Deese, A. J., Dratz, E. A., Dalquist, F. W., & Paddy, M. R. (1981) *Biochemistry 20*, 6420–6427.
Degli Esposti, M., Ferri, E., & Lenaz, G. (1981) *Ital. J. Biochem. 30*, 437–452.
Ernster, L., Lee, I.-Y., Norling, B., & Persson, B. (1969) *Eur. J. Biochem. 9*, 299–310.
Fato, R., Battino, M., Degli Esposti, M., Parenti Castelli, G., & Lenaz, G. (1986) *Biochemistry 25*, 3378–3390.
Gerstein, B. C., & Dybowski, C. R. (1985) *Transient Techniques in NMR of Solids*, Academic, Orlando, FL.
Green, D. E. (1962) *Comp. Biochem. Physiol. 4*, 81–122.
Hauska, G., & Hurt, E. (1982) in *Functions of Quinones in Energy Conserving System* (Trumpower, B., Ed.) pp 87–110, Academic, New York.
Katsikas, H., & Quinn, P. J. (1981) *FEBS Lett. 133*, 230–234.
Kingsley, P. B., & Feigensen, G. (1981) *Biochim. Biophys. Acta 635*, 602–618.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Mackler, B. (1967) *Methods Enzymol. 10*, 261–263.
Mehring, M. (1983) *High Resolution NMR in Solids*, Springer, Berlin.
Mitchell, P. (1976) *J. Theor. Biol. 62*, 327–367.
Nguyen-Dinh-Nguyen & Stenhagen, E. (1966) *Acta Chem. Scand. 20*, 1423–1424.
Oldfield, E., Meadows, M., Rice, D., & Jacobs, R. (1978) *Biochemistry 17*, 2727–2740.
Ondarroa, M., & Quinn, P. (1986) *Eur. J. Biochem. 155*, 353–361.
Peterson, N. O., & Chan, S. I. (1977) *Biochemistry 16*, 2657–2667.
Raison, J. K., & Lyons, J. M. (1970) *Plant Physiol. 45*, 382–385.
Seelig, A., & Seelig, J. (1974) *Biochemistry 13*, 4839–4845.
Shunk, C. H., Wolf, D. E., McPherson, J. F., Linn, B. O., & Folkers, K. (1960) *J. Am. Chem. Soc. 82*, 5914–5918.
Silbert, L. S., & Swern, D. (1959) *J. Am. Chem. Soc. 81*, 2364–2367.
Singleton, W. S., Gray, M. S., Browen, M. L., & White, J. L. (1965) *J. Am. Oil Chem. Soc. 42*, 53–56.
Stidham, M. A., MacIntosh, T. J., & Siedow, J. N. (1984) *Biochim. Biophys. Acta 767*, 423–431.
Strenk, L. M., Westerman, P. W., & Doane, J. W. (1985) *Biophys. J. 48*, 765–773.
Trumpower, B. L. (1981) *J. Bioenerg. Biomembr. 13*, 1–24.
Trumpower, B. L., Ed. (1982) *Function of Quinones in Energy Conserving Systems*, Academic, New York.
Ulrich, E. L., Girvan, M. E., Cramer, W. A., & Markley, J. L. (1985) *Biochemistry 24*, 2501–2509.
Wan, Y. P., & Folkers, K. (1978) *Methods Enzymol. 53*, 591–599.
Westerman, P. W., Keniry, M. A., Robertson, R. N., & Cornell, B. A. (1986) *Biophys. J. 49*, 96a.
Wilczynski, J. J., Daves, G. D., & Folkers, K. (1968) *J. Am. Chem. Soc. 90*, 5593–5598.
Yu, C.-A., Gu, L., Lin, Y., & Yu, L. (1985) *Biochemistry 24*, 3897–3902.

DNA Structure in Which an Adenine–Cytosine Mismatch Pair Forms an Integral Part of the Double Helix†

Mukti H. Sarma, Goutam Gupta, and Ramaswamy H. Sarma*
Institute of Biomolecular Stereodynamics, State University of New York at Albany, Albany, New York 12222
Rolf Bald, Ute Engelke, Shiao Li Oei, Reinhard Gessner, and Volker A. Erdmann
Institut für Biochemie, Fachbereich Chemie, Freie Universität Berlin, 1000 Berlin 33 (Dahlem), FRG
Received April 13, 1987; Revised Manuscript Received July 1, 1987

ABSTRACT: Extensive studies using one- and two-dimensional ^{1}H NMR at 500 MHz revealed that the oligonucleotide d(CGCCGCAGC) in solution at 5 °C forms a double helix under conditions of high salt (500 mM in NaCl, 1 mM sodium phosphate), low pH (pH 4.5), and high DNA concentration (4 mM in duplex). The presence of very strong nuclear Overhauser effects (NOEs) from base H8/H6 to sugar H2′,H2″ and the absence of NOE from base H8/H6 to sugar H3′ suggested that the oligomer under these solution conditions forms a right-handed B-DNA double helix. The following lines of experimental evidence were used to conclude that C4 and A7 form an integral part of the duplex: (i) the presence of a NOESY cross-peak involving H8 of A7 and H8 of G8, (ii) the presence of a two-dimensional NOE (NOESY) cross-peak between H6 of C3 and H6 of C4, (iii) base protons belonging to C4 and A7 forming a part of the H8/H6---H1′ cross-connectivity route, and (iv) the pattern of H8/H6---H2′,H2″ NOESY cross-connectivity based upon a B-DNA model requiring that both C4 and A7 form an integral part of the duplex. The possibility of an A-C pair involving H bonds was also examined. Two possible structural models of the duplex at pH 4.5 are proposed: in one model A-C pairing involves two H bonds, and in the other A-C pairing involves a single H bond.

Until very recently, structural studies on short DNA duplexes were restricted to self-complementary DNA sequences. By single-crystal and solution studies it has been shown that self-complementary DNA sequences can exhibit a wide variety of secondary structures while still retaining the Watson–Crick base-pairing schemes. Structural morphologies of the A-, B-, and Z-DNA have been understood in terms of the nucleotide geometries (Conner et al., 1982; Drew et al., 1980; Shakked

† This research is supported by a grant from the National Institutes of Health (GM29787), the Deutsche Forschungsgemeinschaft (DFG-SFBg-B5), the Fonds der Chemischen Industrie e.V., and a contract from the National Foundation for Cancer Research.

0006-2960/87/0426-7707$01.50/0

et al., 1981; Wang et al., 1979, 1982; Wing et al., 1980). Recently it has been demonstrated by single-crystal (Ho et al., 1985; Kennard, 1986) and solution studies (Kan et al., 1983; Pardi et al., 1982; Patel et al., 1984a,b) that incorporation of wobble G-T/A-G pairs in place of the regular Watson–Crick G-C pair does not significantly alter the morphologies of A-, B-, and Z-DNA. Very recently non-Watson–Crick A-C pairs with two hydrogen bonds have been shown to exist in the single-crystal structure of a DNA double helix (Hunter et al., 1986). Solution studies of oligonucleotides have so far failed to establish the nature of A-C pairing (Patel et al., 1984a,b). In this paper, we examine the possibility if A and C facing each other from opposite strands can really be an integral part of the DNA double helix under solution conditions.

Materials and Methods

The deoxynonanucleotide d(CGCCGCAGC) has been synthesized by the phosphoramidite method (Matteuci & Caruthers, 1981) on a DNA synthesizer (Applied Biosystems Model 380 A). The product was purified on a 4 × 250 mm column of ODS-Hypersil (5 μm) with a linear gradient of 0–80% acetonitrile in 0.1 M triethylammonium acetate (pH 7.0).

