Resonance Assignment of the 500-MHz Proton NMR Spectrum of Self-Complementary Dodecanucleotide d-GGATCCGGATCC: Altered Conformations at *Bam*HI Cleavage Sites†

M. Ravi Kumar,‡ R. V. Hosur,‡ K. B. Roy,§ H. T. Miles,*,∥ and G. Govil‡

Chemical Physics Group, Tata Institute of Fundamental Research, Bombay 400 005, India, Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20205, and Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India

Received February 26, 1985; Revised Manuscript Received July 29, 1985

ABSTRACT: Resonance assignments of nonexchangeable base and sugar protons of the self-complementary dodecanucleotide d-GGATCCGGATCC have been obtained by two-dimensional NMR methods and strategies derived from interproton distance calculations on different secondary structures of nucleic acids. Conformational details about the glycosidic dihedral angle and sugar pucker have been derived from the relative intensities of cross peaks in the two-dimensional *J*-correlated and nuclear Overhauser enhancement correlated spectra in D_2O solution. It is observed that d-GGATCCGGATCC assumes a predominantly B-type conformation with sequence-dependent changes along the chain. The recognition site of *Bam*HI shows a distinctly different geometrical environment. The sugar rings of G1 and G7 assume a C3′-endo geometry while the rest of the sugars possess C2′-endo geometry.

Recognition of nucleic acid base sequences by certain proteins occurs in a highly specific manner, e.g., operators by repressors, promoters by RNA polymerases, DNA restriction sites by restriction endonucleases, etc. (Ohlendorf et al., 1983). This specificity in recognition may sometimes arise from subtle differences in the three-dimensional structure of these molecules. Several techniques have therefore been used for determination of three-dimensional structures of proteins and nucleic acids, and among these nuclear magnetic resonance (NMR) spectroscopy has emerged as a very powerful tool for studies in aqueous solutions (Govil & Hosur, 1982). Development of two-dimensional Fourier-transform NMR (2D NMR) has added a new dimension to this technique, and in recent years it has been applied to structure determination of proteins and nucleic acids. The first step in structure determination of molecules by NMR is the assignment of resonances to individual nuclei in the molecule. In the case of proteins, clear-cut strategies have been evolved to obtain unique resonance assignments, in ^1H NMR spectra (Wagner et al., 1982; Wider et al., 1982; Arseniev et al., 1983; Hosur et al., 1983; Strop et al., 1983). On the other hand, in the case of nucleic acids, there is no unique procedure that will work for all DNA structures, namely, A DNA, B DNA, Z DNA, and single-stranded DNA, and a variety of RNA structures (Feigon et al., 1982; Hare et al., 1983; Scheek et al., 1983; Hasnoot et al., 1984). Pardi et al., (1983) have used a combination of two-dimensional homonuclear correlated spectroscopy (COSY) and heteronuclear ^{31}P–H correlation techniques to arrive at resonance assignments of nonexchangeable protons in the case of a tetranucleotide that exists as single strand in aqueous solutions. Hare et al. (1983) have used COSY and two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) techniques to carry out sequential assignment of nonexchangeable protons along the nucleic acid chain in d-CGCGAATTCGCG. Scheek et al. (1983) have carried out sequential assignment of base and sugar H1′ protons in a mixture of d-TGAGCGG and its complimentary oligomer 4-CCGCTCA entirely from the NOESY spectrum. Westerink et al. (1984) carried out a similar investigation of r-CGCGCG. The COSY experiment has been used to identify the sugar proton spin systems and the cytosine base proton chemical shifts. These, along with the NOESY experiment, have been used to jump from one nucleotide to the next nucleotide via base–H1′ NOE. Such a strategy is designed for B-DNA structures and is not applicable to other types of structures mentioned above. It is not clear whether nucleotide sequence influences the applicability of these strategies. Until now the main difficulty in evolving sequential assignment strategies using NOESY experiments in nucleic acids has been in the identification of suitable interproton distances between successive nucleotide units that can give rise to NOEs and are independent of the overall secondary structure of the molecule. In addition, the overlap of resonances of sugar protons and water (HOD) presents difficulties in the unambiguous assignment of 3′, 4′, 5′, and 5″ protons from a COSY spectrum. Thus, these protons cannot be used for sequential assignment purposes in a general way, and the sugar H1′, H2′, and H2″ and the base protons are the most useful probes for both sequential assignment and structure determination purposes. We have carried out a systematic analysis of interproton distances from the NMR viewpoint and have evolved new strategies for resonance assignment and structure determination of oligonucleotides.

In this paper we describe assignment of nonexchangeable protons in a self-complementary dodecanucleotide d-GGATCCGGATCC using two-dimensional COSY and NOESY experiments. Interstrand NOEs have been observed that prove that the molecule is in double-helical state at 25 °C and pH 7.2, and further, from the pattern of NOESY cross-peak intensities, it is inferred that the molecule adopts a B-type conformation. This is, possibly, the first complex

† Results of this study were presented at the XIth International Conference on Magnetic Resonance in Biological Systems in Goa, India, Sept 17–23, 1984, and at the International Symposium on Quantum Biology and Quantum Pharmacology in Gainsville, FL, March 14–16, 1985.

‡ Tata Institute of Fundamental Research.

§ All India Institute of Medical Sciences.

∥ National Institutes of Health.

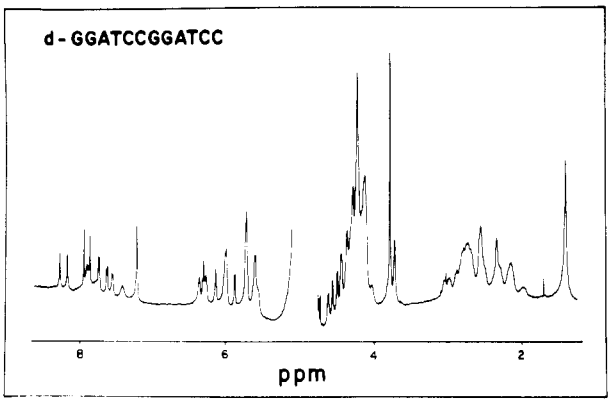

FIGURE 1: 500-MHz proton NMR spectrum of d-GGATCCGGATCC in D_2O solution at 25 °C.

nucleotide sequence for which conformational details have been obtained entirely from the NMR data in aqueous solutions. A striking feature observed in the conformation of d-GGATCCGGATCC is that the sugars in G1 and G7 nucleotides adopt a C3′-endo geometry while the rest of the sugars have a C2′-endo conformation.

Materials and Methods

Synthesis. GGATCCGGATCC was synthesized by the solid-phase phosphotriester method described by Tan et al. (1982). The 3′-terminal C was attached to the polymer support and joined to a blocked monomer C. The succeeding 10 residues were attached as blocked dimers. The deblocked oligomer was purified by chromatography 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 gave a single peak on HPLC (Bondapak C_{18}; gradient 5–10% acetonitrile). The sequence was confirmed by the procedure of Maxam & Gilbert (1980).

