Structure of XhoI Cleavage Site. The molecule has two cleavage sites for the restriction enzyme *Xho*I (Figure 1). The enzyme recognizes the hexanucleotide stretch d-CTCGAG and cleaves the molecule between C and T. It is interesting to note from the results (Table II) that the cleavage sites exhibit a very significant variation in sugar geometry: P around 90° for C1 and C7 and around 144° for T2 and T8. Since the sugar geometry plays a crucial role in determining the overall structure of DNA, the observed variations in sugar geometries at the cleavage sites are expected to bring about substantial variation in the backbone structure of the molecule. We note that the NMR data also indicate substantial variations in the χ angles at these points. Besides, the H2′ resonances are broader in A5 and A11 units, which are near the cleavage sites. At the same time, the AT base pairing is preserved.

To visualize the effects of perturbations due to sugar pucker and glycosidic torsion angle variations in the middle segment of d-CTCGAGCTCGAG, we have built a double-helical model of the molecule using the FRODO program commonly used by X-ray crystallographers (Jones, 1978). This program uses the restrained regulation procedure of Hermans and McQueen (1974) for the refinement of molecular structures. The model-building procedure consisted of generating the DNA molecule of the required sequence and regularizing the model by several cycles of least-squares refinement using restraint information. Thus, structure can be built to satisfy experimental and stereochemical constraints. The starting point consisted of a standard B-DNA geometry (Arnott, 1980). However, the sugar conformation of C7 was changed to the one observed by NMR, namely, $P = 90°$. All the glycosidic torsion angles were also initialized to the values obtained from NMR data. The structure was then allowed to refine for 30 cycles under the constraint that all base pairs remain intact. The molecule converged to a structure that had no bad interatomic distances and showed overall similarity with B-DNA (Figure 11). Table III lists the backbone torsion angles in

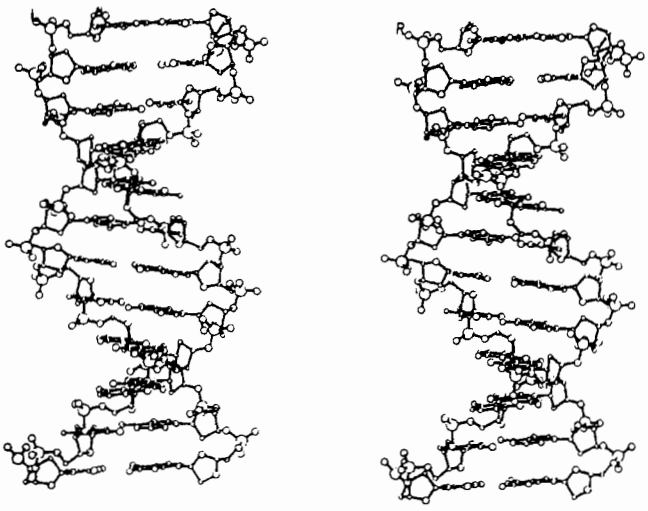

FIGURE 11: Stereopair of the structure of d-CTCGAGCTCGAG. A double helix of the molecule was first built with standard B-DNA geometry and then with the following constraints in FRODO dictionary: (i) base pairs intact; (ii) sequence-specific χ angles as observed by NMR; (iii) C7 sugar in O4′-endo geometry and the others in C2′-endo geometry, the backbone torsion angles were allowed to vary in the FRODO refinement, so that a covalently acceptable structure is obtained.

the central segment of the molecule for the final structure thus arrived at. The converged backbone torsion angles in this segment show considerable differences from those of the starting B-DNA structure. The values of the corresponding torsional angles for the two strands are slightly different (within 20°), essentially reflecting the accuracy of the refinement. The oligonucleotide has a palindromic sequence, and these aspects are preserved (to within 20°) following the FRODO refinement.

Sequence Effects. It is worthwhile to compare the structure of the present molecule (V) with that of d-GGTACGCGTACC (IV) (Chary et al., 1988), which has the same base composition as the present oligonucleotide but a different base sequence. The differences in the base sequences in the two molecules are as follows: (i) Molecule IV does not have the same base sequence symmetry as V. (ii) V has alternating stretches, 3 units long, of pyrimidines and purines, while IV has alternating pyrimidine and purines in the central 8 nucleotide units. The two ends in the latter have purines and pyrimidines, respectively. (iii) In IV, two A·T base pairs are adjacent and are flanked by two G·C base pairs on either side. On the other hand, in V, the A·T base pairs are separated by two G·C base pairs regularly all along the sequence. This difference would manifest itself in different base stacking and tilts owing to the fact that the A·T base pairs exhibit a greater propellar twist than G·C base pairs in double-helical DNA structure (Drew et al., 1981; Dickerson & Drew, 1981).

