Abasic Frameshift in DNA. Solution Conformation Determined by Proton NMR and Molecular Mechanics Calculations†

Ph. Cuniasse,‡ L. C. Sowers,§,∥ R. Eritja,∥ B. Kaplan,∥ M. F. Goodman,⊥ J. A. H. Cognet,# M. Le Bret,# W. Guschlbauer,‡ and G. V. Fazakerley*,‡

Service de Biochimie, Bat. 142, Département de Biologie, Centre d'Etudes Nucléaires de Saclay, F-91191 Gif-sur-Yvette Cédex, France, Department of Molecular Genetics, Beckman Research Institute of City of Hope, Duarte, California 91010, Molecular Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-1481, and Laboratoire de Physicochimie Macromoléculaire, Institut Gustave Roussy, F-94800 Villejuif, France

Received May 4, 1988; Revised Manuscript Received September 9, 1988

ABSTRACT: We have determined the three-dimensional structure of a non-self-complementary oligodeoxynucleotide duplex that contains a model abasic site. The duplex contains six GC base pairs plus the abasic site at the center of one strand and corresponds to an abasic frameshift. Two-dimensional NMR studies on the nonexchangeable protons show that the guanine bases on either side of the abasic site are stacked over each other and that the abasic site is rotated out of the helix. Close proton–proton interactions are observed between the H4′ proton of the abasic site and sugar protons of the guanosine in the 5′ direction, which allows the position of the free sugar to be well-defined. NOE buildup curves from NOESY spectra recorded at very short mixing times were used to calculate a set of interproton distances. This data set was incorporated into the refinement of the oligonucleotide structure by molecular mechanics calculations. Two conformations that differ in the sugar conformation of the guanosine next to the abasic site in the 3′ direction were necessary to fit all the NMR data. One of these two conformations could only be stabilized by addition of counterions at specific sites.

Replication of a DNA containing an abasic site, which is the most frequent damage in DNA (Lindahl & Nyberg, 1972; Lindahl & Karlström, 1973), can give rise to a normal structure that retains the genetic information, as well as two abnormal structures that have lost this information. The abnormal structures and the normal one are presented in Figure 1. It has been shown that polymerases can, in vitro, copy over homopolymeric or natural DNAs containing abasic sites by insertion of incorrect nucleotides opposite the baseless site. Thus, the first abnormal consequence is an abasic site containing one of the four bases (Figure 1B). The structure of an oligonucleotide containing an abasic site with an A opposite the site has been recently reported (Cuniasse et al., 1987). Another study of an oligonucleotide containing an abasic site has been also reported (Raap et al., 1987). The second abnormal structure can be described as an abasic frameshift associated with a deletion on the nonparental strand (Figure 1C). This structure may arise when the parental strand contains the abasic site and the sugar is in an extrahelical conformation at the replication site. Thus, the polymerase may not incorporate a nucleotide opposite the lost base. While during replication the formation of a frameshift may require that the sugar be in the conformation shown in Figure 1C, the abasic site may reenter the helix afterward or be in an equilibrium between the two states. Neither structural nor thermodynamic information has been reported on this abnormal structure whose presence may induce modifications in the double-helix characteristics as in the genetic information.

Materials and Methods

The strand containing the abasic site and the hexanucleotide were synthesized by a classical phosphotriester method (Gait, 1984; Eritja et al., 1987). They were annealed to form the duplex.

5′(C1 p_1 G2 p_2 G3 p_3 dr4 p_4 G5 p_5 G6 p_6 C7)3′ (strand 1)
3′(G13 p_{12} C12 p_{11} C11 - p_{10} - C10 p_9 C9 p_8 G8)5′ (strand 2)

The duplex was 4 mM in strand, dissolved in 10 mM phosphate buffer, 150 mM NaCl, and 0.2 mM EDTA. NMR[1] spectra were recorded in 99.99% D_2O. Chemical shifts were measured relative to the internal reference tetramethylammonium chloride (3.18 ppm).

NMR Spectra. The spectra did not change over the ca. 4 months during which measurements were made. Thus, the duplex is chemically stable.

NMR spectra were recorded on a Bruker WM500 spectrometer at 10 °C. NOESY spectra were recorded with different mixing times (35, 50, 65, 80, and 400 ms) in the phase-sensitive mode (Bodenhausen et al., 1984) with 2K data points in the t_2 dimension and 176 acquisitions per spectrum; 220 free induction decays were collected in the t_1 dimension. After zero filling to give a 2K × 2K matrix, a sine bell (shifted by $\pi/2$ for cross-peak volume determination) was applied to the data in both dimensions prior to Fourier transformation.

The double-quantum spectrum in D_2O at 10 °C was recorded in the absorption mode (Braunschweiler et al., 1983).

HOHAHA spectra in D_2O at 10 °C were recorded in the phase-sensitive mode with 10- and 40-ms mixing times (Davis

† This work was supported in part by Grants GM 33863 and GM 21422 and by Biomedical Research Support Grant 2S07RR05471 from the National Institutes of Health.

* Author to whom correspondence should be addressed.
‡ Centre d'Etudes Nucléaires de Saclay.
§ Present address: Division of Pediatrics, City of Hope National Medical Center, Duarte, CA 91010.
∥ Beckman Research Institute of City of Hope.
⊥ University of Southern California.
Institut Gustave Roussy.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; 2Q, double quantum; HOHAHA, homonuclear Hartman–Hahn; NOESY, 2D nuclear Overhauser effect enhancement spectroscopy.

0006-2960/89/0428-2018$01.50/0

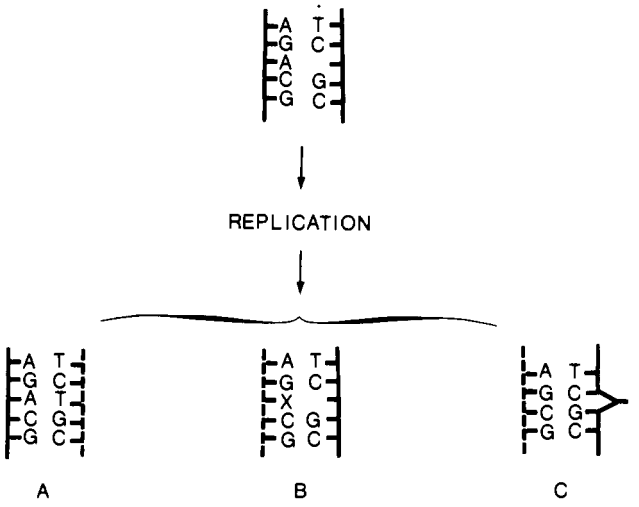

FIGURE 1: Hypothetical structures arising from replication of a DNA containing an abasic site: normal DNA (A); abasic site (B); abasic frameshift (C).

& Bax, 1985). In liquids, rapid molecular reorientation averages out the strong but anisotropic dipolar interactions, leaving only the intramolecular scalar J couplings.

To determine the structure of the oligonucleotide and especially in the region of the abasic frameshift, we have determined a set of interproton distances. To avoid spin diffusion effects and to take into account the apparent correlation time of each different pair of protons, we have used a strategy described previously (Cuniasse et al., 1987). It has been shown that the interproton distances must be determined from short mixing time NOESY spectra. Under these conditions the NOE is dominated by the dipolar contribution, which depends on the inverse of the sixth power of the interproton distance and on the apparent correlation time (Kumar et al., 1981; Chazin et al., 1986). Since the latter is not the same for all interproton interactions, quantitative analysis of the short mixing time NOESY spectra requires, for each interproton distance, the choice of an appropriate reference, a pair of protons for which the distance is known or fixed, and the apparent correlation time can reasonably by assumed to be the same or very similar to the one of the pair of protons considered for which the distance is unknown. We recorded a series of short mixing time NOESY spectra and have measured the NOE buildup rate for each interproton interaction.