NMR. ^1H NMR experiments were carried out on a Bruker AM 500 FT-NMR spectrometer operating at 500-MHz frequency for proton. A total of 5 mg of d-GGATCCGGATCC (~4 mM on DNA strand) was dissolved in 0.5 mL of phosphate buffer (0.02 M) having pH 7.2. Total counterion concentration was 0.08 M. The solution was lyophilized, redissolved in D_2O, again lyophilized, and finally made up to 0.5 mL with D_2O whose pH was adjusted to 7.2 (pH meter reading) with NaOD. Two-dimensional NOESY experiments were carried out with 1024 data points along the t_2 axis and 450 data points along the t_1 axis. The COSY experiment was carried out with 2048 data points and 400 data points along t_2 and t_1 directions, respectively. The COSY pulse sequence ($90-t_1-90-t_2$) was modified by introducing a fixed delay of 0.005 s after each 90° pulse, so as to enhance cross-peak intensities in the two-dimensional spectrum. The HOD signal was suppressed with low-power continuous irradiation by the scheme of Wider et al. (1983) to avoid dynamic range problems. A relaxation delay of 1 s was given in all two-dimensional experiments. The time domain data were multiplied by sine square bell and sine bell window functions prior to Fourier transformation along t_2 and t_1 directions, respectively. The digital resolution along both ω_1 and ω_2 axes was 7.8 Hz. Chemical shifts are expressed with respect to sodium 3-(trimethylsilyl)[2,2,3,3-^2H]propionate (TSP).

Results and Discussion

Figure 1 shows the 500-MHz ^1H NMR spectrum of d-GGATCCGGATCC in D_2O solution at 25 °C. The spectrum consists of well-resolved lines between 6.5 and 9.0 ppm, which belong to the nonexchangeable base protons in the oligonucleotide. The region between 2.0 and 6.0 ppm, which contains sugar proton resonances, shows an extensive crowding of lines. The one-dimensional NMR techniques are clearly insufficient for unambiguous assignment of all the individual proton resonances. In the following subsections, we discuss the use of 2D NMR in obtaining resonance assignment and structure of oligonucleotides with special reference to d-GGATCCGGATCC.

Two-Dimensional NMR Techniques. The principles of two-dimensional COSY (Aue et al., 1976; Jeener, 1971; Nagayama, 1980) and NOESY (Macura et al., 1981; Anil Kumar et al., 1980) have been described in detail. The cross peaks in the COSY spectrum reflect *J*-coupling correlations and most often show two-bond and three-bond correlations. The NOESY spectrum on the other hand reflects dipolar coupling correlations between protons and indicates which pairs of protons are close by in space. (The shorter the distance between protons, the larger is the intensity of cross peaks.) Although several other factors such as internal motions, tumbling, and data processing affect the intensities, the relative intensities can be used to obtain qualitative information about relative distances. For example, the cross-peaks between the H1′ proton and the two protons on the C2′ atom of a sugar ring indicate relative magnitudes of H1′–H2′ and H1′–H2″ distances. Thus in such cases, NOE cross peaks do have structural information; a priori, appearance of a cross peak in NOESY spectrum can indicate a distance less than 5 Å (Kooper & James, 1984). The COSY and the NOESY are the two crucial experiments that enable unique assignments in complex systems.

Resonance Assignment Strategies. The deoxyribose ring protons of the individual nucleotide units constitute a complex coupled-spin network and produce several cross peaks in the COSY spectrum. A typical pattern is schematically shown in Figure 2, where the chemical shifts have been chosen on the basis of results from small nucleotides. Among the two protons on C2′, the one which is cis to C5′ is generally designated as H2′. Often, the H2″ proton appears downfield with respect to the H2′ proton in the NMR spectrum. The H1′ protons and H2′ and H2″ protons occur at widely separated positions in the one-dimensional spectrum and thus produce distinct cross-peaks quite far from the diagonal, which essentially represents the 1D spectrum. Similarly, H2′ and H3′ cross-peaks also appear quite far from the diagonal and are easily identifiable. However, when one is working in aqueous solutions, the intense solvent signal masks the H3′ protons, and in such cases, observation of cross-peaks involving H3′ protons becomes difficult. The cross peaks between H3′, H4′, and H5′ protons occur very close to the diagonal and can be discerned in special situations where they have a better spread in the chemical shifts. Since the sequence of nucleotides plays an important role in determining the chemical shifts of the protons, success in identification of all sugar ring spin systems is heavily dependent on the sequence of nucleotides in the DNA fragment. However, it seems possible that the positions of H1′, H2′, and H2″ protons can always be identified from the COSY spectrum. The spin system thus identified can not, however, be assigned to particular nucleotide units along the sequence, as there are no $^1H-^1H$ *J*-coupling correlations between consecutive residues that can be observed in the COSY spectrum. At this stage, the NOESY spectrum, which reflects distance correlations, serves to resolve the problem in favorable cases. Various cross peaks can be seen connecting intranucleotide protons, and one can obtain additional assignments of sugar ring protons not obtained in the COSY spec-

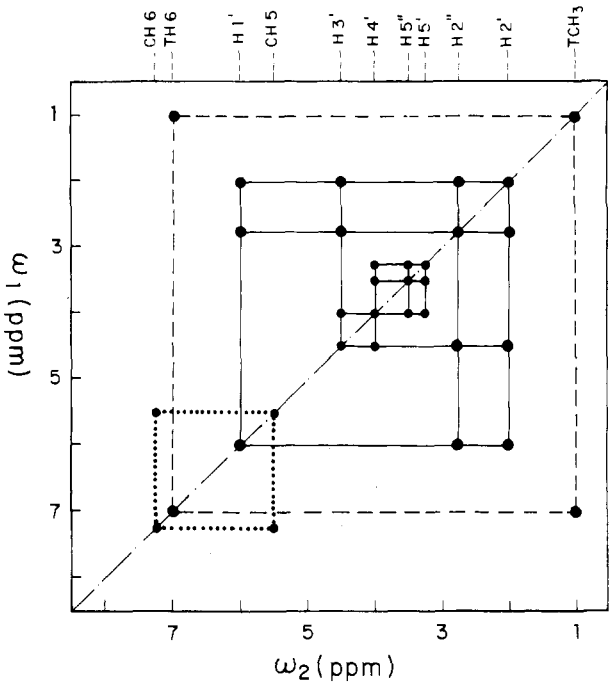

FIGURE 2: Schematic COSY diagram showing typical connectivity pattern of nonexchangeable protons in nucleotides. Chemical shift positions chosen for various protons indicate regions of common occurrence of respective protons. The pattern could look slightly different in the case of overlap or equivalence of certain groups of protons. (—) Connectivities of sugar protons; (--) J connectivity between thymine CH_3 and H6 protons; (···) H6–H5 connectivity of cytosine base protons.

trum. NOESY cross peaks between base protons and sugar ring protons can be used to identify which spin systems are associated with which nucleotides along the sequence of the molecule. Base to base NOEs help in direct identification of base protons belonging to neighboring nucleotide residues. Figure 3 shows distance plots for the following proton pairs; (a) base–H1′, (b) base–H2′, base–H2″, and (c) base proton of the nth nucleotide to H2′ and H2″ protons of the $(n-1)$th residue, as a function of χ dihedral angles for the two base deoxribose conformations C3′-endo and C2′-endo. The backbone dihedral angles have been fixed at the following values: $\chi = 36°$, $\chi' = 83°$, $\phi' = 180°$, $\omega = 314°$, $\omega' = 314°$, and $\phi = 214°$ [for nomenclature, see Govil & Hosur (1982)]. In Figure 3, distances greater than 4 Å have not been plotted, since they would produce only weak NOEs. In Figure 3a, only one curve is shown since both C3′-endo and C2′-endo geometries give rise to identical dependence of the base (H8/H6) to H1′ distance. This curve shows a minimum around $\chi = 240°$, which corresponds to a syn conformation, indicating that strong NOEs should be observed between base and H1′ protons when the nucleotide has a syn conformation with respect to the glycosidic bond. Further, due to the fact that for the entire range of χ angles the interproton distance is within 4 Å, one should, in principle, always observe base–H1′ NOEs. On the other hand, base (H8 or H6)–H2′ and –H2″ curves (Figure 3b) show much deeper minima and span only half of the complete range of χ values. For C3′-endo geometry, both the curves show minimum around $\chi = 140°$, and the two protons have always very different distances. In contrast, for C2′-endo geometry, the two protons (H2′, H2″) show a minimum at values of 100° and 150° for the dihedral angle. In this case, the two curves intersect, suggesting that for χ values in the neighborhood of 150° the two protons are equidistant and should show similar cross-peak intensities, in the NOESY spectrum. Comparison of the curves in Figure 3a,b suggests