The two-dimensional correlated spectroscopy (COSY) spectrum of d(CGCCGCAGC) in D_2O at 15 °C (in 500 mM NaCl, 1 mM sodium phosphate buffer, pH 4.5) was recorded with the pulse sequence $(RD–90°–t_1–90°–Acq)_n$. For each of 512 t_1 values, 64 transients were recorded with 1024 data points with relaxation delay RD = 1.0 s; HDO was presaturated. The two-dimensional nuclear Overhauser effect (NOESY) spectrum of d(CGCCGCAGC) in D_2O was collected under the same solution conditions of the COSY experiment. For NOESY, the pulse sequence was $(RD–90°–t_1–90°–\tau_m–90°–Acq)_n$ with mixing time τ_m = 200 ms, other parameters being the same as in the COSY experiment. Both COSY and NOESY spectra were recorded in the pure absorption mode (States et al., 1982). Time domain data of COSY and NOESY experiments were processed on a VAX-11/750 computer; free induction decays were weighed with an exponential multiplication factor of 10 Hz before Fourier transformation.

System under Study

Two of the likely structural forms that will result from the self-association of the non-self-complementary DNA d(CGCCGCAGC) are shown in Figure 1. In one type, A7 and C4 facing each other on opposite strands loop out (Figure 1A) while in the other they form an integral part of the helix (Figure 1B). In both arrangements the strands are antiparallel and there is an internal 2-fold between G5 and C6 (Figure 1). Notice that in both arrangements the regular Watson–Crick H bonds are between G's and C's, i.e., between G2 and C9, C3 and G8, G5 and C6, etc. The distinguishing features of the two structural arrangements are the spatial relations of C4 and its neighbors C3 and G5 and of A7 with its neighbors C6 and G8. We have employed 1D and 2D NMR measurements at 500 MHz to elucidate the structural arrangement of d(CGCCGCAGC) in the duplex state at pH 4.5. The reason for selecting pH 4.5 is explained later. The unusual sequence d(CGCCGCAGC) is selected to extend oligonucleotide studies to possible natural genomic sequences. The above sequence is the DNA analogue of helix II of 5S RNA from *Thermus thermophilus* (Digweed et al., 1986).

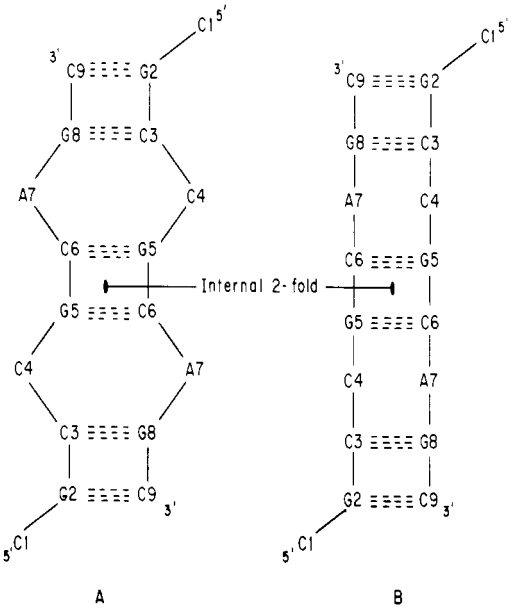


FIGURE 1: Two possible structural arrangements of d(CGCCGCAGC) duplex; in one arrangement (A) C4 and A7 loop out while in the other (B) C4 and A7 are an integral part of the double-helix.

Comments on the pH Used in This Study

A pH of 4.5 along with low temperature (5 °C), high salt (500 mM NaCl), and high concentration of the oligonucleotide (4 mM in duplex) was selected to promote the formation of the double helix. The selection of low pH was motivated by the observation of Hunter et al. (1986) that in crystals of oligomers with AC mismatches the adenine N1 was protonated, even though the crystallization was carried out at neutral pH. There are several other observations in which, when crystallization was carried out at neutral pH, the crystal structure solved contained protonated species. For example, Sundaralingam and co-workers (Sabat et al., 1985; Westhof et al., 1980; Westhof & Sundaralingam, 1985) have observed adenine N1 protonated ATP and a parallel CpA duplex in which cytosine was hemiprotonated. Coll et al. (1987) reported self-paired d(CpG) in which one of the cytosines was protonated. Wang et al. (1986) reported a crystal structure of Hoogsteen GC paired oligonucleotide in which cytosine N3 was protonated.

These observations clearly suggest that even under neutral pH conditions the tiny fraction of the protonated species that is present in the solution nucleates crystallization of the protonated species. One may prima facie dismiss the biological relevance of above protonated structures as well as the structure in this paper because most of the biological systems operate under neutral-pH conditions. However, the observation that crystallization even under neutral conditions results in protonated structures clearly indicates that even at neutral pH a small fraction of the protonated species is present in solution and one cannot dismiss a role for them.

In order to create a condition close to crystal studies, in NMR research one has to adjust the solution conditions. We found that lowering the pH to 4.5 protonated adenine N1, thus facilitating the formation of double hydrogen-bonded A-C pairs [as has been reported by Hunter et al. (1986) in crystal structural studies] as opposed to single hydrogen-bonded A-C pairs involving neutral A (see top of Figure 7, later).

Our own studies at neutral pH under high concentration and high salt clearly indicated that A and C from the opposing strands are not an integral part of the helix and the confor-

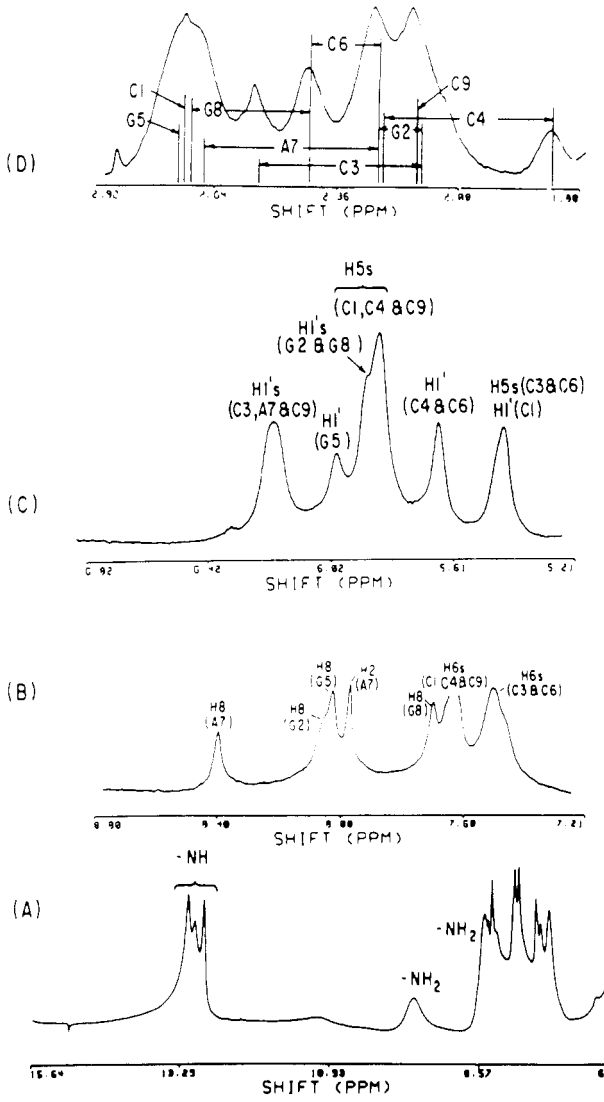


FIGURE 2: The 500-MHz ^1H NMR spectra of d(CGCCGCAGC) in the duplex state (temperature 5 °C, concentration of DNA 4 mM in duplex in 500 mM NaCl, 1 mM sodium phosphate buffer, pH 4.5). Chemical shift values (ppm) are with respect to the internal standard TSP. For identification and assignment of signals, see Figures 3–5 and Table I. (A) Spectrum recorded in H_2O by use of a time-shared long pulse sequence with a notch filter at 3750 Hz; number of scans was 14000. The region of H-bonded exchangeable proton signals involved in Watson–Crick G-C pairs is indicated. (B) Region of the spectrum recorded in D_2O showing H8's of G's, H2 of A, and H6's of C's. (C) Region of the spectrum in D_2O showing the H1''s and H5's of C's. (D) Spectrum in D_2O showing the H2'/H2'' region.

mation is significantly different from that at pH 4.5. At the time this paper was being written, we had not been able to completely solve the complex NOESY data at neutral pH. A referee has raised the question about the structure at pH 5.5. We have not collected detailed data at pH 5.5; but inspection of the NMR titration profiles indicate a mixture of conformations at pH 5.5.

