

Comparison of the B- and Z-Form Hairpin Loop Structures Formed by $d(CG)_5T_4(CG)_5^{\dagger}$

Steven K. Wolk,[†] Charles C. Hardin,[†] Markus W. Germann,[§] Johan H. van de Sande,[§] and Ignacio Tinoco, Jr.*[†]

Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, California 94720, and Department of Medicinal Biochemistry, Faculty of Medicine, The University of Calgary, Calgary, Alberta T2N 4N1, Canada

Received January 15, 1988; Revised Manuscript Received May 10, 1988

ABSTRACT: The partially self-complementary synthetic DNA oligonucleotide $d(CG)_5T_4(CG)_5$ has been studied by using 1H and ^{31}P NMR and circular dichroism. Results show that, under low-salt conditions (120 mM NaCl buffer), an intramolecular hairpin loop exists in which the double-helical stem region is B-form and the thymidine loop residues have predominantly southern (C2'-endo) sugar conformations. The thymidine glycosidic torsion angles are intermediate between syn and anti or exist as an equilibrium mixture of residues in the two extremes. NOESY data indicate that the structure of the loop region is very similar to that found for $d(CG)_2T_4(CG)_2$ [Hare, D. R., & Reid, B. R. (1986) *Biochemistry* 25, 5341–5350]. Under high-salt conditions (6 M NaClO₄ buffer), the dominant form ($\approx 85\%$) is an intramolecular hairpin structure in which the stem region forms a Z-form double helix. As in the B-form, the loop thymidine residues are intermediate between the syn and anti conformations or exist as an equilibrium mixture of the two, but the thymidine sugar conformations differ in that they are biased toward northern (C3'-endo) conformations.

In recent years it has become increasingly clear that DNA has conformational variability far beyond the original B-form double helix proposed by Watson and Crick (1953). Right-handed A-form (Fuller et al., 1965), C-form (Marvin et al., 1961), D-form (Arnott et al., 1974), and left-handed Z-form (Pohl & Jovin, 1972; Wang et al., 1979) have all been reported. Short sequences that are not fully complementary have been shown to form double helices containing mismatched bases (Patel et al., 1982a), bulged bases (Patel et al., 1982b; Morden et al., 1983), and hairpin loops (Haasnoot et al., 1986). Very recently, it has been shown that the *Tetrahymena* telomere DNA sequence forms an intramolecular structure containing only non-Watson–Crick base pairs in 20 mM phosphate buffer (Henderson et al., 1987).

With the large range of accessible conformations, it is likely that more and more of these structures will be found to have biological importance. Detailed studies of these structures should provide clues that facilitate the discovery of these biological roles. We therefore undertook the study of the DNA oligonucleotide $d(5'-CGCGCGCGCTTTTCGCGCGC-GCG-3')$ [henceforth referred to as $d(CG)_5T_4(CG)_5$ or "24-mer"], which was reported to have the unusual property of forming an intramolecular hairpin loop structure (Figure 1a) with a Z-form double-helical stem region (Germann et al., 1985). Hairpin loops are clearly important in maintaining the structure of transfer RNAs. They may also occur in chromosomal regions containing nearly palindromic sequences, where they could play a role in transcription or other in vivo processes (Lilley, 1981; Courey & Wang, 1983; Lilley et al., 1987). The significance of Z-form structures has yet to be clearly established, though immunofluorescence experiments

suggest that Z-form tracts may exist in vivo in both DNA (Rich et al., 1984; Zarlring et al., 1984) and RNA (Zarlring et al., 1987). Possible biological roles of Z-DNA, e.g., in *Rec1*-catalyzed recombination (Kmiec & Holloman, 1986), and inhibition of methylation in plasmids by *EcoRI* methylase (Jaworski et al., 1987) are currently being investigated.

Oligonucleotides with hairpin loops of four thymidine residues and right-handed helical stems have been well characterized. Ikuta et al. (1986) found by NMR methods that $d(CG)_3T_4(CG)_3$ forms a unimolecular hairpin structure over a large range of experimental conditions and that the thymidine in the seventh position [T(7)] stacks upon the neighboring guanine [G(6)] at the loop/stem junction. Hare and Reid (1986) published the detailed structure of $d(CG)_2T_4(CG)_2$, using a combination of NMR and distance geometry methods. They showed that three of four thymidines stacked in a continuation of the stem helix [T(5) stacked upon G(4), and T(7) and T(8) upon C(9)], while the last [T(6)] was rotated away from the center of the loop. All sugar conformations were B-like, but there was considerable variation in the pseudorotation phase angle ($P = 137$ – 238°). Glycosidic torsion angles were also B-like and varied ($\chi = -52^\circ$ to -116°), but the thymidines were biased slightly toward syn conformations.

To our knowledge, only circular dichroism data have been presented as evidence for a Z-form hairpin structure in solution, and only for the sequences $d(CG)_3T_5(CG)_3$ (Xodo et al., 1986) and $d(CG)_5T_4(CG)_5$ (Germann et al., 1985). A Z-hairpin crystal structure of $d(CG)_3T_4(CG)_3$ has recently been reported (Chattopadhyaya et al., 1988) in which all four loop thymidines are rotated away from the center of the loop, with T(9) and T(10) stacking on one another. The authors noted, however, that intermolecular interactions were present in the crystal that may not be important in solution. We have thus used NMR methods to study the Z-form hairpin structure of $d(CG)_5T_4(CG)_5$ to (1) verify with interproton distances and phosphorous chemical shifts that the stem region is indeed Z-form, as CD evidence suggests, and (2) elucidate the structure of the loop region in order to compare the B- and Z-forms. The B-form hairpin structure of $d(CG)_5T_4(CG)_5$

[†] This work was supported by National Institutes of Health Grant GM 10840 (I.T.) and the U.S. Department of Energy Office of Energy Research, under contract DE-FG03-86ER60406 (I.T.). C.C.H. was supported by NIH–NRSA Postdoctoral Fellowship GM11638. M.W.G. was supported by the Alberta Heritage Foundation for Medical Research. J.H.v.d.S. was supported by the Medical Research Council of Canada.

[†] University of California.

[§] The University of Calgary.

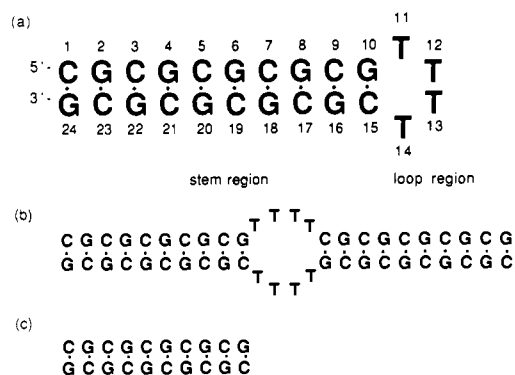


FIGURE 1: (a) Intramolecular hairpin and (b) intermolecular dimer forms of d(CG)₅T₄(CG)₅; (c) intermolecular dimer form of (CG)₅.

was studied to confirm that the longer stem region and slightly different buffer conditions did not significantly affect the loop structure found by Hare and Reid (1986).

