Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5224–5228, June 1993
Biochemistry

5′-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis

(unusual DNA conformation/non-Watson–Crick base pairing/repetitive sequence/structural refinement)

HOWARD ROBINSON AND ANDREW H.-J. WANG†

Biophysics Division and Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Communicated by Hans Frauenfelder, March 8, 1993

ABSTRACT **The structure of the non-self-complementary DNA heptamer d(CGACGAC) at low pH has been determined by the quantitative NMR refinement procedure designated SPEDREF (SPEctral-Driven REFinement). Acid–base titration of the molecule indicated a prominent $n = 2$ pK_a near 6.8. In the pH range up to 6.0, the heptamer forms a remarkably stable double helix, which was conclusively shown to be an unusual homobase-paired parallel-stranded double helix (termed II-DNA). In this II-DNA helix, the 5′-CGA trinucleotide is the structural motif that accounts for the stability, with the C^+·C hemiprotonated base pair (in which C^+ is N^3-protonated cytosine) providing for the alignment site and the unusual interstrand G-A base stack in the GpA step furnishing the additional stabilizing forces. The exchangeable proton data from two-dimensional nuclear Overhauser effect spectroscopy are in total agreement with the refined structure. We conclude that the 5′-CGA or other related sequences (e.g., 5′-CCGA) are powerful motifs in promoting the II-DNA or II-RNA conformations that may play certain biological functions.**

The central role of DNA structure in biology is well established. To correlate the structure and function of DNA more fully, a detailed knowledge of its potential in adopting alternative stable conformations is necessary. Stretches of unusual arrangements of nucleotide sequences, such as $(CG)_n$ or $(TG)_n$, which form Z-DNA readily (1, 2), are often found in genomic DNA (3). Other repetitive sequences like $(TTGGGG)_n$ found in certain eukaryotic telomere DNA (4, 5), $(GGAAT)_n$ in centromere DNA (6), or $(CGG)_n$ in X-chromosome (responsible for the fragile-X syndrome) (7, 8) have been implicated in various biological functions. These unusual repetitive sequences may have specific conformations. This polymorphic nature of DNA conformations depends not only on the nucleotide sequence but also on extrinsic factors such as proteins, counter ions, humidity, or pH. Not surprisingly, the equilibrium between the B-DNA and other DNA conformers is influenced by these factors.

Another important structure, the triple-stranded H-DNA, may be induced in $(dG)_n$·$(dC)_n$ sequences especially at low pH (<5) (9–11). The H-DNA model involves a protonated oligo(dC^+) strand, where C^+ indicates N^3-protonated cytosine, folding back to the major groove side of the double helix and forming the triple (dC)·(dG)·(dC^+) base pairs with Hoogsteen pairing geometry between the dG and the protonated dC^+ (12). The formation of triplex structure is facilitated by the moderately high pK_a of cytosine N^3 in DNA (13). The structural information of a number of DNA triplexes has been provided by recent NMR studies (14–16).

We have been exploring the possible structural consequences on nucleic acids perturbed by the protonation of bases, particularly of cytosine whose pK_a of 4.2 is the highest among the bases (13). However the pK_a of cytosine in DNA seems variable depending on structural influences. An NMR study of calf thymus DNA suggested a value of 3.7 (17), yet it has been shown that poly(dC) has a pK_a of 7.4 (18). This may imply that certain structures of DNA/RNA could push the pK_a of cytosine in nucleic acids up into the "physiological" range. When the cytosine is protonated at N^3, it can no longer participate in the normal Watson–Crick base pair. Other possible base pairing schemes involving C^+ may take place, including the hemiprotonated C·C^+ base pair with three hydrogen bonds. Several double-stranded helical structures involving the C^+·C base pairs have been studied by optical spectroscopic methods (19, 20) and NMR (21, 22).

Recently we observed that a number of DNA oligonucleotides having sequences related to 5′-CGA adopt a structure different from B-DNA at pH values below 5.5 (23). On the basis of the results from the NMR refinement of the d(CGATCG) structure using the observed two-dimensional nuclear Overhauser effect (2D-NOE) intensities and the acid-titration data, we proposed a parallel-stranded double helix in which all base pairs are of the non-Watson–Crick self-pairing type—i.e., A with A, T with T, G with G, and finally, C with C^+ (23). This unusual structure is different from any of the parallel-stranded structures previously reported. A prediction from that work is that a DNA molecule with the sequence containing only 5′-$(CGA)_n$ without its Watson–Crick complementary strand is capable of forming a stable parallel-stranded structure at nearly neutral pH. We propose the name II-DNA for this new, homobase-paired parallel-stranded duplex. This is based on the fact that the upper case Greek **Pi**, II, stands for **P**arallel-stranded helix, and the letter has twofold symmetry vertically with polarity.

