Sayre, M. H., & Geiduschek, E. P. (1988) J. Virol. 62, 3455–3462.
Skibsted, L. H., Hancock, M. P., Magde, D., & Sexton, D. A. (1987) Inorg. Chem. 26, 1708–1712.
Szabo, A. G., & Rayner, D. M. (1980) J. Am. Chem. Soc. 102, 554–563.
Tanaka, I., Appelt, K., Dijk, J., White, S. W., & Wilson, K. S. (1984) Nature (London) 310, 376–381.

Multinuclear NMR Studies of DNA Hairpins. 1. Structure and Dynamics of d(CGCGTTGTTCGCG)†

James R. Williamson‡ and Steven G. Boxer*
Department of Chemistry, Stanford University, Stanford, California 94305
Received September 8, 1988; Revised Manuscript Received December 8, 1988

ABSTRACT: The solution structure of the hairpin formed by d(CGCGTTGTTCGCG) has been examined in detail by a wide variety of NMR techniques. The hairpin was characterized by proton NMR to obtain interproton distances and torsion angle information. An energy-minimized model was constructed that is consistent with these data. The hairpin consists of a B-DNA stem of four C-G base pairs and a loop region consisting of five unpaired bases. Three bases in the 5′ of the loop are stacked over the 3′ end of the stem, and the other two bases in the 3′ of the loop are stacked over the 5′ end of the stem. The phosphorus NMR spectrum revealed a phosphate in the stem region with an unusual conformation, and two phosphates, P_9 and P_{10}, were found to undergo intermediate exchange between conformations. The hairpin was also synthesized with a carbon-13 label in each of the thymidine C6 carbons, and relaxation measurements were performed to determine the extent of internal motions in the loop region. The loop bases are more flexible than the stem bases and exhibit subnanosecond motions with an amplitude corresponding to diffusion in a cone of ~30°.

Hairpins are secondary structural elements for nucleic acids that are found in many structural and functional contexts. Hairpin loops are a particularly important feature of ribonucleic acids. The role of the hairpin loop is to negotiate a complete turn to allow base pairing of a single strand with itself. The properties of loop regions are of interest because they can serve as a site for tertiary interactions or as a recognition site for other molecules.

Advances in deoxyribonucleotide synthesis have made possible the preparation of large quantities of material of defined sequence required for NMR[1] studies (Gait, 1984). Only very recently has progress been made toward the routine synthesis of RNA oligonucleotides by chemical synthesis (Caruthers et al., 1987; Usman et al., 1987) and by in vitro transcription (Milligan et al., 1987). The first studies on nucleic acid hairpins were necessarily made on DNA oligomers as the best available model system for RNA structure. DNA hairpins have been observed from self-complementary oligonucleotides at low concentration (Wemmer et al., 1985; Patel et al., 1982), and the effect of loop size on thermodynamic stability of a series of hairpins has been investigated (Haasnoot et al., 1980, 1983). Distance geometry methods were used in conjunction with NMR to solve the solution structure of the hairpin formed by d(CGCGTTTTCGCG) (Hare & Reid, 1986). An NMR study of a similar hairpin, d(CGCGCGTTTTCGCGCG), has also been performed (Ikuta et al., 1986). An interesting hairpin containing only two bases in the loop region has recently been investigated (Orbons et al., 1987a,b).

We report here a detailed NMR investigation of the structure of the DNA oligomer d(CGCGTTGTTCGCG) that also adopts a hairpin conformation. This oligomer is similar in sequence to those described above, but is unique in that there are five bases in the loop region, one of which is a G residue. This sequence was originally designed to test the solution accessibility of guanosine residues in hairpin loops by using photo-CIDNP (McCord et al., 1984a,b). Initial studies on this molecule revealed several anomalous features in the NMR spectra, so we have attempted to characterize the possibilities of conformational flexibility and internal motions by as many NMR methods as possible. The oligomer has been characterized by proton NMR to provide interproton distances and torsion angle information and by phosphorus NMR to examine phosphate conformations. To probe dynamics in the loop region, the oligomer was resynthesized with thymidine residues labeled at C6 with carbon-13, and the relaxation behavior of the carbons was examined. A less detailed comparative study of two hairpins of similar sequence, d(CGCGTTTGTCGCG) and d(CTGCTCTTGTTGAGCAG), is presented in an accompanying paper. These sequences are related to that of the present study in that they contain the same stem and loop sequences, respectively, but differ in the remainder of the sequence. These changes lead to surprisingly large changes in the hairpin structure.

Materials and Methods

The DNA molecules d(CGCGTTGTTCGCG) and d(CGCGT*T*GT*T*CGCG), in which the thymidine C6

† This work was supported by a grant from the National Institutes of Health (R01 GM27738). S.G.B. is the recipient of a Presidential Young Investigator Award.
‡ Present address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309.

[1] Abbreviations: HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; NOESY, 2D NOE spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; NOE, nuclear Overhauser enhancement; AMBER, assisted model building with energy refinement; TMP, trimethyl phosphate; CSA, chemical shift anisotropy; T, Tesla; rms, root mean square.

0006-2960/89/0428-2819$01.50/0

carbons have been labeled with ^{13}C (T*), were synthesized by the phosphoramidite method on a home-built fully automated synthesizer (Williamson, 1988) and were purified by reverse-phase HPLC. Samples were desalted on a low-pressure reverse-phase column, the desired buffers were added, and the solvent was exchanged for D_2O by repeated concentration from 99.8% D_2O on a Speed-Vac. NMR samples were prepared by dissolving the exchanged sample in 99.996% D_2O in a glovebag under nitrogen. The pH of the sample was measured in the NMR tube with a microelectrode, and the reported pHs are not corrected for the effects of isotope or temperature.

NMR Measurements. Proton spectra were recorded at 500 MHz, phosphorus spectra were recorded at 202 MHz, and carbon spectra were recorded at 125 MHz on a GE GN-500 spectrometer. Carbon spectra were also recorded at 75 MHz on a Nicolet NT-300 spectrometer. Proton chemical shifts are referenced to (trimethylsilyl)propionic-d_4 acid internal standard, phosphorus shifts are referenced to trimethyl phosphate external standard, and carbon shifts are referenced to dioxane external standard.

Spectra in H_2O were measured with a 1331 pulse sequence (Hore, 1983) and were corrected for base-line roll. NOESY spectra were recorded with 1024 complex points in t_2 and 256 points in t_1 by using hypercomplex phase cycling (States et al., 1982). NOESY spectra for quantitation of NOEs were apodized with a pure Gaussian function in both dimensions and zero-filled to 2048 points. The DQF-COSY spectrum with ω-scaling in ω_1 was recorded with 4096 complex points in f_2 and 512 points in f_1 (Chazin et al., 1986). The spectrum was apodized with a 1-Hz Gaussian function in t_2 and with a sine bell shifted by $\pi/3$ in t_1 followed by zero filling to 2048 points. The final resolution was 0.75 Hz/point in f_2 and 1.5 Hz/point in f_1. Phosphorus–proton correlation was performed by using a proton-detected variation of the standard heteronuclear correlation experiment (Sklenar et al., 1986).

Carbon T_1 measurements were performed with the inversion recovery method. A full $10T_1$ relaxation delay was inserted between successive scans. T_1 values were obtained from a three-parameter fit with an extra parameter to allow for incomplete inversion. NOE measurements were made by interleaving scans acquired with and without NOE, allowing $5T_1$ relaxation between acquisitions. The spectra were transformed identically and then fit manually to Lorentzian lines by using the spectrum simulation software on the spectrometer. The peak areas from the fits were used to calculate the NOEs.

NMR data were transferred to a DEC VAX 11/785 computer for processing using the FTNMR program written by Dennis Hare. Energy minimization calculations with the program AMBER were performed on a μ-VAX-II, or a VAX-8550, or on the Cray XMP-II at the San Diego Supercomputer Center. Energy minimizations were performed in the absence of solvent molecules using the all-atom force field with a nonbonding interaction cutoff of 10 Å. Counterions were placed at the bisector of the phosphate oxygens before minimization and were replaced if they strayed from their phosphates, followed by reminimization. Interproton distance restraints were included as harmonic pseudobonding potentials with a force constant of 100 kcal/Å^2.

Results

Proton NMR Data

Assignment of the Proton NMR Spectrum. The 500-MHz proton NMR spectrum of d(CGCGTTGTTCGCG) at 25 °C in D_2O is shown in Figure 1. The spectrum shows one set of resonances corresponding to the 13 residues in the molecule,

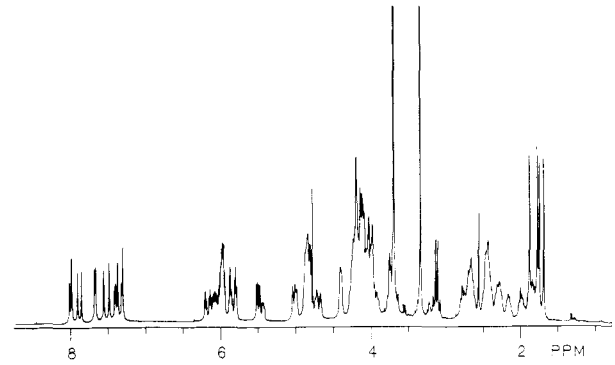

FIGURE 1: 500-MHz ^1H NMR spectrum of d(CGCGTTGTTCGCG) in D_2O at 25 °C. Sample conditions are 70 mM NaCl, 35 mM sodium phosphate, pH 7.0, 5 mM EDTA, and 3 mM oligonucleotide. The residual HDO peak was reduced by presaturation, and the two large singlets between 3 and 4 ppm are EDTA protons.

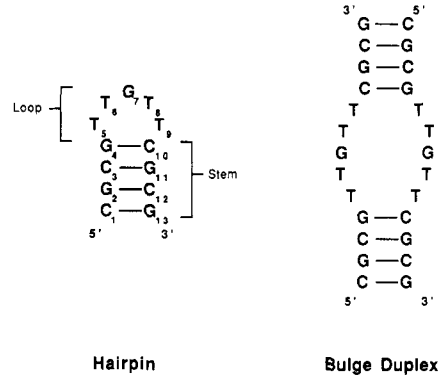

FIGURE 2: Two possible structures for the oligonucleotide d(CGCGTTGTTCGCG).

indicating that it exists as either a monomer or a symmetric dimer. The base proton, H1′, H2′, H2″, and H3′ resonances were completely assigned by using a combination of NOESY and COSY spectra following the sequential assignment procedure (Scheek et al., 1983, 1984; Hare et al., 1983; Wüthrich, 1986). Expansions of the regions in the NOESY spectrum used for assignment are shown in Figures S-1–3 and chemical shifts are given in Table S-I of the supplementary material.

