New Antiparallel Duplex Motif of DNA CCG Repeats That Is Stabilized by Extrahelical Bases Symmetrically Located in the Minor Groove

Xiaolian Gao,*,†,‡ Xuening Huang,† G. Kenneth Smith,‡ Minxue Zheng,† and Hongyu Liu†

Department of Chemistry and Department of Biochemical and Biophysical Sciences University of Houston, Houston, Texas 77204-5641

Received January 30, 1995

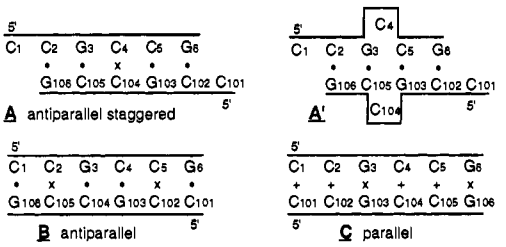
Genetic mutation studies have revealed that abnormal expansion or contraction of certain triplet repeat (TR) sequences, such as CCG/CGG and CTG/CAG, is implicated in more than 10 neurodegenerative diseases.[1] TR sequences are often found within or near relevant gene-coding regions[1b] and thus have promise as diagnostic probes and perhaps as targets for new genetic therapies for related hereditary diseases. To understand the possible mechanisms of dynamic mutations of TR sequences, we have initiated an investigation into the solution conformation of the six CXG and GXC triplets (X = A, C, G, T) using NMR methods. The presence of a stable structure associated with these TR sequences might play a role in disrupting the normal replication or transcription process.[2]

Here we report a novel duplex motif formed by DNA CCG TR sequences in a neutral to basic pH range. The structure is characterized by NMR spectroscopy, electrophoresis, and NMR-based calculations. A series of CCG TR sequences (all sequences refer to DNA, unless otherwise noted), which include the homologous sequences CGCCG and $(CCG)_n$ (n = 2, 3, 4, and 5), the permutation isomers $(CGC)_2$ and $(GCC)_2$, and the linker sequence $(CCG)_2$-xx-$(CCG)_2$ (x = triethylene glycol linker), have been examined to supplement the results pertaining to the $(CCG)_2$ sequence.

NMR spectra of $(CCG)_2$ in aqueous solutions containing 0.1−0.2 M NaCl at pH 6.3−8.5 provide strong evidence for a distinct, stable Watson−Crick base paired structure. The complete ^1H and ^{31}P assignments and a summary of the major spectral features of $(CCG)_2$ are given in Tables S1 and S2, respectively, in the supporting information. $(CCG)_2$ forms a symmetrical duplex, which displays ^1H and ^{31}P resonances corresponding to a single strand of the hexamer (Figure S1, Table S1, supporting information).[3] Twenty percent polyacrylamide nondenaturing gel electrophoresis shows that $(CCG)_2$ and $(CCG)_3$ duplexes migrate faster than the corresponding complementary hexamer and nanomer duplexes at 4 °C,[4] excluding the possible presence of higher order structures. In the spectra recorded in 90% H_2O (Figure S1A) G3 and G106 HN (imino) resonances at 13.04 and 13.48 ppm show an HN−HN NOE and HN−H_2N (amino) NOEs (Figure S1A). These observations, especially the presence of the HN−HN NOE between G3 and G106 residues, are unequivocal evidence for the staggered alignment of $(CCG)_2$ (Scheme 1).

The ^{31}P 1D spectrum of $(CCG)_2$ (Figure S1B) exhibits four well-resolved signals in a chemical shift range of 2.1 ppm [3′-P (ppm) of C2 −2.84; C5 −3.45; C1 −4.04; C4 −4.09; G3 −4.97]. The ^{31}P resonances of canonical DNA duplexes are usually detected in a ~0.6 ppm region.[5] The large dispersion of ^{31}P resonances of $(CCG)_2$ is diagnostic of an untwisted and extended duplex conformation as observed in several antibiotic−DNA complexes.[6] Further evidence for a distorted $(CCG)_2$ duplex has been established by spectral analyses of several other data sets. An unusual characteristic of proton resonances in $(CCG)_2$ is their slower spin−lattice relaxation rate compared to those of a B-DNA hexamer duplex. For instance, the nonselective H1′ T_1 of $(CCG)_2$ is approximately 2.9 s in D_2O solution, which compares to the 2.0 s for the H1′ of the complementary $(CCG)_2$·$(CGG)_2$ duplex under comparable conditions. This is consistent with an increase in interresidue separations (*vide infra*), which result in ineffective interproton dipolar cross relaxation. The sequential NOEs between base and sugar protons are unusually weak and are barely visible in the D_2O NOESY spectra of short mixing times (50 and 120 ms). They do, however, follow the specific connectivities for a right-handed duplex including all anti-glycosidic bonds as revealed by NOESY spectra (Figures 1 and S2).[7] These results are unexpected even for a staggered duplex (Scheme 1A). Spectral analyses of various COSY data suggest that the average sugar conformation of the $(CCG)_2$ duplex is in a C2′-endo range (Figure S3).

