Formation of an i-Motif Structure by a Self-Complementary DNA Sequence

Shanteri Singh and R. V. Hosur*

Tata Institute of Fundamental Research
Homi Bhabha Road, Bombay 400 005, India

Received August 7, 1995

Our knowledge about DNA structure and function has undergone rapid enrichment in recent years with discoveries of several new types of structures being added to the literature. The new structures include triplexes, tetraplexes, parallel stranded structures, etc.,[1-8] the latest being the so-called four-stranded "i-motif" exhibited by cytidine stretches at low pH.[1,2] Several speculations and observations have been made about their possible structure−function relationships, and these have been reviewed recently.[1,9,10] In this communication, we report a puzzling NMR observation of an i-motif structure formed by a self-complementary sequence d-GGAAGATCTTCC under the normal conditions of near neutral pH, temperature, and millimolar ranges of salt concentrations. The structure is formed by interdigitation of two duplexes with antiparallel strands and is characterized by complete sets of sequential H1′−H1′, H1′−H3′ and many other interstrand NOEs. To our knowledge, this is the first such observation for a self-complementary DNA sequence and opens up new possibilities for DNA packing and function inside a living cell.

Figure 1a shows a region of the two-dimensional NOESY[11] spectrum of the DNA segment d-GGAAGATCTTCC in 90% H_2O + 10% D_2O at pH 8.3, displaying cross peaks emanating from the imino protons of the molecule. The peak assignments have been readily obtained following standard procedures,[12] and these indicate that the DNA exists as an antiparallel duplex with Watson−Crick base pairs. Two-dimensional NOESY spectra recorded in D_2O with the same sample showed all the cross peaks expected of such a duplex and permitted assignment of all the non-exchangeable protons in the molecule. The total number of resonances in any of the spectra corresponded exactly to what was expected of a single structure in solution. The relative intensities of cross peaks providing sequential assignment in the NOESY were consistent with a right-handed B-DNA structure for the molecule.

For a self-complementary sequence, these observations are not surprising. However, we observed that the NOESY spectra contained several unusual and intense cross peaks at all mixing times in the range 40−400 ms. These include A6H2−(A6H1′, T7H1′, C8H1′), C11H5−G1H1′, A3H2−(A3H1′, T10H1′, C11H1′), A4H2−A4H1′, C12H6−G1H1′, and G1H8−C12H5, and the most notable are the sequential H1′−H1′, H1′−H3′, and H1′−H4′ connectivities. Of these the H1′−H1′ and H1′−H3′ connectivities are shown in Figure 2, parts a and b, respectively. In a normal duplex, these distances are more than 5 Å and do not produce cross peaks in low mixing time (<100 ms) NOESY spectra. In fact, we have never observed such

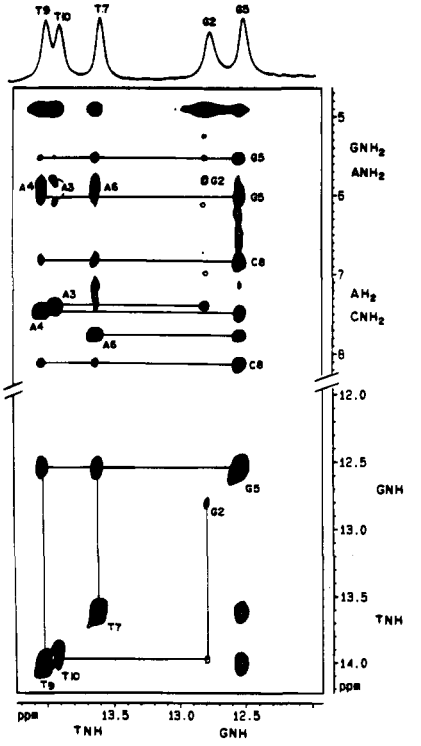

Figure 1. A region of the two-dimensional NOESY spectrum (mixing time, 150 ms) of the DNA segment d-GGAAGATCTTCC in 90% H_2O and 10% D_2O, with 10 mM phosphate buffer, 0.1 mM EDTA, and 100 mM NaCl, pH = 8.3, at 293 K. DNA concentration is 9 mM on a single-strand basis. Sequential imino−imino connections have been drawn in the lower spectral region. Horizontal lines in the upper region identify NOEs between adjacent base pairs. For water suppression, the last 90° pulse in the NOESY pulse sequence was replaced by a "jump and return" [14] sequence. The data was multiplied by phase-shifted sine functions prior to 2D Fourier transformations. Chemical shifts have been expressed with respect to 3-trimethylsilyl[2,2,3,3-$^2H_4$]-propionate.

H1′−H1′, H1′−H3′, and H1′−H4′ NOEs, in any of the self-complementary DNA sequences (more than 10) of similar lengths that we have investigated in the past. Neither have we seen them in the literature in antiparallel duplexes at neutral pH conditions. All these peaks persisted on changing the pH to 5.5 and also on diluting the sample 2.5-fold (data not shown).