Molecular Mechanics Calculations. Energy minimization was carried out by using the program AMBER (Weiner & Kollman, 1981) on a VAX 8600 computer. The parameters were those described by Weiner et al. (1986) and by Singh et al. (1986). All hydrogen atoms are treated explicitly. To simulate the screening effect of the solvent, a gas-phase potential was employed where the dielectric constant D_{ij} is proportional to the distance d_{ij} separating a pair of atoms: $D_{ij} = Cd_{ij}$ (Gelin & Karplus, 1981; Weiner et al., 1984). The proportionality constant C was taken as 4 Å^{-1}. All atom pairs were included in the calculations. Minimizations were carried out with the 1–4 interatomic interactions at full values (not divided by 2). Refinements were terminated when the norm of the energy gradient was less than 0.05 kcal/Å. The negative charges of the phosphate groups of the oligonucleotide were neutralized by spheres with a positive unit charge. These counterions had a van der Waals radius of 2 Å and a wall depth of 0.3 kcal/mol. They were originally set at 3 Å from the phosphate oxygen atoms by the subprogram EDIT of AMBER. Unless specified explicitly, energy values reported are total conformational energies, excluding the contribution due to counterions.

Certain computed structures could be rapidly discarded as they did not satisfy observed sugar conformations or gave very poor quantitative agreement with the NMR data as evaluated by the factor F. The sum $F = \sum(r_{ij} - d_{ij})^2/d_{ij}$, where r_{ij} represents NMR distance measurements and d_{ij} the corresponding distance in a given proposed structure, is used to give an estimate of the overall agreement between computed and NMR data. NMR distance measurements were incorporated into the refinement of the best model structures. To take into account the motional characteristics of the different pairs of protons and to have a better interpretation, the proportionality law between NOE volumes and the inverse of the sixth power of the interproton distance was restricted to pairs of protons expected to have similar motional characteristics. In the simulation, the distance between protons i and j belonging to a pair of type m was forced to the NMR value r_{ij} with a proportionality constant L_m, using the additional energy penalty function $E = \sum k(d_{ij} - L_m r_{ij})^2$. An AMBER-compatible program (Zimmerman, unpublished data) has been written that minimizes the sum of the total energy and of the penalty function, yielding the value of d_{ij} and L_m. The constant k of the penalty function was set equal to 500 kcal/(mol·Å^2). The reference distances used to calibrate the NMR data were taken from an Arnott B-form model. L_m constants were always close to unity at the end of the minimization. The NMR data include some unique interactions involving the dr4 sugar. In these cases the proportionality constant was set to unity.

The molecular structures were displayed on a raster monitor LEX 90 using the program NACAD developed by J. Gabarro and M. Le Bret (unpublished data).

Results and Discussion

NOESY Spectra in D_2O: Assignment of Nonexchangeable Protons. To assign the resonances of the nonexchangeable H5, H6, and H8 base protons and the H1′, H2′, H2″, H3′, H4′, H5′, and H5″ sugar protons, we have recorded NOESY, double-quantum, and HOHAHA spectra of the duplex in D_2O at 10 °C. The spectrum at 23 °C shows that the line widths of the aromatic proton resonances are variable. When the temperature is lowered to 10 °C, those that were broadest at 23 °C decrease in line width. However, even at 10 °C certain resonances remain significantly broad. At lower temperature all the resonances broadened.

We first recorded a NOESY spectrum with a 400-ms mixing time. The cross peaks found in a NOESY spectrum correlate the chemical shifts of nuclei for which cross relaxation occurs during the mixing time and thus are indicative of the proximity of the nuclei involved. The general principles and strategy for the assignment of proton spectra of oligonucleotides from NOESY spectra have been described in detail (Hare et al., 1983; Feigon et al., 1983; Fréchet et al., 1983; Scheek et al., 1984). We have started with the analysis of the H6H8/H1′H5 region (Figure 2) for the sequential assignment of the resonances of the duplex. The six strong cross peaks (× in Figure 2) correspond to six doublets in the region between 5 and 6 ppm of the 1D resolution enhanced spectrum and thus can be assigned to the six cytidine H5–H6 cross peaks.

The most convenient starting point for the sequential assignment is from the resonances of the H6 or H8 proton of the 5′-terminal base of each strand that give rise to only one cross peak in this region of the NOESY spectrum [i.e., the C1(H6) and G8(H8) protons]. For strand 2 of the duplex, the G8(H8) proton at 7.88 ppm gives a cross peak at 5.87 ppm

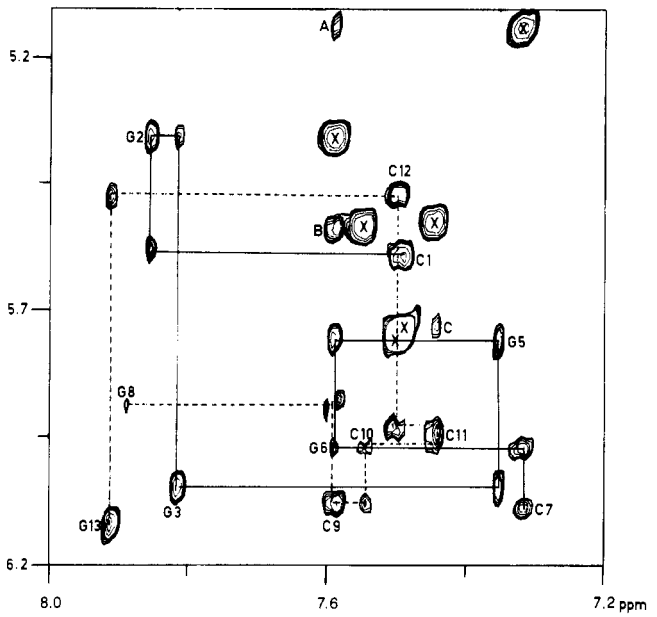

FIGURE 2: Expanded contour plot of the H6,H8/H1′,H5 region of the 400-ms NOESY spectrum of the duplex in D_2O at 10 °C.

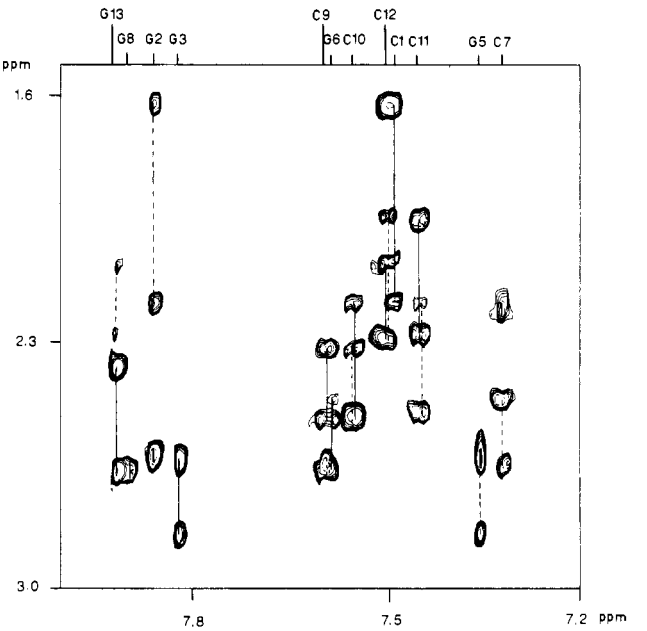

FIGURE 3: Expanded contour plot of the H6,H8/H2′,H2″ region of the 400-ms NOESY spectrum of the duplex in D_2O at 10 °C.