Scheme 1. Possible Alignments of the $(CCG)_2$ Duplex[a]



[a] Base pair symbols: Watson−Crick (●); mismatch (×); and protonated C (+). The distinction between type A and B alignments is that dinucleotides on either side of the unpaired C residue are 5′-CG in type A and 5′-GC in type B, respectively. NMR data support the type A alignment, while the exact form is depicted in A′. This type of extrahelical alignment is similar to what was predicted for polynucleotides. (Fresco, J. R.; Alberts, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **1960**, *46*, 311−321.)

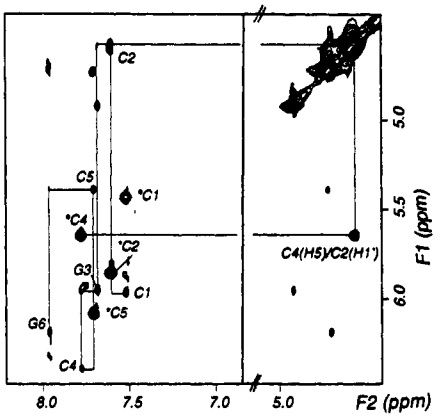


Figure 1. Expanded NOESY spectrum (120 ms mixing time) of $(CCG)_2$ showing base (F2 axis) to sugar H1′ (F1 axis) connectivities. Spectrum was recorded in D_2O containing 0.2 M NaCl, 20 mM sodium phosphate, and 0.2 mM Na_2EDTA, pH 6.3, −10 °C. Sequential NOEs are shown by solid lines in the left panel. Residue numbers are given to intraresidue NOEs. NOEs of C H6 and H5 are indicated by *. The base proton resonances of C in the 7.5−7.8 ppm region are downfield shifted compared to those in a canonical duplex (7.1−7.6 ppm). The interresidue NOEs are much weaker than what are expected for a typical B-type helix. The right panel shows an important NOE between C4 H5 and C2 H1′ protons.

* To whom correspondence should be addressed.

† Department of Chemistry.

‡ Department of Biochemical and Biophysical Sciences.

(1) Reviewed in the following: (a) Caskey, C. T.; Pizzuti, A.; Fu, Y.-H.; Fenwick, R. G., Jr.; Nelson, D. L. *Science* **1992**, *256*, 784−789. (b) Willems, P. J. *Nature Genetics* **1994**, *8*, 212−214.

(2) (a) Kang, S.; Jaworski, A.; Ohshima, K.; Wells, R. D. *Nature Genetics*, in press. (b) Mitas, M.; Yu, A.; Dill, J.; Kamp, T. J.; Chambers, E. J.; Haworth, I. S. *Nucleic Acids Res.*, in the press.

(3) For simplicity in the discussion, most times only one of the symmetrical residues is mentioned.

(4) Data not shown. The migration band for $(CCG)_3$ is relatively broad, which may correspond to a mixture of several conformers as discussed later.

(5) (a) Patel, D. J.; Kozlowski, S. A.; Nordheim, A.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 1413−1417. (b) Chen, C.-W.; Cohen, J. S. In *Phosphorus-31 NMR. Principles and applications*; Gorenstein, D. G., Ed.; Academic Press: New York, 1984; pp 233−264.

(6) Gao, X.; Patel, D. J. *Q. Rev. Biophys.* **1989**, *22*, 751−762.

(7) Wuthrich, K. *NMR of Proteins and Nucleic Acids;* John Wiley & Sons, Inc.: New York, 1986. Chemical shifts of nucleic acids: pp 20−23. NMR of nucleic acids: pp 203−255.

