5′-CGA Motif Induces Other Sequences To Form Homo Base-Paired Parallel-Stranded DNA Duplex: The Structure of $(G\text{-}A)_n$ Derived from Four DNA Oligomers Containing $(G\text{-}A)_3$ Sequence

Howard Robinson,† Jacques H. van Boom,‡ and Andrew H.-J. Wang*,†

Biophysics Division and Department of Cell and Structural Biology University of Illinois at Urbana-Champaign Urbana, Illinois 61801 Gorlaeus Laboratories, Leiden State University Leiden 2300RA, The Netherlands

Received May 3, 1993

Alternating $(G\text{-}A)_n$ sequence is unusually abundant in some higher organisms such as rodents and primates.[1,2] This sequence exhibits interesting properties under the influence of pH. At low pH, plasmids containing $(G\text{-}A)_n$ inserts undergo conformational transitions which have been attributed to the formation of pu·py·py triplexes.[3-6] This leaves the extruding purine strand in an unknown structural state. A number of studies have attempted to elucidate the structures of $(G\text{-}A)_n$ sequences. Earlier studies using synthetic polymers such as poly(G-A) or poly(G-A-A) suggested that multistranded (e.g., duplex) structures are possible.[7,8] Feigon et al. studied the oligonucleotides $(G\text{-}A)_4$ and $(A\text{-}G)_4$ by NMR and found that although there were some indications of ordered structures for the two octamers, structural interpretation was not possible due to the broad lines of their NMR spectra.[9] More recently, two circular dichroism studies of $(G\text{-}A)_n$ oligomers[10,11] showed a negative band at 190 nm which was attributed to G(*syn*) conformation by analogy for Z-DNA. On the basis of pyrene fluorescent quenching experiments, Rippe et al. proposed a homo base-paired parallel-stranded duplex structure for the $(G\text{-}A)_n$ sequence.[10] We have unambiguously shown by NMR analyses that certain sequences (e.g., 5′-CGA) form unusually stable homo base-paired parallel-stranded duplex (Π-DNA) structures.[12,13] In this new structure, all bases are in the *anti* conformation, with the G:G pair using N^3–N^2 hydrogen bonds and the A:A pair using N^6–N^7 hydrogen bonds. Here we show, by analyses of 2D-NOE data of four $(G\text{-}A)_3$-containing DNA oligomers of CGAGAGA, CGAGAGAC, CGA[c^{7}G]AGAC and CGAGA-[c^{7}G]AC, that the contiguous $(G\text{-}A)_n$ sequence adopts the same type of Π-DNA structure as in the Π(CGA) helix.[13]

A pH–titration study of CGAGAGA indicated that one proton was absorbed with a pK_a of 6.0 and that six more protons were absorbed as the pH was lowered to 3.0. The pH-dependent 1D NMR study (Figure 1S, supplementary material) showed that the heptamer adopts a single species conformer at pH below 5.5 as judged by the relatively sharp resonances and its uniquely-interpretable 2D-NOE spectrum. The temperature-dependent study of CGAGAGA at pH 4.0 indicated a reasonable stability of the structure. The intensities of the NOE crosspeaks of the phase-sensitive 2D-NOESY spectra of all four oligomers were measured and used in each 3D structural refinement.[14,15]

In our earlier studies of other Π-DNA structures, we noted that some characteristic NOEs are associated with the Π(CGA) motif. Specifically, the NOE crosspeaks resulting from the neighboring interstrand G(n)H^8–A(n + 1)*H^2 protons unique to a Π duplex are strong.[13] In the 2D-NOE spectra of all four oligomers, the G2H^8–A3*H^2 crosspeak is clearly visible (Figure 1). In addition, H$^{2'}$ and H$^{2''}$ of C1 are very upfield shifted (data not shown), as in CGACGAC.[13] These data strongly suggest that the first three (CGA) nucleotides are in the Π-DNA conformation.