that corresponds to the G8(H1′) sugar proton. This resonance at 5.87 ppm gives also a cross peak at 7.58 ppm that is assigned to the C9(H6) proton resonance. There is an uncertainty for the assignment of the C9(H1′) proton, but this can be easily resolved by analysis of the H6H8/H2′H2″ and H1′/H2′H2″ regions of the NOESY spectrum of the duplex. The C9(H1′) resonance is found at 6.08 ppm. It is then possible to follow unambiguously the connectivities from G8 to G13 on the basis of the analysis of the H6H8/H1′, H6H8/H2′H2″, and H1′/H2′H2″ regions of the NOESY spectrum and to assign the H5, H6, and H8 base proton resonances and the H1′, H2′, and H2″ sugar proton resonances for strand 2 of the duplex.

For strand 1 of the duplex, the starting point of the assignment is the C1(H6) resonance at 7.48 ppm. Following the connectivities from C1(H6) to G3(H8), we observe the latter at 7.81 ppm, which gives a cross peak at 6.04 ppm that corresponds to the G3(H1′) sugar proton. The resonance at 6.04 ppm gives also a cross peak with a base proton at 7.35 ppm that may be assigned to the G5(H8) base proton. The observation that the chain of connectivities can be followed from G3 to G5, across the abasic site, shows qualitatively that G3 and G5 stack over each other. From the resonance at 7.35 ppm we observe also a cross peak with an H1′ sugar proton at 5.74 ppm that corresponds to the G5(H1′) proton. It is then possible to follow without ambiguity the connectivities from G5 to C7. From analysis of the H6H8/H2′H2″ region of the NOESY spectrum the connectivities can also be followed from C1(H6) to G3(H8) at 7.81 ppm. Analysis of the H2′H2″/H1′ region of a 65-ms mixing time NOESY spectrum gives the relative assignment of the H2′ and H2″ proton resonances. We find the G3(H2′) proton resonance at lower field (2.85 ppm) than the G3(H2″) proton resonance (2.65 ppm). This order is observed for only one other residue, the 3′-terminal G13. The G5(H8) proton at 7.35 ppm gives a cross peak with the H2′ proton of G3 in the 400-ms mixing time NOESY spectrum (Figure 3). This interaction may be strongly influenced by spin diffusion, but it confirms that G5 is stacked over G3. In spectra recorded with short mixing times, the intranucleotide NOEs between the base H8, H6 protons and the H2″ protons are small. In these spectra G5(H8) gives two cross peaks, at 2.55 and 2.65, that correspond to the intranucleotide interaction with the H2′ and the internucleotide

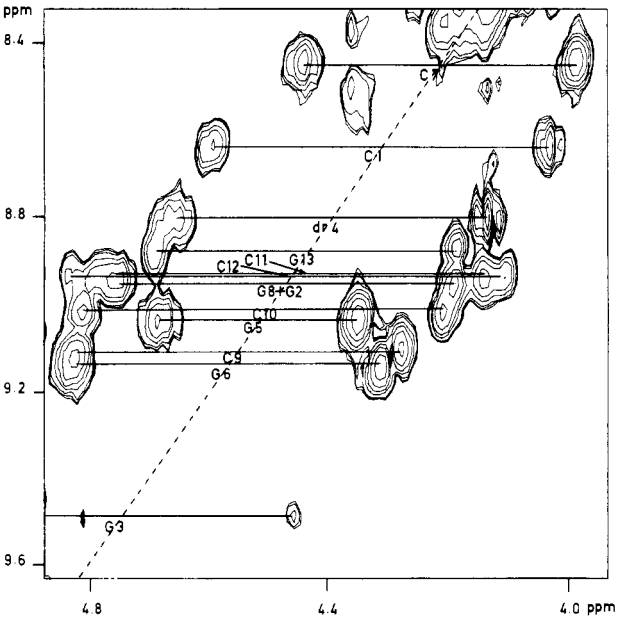

FIGURE 4: Expanded contour plot of the H3′/H4′ region of the double-quantum spectrum in D_2O at 10 °C.

interaction with the G3(H2″).

Analysis of the H2′H2″/H3′ region of the NOESY spectrum gives the assignment of the H3′ sugar proton resonances of the 12 residues. It was not possible to distinguish the H4′, H5′, and H5″ protons of the duplex from analysis of the NOESY spectrum. Thus, we recorded a double-quantum spectrum in D_2O at 10 °C of the duplex (Figure 4). In this spectrum we observe all the cross peaks due to direct connectivities H3′–H4′ for all the nucleotides except G3. The resonances of the H3′ proton of the G3 residue at 5.03 ppm is not observable in this spectrum because of the noise at the frequency of the residual HDO peak. It is well-known that in a double-quantum spectrum when a proton H_a at the frequency Ω_a is J coupled with a proton H_b at the frequency Ω_b, we expect at Ω_a and Ω_b cross peaks at $\Omega_a + \Omega_b$. From the observation of one of the two cross peaks the frequency of the other can be calculated. In the double-quantum spectrum, a resonance at 4.45 ppm gives a cross peak at 9.48 ppm that

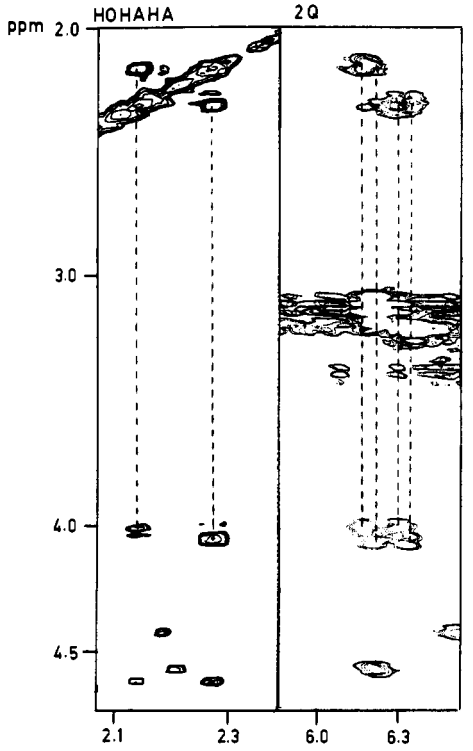

FIGURE 5: Expanded contour plot of the dr4(H1′,H1″,H3′)/dr4-(H2′,H2″) region in both 40-ms mixing time HOHAHA spectrum and double-quantum spectrum recorded with the duplex in D_2O at 10 °C.

corresponds to the sum of 4.45 and 5.03 ppm. As the latter corresponds to the H3′ proton of the G3 residue, the cross peak at 4.45 ppm corresponds to the H4′ resonance of the G3 residue.

Analysis of both the double-quantum spectrum and the NOESY spectrum with a 400-ms mixing time gives the assignment of all the H5′ and H5″ protons of C or G residues of the duplex from the H4′ resonance of each residue. We cannot distinguish the H5′ protons from the H5″ protons. We observe that the H5′ and H5″ proton resonances are resolved only for the G residues and for the C-terminal residues. For the C-nonterminal residues the H5′ and H5″ proton resonances are coincident.