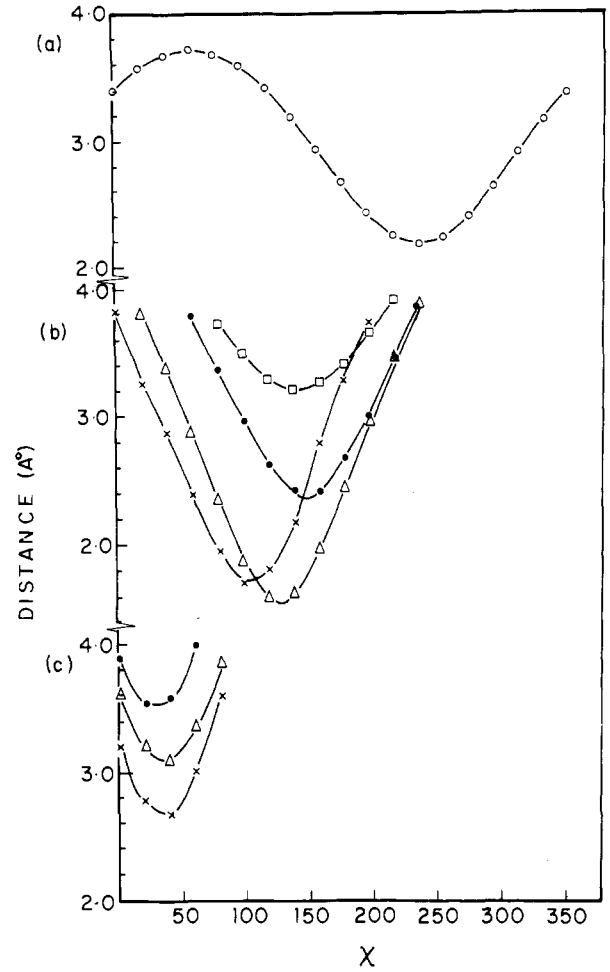

FIGURE 3: Plots of distance as a function of angle between (a) H8/H6 of base and H1′ of sugar and (b) H8/H6 of base and H2′ and H2″ protons of sugar for C3′-endo and C2′-endo sugar puckers: (×) H2′ and C2′-endo geometry; (●) H2″ and C2′-endo geometry. (△ and □) Corresponding curves for C3′-endo geometry. (c) H8/H6 proton of the nth nucleotide to H2′ and H2″ protons for the two sugar geometries of the $(n-1)$th nucleotide: (● and △) H2′ and H2″ protons, respectively, for C2′-endo geometry. For C3′-endo geometry only the H2′ proton shows a distance less than 4 Å.

that for values (30° and 90°) that correspond to an anti conformation the H2′ and H2″ protons should show much stronger NOEs to the base protons than does the H1′ proton in the sugar ring. Plots of distances between base protons and sugar H2′, H2′, and H2″ protons of the previous residue (Figure 3c) show that for C2′-endo geometry both base–H2′ and base–H2″ distances are shorter than 4 Å, whereas for C3′-endo geometry only one of them has a short distance. The base–H1′ distance is always greater than 4 Å, indicating weaker NOEs. Thus, it appears that sequential assignment of base and sugar protons should proceed via H2″ protons for right-handed DNA structures, since these NOEs are expected to be much more intense than the corresponding ones to the H1′ protons. In practice, however, because of the overlap of cross peaks or because of the possibility that some of the NOEs may be weak due to local fluctuations in structure, one may have to obtain additional information through the base (H8/H6)–H1′ connectivity and the base (H8/H6)–H2″ connectivity for the complete assignment of the resonances. The strategies using H1′ and H2″ protons and base protons are schematically shown in Figure 4a.

In the case of Z DNA, the above strategies may not work, and alternative strategies have to be evolved. Calculation of interproton distances from the reported crystal structure of