Determination of the Molecular Form. There are two molecular forms that this molecule might adopt: a bimolecular complex, or bulge duplex, and a unimolecular hairpin, shown schematically in Figure 2. The two possible forms can be distinguished by measuring the rotational correlation time of the molecule. The cross-relaxation rate between the cytosine H5 and H6 protons is proportional to the correlation time for reorientation of the internuclear vector. A plot of the NOE buildup on the H5 protons when H6 is irradiated is shown in Figure S-4 (of the supplementary material) for all four cytosine residues at 25 °C. The cross-relaxation rate is nearly the same for all four residues, and the average value is -0.33 ± 0.03 s^{-1}, which corresponds to a correlation time of 1.4 ± 0.1 ns. This is consistent with a sphere of radius 11.2 Å, which corresponds to the approximate dimensions of the hairpin structure. This value is close to the 2.0 ns (at 18 °C) observed for the slightly larger hairpin formed by d(CGCGCGTTTTCGCGCG) by NMR relaxation (Ikuta et al., 1986). The bulge duplex structure can be approximated by a prolate ellipsoid of axial ratio 2:1. The correlation time for such an ellipsoid is equal to the correlation time of a sphere of equivalent volume times a shape factor, which for a 2:1 axial ratio is nearly unity. The correlation time for bulge duplex structure would be expected to be roughly twice that of the hairpin.

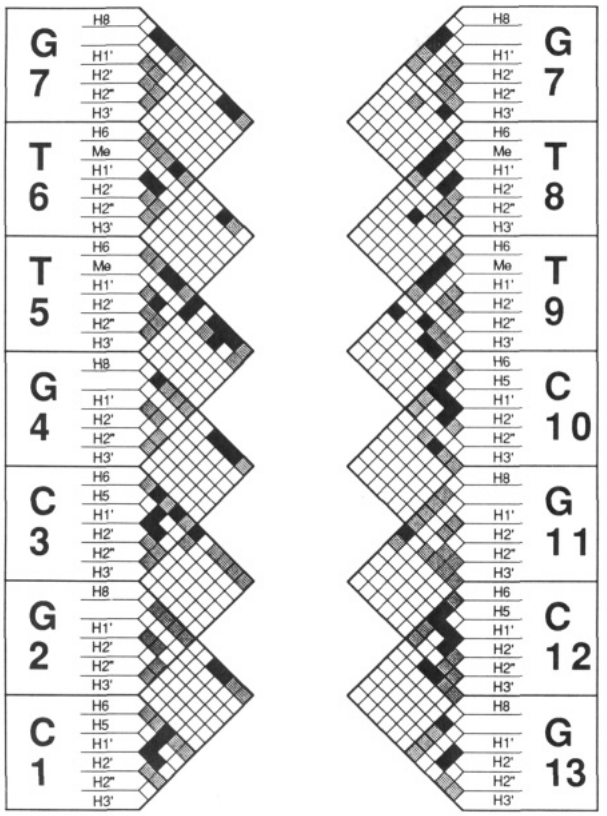

FIGURE 3: Summary of the NOE contacts observed for d(CGCGTTGTTCGCG) at 25 °C. Intraresidue contacts are contained in the triangular area for each base, and the interresidue contacts are contained in the square area between two bases. Small squares filled with black represent NOEs that were quantitated in a 2D NOE buildup to give interproton distances. Gray squares represent NOEs that were observed but could not be quantitated due to overlap of cross peaks. At least two NOEs were observed between each adjacent pair of bases.

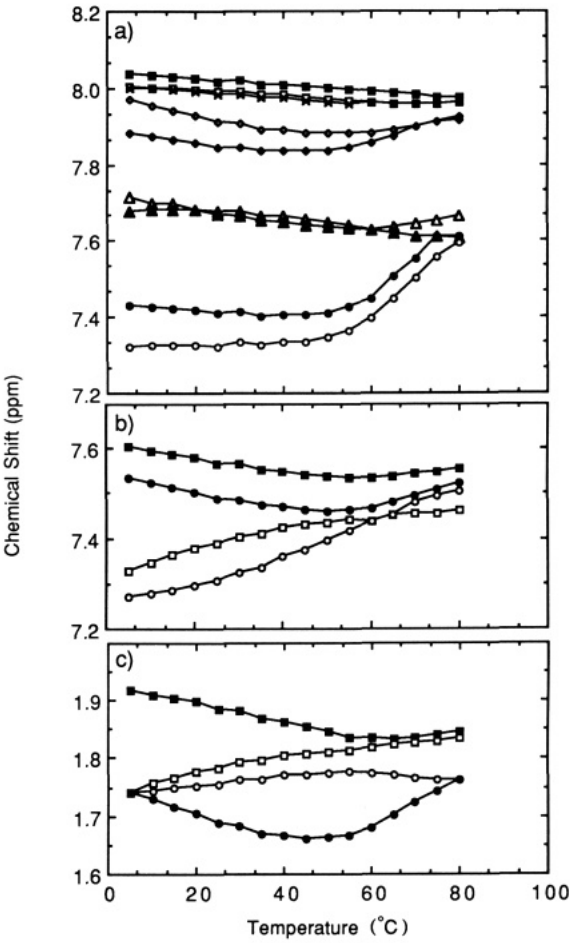

FIGURE 4: Temperature dependence of the chemical shifts of selected protons in d(CGCGTTGTTCGCG). (a) Guanosine H8 protons (■) G_2, (□) G_{11}, (×) G_{13}, (◇) G_4, (◆) G_7; cytosine H6 protons (△) C_1, (▲) C_{10}, (●) C_{12}, (○) C_3. (b) Thymidine H6 protons. (c) Thymidine methyl protons (■) T_9, (●) T_8, (□) T_6, (○) T_5.

The sequential pattern of NOEs can be followed from the 5′ residue C_1 through residue G_7. In each step, an NOE is observed between the H6 or H8 and the H2″ of the preceding residue, and a similar connectivity is observed from residues T_8 through G_{13}. However, the two contacts between residues G_7 and T_8 are atypical. One contact is observed between the T_8 H6 and the G_7 H2″, but this NOE is extremely weak. The other contact between T_8 CH_3 and G_7 H1′ is not unusual but is usually accompanied by a contact to H6 adjacent to the methyl. The turn in the loop of the hairpin appears to occur between these two residues. The pattern of NOEs is consistent with a duplex in the B-DNA conformation in part of the molecule; however, a modification of the normal sequential connectivity is observed between residues G_7 and T_8. A summary of the observed NOE contacts is presented in Figure 3.

Temperature Dependence of the Nonexchangeable Resonances. The temperature dependence of the chemical shifts for the base proton resonances and the thymidine methyl groups is shown in Figure 4. Two of the cytosine H6 resonances show a large downfield shift between 50 and 80 °C, which is characteristic of the melting of a double-helical structure. Similar shifts with temperature are observed in d(CGCG) and d(CGCGCG) (Uesugi et al., 1981; Frechet et al., 1983; Cheng et al., 1984). The endpoint of the transition is above 80 °C, but a lower limit on the estimated melting temperature for the hairpin structure is ~70 °C. In contrast to the shifts of the cytosine H6 protons, which are essentially constant below the melting temperature, the shifts of the thymidine H6 protons change significantly between 10 and 50 °C. The H6 protons of residues T_5 and T_6 show a downfield shift, while those of residues T_8 and T_9 exhibit an upfield shift with increasing temperature. The corresponding methyl resonances parallel this trend. The thymidine H6 and methyl resonances exhibit an upfield shift with increasing temperature well below the melting temperature, suggesting a premelting conformational transition in the loop region occurring in the temperature range 0–50 °C. The chemical shift of the H1′ proton on residue C_{10} also shows a strong temperature dependence in this range, moving downfield from 5.2 to 5.8 ppm. The melting behavior was identical at 0.2 and 2 mM oligonucleotide concentration (data not shown), which is further evidence for existence of the hairpin conformation.

Exchangeable Protons. The 500-MHz proton NMR spectrum in 95% H_2O/5% D_2O at 5 °C is shown in Figure 5 at pH 6.9, 5.8, and 4.7. Three distinct groups of exchangeable protons are visible. A group of four protons at ~13 ppm corresponds to the guanosine imino protons involved in base pairs. Three sharp resonances are observed, and there is a fourth broad resonance underneath them. A second group of five protons at ~11 ppm corresponds to the unpaired imino protons of the thymidines and the guanosine in the loop region. The third group is spread from 6 to 9 ppm and represents the amino protons on the cytosine residues. The imino and amino protons involved in base pairing were assigned by selective NOE experiments (data not shown). Irradiation of a guanosine amino proton produced an NOE to the base-paired amino proton on the cytosine residue opposite. Irradiation of the base-paired imino proton produced an NOE to the un-

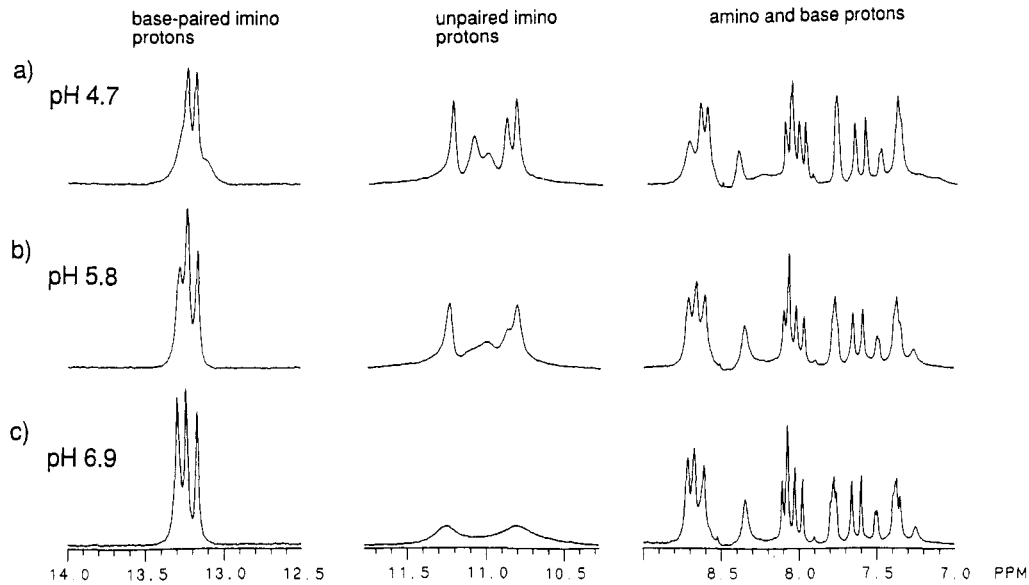

FIGURE 5: 500-MHz NMR spectra of d(CGCGTTGTTCGCG) in H_2O at 5 °C. Spectra were recorded by using the 1331 pulse sequence with the excitation maximum centered at 11 ppm, as a function of pH: (a) pH 4.7, (b) pH 5.8, and (c) pH 6.9.

paired amino proton, which was, in turn, connected to the nonexchangeable H5 proton on the cytosine ring that was already assigned. Internucleotide imino–imino NOEs were also observed, but assignment of the imino protons by these NOEs alone was uncertain due to spillover of the decoupler when the closely spaced imino resonances were irradiated.

