JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

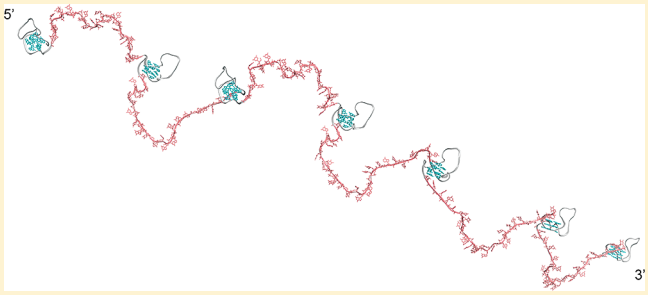
Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite

Samir Amrane,[†,‡,∥] Michael Adrian,[†,∥] Brahim Heddi,[†] Alexandre Serero,[§] Alain Nicolas,[§] Jean-Louis Mergny,*[,‡] and Anh Tuân Phan*[,†]

[†]School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore
[‡]University of Bordeaux, European Institute of Chemistry and Biology, INSERM U869, 33600 Pessac, France
[§]Institut Curie, Centre de Recherche, UMR3244 CNRS, Université Pierre et Marie Curie, 75248 Paris, France

Supporting Information

ABSTRACT: CEB25 is a human minisatellite locus, composed of slightly polymorphic 52-nucleotide (nt) tandem repeats. Genetically, most if not all individuals of the human population are heterozygous, carrying alleles ranging from 0.5 to 20 kb, maintained by mendelian inheritance but also subject to germline instability. To provide insights on the biological role of CEB25, we interrogated its structural features. We report the NMR structure of the G-quadruplex formed by the conserved 26-nt G-rich fragment of the CEB25 motif. In K^+ solution, this sequence forms a propeller-type parallel-stranded G-quadruplex involving a 9-nt central double-chain-reversal loop. This long loop is anchored to the 5′-end of the sequence by an A·T Watson−Crick base pair and a potential G·A noncanonical base pair. These base pairs contribute to the stability of the overall G-quadruplex structure, as measured by an increase of about 17 kcal/mol in enthalpy or 6 °C in melting temperature. Further, we demonstrate that such a monomorphic structure is formed within longer sequence contexts folding into a pearl-necklace structure.



INTRODUCTION

G-quadruplexes are a family of nucleic acid structures formed by G-rich DNA or RNA sequences, built from the stacking of G-tetrads, each being a planar association of four guanines held together by eight hydrogen bonds and stabilized by cations, such as Na^+ or K^+.[1−4] G-quadruplex structures from many human genomic regions including oncogenic promoters and telomeres have been elucidated based on short truncated G-rich fragments derived from natural sequences.[2] These structures show a variety of G-quadruplex folding topologies with respect to strand orientations, glycosidic conformations of guanine bases, and connecting loops. Several G-quadruplex structures of oncogenic promoter sequences exhibit unique folding topologies. For instance, a five-G-tract sequence from the *c-myc* promoter folds into a snap-back configuration with a stable triad-containing diagonal loop,[5] and a G-rich sequence from the *c-kit* promoter forms an unprecedented snap-back parallel G-quadruplex involving an isolated guanine and a 5-nt stem loop.[6] Human telomeric repeats have been found to adopt multiple G-quadruplex conformations depending on the length of the fragments, flanking sequences, and crowding condition.[4,7] Furthermore, the formation of a G-quadruplex structure containing an extended loop has been observed for sequences spanning five or more human telomeric repeats.[8] An array of G-quadruplexes interconnected by short linkers (TTA) could be formed on a long telomeric sequence.[9]

Intramolecular G-quadruplexes adopted by a single DNA strand have attracted much interest since they could be formed within biologically relevant regions of the genome.[10−13] Although the biological role of the numerous potential G-quadruplex-forming sequences in the genome remains to be characterized, such structures which can be formed on single-stranded DNA and RNA, for example, during replication, recombination, and transcription, likely play a role and interfere with these key cellular processes.[13−15] The remarkable G-quadruplex structure adopted by the human CEB25 sequence (locus D10S180), a polymorphic minisatellite in the human population maintained by mendelian inheritance but also subject to germline instability,[16] is under the scope of this paper.