Assignments of Protons Belonging to the Individual Nine Nucleotides of Duplex d(CGCCGCAGC) in Solution

Figure 2A shows the 500-MHz ^1H NMR spectrum of d-(CGCCGCAGC) in H_2O at a low temperature (5 °C), in high salt (500 mM NaCl, 1 mM sodium phosphate buffer, pH 4.5), and at high concentration of the oligonucleotide (4 mM in duplex)—conditions that highly facilitate the duplex formation. The water spectrum shows the exchangeable proton signals,

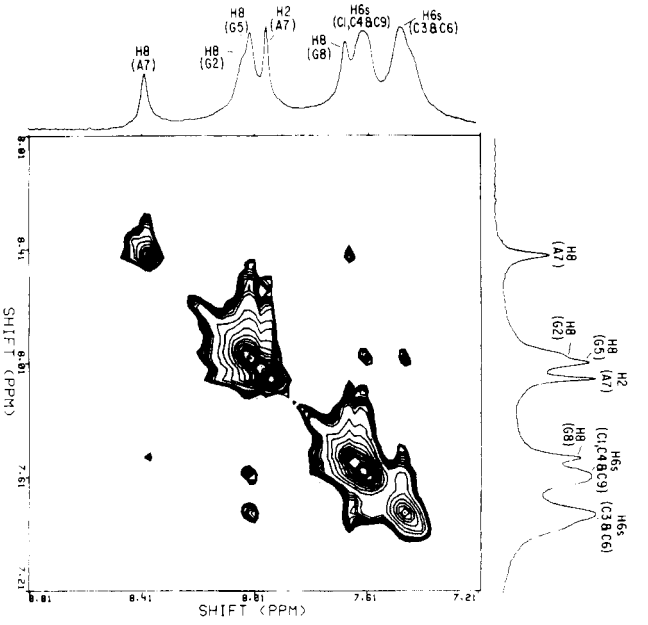


FIGURE 3: Cross section of the NOESY spectrum showing the cross-peaks between the base protons H8/H6. Following the logic of the text, the assignments are as follows: 8.36 (H8 of A7), 8.02 (H8 of G2), 7.99 (H8 of G5), 7.92 (H2 of A7), 7.65 (H8 of G8), 7.59 (H6's of C1, C4, and C9), and 7.43 (H6's of C3 and C6). See text for a detailed explanation.

expected to appear in Watson–Crick-type G-C pairs. The NH signals are observed within 13.2–12.9 ppm belonging to three G-C pairs, viz., G2-C9, C3-G8, and C5-G6, and perhaps $N1^+$-H of A7 if the protonated A7 involved in base pairing (see above); also present are the H-bonded $-NH_2$ signals at 9.8 and 8.5 ppm (McConnell, 1984). These sets of signals disappear in a spectrum of the DNA in D_2O under the same conditions, indicating that they occur as a result of base pairing in the $d(CGCCGCAGC)_2$ duplex. Figure 2B–D shows other region of the NMR spectra in D_2O for the same sample under conditions identical with those in Figure 2A. Figure 2B shows the regions belonging to base protons H8's of G and A, H2 of A, and H6's of C; Figure 2C shows the regions belonging to H1''s and H5's of C, and Figure 2D shows the H2'/H2'' region.

Base Protons in Figure 2B

H6's belonging to C's were identified by a COSY experiment of $d(CGCCGCAGC)_2$ in D_2O; two cross-peaks were observed: one between the signals at 7.59 (H6) and 5.90 (H5) ppm and the other between the signals at 7.43 (H6) and 5.43 (H5) ppm. This suggested that H6's belonging to five C's were located at 7.59 and 7.43 ppm. Inspection of the cross-peak intensity indicated that the COSY cross-peak at (7.59, 5.90) is more intense than the one at (7.43, 5.43). On the basis of this, we have concluded that three H6's are located at 7.59 ppm and two at 7.43 ppm. The other five signals (at 8.36, 8.02, 7.79, 7.92, and 7.65 ppm) should then belong to G's and A. In order to obtain the sequential assignment of the base protons, NOESY measurement with a mixing time $\tau_m = 200$ ms was conducted on $d(CGCCGCAGC)_2$ duplex (in D_2O at 5 °C, 4 mM in DNA duplex in 500 mM NaCl, 1 mM phosphate buffer, pH 4.5). Figure 3 shows a cross section of the NOESY spectrum in which cross-peaks between base protons are depicted.

In a B-DNA duplex, the H8/H6 of i and $i + 1$ residues are about 4.5 Å apart. However, H8(i) and H8($i + 1$) can show NOESY cross-peak for a moderate mixing time τ_m because

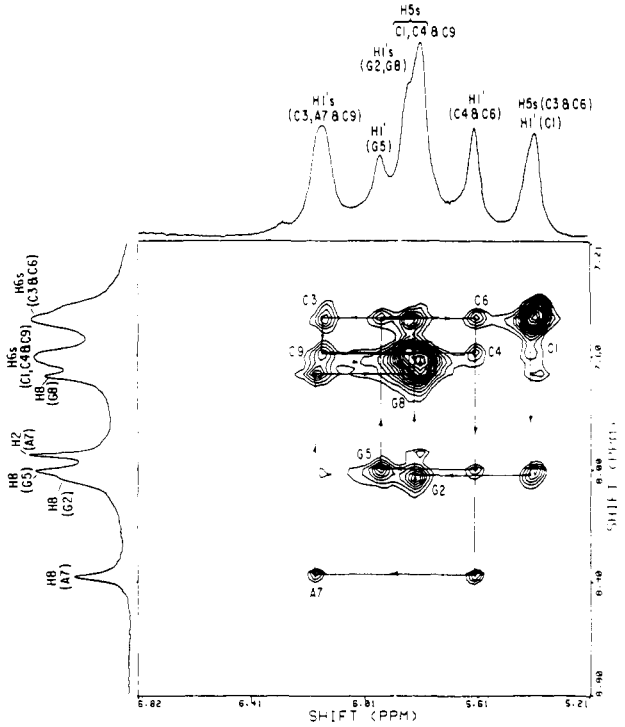


FIGURE 4: Cross section of the NOESY spectrum showing the connectivity among H8/H6's and sugar H1''s. Notice that H5's of C's overlap with H1''s. H5's of C1, C4, and C9 are located at 5.90 ppm, and H5's of C3 and C6 are at 5.45 ppm (directly obtained from the COSY experiment). The following should be noted for clarification. (i) Terminal unpaired C1 is involved in two cross-peaks: one H6–C1---H1'–C1 (weak) and the other H1'–C1---G8-G2 (overlapping with the cross-peak H8–G2---H5–C3). (ii) H6---H5 cross-peaks of C showing up in this cross section do not form a part of the base H8/H6 and sugar H1' connectivity route. (iii) The cross-peak between H6 of C9 and its H1' and the cross-peak between H1' of C3 and H6 of C4 occur at the same ppm value, so much so that the horizontal axis connecting C9 and C4 contains partially overlapping two routes. (iv) Routes from H1' of G2 to H6 of C3 and the route from H1' of G5 to H6 of C6 fall on the same base axis because C3 and C6 are located at the same ppm value. Heavy lines indicate when routes overlap.

of the presence of alternate (other than direct) pathways of relaxation:

$$\text{H8/H6}(i) \xrightarrow{2.3\ \text{Å}} \text{H2}'(i) \xrightarrow{1.8\ \text{Å}} \text{H2}''(i) \xrightarrow{2.1\ \text{Å}} \text{H6/H8}(i+1)$$

Note that, for a pair of protons separated by a long distance (~4.5 Å), the higher order NOEs might appear in a NOESY spectrum at a moderate τ_m if the two protons are connected by relaxation pathways involving short distances (~2 Å).