On comparison of the structures of these molecules it is clear that the sequence plays a crucial role in determining the 3D structure of the molecule. While molecule IV shows a rather uniform structure, molecule V exhibits substantial variations along the sequence. The sugar geometries and backbone torsion angles show significant variations as discussed in previous sections. Considering that the two molecules are recognized by different restriction enzymes, the observed sequence dependence of the structure suggests that the DNA geometry provides important hot points for recognition in addition to the direct specific interactions at the base level.

ACKNOWLEDGMENTS

The facilities provided by 500 MHz FT-NMR National Facility supported by the Department of Science and Technology and located at TIFR, Bombay, are gratefully acknowledged.

Registry No. d-CTCGAGCTCGAG, 110879-74-8; restriction endonuclease *Xho*I, 81295-43-4.

REFERENCES

Anil Kumar, Wüthrich, K., & Ernst, R. R. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–8.

Anil Kumar, Hosur, R. V., & Chandrasekhar, K. (1984) *J. Magn. Reson. 60*, 143–149.

Arnott, S., Chandrasekharan, R., Birdsall, D. L., Leslie, A. G. W., & Ratcliffe, R. L. (1980) *Nature (London) 283*, 743–746.

Bax, A., & Freeman, R. (1981) *J. Magn. Reson. 44*, 542–561.

Bax, A., & Lerner, L. (1988) *J. Magn. Reson. 79*, 429–438.

Chary, K. V. R., Hosur, R. V., Govil, G., Tan, Z. K., & Miles, H. T. (1987) *Biochemistry 26*, 1315–1322.

Chary, K. V. R., Hosur, R. V., Govil, G., Chen, C., & Miles, H. T. (1988) *Biochemistry 27*, 3858–3867.

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

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

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179–2183.

Govil, G., & Hosur, R. V. (1982) *Conformation of Biological Molecules: New Results from NMR*, Springer Verlag, Heidelberg.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Hermans, J. S., & McQueen, J. R. (1974) *Acta. Crystallogr. A30*, 730–739.

Hosur, R. V. (1986) *Curr. Sci. 55*, 597–605.

Hosur, R. V., Ravikumar, M., Roy, K. B., Tan, Z-k., Miles, H. T., & Govil, G. (1985a) in *Magnetic Resonance in Biology and Medicine* (Govil, G., Khetrapal, C. L., & Saran, A., Eds.) pp 243–260, Tata McGraw-Hill, New Delhi.

Hosur, R. V., Ravikumar, M., & Sheth, A. (1985b) *J. Magn. Reson. 65*, 375–381.

Hosur, R. V., Chary, K. V. R., & Ravikumar, M. (1985c) *Chem. Phys. Lett. 106*, 105–108.

Hosur, R. V., Ravikumar, M., Chary, K. V. R., Sheth, A., Govil, G., Tan, Z. K., & Miles, H. T. (1986a) *FEBS Lett. 205*, 71–76.

Hosur, R. V., Sheth, A., Chary, K. V. R., Ravikumar, M., Govil, G., Tan, Z. K., & Miles, H. T. (1986b) *Biochem. Biophys. Res. Commun. 139*, 1224–1232.

Hosur, R. V., Chary, K. V. R., Sheth, A., Govil, G., & Miles, H. T. (1988a) *J. Biosci. 13*, 71–86.

Hosur, R. V., Govil, G., & Miles, H. T. (1988b) *Magn. Reson. Chem. 26*, 927–944.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4554.

Jones, T. H. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95–117.

Maxam, M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Mueller, L. (1987) *J. Magn. Reson. 72*, 191.

Nilges, M., Clore, G. M., Gronenborn, A. M., Brunger, A. T., Karplus, M., & Nilsson, L. (1987) *Biochemistry 26*, 3718–3733.

Ravikumar, M., Hosur, R. V., Roy, K. B., Miles, H. T., & Govil, G. (1985) *Biochemistry 24*, 7703–7711.

Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., &

Williams, D. H. (1983) *Biochemistry 22*, 2019-2025.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621-649.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & Van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914-2916.
Scheek, R. M., Boelens, R., Russo, N., Van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371-1376.
Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z-k., & Miles, H. T. (1987a) *Biochem. Biophys. Res. Commun. 144*, 26-34.
Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z-k., Roy, K. B., & Miles, H. T. (1987b) *Biopolymers 26*, 1301-1313.
Tan, Z-k., Ikuta, S., Huang, T., Dagaiczyk, A., & Itakura, K. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 383-391.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316-336.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wynants, C., & Van Binst, G. (1984) *Biopolymers 23*, 1799-1804.