What about the remaining GAGA sequence? In CGAGAGA and CGAGAGAC, the G4H^8–A5*H^2 and the G6H^8–A7*H^2 crosspeaks are obscured by the diagonal peaks. To remove this ambiguity, we recorded the 2D-NOESY spectra of CGA[c^{7}G]-AGAC and CGAGA[c^{7}G]AC. The additional proton at G–C^7 serves very useful purposes: (1) it proves that N^7 of G is not involved in base pairing, (2) the H^7–H^8 system provides improved chemical shift dispersion, and (3) the additional H^7 helps define the orientation of guanine base in the NOE-constrained structural refinement. Figure 1 shows unequivocally that the [c^{7}G]4H^8–A5*H^2 crosspeaks in CGA[c^{7}G]AGAC and the [c^{7}G]6H^8–A7*H^2 crosspeaks in CGAGA[c^{7}G]AC are present and strong. The corresponding [c^{7}G]4H^7–A5*H^2 and [c^{7}G]6H^7–A7*H^2 crosspeaks are also strong (data not shown). These data prove that the same (G-A) Π conformation in 5′-CGA has propagated into the remaining GAGA sequence. We have refined the structure of CGAGAGA by a quantitative NOE refinement of a model consisting of Π(GA) motifs, and the refined model (Figure 2) produces simulated NOE spectra (Figure 2S, supplementary material) very similar to the experimental NOE spectra.

Although this structure is coerced by the 5′-CGA sequence motif into a Π conformation, it is likely that $(G\text{-}A)_n$ sequence

† University of Illinois.

‡ Leiden State University.

(1) Abbreviations used: Pu, purine; Py, pyrimidine; c^{7}G, 7-deazaguanine; NMR, nuclear magnetic resonance; 2D-NOESY, two-dimensional nuclear Overhauser effect spectroscopy.

(2) Manor, H.; Rao, B. S.; Martin, R. G. *J. Mol. Evol.* **1988**, *27*, 96–101.
(3) Htun, H.; Dahlberg, J. E. *Science* **1988**, *241*, 1791–1796.
(4) Johnston, B. H. *Science* **1988**, *241*, 1800–1804.
(5) Pulleyblank, D. E.; Haniford, D. B.; Morgan, A. R. *Cell* **1985**, *42*, 271–280.
(6) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.
(7) Lee, J. S. *Nucleic Acids Res.* **1990**, *18*, 6057–6060.
(8) Lee, J. S.; Johnson, D. A.; Morgan, A. R. *Nucleic Acids Res.* **1979**, *6*, 3073–3091.
(9) Feigon, J.; Gilbert, D.; Rajagopal, P.; Wang, E.; van der Marel, G. A.; van Boom, J. H. In *Structure and Methods, Vol. 3: DNA and RNA*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Schenectady, NY, 1990; pp 207–224.
(10) Rippe, K.; Fritsch, V.; Westhof, E.; Jovin, T. M. *EMBO J.* **1992**, *11*, 3777–3786.
(11) Dolinnaya, N. G.; Fresco, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9242–9246. Dolinnaya, N. G.; Braswell, E. H.; Fossella, J. A.; Klump, H.; Fresco, J. R. *Biochemistry* **1993**, *32*, 10263–10270.
(12) Robinson, H.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 10510–10517.
(13) Robinson, H.; Wang, A. H.-J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5224–5228.
(14) Robinson, H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 3524–3533.