MATERIALS AND METHODS

The DNA heptamer d(CGACGAC) was synthesized on an Applied Biosystems DNA synthesizer and was purified by reverse-phase HPLC. The 2.6 mM duplex solution of d(CGACGAC) for NMR studies was prepared as described (24). The solution was unbuffered and contained 0.15 M NaCl. The indicated pH values were recorded after the acquisitions of the spectra with an electrode that was corrected for measurement in ^2H_2O. The NMR spectra were recorded on a GE GN500 500-MHz spectrometer. The chemical shifts (in ppm) are referenced to the H^2HO peak, which was calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).

The phase-sensitive spectra from NOE spectroscopy (NOESY) were recorded at 5°C as 2×512 t_1 blocks of 2048 complex points each and were averaged for 16 scans per block. The recycle delay was 2.43 s, and the mixing time was 200 ms. The 2D data sets were processed with the program

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 1D and 2D, one and two dimensional; C^+, N^3-protonated cytosine; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; SPEDREF, SPEctral-Driven REFinement.

†To whom reprint requests should be addressed.

FELIX v1.1 (Hare Research, Woodinville, WA). The spectra were processed as described (24) with 4-Hz exponential multiplication and truncation apodization. The structure refinement of the complex has been carried out by an improved version of the procedure SPEDREF (24). The mixing time of 200 ms is sufficiently long for some spin diffusion to occur. By inclusion of relaxation matrix analysis in the refinement procedure, the conformational dependence of this spin diffusion contributes additional information to the refinement. The correlation time τ_c of 5 ns was selected by consensus between analysis of errors for several structural features.

The experimental proton 2D-NOESY of d(CGACGAC) in 90% H_2O/10% ^2H_2O, pH 4.2, at 5°C was collected by using H_2O presaturation and a mixing time of 200 ms. The exchangeable proton one-dimensional (1D) spectrum was carried out by a $1\bar{3}3\bar{1}$ NOE difference experiment (overbars signify inverse phasing) where the signal at 14.9 ppm was alternately saturated to identify the protons that are close to the proton associated with this signal. There were 1024 cycles of saturated and unsaturated scans that were subtracted to produce the spectrum.

Circular dichroism spectra were recorded on a Jobin–Yvon CD-6 spectrometer. The concentrations of the samples were 10 μM DNA duplex in 150 mM NaCl with varying pH at 22°C and a 1-cm path-length cuvette.

RESULTS AND DISCUSSION

A pH-titration study indicates a well-defined $n = 2$ pK_a of 6.8 at a DNA concentration of 2.6 mM duplex (Fig. 1 *Inset*). The 1D NMR pH study at the same concentration of duplex also confirms this unusually high pK_a value (Fig. 1). At pH values greater than 7.5, the spectrum consists of resonances from at least two species of DNA structures. However, a dramatic change occurred when the pH was decreased to ≈6.0. The NMR spectrum becomes significantly simplified (consistent with a single DNA structure). Decreasing the pH to 3.8 maintains the same basic spectrum, but the change in chemical shifts for C7 suggests that C1 and C4 are hemiprotonated at the $n = 2$ pK_a of 6.8 and that C7 becomes hemiprotonated with a pK_a around 5.5. The temperature-dependent study of d(CGACGAC) at pH 3.8 suggested a remarkable stability of the structure as the 1D NMR spectrum changes very little up to 40°C (data not shown).