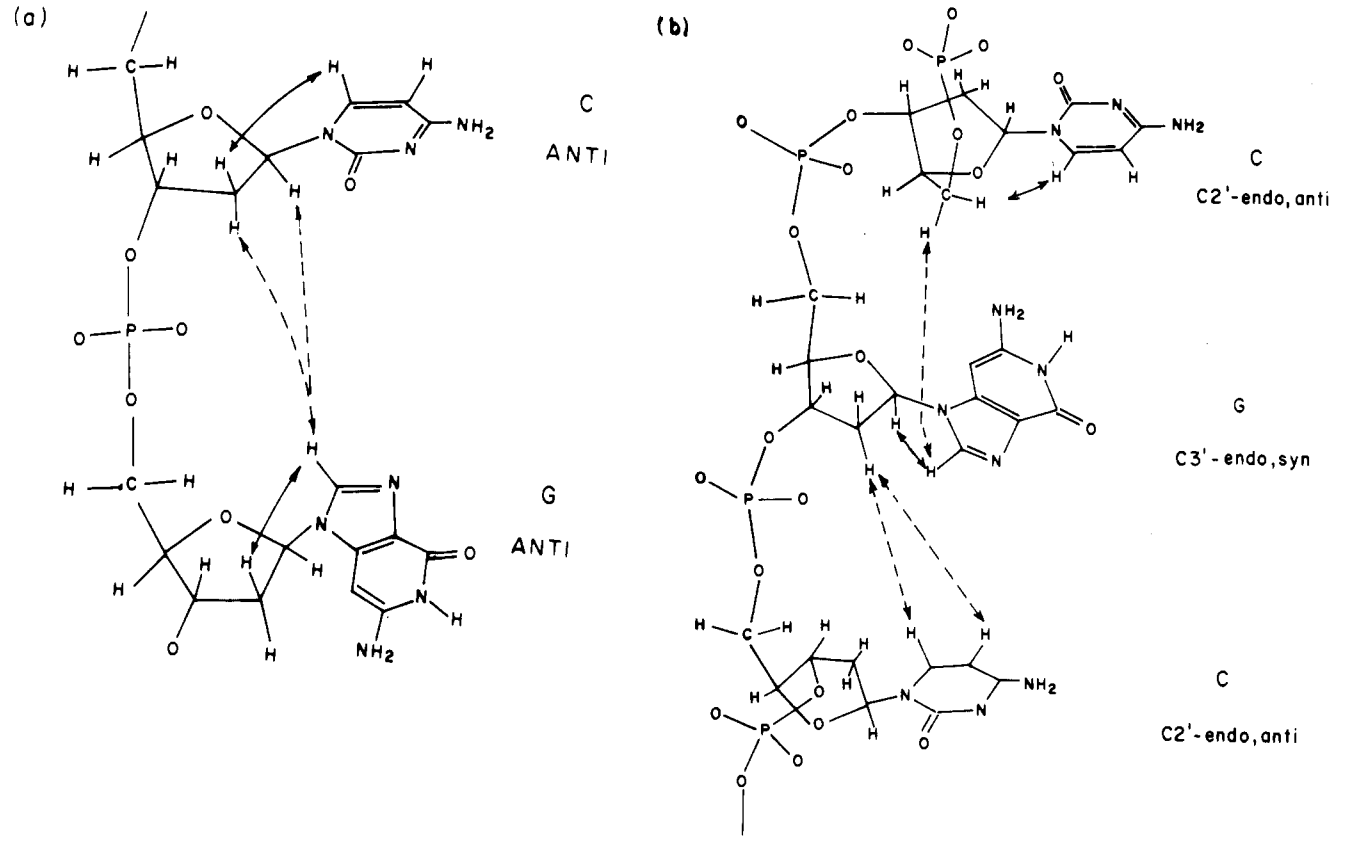

FIGURE 4: Strategies of sequential resonance assignments for (a) B-DNA/A-DNA and (b) Z-DNA structures. Double-pointed arrows indicate observable NOE correlations that are to be used in sequential assignment. Thick arrows indicate intranucleotide NOEs, and dotted arrows indicate internucleotide NOEs. In part b, a trinucleotide is shown, since a dinucleotide is the repeating unit for Z-DNA structure.

d-CGCGCG (Wang et al., 1979, 1981) in the Z conformation suggests a number of distances that are shorter than 4 Å. A list of such short distances is as follows: CH6 (n)–H5′ (n), –H2′ (n), –H2′ ($n - 1$), and –H3′ ($n - 1$); CH5′ (n)–H2′ ($n - 1$); GH8 (n)–H5′ ($n - 1$) and –H1′ (n); H5′ (n)–H1′ ($n - 1$), –H3′ ($n - 1$), and –H4′ ($n - 1$). From these results, a convenient strategy employing NOESY connectivities for sequential assignment has been developed, and this is shown in Figure 4b. Since in Z DNA the repeating unit is a dinucleotide, there are two sets of internucleotide connectivities. Thus in Z DNA sequential connectivity would proceed in the following manner: base ($2n - 1$)–H5′ ($2n - 1$)–base ($2n$)–H1′ ($2n$)–H2′ ($2n$) and H2″ ($2n$)–base ($2n + 1$).

The fact that the sugar H5′ and H5″ protons have to be used for sequential assignment poses a slight problem since sometimes the COSY spectrum does not yield unambiguous assignment of H5′ and H5″ protons due to extensive overlap in this region. However, base (H6 or H8) to H5′ and H5″ NOEs within the same nucleotide can be used to circumvent this problem.

Assignments in d-GGATCCGGATCC. With these strategies, we have attempted resonance assignment and structure determination of a dodecanucleotide having the sequence d-GGATCCGGATCC. Figure 5 shows the 500-MHz COSY spectrum of d-GGATCCGGATCC at 25 °C. All the four cytosines show distinct cross-peaks between H6 and H5 protons. A cross peak is also seen between methyl and H6 protons of the thymine nucleotides. Appearance of a single cross peak suggests that both the methyls as well as the H6 protons are nearly equivalent. This is also evident from NOESY spectra, as will be seen later. The H1′–H2′, H1′–H2″ and H2′–H2″ cross peaks are clearly seen. However, several cross peaks between H2′ (H2″) and H3′ protons are not clearly seen, probably because these are buried under the water signal and are saturated owing to continuous irradiation of the water resonance. Thus, the spin system connectivities could not be extended beyond H2′ and H2″ protons from the COSY spectrum as illustrated in Figure 6.

NOESY experiments have been carried out with two different mixing times: 200 and 300 ms. Most of the cross peaks have higher intensity in the 300-ms spectrum compared to the 200-ms spectrum. Besides, the 300-ms spectrum shows a greater number of cross peaks. We have therefore used the 300-ms NOESY spectrum (Figures 7 and 8) for resonance assignment purposes. Figure 8 shows base–base connectivities between H6 protons of T4 and T10 with H8 protons of A3 and A9 residues, respectively. H6 protons of T4 and T10 also show NOEs to H6 protons of C5 and C11 residues. Finally, an NOE is also seen between H8 protons of A9 and G8 residues. At the outset, these NOEs helped only in associating certain base proton chemical shifts with particular types of bases, and the assignment of particular positions along the nucleotide sequence has been obtained later, during the process of sequential assignment. The spectrum in Figure 8 also shows clearly that the H6 protons of the two thymines have nearly the same chemical shifts, and since the methyl protons of T4 and T10 are also equivalent, only one cross-peak is seen (ω_1 = 1.39 ppm; ω_2 = 5.93 ppm) between H6 and CH_3 protons of the thymine nucleotides in the COSY spectrum (Figure 5). With the identification of H1′, H2′, and H2″ spin systems from the COSY spectrum and of some base proton connectivities in the NOESY spectrum, sequential assignment could be carried out via H1′ protons and H2′ and H2″ protons as discussed in the previous section. These connectivities in the NOESY spectrum are shown in Figures 9 and 10. A summary of the observed sequential NOE connectivities is shown

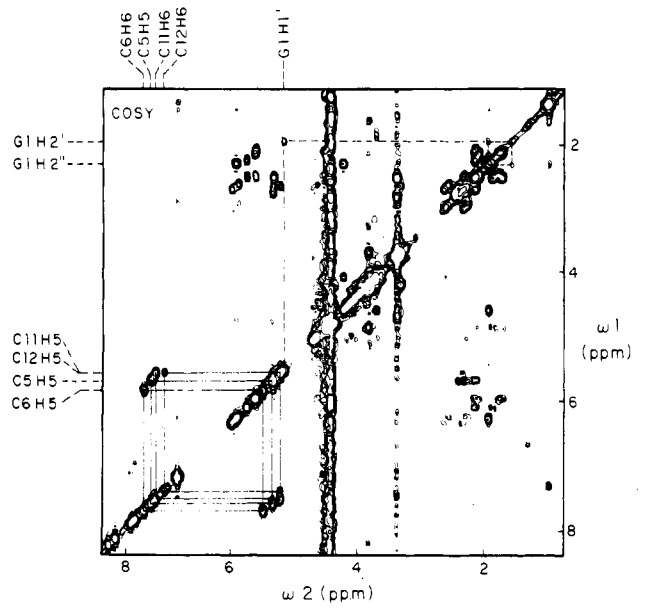

FIGURE 5: 500-MHz COSY spectrum of d-GGATCCGGATCC at 25 °C in D_2O solution at pH 7.2. The time-domain data consists of 2048 data points along the t_2 axis and 400 data points along the t_1 axis. The data were zero-filled to 1024 points along the t_1 axis and then multiplied by sine-squared bell and sine bell window functions along the t_2 and t_1 axes, respectively, prior to respective Fourier transformation. A delay period of 0.005 s was used after each of the 90 pulses in the COSY sequence. Base proton connectivities of the four cytosines are indicated. Only one of the sugar H1′–H2′ (H2″) connectivity patterns is shown (––) in this figure, to avoid overcrowding. The connectivity shown is for the G1 nucleotide, which includes H1′–H2′ and the H2′–H2″ cross peaks. The absence of the H1′–H2″ cross peak at ω_2 = 5.48 ppm and ω_1 = 1.97 ppm has been interpreted as indicative of C3′-endo sugar geometry.

in Figure 11. No connectivity could be established for G1 and C12, while only a weak connectivity from H2′ of C11 to H6 of C12 was observed. The 200-ms NOESY spectrum showed a weak NOE from H6 of C12 to H1′ of the C11 nucleotide. The small intensities of these peaks may be attributed to higher mobility at the ends of the chain. Figures 9 and 10 contain many other NOEs, all of which, except four peaks in Figure 9, are assigned to be intranucleotide NOEs. The four peaks in Figure 9 were attributed to interstrand NOEs between H2 protons of adenine residues to H1′ protons of thymines and to two of the four cytosine residues. This leads to the assignment of H2 base protons of the two adenine nucleotides. These interstrand connectivities are indicated by dotted lines in Figure 9. The observation of interstrand NOEs confirms, in the first place, the sequential assignment and, second, that the molecule is in the double-helical state under the present experimental conditions.