H2 of A7 was assigned as the signal at 7.92 ppm, which gave no NOESY cross-peak with any proton (base/sugar, see, for example, Figure 4). One cross-peak in Figure 3 linked two base protons belonging to two purines. The signals were located at 8.36 and 7.65 ppm. The signal at 8.36 ppm was assigned to H8 of A7 and that at 7.65 to H8 of G8 because A7 and G8 are the only two purines adjacent in sequence. The observed cross-peaks between purines can be rationalized only on the basis of a B-DNA duplex in which A7 and G8 are part of the helix and the cross-peak originates through the pathway mentioned above. The specific assignment of the low-field signal to A7 and the high-field one to G8 makes internal sense during NOESY base–H1' cross-connectivity walk (see later, Figure 4). Also, the frequency distribution of H8 and H2 of A7 is in accordance with the general occurrence of those signals in A residues in several other oligonucleotide duplexes, i.e., H2 of A occurring at a higher field than H8 of A. The signals of 8.02/7.99 ppm belong to H8's of G2/G5. The base protons are accounted as follows: 8.36 (H8 of A7), 8.02/7.99 (H8's of G2/G5), 7.92 (H2 of A7), 7.65 (H8 of G8), 7.59 (H6's of three C's), and 7.43 (H6's of two C's). In a tight B-DNA duplex of $d(CGCCGCAGC)_2$, H6's of C3 and C4 are expected to show cross-peaks with each other following the criteria mentioned above for A7 and G8. The precise identification of cross-peaks between H6's of C3 and C4 in Figure 3 was not possible due to extensive overlap. However, 1D slices through 7.59 and 7.43 ppm revealed that the 7.43 ppm signal was a site of NOE from the 7.59 ppm signal and vice versa. C3 and C4 being the only two C's close in sequence to produce a H6–H6 cross-peak, it followed that H6 of C3 is located at 7.43 ppm and H6 of C4 is at 7.59 ppm or vice versa. The assignments of H6's of C3 and C4 and hence the distribution of other H6's to the signals at 7.59 and 7.43 ppm were arrived at in the following manner: As shown in Figure 3, H8's of G2/G5 (at 8.02/7.99 ppm) show cross-peaks with the signals at 7.59 and 7.43 ppm. H6 of C3 is expected to show a cross-peak with H8 of G2, and H6's of C4 and C6 are likely to show cross-peak with H8 of G5. Because the signal at 8.02 ppm as well as the one at 7.99 ppm shows cross-peaks with the signals at 7.59 and 7.43 ppm, H6's of C3 and C6 should be located at one site while H6's of C1 and C4 ought to be at the other site. As explained later, in order to meaningfully interpret the cross-connectivity between H8/H6---H1' and H8/H6---H2'/H2'', the signal at 8.02 ppm had been assigned as H8 of G2, and the one at 7.99 ppm was that of H8 of G5. Once the locations of H8's of G2 and G5 are known, the H6's of five C's were readily assigned as H6's of C1, C4, and C9 at 7.59 ppm and H6's of C3 and C6 at 7.43 ppm. Having assigned the H6's to individual C's, H5's of C's were readily assigned, i.e., H5's of C1, C4, and C9 at 5.90 ppm and H5's of C3 and C6 at 5.43 ppm (from the COSY experiment).

Sugar Protons H1' and H2'/H2''s in Figure 2C,D

Figure 4 shows the NOESY cross section showing the cross-peaks between H8/H6 and H1'. In a regular duplex (A or B form) H8/H6 is close to its own sugar H1' (distance ~3.7 Å) and to the sugar H1' of the 5'-neighbor (distance ~3.0 Å in B-DNA and 4.0 Å in A-DNA); for example, in the present case, H6 of C3 should be close to its own H1' and to H1' of G2. On the basis of this criterion, we are able to sequentially assign all the base H8/H6 and sugar H1' belonging to all the nine individual nucleotides of d-(CGCCGCAGC) in the duplex state. The NOESY H8/H6---H1' connectivity route shown in Figure 4 is internally consistent and is in agreement with base proton–base proton cross-peaks in Figure 3. As is observed in the connectivity route, C1 shows a very weak cross-peak with its own sugar H1' (as expected for a terminal residue). The cross-peak between H1'-C1 and H8-G2 is also expected to be weak; however, the latter overlaps with the cross-peak between H8-G2 and H5-C3, which is expected to be medium in intensity as stated below. In addition, Figure 4 demonstrates the spatial interaction between H8 of G2 and H5 of C3, H8 of G5 and H5 of C6, and H8 of G8 and H5 of C9; i.e., there are cross-peaks between signals at 8.02 (H8 of G2) and 5.45 ppm (H5 of C3), between signals at 7.99 (H8 of G5) and 5.45 ppm (H5 of C6), and between signals at 7.65 (H8 of G8) and 5.90 ppm (H5 of C9); in a regular duplex, for G-C sequence the distance between H8G and H5C ~ 3.4 Å.

Having assigned the nine individual H1''s, corresponding H2'/H2'' were assigned by monitoring the NOESY cross-peaks between H1' and H2'/H2'' as shown in Figure 5.

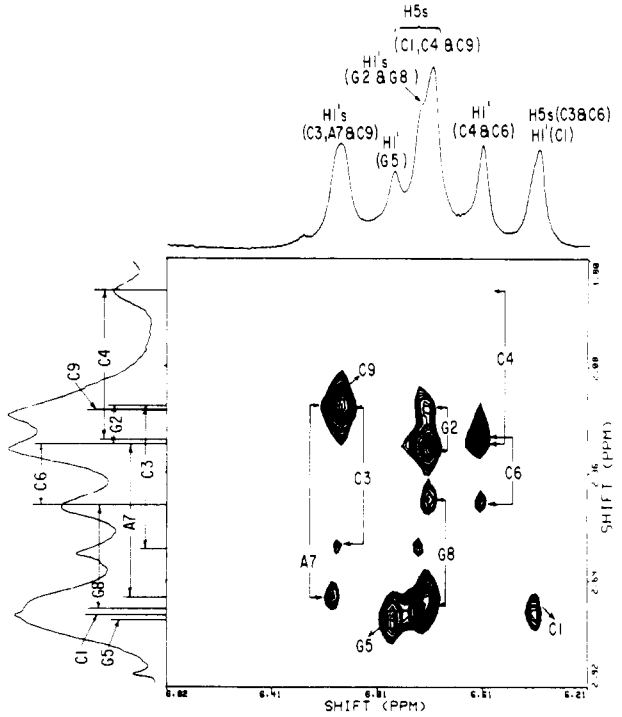


FIGURE 5: Cross section of the NOESY spectrum showing the cross-peaks between H1"s and H2'/H2"s; except for one H2'/H2" signal at 1.85 ppm, all the expected cross-peaks are observed. H2'---H2" cross-connectivity was established by examining the relevant NOESY cross section. The assignment of H2' and H2" is as follows: C1 (2.73, 2.73), G2 (2.30, 2.20), C3 (2.19, 2.55), C4 (2.28, 1.85), G5 (2.74, 2.74), C6 (2.28, 2.43), A7 (2.68, 2.21), G8 (2.71, 2.43), and C9 (2.19, 2.19).

Table I lists the chemical shift values (in ppm) of the base protons, H1' and H2'/H2" of nine nucleotides of d-(CGCCGCAGC) in the duplex state. Even though the general positions of the H3', H4', H5'/H5" signals are apparent, no attempt was made to assign these signals to individual nucleotides.

Our assignment procedure was as follows: (i) Strong NOE from H8/H6 to the H2', H2" region and the absence of NOE from H8/H6 to the H3' region suggested that the oligomer adopted a B-DNA duplex in solution. (ii) Hence, the NOESY cross sections involving *H8/H6---H1'* and *H1'---H2',H2"* were interpreted in terms of a B-DNA model to arrive at a sequential assignment of H8/H6, H1', H2' and H2" of the nine residues; this was supplemented by the COSY data to arrive at the assignment of H5's of C.

There is an alternative approach that does not require the assumption of any structure for assignment; the procedure involves the following steps: (i) assignment of H1', H2', H2", H3', etc. of each sugar from the COSY data; (ii) then, the sequential assignment from the ^1H–^{31}P coupling data, which reveal the connectivity H3' (i − 1) → P(i) → H5'/H5"(i) → and so on.

We were unable to perform the latter procedure in view of the fact that the F. Bitter National Laboratory, MIT (where we conducted our experiment), at the present moment is not equipped with a facility to conduct ^1H–^{31}P correlated spectroscopy.