Three of the base-paired imino protons exchange more slowly than the fourth, which is assigned to the terminal C_1-G_{13} base pair. This proton exchanges more rapidly presumably due to fraying at the end of the stem and is much broader than the others at all pHs and temperatures studied. It is important to note that the imino proton in the base pair at the junction of the loop and the stem has an exchange rate roughly equal to that for the central C-G base pairs, as evidenced by their similar line widths. This implies that there is no fraying of the base pair at the stem–loop junction and that the G_4-C_{10} base pair is less accessible to solvent than the terminal C_1-G_{13} base pair. The base-paired imino protons are completely exchanged at 30 °C at pH 7.0, a temperature well below the melting temperature. The imino proton exchange rate at 5 °C increases with pH up to pH 9.0, by which point the resonances are broadened into the base line, indicating that the exchange rate is not limited by the opening of the base pairs.

The unpaired imino resonances for bases in the loop region exchange much more rapidly than the base-paired imino protons and are barely visible at pH 6.9. At pH 4.7, all five unpaired imino resonances are visible, while the base-paired imino protons are somewhat broadened due to acid-catalyzed exchange. It was not possible to assign the loop imino protons as no NOEs were observed at any pH. It is reasonable to assume, however, that the two resonances that are sharpest at higher pH correspond to the imino protons of the T_5 and T_9 residues. These residues appear to be stacked over the stem from the NOE data (see below), and it is probable that they are less exposed to solvent than those residues in the outer portion of the loop region.

Analysis of the Sugar Puckering. The coupling constants between the H1′, H2′, H2″, and H3′ protons at 25 °C for all of the residues except G_2, G_4, G_7, and G_{11} were determined from a DQF-COSY spectrum, and the measured values were analyzed to obtain the preferred sugar puckering (Rinkel & Altona, 1987). An expansion of the region containing cross peaks between the H1′ and H2′/H2″ protons is shown in Figure S-5 of the supplementary material. The coupling constants were estimated directly from the spectrum, and these values were used as the starting point for a nonlinear least-squares procedure to fit the coupling constants. Cross sections through the cross peaks from the DQF-COSY were fit to sums of Lorentzian lines of the appropriate antiphase character. The adjustable parameters in the fits were the coupling constants, the peak amplitude, position, and width, and a base-line offset from zero. Use of this procedure helps to minimize systematic errors in the measurement of coupling constants from antiphase lines. The cross peak between the H1′ and H2′ of residue G_{13} is shown in Figure 6a with a cross section through the peak and the fit to the cross section in Figure 6b. Coupling constants determined from several different cross peaks for the same proton generally agreed to within 0.3 Hz, and in these cases the values were averaged. For all of the residues determined, the sugars existed predominantly in the C2′-endo conformation. This is readily determined from the sum of the $J_{H1'H2'}$ and $J_{H1'H2''}$ coupling constants, which were in all cases greater than 13.5 Hz. The data fit well to a mixture of 70–90% south conformation Φ_S = 135–171°, the remainder being modeled as C3′-endo, Φ_N = 9°, with Φ_m = 35°. The only exception to this was residue C_{10}, which exists in a pure C2′-endo conformation.

The DQF-COSY spectrum was recorded on the ^{13}C-labeled oligomer, permitting the measurement of carbon–proton coupling constants for the thymidine residues. The C6 carbon is coupled through the glycosidic bond to the H1′ proton of the deoxyribose, and this three-bond coupling constant is related to the glycosidic torsion angle by a Karplus relationship (Davies et al., 1985). The C6–H1′ coupling constants were determined by examination of the fine structure of the H1′ cross peaks in the DQF-COSY spectrum. An expansion of the H1′ to H2′ cross peak of residue T_5 is shown in Figure 6c, with the cross section and fitted spectrum shown in Figure 6d. The extra coupling due to the C6 is readily visible. The values of the C6–H1′ coupling constants for all four thymidine residues are 3.4 ± 0.2 Hz. The magnitude of the H6 to H1′ and H2′ NOEs for the thymidine residues indicates that they must be in the anti conformation, and there are two values of the glycosidic torsion angle that can give rise to the observed coupling constant, χ = 185° and 295°. The expected distances between the H6 protons and the H2′ and H2″ protons are both

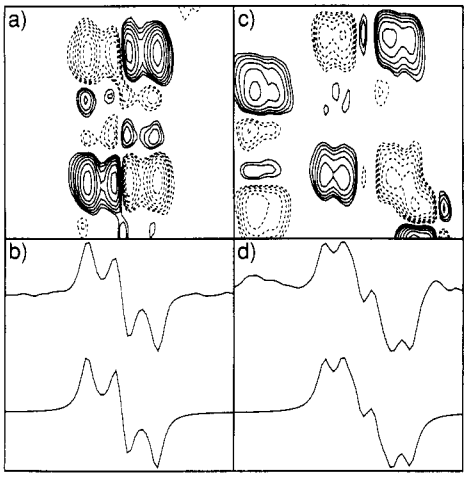

FIGURE 6: Determination of proton–proton coupling constants in d(CGCGT*T*GT*T*CGCG). The fine structure of the ω-scaled DQF-COSY spectrum was examined to extract coupling constants. Positive contours are solid lines, and negative contours are dashed lines. (a) Expansion of the cross peak between the H1′ and H2′ of G_{13}. (b) (Upper trace) Horizontal section through the cross peak in (a); (lower trace) calculated from a nonlinear least-squares fit of a sum of antiphase Lorentzian lines to the cross section. Fitted parameters are $J_{H1'-H2'}$ = 8.4 Hz, $J_{H1'-H2''}$ = 6.0 Hz, width = 3.0 Hz, amplitude = −7.68, and base-line offset = −0.037. (c) Expansion of the cross peak between the H1′ and H2′ of residues T_5. An extra coupling is observed in the H1′ direction due to coupling with ^{13}C at the C6 carbon. (d) (Upper trace) Horizontal cross section through the cross peak in (c); (lower trace) calculated from a nonlinear least-squares fit of a sum of antiphase Lorentzian lines to the cross section. Fitted parameters are $J_{H1'-H2'}$ = 7.2 Hz, $J_{H1'-H2''}$ = 6.5 Hz, $J_{H1'-C6}$ = 3.8 Hz, width = 3.1 Hz, amplitude = −4.14, and base-line offset = −0.02.

at least 1 Å longer for χ = 185° than is observed, while the expected distances for χ = 295° agree much more closely.

Interproton Distances. The time course of the buildup of interproton NOEs was determined by acquisition of NOESY spectra with successively longer mixing times. Five spectra with mixing times of 100, 200, 300, 400, and 500 ms were recorded and the cross-peak volumes integrated. The cross-relaxation rate is inversely proportional to the sixth power of the internuclear distance and is obtained from the initial slope of the NOE buildup curve. The cross-peak volumes as a function of mixing time were plotted for a number of cross peaks to determine the extent of spin diffusion. The 400- and 500-ms mixing time spectra showed significant spin diffusion, and these spectra were not used in the distance determinations. The longest mixing time used here is much longer than the ~100 ms recommended to avoid serious spin diffusion (Chazin et al., 1986). However, the rotational correlation time for the hairpin molecule is approximately half that of the decamer duplex used in those studies, and so the cross-relaxation rate will be approximately half as well. Thus, the mixing time may be doubled over the recommended 100-ms value without an increase in spin diffusion. The distances were calibrated by comparison to the cross-peak volume for the cytosine H6 to H5 using the relationship

$$r_{ij} = r_{ref}(\sigma_{ref}/\sigma_{ij})^{1/6}$$

where r_{ij} is the distance between hydrogens, r_{ref} is 2.47 Å, the distance between cytosine H5 and H6, σ_{ij} is the observed rate of increase of the volume of the cross peak, and σ_{ref} is the observed rate of increase of the volume of the H5/H6 cross peak. The cross-relaxation rate between the cytosine H5 and H6 protons determined from the 1D NOE experiments was used to scale the cross-peak volumes. The distances were used as restraints in energy minimization calculations using the program AMBER (Weiner & Kollman, 1981), to arrive at a model structure consistent with the NMR data (vide infra). A set of 60 distances, 37 intranucleotide and 23 internucleotide, was obtained and used for the model building.