0002-7863/95/1517-8883$09.00/0

The presence of a number of long-range NOEs in $(CCG)_2$ as shown in Figures 1, S1, and S2 and Table S3, especially the NOEs connecting H1′, H2′, and H2″ protons of C2 with base protons of C4 residues, indicates that CCG repeats form a unique structure. These NOEs, in conjunction with the unusual NOEs involving the C2, G3, and C4 residues (Figure S2, Table S3), reveal a major interruption in the helical conformation. C4 residues in the symmetrical strands of the $(CCG)_2$ duplex do not stack-in to form a C × C mismatch[8] but protrude into the minor groove pointing their base moieties in the direction of the $i-2$ residue (i is the residue number). The bond angle strain energies which arise from the extrahelical conformation of C4 may be compensated by the new contacts between C4 and other residues in the minor groove and by the stacking interactions between G3•C105 and G103•C5 base pairs at adjacent sites (Scheme 1). Indeed, these G and C residues show a number of close contacts (G3 sugar to C5 base) in the NOESY spectra (Figure S1 and Table S3). Additional stabilization of the $(CCG)_2$ duplex can be attributed to the terminal C1 residues, which adopt a stacked-in conformation. This is demonstrated by the far upfield shifted C1 H2′ (1.16 ppm), the NOEs to the adjacent C2, and the contacts linking C1 H5 and H6 to the cross strand G106 HN (Figure S1A and Table S3). Thus, the central component of the $(CCG)_2$ duplex is a base-paired, right-handed, complementary tetramer helix, C2G3−C5G6•C102G103−C105G106, which contains two extrahelical C4 residues in the minor groove symmetrically positioned proximal to the C2 residue of the same strand. Several CCG related sequences (*vide supra*) have been examined to (a) verify the antiparallel orientation of the $(CCG)_2$ duplex, (b) discern the CCG repeat that is responsible for the new structure, and (c) probe the presence of the unusual duplex conformation in longer (three to five copies) CCG repeats (Figures S2 and S4).

Structures drawn in Figure 2 demonstrate the unique features of the $(CCG)_2$ duplex. The structures were derived from distance geometry and restrained molecular simulations using the XPLOR program (Molecular Simulations Inc.).[9] The stick drawing in Figure 2A shows a positive view into the minor groove of the antiparallel duplex. The four G•C base pairs are well-aligned, forming a continuous helix with the terminal C1 residues stacked on the terminal G•C base pairs. The "derailed" C4 residues are accommodated in the minor groove, and the base moieties are oriented such that the hydrophilic O2 and N3 groups face away from the groove floor and point toward the solvent. The C4 bases are tilted relative to the helical axis, and the N4 amino groups are located proximally to the C2 residue of the same strand. Despite the dramatic turns and twists around C4 sites in the $(CCG)_2$ duplex, the overall layout of the structure appears to be quite compact and smooth as shown by the space-filling drawing in Figure 2B. The central two G•C base pairs are well stacked. The structure also features wider minor and major grooves, so that the distinctions between the major and minor grooves as observed in canonical A- and B-form DNA duplexes, such as groove depth and width, become insignificant. The change in the shape of the helical grooves is concomitant with the unwinding of the duplex, which appears to be required by the extrahelical C4 residues for interaction with the C2 residues (Figure 2).

In summary, a set of distinct NMR spectral features characterizing a new form of antiparallel duplex is presented. This new duplex motif, typified by the $(CCG)_2$ duplex, differs from any known DNA structure in its stable, symmetrical disposition of extrahelical C residues in the minor groove, hence it is termed the *e-motif*. This motif is present in a conformational equilibrium mixture for sequences of three, four, or five CCG repeats.