Figure 6 shows the base H8/H6---sugar H2'/H2" cross-peaks of d(CGCCGCAGC) in the duplex state. Strong NOEs at H2'/H2" from H8/H6 bear the signature of a right-handed double helix, i.e., either the A or the B form. The A and B forms are distinguished from each other by a set of interproton distances involving base H8/H6 and sugar H2'/H2" and H3'. *In the A-DNA*, the distance between base H8/H6 and its own

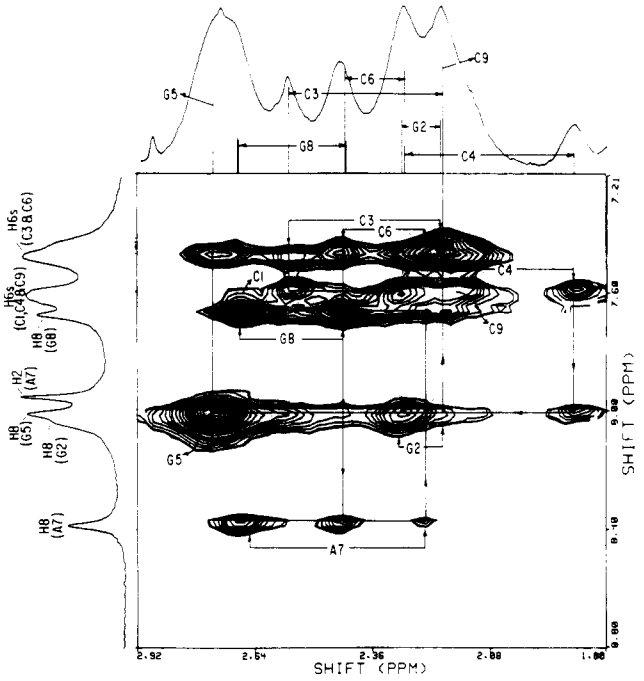


FIGURE 6: NOESY cross section of d(CGCCGCAGC)$_2$ duplex showing the H8/H6---H2'/H2". The assignment of H2'/H2" given here is the one which is consistent with the right-handed B-DNA model of d(CGCCGCAGC); i.e., for a B-DNA model of the DNA, the assignment of H2'/H2" as given made meaningful chemical sense.

sugar H2' is ~3.9 Å, and a weak NOE is expected at sugar H2' from its base H8/H6, but the distance between H8/H6 of a base and sugar H2' of its 5'-neighbor is ~2.0 Å, and thus a strong NOE is expected from H8/H6 to the H2' of its 5'-neighbor. In addition *in the A form* base H8/H6 and its own sugar H3' are 2.9 Å apart, which should result in a medium NOE from base H8/H6 to its own sugar H3'. *In the B-DNA*, the distance between base H8/H6 and its own sugar H2' is ~2.4 Å, and therefore, a strong NOE is expected from base H8/H6 to its own sugar H2'; also, base H8/H6 and sugar H2" of the 5'-neighbor being 2.1 Å apart, a strong NOE from base H8/H6 to sugar H2" of its 5'-neighbor is expected. *In the B-DNA*, the distance between base H8/H6 and its own sugar H3' is beyond 4.5 Å, and thus a weak or no NOE is expected from base H8/H6 to its own sugar H3'. In summary, the expected NOE pattern for the A form is strong NOEs between base H8/H6 and sugar H2' and medium NOEs between base H8/H6 and sugar H3'. For the B form, the expected NOE pattern is strong NOEs between base H8/H6 and sugar H2'/H2" and a weak or no NOE between H8/H6 and H3'. For a detailed discussion of manifestation of the A and B forms of DNA in the base–sugar NOE pattern, see Gupta et al. (1985, 1986) and Sarma et al. (1986).

The NOESY spectrum of d(CGCCGCAGC) showed a weak NOE between base H8/H6 and sugar H3' (data not shown); this rules out the possibility of the A form. Figure 6 displays a H8/H6---H2'/H2" NOE pattern quite characteristic of the right-handed B-DNA.

The assignment of H2', H2" given in Figures 5 and 6, Table I, is the one that meaningfully explained the base H8/H6---H2', H2" cross-connectivity on the basis of a B-DNA model for the duplex d(CGCCGCAGC).

It should be emphasized that even though we cannot prevent diffusion from H2' to H2" (in a given residue) under conditions of NOESY (τ_m = 200 ms), this does not affect any of our assignments or structural conclusions because we are not measuring interproton distances (for quantitative purpose).

Table I: Chemical Shift Values (ppm) of Signals in the Duplex $d(CGCCGCAGC)_2$ in Solution at 5 °C[a]

residue	protons			
	H6/H8	H2/H5	H1′	H2′/H2″
C1	7.59	5.90	5.45	2.73, 2.73
G2	8.02		5.90	2.30, 2.20
C3	7.43	5.45	6.19	2.19, 2.55
C4	7.59	5.90	5.67	2.28, 1.85
G5	7.99		6.00	2.74, 2.73
C6	7.43	5.45	5.67	2.28, 2.43
A7	8.36	7.92	6.19	2.68, 2.21
G8	7.65		5.90	2.71, 2.43
C9	7.59	5.90	6.19	2.19, 2.19

[a] Chemical shift values are given with respect to TSP as an internal standard.

Theoretically in B-DNA the connectivity pattern is H2′–base H8/H6---H2″ [(–) indicates intranucleotide; (---) denotes 5′-neighbor]. This means that excluding the terminal residues (5′ and 3′) each of the base protons should have three cross-peaks along the H2′/H2″ axis provided H2′ and H2″ of the same residue have different chemical shifts. In Figure 6 the signature of B-DNA is observed. For example, in the case of A7 in Figure 6 the first cross-peak is between A7H8 and A7H2′, the second one between A7H8 and C6H2″, and the third one between A7H2″ and G8H8. The point we must emphasize is that, because of H2′ → H2″ diffusion, along the H2′/H2″ axis, one will always see cross-peaks for H2′ and H2″ corresponding to the same base proton; but the H2″ cross-peak will be isolated and *will not make the required connectivity*, if the system is not B-DNA.

C4 and A7 Are an Integral Part of the Helix

The following is the summary of arguments in favor of our hypothesis that C4 and A7 form an integral part of the B-DNA duplex of d(CGCCGCAGC) in solution: (i) Irrespective of any assignment, the data in Figure 3 clearly showed a cross-peak between two purines; this is only possible if purines adjacent in sequence are also close in space as in B-DNA; i.e., A7 (as per sequence) must be a part of the helix. (ii) Irrespective of any assignment the cross-peaks between two pyrimidines in Figure 3 are possible only if C3 and C4 are close in space as in B-DNA; i.e., C4 forms a part of the helix. (iii) In order to explain the connectivity pattern involving base H8/H6 and sugar H1′ (Figure 4), it was obligatory that both C4 and A7 formed an integral part of the helix. (iv) H8/H6---H2′/H2″ cross-connectivity (Figure 6) was only consistent with both C4 and A7 being a part of the helix and not looped out.

Cross-peaks are observed in the d(GpC) sequence between H8 of G and H5 of C (distance H8 of G---H5 of C is ~3.4 Å), see Figure 4. However, in the d(CpG) sequence the distance H5 of C---H8 of G is beyong 4.5 Å, and no cross-peak is expected between H5 of C and H8 of G.