MATERIALS AND METHODS

DNA oligonucleotides d(CG)₅T₄(CG)₅ and d(CG)₅ were synthesized by standard phosphoramidite methods, and subsequently purified by high-pressure liquid chromatography using a Hamilton PRP-1 column maintained at 60 °C. A gradient from 50 mM NaOH to 50 mM NaOH/50% acetonitrile was used as described previously (Germann et al., 1987).

Conditions referred to as "low salt" are 120 mM NaCl, 30 mM sodium phosphate, pH 7, and 0.5 mM EDTA.¹ "High-salt" conditions are 6 M NaClO₄, 10 mM sodium phosphate, pH 7, and 1 mM EDTA. All experiments were done at 37 °C. DNA concentrations were estimated by using approximate extinction coefficients ($\epsilon_{260} = 7000 \text{ M}^{-1} \text{ cm}^{-1}$ for the 24-mer).

NMR samples were prepared as follows: DNA samples were combined with 0.4 mL of the appropriate buffer, lyophilized repeatedly from 99.8% D₂O (Cambridge Isotope Laboratories), resuspended in 0.4 mL of 99.96% D₂O (Aldrich) under N₂, and placed in a 5-mm NMR tube (Aldrich, gold label). Final DNA concentrations were approximately 2 mM in strands. Partially deuterated sodium 3-(trimethylsilyl)-1-propanesulfonate (TSP) was used as the internal ¹H chemical shift reference.

NMR data were collected on either a Bruker AM-500 spectrometer equipped with an Aspect 3000 computer or a General Electric GN-500 spectrometer. Some spin-lattice relaxation experiments were repeated on both instruments to ensure reproducibility.

Selective relaxation data were acquired by using the standard $(180^\circ_{\text{selective}} - \tau - 90^\circ_{\text{nonselective}} - \text{observe})_n$ pulse sequence and data sets containing 8K points. Peaks of interest were irradiated by using a decoupler pulse of approximately 5 ms, resulting in 40–80% inversion. Experiments were actually semiselective in that many neighboring resonances were also partially irradiated. An 8-s relaxation delay was used to ensure equilibration between scans. Delay times between 0.1 ms and 10 s (τ_∞) were used, with most of the delay times shorter than 100 ms to facilitate initial slope calculations. Approximately 256 scans were acquired per delay time. One-dimensional nuclear Overhauser effect (NOE) data were collected in the same manner, alternating the selective pulse between on and off resonance (control spectrum) in cycles of 16 scans. Percent

NOE values were calculated as the ratio of the peak intensities in the NOE difference and control spectra.

³¹P NMR data were collected on the AM-500 spectrometer by using a broad-band probe tuned to the phosphorus frequency. Protons were broad band decoupled during the 1-s relaxation delay. Approximately 1200 scans were collected in 8K data sets, with either internal or external trimethyl phosphate (TMP) as the ³¹P chemical shift reference.

Circular dichroism spectra were collected on a Jasco J600 spectropolarimeter interfaced to an IBM PC computer. Samples were maintained at 37 °C in a 1-cm path length cylindrical cell surrounded by an external jacket for water circulation. Absorbance measurements were collected on a Shimadzu UV-160 spectrophotometer, maintained at 37 °C with a TCC-240 temperature-controlled cell holder. All CD and absorbance measurements were base line corrected for signals from the cell and buffer. DNA concentrations were approximately 20 μM in strands.

Two-dimensional phase-selective NOESY and magnitude COSY spectra were collected by using standard pulse sequences (Ernst et al., 1987) and TPPI phase cycling (Bodenhausen et al., 1984). NOESY data sets (2K points) were collected for each of 300–500 t_1 values, with 64–112 scans per t_1 . The relaxation delay between scans was 2 s. Data were transferred via magnetic tape to either a VAX 11/785 or a MicroVAX computer and processed by using software kindly provided by Dr. Dennis Hare. All data sets were Fourier transformed in the t_2 dimension, zero filled, and transformed in the t_1 dimension, yielding 1K \times 1K data sets. Data for figures were apodized in both dimensions prior to Fourier transformation by using a skewed phase-shifted sine bell that decayed to zero at the end of the data points. Data for cross-peak integration were not apodized in t_2 and were apodized by 90° phase-shifted sine bell in t_1 . Cross-peaks were integrated by using the elliptical fit routine provided with the processing software. COSY data sets (1K points) were collected for 512 t_1 values, 64 scans per t_1 ; they were apodized with an unshifted sine bell of length 512 points in both dimensions and were Fourier transformed to yield a 0.5K \times 0.5K data set.

RESULTS

d(CG)₅T₄(CG)₅ Forms Two Different Intramolecular Hairpin Structures. Portions of the low- and high-salt proton NMR spectra of d(CG)₅T₄(CG)₅ are shown in Figure 2. The upfield region contains the H2' and H2'' sugar proton resonances, and the thymidine methyl (T-CH₃) resonances. The presence of four large, sharp methyl peaks for the 24-mer under both low- (120 mM NaCl) and high-salt (6 M NaClO₄) conditions demonstrates the predominance of one form in each of these solutions. Smaller peaks in the high-salt spectrum demonstrate a contribution ($\approx 15\%$) from a second form. The percent contribution from this second form does not vary with temperature between –5 and 60 °C. Thus, the transition enthalpy is very small, or the kinetics of interconversion are slow (i.e., hours). The melting temperature (T_m) of the hairpin in high-salt conditions is about 75 °C (data not shown), suggesting that this second form is not a single strand. A 10-fold reduction in oligonucleotide concentration results in a 9% contribution of this second form.

The downfield region contains the aromatic GH8, CH6, and TH6 resonances. The 10 GH8 peaks in each spectrum overlap considerably around 7.9 ppm but are well resolved from the CH6 and TH6 peaks. However, the pyrimidine aromatic peaks overlap considerably with each other, limiting some aspects of structural analysis. The large peak at 7.30 ppm in

¹ Abbreviations: CD, circular dichroism; COSY, correlated spectroscopy; EDTA, ethylenediaminetetraacetic acid; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TMP, trimethyl phosphate; TPPI, time-proportional phase incrementation; TSP, sodium 3-(trimethylsilyl)-1-propanesulfonate.