The NOESY spectrum in Figure 7 also contains cross peaks showing connectivities between H1′ protons and H4′ protons of the 12 sugar rings. Since the assignment of H1′ protons is known, the H1′–H4′ cross peaks could be readily assigned to particular nucleotides along the chain (Figure 12). Table I gives a list of the assignments.

Structure of d-GGATCCGGATCC. As is evident from Figure 3, information about the sugar pucker and base orientation (glycosidic torsion angle) can be derived from the relative intensities of the cross peaks between base protons and H2′ and H2″ protons and from the type of internucleotide connectivities. For C3′-endo geometry, H2′ and H2″ protons are always at widely different distances from base H8 and H6 protons throughout the entire range of values and thus would give NOESY cross peaks having very different intensities. On the other hand, for a C2′-endo geometry, the two protons H2′ and H2″ would produce equally intense cross peaks for a χ value of about 150°. On either side of 150°, the two cross peaks will have different intensities. The domain of χ values can be determined by looking at whether the same C2′ proton gives strong NOEs to the base proton of the same nucleotide as well as to that of the next nucleotide. If that is true, then the conformation must belong to the high-anti or syn domain. In summary, the NOESY data can be used to discriminate between anti, high-anti, and syn conformations with respect to the glycosidic dihedral angle, by following the guidelines given below.

(i) For a syn conformation, a strong NOE between base (H8/H6) and the H1′ proton should be observed. At the same time, the NOEs from base to H2′ and H2″ protons will be relatively weak and will have different intensities.

(ii) In the anti conformation, NOEs from base (H8/H6) to H2′ and H2″ protons of the same nucleotide will have different intensities. Also, the proton corresponding to the

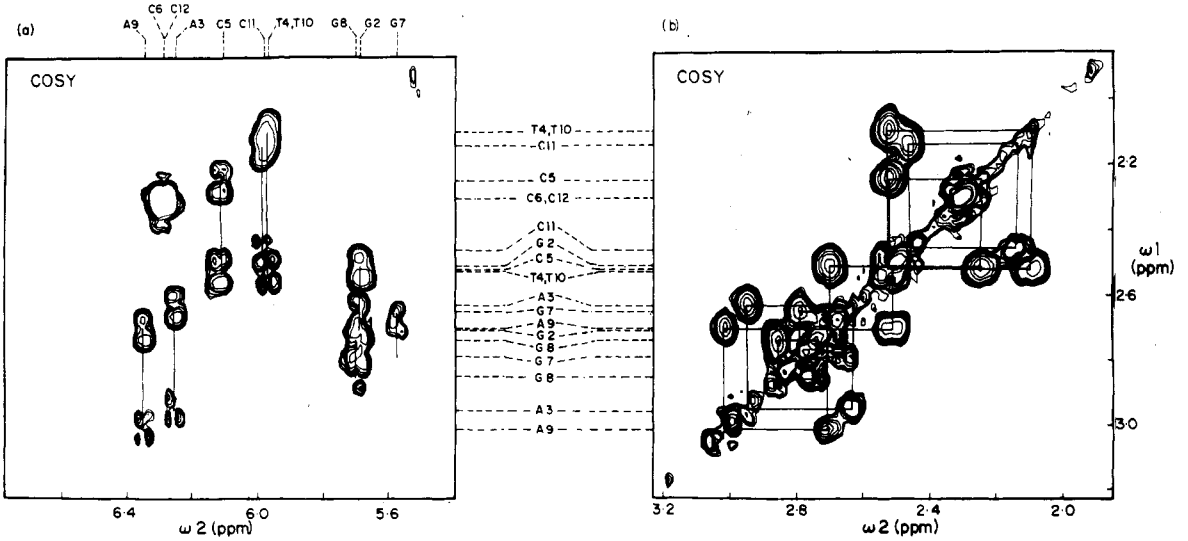

FIGURE 6: Expansions of selected regions of spectrum in Figure 5. Spin system connectivities of the sugar ring for 11 nucleotides are shown. Part a shows H1′–H2′ (H2″) cross peaks, and part b shows H2′–H2″ cross peaks. In (a), peaks belonging to the same H1′ protons are joined by a line. Assignment of H1′ protons is given at the top of part a and that of H2′ and H2″ protons is given in between parts a and b. For G7, only the H1′–H2′ cross peak (at ω_2 = 5.54 ppm and ω_1 = 2.66 ppm) is seen. The absence of the H1′–H2″ cross peak is attributed to C3′-endo sugar geometry. In the case of C6 and C12, the two protons on C2′ are chemically equivalent.

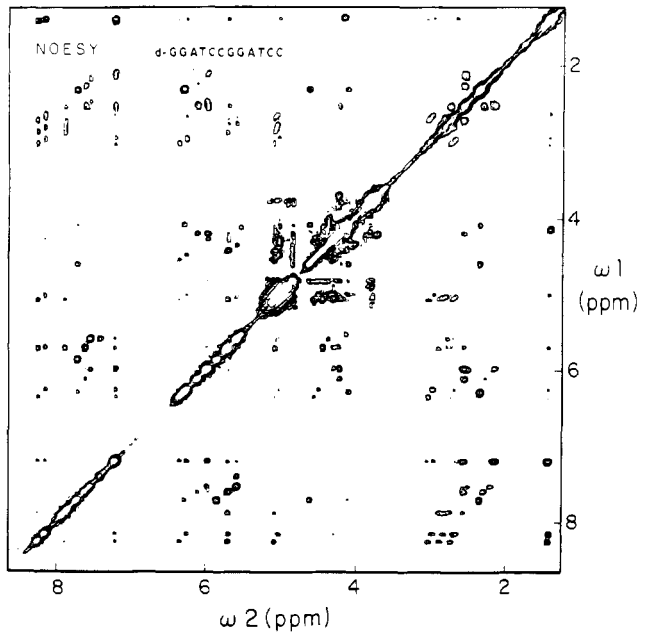

FIGURE 7: Symmetrized 500-MHz NOESY spectrum of d-GGATCCGGATCC recorded under identical conditions as in the COSY spectrum. Mixing time is 300 ms. Time-domain data consisted of 1024 points along the t_2 axis and 450 points along the t_1 axis. The data were manipulated in the same way as for the COSY spectrum except that the data were zero-filled up to 2048 points along the t_2 axis before window multiplication.