Probable H-Bonded Pairings between A and C

Even though it is clear that C4 and A7 are integral parts of the double helix, it is of interest to examine if they indeed form H-bonded pairs. In view of the fact that we are working at pH 4.5, the possibility of A being protonated is very real. Comparison of the D_2O spectra of the oligomer at neutral pH and at pH 4.5 indicated that H8 of A7 moved downfield by about 0.3 ppm upon lowering the pH. This is consistent with a recent report (Sowers et al., 1986) that upon protonation (in the pH range 4–5) of A H8 moves downfield by 0.3 ppm. When A7 is protonated at pH 4.5, a doubly H-bonded A-C

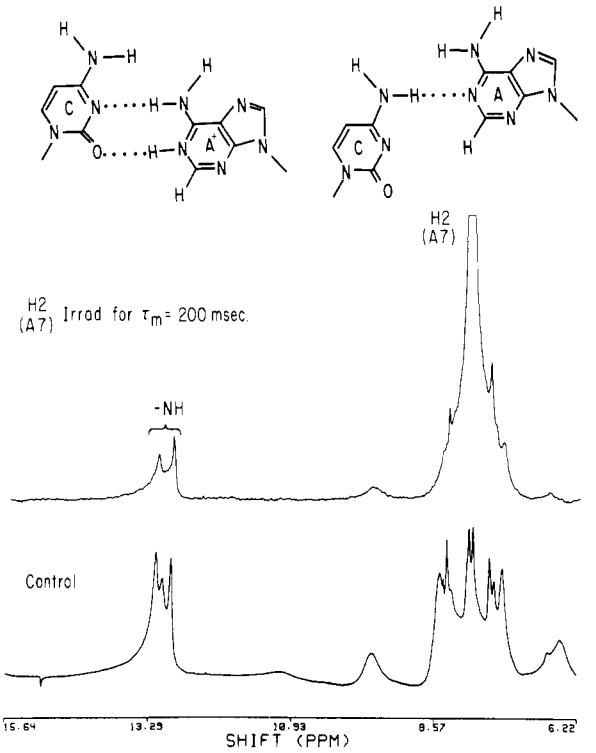


FIGURE 7: 1D NOE spectrum of d(CGCCGCAGC) duplex in H_2O with other solution conditions same as in Figures 2–6. Spectrum on top shows NOE sites at two –NH signals with H2 of A7 as the site of irradiation (τ_m = 200 ms, 20 dB); the control spectrum is also shown at the bottom. The NOEs at two sites within 13.2–12.9 ppm can be explained by one of two ways: (i) A7 and C4 form a double H-bonded A^+=C pair with A being protonated at N1 as shown, or (ii) A7 and C4 form a single H-bonded A-C pair with both A7 and C4 in their neutral keto form. In A^+=C pairs, two sites of NOE should originate as A7H2 → A7N1$^+$-H and as A7H2 → A7N1$^+$-H → G8N1-H (see text). In A-C pairing, two sites of NOE are explained as A7H2 → G5N1–H and as A7H2 → G8N1-H (see text). Under this experimental condition along with H2 of A7, the neighboring signal, i.e., H8's of G2 and G5, is partially irradiated due to an unavoidable radio-frequency spill. This results in the observation of NOEs (appearing as the shoulder near the site of irradiation) at H6's of C3 and C6. We did not have the capability to perform a NOESY of d(CGCCGCAGC) in H_2O such that these artifacts could be removed. Also included in the diagram are the pairing schemes involving C4 and A7, which would explain the data (see text for details).

pair of the type shown in Figure 7 is expected to be present. If this doubly H-bonded species is present, in the water spectrum we should be able to isolate the hydrogen-bonded imino proton of the protonated A in the ≃13 ppm range. The water spectrum in Figures 2A and 7 at pH 4.5 show that we have been unsuccessful in locating the expected resonance unequivocally even though the expected resonance could be merged with the GN1–H resonances in the 13.2–12.9 ppm. We have attempted to decipher the resonance by comparing the water spectra of the oligomer at pH 7.0 and 4.5. At these two pH values, not only the state of protonation of A7 but also the 3D structures of the oligomer are different. Hence, one cannot use this approach to decipher profitably the position of N1$^+$–H of A7 (if any) at pH 4.5. The details of our studies on d(CGCCGCAGC) at pH 7 will be reported at a later time.

There are two very distinctive features that emerge when A^+=C pairs are incorporated inside the core of a B-DNA double helix: (i) in the A^+=C pair (Figure 7), H2 is only 2.5 Å apart from N1$^+$-H of A, and (ii) in the double-helix, N1$^+$-H of A7 is about 3.4 Å away from N1-H of G8 (see later in Figure 9).

We tried to determine the proximity of these interproton distances by conducting an 1D NOE experiment on d-

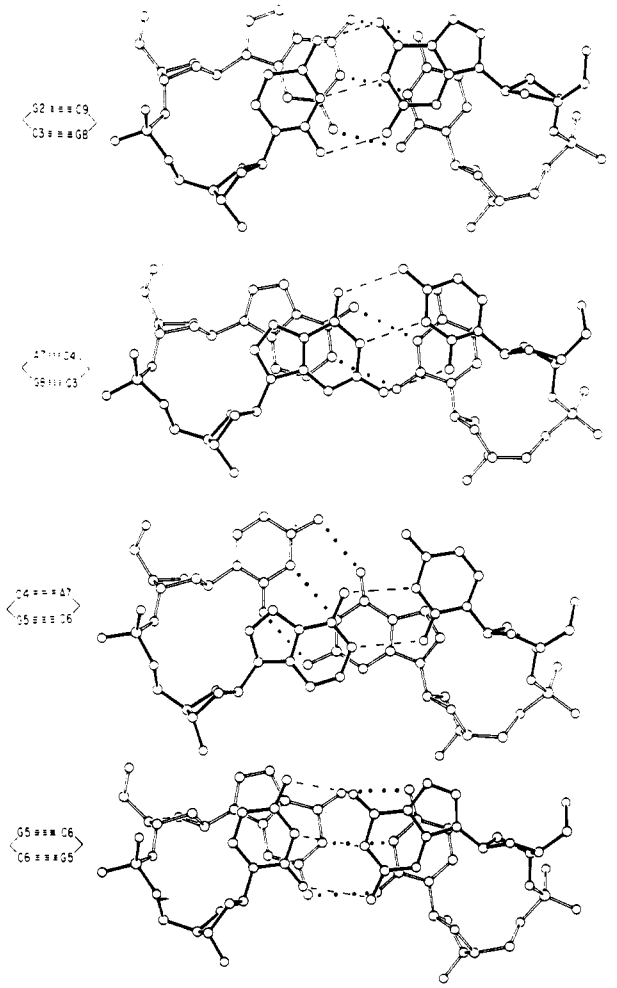


FIGURE 8: Stacking arrangement of various doublets in the B-DNA model of d(G2-C3-C4-G5-C6-A7-G8-C9)$_2$ in which A$^+$=C pairs are present and C4 and A7 form an integral part of the double helix. For conformational parameters of the model, see Table II. In the doublet $<^{A7^+\ C4}_{G8\ C3}>$, N1$^+$ of A7 is about 3.4 Å away from N1-H of the G8-C3 pair, which is consistent with the NOE data (Figure 7). A view along the helix axis is available from the authors.

(CGCCGCAGC)$_2$ in H_2O at pH 4.5 at 5 °C. In the experiment we irradiated H2 of A7 and monitored NOEs in the H-bonded –NH region. Figure 7 shows the results of such an experiment. NOE is observed at two sites in the –NH region. This is consistent with two H-bonded A$^+$=C pairs: the stronger site of NOE is from H2 to N1$^+$-H in A7 (Figure 7), and the second one originates as a second-order NOE as H2(A7) → N1$^+$-H(A7) → N1-H(G8) (see the doublet $<^{A7\ C4}_{G8\ C3}>$ in Figure 8).

There is an alternate scheme of single H bonding between C4 and A7 that could also explain the 1D NOE data of Figure 7. In this H-bonding scheme (Figure 7), both C and A are in neutral keto form and have only one H bond. In a B-DNA duplex when A-C pairs form the core of the structure there could be two direct NOEs from H2 of A7, i.e., *H2 of A7 to N1-H in G5-C6* and *H2 of A7 to N1-H in G8-C9* (the distance is ~3.4 Å; see Figure 9 for details).

In the water spectra in Figure 7, at the AH2 position, some NH2 resonances are present. Hence, some of the observed magnetization transfer from AH2 to NH resonances originates from amino proton transfer. This explains why the observed NOEs at NH appear larger than one would expect for adenine H2 irradiation. Because we are not doing actual distance measurements in this instance but only checking the proximity of AH2 to other protons, our conclusions are not compromised.