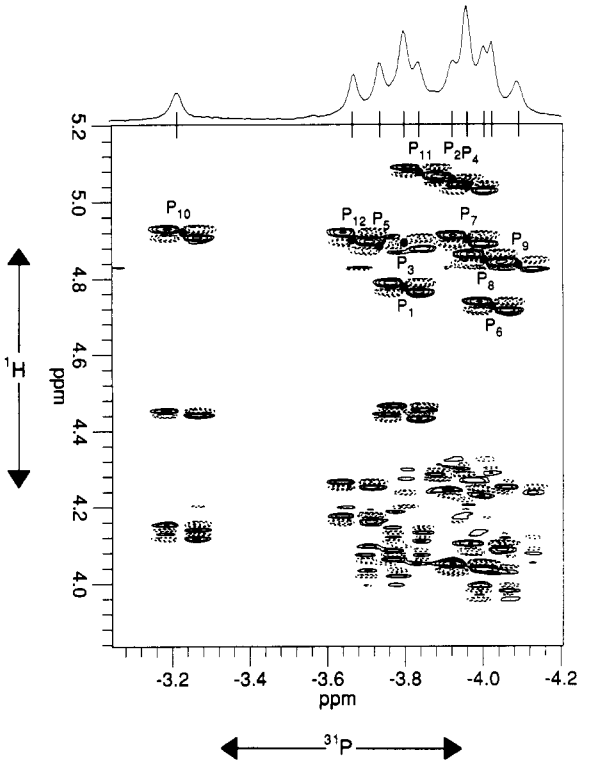

FIGURE 7: Phosphorus NMR spectrum of d(CGCGTTGTTCGCG) at 25 °C. The 202-MHz ^{31}P spectrum is shown along with the ^1H–^{31}P heteronuclear correlation spectrum used to assign the spectrum. The centers of the cross peaks between each phosphate and H3′ are marked by dots, and the assignments of the phosphates are indicated above or below the dots. Additional cross peaks are observed between the H4′ and H5′/H5″ protons and the phosphates, but with few exceptions these protons were not assigned. A total of 128 accumulations for each of 128 values of t_1 were averaged; 4K points were taken in t_2, and the apodization was 8-Hz Gaussian function. A pure sine bell apodization was used in t_1, and the data were zero-filled to 512 points.

Phosphorus NMR Data

The 202-MHz phosphorus NMR spectrum of d(CGCGTTGTTCGCG) at 25 °C is shown in Figure 7. The resonances, corresponding to the 12 phosphate groups in the molecule, are spread out over 1 ppm, but are clustered around −4.0 ppm. One resonance is particularly shifted downfield from the average shift. The phosphorus resonances were completely assigned by a 2D heteronuclear correlation experiment, also shown in Figure 7. In this experiment the chemical shift of a phosphate group is correlated to the chemical shift of the protons to which it is scalar-coupled. The phosphate is coupled via a three-bond coupling to the H3′ on the 5′ neighboring residue and to the H5′/H5″ on the 3′ neighboring residue. In addition, there is a four-bond coupling to the H4′ on the 3′ neighboring residue. Resolved cross peaks are observed for each of the possible connectivities to the downfield shifted phosphate, which may be assigned to the H3′, H4′, and H5′/H5″. Assignment of the H5′/H5″ and most of the H4′ resonances was not possible, so phosphorus assignments were made primarily on the basis of the connectivity to the H3′ resonances that were previously assigned from the proton NOESY spectrum. However, a group of H4′ resonances from residues G_2, G_4, and G_{11}, which are shifted downfield from the pack, were also readily assigned from the NOESY spectrum. The anomalously shifted phosphate was

clearly coupled to the H4′ of residue G_{11}, assigning that phosphate to P_{10}, which was confirmed by the coupling to the H3′ of residue C_{10}. The phosphates are numbered sequentially from the 5′ end of the molecule, so that the phosphate in C_1-p-G_2 is labeled P_1. The phosphate resonances P_1 and P_3 were also assigned by their connectivity to the H4′ protons. The remaining phosphates were unambiguously assigned on the basis of their coupling to the H3′ protons as indicated in Figure 7.

Conformation of P_{10}. The chemical shifts of phosphate residues are sensitive to the torsion angles about the P–O bonds (α and ζ) (Gorenstein & Luxon, 1979), to a lesser extent to the C–O torsions (β and ϵ), and also to the O–P–O bond angle (Giessner-Prettre et al., 1984). The deviation of the shift of a phosphate resonance from the average value can be interpreted in terms of changes in the populations of the rotamers about the P–O bonds. Alternate phosphate conformations are precedented in the different families of helices adopted by DNA. The phosphates in B-DNA are usually found in the g^-,g^- conformation, and the chemical shift of these phosphates is ~−4.26 ppm (TMP) (Chen & Cohen, 1984). There are two distinct phosphate conformations in Z-DNA, the GpC phosphate in the g^-,g^+ conformation and the CpG phosphate in the t,g^+ conformation. The chemical shifts of these phosphates are ~−4.4 and ~−3.0 ppm, respectively. Two different phosphate conformations are also found in the alternating B-DNA structure adopted by poly(dAdT)·poly(dAdT). The ApT phosphate in the g^-,g^- conformation is observed 0.2 ppm upfield of the TpA phosphate in the t,g^- conformation. The chemical shift of a phosphodiester as a function of the α and ζ torsion angles has been calculated (Gorenstein & Luxon, 1979), and the calculated chemical shifts of the t,g^- and the t,g^+ conformations are the same, as are the calculated shifts for g^-,g^- and g^-,g^+.

We propose that the downfield shifted phosphate P_{10} exists in the t,g^- conformation or a mixture of t,g^- and g^-,g^- rotamers. The t rotamer for the ζ torsion is correlated with the C2′-endo sugar pucker in alternating B-DNA (Gupta et al., 1980) and in the crystal structure of the dodecamer d(CGCGAATTCGCG) (Dickerson & Drew, 1981). The observed deoxyribose pucker for residue C_{10} is also pure C2′-endo, in contrast to the other sugars that are a mixture of S and N conformers, which is additional evidence in support of the t,g^- conformation for P_{10}. The temperature dependence of the chemical shifts for several phosphate groups is shown in Figure 8a. The downfield-shifted resonance corresponding to P_{10} exhibits a strong temperature dependence well below the melting temperature, shifting upfield by 0.4 ppm over a range of 50 °C. This is consistent with a change in the dominant population of the P–O rotamers of P_{10} from t,g^- to g^-,g^-. The resonances for all of the other phosphates shift downfield with increasing temperature. This is usually observed in the melting of B-DNA as the population of the t rotamers increases.

The β and ϵ torsion angles can be related to phosphorus coupling constants to the H5′/H5″ and H3′, respectively, via a Karplus relationship (Lankhorst et al., 1984). Unfortunately, the H5′/H5″ protons are difficult to assign, obscuring information about the β torsion. The H3′ protons, however, have been assigned, and approximate (±0.5 Hz) H3′–P coupling constants were determined from cross sections of a selective phosphorus–proton correlation experiment (Sklenar & Bax, 1987) (data not shown). The observed coupling constants for all residues were in the range 5.6–7.7 Hz, indicating that all residues exhibit an ϵ torsion in the range 150–162°. The ϵ torsion for residue C_{10} was determined to be 150° from the coupling constant of 7.7 Hz, which is within the normally observed range. The possibility of an unusual ϵ torsion is not rigorously excluded for residue C_{10}, or any other residue, since there are three other possible torsions that could give the same observed coupling constant: 2°, 90°, and 238°.

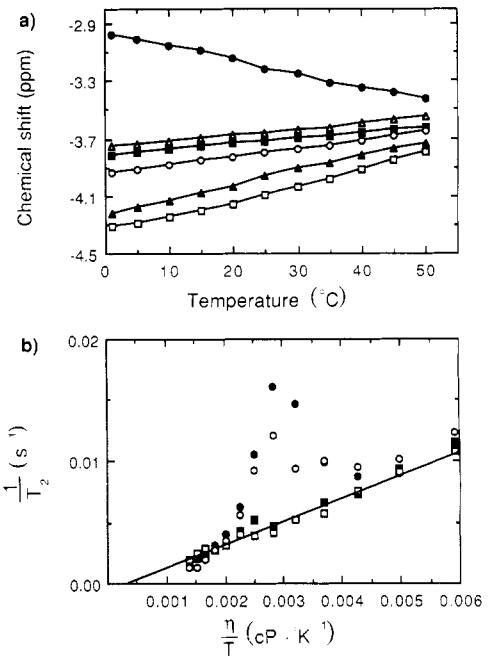

FIGURE 8: (a) Temperature dependence of selected phosphate chemical shifts for d(CGCGTTGTTCGCG): (●) P_{10}, (△) P_{12}, (■) P_5, (○) P_3, (▲) $P_{4'}$ (□) P_9. (b) Stokes–Einstein plot for T_2 relaxation of selected phosphates in d(CGCGTTGTTCGCG): (●) P_{10}, (○) P_9, (■) P_4, (□) P_{12}.

Phosphorus Relaxation. The temperature dependence of the relaxation of individual phosphates was also examined. The T_2 relaxation time as a function of temperature was measured for resolved phosphate resonances. A plot of transverse relaxation rate versus viscosity divided by temperature is shown for phosphates P_{12}, P_{10}, and P_9 in Figure 8b. This form of the temperature dependence is plotted because for molecules of this size T_2 is nearly inversely proportional to the rotational correlation time, which is in turn directly proportional to η/T. The temperature dependence for P_{12} is indeed nearly linear, while dramatic deviations from linearity are observed for the two phosphates P_9 and P_{10}. Such an increase in the transverse relaxation rate is indicative of an intermediate exchange process. This exchange process is only observed for the two phosphates P_9 and P_{10}, which are at the extremes of chemical shift.

The magnitude of the T_1 and T_2 relaxation rates can be interpreted in terms of the overall rotational motion of the molecule. Previous studies have indicated that dipolar proton relaxation and chemical shift anisotropy (CSA) relaxation both contribute to DNA phosphodiester relaxation at high fields (Gueron & Shulman, 1975; Hayashi et al., 1977; Gorenstein & Luxon, 1979; James, 1984). The T_1 and T_2 relaxation rates were calculated for the hairpin from the standard expressions for dipolar and CSA relaxation at 11.74 T, considering three protons at a distance of 2.6 Å, the chemical shift anisotropy of 136 ppm (Gueron & Shulman, 1975), and the rotational correlation time of 1.4 ns determined from the proton cross-relaxation rate. These calculations indicated that <3% of the total relaxation was due to dipolar interaction with the protons. This calculation is supported by the experimental fact that there was no observable steady-state ^1H–^{31}P NOE (<1%).