The double-quantum spectrum (Figure 4) shows a resonance at 4.65 ppm, which is *J* coupled with a proton at 4.15 ppm. This pair of resonances does not belong to one of the C or G residues for which all the sugar proton resonances have already been assigned. They must therefore correspond to protons of the dr4.

This is confirmed in the HOHAHA spectra of the duplex that were recorded at 10 °C. This method, which relies on cross polarization, was found to give better results for the dr4 sugar than a COSY experiment at this temperature. We recorded HOHAHA spectra with 10- and with 40-ms mixing times. In the 40-ms mixing time HOHAHA spectrum (Figure 5), we observe a cross peak between the resonances at 2.14 and 2.28 ppm and that each of these resonances is also coupled to a proton at 4.65 ppm. The resonance at 2.14 ppm gives a cross peak at 4.02 ppm, and the resonance at 2.28 ppm gives a cross peak at 4.06 ppm. Further, the double-quantum spectrum (Figure 5) shows that the resonances at 2.14 and 2.28 ppm are both *J* coupled to the resonances at 4.02 and 4.06 ppm. We observe that the proton at 2.14 ppm is *J* coupled to four protons, at 2.28, 4.02, 4.06, and 4.65 ppm, and similarly that the proton at 2.28 ppm is *J* coupled to four protons. Excluding the possibility of four-bond coupling, which we do not observe for any residue in these experiments, the resonances at 2.14 and 2.28 ppm must be assigned to the H2′ and H2″ protons of dr4. Hence, those at 4.02 and 4.06 ppm can be assigned to H1′ and H1″ and that at 4.65 ppm to H3′. We note that these protons are found in a similar chemical shift range to that observed for an intrahelical abasic site (Cuniasse et al., 1987).

In the 50-ms NOESY spectrum we observe that the cross peak between the resonances at 4.65 and 2.14 ppm is much less intense than that between the resonances at 4.65 and 2.28 ppm. In both C2′-endo and C3′-endo conformations the distance H3′–H2′ is smaller than the distance H3′–H2″. Thus, the resonance at 2.28 ppm can be assigned to H2′ of dr4. Similarly, as the distance H1′–H2″ is always smaller than H1″–H2″, we can assign the resonance at 4.02 ppm to H1′ of dr4 and that at 4.06 ppm to H1″. We are able to follow the connectivities in the double-quantum spectrum from the H3′ resonance at 4.65 ppm to the H4′ proton at 4.15 ppm and from this resonance to the H5′/H5″ protons coincident at 4.06 ppm. Table I summarizes the observed chemical shifts for all the proton resonances of the duplex.

Having assigned the spin system of dr4 we have searched for NOE connectivities between these protons and those of the adjacent residues. As stated above, we do not observe any dipolar connectivities between protons of dr4 and base protons. We have therefore searched for connectivities between protons of dr4 and protons of the adjacent sugars. Figure 6 shows the region of the 400-ms NOESY spectrum giving connectivities between H3′ protons and H4′, H5′, and H5″ protons. All H3′ protons show cross peaks with their H4′ protons and in certain cases with H5′/H5″ protons. We observe that G3(H3′) shows

Table I: Chemical Shifts of Nonexchangeable Protons at 10 °C

residue	H8/H6	H5	H1′	H2′	H2″	H3′	H4′	H5′/H5″
C1	7.48	5.72	5.56	1.65	2.20	4.59	4.03	3.68/3.62
G2	7.85		5.33	2.61	2.66	4.75	4.23	4.19/3.81
G3	7.81		6.04	2.85	2.65	5.03	4.45	4.03/3.88
dr4			4.02/4.06	2.28	2.14	4.65	4.15	4.06/4.06
G5	7.35		5.74	2.55	2.66	4.68	4.35	4.14/4.12
G6	7.59		5.97	2.47	2.65	4.81	4.31	4.15/4.11
C7	7.32	5.10	6.09	2.18	2.23	4.45	3.99	4.20/3.74
G8	7.88		5.87	2.63	2.69	4.75	4.19	3.84/3.88
C9	7.58	5.32	6.08	2.32	2.53	4.82	4.29	4.19/4.19
C10	7.54	5.50	5.98	2.20	2.51	4.80	4.21	4.15/4.15
C11	7.44	5.50	5.91	1.96	2.30	4.79	4.14	4.11/4.11
C12	7.49	5.72	5.45	2.08	2.28	4.82	4.11	3.99/3.99
G13	7.90		6.12	2.67	2.38	4.69	4.19	4.04/3.75

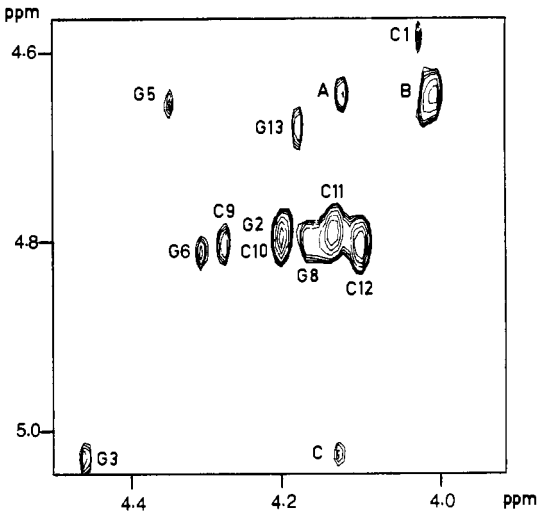

FIGURE 6: Expanded contour plot of the H3′/H4′,H5′,H5″ region of the 400-ms NOESY spectrum of the duplex in D_2O at 10 °C.

a cross peak at 4.15 ppm (peak C, Figure 6) that does not correspond to a proton of this residue. As we find an overall B conformation for the six base pairs, we can exclude that this cross peak arises from an interaction between G3 and G5. It thus must arise from an interaction with dr4, and we have already assigned the H4′ resonance of dr4 at 4.15 ppm. We note that this cross peak remains strong even at short mixing times. We also observe a similarly strong cross peak between a resonance at 4.15 ppm and another at 2.65 ppm (not shown). This could arise from an interresidue interaction dr4(H4′)–G3(H2″) but could also be an intraresidue H2″–H5′/H5″ interaction as several resonances overlap in these regions. In the NOESY spectra recorded at short mixing times, connectivities of the type H2′/H2″–H5′/H5″ disappear as they can only arise from spin diffusion. However, this cross peak remains strong and corresponds to a very short interproton distance. We can thus unambiguously assign it to the interaction dr4(H4′)–G3(H2″). Observation of these two short distances between dr4(H4′) and the sugar of G3 are sufficient to define the location of dr4 (see below). The dr4(H3′) showed cross peaks with the dr4(H4′) and the coincident dr4(H5′, H5″), peaks A and B, respectively (Figure 6).