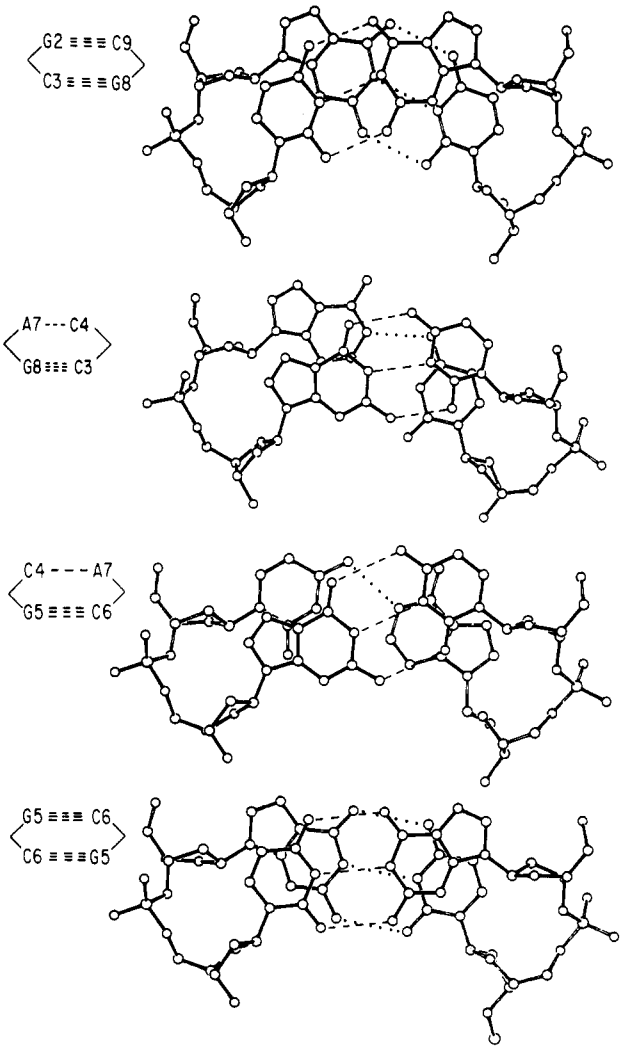


FIGURE 9: Stacking arrangement of various sequences as viewed down the helix axis for the B-DNA model containing d(G2-C3-C4-G5-C6-A7-G8-C9)$_2$ in which C4 and A7 form an integral part of the helix without any major distortion in the rest of the helix. We did not attempt to determine nucleotide geometry of the dangling C1, and hence, this is not included in the diagram. In this model, there is a small difference in the stacking arrangements of the doublets of $<^{C3\ G8}_{G2\ C9}>$ and $<^{C6\ G5}_{G5\ C6}>$ even though both involve G-C sequences: In the $<^{C6\ G5}_{G5\ C6}>$ sequence the separation between the bases along the helix axis is 3.4 Å (an optimum base–base distance for substantial geometric overlap) while in the $<^{C3\ G8}_{G2\ C9}>$ sequence the separation between the bases along the helix axis is 3.2 Å. As a result, in $<^{C6\ G5}_{G2\ C9}>$ there is a subatantial geometric overlap between G5 and C6. While in the $<^{C3\ G8}_{G2\ C9}>$ sequence there is a sliding between the base planes of G2 and C3 to allow favorable stacking. A view along the helix axis is available from the authors.

B-DNA Model for Duplex d(CGCCGCAGC) in Solution Derived from NMR Data

Molecular modeling was performed for the duplex involving (G2-C3-C4-G5-C6-A7-G8-C9)$_2$ subject to the following constraints: (i) there are spatial interactions of H8/H6 in the neighboring bases (distances being kept within 5 Å) (Figure 3); (ii) there is interrelation between H8/H6 and H1′; i.e., the distance between H8/H6 of a base and its own sugar H1′ fixed within 3.5–4.0 Å and the distance between H8/H6 and sugar H1′ of its 5′-neighbor is ~3.0–3.7 Å (Figure 4); (iii) distances between H8 of G2 and H5 of C3, H8 of G5 and H5 of C6, and H8 of G8 and H5 of C9 are kept around 3.5 Å (Figure 4); (iv) there is very close proximity between H8/H6 and H2′/H2″; i.e., H8/H6 of a base and its own sugar H2′ distance is 2.4 Å while H8/H6 and sugar H2″ of the 5′-neighbor being

Table II: Torsion Angles (deg) of the Duplex d(G2-C3-C4-G5-C6-A7-G8-C9)$_2$ in Solution[a]

residue	α	β	γ	δ	ϵ	ξ	χ
G2			50	135	215	210	260
			(50)	(130)	(222)	(220)	(250)
C3	306	144	45	140	230	224	262
	(294)	(166)	(45)	(153)	(205)	(225)	(260)
C4	309	136	37	130	221	220	250
	(317)	(148)	(38)	(135)	(221)	(214)	(247)
G5	312	138	40	125	222	218	242
	(311)	(144)	(41)	(144)	(222)	(212)	(245)
C6	314	140	40	135	222	216	252
	(313)	(144)	(41)	(144)	(222)	(214)	(242)
A7	312	152	37	140	222	228	250
	(317)	(152)	(40)	(140)	(191)	(219)	(257)
G8	312	146	45	140	207	219	243
	(328)	(148)	(42)	(144)	(212)	(220)	(256)
C9	313	141	42	141			235
	(314)	(154)	(43)	(145)	(212)	(220)	(250)

[a] For definition of the torsion angles, see Saenger (1984). The atomic coordinates can be obtained from the authors upon request. The values within the parentheses correspond to the B-DNA model with single H-bonded A-C pairs (Figure 9).

2.1 Å (Figure 6); (v) G2 and C9, C3 and G8, and G5 and C6 form regular Watson–Crick pairs while C4 and A7 form a single, two H-bonded A-C pair as in Figure 7.

Figure 8 shows the successive base pairs in the duplex down the helix axis, i.e., the doublets

$$\left\langle \begin{matrix} G2 & C9 \\ C3 & G8 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} C4 & A7 \\ C3 & G8 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} C4 & A7 \\ G5 & C6 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} G5 & C6 \\ C6 & G5 \end{matrix} \right\rangle$$

The model satisfies all the criteria required by the NMR data and is free of any steric compression. It can be seen from the figure that that A^+=C pairs are incorporated in the model with a minimum distortion in the double helix; i.e., as in a regular B-DNA duplex, the bases are perpendicular to the helix axis and situated close to the helix center with adjacent base planes separated by about 3.2–3.5 Å and rotated by about 32°–38°.

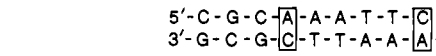
Figure 9 describes the alternative B-DNA model of d(G2-C3-C4-G5-C6-A7-G8-C9)$_2$ with C4 and A7 involved in a single H-bonding scheme (Figure 7). In this model, too, both C4 and A7 form an integral part of the B-DNA duplex. In the model all eight residues belong to average (C2′-endo,anti) conformation in accordance with the NOE data. Notice that H2 of A7 is close to N1–H of G5 and G8, which would explain the 1D NOE data of Figure 7.

Table II lists the conformational parameters of two B-DNA models for d(G2-C3-C4-G5-C6-A7-G8-C9)$_2$ as described in Figures 8 and 9. It is evident from the table that, even though all eight nucleotides belong to the C2′-endo,anti domain, their conformations are not exactly the same. In other words, for molecular model building the conformation of each residue was varied within the C2′-endo,anti domain to arrive at the final model; however, in the final model the conformations of C4 and A7 (involved in A-C pair) are not significantly different from those of G2, C3, G5, C6, G8, and C9 (involved in Watson–Crick G-C pairs). This is a clear demonstration that A-C pairing could be stereochemically accommodated in a B-DNA duplex model of d(CGCCGCAGC)$_2$ in solution without any major distortion in the rest of the helix.

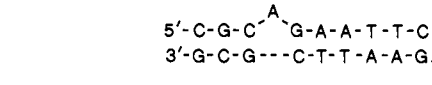
The structures described in Figures 8 and 9 are representative working models that are consistent with the NMR data of Figures 1–7. The refined model could be obtained by simulation of *theoretical* NOESY data for a given model with respect to experimental NOESY data at various mixing time. Efforts are in progress in this direction in our and other laboratories.