(15) Solutions (~5.4 mM with 0.15 M NaCl) of the four DNA oligomers were prepared and titrated to an equivalent of pH ~ 4.0 in 0.5 mL of D_2O as described earlier.[14] NMR spectra were collected on a GE GN500 or a Varian VXR500 500-MHz spectrometer, and the data were processed with FELIX v1.1 (Hare Research, Woodinville, WA). The 2D-NOE spectra were collected at 5 °C at a mixing time of 200 ms and a total recycle delay of 4 s. The data were collected by the States technique[20] with 512 t_1 increments and 2048 t_2 complex points, each the average of 16 transients. Apodization of the data in the t_1 and t_2 dimensions consisted of 4-Hz exponential multiplication with one-half of a sine-squared function for the last one-fourth of the data to reduce truncation artifacts. Integrals from the 2D-NOE data set were extracted by evaluation with the observed crosspeak shapes of each spin in the f_1 and f_2 dimensions. These shapes were determined by spectral analysis with the program MYLOR. Refinement of the starting model was carried out by the sequence of procedures comprising the SPEDREF package.[14] This includes a full matrix relaxation calculation of the NOEs for the model with comparison of the experimental and simulated spectra to deconvolute overlapped areas of the spectra. Minimization of the residual errors within the program X-PLOR is then performed by conjugate gradient minimization of the NOE-derived force-springs together with the chemical force field. The initial model was derived from the model for $(C\text{-}G\text{-}A)_n$.[13] Segments of the start model were systematically overwound or underwound to evaluate convergence to the resulting model. At various stages, sections of refined models were concatenated to compile new models with the best aspects of several models as evaluated by examination of residual errors for section of the model.

0002-7863/94/1516-1565$04.50/0

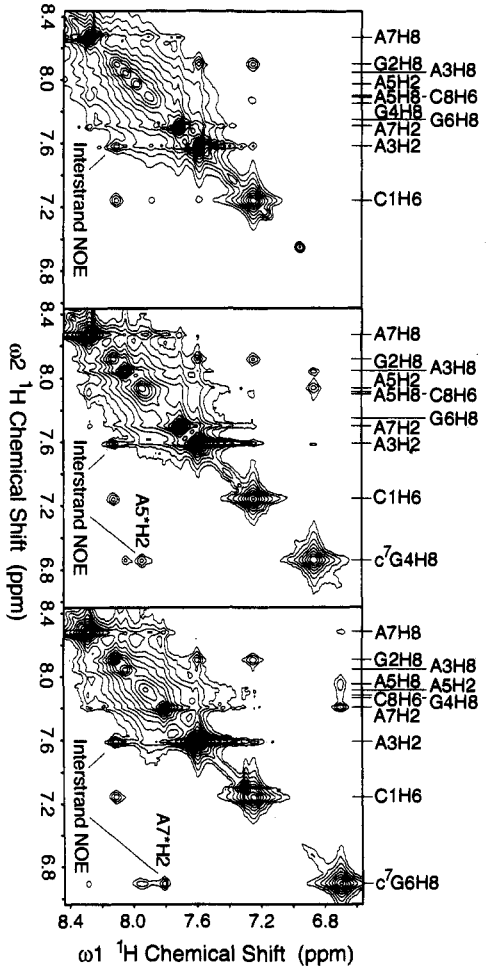

Figure 1. Expanded aromatic–aromatic region of the nonexchangeable proton 2D-NOESY spectra of the CGAGAGAC (left), CGA[c^{7}G]AGAC (middle), and CGAGA[c^{7}G]AC (right) duplexes at pH ~ 4.0. The indicative experimental NOE crosspeaks between the interstrand G(n)-H^8–A(n + 1)*H^2 are marked.

adopts this structure even without the assistance of the protonated C$^+$:C base pair. Note that the GpA step is significantly stabilized by the interstrand G–A stacking interactions. Using the GpA motif from the refined Π(CGAGAGA) helix, we have constructed a model for the Π(G-A)$_n$ helix (Figure 3S, supplementary material). In addition to the two hydrogen bonds between the two purine bases, there are additional interstranded hydrogen bonds from the N^6 of A to a phosphate oxygen and from the N^1 of G to the O$^{4'}$ of an opposite A, respectively. This model is tightly wound with eight base pairs per turn, resulting in an unusual square-tube shape. The Π(GA) structure explains the data of Rippe et al.[10] Specifically, it is a parallel structure, as shown by the pyrene fluorescent quenching data. The reactivity toward DMS is not affected for guanine, since the N^7 of guanine is not involved in the hydrogen-bonding interaction. However, our