Minisatellites consist of head-to-tail arrays of identical or slightly polymorphic 10−100-bp long motifs. They are present in prokaryote and eukaryote genomes and might constitute as much as 10% of the human genome.[17] Often, the overall size of the minisatellite loci varies between individuals and can range from 0.5 to >15 kbp. These variations correspond to simple or complex gain or loss of repeat units and have initially been used as genetic DNA typing markers.[16,18] The large number of minisatellite loci, the variety of their nucleotide sequences, and their persistence in the genome suggest that they likely possess

Received: September 24, 2011
Published: February 29, 2012

   dx.doi.org/10.1021/ja208993r | J. Am. Chem. Soc. 2012, 134, 5807−5816

biological roles. Interestingly, the yeast natural minisatellites are located in the coding region of cell-wall proteins and variation of the intragenic repeats may have a role in adaptation within environmental changes.[19] In the human genome, numerous G-rich minisatellite loci are located in subtelomeric locations, but internal minisatellite repeats also exist and are associated with numerous disorders, such as epilepsy related to the EPM1 gene,[20] diabetes associated with the insulin gene,[21] and bipolar disorders linked to the serotonin transporter gene 5-HTT,[22] and might also predispose to cancer related to the function of the H-ras gene.[23] On a mechanistic perspective, linked to the underlying mechanisms promoting size variation in somatic and/or germline cells by replication, repair, or recombination mechanisms,[24] it is also suspected that the nature of the DNA sequence itself carries structural properties at the DNA and chromatin levels,[25] as observed for DNA sequences implicated in trinucleotide-repeat expansion diseases,[26−28] able to form noncanonical DNA structures such as hairpin, triplex, quadruplex, or sticky DNA.[29,30] Therefore, structural study of minisatellite sequences might provide fundamental insights to elucidate their functions.

Along this line, here we characterize the sequence polymorphism of the G-rich human CEB25 minisatellite loci and uncover the remarkable structural features of the G-quadruplex forming motif which contains successive guanines prone to G-quadruplex formation, in particular, a segment, AAGGGTGGG-TGTAAGTGTGGGTGGGT, composed of four GGG tracts (underlined) separated by linkers of 1, 9, and 1 nucleotides, respectively. We show that this sequence forms in K^+ solution a propeller-type parallel-stranded G-quadruplex involving a 9-nucleotide (nt) central double-chain-reversal loop. This long loop is anchored to the 5′ end of the sequence by an A·T Watson−Crick base pair and a potential G·A noncanonical base pair, contributing to the stability of the overall G-quadruplex structure. Further, we demonstrate that such a structure is formed in longer sequence contexts and, along the line of its in vivo organization in tandem arrays, fold into a monomorphic pearl-necklace structure.

METHODS

DNA Sample Preparation. Unlabeled and site-specific labeled DNA oligonucleotides (Tables 1 and S1, Supporting Information) were chemically synthesized on an ABI 394 DNA/RNA synthesizer or purchased from Eurogentec (Seraing, Belgium). Samples were purified and dialyzed successively against potassium chloride solution and water as previously described.[8] Unless otherwise stated, DNA oligonucleotides were dissolved in solution containing 70 mM potassium chloride and 20 mM potassium phosphate, pH 7. DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.[31]

Sequencing of Human CEB25 Alleles. CEB25 regions were amplified from human genomic DNA with primers specific to the flanking chromosomal regions (CEB25Up 5′-AAAGACAATGACT-CAGGGTGG-3′ and CEB25LowII 5′-CCGGCACAAACCCTG-CTGCTGGGAGTAAGAGGG-3′). Due to the high GC content of the CEB25 sequence, PCR reactions were performed as previously described[32] and resuspended in a final volume of 15 μL. The PCR products were analyzed by electrophoresis on 1.0% agarose gels to verify the size determined by prior Southern blot analysis, purified, cloned into pGEMT-Easy (Promega), and sequenced by using the Big Dye Terminator version 3.1 kit (Perkin-Elmer) and 20% betaine (Sigma).