Interproton Distances. We recorded a series of short mixing time NOESY spectra, and we have measured the NOE buildup rates for all interproton interactions. As described previously (Cuniasse et al., 1987), we calibrate using interactions outside the deformed part of the molecule. The dominant sugar conformation was qualitatively determined by the relative magnitude of the NOE from H6 or H8 to H2′ and from H8 or H6 to H3′ in a NOESY spectrum recorded with a short mixing time. It is well-known that the H6H8–H2′ and H6H8–H3′ intranucleotide distances, differ greatly between the C2′-endo conformation and the C3′-endo conformation. More precisely, the ratio of the distance $R = I(\text{H6H8–H2}')/I(\text{H6H8–H3}')$ is inverse between the C2′-endo conformation and the C3′-endo conformation. This inversion causes a great difference in the ratio R of the NOE intensities between these conformations. The value of R for all nonterminal residues is in the range 3–4, while that of the G5 residue is 0.4. The values do not correspond to the exact values of the sixth power of the distance ratio for a pure C2′-endo or C3′-endo conformation. Although the apparent correlation times for H6H8–H2′ and H6H8–H3′ may not be very different, they are not strictly identical. For a reasonable signal to noise ratio the data are taken from the 80-ms mixing time NOESY spectrum in which spin diffusion is not completely eliminated. In the 50-ms NOESY spectrum the observed values of the ratio R are increased for the G2, G3, G6, C9, C10, C11, and C12 residues, while it is decreased for the G5 residue. Further, the observed value of the ratio R may correspond to an equilibrium between C2′-endo and C3′-endo conformation. Thus, for G2, G3, G6, C9, C10, C11, and C12 we find the major conformational type as C2′-endo, while this is not the case for the G5 residue. Unfortunately, from our NMR data we could not determine the dr4 sugar conformation. From a total analysis of the NMR data, we have been able to evaluate a set of interproton distances that allowed us to define a time-averaged structure of the oligonucleotide.

Model Building and Molecular Mechanics Calculations. It has previously been shown that the combined use of NMR and molecular mechanics calculations is a powerful tool for the determination of oligonucleotide conformation in solution (Gronenborn et al., 1984; Clore et al., 1985; Suzuki et al., 1986; Cuniasse et al., 1987; Zhou et al., 1987). To study by molecular mechanics the oligomer in a conformation in which G3 is stacked over G5, as shown by the NMR data, the extra residues dr4 and phosphate p_4 were introduced into a classical Arnott d(CGGGGC)·d(GCCCCG) structure (Arnott et al., 1976). The two residues, dr4 and p_4, were then pulled ca. 2 Å out of the helix. This creates considerable steric hindrance that can be relieved in two ways. The first consists of rotating dr4 180° about an axis perpendicular to its mean plane. This puts C1′ furthest from the helix. From this starting structure only poor agreement with the NMR data could be attained. Alternatively, dr4 could be rotated sideways 90° about an axis through C2′,C4′, which is roughly parallel to the helix axis, toward G3. This puts the plane of dr4 approximately tangential to the helix. From this initial structure refinement gave good agreement with the NMR data.

While experimenting with the basic structure built above in which dr4 is outside the helix, it rapidly became apparent that the counterions around the dr4 had a propensity for moving off their initial positions, set by the subprogram EDIT or AMBER, to specific locations. The introduction of the extra residues dr4 and p_4 in the regular oligonucleotide d(CGGGGC)·d(GCCCCG) forces the three phosphates p_3, p_4, and p_5 to be much closer to one another than usual. This situation creates potential counterion traps. The presence or absence of counterions does not change the overall structure except in the region $p_3$dr4p_4G5p_5. These interactions could be important in the crystalline state but could be attenuated in solution as molecular Brownian motion could displace them off specific locations. We observe (see below) that structures with and without ions are required to explain NMR data, and thus both structures have been refined. Oligonucleotide models described as "without ions" serve as a reference to understand the effects of internal forces that determine their geometry. Oligonucleotide models described "with ions" allow evaluation of the influence of counterions and more generally of solvent molecules that may play a similar role by bridging strongly charged residues.

We have investigated the effect of the extra sugar phosphate on the helical twist at the junction and on the pucker of the G5 residue as the NMR data showed that this sugar does not adopt a typical C2′-endo conformation, but may have a significant C3′-endo contribution. These two variables are not independent. When a nucleotide changes from a C2′-endo conformation to a C3′-endo conformation, the torsion angle δ (C5′–C4′–C3′–O3′) decreases and, to a first approximation, the 5′ phosphate residue moves outside the helix and the helical twist is modified (Dickerson et al., 1985). To study completely the effect of introducing dr4 and p_4 into the structure, we have carried out the same calculations varying the sugar pucker of G3. Three series of models have been studied with different initial structures in which the sugar puckers are fixed by constraining the angle δ. In series A, G3 is C3′-endo ($\delta = 82°$) and all other nucleotides are C2′-endo ($\delta = 144°$); for series B, G5 is C3′-endo, all others C2′-endo; and for series C, all nucleotides are C2′-endo. For all three models an initial helical twist, θ, between G3 and G5 was imposed. This was varied in steps of 5° between 21° and 76°.

This operation distorts phosphate oxygen bond lengths, and thus the refinement was carried out in three steps. The first involved only the central two base pairs, dr4 and the neighboring phosphate residues (residues p_2–p_5 and p_3–p_{11}). This was followed by refinement of the entire oligonucleotide. Finally, refinement was carried out with the sugar pucker constraints relaxed. The entire procedure was carried out twice, either with counterions or in the absence of counterions. The structures obtained depended upon the starting point

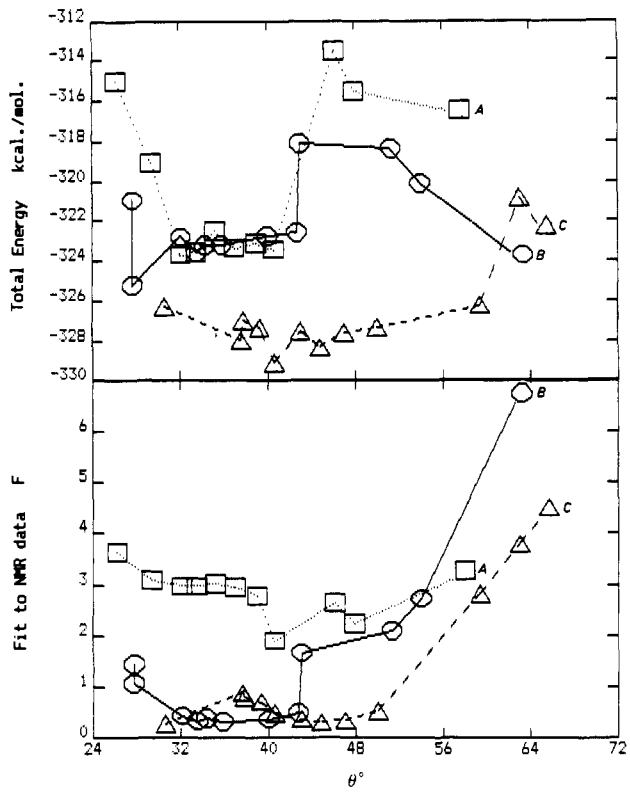

FIGURE 7: Variations of total energy of the oligonucleotide d-(CGGdrGGC)·d(GCCCCG) with θ (top) and corresponding values of the NMR distance fit criterium F (bottom). Counterions were not included in the final refinement with θ. Series A, G3 is C3′-endo, all other C2′-endo; series B, G5′ is C3′-endo, all others C2′-endo; series C, all C2′-endo.

