B-DNA backbone. Currently, we are investigating the antibiotic properties of the new compound for its therapeutic potential.

ACKNOWLEDGMENTS

We thank Pradipkumar Parrack, Indira Chandrasekhar, and B. Gopalakrishnan for their help with the computer drawings of the backbone structures of the ligands, Dr. Probal Banerji for helping in synthesis, and Rajani Srinivasan and S. L. Nagaraj (Department of Biochemistry) for their technical assistance. We also thank the Department of Biochemistry for the use of CD spectropolarimeter and spectrofluorometer facilities. We gratefully acknowledge Prof. H. T. Miles, NIAAKD, Bethesda, MD, for the kind gift of poly($^2NH_2$-dA-dT) to Prof. V. Sasisekharan.

Registry No. Dst, 636-47-5; mPD, 113474-83-2; NSC 101327, 16758-33-1; poly(dG-^{5}MedC), 51853-63-5.

REFERENCES

Atwell, G. J., & Cain, B. F. (1968) *J. Med. Chem. 11*, 295-300.
Baguley, B. C. (1982) *Mol. Cell. Biochem. 43*, 167-181.
Behe, M., & Felsenfeld, G. (1981) *Proc. Natl. Acad Sci. U.S.A. 78*, 1619-1623.
Borah, B., Cohen, J. S., Howard, F. B., & Miles, H. T. (1985) *Biochemistry 24*, 7456-7462.
Dasgupta, D., Rajagopalan, M., & Sasisekharan, V. (1986) *Biochem. Biophys. Res. Commun. 140*, 626-631.
Dasgupta, D., Parrack, P., & Sasisekharan, V. (1987) *Biochemistry 26*, 6381-6387.
Goodsell, D., & Dickerson, R. E. (1986) *J. Med. Chem. 29*, 727-733.
Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D., & Schyolkina, A. K. (1974) *J. Mol. Biol. 87*, 817-833.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1372-1380.
Le Pecq, J. B., & Paoletti, C. (1967) *J. Mol. Biol. 27*, 87-106.
Lown, J. W., Krowicky, K., Bhat, U. G., Skorobogaty, A., Ward, B., & Dabrowiack, C. (1986) *Biochemistry 25*, 7408-7416.
Luck, G., Zimmer, Ch., & Baguley, B. C. (1984) *Biochim. Biophys. Acta 782*, 41-48.
Parrack, P., Dasgupta, D., Ayyer, J., & Sasisekharan, V. (1987) *FEBS Lett. 212*, 297-301.
Pullman, B. (1986) in *Molecular Mechanisms of Carginogenic and Antitumor Activity* (Pullman & Chagas, Eds.) Pontificiae Academiae Scientiarvm Documenta, Vatican Press.
Scatchard, G. (1949) *Ann. N.Y. Acad. Sci. 51*, 660-663.
Silverstein, R. M., Bassler, G. C., & Morrill, T. C. (1981) in *Spectrometric Identification of Organic Compounds*, p 329, Wiley, New York.
Wade, W. S., & Dervan, P. B. (1987) *J. Am. Chem. Soc. 109*, 1574-1575.
Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E., & Cantor, C. R. (1970) *J. Mol. Biol. 54*, 465-497.
Zimmer, Ch. (1975) *Prog. Nucleic Acid Res. Mol. Biol. 15*, 285-318.
Zimmer, Ch., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31-112.
Zimmer, Ch., Reinert, K. E., Luck, G., Wahnert, U., Lober, G., & Thrum, H. (1971) *J. Mol. Biol. 58*, 329-348.

Hairpin and Duplex Formation in DNA Fragments CCAATTTTGG, CCAATTTTTTGG, and CCATTTTTGG: A Proton NMR Study†

Pradip Pramanik, Neera Kanhouwa,‡ and Lou-sing Kan*
Division of Biophysics, SHPH, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, Maryland 21205
Received September 14, 1987; Revised Manuscript Received December 21, 1987

ABSTRACT: Three DNA fragments, CCAATTTTGG (**1**), CCAATTTTTTGG (**2**), and CCATTTTTGG (**3**), were studied by proton NMR spectroscopy in aqueous solution. All these oligodeoxyribonucleotides contain common sequences at the 5′ and 3′ ends (5′-CCA and TGG-3′). **2** as well as **3** forms only hairpin structures with four unpaired thymidylyl units, four and three base pair stems, respectively, in neutral solution under low and high NaCl concentrations. At high salt concentration the oligomer **1** forms a duplex structure with -TT- internal loop. On the other hand, the same oligomer forms a stable hairpin structure at low salt and low strand concentrations at pH 7. The hairpin structure of **1** has a stem containing only three base pairs (CCA·TGG) and a loop containing four nucleotides (-ATTT-) that includes a dissociated A·T base pair. The two secondary structures of **1** coexist in an aqueous solution containing 0.1 M NaCl, at pH 7. The equilibrium shifts to the hairpin side when the temperature is raised. The stabilities and base-stacking modes of all three oligonucleotides in two different structures are reported.

Hairpin loops in DNA, though rare, play an important role in gene control mechanisms (Muller & Fitch, 1982; Weaver & DePamphilis, 1984). Inverted repeat sequences in superhelical DNA extrude into cruciform structures (Sinden & Pettijohn, 1984; Frank-Kamenetskij & Vologodskij, 1984; Lilley, 1981; Sheflin & Kowalski, 1985) that appear to be involved in transcription and replication (Gierer, 1966). In hybridization experiments used in molecular cloning, short oligodeoxynucleotides are used. In the presence of appropriate complementary sequence and the low-concentration conditions used in hybridization experiments, hairpin loop formation may be important. A thorough knowledge of the conditions that favor loop formation is also expected to give a better under-

†This work was supported by NIH Grant GM 34252-08. Preliminary results of this work were presented at the Fifth Conversation in the Discipline of Biomolecular Stereodynamics, SUNY, Albany, NY, June 2-6, 1987.
‡Present address: Tufts University School of Medicine, Boston, MA.