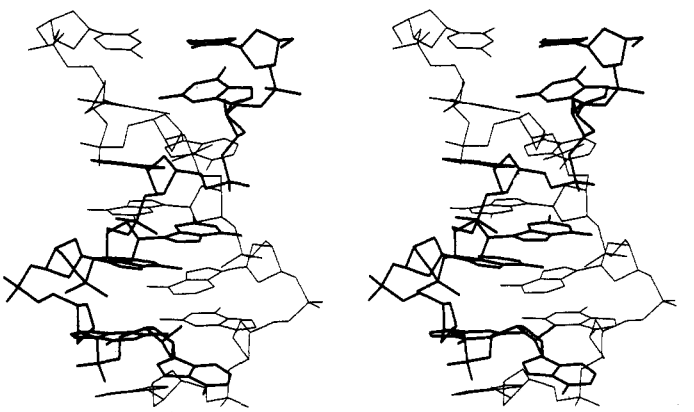

Figure 2. Molecular model of the CGAGAGA refined structure. There are relatively large propeller twists in the purine–purine base pairs. This model represents a consensus of an ensemble of related conformers which may fit the NOE data to a similar degree. Other models have also been tested. For example, the fully-intercalated tetrastranded structure proposed by Gehring et al.[21] was considered but deemed to be inconsistent with the NOE data.

structure differs from that of Rippe et al., whose proposed model is a PS-duplex with G(*syn*)-G(*syn*) and A(*anti*)-A(*anti*) base pairs.[10] No evidence of any *syn*-G is observed in our NMR data. The model is also different from the single helix proposed by Dolinnaya et al.[11]

An important general implication in this work is that we can attach the strong Π-DNA-forming motif, 5′-CGA, to numerous sequences to probe their potential for forming the Π-DNA structure. Here we show that alternating (G-A)$_n$ sequence adopts the Π(GA) helix with G(*anti*)-G(*anti*) pairs and A(*anti*)-A(*anti*) pairs having N^3–N^2 and N^6–N^7 hydrogen bonds, respectively. A sequence such as C(G-A-A)$_3$ may give us a clue as to a possible Π structure for (G-A-A)$_n$.

It has been noted previously that Π-DNA may be involved in recombination processes.[10,12,13,16] Further, these stable Π-helices differ from B-DNA or Z-DNA[17] such that they may be recognized by specific proteins. Indeed, some poly(G-A) binding proteins have been identified.[18,19] In conclusion, we showed that 5′-CGA sequence can induce a Π-helix for other sequences, including (G-A)$_n$ sequences. This simple yet powerful finding will permit us to systematically explore a broader landscape of conformational space of nucleic acids.

Acknowledgment. This work was supported by a grant from the NIH to A.H.-J.W.

Supplementary Material Available: Three figures showing the pH-dependent 1D NMR spectrum, expanded regions of the 2D-NOESY of CGAGAGA, and a model of parallel-stranded homo base-paired poly(dG-dA) duplex (4 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

(16) Rippe, K.; Kuryavyi, V. V.; Westhof, E.; Jovin, T. M. *Structural Tools for the Analysis of Protein Nucleic Acid Complexes*; Birkhauser Verlag: Basel, 1992, p 81.
(17) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G. A.; Rich, A. *Nature* **1979**, *282*, 680–686.
(18) Davis, T. L.; Firulli, A. B.; Kinniburgh, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9682–9686.
(19) Hoffman, E. K.; Trusko, S. P.; Murphy, M.; George, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2705–2709.
(20) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*, 286–292.
(21) Gehring, K.; Leroy, J.-L.; Gueron, M. *Nature* **1993**, *363*, 561–565.