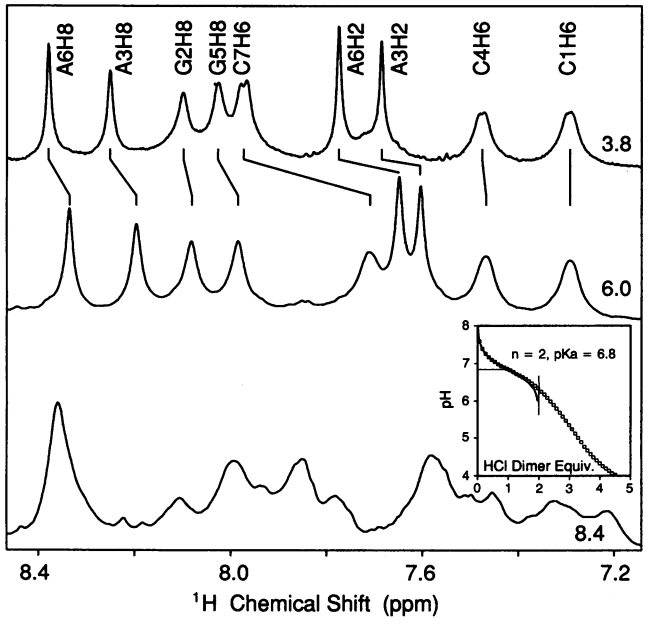

FIG. 1. 1D NMR spectra of nonexchangeable DNA aromatic protons of d(CGACGAC) in ^2H_2O at the indicated pH values. The spectrum of the low-pH form is that of the Π-DNA helix as explained in the text. The differences between the pH 3.8 and pH 6.0 spectra for the adenine resonances are consistent with protonation at the N^1 position (pK_a 3.5) of the four adenine residues per duplex. Since these shifted resonances do not split (indicating rapid exchange), the N^1 sites are likely not involved in base-pair hydrogen bonds. The spectrum at pH 8.5 is likely due to two possible B-DNA helices, one with a six-base-paired duplex (with two A·A mismatches and a 5′ dangling C) and the other with a five base-paired duplex (with one A·A mismatch and a 3′ A-C dangling dinucleotide). (*Inset*) pH titration of d(CGACGAC).

Optical studies at low DNA concentration (10 μM duplex) also show a similar dramatic change in structure (Fig. 2) between pH 6.0 and 6.9. Intense bands in the CD spectrum are evident between 240 and 270 nm when the pH is below 6. These bands are indicative of a tightly stacked structure even at these low concentrations. Some of the absorption bands also shift toward longer wavelength below pH 6.

To determine the three-dimensional structure unambiguously, we collected 2D-NOESY (Fig. 3 *Upper*) and purge-correlated spectroscopy (PE-COSY) data for structural refinement. All resonances (except for the ambiguity of some 5′ and 5″ protons) were assigned in a straightforward manner starting with a sequential assignment of the cross peaks between H1′ to aromatic protons. Most of the NOE cross peaks are well separated, ensuring a reliable measurement of the NOE intensities by our SPEDREF package (24) aided by the lineshapes determined from our program MYLOR. Some of the key NOEs are very evident. In particular, the revealing NOE cross peaks resulting from the proximity of the interstrand $G2H^8$–$A3*H^2$, $G5H^8$–$A6*H^2$ protons and the intrastrand $C4H^{1'}$–$A3H^2$, $C7H^{1'}$–$A6H^2$ protons in a Π-DNA duplex are clearly visible (23). These data gave us confidence that the structure of d(CGACGAC) at pH 3.8 is a Π(CGA) structure with homobase pairs. A starting Π(CGA) model for the 7-mer was constructed by using the principles derived from our previous work on d(CGATCG) (23). The model was subjected to the SPEDREF refinement using ≈620 NOE

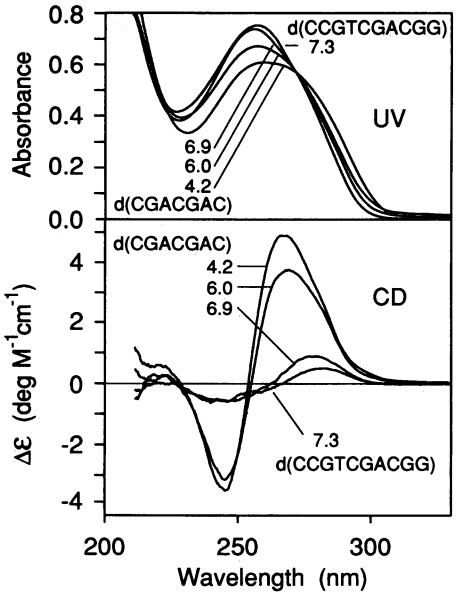

FIG. 2. Influence of pH on the circular dichroism spectra of d(CGACGAC). The spectrum of self-complementary d(CCACGCGTGG) at pH 7.3 was used as a B-DNA control. The pH of the sample is indicated on the figure. The ellipicity is calculated per nucleotide. Note the large change in the ellipicity in going from pH 6.9 to 6.0. The spectrum at pH 6.0 has nearly a 10-fold increase in ellipicity compared with that of B-DNA. This suggests a structure very different from B-DNA.