These observations indicated that the DNA had a more complex structure than a simple duplex under our experimental conditions. A qualifying candidate which keeps the integrity of antiparallel base pairing and satisfies the unusual NOEs is a four-stranded i-motif, similar to the one seen by Gehring et al.[2] The difference, however, is that the strands in each duplex are antiparallel as against the parallel strands in the cytidine i-motif. Independent evidence for such a conclusion came from the PAGE analysis of the DNA under nondenaturing conditions, (Figure 3) for different concentrations, ranging from 33 to 200 μM. In all cases, four bands are seen. The strongest band corresponds to a duplex, and the one above belongs to a higher molecular weight structure, which can be identified as a tetramer. The two bands on the lower side of the duplex band belong to lower molecular weight structures and may be identified as hairpin and extended single strand structures. These bands have been labeled as T (tetramer), D (duplex), H (hairpin), and S (extended single strand)), respectively. Under denaturing conditions of the gel a single band corresponding to 12-mer DNA was observed. These observations indicate that at low DNA concentrations, as used in the nondenaturing gels, a complex equilibrium of DNA structures exists in solution, and the intensities of the different bands in Figure 3 suggest that the duplex is the most dominant structure, but others also make significant contributions. At the NMR concentrations, on the

(1) Chen, L.; Cai, L.; Zhang, X.; Rich, A. *Biochemistry* **1994**, *33*, 13540−13546.
(2) Gehring, K.; Leroy, J.-L.; Gueron, M. *Nature* **1993**, *363*, 561−565.
(3) Skelnar, V.; Feigon, J. *Nature* **1990**, *345*, 836−838.
(4) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164−168.
(5) Olivas, W. M.; Mather, L. J. *Biochemistry* **1995**, *34*, 278−284.
(6) Radhakrishnan, I.; de Los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1992**, *221*, 1403−1418.
(7) Strobel, S. A.; Dervan, P. B. *Nature* **1991**, *350*, 172−174.
(8) Cheng, Y. K.; Pettit, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465−4474.
(9) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 17−34.
(10) Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlraub, F. *FSSEB J.* **1988**, *2*, 2939−2949.
(11) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6.
(12) Wuthrich, K. *NMR of Proteins and Nucleic acids*; Wiley: New York, 1986.

0002-7863/95/1517-12637$09.00/0

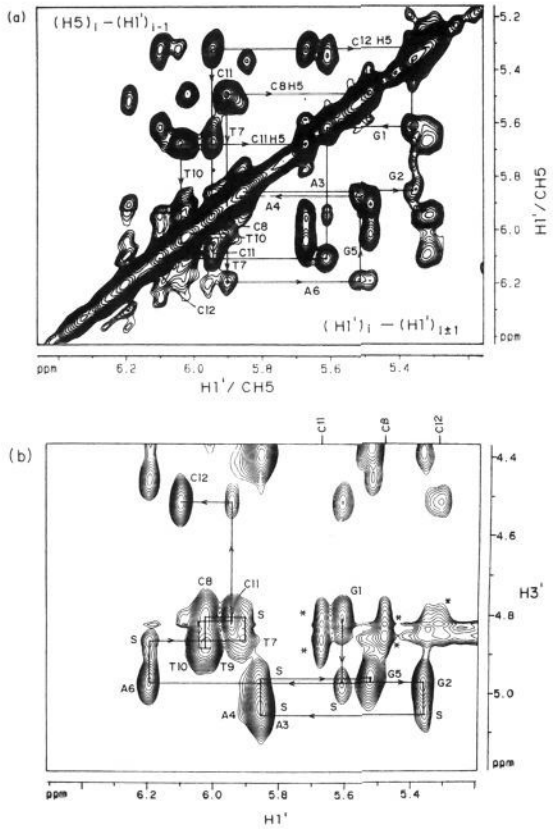

Figure 2. Regions of the two-dimensional NOESY spectrum (mixing time, 150 ms) in 99.996% D_2O at 293 K, showing (a) H1′−H1′ pseudosequential connectivities (lower triangle) and H5−H1′ connections (upper triangle) and (b) H1′−H3′ sequential connectivities. Self-peaks have been labeled by nucleotide labels, and sequential peaks have been marked "S". The peaks marked with asterisks originate from cytosine H5's.