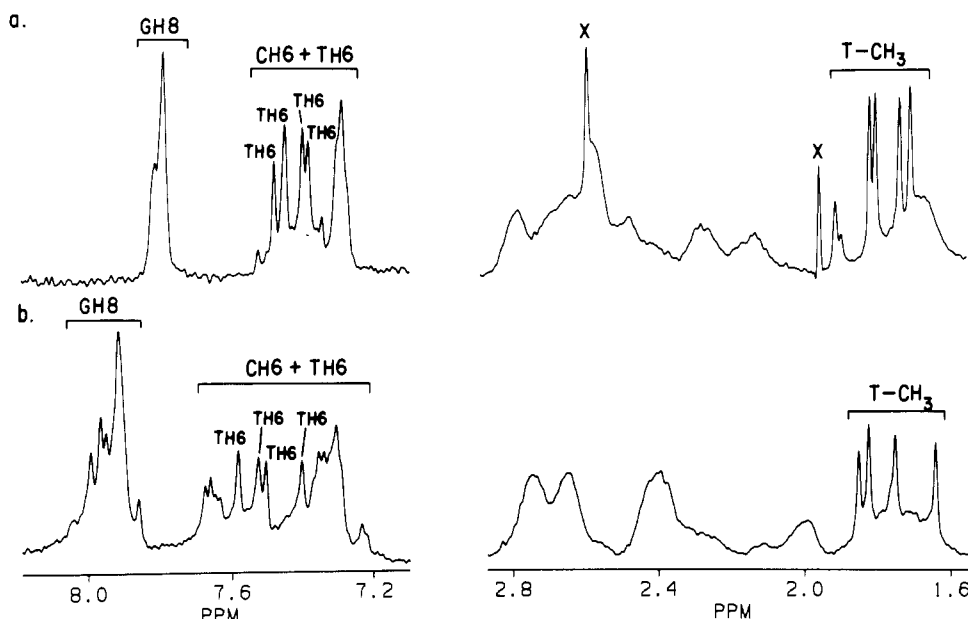


FIGURE 2: Aromatic and H2',H2'',methyl regions of the 500-MHz proton NMR spectrum of d(CG)₅T₄(CG)₅ in (a) 6 M NaClO₄ buffer and (b) 120 mM NaCl buffer at 37 °C. Samples were 2 mM in oligonucleotides. Impurity peaks are marked with an "X". Regions are not equally scaled.

the high-salt spectrum is due to six CH6 resonances; these peaks probably correspond to central residues in the stem structure, since they are expected to have similar magnetic environments.

It must first be determined whether the dominant form under low- and high-salt conditions is the intramolecular hairpin (Figure 1a). Gel electrophoresis results show this to be the case in 90 mM Tris-borate (pH 8.3) buffer (Germann et al., 1985). It is unclear, however, whether the same will be true in 6 M NaClO₄ buffer and/or at the oligonucleotide concentrations used in NMR experiments. Thus, NMR relaxation measurements were used to determine the size of the structure formed by the 24-mer. Because cross-relaxation with other protons is the dominant relaxation pathway (Abragam, 1961; Kearns, 1984), the selective spin-lattice relaxation rates (R_{1s}) are approximately proportional to the correlation time (τ_c) of the molecule, which is proportional to the length of short helices. Thus, if the 24-mer forms a hairpin, its correlation time should be roughly equivalent to that of d(CG)₅ (the 10-mer), though slightly larger. If a duplex with an internal loop (Figure 1b) is the dominant structure, then the correlation time should be about a factor of 2 larger. Wemmer and co-workers used this reasoning to explain observed R_{1s} values for the hairpin-to-duplex transition in an oligonucleotide containing twelve bases (2.0 s⁻¹ vs 4.2 s⁻¹; Wemmer et al., 1985).

Selective R_{1s} values for both oligonucleotides were determined by using the 180°_{selective}- τ -90°_{nonselective}-observe pulse sequence, and the intensities of the irradiated peaks measured as a function of the delay time τ . The CH6 resonances were irradiated, since they should be in very similar spin systems in both the 10-mer and the 24-mer, and all CH6 resonances were irradiated simultaneously. Some overlapping TH6 resonances were also irradiated, but this should not affect the results since the relatively large internuclear distance (>5 Å) prohibits significant cross-relaxation. Figure 3b shows a plot of $\ln(I_\infty - I_\tau)$ vs τ for an irradiated CH6 peak in the spectrum of the 24-mer in high salt; the initial slope is equal to $-R_{1s}$. Table I lists the R_{1s} values for the CH6 resonances that were resolved from TH6 resonances in the spectrum of the 24-mer in low and high salt and CH6 resonances in the control duplex

Table I: Spin-Lattice Relaxation Rates of Cytosine H6 Protons in d(CG)₅T₄(CG)₅ and d(CG)₅ at 37 °C

	high salt R_{1s} (s ⁻¹)	low salt R_{1s} (s ⁻¹)		high salt R_{1s} (s ⁻¹)	low salt R_{1s} (s ⁻¹)
d(CG) ₅	-4.8	-2.5	d(CG) ₅ T ₄ (CG) ₅	-4.9	-2.3
	-3.4	-1.8			
	-4.7	-2.5			

(10-mer) spectrum. As expected, both molecules give similar R_{1s} values in low salt, where the 24-mer is presumed to be a hairpin. In high salt, the rates for both molecules are larger, presumably due to the effect of viscosity on the molecular correlation times. As in the low-salt experiments, the rates of both molecules are the same within experimental error. Thus, the hairpin is also the dominant form in high-salt buffer. Several R_{1s} values seem uncharacteristically small, but these rates probably correspond to cytidine residues located near the end of the helix. These residues probably have additional motional freedom due to fraying (Patel et al., 1982b), which may account for the slower relaxation rates of these nuclei.

Since the 24-mer is a hairpin structure in high-salt solution, circular dichroism, ³¹P NMR, and ¹H NMR experiments were performed to verify that the double-stranded stem adopts a left-handed Z conformation and to elucidate details of the T₄ loop conformation.

Circular Dichroism and UV Absorption Spectroscopies.

Circular dichroism measurements of nucleic acids are sensitive to the stacking geometry of the bases, which is dependent on the helical conformation. A-form, B-form, and Z-form spectra each have unique characteristics. The CD spectra of the 24-mer in low and high salt are shown in Figure 4. These characteristics closely match those first observed by Pohl and Jovin (1972) for poly(dG-dC) in 0.2 M NaCl and 2.5 M NaCl solutions and those observed by Germann et al. (1985) for the 24-mer in 5 M NaCl. Due to the strong absorbance of chloride ion below 200 nm, data for the 24-mer in the 187–200-nm range in low-salt solution could not be obtained. Thus, the CD spectrum of the 24-mer was obtained in distilled deionized water (with residual Na⁺ concentration) and was compared to the low-salt spectrum. Since the low salt and H₂O spectra of the 24-mer in the 210–320-nm range match within ex-