Table I: Observed and Calculated Temperature Dependence of the Relaxation of P_5, $\Delta\sigma$ = 160 ppm

temp (°C)	T_1 exptl	T_2 exptl	τ_c ratio	τ_c calcd	T_1 calcd	T_2 calcd
10	0.86	0.083	2.9	3.0	0.95	0.085
15	0.76	0.106	2.4	2.6	0.82	0.099
20	0.71	0.119	2.1	2.2	0.73	0.114
25	0.64	0.129	1.9		0.65	0.129

Working under the assumption that the phosphate relaxation is completely due to CSA, it is possible to independently determine the rotational correlation time from the ratio of T_1 to T_2. The T_1 and T_2 relaxation rates due to the CSA mechanism for an isotropically tumbling molecule that has an axially symmetric CSA tensor are given by (Abragam, 1961)

$$1/T_{1csa} = {}^1/_{15}\omega_P^2\Delta\sigma^2 J(\omega_P)$$

$$1/T_{2csa} = {}^1/_{90}\omega_P^2\Delta\sigma^2[3J(\omega_P) + 4J(0)]$$

where $J(\omega)$ is the spectral density:

$$J(\omega) = 2\tau_c/(1 + \omega^2\tau_c^2)$$

Taking the ratio of these two rate expressions and rearranging the result, the following simple expression can be derived

$$\tau_c = \frac{1}{\omega_P}\left[\frac{3}{2}\left(\frac{T_1}{T_2}\right) - \frac{7}{4}\right]^{1/2}$$

which is true only when the extreme narrowing condition does not hold (i.e., $\omega_P\tau_c > 1$). This expression is particularly useful because it requires no knowledge of the value of the chemical shift anisotropy ($\Delta\sigma$).

The phosphate resonance for P_5 is clearly resolved below the melting temperature, and the T_1 and T_2 determined at 10, 15, 20, and 25 °C are listed in Table I. The approximate rotational correlation time at these temperatures was calculated from the ratio T_1/T_2 by using the above expression and is also listed. The value of 1.9 ns at 25 °C is close to the value of 1.4 ns obtained from the proton cross-relaxation rates. The uncertainty in the correlation time determined from phosphorus relaxation at 25 °C is ±0.3 ns, assuming an error of ±20 ms in the T_1 and T_2 values. The measured T_1 and T_2 values and the calculated correlation time can be used to estimate the size of $\Delta\sigma$ from the expressions for the CSA relaxation rates. From the T_1 and T_2 of P_5 at 25 °C a value of $\Delta\sigma$ of 160 ppm is obtained. This value is intermediate to the extremes reported in the literature of 136 (Gueron & Shulman, 1975) and 180 ppm (Gorenstein & Luxon, 1979), demonstrating the self-consistency of the analysis.

The rotational correlation times at 10, 15, and 20 °C were calculated from the value at 25 °C on the basis of the known temperature dependence for the viscosity of D_2O (Millero et al., 1971), and these values are listed in Table I. These predicted values agree well with those determined from the T_1/T_2 ratio. The T_1 and T_2 relaxation times were then calculated by using these correlation times and the value for $\Delta\sigma$ of 160 ppm, and these values are also shown in Table I. The agreement of the predicted temperature dependence of two different relaxation parameters based on the measurements at a single temperature is a good indication that the assumptions of the model for relaxation are valid. The effects of internal motions were not taken into consideration in these calculations, and they are not required to explain the relaxation data. The relaxation rates are extremely sensitive to internal motions because the value of $\omega_P\tau_c \simeq 1$, and the relaxation rates are near the maximum value possible. Internal motions would reduce the effective correlation time and produce longer relaxation times. Internal motions have been reported on the nanosecond time scale for much larger fragments of DNA (James, 1984), but these motions are on the order of the rotational correlation time for this molecule and so would be obscured. There seems to be no evidence for substantial internal motions for P_5 on the subnanosecond time scale from the present work.

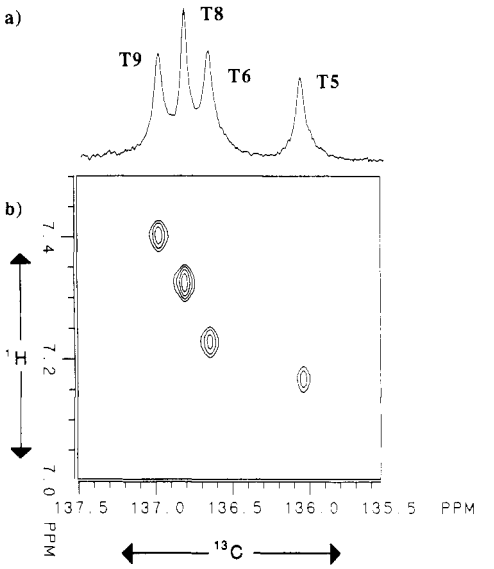

FIGURE 9: (a) 125-MHz carbon NMR spectrum of d(CGCGT*T*GT*T*CGCG). Sample conditions are the same as in Figure 1 except that the oligonucleotide concentration is 7 mM. Four resolved resonances are observed, one for each of the four thymidine residues labeled at C6 with ^{13}C, and the assignments are indicated. This spectrum is 1000 averaged accumulations and required 15 min to acquire. (b) Heteronuclear correlation spectrum used to assign the carbon resonances. A total of 128 FIDs were collected with 104 acquisitions each, for a total acquisition time of 7.5 h. Apodization was a sine bell shifted by $\pi/4$ in f_2 and an unshifted sine bell in f_1, and the spectrum is shown in absolute value mode. One cross peak is observed from the C6 to the H6 proton for each thymidine residue.

Carbon NMR Data

The 13-mer d(CGCGTTGTTCGCG) was resynthesized with a thymidine phosphoramidite labeled in the thymine ring C6 with carbon-13. The details of the synthesis of the labeled phosphoramidite and the analysis of the relaxation mechanisms for the C6 carbon are published elsewhere (Williamson & Boxer, 1988). The 125-MHz carbon NMR spectrum of d(CGCGT*T*GT*T*CGCG), which contains a carbon-13 label at each of the four thymidine C6 carbons, is shown in Figure 9, together with a plot of the heteronuclear correlation spectrum used to assign the resonances. The purpose of the carbon labeling was to determine the extent of internal motions in the loop region by measurement of carbon relaxation parameters. The T_1 and NOE for each of the four carbons were measured at both 7.04 and 11.74 T, at 5, 15, 25, 35, and 45 °C, and these results are summarized in Table II. Also measured was the NOE buildup rate between the four cytosine H6 to H5 protons at each temperature, and the average value was used to calculate the rotational correlation time that is plotted in Figure 10.

The relaxation of the C6 carbon is mainly due to the dipolar interaction with H6, with a lesser contribution due to chemical shift anisotropy (Williamson & Boxer, 1988). The value of $\Delta\sigma$ obtained for the C6 carbon from relaxation measurements on a model compound was 185 ppm. The total longitudinal relaxation rate ρ due to both mechanisms and the cross-re-

Table II: T_1 and NOE for the C6 Carbons in d(CGCGT*T*GT*T*CGCG)

	temp (°C)				
	5	15	25	35	45
			NOE 500 MHz		
T_9	0.213	0.215	0.290	0.285	0.348
T_8	0.356	0.409	0.380	*a*	0.455
T_6	0.265	0.365	0.370	*a*	0.390
T_5	0.161	0.180	0.240	0.262	0.317
			T_1 500 MHz (s)		
T_9	0.342	0.276	0.210	0.235	0.232
T_8	0.344	0.250	0.228	*a*	0.260
T_6	0.348	0.241	0.234	*a*	0.239
T_5	0.365	0.244	0.209	0.242	0.221
			NOE 300 MHz		
T_9	0.197	0.297	0.343	0.489	0.545
T_8	0.412	0.477	0.482	*a*	0.710
T_6	0.339	0.473	0.506	*a*	0.666
T_5	0.216	0.310	0.425	0.534	0.621
			T_1 300 MHz (s)		
T_9	0.173	0.156	0.150	0.163	0.184
T_8	0.193	0.185	0.172	*a*	0.213
T_6	0.186	0.169	0.177	*a*	0.191
T_5	0.166	0.154	0.145	0.161	0.176

[a] Not measured due to peak overlap.

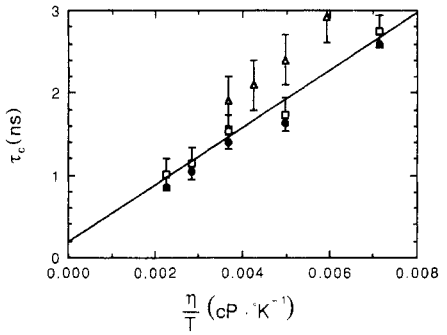

FIGURE 10: Stokes–Einstein plot for the rotational correlation times from three different nuclei: (●) ^{1}H, (□) ^{13}C, (△) ^{31}P. The rotational correlation time was experimentally determined from the cross-relaxation rate between cytosine H5 and H6, from T_1 and NOE measurements on the thymidine C6 carbon, and from T_1 and T_2 measurements on P_5. Error bars are indicated for the data from phosphorus relaxation, and the error bars on the carbon data points are the errors from the least-squares fit. The line drawn is the least-squares line fit to the correlation time obtained from the carbon data.

laxation rate σ due only to the dipolar mechanism can be expressed in terms of spectral densities (Abragam, 1961)

$$\rho = \frac{1}{T_{1dd}} + \frac{1}{T_{1csa}} = \frac{\gamma_C^2\gamma_H^2\hbar^2}{20r_{CH}^6}[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H - \omega_C)] + \tfrac{1}{15}\omega_C^2\Delta\sigma^2 J(\omega_C)$$

$$\sigma = \frac{\gamma_C^2\gamma_H^2\hbar^2}{20r_{CH}^6}[6J(\omega_H + \omega_C) - J(\omega_H - \omega_C)]$$

where γ_C and γ_H are the carbon and proton gyromagnetic ratios, h is Planck's constant, r_{CH} is the length of the carbon–proton internuclear vector, $\Delta\sigma$ is the asymmetry parameter of the chemical shift tensor, and the $J(\omega)$ are the spectral densities. The T_1 relaxation time and NOE are given by

$$T_1 = 1/\rho \qquad \text{NOE} = \gamma_H\sigma/\gamma_C\rho$$

The carbon relaxation data cannot be fit to a single correlation time and require a more detailed model including internal motions. We chose to use the model-free order parameter model (Lipari & Szabo, 1982a,b) to describe the relaxation, since it has a simple form with only two additional adjustable parameters. The spectral densities in this model include a term to describe the effects of internal motion

$$J(\omega) = \frac{S^2 2\tau_c}{1 + \omega^2\tau_c^2} + \frac{(1 - S^2)2\tau}{1 + \omega^2\tau^2}$$

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_i}$$

where τ_i is the correlation time for the internal motion and S is the generalized order parameter. The internal motions may be complex, but in this model they are grouped together into an effective internal correlation time that approximates the effects of the true motions. The order parameter is a measure of the spatial restriction of the internal motions.