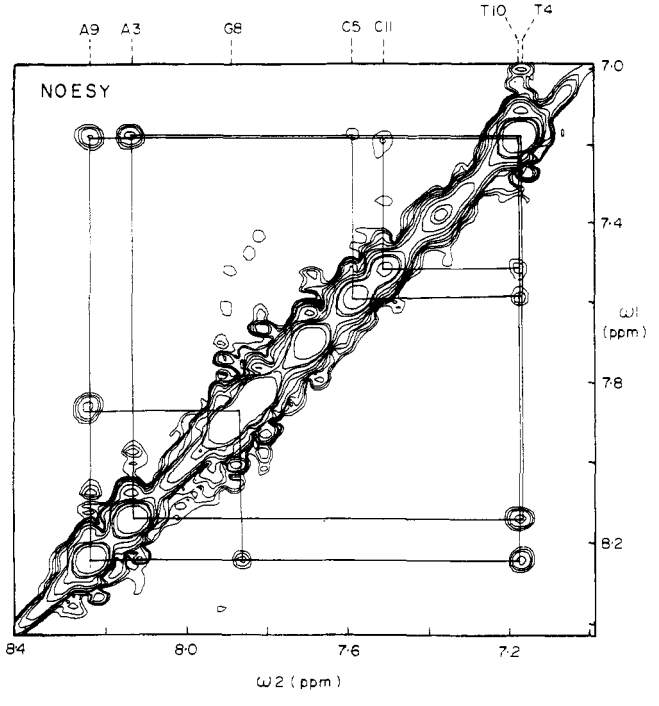

FIGURE 8: Expansion of selected region of unsymmetrized NOESY spectrum with mixing time of 300 ms, showing direct base–base connectivities between adjacent nucleotide units. Base protons (H8/H6) at different chemical shifts have been identified by respective bases at the top of the figure.

weaker NOE will show a strong NOE to the base proton of the next nucleotide.

(iii) In the high-anti conformation, the NOEs from base (H8/H6) to H2′ and H2″ will have similar intensities for C2′-endo sugar geometry and will have different intensities for C3′-endo sugar geometry. Further, the H2″ proton that would show a strong NOE to base proton of the same nucleotide would also show a strong NOE to the base proton of the following nucleotide.

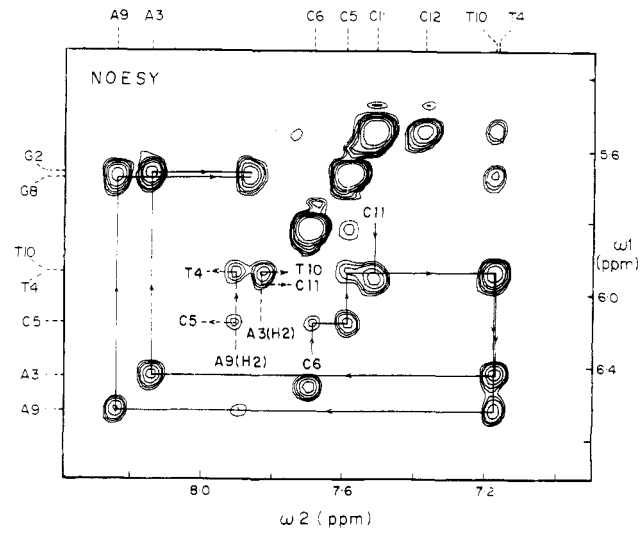

FIGURE 9: Selected region of unsymmetrized NOESY spectrum (mixing time 300 ms) showing sequential connectivities using H8 or H6 proton and H1′ protons. A vertical line to a cross peak indicates an NOE from the base proton of the nth residue to H1′ proton of the $(n-1)$th residue. A horizontal line joins NOE cross peaks from the H1′ proton of the nth residue to its own base proton and to the base proton of the $(n+1)$th residue. NOEs, observed across the two strands in the double helix, are indicated by dotted lines. Identification of H1′ protons is given on the left side of the figure and that of base protons is given at the top of the figure.

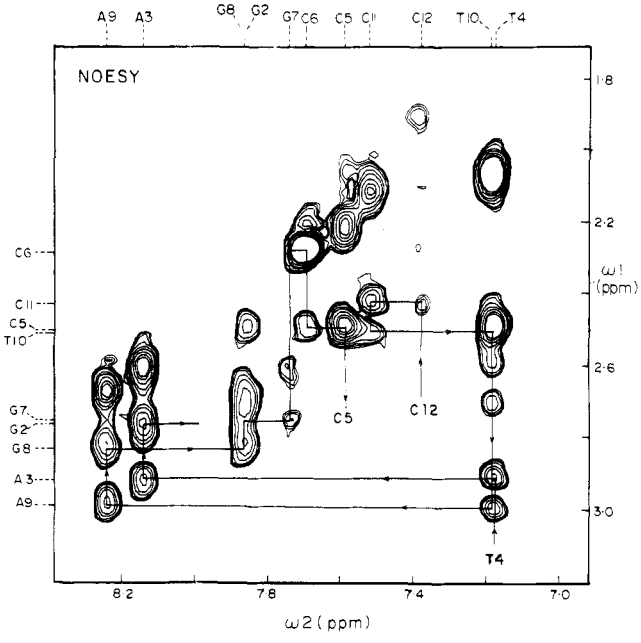

FIGURE 10: Sequential connectivities using the base (H6/H8) proton and H2′/H2″ protons. Vertical and horizontal lines have similar meaning as in Figure 9. Identification of the H2″ proton is given on the left side of the figure, while that of base protons is given at the top of the figure.

Qualitative information about sugar geometry can be obtained from the COSY spectrum by looking at the relative intensities of H1′–H2′ and H1′–H2″ cross peaks. For C2′-endo geometry, both the H1′–H2′ and H1′–H2″ coupling constants are large (between 6 and 10 Hz) and would produce strong cross peaks in the COSY spectrum. On the other hand, for C3′-endo geometry one of the coupling constants is close to zero and the other is about 7–10 Hz (Hosur et al., 1981). Thus in the COSY spectrum, the two H1′–H2′ and H1′–H2″

Table I: Chemical Shifts[a] of Nonexchangeable Protons in d-GGATCCGGATCC

base	H8/H6	H1′	H2′	H2″	H4′	H5/H6/CH_3
G	7.39	5.48	1.97	2.31	4.09	
G	7.87	5.65	2.50	2.70	4.26[b]	
A	8.15	6.22	2.62	2.94	4.46	7.84
T	7.18	5.93	2.07	2.52	4.18	1.39
C	7.59	6.07	2.23	2.52	4.18	5.66
C	7.69	6.26	2.28		4.06	5.81
G	7.74	5.54	2.66	2.79	4.33	
G	7.87	5.66	2.74	2.87	4.40[b]	
A	8.25	6.23	2.69	3.01	4.52	7.90
T	7.19	5.93	2.07	2.52	4.25	1.40
C	7.52	5.95	2.12	2.44	4.17	5.54
C	7.39	6.26	2.28		4.06	5.54

[a] Chemical shifts are in ppm with respect to TSP. [b] Assignment is based on analogy of relative chemical shifts of H2′ and H2″ protons. These protons of G2 are upfield compared to those of G8.

Table II: Structural Information from Relative Intensities in Selected Regions of COSY and NOESY Spectra[a]

	NOESY, NOE from base to			COSY, J coupling with H1′		
base	H2′	H2″	inference	H2″	H2″	inference, sugar pucker
G	w			s		3′-endo
G	w	o[b]	anti	s	s	2′-endo
A	s	s[b]	high-anti	s	s	2′-endo
T	w			s	s	2′-endo
C	s	s[b]	high-anti	s	s	2′-endo
C	w	s[b]	high-anti	s		2′-endo[c]
G	w	w[b]	anti	s		3′-endo
G	o	o[b]	anti	s	s	2′-endo
A	s	s[b]	high-anti	s	s	2′-endo
T	s	s[b]	high-anti	s	s	2′-endo
C	s	w[b]	anti	s	s	2′-endo
C				s		2′-endo

[a] w = weak; s = strong; o = overlapping. [b] Used in sequential assignment. [c] Equivalence of H2′ and H2″ protons in C6 and C12 did not allow establishment of sugar geometry from COSY spectrum. The sugar conformations in these are thus determined from 2D J-resolved experiment.