0006-2960/88/0427-3024$01.50/0

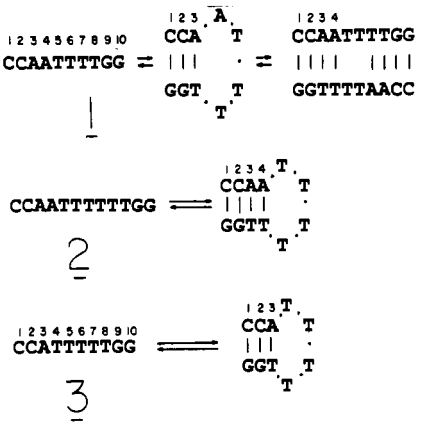

FIGURE 1: Scheme of primary and secondary structures of d-CCAATTTTGG (**1**), d-CCAATTTTTTGG (**2**), and d-CCATTTTTGG (**3**). Oligomer **1** is capable of forming both duplex and hairpin structures, whereas **2** and **3** form only hairpins.

standing of many biological processes. These and other reasons have spurred interest in the study of DNA hairpin structure in recent years (Haasnoot et al., 1986; Ikuta et al., 1986; Summers et al., 1985; Wemmer et al., 1985; Patel et al., 1983). Here, we report our NMR[1] study on three DNA fragments, CCAATTTTGG (**1**), CCAATTTTTTGG (**2**), and CCATTTTTGG (**3**), containing stem sequences that are different from those used by others.

Experimental Procedures

The oligodeoxynucleotides used here were synthesized by a standard solid-phase phosphoramidite triester synthesis procedure (Atkinson & Smith, 1984; Yuhasz et al., 1987). NMR measurements were performed with samples in 5-mm NMR tubes containing either 0.35 or 0.40 mL of solution. The solutions of the oligonucleotides each contained 0.01 M sodium phosphate buffer (pH 7.0) and 0.1 mM EDTA solutions. The single-strand concentrations of the oligonucleotides were 1.3 mM. Exceptions are noted in the figure legends. Proton NMR spectra of the oligonucleotide solutions were recorded with a homemade 498-MHz instrument located at the Francis Bitter National Magnet Laboratory at MIT and a Bruker WM-300 NMR spectrometer located in Biophysics NMR Facility Center, Johns Hopkins University. Imido proton NMR spectra of the oligonucleotide solutions in 90% H_2O/10% D_2O (v/v) were recorded by using a time-shared Redfield long pulse to suppress the water signal (Redfield et al., 1975; Haasnoot & Hilbers, 1983). Nuclear Overhauser enhancement (NOE) experiments on the exchangeable protons were carried out by collecting free induction decays (FIDs) in the interleave mode for a set of decoupling frequencies. NOEs were then determined by subtracting each on-resonance FID from the off-resonance FID. A base-line correction routine was used to straighten the wavy base line that is associated with the 214-pulse sequence. The chemical shifts were calculated with respect to that of the internal H_2O/HDO peak. Details of the selective inversion recovery using a DANTE 180° pulse (Morris & Freeman, 1978), magnetization transfer, and 2-D COSY (Ernst et al., 1987a) and NOESY (Ernst et al., 1987b) experiments are given in the figure legends. The samples used in these experiments were prepared by lyophilizing and dissolving in 99.5% D_2O and then lyophilizing again

[1] Abbreviations: 2D-COSY, two-dimensional correlated spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; FID, free induction decay; EDTA, (ethylenedinitrilo)tetraacetic acid.

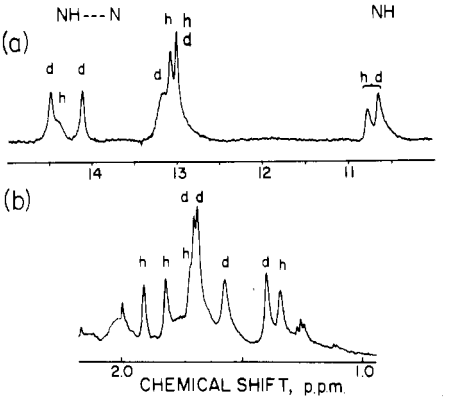

FIGURE 2: ^{1}H NMR spectra of **1** containing 0.1 M NaCl, at 498 MHz: (a) imido proton signal region at 5 °C; (b) methyl proton signal region at 10 °C. Symbols d and h represent duplex and hairpin, respectively.

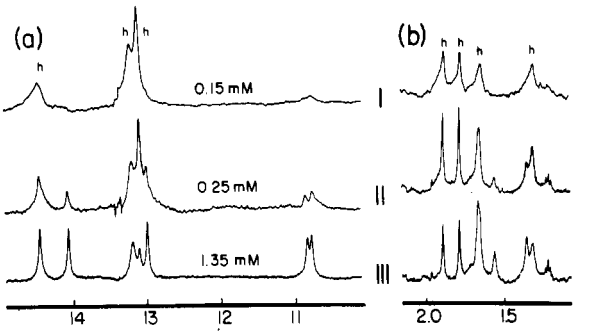

FIGURE 3: 498-MHz ^{1}H NMR spectra of **1** with variation of oligonucleotide concentration at 0 °C. Rows I, II, and III are 0.15, 0.25, and 1.35 mM of single-strand concentration, respectively. The contents in panels a and b are the same as in Figure 2. Only the hairpin species (h) exists at low concentration. The free NH resonances are very broad at low strand concentration due to the fast exchange rate.

and dissolving in 99.96% D_2O obtained from the Sigma Chemical Co. The 2D-COSY data were collected in the magnitude mode. The NOESY data were acquired in the absorption mode by using phase-sensitive methods (States et al., 1982). The 2D spectra presented here were symmetrized. Symmetrized and unsymmetrized spectra were compared to distinguish the artifacts from the real cross peaks. The chemical shifts were calculated with respect to sodium 4,4-dimethyl-4-silapentane-1-sulfonate.

Results and Discussion

The oligonucleotides reported here were found to have the structures shown in Figure 1 under the conditions of studies in this paper. (The nonpalindromic parts of the oligonucleotides are enclosed in parenthesis for ease of understanding.)

(*1*) *CCAA(TT)TTGG* (**1**). (*a*) *Imido and Thymine CH_3 Proton Resonance.* From examination of the sequence of the oligonucleotide it is expected that it would form a duplex secondary structure (Tinoco et al., 1973; Breslauer et al., 1986) in solution. There are, however, six peaks in the low-field interstrand hydrogen-bonding region, 12–15 ppm (instead of the four expected for the duplex structure) and two peaks in the nonhydrogen-bonded imido proton region, 10–11 ppm (Haasnoot et al., 1983) in the spectrum of the oligonucleotide at low temperature (Figure 2a). Neither the duplex with an internal loop nor the hairpin structure of oligonucleotide **1**, shown in Figure 1, could alone account for all the peaks seen in the spectrum. Eight thymine methyl (T-CH_3) resonances are observed in the high-field region, 0.5–2.5 ppm, of the spectrum of **1** (Figure 2b). Since the oligonucleotide has only four thymines, two different species are probably present in

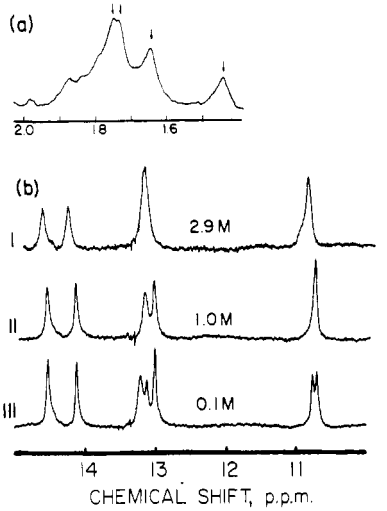

FIGURE 4: (a) 300-MHz ^{1}H NMR spectrum of the methyl protons of **1** containing 1.0 M NaCl in D_2O, at 0 °C. Only the duplex species (d) (indicated by arrows; also see Figure 2) exists at high-salt solution. (b) 498-MHz ^{1}H NMR spectra of the imido protons of **1** with variation of NaCl concentration from 0.1 (row I) to 1.0 (row II) to 2.9 M (row III), at 0 °C. The equilibrium shifts from h to d as the salt concentration increases.