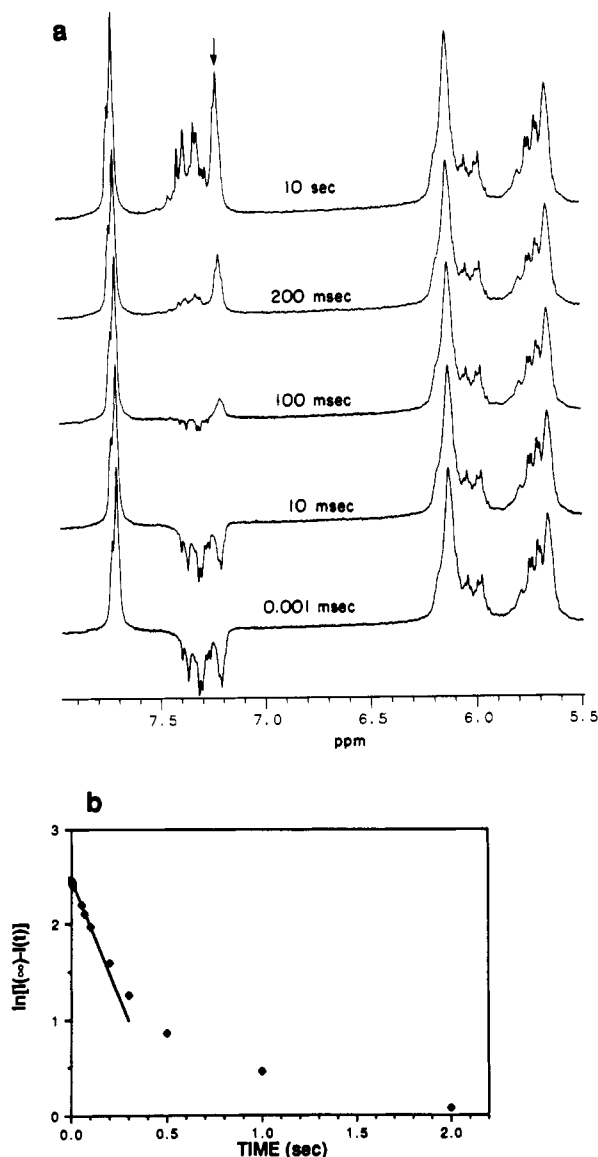


FIGURE 3: (a) Selective inversion-recovery data for $d(CG)_5T_4(CG)_5$ in 6 M $NaClO_4$ buffer. CH6 and TH6 resonances are inverted and plotted as a function of the recovery time (τ). (b) Plot of $\ln(I_\infty - I_\tau)$ vs τ for the group of CH6 resonances (denoted by an arrow in panel a). The initial slope is equal to $-R_{1\rho}$. (See text.)

perimental error (Figure 4a), it is reasonable to assume that the structure of the 24-mer is very similar under each of these conditions. Thus, for comparison of the 187–210-nm region of the CD spectra, the H_2O spectrum is contrasted with the high-salt spectrum. These results are very similar to the B- and Z-form spectra of poly(dG-dC) observed in 10 mM phosphate and 2 M $NaClO_4$ (Riazance et al., 1985). Thus, CD evidence suggests that the low-salt form of the 24-mer has a B-form stem region, while the high-salt form has a Z-form stem region. Figure 4a also includes the CD spectrum of the 24-mer in 90% trifluoroethanol, 1 mM NaCl, and 0.1 mM EDTA, pH 7, which has characteristics of a molecule with an A-form stem region. Further experiments to confirm this possibility are in progress.

Figure 4c shows the ratio of absorbance values at 295 and 260 nm in the spectrum of the 24-mer obtained as a function of $NaClO_4$ concentration. This ratio has been used as an indicator of the B to Z conformational transition in DNA (Pohl & Jovin, 1972). The A_{295}/A_{260} ratio indicates that at 37 °C the transition midpoint occurs at 2.1 M $NaClO_4$, and the transition is virtually complete by 4 M.

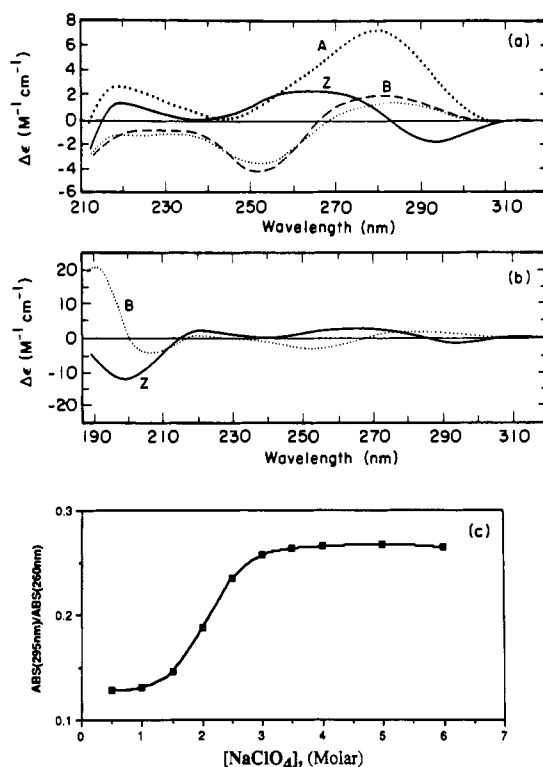


FIGURE 4: Circular dichroism spectra of $d(CG)_5T_4(CG)_5$. (a) In 6 M $NaClO_4$ buffer [Z-form (—)], 120 mM NaCl buffer [B-form (---)], H_2O [B-form (···)], and 90% trifluoroethanol, 1 mM NaCl, and 0.1 mM EDTA, pH 7 [A-form (•••)], in the 220–320-nm region. (b) In 6 M $NaClO_4$ buffer [Z-form (—)] and H_2O [B-form (···)], in the 187–320-nm region. (c) Transition profile of the B- to Z-form transition in $d(CG)_5T_4(CG)_5$ as a function of $NaClO_4$ concentration in 10 mM sodium phosphate (pH 7) and 1 mM EDTA at 37 °C. Transition is monitored by the A_{295}/A_{260} ratio.

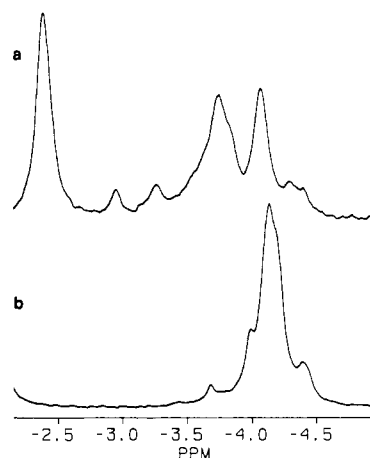


FIGURE 5: Proton-decoupled 202-MHz ^{31}P NMR spectra of $d(CG)_5T_4(CG)_5$ in (a) 6 M $NaClO_4$ buffer and (b) 120 mM NaCl buffer at 37 °C. Samples were 2 mM in oligonucleotides.