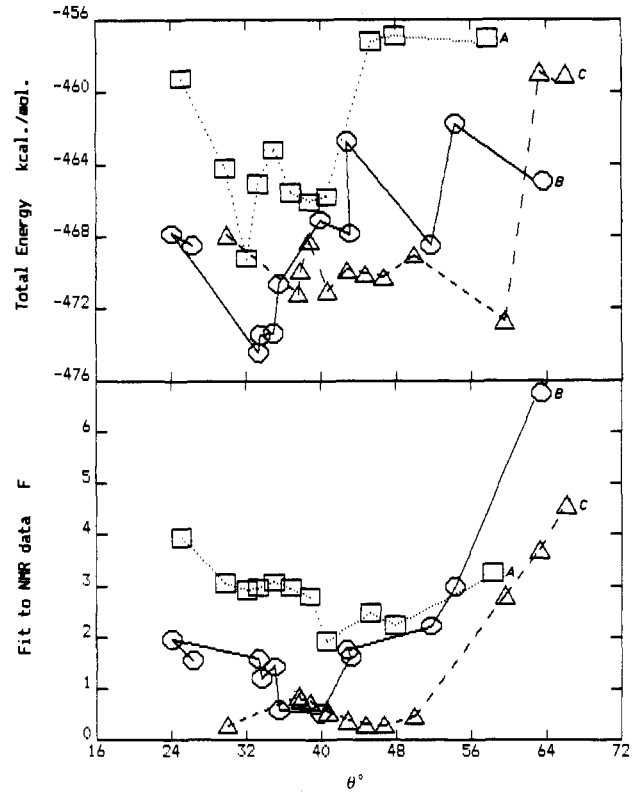

FIGURE 8: Variations of total energy of the oligonucleotide d-(CGGdrGGC)·d(GCCCCG) with θ (top) and corresponding values of the NMR distance fit criterium F (bottom). Counterions were included in the final refinement with θ. Series A, G3 is C3′-endo, all others C2′-endo; series B, G5′ is C3′-endo, all others C2′-endo; series C, all C2′-endo.

because of the multiplicity of local minima. It does, however, allow us to explore a variety of oligonucleotide conformations.

The successive refinements affected not only the two central base pairs but also the base pairs on either side that slid upon G3·C11 or G5·C10. The twist of G2·C12 relative to G3·C11 and that of G5·C10 relative to G6·C9 fluctuated in an apparently random way with a root mean square deviation of 2°.

In the absence of counterions we observed that series C showed the lowest energy values (Figure 7), and over a wide range of θ, indicating lower free energy. This series showed a good fit to the NMR distance measurements. Starting from the same imposed helical twist, after refinement, series C showed typically a ca. 5° shift relative to series A and B.

In the presence of counterions the lowest energy was observed for series B with θ [33°, 35°] (Figure 8); however, more scatter is observed than in Figure 7. Generally, series C fits the NMR data better. The best fit in series B is with θ [36°, 40°]. We observed that these two structures converged to an O1′-endo conformation if the pucker angle of G5 was relaxed. It is noteworthy that the O1′-endo conformation is halfway between C2′-endo and C3′-endo in the pseudorotation pathway (Altona & Sundaralingam, 1973) and is the obligatory transition between the two canonical forms. Only in these two cases did a sugar pucker change during the course of the refinement. We note that the value of F in Figures 7 and 8 in the range θ [27°, 42°] for series A and B and θ [33°, 47°] for series C is determined largely by contributions from residues G3, dr4, and G5. From both the criteria of total energy and fit to the NMR data we find that θ must lie in the range [26°, 50°], indicating near-normal stacking of G3 over G5. The uncertainty in the helical twist is similar to the variations observed in the dodecamer of Dickerson (Fratini et al., 1982) and in the global twist in short DNA (Shore & Baldwin, 1983).

Table II: Interphosphate Distances (Å) Less than 9.0 Å

distance	S1 with counterions	S2 without counterions
p_1–p_2	7.1	7.0
p_2–p_3	6.3	6.3
p_2–p_4	8.8	8.8
p_3–p_4	5.2	5.0
p_3–p_5	8.5	9.1
p_4–p_5	5.1	6.5
p_5–p_6	7.1	7.0
p_8–p_9	6.9	7.0
p_9–p_{10}	7.1	7.1
p_{10}–p_{11}	7.0	7.0
p_{11}–p_{12}	7.0	7.0

The NMR data showed that the sugar of G5 does not adopt a normal C2′-endo pucker. We have therefore varied the torsion angle δ of G5 and calculated the total energy of the system with and without counterions (Figure 9). Without counterions the minimum was observed for a C2′-endo conformation. In the presence of counterions the best minimum corresponded to a C3′-endo conformation with a local minimum at C2′-endo. In all cases the preferred conformation of dr4 was C2′-endo.

To obtain a stable C3′-endo conformation for G5 it was necessary to include counterions during the refinement. The best structure in this series, S1, is shown in Figure 10. Forcing G5 into a C3′-endo conformation pulls p_4 close to p_5. This interphosphate distance is 5.1 Å (Table II). The close contact is stabilized by the presence of two counterions that salt-bridge the two phosphates rather than being each associated with one phosphate. Further, the sodium ion of p_3 moves to bridge the phosphate and O1′ of dr4. These two locations correspond

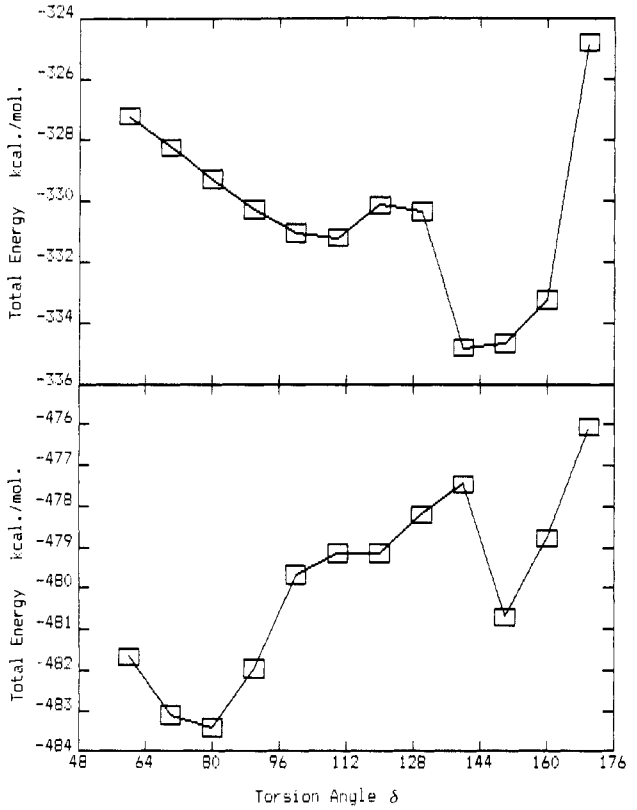

FIGURE 9: Total energy of d(CGGdrGGC)·d(GCCCCG) calculated as a function of the torsion angle δ of G5. Counterions were not (top, $\theta = 39°$) or were (bottom, $\theta = 34°$) included in the final refinement.

Table III: Characteristics of Structures S1 and S2[a]

residue	pucker	amplitude	α	β	γ	δ	ϵ	ζ	χ
Model S1: B Series, $\theta = 34$, with Counterions, $F = 1.2$									
G3	146	38	−63	−177	55	135	−81	−93	−110
dr4	161	40	68	−95	−73	150	−97	49	
G5	47	37	−152	−84	78	77	−172	−78	−163
G6	170	36	−63	179	67	145	178	−96	−108
Model S2: C Series, $\theta = 39$, without Counterions, $F = 0.7$									
G3	141	38	−66	−178	56	131	−68	−124	−109
dr4	10	40	87	−75	−63	81	−178	168	
G5	166	34	−59	−149	61	142	−175	−93	−128
G6	152	39	−72	−179	55	136	175	−101	−105

[a] For both structures, S1 and S2, the remaining residues take standard B DNA values. The best fit to NMR data was found with a 1:1 ratio of the two structures, $F = 0.33$.

to the major counterion traps observed. This structure gives a fit to the NMR data with $F = 1.2$. The torsion angles for the segment G3·G6 are shown in Table III.