Analysis of the CEB25 Size Polymorphism in CEPH Families. Lyophilized DNA from five CEPH families (#1334, 1354, 1355, 1357, and 1417) was suspended in 200 μL of TE buffer. The genomic DNA was purified by phenol/chlorophorm/isoamyl alcohol extraction and ethanol precipitation and resuspended in 20 μL of TE. For each individual, 15 μg of genomic DNA was digested with *Apa*I (3 h at 37 °C) and treated with RNase A for 30 min. The resulting fragments were separated by electrophoresis on 0.8% agarose gels and transferred under vacuum (Qbiogene) onto Hybond N^+ membranes (Pharmacia). The membranes were hybridized with the radiolabeled natural CEB25−1.7 fragment. According to the data provided by the NCBI resources, human CEB25 sequence is flanked by *Apa*I restriction sites, 36 bp upstream and 843 bp downstream, respectively. Therefore, the size of the CEB25 locus is 879 bp smaller than the *Apa*I band detected by Southern blot.

UV-Melting Experiments. The thermal stability of DNA structures was characterized in heating/cooling experiments by recording the UV absorbance at 295 nm as a function of temperature[33] using a UVIKON XL UV/vis spectrophotometer. UV-melting experiments were conducted as previously described[34] at variable DNA strand concentrations ranging from 1 to 125 μM in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl. The heating and cooling rates were 0.2 °C/min. Experiments were performed with 1 and 0.2 cm path length quartz cuvettes.

Thermal Difference Spectra. Thermal difference spectra (TDS) were obtained by taking the difference between the absorbance spectra

Table 1. List of DNA Sequences Analyzed in This Work[a]

sequence name	sequence (5′−3′)
26CEB	AAGGGTGGGTGTAAGTGTGGGTGGGT
m1T	**T**AGGGTGGGTGTAAGTGTGGGTGGGT
m2T	A**T**GGGTGGGTGTAAGTGTGGGTGGGT
m10C	AAGGGTGGG**C**GTAAGTGTGGGTGGGT
m11T	AAGGGTGGGT**T**TAAGTGTGGGTGGGT
m12A	AAGGGTGGGTG**A**AAGTGTGGGTGGGT
m13T	AAGGGTGGGTGT**T**AGTGTGGGTGGGT
m14T	AAGGGTGGGTGTA**T**GTGTGGGTGGGT
m15T	AAGGGTGGGTGTAA**T**TGTGGGTGGGT
m16A	AAGGGTGGGTGTAAG**A**GTGGGTGGGT
m17T	AAGGGTGGGTGTAAGT**T**GGGTGGGT
m18C	AAGGGTGGGTGTAAGTG**C**GGGTGGGT
m(1,2)T	**TT**GGGTGGGTGTAAGTGTGGGTGGGT
m(11,13−15)T	AAGGGTGGGT**TTTTT**GTGGGTGGGT
m(11,13−15,17)T	AAGGGTGGGT**TTTTT**T**T**GGGTGGGT
m(1,2,11,13−15,17)T	**TT**GGGTGGGT**TTTTT**T**T**GGGTGGGT
m1Δ	-AGGGTGGGTGTAAGTGTGGGTGGGT
m(1,2,26)Δ	--GGGTGGGTGTAAGTGTGGGTGGG-
m1Δ2T	-**T**GGGTGGGTGTAAGTGTGGGTGGGT
I7	AAGGGT**I**GGTGTAAGTGTGGGTGGGT
26CEBmut	AAG**T**GTG**T**GTGTAAGTGTG**T**GTG**T**GT
52CEB(5′)	*26CEB*−GTGAGTGTGGGTGTGGAGGTAGATGT
52CEB(3′)	GTGAGTGTGGGTGTGGAGGTAGATGT-*26CEB*
57CEB(mid)	TGGGTGTGGAGGTAGATGT-*26CEB*-GTGAGTGTGGGT
78CEB(5′,3′)	*52CEB(5′)−26CEB*
78CEB(5′,3′I7)	*52CEB(5′)−I7*
78CEB(5′,3′mut)	*52CEB(5′)−26CEBmut*
78CEB(5′mut,3′)	*26CEBmut−52CEB(3′)*
78CEB(mid)	GTGAGTGTGGGTGTGGAGGTAGATGT−*52CEB(5′)*
130CEB(5′,mid,3′)	*52CEB(5′)−26CEB−52CEB(3′)*
130CEB(5′I7,mid,3′)	*I7−52CEB(3′)−52CEB(3′)*
130CEB(5′,midI7,3′)	*52CEB(5′)−I7−52CEB(3′)*
130CEB(5′,mid,3′I7)	*52CEB(5′)−52CEB(5′)−I7*