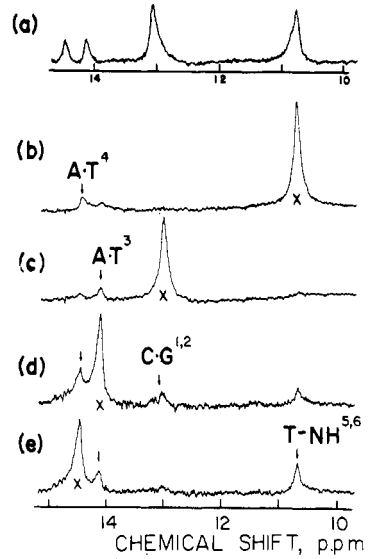

FIGURE 5: (a) 498-MHz ^{1}H NMR spectrum, 10–15.5 ppm, of the imido protons of the duplex form of **1** in 3.8 M NaCl solution at 0 °C. Difference spectra following 0.6-s irradiation of (b) unpaired thymine imido protons (T-NH), (c) guanosine imido protons in G·C(1) and G·C(2), (d) A·T(3), and (e) A·T(4). Irradiation power level resulted in ca. 50% saturation of the irradiated resonance, designated by ×. A relaxation delay time of 3 s and no delay time between the presaturation and observed pulses were used. The peak(s) showing the primary NOE is (are) designated by arrow(s).

the solution. Each species is represented by four T-CH_3 peaks.

(*b*) *Hairpin Formation by Dilution.* The low-field (imido protons) and the high-field (thymine methyl protons) parts of the spectra with variation of oligonucleotide concentration are shown in parts a and b of Figure 3, respectively. When the solution is diluted, the intensity of peaks due to one of the two species gradually decreases. At 0.15 mM single-strand oligonucleotide concentration, only four thymine methyl peaks are seen (Figure 3b). Therefore, only one species is present at this concentration under the condition of study. This species is assigned to be hairpin because only three H-bonded imido proton peaks due to three base pairs are seen at 0.15 mM oligonucleotide concentration (Figure 3a). This assignment is further consolidated by selective spin-lattice relaxation measurement described below. Thus, **1** forms a hairpin with a stem containing three base pairs (CCA·TGG) and a loop containing four bases (-ATTT-) that includes a dissociated A·T base pair.

(*c*) *Duplex Formation at High NaCl Concentration.* When the NaCl concentration is increased, the other species becomes the dominant one. At 1.0 M NaCl concentration only one species, a duplex containing an internal loop, exists almost exclusively, as evident from the four T-CH_3 resonances (Figure 4a). The shift of equilibrium from the mixture to the duplex when the NaCl concentration is increased can be vividly shown in the imido proton resonance region (Figure 4b). Thus, a duplex with eight base pairs (only four resonances in the low-field region due to symmetry) and four unpaired thymines is formed by **1** at high salt concentration. The imido proton spectrum is explicable on the basis of the duplex structure shown in Figure 1. Two imido proton resonances due to two A·T base pairs, as opposed to one A·T base pair expected for the hairpin structure, are seen. The imido proton resonances were assigned to the specific base pairs by 1D-NOE studies done at high NaCl concentration to ensure that only duplex was present in the solution. The results are shown in Figure 5. The peak specifications are indicated in the figure. The numbers correspond to the base pair numbers shown in Figure 1. From chemical shifts we can assign the peaks below 13.5 ppm to the A·T base pairs, the peak upfield from 13.5 ppm to the G·C base pairs (Kearns & Shulman, 1974; Hilbers et al., 1979), and the peak between 10 and 11 ppm to the unpaired thymines (Haasnoot et al., 1983). The resonance at 10.9 ppm due to four protons can be easily assigned to the free NH of the thymines in the loop. An NOE is observed at 14.5 ppm when these free NH protons are irradiated (Figure 5b). Therefore, the peak at 14.5 ppm can be assigned to the A·T base pair next to the thymine in the loop [i.e., A·T(4) in Figure 1]. Consequently, by irradiation of the imido proton of A·T(4), the peak at 14.1 ppm can be assigned to the hydrogen-bonded resonance of A·T(3) (Figure 5e) and the peak at 13.0 ppm (one peak due to two different sets of imido protons) to the C·G(1) and C·G(2) by irradiating the A·T(3) peak (parts d and c of Figure 5). The species identification is confirmed by the selective longitudinal relaxation studies described below.

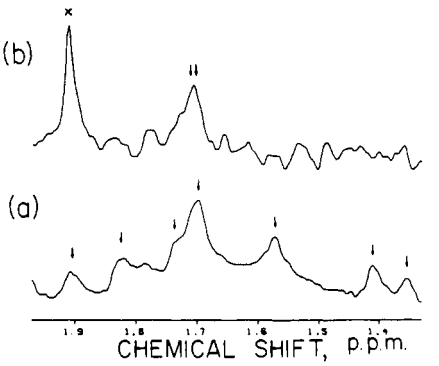

FIGURE 6: (a) 300-MHz ^{1}H NMR spectra of the methyl resonances at 15 °C illustrating transfer of thymine methyl proton magnetization between the duplex and the hairpin forms of **3** in D_2O. (b) × indicates the peak irradiated, and the double arrow indicates the peak to which magnetization was transferred. The peak was irradiated for 2 s before acquisition of the FID, the memory was negated, and then the off-resonance FID was collected in the same block. Please note that the spectrum of (a) is very similar to that in Figure 2b except the signals (indicated by arrows) are broader at 300 MHz.