Phosphodiester Backbone Conformations. Figure 5 shows the phosphorus NMR spectrum of the 24-mer under low- and high-salt conditions. The low-salt spectrum (Figure 5a) shows that all of the backbone phosphodiester resonances are near -4 ppm, which is characteristic of right-handed oligo- and polynucleotides (Chen & Cohen, 1984). The resonances (-3.7, -4.0, and -4.4 ppm) resolved from the main group (-4.2 ppm) are likely to be from the phosphates in the loop region, since ^{31}P chemical shifts are known to be sensitive to the backbone torsion angles (Gorenstein, 1981), and the phosphodiester conformations in the loop are likely to be distorted relative to those in the stem region. The ^{31}P spectrum of the 24-mer

in high salt is shown in Figure 5b. Two main groups of peaks are observed at -2.4 and -3.9 ppm. This splitting pattern and these chemical shift values are characteristic of Z-form nucleic acids (Patel et al., 1982d; Sklenar et al., 1987). As in the low-salt spectrum, other peaks (-2.95, -3.3, -4.3, and -4.4 ppm) are dispersed from the main groups and are likely to be from the loop region. However, this spread is more pronounced in the Z-form spectrum. This may reflect greater variation in the average phosphodiester conformations that could be required to form a hairpin loop of four nucleotides spanning the strands terminating a Z-form stem.

Studies in Sodium Chloride and Sodium Bromide. Circular dichroism and ^{31}P NMR results suggest that the 24-mer also forms structures with a Z-form stem region in 5 M NaCl or 6 M NaBr buffers. However, the methyl region of the 500-MHz ^1H NMR spectra indicates the presence of two or more conformations in equilibrium under these conditions.

Structural Details of Z-Form $d(\text{CG})_5\text{T}_4(\text{CG})_5$. The most detailed structural information was obtained from a 2D NOESY experiment (Ernst et al., 1987). Interproton distances were calculated by using the approximation

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

where σ_{ij} and σ_{kl} are cross-relaxation rates between nuclei i and j and between nuclei k and l , respectively; r_{ij} and r_{kl} are the respective internuclear distances (Wagner & Wüthrich, 1979). Using two measured cross-relaxation rates, one corresponding to a known fixed distance within the molecule of interest, the remaining distance can be calculated. For experiments on DNA oligonucleotides, the cytosine H5/H6 distance (2.4 Å) is typically used as the reference; the sugar H2'/H2'' or thymidine H6/CH₃ distances can also be used (Hare & Reid, 1986). Cross-relaxation rates are estimated from the initial slopes of NOE buildup curves, i.e., plots of % NOE vs the time that the interaction was allowed to occur (the mixing time). The NOE intensity is measured as a cross-peak volume in the NOESY spectrum.

In order to determine mixing times appropriate for calculating σ values, one-dimensional NOE buildup curves were generated by irradiating the cytidine H6 protons of the six peaks corresponding to the most central cytidine residues in the stem region. Magnitudes of the NOEs to the cytidine H5 and H1' protons were measured. As mentioned previously, these protons all have the same chemical shift. From these data (not shown), it is evident that the linear region of the CH6/CH5 buildup curve occurs below 100 ms. However, for shorter times, the intensity of the CH6/CH1' NOE is close to the limit of reasonable signal to noise. For slightly larger distances, i.e., 4 Å, the signal to noise ratio becomes so small that accurate integration is unlikely. Thus, distance calculations for the Z-form hairpin were made from one NOESY experiment with a mixing time of 100 ms, with the average cytosine H5/H6 NOE intensity as the reference. To estimate the errors introduced in the latter case, a comparison was made between the σ value obtained from the 100-ms time point for the H5/H6 buildup and from the slope through all time points less than 50 ms. It was found that the latter calculation yielded a σ value 30% larger. When used in eq 1, this yields distance errors of about 5%, which are well within the inherent error in these measurements estimated by other authors (Fesik et al., 1986; Hare & Reid, 1986).

Assignments of the Z-form resonances were made from the appropriate regions of the NOESY spectrum. Each thymidine CH₃ resonance shows a cross-peak to the H6 resonance on the same residue, which in turn shows NOEs to the intrasidue H1', H2', H2'', and (in this case) H3' protons. Cross-peaks

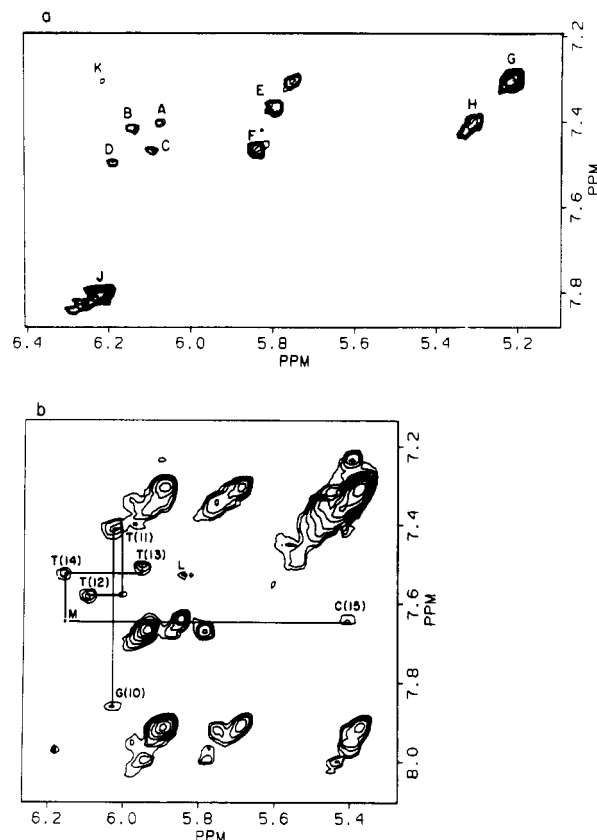


FIGURE 6: (a) Aromatic/H1',H5 cross-peak region of the 500-MHz NOESY spectrum of $d(\text{CG})_5\text{T}_4(\text{CG})_5$ in 6 M NaClO_4 buffer at 37 °C. The mixing time is 100 ms. Peaks A–D are intranucleotide TH6/TH1' cross-peaks. Peaks E, F, G, and H are 1, 1, 2, and 6 CH6/CH5 cross-peaks, respectively. Peak J is 10 nearly overlapping GH8/GH1' cross-peaks. Peak K is six overlapping CH6/GH1' cross-peaks. (b) Aromatic/H1',H5 cross-peak region of the 500-MHz NOESY spectrum of $d(\text{CG})_5\text{T}_4(\text{CG})_5$ in 120 mM NaCl buffer at 37 °C. Solid lines show the $[\text{H}1'(n) \rightarrow \text{H}6(n) \rightarrow \text{H}1'(n-1)]$ B-form connectivity path (Hare et al., 1983) which is broken between T(12) and T(13). Peak L is the C(15)H5/T(14)/H6 cross-peak. Peak M is the C(15)H6/T(14)H1' cross-peak.

between these deoxyribose resonances are also observed, allowing assignment of the H6, CH₃, H1', H2', H2'', and H3' resonances for each thymidine residue. These results are listed in Table II. The thymidine protons of one of these groups (1.81 ppm) were found to be in NOE contact with an H1' and an H2'' proton belonging to a guanosine residue (data not shown); thus, at least one thymidine residue is proximal to the G(10) sugar moiety. No NOEs were found that could unambiguously assign the thymidine residues.