An entire set of relaxation data for a given temperature was fit by using a least-squares procedure to an overall correlation time and individual internal correlation times and order parameters for each of the four residues. The overall correlation time determined for this procedure agrees extremely well with the values determined from the H6/H5 cross-relaxation rates from completely independent experiments. A plot of the overall rotational correlation time determined from relaxation measurements on three different nuclei as a function of temperature is shown in Figure 10. The correlation times are plotted versus η/T (viscosity over absolute temperature), which results in a linear plot. The viscosity for pure D_2O was used for all calculations (Millero et al., 1971), and the effect of buffer salts on viscosity is estimated to be <3%. The least-squares line is plotted through the carbon data.

The slope of this line is proportional to the molecular volume

$$\tau_c = \frac{V_{mol}}{k}\frac{\eta}{T} = \frac{4\pi r^3_{sphere}}{3k}\frac{\eta}{T}$$

where k is Boltzmann's constant and V_{mol} is the hydrated molecular volume. The molecular volume calculated from the slope of the line through the carbon data in Figure 10 corresponds to a sphere of radius 11.5 Å. The rotational correlation time for a sphere of this size would be 1.5 ns, which is the measured value within experimental error. The correlation times plotted in the Stokes–Einstein plot were obtained by interpretation of the relaxation data, whereas the correlation time obtained from the slope of the Stokes–Einstein plot is derived from hydrodynamic theory. Agreement of these two independent results is a good indication that the relaxation data have been interpreted correctly and that the molecule is well approximated by a sphere over the temperature range measured. The correlation time data were also plotted in the Arrhenius form, and the slope of that line is equal to the apparent activation energy for tumbling in water. The value obtained was 4.0 kcal/mol, which is close to the expected value due to the viscosity of D_2O (Wilbur et al., 1976).

Plots of the order parameter and internal correlation time determined for each of the four residues are shown in Figure 11. The general trend is for an order parameter of ~0.6 with internal correlation times from 10 to 50 ps for the loop residues. Closer examination of these data shows that the loop residues divide themselves into two groups. Residues T_5 and T_9 have an order parameter slightly larger than do residues T_6 and T_8, which may be interpreted as a more restricted motion for the two loop residues closing the loop, while the residues in the extremities of the loop are less restricted.

Model Building

The purpose of these model-building exercises was to construct a model structure that was a consistent as possible with

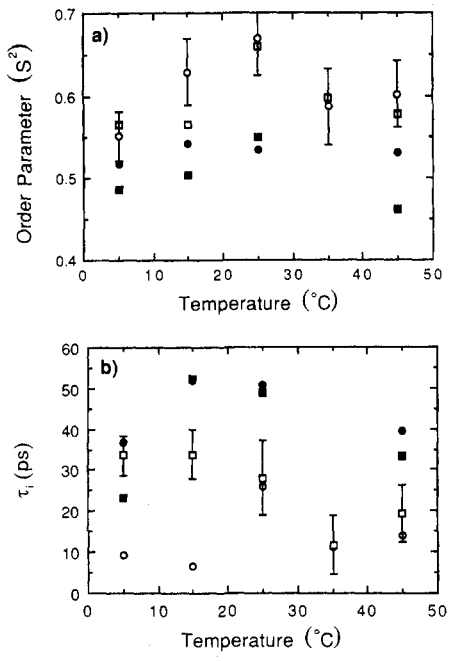

FIGURE 11: (a) Order parameters and (b) internal correlation times for the thymidine C6 carbons: (□) T_9, (■) T_8, (●) T_6, (○) T_5. The order parameters and the correlation times for internal motions were obtained from a least-squares fit to the relaxation data. Error bars indicated for the error in the order parameter are for residue T_9 and are the largest errors among the four residues. The errors for residue T_5 are the smallest and are approximately 50% smaller. The error bars for the internal correlation time are for residue T_8 and are the largest errors.

the available NMR data. Although the data were clearly insufficient to do a solution structure determination, we wished to determine if the data could be used to distinguish among several plausible model structures. Four initial models were constructed by using either A- or B-DNA ideal coordinates for the stem region, and two different arrangements of the loop bases stacked over the stem in a continuation of the A or B geometry. The two loop stacking arrangements were three bases stacked on the 5′ end of the stem with two bases stacked on the 3′ end, and three bases stacked on the 3′ end of the stem, with two bases stacked on the 5′ end. In addition, a fifth initial model was constructed as a control by starting with the bases all stacked in a single strand of B-DNA and subjecting this structure to energy minimization with hydrogen-bonding restraints to fold the structure into a hairpin with four base pairs. In the first four initial structures, there was a gap between a phosphorus atom and a 5′ hydroxyl atom of 14–20 Å that was to be closed during the energy minimization. Although this is a somewhat drastic procedure, it proved to be the most effective method for closing the loop while maintaining nearly ideal geometry in the loop residues without interactively manipulating individual torsions.

The model building was performed in two stages. In the first, all five initial model structures were minimized with the 60 NOE distance restraints and, as a control, without distance restraints. A table (Table S-II) of the distances used in the model building is available in the supplementary material. The large force constant of 100 kcal/Å^2 was used for the restraints as the structure changed very little unless the restraint energy was a noticeable fraction of the total energy. In the initial steps of the minimizations, the top base pair of the stem became disrupted as the gap in the initial loop structure was closed. The effects of this artifact were greatly reduced by inclusion of base-pairing hydrogen bond restraints in the stem to assist in maintaining the stem geometry in the initial steps

Table III: Energy Minimization of Different Model Hairpin Structures

	no restraints		with NOE restraints		
	energy (kcal × 10^3)	rms[a] (Å)	energy (kcal × 10^3)	rms[a] (Å)	model
model 1	−1.20	0.785	−1.15	0.462	B-DNA 5′ stack
model 2	−1.18	1.08	−0.972	0.448	B-DNA 3′ stack
model 3	−1.16	1.56	−0.933	0.657	A-DNA 5′ stack
model 4	−1.16	1.76	−1.01	0.610	A-DNA 3′ stack
model 5	−1.05	1.85	−0.798	0.551	control

[a] Root mean square deviation from the exact NMR distances.

of the calculation. The final energy and the rms deviation from the NMR distances are given for even of the five models refined with and without distance restraints in Table III. The two models with B-DNA stems gave the lowest energy and the best agreement with the NMR data. Of these two models, the model in which the three 5′ loop bases were stacked was chosen for further refinement.

The second stage of the model building was refinement of this model using additional torsion restraints that could be derived from other NMR data. The sugar pucker for all residues was restrained to $P = 153°$, and the glycosidic torsion angles for the four thymidine residues were restrained to $\chi = 295°$. Also the conformation of P_{10} was restrained to be in the t,g$^-$ conformation. Examination of the initially refined model indicated that several of the loop torsion angles were outside the range normally found in B-DNA. We found that the loop region was quite flexible, and the turn in the loop could be accommodated with most of the loop torsions in their normal ranges. Addition restraints were applied to reset any loop torsions to B-DNA values, and the initially refined structure was reminimized. Finally, all distance and torsion restraints were removed, and the structure was minimized again. The rationale for this procedure was to allow the restraints to initially shape the structure and pull it into a local minimum. Once in the local minimum, the restraints are dropped and unfavorable contacts imposed by the restraints are allowed to relax. During this minimization, the P_{10} torsion reverted to g$^-$,g$^-$, and so the ζ torsion for P_{10} was subsequently forced into the t conformation with a single restraint and a brief minimization.

The final model is shown in Figure 12 as a stereo pair. The final energy was -1.20×10^3 kcal, and the final rms deviation from the NMR data was 0.54 Å. A table (Table S-III) of the torsion angles for the final model is available in the supplementary material. For the final model, minimized without any restraints, there were 33 distances that differed from the measured value by <0.25 Å, 11 that differed by 0.25–0.5 Å, 12 that differed by 0.5–1.0 Å, and 4 that differed by >1.0 Å. Of the 16 distances that differed by more than 0.5 Å, half were distances in the loop region and half were distances in the stem region. This suggests there is no localized gross inaccuracy in the model. It is probable that these large distance errors are at least in part due to inaccuracies in the NOE data.

We attempted to sample the conformational space for this molecule by choosing several different starting structures. While the initial structures could be differentiated on the basis of agreement with the NOE distance data, the distance restraints were not sufficient to pull any of the structures out of their local minimum, and so the final structure was not independent of the starting structure. This is due in part to

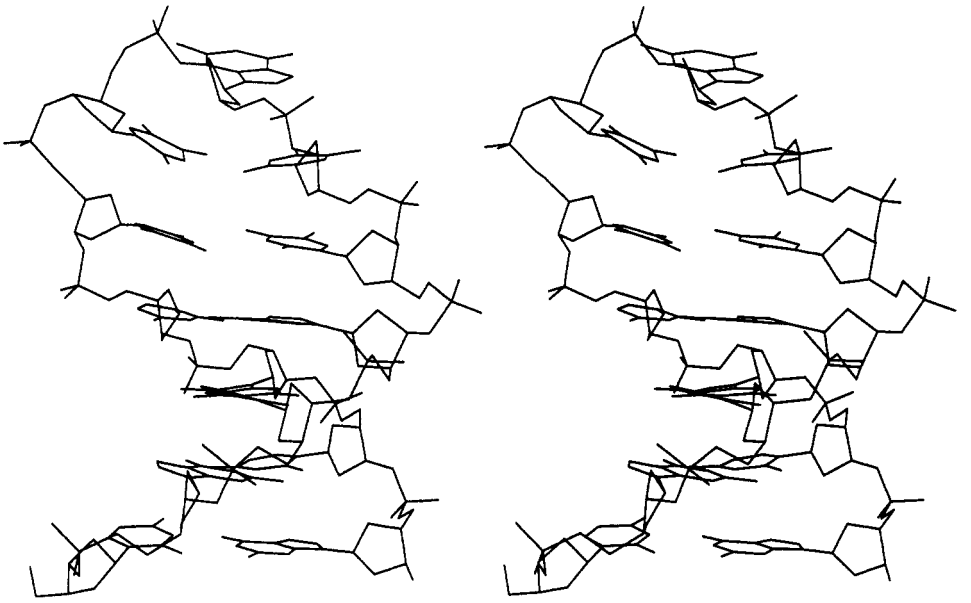

FIGURE 12: Stereo pair diagram of the energy-minimized model of the hairpin formed by d(CGCGTTGTTCGCG).

a fundamental limitation of the model-building procedure we used and in part to a limitation of the information content in the set of 60 NOE distances.