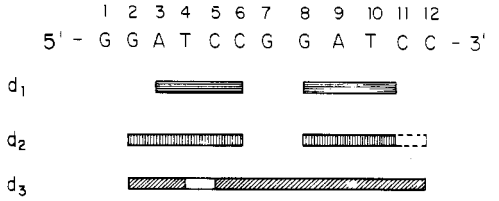

FIGURE 11: Summary of all sequential connectivities. d_1 refers to base–base connectivity, d_2 refers to base–H2′ connectivity, and d_3 refers to base–H2″ connectivity. An open box in the d_3 stretch indicates that connectivity is not unambiguous due to overlap of peaks. The dotted line box in the d_2 stretch indicates that weak connectivity was seen in the 200-ms spectrum.

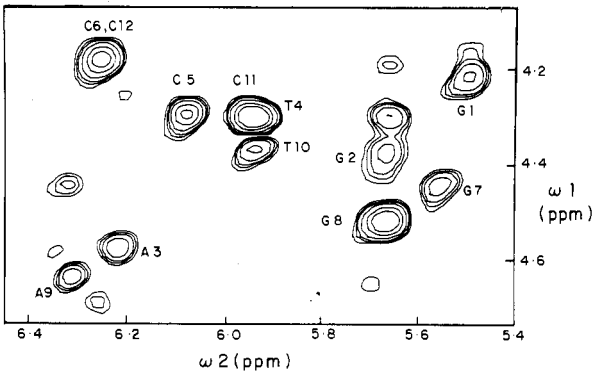

FIGURE 12: Spectral region of NOESY spectrum (mixing time 300 ms) showing NOEs between H1′ protons and H4′ protons of different nucleotides. The peaks have been identified by respective nucleotide labels.

cross peaks would have very different intensities for a C3′-endo geometry. The above conclusions are actually valid for a wide range of related sugar conformations that are grouped into the two categories N and S. The N family is centered around C3′-endo geometry while the S family is centered around C2′-endo geometry. Table II gives a qualitative calibration of NOESY and COSY cross-peak intensities and inferences derived therefrom about sugar geometry and glycosidic dihedral angles in d-GGATCCGGATCC. The conclusions about sugar geometry are also supported by direct measurements of coupling constants by 2D J-resolved experiments. The detailed results will be published separately.

The fact the sequential assignment could only be completed by using the strategy designed for right-handed DNA suggests that the molecule adopts a right-handed conformation. We note that in principle this sequence is capable of hairpin formation of the individual single strands (e.g., with GGAT paired to ATCC at the 3′ end and CCGG looped out) or the formation of end to end aggregation with the first half of one chain paired to the second half of another chain [cf. Scheffler et al. (1968, 1970)]. Both these possibilities are clearly excluded by the spectra, and in addition, there is an energetic disadvantage of losing two G-C pairs in the hairpin structure. For example, the observed NOEs from the H2 proton of A9 to H1′ protons of T4 and C5 eliminate the possibility of hairpin loop. Similarly, end to end aggregation resulting in a higher molecular weight species is excluded from the fact that the NMR spectrum consists of sharp lines. In the event of aggregation of a few molecules in linear fashion, the unmatched nucleotides at the end would give rise to additional cross peaks in the NOESY spectrum. The fact that all cross peaks have been assigned in the NOESY spectrum is consistent with the existence of a single double helix in aqueous solution. In principle, a closed circular double helix could be formed by pairing of sticky ends of an end to end aggregate, though the size of the aggregate would presumably need to be rather large

(e.g., >200 bp) for ring closure to begin to be significant [cf., Shore et al. (1981)]. The spectroscopic results are, in any event, not consistent with such a cyclic structure. Because of the higher molecular weight, the line widths would be much greater than those observed. The chemical shifts of H8 in G1 and G7 show a greater difference than would be expected in a cyclic structure. And finally, we do not see NOESY cross peaks between G1 and C12 as expected for such a structure. Thus, our results on d-GGATCCGGATCC are consistent with a predominantly B type of conformation of the molecule.

The strategies discussed above are quite general and do not require prior knowledge about the conformation of the oligonucleotide. Although no crystal structure data on d-GGATCCGGATCC are presently available, almost complete assignment and qualitative information about conformation could be obtained from the NMR data alone. The structural diversity in nucleic acids, unlike that in proteins, is limited, and therefore, one does not expect to see long-range NOEs. Thus, sequential assignment itself gives substantial information about the structure of the DNA fragment.

Examination of chemical shifts in Table II reveals that the information about base type is reflected in the chemical shifts of the H1′ protons in d-GGATCCGGATCC. H1′ protons of guanines appear upfield followed by those of thymines, cytosines, and adenines. If this trend is general, then it will provide very useful guidelines for resonance assignment.

It is interesting to note from the qualitative conclusions in Table II that G1 and G7 sugars have different geometries compared to the rest of the nucleotides. The reason for the locally different C3′-endo conformation at G1 and G7 is not clear. Possibly by assuming this geometry the 5′-G residues can assume a more favorable stacking interaction in the duplex. The structural features of these sugars may facilitate recognition for the endonuclease enzyme *Bam*HI, which cleaves the molecule between the G residues. The variation of the structure at the G1 and the G7 site may also result in a higher flexibility of the DNA duplex at these positions. In fact, it is observed that the H8 proton of G7 shows weaker NOEs to its own H2′/H2″ protons as compared to the corresponding cross peaks from other nucleotide units.

In addition, Table II also depicts certain diversity in the glycosidic dihedral angle at various nucleotide positions. These variations show the influence of sequence on the conformation of the DNA fragments.

ACKNOWLEDGMENTS

The help provided by the 500-MHz FT-NMR National Facility at the Tata Institute of Fundamental Research is gratefully acknowledged.

Registry No. *Bam*HI, 81295-09-2; d-GGATCCGGATCC, 98760-09-9; deoxyribose, 533-67-5.

REFERENCES

Arseniev, A. S., Wider, G., Joubert, F. J., & Wüthrich, K. (1982) *J. Mol. Biol. 159*, 323–351.

Aue, P., Bartholdl, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540–5541.

Feigon, J., Wang, A. H. J., Van der Marel, G. A., van Boom, J. H., & Rich, A. (1984) *Nucleic Acids Res. 12*, 1243–1263.

Frechet, D., Cheng, D. M., Kan, L. S., & Tso, P. O. P. (1983) *Biochemistry 22*, 5194–5200.

Giessner-Prettre, C., Pullman, B., Tran-Dinh, S., Neumann, J. M., Huynh-Dinh, T., & Egolon, J. (1984) *Nucleic Acids Res. 12*, 3271–3281.

Govil, G., & Hosur, R. V. (1980) in *Conformation of Biological Molecules: New Results From NMR*, Springer-Verlag, Heidelberg.

Gurevich, A. Z., Barsukov, I. L., Arseniev, A. S., & Bystrov, V. F. (1984) *J. Magn. Reson. 56*, 471–478.