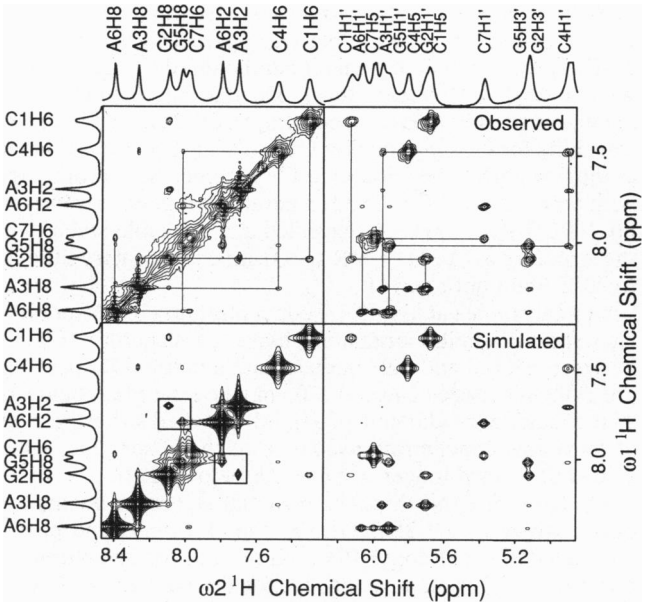

FIG. 3. Portions of the nonexchangeable proton phase-sensitive 2D-NOESY spectra of the d(CGACGAC) duplex at pH 4.2, which provide key structural information including the glycosyl conformation, sugar puckers, and base-base stack. (*Upper*) Experimental NOESY cross peaks between the aromatic (6.2–8.6 ppm)-to-aromatic protons and the aromatic protons to H1′/H5 (5.5–6.2 ppm). (*Lower*) Simulated NOE spectra of the same regions based on the refined Π-DNA model (*R* factor = 21.3%). The strong experimental NOE cross peaks between the interstrand G2H^8–A3*H^2, G5H^8–A6*H^2 protons (in boxed regions) and the intrastrand C4H$^{1'}$–A3H^2, C7H$^{1'}$–A6H^2 protons are evident.

integrals. The refined model has a current NMR *R* factor of 21.3%. The refined model produces the simulated NOE spectra (Fig. 3 *Lower*) very similar to the experimental NOE spectra.

The refined d(CGACGAC) Π-DNA helix is depicted in Fig. 4. As can be seen, this helix possesses an exact twofold symmetry coinciding with the helix axis. The two grooves are identical. Note that because of the unusual base-pairing scheme and the opposite-strand alignment, the disposition of various atoms on the surface of the grooves is completely different from that in B-DNA.

This new PS structure is favored by the 5′-CGA sequence motif because at low pH the cytosines provide the highly stable C$^+$·C pairings and the GpA step is significantly stabilized by the interstrand G-A stacking interactions. Li *et al.* (26, 27) have shown that if two consecutive G·A mismatched base pairs in B-DNA are arranged in a specific way—i.e., embedded in a tetranucleotide of the type of 5′-YGAR, where Y = pyrimidine nucleotide and R = purine nucleotide—the helix is remarkably stable despite the GA mismatches. This was attributed to the unusual interstrand A-over-A and G-over-G stacking interactions, not unlike those in the Π-DNA helix.

Since the Π(CGA) helix contains unusual base pairs with interdigitated strands, the exchangeable proton NMR data should provide important confirmation for the structure. Fig. 5 shows portions of the 2D-NOESY spectrum of d(CGACGAC) in 90% H_2O/10% ^2H_2O at 5°C. Most of the exchangeable (amino and imino) proton resonances were detected and have been unambiguously assigned. All of the observed NOE cross peaks are consistent with the refined structure. These new data, which are independent of the refinement process, lend further proof to the validity of the structure. Interestingly at pH 4.2, we detect only a single resonance at 14.98 ppm, which is unambiguously assigned to the imino proton of the first C·C$^+$ base pair (see Fig. 5). This is in contrast to the readily observable H*N*^{3} imino protons from the oligo(dC) molecules (21, 22) or in the triplex structures (14–16) at low pH. It is not clear why the imino protons of the other two C·C$^+$ base pairs of this stable Π-DNA helix are exchanging too rapidly to be detected.