(*d*) *Equilibrium between Duplex and Hairpin Structures of* **1** *in Low-Salt Neutral Solution by Magnetization-Transfer Experiment.* At 15 °C magnetization is transferred from the

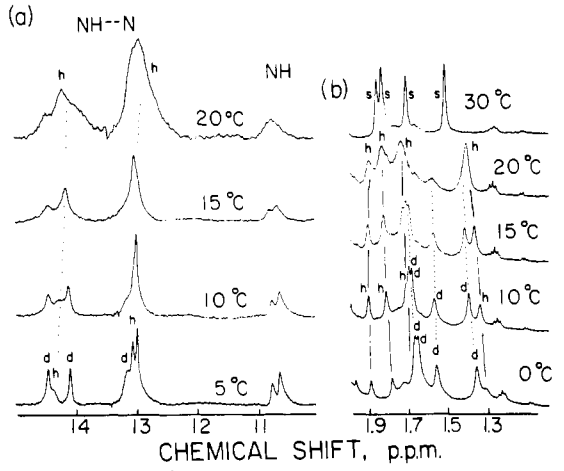

FIGURE 7: 498-MHz ^{1}H NMR spectra, 10–15 ppm, of the imido protons of **1** at varying temperatures, 0–20 °C (a), and of the methyl protons, 1–2 ppm, at varying temperatures, 0–30 °C (b). Symbol s represents the single-stranded state.

methyl groups of one species to the corresponding methyl groups of the other species. An illustrative example of the process is shown in Figure 6. Figure 6a shows the same region in Figure 2b by a 300-MHz NMR spectrometer. It is interesting to note that the methyl proton resonances are broader in Figure 6a than those in Figure 2b. This may be due to the faster exchange rate between the two structures at 15 °C compared to that at 10 °C. Saturation of the peak at 1.91 ppm, due to one of the four T-CH_3 groups of the hairpin structure, causes magnetization to be transferred to the peak at 1.70 ppm, due to the corresponding T-CH_3 group of the duplex structure. Therefore, the two species coexist and are in equilibrium with each other.

(*e*) *Transition of Single Stranded to Hairpin Then to Internal Loop Duplex of* **1**: *A Temperature Variation Study.* Spectra obtained by varying the temperature are shown in parts a (imido protons) and b (methyl group protons) of Figure 7. It is clear that the duplex is the dominant form at very low temperature (0 °C) (Figure 7b). Concentration of the hairpin increases when the temperature is raised, and it becomes the dominant species at higher temperature (e.g., 20 °C) (Figure 7b). The duplex that dominates at high NaCl concentration is the dominant species at low temperature. When the temperature is raised even further, a set of four new peaks emerges clearly at 30 °C. These four peaks represent the thymines of the single-stranded coil form of **1**. The eight peaks reappeared when the sample was cooled to 10 °C. Thus, the observed change is reversible. Similar results can be observed on the imido resonances in Figure 7a. It is clear that only hairpin structure exists in the solution at 20 °C. The transition of the structure of **1** is from single stranded to hairpin and then to duplex with lowering of the temperature as shown in Figure 1.

Because the hairpin structure of **1** contains a dissociated A·T base pair, we decided to synthesize an oligonucleotide such as CCAA(TTTT)TTGG, whose loop contained no dissociated base pair, by putting four thymines in the nonpalindromic part of the sequence to find out how many secondary structures such a sequence would assume.

(*2*) *Hairpin Formation by CCAA(TTTT)TTGG* (**2**). The imido proton and thymine methyl proton regions of the spectrum of this oligonucleotide were studied varying the NaCl concentration (up to 0.7 M). The spectra were the same under all these conditions. Sample spectra are shown in Figure 8, parts a (imido protons) and b (methyl protons). On the basis

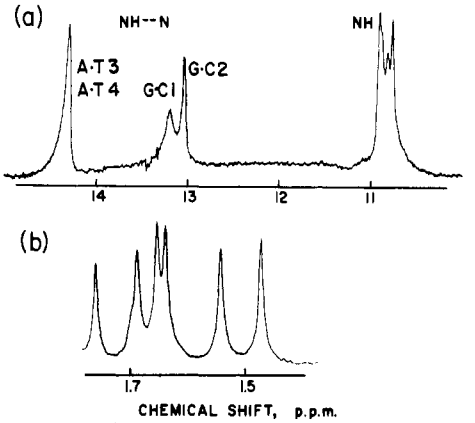

FIGURE 8: 498-MHz ^{1}H NMR spectra of 0.85 mM **2** containing 0.1 M NaCl at 0 °C. (a) Imido proton resonances in the 10–15 ppm region. (b) Methyl proton resonances in the 3–2 ppm region.

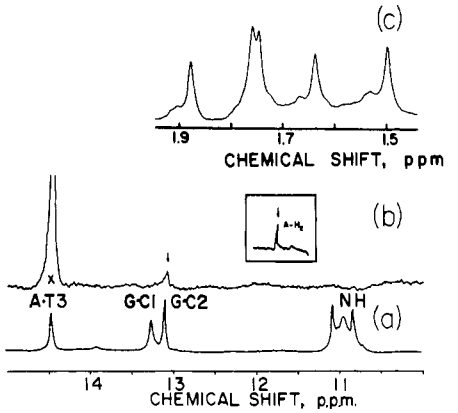

FIGURE 9: NOE spectrum (indicated by arrows) of the imido proton resonance region of **3**. Difference following 1.0-s irradiation of peak at 14.5 ppm (indicated by ×). (a) Unirradiated spectrum. (b) Irradiation power level resulted in about complete saturation of the irradiated peak. A relaxation delay and no delay between the presaturation and observed pulses were used. The NOE observed in the base region (an A-H_2 signal, 8.25 ppm) is shown in the box. (c) Methyl proton resonances at the upfield region.

of these spectra and the spin–lattice relaxation measurements described below, we identify the species as hairpin. The imido proton resonances were assigned by 1D-NOE studies (data not shown).

Thus, it was found that **2** formed only hairpin with no dissociated base pair in the loop. We then wanted to find out if shortening of the stem length of the hairpin would destabilize the formation of the hairpin and thereby favor the formation of the duplex structure. Therefore, we studied the nucleotide CCA(TTTT)TGG.