Hanstock, C. C., & Lown, J. W. (1984) *J. Magn. Reson. 58*, 167–172.

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

Hasnoot, C. A. G., Van de Ven, J. M., & Hilbers, C. W. (1984) *J. Magn. Reson. 56*, 343–349.

Hosur, R. V., Govil, G., Hosur, M. V., & Viswamitra, M. A. (1981) *J. Mol. Struct. 72*, 261–267.

Hosur, R. V., Wider, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 130*, 497–508.

Jeener, I. (1971) AMPERE International Summer School, Basko Polje, Yugoslavia.

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

Keller, R. M., Baumann, R., Hunziker-kwik, E. H., Joubert, F. J., & Wüthrich, K. (1983) *J. Mol. Biol. 163*, 623–646.

Kumar, A., Wagner, G., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 96*, 1156–1163.

Kumar, A., Hosur, R. V., & Chadrasekhar, K. (1984) *J. Magn. Reson. 60*, 143–148.

Lown, J. W., Hanstock, C. C., Bleackly, R. C., Imbach, J. L., Rayner, B., & Vasseur, J. J. (1984) *Nucleic Acids Res. 12*, 2519–2533.

Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.

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

Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321–334.

Ohlendorf, D. H., & Matthews, B. W. (1983) *Annu. Rev. Biophys. Bioeng. 12*, 259–284.

Pardi, A., Walker, R., Rapoport, H., Wider, G., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 1652–1653.

Patel, D. J., Kozlowski, S. A., Nordheim, A., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1413–1417.

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1968) *J. Mol. Biol. 36*, 291–304.

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1970) *J. Mol. Biol. 48*, 145–171.

Shore, D., Langowski, J., & Baldwin, R. L. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4833–4837.

Strop, P., Wider, G., & Wüthrich, K. (1983) *J. Mol. Biol. 166*, 641–665.

Tan, Z.-K., Ikuta, S., Huang, T., Dugaiczyk, A., & Itakura, K. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 383–391.

Wagner, G. (1983) *J. Magn. Reson. 55*, 151–156.

Wagner, G. (1984) *J. Magn. Reson. 57*, 497–505.

Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–366.

Wagner, G., Kumar, A., & Wüthrich, K. (1981) *Eur. J. Biochem. 114*, 375–384.

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

Wang, A. H. J., Quigley, G. J., Kolpak, R. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) *Science (Washington, D.C.) 211*, 171–176.
Westerink, H. P., van der Marel, G. A., van Boom, J. H., & Haasnoot, C. A. G. (1984) *Nucleic Acids Res. 12*, 4323–4338.
Wider, G., Lee, K. H., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 367–388.
Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.
Wider, G., Macura, S., Kumar, A., Ernst, R. R., & Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207–234.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
Zuiderweg, E. R. P., Kaptein, R., & Wuthrich, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837–5841.

Proton Magnetic Resonance Studies of the States of Ionization of Histidines in Native and Modified Subtilisins†

Frank Jordan,*,‡ Laszlo Polgar,§ and Guillermo Tous‡

Department of Chemistry, Rutgers University, Newark, New Jersey 07102, and Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest H-1502, Hungary

Received January 30, 1985

ABSTRACT: A technique was developed to exchange the backbone –N–H protons in D_2O in the native subtilisins Carlsberg and BPN (Novo) that resulted in clearly resolved proton resonances in the aromatic region of the nuclear magnetic resonance spectrum. pH titration curves for four of the five histidine C2–H resonances in subtilisin Carlsberg and five of the six in subtilisin BPN between 7.5 and 8.8 ppm downfield from 4,4-dimethyl-4-silapentane-1-sulfonic acid sodium salt provided microscopic pK_a's between 6.3 and 7.2 for both sources of the enzyme at ambient (~22 °C) probe temperature. A resonance that titrated with a pK_{app} of 7.35 ± 0.05 was observed in the ^1H spectra only of the diisopropylphosphoryl derivatives of the subtilisins from both sources. The ^{31}P NMR pH titration of the same preparations under identical conditions of solvent (D_2O) and temperature gave a pK_{app} = 7.40 ± 0.05 of the single titratable resonance. Both observations must pertain to His-64 at the active center. A resonance smaller than the others and titrating with a pK_{app} of 7.2 could also be observed in the native enzymes. This resonance was assigned to the catalytic center histidine since its pK corresponded to that derived from kinetic studies. No major perturbations in the chemical shifts or the pK's derived from the pH dependence of the observed resonances were apparent in the presence of saturating concentrations of the two putative transition-state analogues phenylboronic acid and bis[3,5-(trifluoromethyl)phenyl]boronic acid and in monoisopropylphosphorylsubtilisin. It can be concluded that the C2–H resonance corresponding to His-64 in native subtilisins is difficult to observe perhaps on account of the limited mobility of this side chain compared to its mobility in the diisopropylphosphoryl derivative.

A number of nuclear magnetic resonance (NMR) approaches have been applied to serine proteases [see review in Steitz & Shulman (1982)]. An early study by Robillard & Shulman (1972, 1974a,b) on chymotrypsin and other serine proteases found a ^1H resonance at 15–18 ppm downfield from 4,4-dimethyl-4-silapentane-1-sulfonic acid sodium salt (DSS)[1] that was attributed to the hydrogen located between the Asp and His residues at the catalytic center, residing very likely on the nitrogen (Bachovchin & Roberts, 1978; Kossiakoff & Spencer, 1981). Surprisingly, no such resonance was found in the spectra of native subtilisins (Jordan & Polgar, 1981; Jordan et al., 1982). The resonance, however, was observed when there was located a second negative charge in the catalytic center so as to create an $Asp^-His^+X^-$ charge distribution (where X^- represents a thiolate anion produced by conversion of the serine enzyme to its chemically mutated thiol form, or of a covalently attached putative transition-state analogue phenylboronic acid on the serine enzyme, or even a noncovalent interaction between the serine enzyme and the reversible competitive inhibitor *N*-acetyl-L-tryptophan). To explain these results, it was suggested that the His imidazole is more mobile in the native enzyme than in the transition-state analogue-like –+– charge distribution state. The resonance corresponding to the N–H proton located between Asp and His is subject to exchange broadening, and apparently the rate of exchange with solvent slows down by placing a negative charge on both sides of the histidinium ion. An independent technique to enable observation of the His imidazoles directly is to monitor the resonance of the C2–H in the aromatic region of the ^1H spectrum. Extensive results were published by Markley and co-workers on this spectral region of trypsin, chymotrypsin, and α-lytic protease (Markley & Porubcan, 1976; Markley & Ibanez, 1978; Markley, 1979). Since exchange of the C2–H with solvent is very slow ($t_{1/2}$ of several days), these resonances are not subject to exchange broadening with solvent and

† Work at Rutgers was supported by a Charles and Johanna Busch Biomedical Grant and an unrestricted grant from Hoffmann-La Roche Inc., Nutley, NJ.
‡ Rutgers University.
§ Hungarian Academy of Sciences.

[1] Abbreviations: DSS, 4,4-dimethyl-4-silapentane-1-sulfonic acid sodium salt; MES, 2-(*N*-morpholino)ethanesulfonic acid; Tris, tris(hydroxymethyl)aminomethane; DIFP, diisopropyl fluorophosphate; PMSF, phenylmethanesulfonyl fluoride; EDTA, ethylenediaminetetraacetic acid.

0006-2960/85/0424-7711$01.50/0