Refinement in the absence of counterions and G5 in C2′-endo conformation leads to structure S2 (Figure 10). The short interphosphate interaction p_4–p_5 is relaxed in this structure (Table II). The fit to the NMR data is better than with S1 as we find $F = 0.7$ (Table III). Addition of the extra residues, dr4 and p_4, lengthens the sugar phosphate backbone. If the backbone were straight, θ would have had to increase. To maintain an overall B DNA conformation for the base pairs, the extra residues have to be pulled out of the helix and dr4 has to be turned such that the plane of the sugar is roughly tangential to the helix. The mean plane of dr4 is turned by ca. 70° between S1 and S2. For both structures the sugar backbone is forced to loop backward on both sides of dr4, bringing p_3 close to p_4. The loop about p_3 is tighter than that about p_4 (Figure 10). If G3 is forced into a C3′-endo conformation, the loop about p_3 is constrained to be even tighter, and this unfavorable situation is reflected in Figures 7 and 8.

Table IV: Proton–Proton Distances Determined by NOE and Computed for Two Structures, S1 and S2, and Their 1:1 Mixture

distance	NMR	S1	S2	S1/S2 = 1	% error
G3(H8)–G3(H1′)	3.7	3.93	3.93	3.93	6
G5(H8)–G5(H1′)	3.4	3.78	3.92	3.85	13
G6(H8)–G6(H1′)	3.8	3.92	3.93	3.92	3
C9(H6)–C9(H1′)	3.7	3.75	3.75	3.75	1
C10(H6)–C10(H1′)	3.9	3.76	3.76	3.76	4
C11(H6)–C11(H1′)	3.6	3.76	3.76	3.76	4
C12(H6)–C12(H1′)	3.8	3.76	3.77	3.76	1
G3(H8)–G3(H2′)	2.4	2.40	2.37	2.38	1
G5(H8)–G5(H2′)	3.0	3.86	2.68	2.97	1
G6(H8)–G6(H2′)	2.3	2.32	2.31	2.32	1
C9(H6)–C9(H2′)	2.0	2.28	2.29	2.29	14
C10(H6)–C10(H2′)	2.1	2.22	2.23	2.22	6
C11(H6)–C11(H2′)	2.0	2.26	2.26	2.26	13
C12(H6)–C12(H2′)	2.0	2.22	2.17	2.20	10
G5(H8)–G3(H1′)	3.6	3.76	3.77	3.76	4
G6(H8)–G5(H1′)	4.4	5.32	4.08	4.45	1
C10(H6)–C9(H1′)	3.5	3.45	3.47	3.46	1
C11(H6)–C10(H1′)	3.6	3.32	3.32	3.32	8
C12(H6)–C11(H1′)	3.5	3.44	3.60	3.51	0
G5(H8)–G3(H2″)	2.8	2.45	3.08	2.64	5
G6(H8)–G5(H2″)	2.2	3.66	2.22	2.48	12
C10(H6)–C9(H2″)	2.3	2.29	2.29	2.29	0
C11(H6)–C10(H2″)	2.4	2.28	2.31	2.30	4
C12(H6)–C11(H2″)	2.2	2.26	2.27	2.27	3
dr4(H4′)–G3(H2″)	2.4	2.61	3.58	2.86	19
dr4(H4′)–G3(H3′)	2.4	2.28	2.27	2.28	5

Neither structure S1 nor S2 provides a totally satisfactory fit to the NMR data. For the central part of strand 1 several distances calculated from the NMR data show poor agreement with either structure, although in the opposite sense (Table IV). We stated above that an O1′-endo conformation for G5 fits well the NMR data, although not all of them. The observed intranucleotide NOEs between G5(H8) and H2′,H3′ fit poorly those expected for a pure O1′-endo conformation. While this conformation would be unusual for a nucleotide, a contribution from this conformation cannot be excluded here. We have tried another approach to fit the data starting from the more commonly observed C2′-endo and C3′-endo conformations. The NMR data do not suggest a dominantly C3′-endo conformation for G5 but, at least, a significant contribution. The distance errors for S1 and S2 suggest that the NMR data may correspond to a time-averaged structure of the two. We have varied the population ratio of the two structures, and we observe a significantly better fit, $F = 0.33$, with 50 ± 20% C3′-endo (Table IV). The model with counterions located at specific positions and that without counterions may represent two extremes of a large family. In reality, Brownian movement could displace ions off specific positions, averaging their effect. The proposed C2′-endo–C3′-endo equilibrium only involves conformational changes in the central region (results shown in Table III) and stacking interactions are highly conserved, as shown by the superposition of the two structures (Figure 10). The choice of averaging C2′-endo and C3′-endo conformations is, of course, arbitrary. While we observed a poor fit for the NMR data to any single sugar conformation, the time-averaged conformation we observe may include a contribution from an O1′-endo conformation.

Conclusion

While the presence of the abasic frameshift strongly destabilizes the helix, we find that the six base pairs adopt a conformation very close to that of B DNA. From the distances measured by NMR the stacking of G3 over G5 and of C10