The final model structure agrees quite well with the qualitative observations that can be made on the basis of the NMR data. The stem fits well to B-DNA, and the top base pair in the stem is not disrupted by the steric constraints of loop closure. The loop bases can be stacked well over the stem, although the stacking between residues G7 and T6 appears weaker in the model than do the other loop stacks. The P_{10} conformation in the final model is t,g^+, which is consistent with the phosphorus NMR data. Although the structure is permissive to this torsion, it reverts to the normal g^-,g^- conformation without a restraint. The fact that the model does not predict this unusual phosphate torsion is probably due to a combination of inaccuracy in the model and an insensitivity of the AMBER force field to subtleties in the stacking of the loop and adjacent bases.

Discussion

The proton NOE data indicate that the oligomer d(CGCGTTGTTCGCG) exists in a hairpin conformation with a stem region consisting of four C-G base pairs stacked in approximate B-DNA geometry and a loop region of five unpaired bases in which stacking interactions are present. The stacking pattern is such that the three bases in the 5′ portion of the loop region are stacked over the stem in a continuation of the helix, while the fourth base is loosely stacked over the fifth, which is in turn stacked over the 3′ end of the stem region. This pattern of stacking is similar to that observed in other DNA hairpins, but is opposite to that observed in the anticodon loop of tRNA.

On the basis of the carbon relaxation data, the loop region appears to be somewhat flexible, exhibiting internal motions on the order of several tens of picoseconds. Although the generalized order parameters determined from the relaxation data need not be interpreted in terms of a specific model, it is convenient for discussion to relate the generalized order parameter to the order parameter describing diffusion in a cone as a measure of the spatial restriction of the motion. The order parameters for the inner loop bases of 0.63 and for the outer bases of 0.53 correspond to cone angles of 31° and 36°, respectively. The order parameters are relatively constant over the temperature range measured.

The cone model is useful if no explicit model is available for the internal motions, but a least one reasonable model for a simple internal motion that has structural significance can be made. If all of the internal motion is assumed to arise from motion about the glycosidic bond, the experimental order parameter can be used to extract a value for the distribution of glycosidic torsion angles. The complete autocorrelation function, $C(t)$, can be factored into an external contribution due to the overall tumbling of the molecule, $C_o(t)$, and a contribution of the internal motions, $C_i(t)$, if internal motion and the tumbling are assumed to be independent:

$$C(t) = C_o(t)C_i(t)$$

The order parameter is the asymptote of the normalized internal correlation function (Lipari & Szabo, 1982a,b). If the assumption is made that the ensemble average of the orientations about the glycosidic bond is equal to the time average for a single molecule, the value of the order parameter can be used to calculate the width of the distribution of glycosidic torsions. The expression for the asymptote of the internal correlation function is given in terms of averages of the second-order spherical harmonics by (Olejniczak et al., 1984)

$$S^2 = \langle C_i(\infty)\rangle^2 = (4\pi/5)\sum_{m=-2}^{2} |\langle Y_m^2(\theta,\phi)\rangle|^2$$

A simple assumption about the distribution of glycosidic torsions is that they are normally distributed about a given mean. The probability of a given value of χ is given by

$$P(\chi) = \frac{1}{(2\pi)^{1/2}\alpha}\exp\left(\frac{-(\chi-\chi_0)^2}{2\alpha^2}\right)$$

where α is the standard deviation of the distribution. A particular value of χ gives rise to a particular orientation of the C6–H6 bond in the molecular axis frame. The polar angles of the bond vector for various values of χ can be used to calculate the averages of the spherical harmonics given above. The order parameters can be obtained from

$$S^2 = (4\pi/5)\int \sum_{m=-2}^{2} |Y_m^2(\theta,\phi)|^2 P(\chi)\, d\chi$$

which averages the spherical harmonics weighted by the dis-

tribution of the glycosidic torsion angles. This expression was evaluated numerically for different values of the width of the distribution of glycosidic torsions, α, by determining from the coordinates of a model the θ and ϕ values for incremental values of χ. The experimental values of the order parameter of 0.53 and 0.63 correspond to distributions about the glycosidic bond of ±33° and ±27°, respectively. These values are an upper limit to the true value of the root mean square glycosidic torsion angles, since in this calculation all of the internal motion is attributed to motion about the glycosidic bond.

Interpretation of the values for the internal correlation time is not straightforward. For molecules of this size, the magnitude of the spectral density term for the internal motions is quite small compared to the term for the overall motion if the internal motions are fast. As a result, the relaxation parameters are most sensitive to the amplitude of the internal motion and relatively insensitive to the frequency of the internal motion. Furthermore, the order-parameter model cannot distinguish two or more internal motions that are in the extreme narrowing limit. The true motions of the C6 carbons are undoubtedly complex, including motion about the glycosidic bond and backbone motions. The internal correlation times obtained from the fit to the data are the best compromise to describing a multiexponential relaxation by a single exponential and cannot be interpreted too strongly. The large error bars on the internal correlation times reflect the insensitivity of T_1 and NOE to the frequency of the internal motion for molecules of this size. This drawback to the model for internal motion is offset by the knowledge that the model correctly fits the relaxation data and the overall rotational motion.

There is a close agreement between the overall rotational correlation times determined from the cross-relaxation data and from the carbon relaxation data. The rotational correlation time of this hairpin has also been determined by dynamic depolarized light scattering to be 1.7 ± 0.1 ns (W. Eimer, unpublished results). This is excellent agreement from two extremely different physical techniques for the overall rotational correlation time. The fact that the correlation time obtained from the cytosine H5 to H6 cross relaxation is consistent with the value obtained from both light scattering and carbon relaxation implies that the motion of the cytosine bases is quite restricted and does not contribute significantly to the cross relaxation. If there are motions of the cytosine bases that affect proton relaxation, then the reported overall correlation time will be shorter than the true one. An estimate of the maximum size of this error can be made by assuming that the cytosine bases have motional parameters similar to those of the loop bases. If the cytosine is assumed to have an internal correlation time of 30 ps and an order parameter of 0.63, an overall correlation time of 2.3 ns is required to give the observed cross-relaxation rate between H5 and H6.

The rotational correlation time obtained from phosphorus relaxation is somewhat longer than that obtained from proton or carbon relaxation. The correlation time was calculated by assuming that all of the relaxation was due to CSA. If there were a small contribution to T_2 from chemical exchange, the value of the rotational correlation time obtained from the T_1/T_2 ratio would be overestimated. The data in Table I at 25 °C can be fit equally well to a correlation time of 1.5 ns if $\Delta\sigma$ is changed to 150 ppm and 30% of the T_2 relaxation is due to chemical exchange. Since the phosphate resonances shift downfield with increasing temperature, it is reasonable that some chemical exchange might occur. Additional relaxation measurements at different magnetic fields would be required to determine the contributions of various mechanisms to the phosphate relaxation.

It is possible that the internal motions affect the values obtained for internuclear distances. To estimate the magnitude of this effect, we can calculate the expected cross-relaxation rate between two protons separated by a distance of 3 Å with an order parameter of 0.6, an internal correlation time of 30 ps, and an overall correlation time of 1.5 ns. If this calculated cross-relaxation rate is interpreted in terms of isotropic motion, the value for the internuclear distance will be overestimated by 0.3 Å. Internal motions will also affect cross-relaxation rates by generating a distribution of internuclear distances. For a distance of 3.0 Å with a range of motion of ±0.4 Å, the average internuclear distance obtained from the cross-relaxation rate will be underestimated by 0.2 Å. In this particular example, these two effects of the internal motions on the distance determination tend to cancel. Even with the significant motion of the loop bases, the largest errors in the distance determination are probably due to spin diffusion and not the effects of internal dynamics.

The existence of slow conformational transitions near the stem–loop junction is an interesting feature of this hairpin. The phosphate groups P_9 and P_{10} are in fast exchange between two or more conformations. Limits can be placed on the rate of conformational transitions if several assumptions are made. The range of chemical shifts for phosphodiesters has been calculated to be ~2 ppm (Gorenstein & Luxon, 1979). If the exchange of P_{10} is modeled as a two-state exchange between g^-,g^- and t,g^- conformations, the upper limit to the difference in chemical shifts is ~2 ppm, or ~400 Hz at 11.74 T. The exchange rate may then be estimated from the broadening of the P_{10} resonance compared to phosphates not undergoing chemical exchange. The exchange broadening of P_{10} is maximal at 35 °C. At that temperature, the exchange relaxation is calculated to be 13 s^{-1}, which for a $\Delta\delta$ of 400 Hz corresponds to a conformational transition rate of 1600 s^{-1}. Exchange broadening has also been observed for some phosphate groups in tRNA, but the resonances were not assigned (Gorenstein & Luxon, 1979).

Both P_9 and P_{10} undergo exchange broadening in the temperature range 20–50 °C. Several attempts were made to change the P_9 and P_{10} torsion angles in the model, and in general the model structure was permissive to these changes; however, there was no obvious resulting energetic advantage. Phosphate P_9 is quite close in chemical shift to that expected for the g^-,g^- conformation and is still observed to undergo exchange broadening in concert with P_{10}. One possible two-state model to explain the observed data is that P_9 exists in the g^-,g^- conformation and P_{10} in the t,g^- conformation at low temperatures. At higher temperatures, P_{10} exchanges with an increasing proportion of g^-,g^- conformation, and P_9 exchanges with an increasing proportion of the t,g^- rotamer, in a concerted motion of the backbone. Although there are clearly insufficient data to unambiguously prove this model, it is consistent with the observations of chemical exchange and direction of shifts with temperature.