(*3*) *CCA(TTTT)TGG* (**3**) *Forms Hairpin.* The imido proton and thymine methyl proton spectra were essentially unchanged under different concentrations of NaCl (up to 0.7 M) and oligonucleotide (1.35–2.90 mM). Typical spectra are shown in Figure 9 parts a (imido protons) and c (methyl protons). There are three hydrogen-bonded imido proton resonances in the 12–15 ppm region of **1** (Figure 2a). The peak around 14.5 ppm can be assigned to an A·T base pair by its chemical shift (Kearns et al., 1971; Hilbers et al., 1979) and also by NOE experiment. Namely, the resonance of A-H_2 shows an NOE when the peak at 14.5 ppm is irradiated (Kan et al., 1982) (Figure 9b). The peaks at 13.1 and 13.3 ppm can be assigned to the hydrogen-bonded imido protons of the C-G's. The former can be recognized as the base pair next to the A·T base pair, as evident from the NOE experiment (Figure 9b). There are three signals (integrated as four

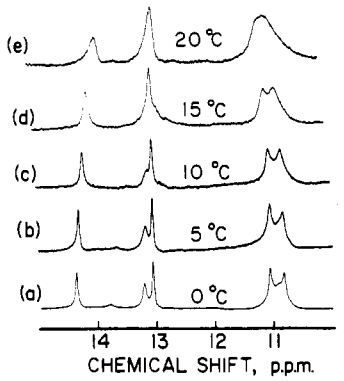

FIGURE 10: Temperature dependence of imido proton resonances of 3 at 498 MHz, 0–20 °C.

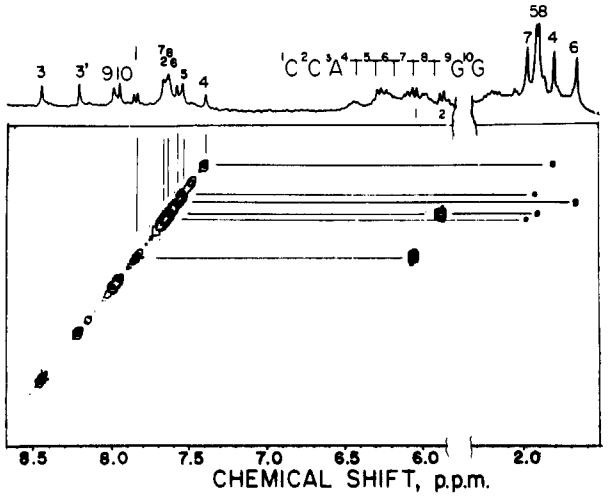

FIGURE 11: 300-MHz COSY proton NMR spectrum of CCA-(TTTT)TGG at 22 °C. The assignments of base protons H_2, H_5, H_6, and H_8 and CH_3 are designated by the sequence numbers, where 3′ is the A(3)-H_2. The assignments of 5 and 8 are tentative. The data were collected over 1024 points in t_2 dimension for 300 t_1 values with 64 scans for each t_1 using a recycle delay of 2.5 s between scans. The data in both dimensions were treated with a sine bell window with no phase shift.

protons) around 11 ppm (Figure 9a) that can be assigned to the free thymidine NH protons (Haasnoot et al., 1983). Thus, 3 forms only one secondary structure, probably a hairpin, in aqueous solution with a three base pair stem (CCA·TGG) and four T's in the loop. This is further supported by the observation of only one set of (five) thymidine methyl proton resonances (Figure 9c). The small peak at 13.9 ppm could be due to another species, but the intensity of this peak did not increase when the NaCl concentration was increased from 0.1 to 0.7 M. Temperature variation study of the imido proton resonances of 3 reveals a transition of simple hairpin to single-stranded form (Figure 10). The hydrogen-bonded imido proton resonances become broader and lose intensity as the hairpin dissociates. The end base pair, G·C(1), dissociates first. Then the other two dissociate simultaneously. This is in contrast to the observation made in CGCGCG(TTTT)-CGCGCG (Ikuta et al., 1986), where the base pairs dissociated sequentially from the stem end toward the loop, but similar to that in ATCCTA(TTTT)TAGGAT (Haasnoot et al., 1983), where the base pairs from the loop and stem ends dissociated simultaneously with increasing temperature. The free thymidine NH resonances also become broader as the exchange rate with water becomes faster at high temperature.

These results therefore clearly indicate that the formation of a hairpin with only a three base pair long stem (CCA·TGG) is more favorable than the formation of the corresponding

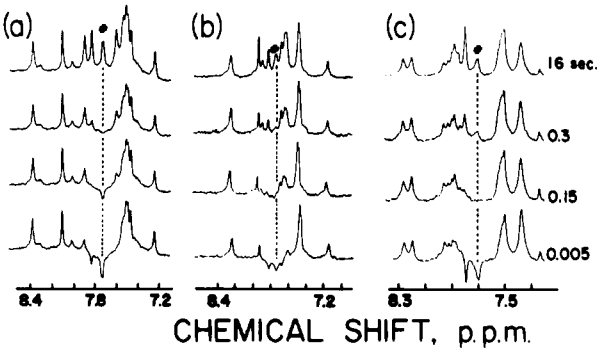

FIGURE 12: Selected proton spin–lattice relaxation rates at 498 MHz and 0 °C: (a) 3 (0.1 M NaCl); (b) 2 (0.7 M NaCl); and (c) 1 (0.2 M NaCl). The selective inversion recovery experiment was carried out by using a DANTE pulse sequence to selectively invert the C(1)-H_6 peak in these three oligomers. The transmitter frequency was placed 1 kHz away so that the C(1)-H_6 peak was at the first side band of the DANTE excitation pulse sequence. Six hard pulses (3.7 μs) spaced with 996.3 μs were applied with a recycle delay of 16 s and a 180° pulse length of 22.2 μs. A 90° nonselective pulse was applied after the variable recovery delay. The recovery delays are indicated on the figure. The # indicate C(1)-H_6 peaks.

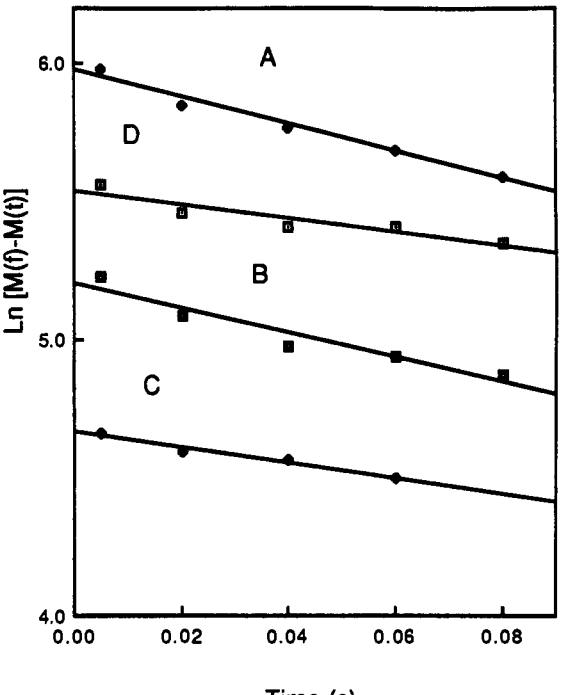

FIGURE 13: Semilogarithmic plots showing the ^{1}C-H_6 magnetization recovery (see Figure 12). $M(t)$ and $M(f)$ are magnetization intensities after recovery times t and 16 s, respectively. The relaxation rate (R_{ss}) equals the negative of the slope. (A) 1 in 1.0 M NaCl; (B) 1 in 0.2 M NaCl; (C) 2 in 0.7 M NaCl; (D) 3 in 0.2 M NaCl. All samples were dissolved in D_2O.

duplex structure. Before studying the base stacking in such an interesting hairpin, it was necessary to ascertain the species identification by NMR relaxation studies as described below.