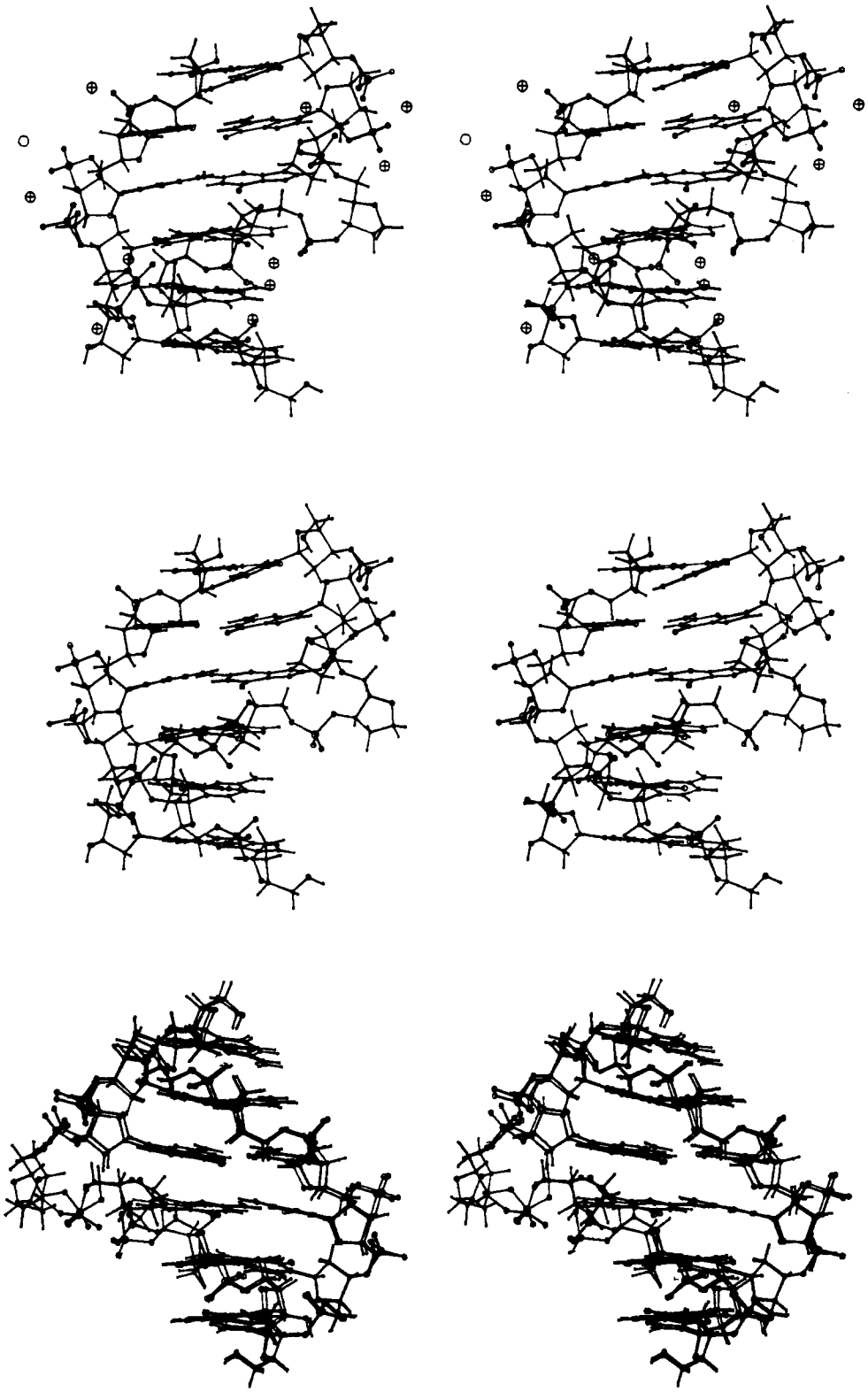

FIGURE 10: Stereoscopic view of the structure S1, in the presence of counterions (top), viewed from the minor groove; structure S2, viewed from the minor groove (middle); superposition of S1 and S2, viewed from the major groove (bottom).

over C11 is normal, very little influenced by the presence of dr4. The energy associated with base stacking is such that dr4 lies outside the helix. The helical twist at the junction G3–G5 is close to a normal value. We are unable to detect any conformational change on strand 2 relative to normal B DNA. On strand 1 the only modification occurs for the sugar of G5. Our data show clearly that this sugar does not adopt uniquely a C3′-endo or C2′-endo conformation. The H3′ is slightly closer to the base H8 proton than is the H2′ proton. This does not correspond to any reasonable known stable sugar conformation. It rather indicates that the usual C2′-endo–C3′-endo sugar conformation equilibrium has been displaced and a greater than normal fractional contribution of C3′-endo conformation is present. We have not attempted to calculate the position of the equilibrium on the basis of the NOEs, as the effective correlation time for the H8–H3′ interaction is uncertain. Rather, we have varied the contribution of the two conformations, S1 and S2, and calculated the fit to the NMR

data. The fit is found to be relatively sensitive to the position of the equilibrium, with the best fit being found at 50 ± 20% C2′-endo.

In general, the combination of NMR distance geometry calculations and molecular mechanics calculations gives rise to a family of fairly closely related structures. When this is done, that is, searching for the best single minimum, we find for the sugar of G5 a C2′-endo and possibly an O1′-endo conformation that, as stated above, we do not favor. The conformational equilibrium that we propose for the sugar of G5 has a consequence on the position of dr4 relative to the helix. The NMR data show two short proton–proton interactions between dr4(H4′) and H2″ and H3′ of G3, but we are unable to define the conformation of the ring of dr4 from NMR data.

We have also observed that the most energetically favored conformation for the sugar of G5 is dependent upon whether phosphate charge neutralization is achieved by counterions at defined positions or not. It is clear that the presence of counterions (or solvent interactions) is necessary for stabilizing the C3′-endo conformation of G5 as in this case the interphosphate distance p_4–p_5 is unusually short. We do not suggest that in solution counterions have defined positions. Even though mobile, they may occupy these positions part of the time, which would be in agreement with the observed conformational equilibrium of the sugar of G5.

Acknowledgments

We are most grateful to Andrée Dutreix for access to the Lexidata system. We thank Karel Zimmerman and Jacques Gabarro for their assistance during this project and Werner Leupin for much appreciated constructive criticism.

Registry No. Duplex, 118139-70-1.

References

Altona, C., & Sundaralingam, M. (1973) *J. Am. Chem. Soc. 95*, 2333–2344.

Arnott, S., Campbell-Smith, P., & Chandrasekharan, P. (1976) *CRC Handb. Biochem. 2*, 411–414.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.

Calos, M. P., & Miller, J. H. (1981) *J. Mol. Biol. 153*, 39–66.

Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin, W. (1986) *J. Mol. Biol. 190*, 439–453.

Clore, G. M., Gronenborn, A. M., Moss, D. S., & Tickle, I. J. (1985) *J. Mol. Biol. 185*, 219–226.

Cuniasse, Ph., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., Guschlbauer, W., & Fazakerley, G. V. (1987) *Nucleic Acids Res. 15*, 8003–8022.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

Dickerson, R. E., Kopka, M., & Pjura, P. (1985) *Biol. Macromol. Assem. 2*, 37–126.

Doubleday, O. P., Lecomte, P., & Radman, M. (1983) in *Cellular Response to DNA Damage* (Friedberg, E. C., & Bridges, B. A., Eds.) pp 489–499, Alan Liss, New York.

Eritja, R., Walker, P. A., Randall, S. K., Goodman, M. F., & Kaplan, B. E. (1987) *Nucleosides Nucleotides 6*, 803–814.

Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983) *Biochemistry 22*, 5943–5951.

Fink, T. R., & Crothers, D. M. (1972) *J. Mol. Biol. 66*, 1–12.

Fratini, A. V., Kopka, M. L., & Drew, H. R. (1982) *J. Biol. Chem. 257*, 14686–14707.

Fréchet, D., Cheng, D. M., Kan, L. S., & Ts'o, P. O. P. (1983) *Biochemistry 22*, 5194–5200.

Gait, M. J. (1984) *Oligonucleotide synthesis, a practical approach*, IRL Press, Oxford, U.K.

Gelin, B., & Karplus, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 77*, 801–805.

Gronenborn, A. M., Clore, G. M., & Kimber, B. J. (1984) *Biochem. J. 221*, 723–736.

Hare, D., Shapiro, L., & Patel, D. L. (1986) *Biochemistry 25*, 7456–7464.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 181*, 319–336.

Kumar, A., Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654–3658.

Kunkel, T. A., & Loeb, L. A. (1979) *J. Biol. Chem. 254*, 5718–5725.

Levitt, M., & Warshel, A. (1978) *J. Am. Chem. Soc. 100*, 2607–2613.

Lindahl, T., & Nyberg, B. (1972) *Biochemistry 11*, 3610–3618.

Lindahl, T., & Karlström, O. (1973) *Biochemistry 12*, 5151–5154.

Raap, J., Dreef, C. E., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1987) *J. Biomol. Struct. Dyn. 5*, 219–247.

Scheek, R. R., Boelens, R., Russo, N., van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371–1376.

Shearman, C. W., & Loeb, L. A. (1979) *J. Mol. Biol. 128*, 197–218.

Shore, D., & Baldwin, R. L. (1983) *J. Mol. Biol. 170*, 957–982.

Singh, U. C., Weiner, P. K., Caldwell, J. W., & Kollman, P. A. (1986) AMBER 3.0, University of California, San Francisco.

Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–6865.

Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. K. (1984) *J. Am. Chem. Soc. 106*, 765–784.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230–252.

Zhou, N., Bianucci, A. M., Pattabiraman, N., & James, T. L. (1987) *Biochemistry 26*, 7905–7913.