Cytidine H6 and H5 resonances were assigned from the aromatic/H5,H1' region of the COSY spectrum (data not shown). From the intensities of these same cross-peaks in the NOESY spectrum (Figure 6a), it is clear that these peaks (E, F, G, and H) represent 1, 1, 6, and 2 H6/H5 cross-peaks, respectively. This supports our earlier assumption that the six most internal cytidine residues have virtually identical magnetic environments. Peak J (Figure 6a) contains the overlapping guanosine H8/H1' cross-peaks; its large magnitude is indicative of the syn conformation expected for guanosine residues in a Z-form stem region (see below).

Cross-peak intensities were determined and distances were calculated by using eq 1 and the above assignments. For equivalent cross-peaks from superimposed residues (i.e., the six superimposed H6 protons giving NOEs to the corresponding six superimposed H5 protons), cross-peak volumes were divided by the number of contributing resonances and

Table II: Chemical Shift Values of $d(CG)_5T_4(CG)_5$ Peaks in Low- and High-Salt Conditions

	Low Salt (B-Form)					C(15)
	T(11)	T(12)	T(13)	T(14)	G(10)	
H6/H8	7.40	7.58	7.50	7.53	7.86	7.63
H1'	6.00	6.09	5.95	6.16	6.02	
H2''	2.43	2.39	2.28	2.55		
H2'	2.09	2.12	2.05	2.32		
H3'	4.91	4.65	4.73	4.72		
CH ₃ /H5	1.75	1.82	1.64	1.85		5.84

	High Salt (Z-Form)					6C(i) ^a
	T(D)	T(A)	T(B)	T(C)	G(10)	
H6/H8	7.48	7.39	7.40	7.46		7.30
H1'	6.18	6.06	6.13	6.08	6.27	5.75
H2''	2.48	2.27	2.24	2.30	2.61	2.57
H2'	2.28	2.17	2.13	2.12		1.66
H3'	4.78	4.73	4.54	4.68		
CH ₃ /H5	1.81	1.70	1.73	1.80		5.21

^aChemical shift values for the overlapping cytidine resonances C(3) through C(21) (see text).

compared to single cross-peak intensities. Therefore, the calculated distances represent an average for these residues. When this NOESY experiment was done, the GH8 protons had already exchanged approximately 40% with D₂O. Thus, cross-peak intensities involving the GH8 peaks were scaled by a factor of 1/0.6 to correct for this. Table III shows a summary of the distances calculated from the 100-ms NOESY and those same distances from X-ray crystallographic studies of Z-form oligonucleotides [Wang et al., 1981; Orbons et al. (1986) and references cited therein]. With a few exceptions, these distances are in excellent agreement with the crystallographic results, providing strong evidence that the stem region of the 24-mer is in the Z conformation under high-salt conditions. In the cases that showed poor agreement, the distances calculated from NMR data were usually shorter. This may reflect a contribution from spin diffusion or dynamic motion of the molecule in solution, since the inverse sixth power distance dependence of the NOE biases a time-averaged distance toward the shorter extreme.

Although the lack of internucleotide NOEs in the loop region precludes assignments, intranucleotide NOEs do allow estimation of the glycosidic torsion angle (χ) and pseudorotation phase angle (P) [Wüthrich (1986), Chapter 11] in these nucleotides. Computed distances are listed in Table III. The H6/H1' distance is a very sensitive measure of χ and is nearly independent of P . For a nucleotide in the anti range [$\chi \cong -140^\circ$, according to the convention of Saenger (1984)], this distance is about 3.7 Å. For the syn conformation ($\chi \cong +60^\circ$), this distance decreases to 2.3 Å. Table III shows that the H6/H1' distances in all of the Z-form loop thymidine residues are in the 2.7–2.9-Å range, clearly intermediate between the distances listed above. These nucleotides therefore have nonstandard glycosidic conformations ($\chi \cong 0^\circ$) or are present as an equilibrium mixture of syn and anti conformations in fast exchange.

Intranucleotide distances can also be indicative of the pseudorotation phase angle, P . One informative distance for distinguishing northern conformations [$P = 0 \pm 90^\circ$, including C3'-endo ($P = 18^\circ$)] from southern conformations [$P = 180 \pm 90^\circ$, including C2'-endo ($P = 162^\circ$)] is the H6/H3' distance, though it is also sensitive to χ . For the χ values consistent with measured H6/H1' distances, the H6/H3' distance changes from 3.0 Å (north) to 4.5 Å (south). From the H6/H3' distances listed in Table III, all four Z-form thymidine loop sugars appear to be in predominantly northern (C3'-endo) conformations. The H2''/H4 distance is also a good indicator

Table III: Calculated Distances in $d(CG)_5T_4(CG)_5$ in High-Salt Solution (Z-Form) Compared to Those Obtained from Crystallographic Studies^a

	Stem Region Distances (Å)		
	NMR ^b	$d(CG)_3$ ^c	$d(m^5C-G)_3$ ^d
CH6/GH1'	4.1 ^e	5.6	4.9
CH6/CH1'	3.4 ^e	3.5	3.7
CH6/CH2'	2.9 ^e	3.3	3.1
CH6/GH2'	2.7 ^e	3.0	2.6
CH6/GH2''	3.1 ^e	4.2	4.0
CH5/GH2''	3.1 ^e	4.2	4.1 ^g
CH5/GH2'	2.7 ^e	3.1	2.9 ^g
GH8/GH1'	2.3 ^f	2.3	2.3
CH5/GH8	4.4 ^e	4.9	4.3 ^g

Intraresidue Loop Thymidine Distances, Z-Form (Å)				
	T(D) = T(11)	T(A)	T(B)	T(C)
TH6/TH1'	2.8	2.9	2.9	2.7
TH6/TH2''	3.3	2.7 ^h	2.7 ^h	
TH6/TH2'		2.3 ^h	2.3 ^h	2.3
TH6/TH3'	3.3	2.8	3.1	3.0

^aInternucleotide distances are $n/(n-1)$. ^bErrors are approximately ± 0.2 Å for distances less than 3 Å and ± 0.5 Å for distances greater than 3 Å. ^cFrom Wang et al. (1981). ^dFrom Orbons et al. (1986) and references cited therein. ^eAverage of six internal cytosine residues with identical chemical shifts (see text). ^fAverage of ten guanosine residues, scaled for exchange with D₂O (see text). ^gDistances are to methyl protons of m⁵C instead of CH5. ^hAverage values of residues T(A) and T(B).

of sugar pucker, varying from 2.3 to 4.0 Å in the C3'-endo and C2'-endo conformers, respectively. However, these cross-peaks were not sufficiently resolved to be interpreted.