(4) Duplex versus Hairpin: A Relaxation Study of C-(1)-H_6. The semiselective spin–lattice relaxation rate of the H_6 proton in cytosine 1 (C-H_6) was measured for CCAA-(TT)TTGG, CCAA(TTTT)TTGG, and CCA(TTTT)TGG species by inversion recovery technique using a DANTE 180° pulse (Morris & Freeman, 1978) to selectively invert C-H_6 peaks in the spectra. C-H_6 protons in the spectra were identified by COSY experiment (Figure 11), where cross peaks are seen in the low-field region due to scalar coupling between H_5 and H_6 protons of cytosines. Some spectra from the in-

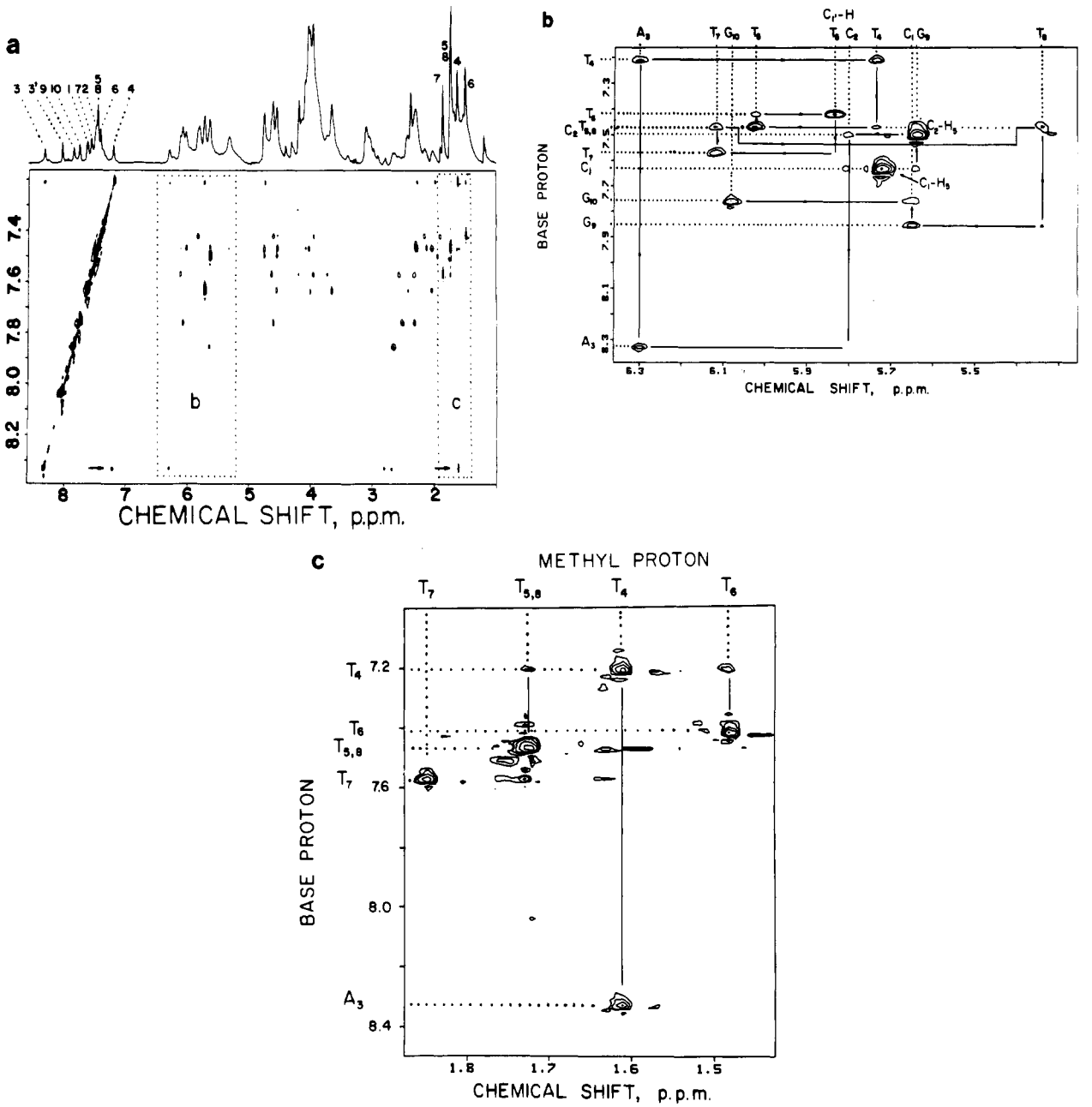

FIGURE 14: 498-MHz NOESY proton NMR spectrum of 2.90 mM 3 containing 0.2 M NaCl, at 2 °C. The data were collected over 1024 points in t_2 dimension for 232 t_1 values with 48 scans for each t_1 using a recycle delay of 3 s between scans and a mixing time of 400 ms. The t_1 dimension was zero filled to 1024 points to create a square data matrix. The data in t_2 dimension were treated with exponential modulation (line width = 3 Hz). A convolution difference with parameters 3, 50, and 0.9 was applied in t_1 dimension. The spectrum was symmetrized. The NOE between the H_8 of A to H_6 and methyl proton of neighboring T are indicated by arrows (a). Closeup view of the content in box b (b) and c (c). All solid straight lines (with arrows) were added for the simplicity of reading. The assignments are connected by the dotted lines on the top and left sides of the figure.