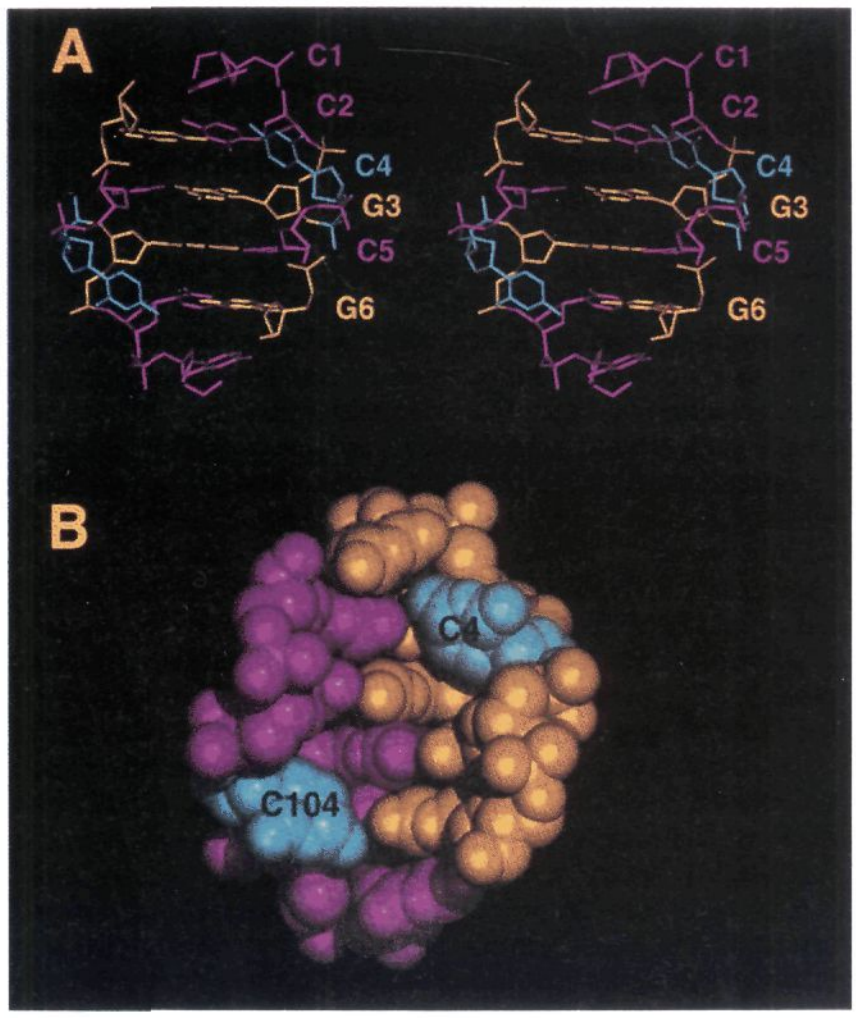


Figure 2. Drawings of the $(CCG)_2$ duplex. Structure refinement of the $(CCG)_2$ duplex using distance geometry and restrained molecular dynamics simulations produces 14 structures. Extraction of distance and dihedral angle restraints from NMR data and calculation protocols are as previously described;[9] 430 NOEs based on 100 or 120 ms NOESY (236 intramolecular, 152 intermolecular, and 42 involving exchangeable protons) and 12 dihedral angle restraints were used in the calculation. The resulting structures converge with respect to the major feature as described in the text. The Cartesian coordinate root mean square deviation (RMSD) of the 14 structures is 1.6196 ± 0.4461 Å (pairwise) and is 1.0999 ± 0.3177 Å (compared to an averaged structure). Fine-tuning of calculated structures and final refinement using relaxation matrix calculations will be reported in due course. (A) Representative stick drawing of the $(CCG)_2$ duplex as viewed into the minor groove. Colors are magenta for base-paired C residues, light blue for the extrahelical C4 residues, and yellow for G residues. The new CCG duplex motif is termed the *e-motif* for its characteristic extrahelical C conformation. (B) Space-filling model of the $(CCG)_2$ duplex. The two extrahelical C4 residues are labeled.

The highly untwisted nature and the intramolecular recognition of the minor groove by the extrahelical C residues in the CCG repeats bring new perspectives to the dogma of duplex formation. Thus far, only A-, B-, and Z-form structures are considered to be typical duplex motifs.[10] The formation of the stable e-motif by short CCG repeats under close to physiological conditions is remarkable and demonstrates the extreme flexibility of ostensibly inflexible DNA duplexes. The molecular origin for formation of the e-motif and its biological relevance in dynamic mutations of CCG triplet repeats in genomic loci of TR expansion diseases remain to be investigated.

Acknowledgment. The 600 MHz NMR spectrometer at the University of Houston is funded by the W. M. Keck Foundation. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society (XG, 26519-G4), and to the Robert A. Welch Foundation (XG, E-1270). G.K.S. and H.L. thank Professors G. E. Fox and H. L. Kohn for their support. We thank the W. M. Keck Center for Computational Biology for computer resource support.

Supporting Information Available: Tables S1−3 and Figures S1−4 containing ^{31}P NMR data and ^{1}H 1D and 2D NOESY and DQF-COSY spectra of $(CCG)_2$ (12 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA950300U

(8) Kouchakdjian, M.; Li, B. F. L.; Swann, P. F.; Patel, D. J. *J. Mol. Biol.* **1988**, *202*, 139−155.

(9) (a) Brunger, A. T. *X-PLOR Version 3.1: A system for X-ray crystallography and NMR*; Yale Univ. Press: New Haven, CT, 1993. (b) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317−324. (c) Kuszewski, J. K.; Nilges, M.; Brünger, A. T. *J. Biol. NMR* **1992**, *2*, 33−56. (d) Gao, X. *J. Mol. Biol.* **1992**, *225*, 125−135.

(10) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, 1984.