Structural Details of B-Form $d(CG)_5T_4(CG)_5$. Figure 6b shows the aromatic/H1',H5 region of the NOESY spectrum collected in low-salt solution. Standard assignment methods were used (Hare et al., 1983), and chemical shifts are listed in Table II. Poor spectral resolution resulted in significant cross-peak overlap and restricted the number of distances that could be calculated from the data. However, the qualitative similarity of chemical shifts and cross-peaks between this spectrum and that of $d(CG)_2T_4(CG)_2$ in 20 mM sodium phosphate, pH 7, at 25 °C (Hare & Reid, 1986) indicates that the structure of the T₄ loop region is essentially the same in these molecules.

DISCUSSION

One- and two-dimensional NMR experiments indicate that $d(CG)_5T_4(CG)_5$ exists in two slowly exchanging forms at millimolar DNA concentrations in high-salt solution (6M NaClO₄, 10 mM sodium phosphate, pH 7, and 1 mM EDTA) and in one form in low-salt solution (120 mM NaCl, 30 mM sodium phosphate, pH 7, and 0.5 mM EDTA) at 37 °C. Selective spin-lattice relaxation measurements indicate that the dominant form ($\cong 85\%$) in high salt is the hairpin monomer shown in Figure 1a. Because the T_m of the hairpin is approximately 75 °C under these conditions, the second form is not likely to be a single-stranded state. Simple two-state theory for a monomer/dimer equilibrium (Figure 1a,b) predicts that a 10-fold decrease in concentration should result in a 2% contribution of the dimer form; empirically, a 9% contribution is observed. Thus, the concentration dependence implies that the minor conformer is a nonmonomeric species such as the dimer; however, the system cannot be described by a simple two-state equilibrium.

Circular dichroism, ³¹P NMR, and ¹H NMR results all indicate that in low-salt conditions the stem region forms a B-family double helix. In the high-salt conditions, the results of these experiments are consistent with a Z-form stem. CD spectra were previously published for this molecule in 5 M

NaCl (Germann et al., 1985); the results in 6 M NaClO₄ were comparable.

One of the most interesting questions, which to our knowledge has not been studied previously, addresses how the structural transition in the stem region affects the loop conformation in solution. NOESY data indicate that the T₄ loop structure of (CG)₅T₄(CG)₅ in 120 mM NaCl buffer is similar to the T₄ loop structure of (CG)₂T₄(CG)₂ in 10 mM sodium phosphate buffer found by Hare and Reid (1986), demonstrating that the changes in stem length and solvent conditions had little effect. Hare's distance geometry structures show that T(8) and T(7) stack on C(9) in an apparent continuation of the 5' side of the helix, while T(5) stacks on G(4) to start the 3' side of the stem helix. T(6) is rotated away from the center of the helix; it is still in NOE contact with T(5), but not with T(7). In contrast, the lack of NOE contacts in the Z-hairpin loop suggests that it is considerably less compact or more flexible than the B-form loop. It is likely that all but one of the thymine residues are rotated away from the center of the loop region, thus decreasing thymidine-thymidine interactions and increasing their exposure to the solvent. This is consistent with the absence of thymidine imino proton resonances for the Z-hairpin loop, even at pH 5.5.

Intranucleotide NOE intensities allow comparison of the glycosidic torsion angles and pseudorotation phase angles of the thymidine residues in the two forms of the DNA hairpin. Distances calculated from NOE data, especially from one mixing time, should be interpreted cautiously. For example, spin diffusion effects can cause intranucleotide H6/H1' and H6/H3' distances to appear shorter than they actually are due to transfer of magnetization through the H2' proton. In addition, dynamic effects such as pseudorotation of the sugar residue give a time-averaged inverse sixth power distance value, which is biased toward the shorter distances. Errors in these distances have been suggested to be ± 0.2 Å for strong NOEs (less than about 3 Å) and ± 0.5 Å for weak NOEs (Hare & Reid, 1986). Thus we have limited our structural conclusions on the basis of these distances to a qualitative level that includes consideration of these uncertainties.

The H6/H1' distance is most sensitive to changes in the glycosidic torsion angles. For a nucleotide in the anti conformation, this distance should be approximately 3.7 Å; in the syn conformation, it shortens to 2.3 Å. In the Z-form loop, all thymidine H6/H1' distances were in the 2.7–2.9-Å range, consistent with χ values of $\approx 0^\circ$. Hare and Reid (1986) report χ values that ranged from -53° to -90° for B-form loop residues. Thus, in both the B- and Z-form hairpin loops the thymidine glycosidic torsion angles appear to be intermediate between the syn and anti conformations or exist as an equilibrium between the two. In the Z-form hairpin, the T residues are more biased toward the syn conformation.

The pseudorotation phase angle of the deoxyribose moieties can be estimated from aromatic-to-sugar and sugar-to-sugar interproton distances. The most sensitive distances among the assigned protons are the H6/H3' and the H2''/H4' (Wüthrich, 1986). The H2''/H4' cross-peaks were not sufficiently resolved to allow interpretation. Consideration of the magnitudes of the H6/H3' cross-peaks for the Z-form loop (distances of about 3 Å) is consistent with a larger contribution from northern (C3'-endo) conformations. This is in contrast with the B-form results, which indicate predominantly southern (C2'-endo) conformations.

ADDED IN PROOF

We have recently become aware of work by Wang et al. (*Biochemistry*, submitted for publication), who have used

NMR and circular dichroism to study the Z-hairpin formed by (CG)₃T₄(CG)₃.

Registry No. 24-mer, 97950-79-3; d(CG)₅, 114742-74-4.