version recovery experiment are shown in Figure 12. Semilogarithmic plots of initial recovery of the magnetization are shown in Figure 13. The initial rate of recovery of semiselective inversion of a proton resonance peak is roughly proportional to the molecular mass (or size) of the oligonucleotide. From simple visual examination of Figure 12 it is clear that the inversion recovery rate of the C-H_6 peak in the oligonucleotide CCAA(TT)TTGG containing 0.2 M NaCl is higher than that of C-H_6 in CCA(TTTT)TGG as well as CCAA(TTTT)TTGG. At 1.0 M NaCl concentration one species existed almost exclusively in the solution of CCAA(TT)TTGG (Figure 4a,b). The relaxation rate, R_{ss} (slope A in Figure 13), for this species, 4.9 (±0.03) s^{-1}, is about twice that for the 10-mer CCA(TTTT)TGG, 2.4 (±0.03) s^{-1}. Therefore, if the species in the solution of CCA(TTTT)TGG is a 10-mer, then the species in the solution of CCAA(TT)TTGG containing 1.0 M NaCl should be equivalent to a 20-mer. From this and the spectra of imido and methyl protons we identify the former species as hairpin and the latter species as duplex. Since when the NaCl concentration in the solution of CCAA(TT)TTGG was increased from 0.2 to 1.0 M the R_{ss} increased from 4.4 (±0.04) to 4.9 (±0.03) s^{-1}, the other species seen in the solution at low NaCl concentration should be a hairpin (which should have a lower R_{ss} value). The R_{ss} value for the 12-mer CCAA(TTTT)TTGG, 2.8 (±0.01) s^{-1}, is between those of the 10-mer hairpin (2.4 s^{-1}) and the duplex (4.9 s^{-1}). Therefore, the oligonucleotide CCAA(TTTT)TTGG assumes a hairpin structure in solution.

(5) *Base Stacking in Duplex and Hairpin.* 1D-NOE spectra of the imido protons of the duplex form of CCAA(TT)TTGG

at 0 °C (Figure 5) indicate that irradiation of the free (not H-bonded) as well as the adjacent H-bonded thymine imido protons shows NOE to each other. Therefore, the thymines in the internal loop of the duplex are well stacked with the stem bases. The base stacking is perhaps not very different from that in the duplex containing no internal loop. On the other hand, 1D-NOE of the imido protons in the hairpin structure of CCA(TTTT)TGG shows an interesting result. Irradiation of the A·T(3) peak (Figure 9b) shows no NOE to the adjacent thymine in the loop [this is in contrast to the observations made for other hairpins such as CGCGCGTTTTCGCGCG (Ikuta et al., 1986) and ATCCTATTTTTAGGAT (Haasnoot et al., 1983) where such NOE was observed]. Therefore, the stacking of the free thymine adjacent to the stem in hairpin is different from that in the duplex.

In order to get some idea about the base stacking in the hairpin structure, a 2D NOESY spectrum of the nonexchangeable protons was recorded. The methyl and base proton assignments were made in the following manner. All thymine and cytosine H_6 protons were identified from 2D COSY (Figure 11). The H_6 resonance of a thymine shows a cross peak with its own CH_3 peak, in the 1–2 ppm region, due to the four bond scalar coupling between these protons. Cytosine H_6 resonances show cross peaks with its own H_5 peaks in the 5.2–6.5 ppm region. By elimination, the other four peaks in the base proton region, 7.0–8.5 ppm, are due to three H_8 protons of two G's and one A and one H_2 proton of the same A. The resonances were then assigned to the specific bases in the oligonucleotide sequence by means of NOESY (Figure 14). Since the maximum possible separation between the H_6 of a pyrimidine or H_8 of a purine with its own $H_{1'}$ is 3.97 Å (Hare & Reid, 1986), it is expected that the cross peaks between these protons should be observed irrespective of the relative orientations of the base and sugar rings in DNA. Assignment of $H_{1'}$ peaks are thus made from NOESY cross peaks in Figure 14a. Almost all cross peaks between base protons and $H_{1'}$ of its preceding sugar in the 5′ end are seen. The peaks at 7.64 and 8.33 ppm were assigned to C(1)-H_6 and A(3)-H_8, respectively, on the basis of their chemical shifts (Kan & Ts'o, 1986). The remaining assignments were made following the NOESY connectivity. The T(4)-CH_3 peak was easily assigned from its cross peak with the A(3)-H_8. The remaining T-CH_3 peaks were assigned from the cross peaks with their own H_6 protons already identified in Figure 14a.

In B-form DNA the cross peak between a base proton with $H_{1'}$ of the preceding sugar in the 5′ end provides an important clue in the sequence determination (Reid, 1986). Such NOESY connectivity breaks down at two places in the case of this hairpin. No cross peak was observed between T(6)-$H_{1'}$ and T(7)-H_6 as well as C(2)-$H_{1'}$ and A(3)-H_8. In addition, the cross peak between T(8)-$H_{1'}$ and G(9)-H_8 is very weak. These indicate that the stem portion of the hairpin is distorted from the ideal B-DNA conformation, and the T(6) and T(7) bases in the loop have twisted away from each other. The NOESY cross peaks between T-CH_3 and base protons (Figure 14b) yield further insight into the orientation of bases in the loop. A(3)-H_8 is found to be close to both the H_6 and CH_3 of T(4) (Figure 14a,c). This indicates that the thymine ring turns toward the five-membered ring of A. The fact that T(6)-CH_3 is close to neither CH_3 of T(5) or T(7) but is close to that of T(4) indicates that T(5) and T(7) have been twisted out of the interior of the loop with T(6) inward to make the turn of the loop complete. Similar observations were also made by Hare and Reid (1986) for the hairpin structure of CGCGTTTTCGCG.

CONCLUSIONS

We have observed two secondary structures for the oligonucleotide CCAA(TT)TTGG in solution at temperatures of 15 °C and below. One species is a hairpin containing a dissociated A·T base pair in the loop, and the other is a duplex with an internal loop. The two species were in dynamic equilibrium with each other. Hairpin is the favored structure at high temperature and low oligonucleotide and NaCl concentrations, whereas the duplex is the favored structure at low temperature and high oligonucleotide and NaCl concentrations. We also studied the secondary structures of the sequentially related oligonucleotides CCA(TTTT)TGG and CCAA(TTTT)TTGG, both of which exhibited only the hairpin structure under various conditions of studies used here. It appears that the number of species formed by the oligonucleotides studied here depended on the number of thymines in the nonpalindromic part, enclosed in parentheses, of the chain. The NOE experiments indicated that the stacking of the H-bonded stem bases with their adjacent bases in the loop in the hairpin was different from that in the duplex.

ACKNOWLEDGMENTS

We acknowledge The JHU Biophysical NMR Facility Center and Dr. Leo Neuringer of the Francis Bitter National Magnet Laboratory, Cambridge, MA, for use of a Bruker WM-300 and a homemade 498-MHz NMR spectrometer, respectively. We are indebted to Dr. David J. Ruben for his technical assistance.

REFERENCES

Atkinson, T., & Smith, M. (1984) *Oligonucleotide Synthesis: A Practical Approach* (Gait, M. J., Ed.) pp 35–81, IRL, Oxford, England.