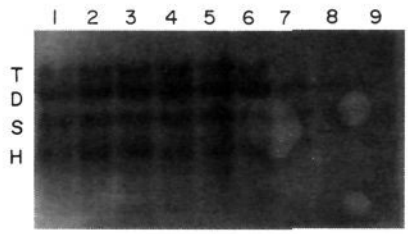

Figure 3. Gel electrophoresis of d-GGAAGATCTTCC, under nondenaturing conditions. Different lanes indicate different concentrations, in the range 200 μM (lane 1) to 33 μM (lane 9). Method: The oligomer was 5′ labeled by incubating the oligomer in a solution containing [^{32}P]-ATP, kinase enzyme, and kinase buffer for 2 h. Gel electrophoresis was carried out with a 20% nondenaturing polyacrylamide gel at room temperature. The labels T, D, S, and H represent the tetramer, duplex, extended single strand, and hairpin structures, respectively.

other hand, the number of peaks is indicative of only one structure and this must be the tetrameric structure. A simple aggregation of duplexes does not produce the unsusual NOEs described above. The formation of the tetramer must arise from destabilization of the hairpin and consequent interaction of the single strands with the duplex to shift the equilibrium toward the tetrameric structure.

A schematic of the i-motif structure formed by the DNA segment is shown in Figure 4a. Every base pair from one duplex has an identical base pair from the other duplex as one of its neighbors, and one duplex is vertically shifted with respect to the other to permit interdigitation of the two duplexes. We have built a model of such a tetrameric structure starting with B-DNA geometries, and we have energy minimized it using base pair constraints alone given in the form of distance constraints in

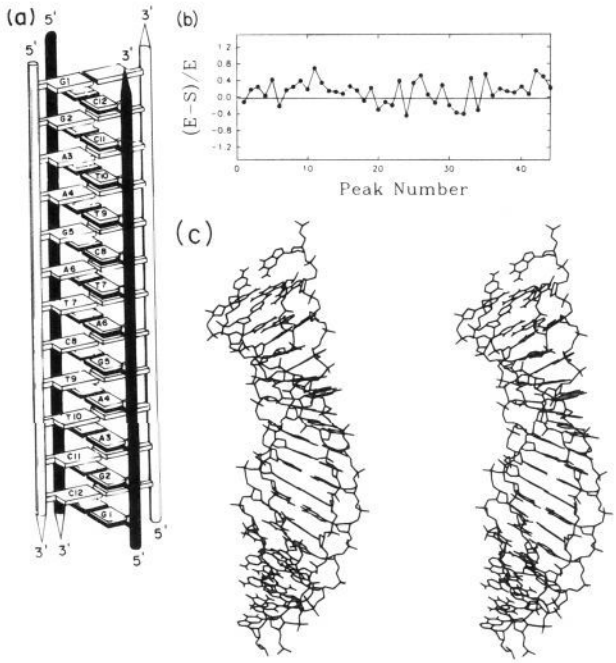

Figure 4. Model of the oligomer d-GGAAGATCTTCC. (a) A schematic of the tetraplex with two intercalating antiparallel duplexes. (b) Peak-to-peak fit for experimental (E) and calculated (S) NOEs for distinct peaks only from all the regions. (c) Stereoview of the structure obtained after iterative relaxation matrix calculation and structure optimization by restrained simulated annealing refinement.

the X-PLOR force field.[13] Indeed we observed several of the unusual interstrand sugar−sugar short distances in the structure. Next, we quantified the distinct peaks in a 200 ms NOESY spectrum and carried out iterative relaxation matrix calculation[15,16] and structure optimization, by restrained simulated annealing refinement to fit these peaks. The resultant peak-to-peak fit between the calculated (S) and the experimental (E) peaks is shown in Figure 4b. Most of the peaks are seen to fit within 30%, and this indicates that the corresponding structure is a good preliminary structure for further search, with additional peaks which were eliminated earlier as overlapping peaks. The preliminary structure so derived for the tetramer is shown in Figure 4c. We observe that one duplex sits in the minor groove of the other duplex with a slight vertical shift along the helical axis and the stacking of the adjacent base pairs is limited to exocyclic atoms. Both duplexes are stretched to some extent but remain right-handed. More detailed investigations are in progress, which will form the subject matter of a future elaborate paper.

What is so special about this sequence is not clear to us now. It has a six unit long purine stretch followed by a six unit long pyrimidine stretch. Whether this has any special significance is hard to say. Nonetheless, the observation of such an unusual structure for a self-complementary sequence at near neutral pH is very fascinating.

Acknowledgment. The facilities provided by the "National Facility for high field NMR" are gratefully acknowledged. We thank Bhabha Atomic Research Centre for the X-PLOR usage. We also thank Dr. Anup Madan for help with gel electrophoresis.

JA9526473

(13) Brunger, A. *X-PLOR (Version 3.1) Manual*; Fellows of Harvard University: New York, 1992.
(14) Plateau, P.; Gueron, M. J. *J. Am. Chem. Soc.* **1982**, *104*, 7310−7311.
(15) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95−117.
(16) Keepers, J. W.; James, T. L. *J. Magn. Reson.* **1984**, *57*, 404−426.