REFERENCES

- Abraham, A. (1961) *Principles of Nuclear Magnetism*, pp 289–305, Oxford University Press, Oxford.
- Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C., & Watts, L. (1974) *J. Mol. Biol.* 88, 523–533.
- Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys.* 64, 2229–2246.
- Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson.* 58, 370–388.
- Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K., & Dickerson, R. E. (1988) *Nature (London)* 334, 175–179.
- Chen, C.-W., & Cohen, J. S. (1984) in *Phosphorous-31 NMR, Principles and Applications* (Gorenstein, D., Ed.) pp 233–263, Academic, Orlando.
- Courey, A. J., & Wang, J. C. (1983) *Cell (Cambridge, Mass.)* 33, 817–829.
- Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Chapters 8 and 9, Clarendon, Oxford.
- Fesik, S. W., O'Donnell, T. J., Gampe, R. T., Jr., & Olejniczak, E. T. (1986) *J. Am. Chem. Soc.* 108, 3165–3170.
- Fujii, S., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1982) *Nucleic Acids Res.* 10, 7879–7892.
- Fuller, W., Wilkins, M. H. F., Wilson, H. R., & Hamilton, L. D. (1965) *J. Mol. Biol.* 12, 60–80.
- Germann, M. W., Schoenwalder, K.-H., & van de Sande, J. H. (1985) *Biochemistry* 24, 5698–5702.
- Germann, M. W., Pon, R. P., & van de Sande, J. H. (1987) *Anal. Biochem.* 165, 399–405.
- Gorenstein, D. G. (1981) *Annu. Rev. Biophys. Bioeng.* 10, 355–386.
- Gronenborn, A. M., Clore, G. M., & Kimber, B. J. (1984) *Biochem. J.* 221, 723–736.
- 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.
- Henderson, E., Hardin, C. C., Wolk, S., Tinoco, I., Jr., & Blackburn, E. H. (1987) *Cell (Cambridge, Mass.)* 51, 899–908.
- Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., & Kearns, D. R. (1986) *Biochemistry* 25, 4840–4849.
- Jaworski, A., Hsieh, W.-T., Blaho, J. A., Larson, J. E., & Wells, R. D. (1987) *Science (Washington, D.C.)* 238, 773–777.
- Kearns, D. R. (1984) *CRC Crit. Rev. Biochem.* 15, 237–290.
- Kmiec, E. B., & Holloman, W. K. (1986) *Cell (Cambridge, Mass.)* 44, 545–554.
- Lilley, D. M. J. (1981) *Nucleic Acids Res.* 9, 1271–1289.
- Lilley, D. M. J., Sullivan, K. M., & Murche, A. I. H. (1987) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) pp 126–137, Springer-Verlag, New York.
- Marvin, D. A., Spencer, M., Wilkins, M. H. F., & Hamilton, L. D. (1961) *J. Mol. Biol.* 3, 547–565.
- Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry* 22, 5557–5563.

- Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1986) *Eur. J. Biochem.* 160, 131-139.
- Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., & Breslauer, K. (1982a) *Biochemistry* 21, 437-444.
- Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. (1982b) *Biochemistry* 21, 428-436.
- Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. (1982c) *Biochemistry* 21, 445-451.
- Patel, D. J., Kozlowski, S. A., Nordheim, A., & Rich, A. (1982d) *Proc. Natl. Acad. Sci. U.S.A.* 79, 1413-1417.
- Pohl, F. M., & Jovin, T. M. (1972) *J. Mol. Biol.* 67, 375-396.
- Riazance, J. H., Baase, W. A., Johnson, W. C., Jr., Hall, K. H., Cruz, P., & Tinoco, I., Jr. (1985) *Nucleic Acids Res.* 13, 4983-4989.
- Rich, A., Nordeim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem.* 53, 791-846.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*, p 22, Springer-Verlag, New York.
- Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc.* 109, 2221-2222.
- Wagner, G., & Wüthrich, K. (1979) *J. Magn. Reson.* 33, 675-680.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature (London)* 282, 743-745.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) *Science (Washington, D.C.)* 211, 171-176.
- Watson, J. D., & Crick, F. H. C. (1953) *Nature (London)* 171, 737-738.
- Wemmer, D. E., Chou, S. H., Hare, D. R., & Reid, B. R. (1985) *Nucleic Acids Res.* 13, 3755-3772.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, pp 203-223, Wiley, New York.
- Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. A. (1986) *Nucleic Acids Res.* 14, 5389-5397.
- Zarling, D. A., Arndt-Jovin, D. J., Robert-Nicoud, M., McIntosh, L. P., Thomae, R., & Jovin, T. M. (1984) *J. Mol. Biol.* 176, 369-415.
- Zarling, D. A., Calhoun, C. J., Hardin, C. C., & Zarling, A. H. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 6117-6121.

Phased Psoralen Cross-Links Do Not Bend the DNA Double Helix[†]

Tali E. Haran and Donald M. Crothers*

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Received December 22, 1987; Revised Manuscript Received April 21, 1988

ABSTRACT: Although the chemical reaction of psoralens with nucleic acids is well understood, the structure of psoralen-DNA cross-linked products is still not clear. Model building studies based on the crystal structure of the psoralen-thymine monoadduct suggest that each cross-link bends the DNA double helix by 46.5° [Pearlman, D. A., Holbrook, S. R., Pirkle, D. H., & Kim, S.-H. (1985) *Science (Washington, D.C.)* 227, 1304-1308]. On the other hand, Sinden and Hagerman [Sinden, R. R., & Hagerman, P. J. (1984) *Biochemistry* 23, 6299-6303] find that, in solution, psoralen cross-linked DNA is not bent. Here we use gel electrophoresis to test the validity of the current models. We have synthesized a series of DNA fragments (21-24 base pairs in length), each containing one unique T-A site for 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen (HMT) cross-linking. Because of an estimated 28° unwinding of the helix by HMT [Wiesehahn, G., & Hearst, J. E. (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 2703-2707], one expects that the 22-bp cross-linked fragment will be repeated nearly in phase with the average helical screw when multimerized. In that sequence ligation will maximally amplify any deformation to the double helix. We find that the ligated multimers of cross-linked DNA migrate close to the multimers of non-cross-linked DNA on polyacrylamide gels. Our observations place an upper limit of 10° on DNA bending induced by psoralen cross-linking and indicate unwinding by about 1 bp, as well as stiffening of the double helix. These properties are not unexpected for classical intercalators.

The use of psoralens in medicine started with the ancient Hindus, Turks, and Egyptians more than 3000 years ago and has been practiced ever since. In particular, the psoralens (furocoumarins) have been used in the treatment of vitiligo, a disease in which the skin pigment melanin is lost from certain skin areas, and psoriasis, a disorder showing an increased turnover rate of the epidermal tissue [reviewed by Scott et al. (1976) and Ben Hur and Song (1984)]. The successful use of psoralens in medicine has been linked to their ability to

cross-link adjacent pyrimidine bases on two strands of the DNA double helix [reviewed by Scott et al. (1976) and Song and Tapley (1979)]. As cross-linking agents psoralens are popular probes of the structure of DNA, RNA, and chromatin [reviewed by Cimino et al. (1985)]. At present we understand the chemical reactions of psoralens with nucleic acids [reviewed by Hearst (1981) and Cimino et al. (1985)], but the spatial arrangement of psoralen cross-linked DNA is not clear.

Understanding the structure of psoralen-DNA complexes is important for the design and synthesis of new and better psoralen derivatives, with greater therapeutic potential and less severe side effects. They can be used in the design of more sensitive probes of nucleic acid conformation and for eluci-

[†]This work was supported by National Institutes of Health Grant CA15583 (to D.M.C.) and by a Rothschild postdoctoral fellowship (to T.E.H.).