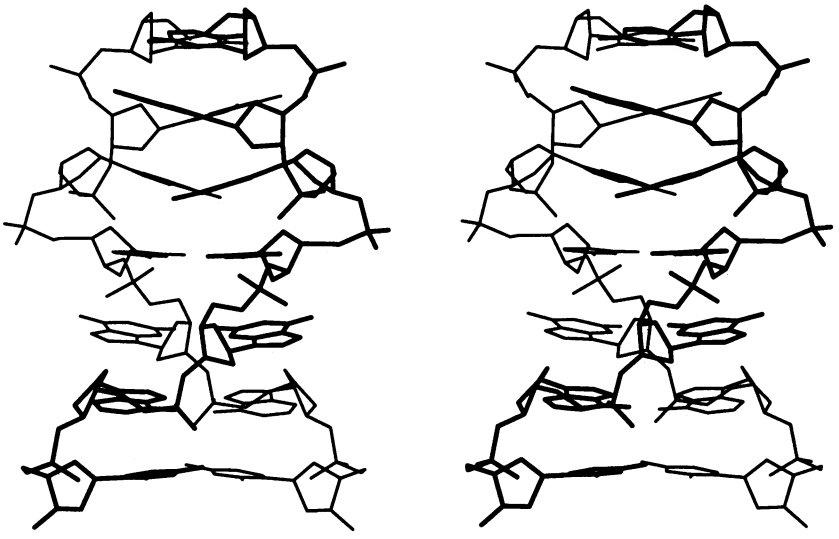

FIG. 4. Molecular model of the d(CGACGAC) structure. The groove is very shallow at the G·G base pair step. There are relatively large propeller twists in the purine·purine base pairs and an interstrand hydrogen bond from the adenosine's amino group to the other strand's phosphate oxygen. Many of the interproton relationships in this structure are unique to this structure—e.g., those between the interstrand G2H^8–A3*H^2, G2HN^1–A3*H^8, G5H^8–A6*H^2, G5HN^1–A6*H^8 protons and the intrastrand C4H$^{1'}$–A3H^2, C7H$^{1'}$–A6H^2 protons. It should be noted that this model represents an average structure based on the dynamic averaging of the NOE measurement. We have constructed a polymer Π(CGA) helix with the following respective conformational angles (in degrees) for cytidine, guanosine, and adenosine nucleotides: α(274.7, 279.9, 289.1), β(153.2, 179.5, 152.1), γ(100.9, 50.8, 55.1), δ(139.1, 150.5, 93.4), ε(202.3, 193.8, 230.1), ζ(294.9, 162.1, 307.4), pseudorotation angle P(169.4, 169.0, −1.9), and glycosyl angle χ(221.1, 289.1, 189.3). An interesting feature of the Π-DNA helix is that the continuous stack of the bases actually run in a left-handed helical path with twice the pitch (≈18 base pairs) of the backbone (≈9 base pairs). We suggest that the Π(GA) helix (25) may adopt a similar type of interdigitating base-pairing pattern as in the Π(CGA) helix, as shown by our NMR data of d(CGAGAGA) (unpublished data).