Discussion

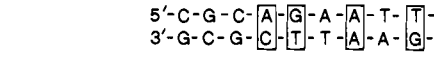
In summary, the NMR data on the d(CGCCGCAGC)$_2$ duplex give direct evidence that C4 and A7 form an integral part of the B-DNA duplex and they are not definitely looped out. It is encouraging to note that the A^+=C pairs have been implicated (Hunter et al., 1986) for the single crystal structure of



An interesting observation can be made on the basis of our solution structure of the d(CGCCGCAGC)$_2$ duplex and the dodecamer duplex in the single crystal (Hunter et al., 1986). In both cases, A-C pairs are surrounded by a host of regular Watson–Crick pairs that might help the helix adapt to a single Watson–Crick base pair replacement by a non-Watson–Crick pair. This hypothesis is strengthened by the observation that the tridecamer d(CGCAGAATTCGCG) forms a duplex without A-C pairs in crystal (Saber et al., 1986); i.e., the tridecamer forms a duplex as



with A being looped out rather than a structure with A-C pairs as in



The latter arrangement with A-C pairs also involves the formation of non-Watson–Crick G-T and A-A pairs, which might significantly reduce the probability of such a structure occurring in a crystal form or solution.

Acknowledgments

The high-field NMR experiments were performed at the NMR facility for Biomolecular Research located at the F. Bitter National Laboratory, MIT. The NMR Facility is supported by Grant RR00995 from the Division of Research Resources of the NIH and by the National Science Foundation under Contract C-670.

Registry No. d(CGCCGCAGC), 110825-97-3.

References

Coll, M., Solans, X., Font-Altaba, M., & Subirana, J. (1987) *J. Biomol. Struct. Dyn. 4*, 797–811.

Conner, B. N., Takano, T., Takano, S., Itakura, K., & Dickerson, R. E. (1982) *Nature (London) 295*, 294.

Digweed, M., Pieler, T., & Erdmann, V. A. (1986) in *Structure and Dynamics of RNA* (Knippenberg, P. H. V., & Hilbers, C. W., Eds.) p 205, Plenum, New York.

Drew, H., Takano, T., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 286*, 567.

Gupta, G., Sarma, M. H., & Sarma, R. H. (1985) *J. Mol. Biol. 186*, 463.

Gupta, G., Sarma, M. H., & Sarma, R. H. (1986) *Int. J. Quantum Chem., Quantum Biol. Symp. 12*, 183.

Ho, P. S., Frederick, C. A., Quigley, G. A., van der Marel, G. A., van Boom, J. A., Wang, A. H.-J., & Rich, A. (1985) *EMBO J. 4*, 3617.

Hunter, W. N., Brown, T., Anand, N. N., & Kennard, O. (1986) *Nature (London) 320*, 552.

Kan, L. S., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263.

Kennard, O. (1985) *J. Biomol. Struct. Dyn. 3*, 205.

Matteuci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185.

McConnell, B. (1984) *J. Biomol. Struct. Dyn. 1*, 1407.
Pardi, A., Morden, K. M., Patel, D. J., & Tinoco, I., Jr. (1982) *Biochemistry 21*, 6567.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, I. (1984a) *Biochemistry 23*, 3207.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, I. (1984b) *Biochemistry 23*, 3218.
Sabat, M., Cini, R., Haromy, T., & Sundaralingam, M. (1985) *Biochemistry 24*, 7827–7833.
Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Saper, M. A., Eldar, H., Mizuuchi, K., Nickol, J., Apella, E., & Sussman, J. L. (1986) *J. Mol. Biol. 188*, 111.
Sarma, M. H., Gupta, G., & Sarma, R. H. (1986) *Biochemistry 25*, 3659.
Shakked, Z., Rabinovch, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A., & Viswamitra, M. A. (1981) *Proc. R. Soc. London, B 213*, 479.
Sowers, L. C., Fazakerley, G. V., Kim, H., Dalton, L., & Goodman, M. (1986) *Biochemistry 25*, 3983.
Wang, A. H.-J., Quigley, G., Kolpack, F. J., Crawford, L., van Boom, J. H., van der Marel, & Rich, A. (1979) *Nature (London) 282*, 680.
Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968.
Wang, A. H.-J., Ughetto, G., Quigley, G. J., & Rich, A. (1986) *J. Biomol. Struct. Dyn. 4*, 319–342.
Westhof, E., & Sundaralingam, M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1852–1856.
Westhof, E., Rao, S. T., & Sundaralingam, M. (1980) *J. Mol. Biol. 142*, 331–361.
Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755.

DNA Hairpin Structures in Solution: 500-MHz Two-Dimensional ^{1}H NMR Studies on d(CGCCGCAGC) and d(CGCCGTAGC)†

Goutam Gupta, Mukti H. Sarma, and Ramaswamy H. Sarma*
Institute of Biomolecular Stereodynamics, State University of New York at Albany, Albany, New York 12222
Rolf Bald, Ute Engelke, Shiao Li Oei, Reinhard Gessner, and Volker A. Erdmann
Institut für Biochemie, Fachbereich Chemie, Freie Universität Berlin, 1000 Berlin 33 (Dahlem), FRG
Received April 13, 1987; Revised Manuscript Received July 1, 1987

ABSTRACT: A hairpin structure contains two conformationally distinct domains: a double-helical stem with Watson–Crick base pairs and a single-stranded loop that connects the two arms of the stem. By extensive 1D and 2D 500-MHz ^{1}H NMR studies in H_2O and D_2O, it has been demonstrated that the DNA oligomers d(CGCCGCAGC) and d(CGCCGTAGC) form hairpin structures under conditions of low concentration, 0.5 mM in DNA strand, and low salt (20 mM NaCl, pH 7). From examination of the nuclear Overhauser effect (NOE) between base protons H8/H6 and sugar protons H1′ and H2′/H2″, it was concluded that in d(CGCCGCAGC) and d(CGCCGTAGC) all the nine nucleotides display average (C2′-endo,anti) geometry. The NMR data in conjunction with molecular model building and solvent accessibility studies were used to derive a working model for the hairpins.

Self-complementary DNA oligomers, in general, adopt double-helical structures. Double helices are stabilized by Watson–Crick base pairing, intra- and interstrand stacking, and favorable backbone and base–backbone conformations. The double-helical structures of self-complementary DNA oligomers are formed only when negatively charged phosphate ions are neutralized by counterions of salt and when the concentration of DNA is sufficiently high such that every oligomer finds its complementary partner. But high salt and DNA concentration are not merely enough to form a double helix; the temperature should also be significantly lower than the melting temperature T_m of the corresponding oligomer. Thus, salt, DNA concentration, and temperature are three physical features that are often altered to monitor helix → coil transitions. It has been recently (Haasnoot et al., 1983, 1986; Orbons et al., 1986) reported that in the pathway of helix → coil there could be a structural intermediate for self-complementary DNA oligomers. In other words, at a particular salt and DNA concentration and within a small range of temperature it has been observed that oligomers display a novel kind of structure called the "hairpin" (Haasnoot et al., 1983, 1986; Orbons et al., 1986). A hairpin structure contains two distinct domains—a double-helical stem with Watson–Crick paired bases and a single-stranded loop that joins the two arms of the stem and the stacking interaction of the bases in the loop and at the junctions between the stem and the loop (Haasnoot et al., 1986). The possibility of the formation of a hairpin structure as a stable thermodynamic entity is markedly enhanced for non-self-complementary DNA oligomers because mismatch pairs significantly lower the stability of a rigid duplex under ordinary conditions of salt and DNA concentration. In this paper, we discuss the identification and the stereochemical characterization of hairpin structures as obtained for two non-self-complementary DNA segments each nine nucleotide long; these DNA segments are DNA analogues

† This research is supported by a grant from the National Institutes of Health (GM29787), the Deutsche Forschungsgemeinschaft (DFG-SFBg-B5), the Fonds der Chemischen Industrie e.V., and a contract from the National Foundation for Cancer Research.

0006-2960/87/0426-7715$01.50/0