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–50.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987a) in *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Chapter 8, Clarendon, Oxford, England.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987b) in *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Chapter 9, Clarendon, Oxford, England.

Frank-Kamenetskij, M. D., & Vologodskij, A. V. (1984) *Nature (London) 307*, 480–481.

Gierer, A. (1966) *Nature (London) 212*, 1480–1481.

Haasnoot, C. A. G., & Hilbers, C. W. (1983) *Biopolymers 22*, 1259–1266.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 3*, 115–129.

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*, 843–858.

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

Hilbers, C. W. (1979) in *Biological Applications of Magnetic Resonance* (Shulman, R. G., Ed.) pp 1–44, Academic, New York.

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

Kan, L.-S., & Ts'o, P. O. P. (1986) in *NMR in Biology and Medicine* (Chien, S., & Ho, C., Eds.) pp 83–116, Raven, New York.

Kan, L.-S., Cheng, D. M., Jayaraman, K., Leutzinger, E. E., Miller, P. S., and Ts'o, P. O. P. (1982) *Biochemistry 21*, 6723–6732.

Kearns, D. R., & Shulman, R. G. (1974) *Acc. Chem. Res. 7*, 33–39.
Lilley, D. M. (1981) *Nucleic Acids Res. 9*, 1271–1289.
Morris, G. A., & Freeman, R. (1978) *J. Magn. Reson. 29*, 433–462.
Muller, U. R., & Fitch, W. M. (1982) *Nature (London) 298*, 582–585.
Patel, D. J., Kozlowski, S. A., Ikuta, S., Itakura, K., Bhatt, R., & Hare, D. R. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 197–206.
Redfield, A. G., Kunz, S. D., & Ralph, E. K. (1975) *J. Magn. Reson. 19*, 114–117.
Reid, B. R. (1986) in *NMR in Biology and Medicine* (Chien, S., & Ho, C., Eds.) pp 83–116, Raven, New York.
Sheflin, L. G., & Kowalski, D. (1985) *Nucleic Acids Res. 13*, 6137–6154.
Sinden, R. R., & Pettijohn, D. E. (1984) *J. Mol. Biol. 259*, 6593–6600.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286.
Summers, M. F., Andrew Byrd, R., Gallo, K. A., Samson, C. J., Zon, G., & Egan, W. (1985) *Nucleic Acids Res. 13*, 6375–6386.
Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature (London), New Biol. 246*, 40–41.
Weaver, D. T., & DePamphilis, M. L. (1984) *J. Mol. Biol. 180*, 961–986.
Wemmer, D. E., Chou, S. H., Hare, D. R., & Reid, B. R. (1985) *Nucleic Acids Res. 13*, 3755–3772.
Yuhasz, S. C., Kanhouwa, N., Shida, T., & Kan, L.-S. (1987) *Bio/Technology 5*, 54–59.

Sequence-Targeted Photosensitized Reactions in Nucleic Acids by Oligo-α-deoxynucleotides and Oligo-β-deoxynucleotides Covalently Linked to Proflavin†

Danièle Praseuth,‡ Trung Le Doan,‡ Marcel Chassignol,§ Jean-Luc Decout,‖ Noureddine Habhoub,‖ Jean Lhomme,‖ Nguyen T. Thuong,§ and Claude Hélène*,‡

Laboratoire de Biophysique, INSERM U.201, CNRS UA.481, Muséum National d'Histoire Naturelle, 61 Rue Buffon, F-75005 Paris, France, Centre de Biophysique Moléculaire, CNRS, F-45071 Orléans Cedex 02, France, and Laboratoire de Chimie Organique et Macromoléculaire, CNRS UA.351, Université des Sciences et Techniques de Lille I, UER de Chimie, F-59665 Villeneuve d'Ascq Cedex, France

Received August 21, 1987; Revised Manuscript Received December 8, 1987

ABSTRACT: Proflavin was covalently linked to the 3′-end or to the 5′-end of an octadeoxythymidylate. This oligonucleotide was synthesized with either the natural β-anomer of thymidine or its synthetic α-anomer. A polymethylene chain was used to link one of the amino groups of proflavin to a terminal thiophosphate group of the oligonucleotide. A 27-mer oligodeoxynucleotide containing an octadeoxyadenylate sequence was used as a target for the proflavin-substituted octadeoxythymidylates. Upon irradiation with visible light, photo-cross-linking reactions induced the formation of branched species that migrated more slowly than the 27-mer on denaturing polyacrylamide gels. Piperidine treatment of the photo-cross-linked species induced strand breaks in the 27-mer. In addition, proflavin induced photosensitized reactions at guanine residues in the 27-mer sequence which were converted to strand breaks following piperidine treatment. Triple-helix formation by the oligothymidylates with their complementary oligodeoxyadenylate sequence at high salt concentration led to photo-cross-linking and cleavage reactions on both sides of the target sequence. These results show that it is possible to target photosensitized reactions to specific sequences on nucleic acids. This opens new possibilities for site-directed mutagenesis and the development of photoactive anti-messenger oligodeoxynucleotides.

The recent development of anti-messenger oligonucleotides or anti-sense RNAs has opened new possibilities for the specific regulation of gene expression (Green et al., 1986). Highly specific complexes can be formed by short oligonucleotides with their complementary sequence. A 12-mer should find only one target sequence in the *Escherichia coli* genome, assuming a random distribution of base pairs and an equal number of A·T and G·C base pairs. The human genome contains more A·T than G·C base pairs. To specify a site that reappears at random with a frequency less than unity, the minimal oligonucleotide length varies between 15 and 19, depending on whether the oligonucleotide contains only G and C or A and T, respectively (Hélène & Thuong, 1987). Only a small fraction of the genome is transcribed as messenger RNA in a living cell. If the oligonucleotide is targeted to a mRNA, its length can be further decreased without loss of specificity. Assuming that about 0.5% of the genome is expressed as messages in a given cell type, the minimal oligonucleotide

† The Association pour la Recherche sur le Cancer is gratefully acknowledged for providing a research fellowship to one of us (D.P.). This work was supported in part by the Ligue Nationale Française contre le Cancer, the Fondation pour la Recherche Médicale, and Rhône-Poulenc-Santé.

‡ Muséum National d'Histoire Naturelle.
§ Centre de Biophysique Moléculaire, CNRS.
‖ Université des Sciences et Techniques de Lille I, UER de Chimie. Present address: Laboratoire d'Etudes Dynamiques et Structurales de la Sélectivité, CNRS UA.332, Université de Grenoble I, B.P. 68, F-38402 Saint-Martin d'Hères, France.

0006-2960/88/0427-3031$01.50/0