NMR Studies of DNA $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ Triple Helices in Solution: Imino and Amino Proton Markers of T·A·T and C·G·C^+ Base-Triple Formation†

Carlos de los Santos, Mark Rosen, and Dinshaw Patel*
Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032

Received April 14, 1989; Revised Manuscript Received May 23, 1989

ABSTRACT: High-resolution exchangeable proton two-dimensional NMR spectra have been recorded on 11-mer DNA triple helices containing one oligopurine $(R)_n$ and two oligopyrimidine $(Y)_n$ strands at acidic pH and elevated temperatures. Our two-dimensional nuclear Overhauser effect studies have focused on an 11-mer triplex where the third oligopyrimidine strand is parallel to the oligopurine strand. The observed distance connectivities establish that the third oligopyrimidine strand resides in the major groove with the triplex stabilized through formation of T·A·T and C·G·C^+ base triples. The T·A·T base triple can be monitored by imino protons of the thymidines involved in Watson–Crick (13.65–14.25 ppm) and Hoogsteen (12.9–13.55 ppm) pairing, as well as the amino protons of adenosine (7.4–7.7 ppm). The amino protons of the protonated (8.5–10.0 ppm) and unprotonated (6.5–8.3 ppm) cytidines in the C·G·C^+ base triple provide distinct markers as do the imino protons of the guanosine (12.6–13.3 ppm) and the protonated cytidine (14.5–16.0 ppm). The upfield chemical shift of the adenosine H8 protons (7.1–7.3 ppm) establishes that the oligopurine strand adopts an A-helical base stacking conformation in the 11-mer triplex. These results demonstrate that oligonucleotide triple helices can be readily monitored by NMR at the individual base-triple level with distinct markers differentiating between Watson–Crick and Hoogsteen pairing. Excellent exchangeable proton spectra have also been recorded for $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ 7-mer triple helices with the shorter length permitting spectra to be recorded at ambient temperature. Our NMR studies open up the possibility of studying triple helices containing errors and lesions, as well as the interaction of ligands with this DNA structural motif in aqueous solution.

There is currently great interest in the chemistry and biology of triple-stranded DNA helices and their potential role in the regulation of the eukaryotic genome [reviewed in Wells et al. (1988), Htun and Dahlberg (1989) and Broitman and Fresco (1989)]. The early characterization of the poly(U·A·U) triple helix where the third strand was an oligopyrimidine (Felsenfeld et al., 1957) has been complemented by the more recent characterization of the poly(A·U·A) triple helix where the third strand is an oligopurine (Broitman et al., 1987). Related studies established the sequence requirements (Riley et al., 1966; Morgan & Wells, 1968) and the role of pH (Lee et al., 1984) for triple-helix formation.

Much of the recent interest in triple-helix formation involving (deoxypurine)$_n$·(deoxypyrimidine)$_n$ sequences [designated $(R)_n\cdot(Y)_n$] results from the observation of such tracts at recombination hot spots and upstream from eukaryotic genes [reviewed in Wells et al. (1988)]. The detection of S1 nuclease hypersensitivity in active chicken globin chromatin (Larson & Weintraub, 1982) and its fine mapping to $(R)_n\cdot(Y)_n$ stretches (Nickol & Felsenfield, 1983; Schon et al., 1983) focused much attention on the unusual DNA structure at such sites. The observed hyperreactivity was explained through $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ triple-helix formation (the signs indicate the directionality of the strands) with the remaining single-stranded $(R-)_n$ segment sensitive to single-stranded nucleases (Christophe et al., 1985; Lyamichev et al., 1986). Triple-helix formation in $(R)_n\cdot(Y)_n$ mirror repeats is favored by both superhelical stress and low pH (Mirkin et al., 1987) and can be readily monitored by two-dimensional gel electrophoresis (Mirkin et al., 1987; Collier et al., 1988) and chemical probes (Hanvey et al., 1988a,b; Voloshin et al., 1988; Htun & Dahlberg, 1988; Johnston, 1988). Finally, chemical footprinting studies on $(G)_n\cdot(C)_n$ sequences in supercoiled plasmid DNA suggest a switch from a $(G+)_n\cdot(C-)_n\cdot(C+)_n$ triple helix in the absence of Mg ions to a $(G+)_n\cdot(C-)_n\cdot(G-)_n$ triple helix

† The research was supported by NIH Grant GM34504. C.d.l.S. is supported, in part, by a fellowship from the CONICET, Republic of Argentina. The NMR spectrometers were purchased from funds donated by the Robert Woods Johnson Trust toward setting up an NMR Center in the Basic Medical Sciences at Columbia University.

0006-2960/89/0428-7282$01.50/0