For a loop of five bases, model building indicates that no unusual or strained torsion angles are required to complete the turn of the loop. However, there is strong evidence from the phosphorus NMR data for unusual torsion angles in the region of the stem–loop junction. The chemical shift evidence from the proton spectrum indicates that there are changes in the stacking of the bases T_8 and T_9 with temperature, and there is also a large shift of the C_{10} H1′ proton. A simple model

to explain all of the data would be that there are two slightly different stacking arrangements among the bases T_8, T_9, C_{10}, and G_{11}. The relative populations of these two conformations change with temperature as the relative free energy of stacking and hydrophobic interactions changes.

It has been suggested that the stacking preferences in hairpin loops can be rationalized on the basis of the shortest interstrand phosphate distance in B-DNA or A-RNA (Haasnoot et al., 1986). By use of this model, the prediction is that three bases will stack in the 5′ portion of the loop, and two will stack in the 3′ portion of the loop. This is the stacking pattern observed in this hairpin. Recent work on the free energy of single bases stacked on a double helix indicates that extra bases stacked on the 5′ of the helix are thermodynamically more stable than extra bases stacked on the 3′ end of the helix (Senior et al., 1988). Thus, there appear to be two opposing trends that shape the loop region. There is a steric constraint for loop closure that favors stacking in the 5′ end of the loop over the 3′ end of the stem, but thermodynamics favors stacking in the 3′ end of the loop over the 5′ end of the stem. The exchange phenomenon observed at the stem–loop junction could well be a manifestation of the interplay between these two opposing forces.

In summary, the dynamic aspects of this hairpin structure fall into two categories. There are slow conformational transitions on the millisecond time scale and rapid motions of the loop residues on the subnanosecond time scale. The structural requirements for closure of the hairpin loop are easily met by torsion angles characteristic of B-DNA. Stacking interactions are present in the loop and between the loop and the stem, and the pattern of stacking follows that expected for B-DNA. Comparison with related hairpins in the accompanying paper will identify which of the features of this hairpin are truly unique and which are general features of hairpin structure.

ACKNOWLEDGMENTS

The 500-MHz NMR spectrometer was funded by grants from the National Science Foundation (DMB-8515942) and from the National Institutes of Health (1-51-RR-2733-01). We thank Drs. Phil Borer and Art Pardi for helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL AVAILABLE

Figures showing expansions of regions of the NOESY spectrum of d(CGCGTTGTTCGCG) (H1′ to H2′/H2″ region, base proton to H1′ region, and base proton to H2′/H2″ region), cross-relaxation rate determination for cytosine H5 to H6, and expansion of the H1′ to H2′/H2″ cross-peak region of the ω-scaled DQF-COSY for d(CGCGTTGTTCGCG) and tables giving the proton chemical shifts for d(CGCGTTGTTCGCG) at 25 °C, the interproton distances used in energy minimization, and the torsion angles from the final energy-minimized hairpin model (9 pages). Ordering information is given on any current masthead page.

Registry No. d(CGCGTTGTTCGCG), 115427-43-5; d(CGCGT*T*GT*T*CGCG), 119242-87-4.

REFERENCES

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Oxford Science Publications, New York.

Caruthers, M. H., Kierzek, R., & Tang, J. Y. (1987) *Bioact. Mol. 3*, 3–21.

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

Chen, C., & Cohen, J. S. (1984) in *Phosphorus-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Chapter 8, Academic Press, Orlando, FL.

Cheng, D. M., Kan, L., Frechet, D., Ts'o, P. O. P., Uesugi, S., Shida, T., & Ikehara, M. (1984) *Biopolymers 23*, 775–795.

Davies, D. B., Rajani, P., MacCoss, M., & Danyluk, S. S. (1985) *Magn. Reson. Chem. 23*, 72–77.

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

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

Gait, M. J., Ed. (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL, Oxford, England.

Giessner-Prettre, C., Pullman, B., Prado, F. R., Cheng, D. M., Iuorno, V., & Ts'o, P. O. P. (1984) *Biopolymers 23*, 377–388.

Gorenstein, D. G., & Luxon, B. A. (1979) *Biochemistry 18*, 3796–3804.

Gueron, M., & Shulman, R. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3482–3485.

Gupta, G., Bansal, M., & Sasisekharan, V. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6486–6490.

Haasnoot, C. A. G., def Hartog, J. H. J., de Rooij, J. F. M., van Boom, J. H., & Altona, C. (1980) *Nucleic Acids Res. 8*, 169–181.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., & Hilbers, C. W. (1983) *J. Biomol. Struct. Dyn. 1*, 115–128.

Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843–857.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

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

Hayashi, F., Akasaka, K., & Hatano, H. (1977) *Biopolymers 16*, 655–657.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283.

Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., & Kearns, D. R. (1986) *Biochemistry 25*, 4840–4849.

James, T. L. (1984) in *Phosphorus-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Chapter 12, Academic Press, Orlando, FL.

Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387.

Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546–4559.

Lipari, G., & Szabo, A. (1982b) *J. Am. Chem. Soc. 104*, 4559–4570.

McCord, E. F., Morden, K. M., Pardi, A. S., Tinoco, I., & Boxer, S. G. (1984a) *Biochemistry 23*, 1926–1934.

McCord, E. F., Morden, K. M., Pardi, A. S., Tinoco, I., & Boxer, S. G. (1984b) *Biochemistry 23*, 1935–1939.

Millero, F. J., Dexter, R., & Hoff, E. (1971) *J. Chem. Eng. Data 16*, 85–87.

Milligan, J. F., Groebe, D. R., Witherall, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.

Olejniczak, E. T., Dobson, C. M., Karplus, M., & Levy, R. M. (1984) *J. Am. Chem. Soc. 106*, 1923–1930.

Orbons, L. P. M., van Beuzekom, A. A., & Altona, C. (1987a) *J. Biomol. Struct. Dyn. 4*, 965–987.

Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1987b) *J. Biomol. Struct. Dyn. 4*, 939–963.

Patel, D. J., Kozlowski, S. A., Ikuta, S., Itakura, K., Bhatt, R., & Hare, D. H. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 197–206.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–649.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & Van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.
Scheek, R. M., Boelens, R., Russo, N., van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371–1376.
Senior, M., Jones, R. A., & Breslauer, K. J. (1988) *Biochemistry 27*, 3879–3885.
Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525–7526.
Sklenar, V., Miyashiro, H., Zon, G., Miles, T. A., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.
Uesugi, S., Shida, T., & Ikehara, M. (1981) *Chem. Pharm. Bull. 29*, 3573–3585.
Usman, N., Ogilvie, K. K., Jiang, M. Y., & Cedergven, R. J. (1987) *J. Am. Chem. Soc. 109*, 7845–7854.
Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.
Wemmer, D. E., Chou, S., Hare, D. R., & Reid, B. R. (1985) *Nucleic Acids Res. 13*, 3755–3772.
Wilbur, D. W., DeFries, T., & Jonas, J. (1976) *J. Chem. Phys. 65*, 1783.
Williamson, J. R. (1988) Ph.D. Thesis, Stanford University.
Williamson, J. R., & Boxer, S. G. (1988) *Nucleic Acids Res. 16*, 1529–1540.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Multinuclear NMR Studies of DNA Hairpins. 2. Sequence-Dependent Structural Variations†

James R. Williamson‡ and Steven G. Boxer*
Department of Chemistry, Stanford University, Stanford, California 94305
Received September 8, 1988; Revised Manuscript Received December 8, 1988

ABSTRACT: The solution conformation of three related DNA hairpins, each with five bases in the loop, is investigated by proton and phosphorus 2D NMR methods. The sequences of the three oligomers are d(CGCGTTGTTCGCG), d(CGCGTTTGTCGCG), and d(CTGCTCTTGTTGAGCAG). One pair of hairpins shares the same stem sequence but differs in the loop, and the appearance of an unusual phosphate torsion in the stem is found to depend on the sequence in the loop of the hairpin. The second pair of hairpins shares the same loop region but differs in the stem sequence in that the base pair which closes the loop is a C-G or G-C pair. The pattern of NOEs reveals that the stacking arrangement in the loop region depends on the base pair that closes the stem. These results suggest that hairpin loop conformation and dynamics are sensitive to small changes in the loop and adjacent stem sequences. These findings are discussed in relation to sequence-dependent thermodynamic changes that have been observed in RNA hairpins.

In the preceding paper, the solution structure of the DNA hairpin formed by d(CGCGTTGTTCGCG) was studied in detail. Some unusual features in this structure were revealed by multinuclear NMR studies in the region of the junction between the stem and the loop regions. A phosphorus resonance, assigned to P_{10}, was shifted downfield, and this was interpreted as a phosphate in the t,g$^-$ conformation. This phosphate, located one base away from the junction of the stem and loop, exhibits relatively slow exchange between two or more conformations at intermediate temperatures. This is accompanied by upfield shifts of the nearby base protons in the loop region. The stacking of the two loop bases T_8 and T_9 over the stem changed slightly in a premelting transition. The other three bases in the loop of this hairpin were also found to stack over the other side of the stem region.

To investigate the possibility of a sequence effect on hairpin conformation, two new sequences were designed and synthesized. For purposes of discussion, the three hairpin sequences will be referred to as follows:

hairpin I

d(CGCG-TTGTT-CGCG)

hairpin II

d(CGCG-TTTGT-CGCG)

hairpin III

d(CTGCTC-TTGTT-GAGCAG)

Hairpin I is the subject of a detailed structural investigation (Williamson & Boxer, 1989). Hairpin II is identical with hairpin I except that the G in the loop has been shifted one position toward the 3′ end of the sequence. This sequence was designed to examine the effect of loop sequence on the conformation of the hairpin. Hairpin III was designed to test the effect of the stem sequences on the loop conformation. The sequence in the loop (TTGTT) is the same as that of hairpin I, but the important difference is that the loop in hairpin III is closed by a G-C pair, while the loop in hairpin I is closed by a C-G pair. All of the sequence share the common feature of a loop region consisting of five bases.

Hairpins II and III were not characterized in as great detail as hairpin I (see preceding paper). Proton and phosphorus NMR spectra were partially assigned by using 2D methods. NOEs[1] were not quantitated to obtain distances, but rather

† This work was supported by a grant from the National Institutes of Health (R01 GM27738). S.G.B. is the recipient of a Presidential Young Investigator Award.
‡ Present address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309.

0006-2960/89/0428-2831$01.50/0