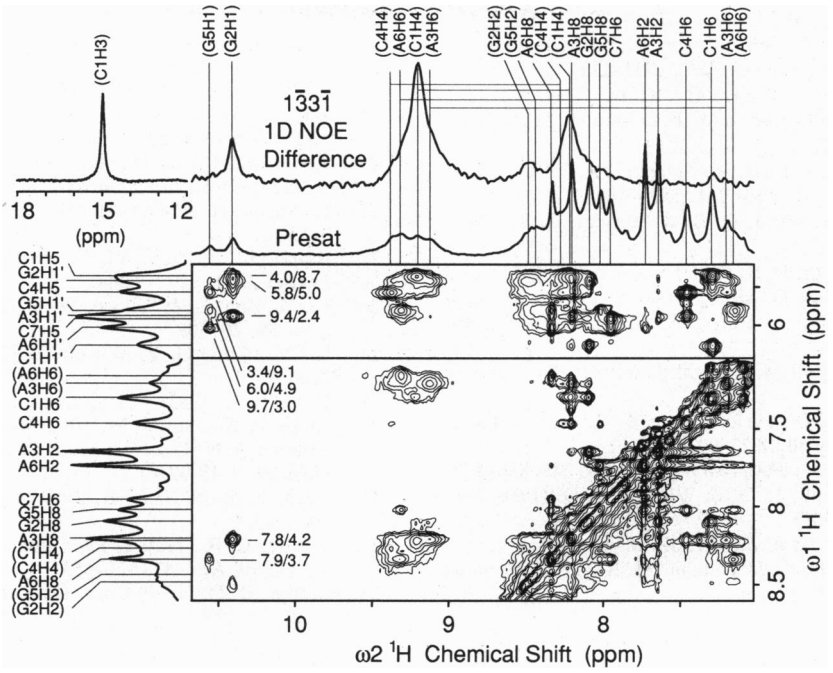

FIG. 5. The experimental proton NOESY of d(CGACGAC) in 90% H_2O/10% ^2H_2O, pH 4.2, at 5°C. Most of the amino and imino proton resonances were detected, except for the imino protons from the C^+·C base pairs. The exchangeable protons are labeled in parentheses on the spectra. The NOE cross-peak intensities between the exchangeable to the exchangeable protons and the exchangeable to the nonexchangeable protons agree with the calculated distances (marked in the spectra in Å) from the refined structure. Presaturation of the H_2O resonance did not allow observation of the resonance at 14.98 ppm; however, a single resonance is easily observed with a 1331 pulse sequence (28), where the excitation maximum is near this frequency (spectrum at the upper left). The adjacent 1D spectrum is a 1331 NOE difference experiment where the signal at 14.98 ppm was alternately saturated. There were 1024 cycles of saturated and unsaturated scans that were subtracted to produce the spectrum. The identity of nearby protons substantiates the assignment of the 14.98 ppm resonance being the imino proton of the first C·C^+ base pair in the helix.

The observation of the highly stable Π(CGA) helix invokes the issue of the generality of parallel-stranded nucleic acid structures. Another type of parallel duplex with reverse Watson–Crick base pairs [may be termed PS(rWC)] exists in molecules with specially designed sequences (29). Two recent studies of oligonucleotides with $(GA)_n$ sequences have appeared (25, 30), and one of them suggested a parallel-stranded duplex with G(syn)·G(syn) and A(anti)·A(anti) base pairs (25). The Π(CGA) helix adopts non-Watson–Crick symmetric homobase pairings with different hydrogen bonding schemes. It is worth noting that the Π-motifs we propose here offer great advantages in providing a highly defined structural motif. Specifically, while the oligo(dC) or oligo[d(GA)] sequence possibly adopts the parallel-stranded structure, two strands of the $Π(C_n)$ or Π(GA) helix presumably can "slide" against each other. In contrast, the 5′-CGA sequence can only pair with the identical 5′-CGA sequence due to the sequence-dependent structural arrangement. This can perhaps be viewed as a special kind of sequence complementarity. Therefore, the Π(CGA) or Π(CCGA) motifs may be used to align two molecules into a parallel-stranded duplex in precise registration. This may be useful in bringing two homologous sequences, distant in locales of a nucleic acid molecule, to pair and possibly initiate a new structural folding. This may have interesting implications in the recombination process. Another likely place to find such structures is in the highly folded RNA molecules, such as ribozyme (31) or ribosomal RNA (32).

These distinctive types of stable Π-DNA helices are completely different from B-DNA helices. Consequently it may be possible to generate antibodies against polymers like poly[d(CGA)] or poly[d(CCGA)] (33). Π-DNA-binding proteins may be searched by using the affinity column of poly[d(C_nGA)]. Additionally, one may use the *in situ* hybridization technique to probe the existence of the $(CGA)_n$ or $(CCGA)_n$ sequences in chromosomes in ways similar to the case of Z-DNA (1) or the centromere DNA (6). Other intriguing questions include (*i*) the reactivity toward nuclease and chemical agents including alkylating agents, (*ii*) the use as antisense DNA because of the high stability of Π-DNA helix, and (*iii*) the behavior of sequence $(C_mGA)_n$ in supercoiled DNA. Answers to many of these questions will help to clarify the possible roles of this unique family of nucleic acid structures.

In conclusion, the surprising finding of a simple sequence like 5′-CGA being able to form the remarkably stable homobase-paired Π-DNA helix substantially increases the repertoire of alternative nucleic acid conformations. Similar structures may be adopted by $(GA)_n$ sequences. It is likely that other unusual structures will be uncovered in the future.

We thank A. J. Wand for use of his Bruker AMX500 spectrometer for the 1331 H_2O NOE difference experiment. The assistance of Ms. Y.-K. Li for the purification of the DNA oligonucleotides is greatly appreciated. Technical help from V. Mainz of the School of Chemical Sciences Molecular Spectroscopy Lab is gratefully acknowledged. This work was supported by National Institutes of Health Grants GM-41612 and CA-52506 to A.H.-J.W.

1. Rich, A., Nordheim, A. & Wang, A. H.-J. (1984) *Annu. Rev. Biochem.* **53**, 791–846.
2. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A. & Rich, A. (1979) *Nature (London)* **282**, 680–686.
3. Palecek, E. (1991) *Crit. Rev. Biochem. Mol. Biol.* **26**, 151–226.
4. Blackburn, E. H. (1992) *Annu. Rev. Biochem.* **61**, 113–129.
5. Sen, D. & Gilbert, W. (1991) *Curr. Opin. Struct. Biol.* **1**, 435–438.
6. Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. & Moyzis, R. K. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1695–1699.

7. Kremer, E., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richands, R. I. (1991) *Science* **252,** 1711–1714.
8. Oberle, I., Rousseau, F., Heitz, C., Kertz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F. & Mandel, J. L. (1991) *Science* **252,** 1097–1102.
9. Htun, H. & Dahlberg, J. E. (1988) *Science* **241,** 1791–1796.
10. Johnston, B. H. (1988) *Science* **241,** 1800–1804.
11. Pulleyblank, D. E., Haniford, D. B. & Morgan, A. R. (1985) *Cell* **42,** 271–280.
12. Moser, H. E. & Dervan, P. B. (1987) *Science* **238,** 645–650.
13. Bloomfield, V., Crothers, D. M. & Tinoco, I. (1974) in *Physical Chemistry of Nucleic Acids* (Harper & Row, New York), pp. 332–371.
14. de los Santos, C., Rosen, M. & Patel, D. (1989) *Biochemistry* **28,** 7282–7289.
15. Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. (1991) *Biochemistry* **30,** 9022–9030.
16. Sklenar, V. & Feigon, J. (1990) *Nature (London)* **345,** 836–838.
17. Banville, D. L., Marzilli, L. G. & Wilson, W. D. (1986) *Biochemistry* **25,** 7393–7401.
18. Inman, R. B. (1964) *J. Mol. Biol.* **9,** 624–637.
19. Gray, D. M., Ratliff, R. L. & Vaughn, M. R. (1992) *Methods Enzymol.* **211,** 389–406.
20. Edwards, E. L., Patrick, M. H., Ratliff, R. L. & Gray, D. M. (1990) *Biochemistry* **29,** 828–836.
21. Kallenbach, N. R., Daniel, W. E., Jr., & Kaminker, M. A. (1976) *Biochemistry* **15,** 1218–1224.
22. Luo, J., Sarma, M. H., Yuan, R.-d. & Sarma, R. H. (1992) *FEBS Lett.* **306,** 223–228.
23. Robinson, H., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1992) *Biochemistry* **31,** 10510–10517.
24. Robinson, H. & Wang, A. H.-J. (1992) *Biochemistry* **31,** 3524–3533.
25. Rippe, K., Fritsch, V., Westhof, E. & Jovin, T. M. (1992) *EMBO J.* **11,** 3777–3786.
26. Li, Y., Zon, G. & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 26–30.
27. Li, Y., Zon, G. & Wilson, W. D. (1991) *Biochemistry* **30,** 7566–7572.
28. Hore, P. J. (1983) *J. Magn. Reson.* **55,** 283–300.
29. Rippe, K. & Jovin, T. M. (1992) *Methods Enzymol.* **211,** 199–220.
30. Dolinnaya, N. G. & Fresco, J. R. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9242–9246.
31. Cech, T. R. & Bass, B. L. (1986) *Annu. Rev. Biochem.* **55,** 599–629.
32. Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. (1983) *Microbiol. Rev.* **47,** 621–669.
33. Stollar, B. D. (1990) *Curr. Opin. Immunol.* **2,** 607–612.