[a]Long sequences can be read by connecting DNA fragments of respective names and sequences. Mutations (variations from *26CEB*) are shown in boldface.

from unfolded and folded oligonucleotides that were respectively recorded much above and below its melting temperature (T_m). TDS provide specific signatures of different DNA structural conformations.[35] Spectra were recorded between 220 and 320 nm on a JASCO V-650 UV/vis spectrophotometer using 1 cm path length quartz cuvettes. The DNA oligonucleotides (~5 μM) were prepared in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl. For each experiment, an average of three scans was taken, and the data were zero-corrected at 320 nm.

Circular Dichroism. Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using 1 cm path length quartz cuvettes with reaction volume of 600 μL. The DNA oligonucleotides (~5 μM) were prepared in a 20 mM potassium phosphate (pH 7) buffer containing 70 mM KCl. For each experiment, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm.

Differential Scanning Calorimetry (DSC). Microcalorimetry experiments were performed on a Nano DSC-II microcalorimeter as previously described.[26−28] The oligonucleotides were prepared at concentrations ranging from 100 to 150 μM in a 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl. An average of six differential scanning calorimetric (DSC) heating and cooling profiles was taken.

NMR Spectroscopy. NMR experiments were performed on 400, 600, 700, and 800 MHz Bruker spectrometers at 25 °C, unless otherwise specified. The strand concentration of the NMR samples was typically 0.2−2.0 mM. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling,[36] site-specific ^{2}H labeling,[37] and through-bond correlations at natural abundance.[38] Spectral assignments were completed by NOESY, TOCSY, {^{13}C−^{1}H}-HMBC, {^{13}C−^{1}H}-HSQC and {^{31}P−^{1}H}-HSQC, as previously described.[39] Interproton distances were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and $^{2}H_2O$ (mixing times, 100, 150, 200, and 350 ms).

One-bond C1′−H1′, C8−H8, C6−H6, and C2−H2 residual dipolar couplings (RDC) were obtained by comparing the ^{13}C−^{1}H splitting in nondecoupled {^{13}C−^{1}H}-HSQC spectra in aligned and isotropic conditions (see example in Figure S1, Supporting Information) measured on unlabeled DNA. The aligned medium was obtained by adding 20 mg/mL of Pf1 bacteriophage (ALSA Biotech) in an unlabeled DNA sample.

All spectral analyses were performed using the FELIX (Felix NMR, Inc.) and SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) programs.

Structure Calculation. Structure computations were performed using the XPLOR-NIH program with RDC restraints. The procedure lies in two general steps essentially as previously described:[40] (i) distance geometry simulated annealing and (ii) distance-restrained molecular dynamics refinement. RDC restraints, hydrogen-bond restraints, interproton distance restraints, dihedral restraints, planarity restraints, and repulsive restraints were imposed during structure calculations.

The components of the alignment tensor were calculated using optimized structures derived from NMR distance restraints. Only the residues forming the core of the G-quadruplex were used to find the tensor parameters. Structure refinement has improved the correlation between the calculated and measured RDCs (Figure S2, Supporting Information). The dihedral restraints were based on intraresidues NOE of H1′-H6/8 cross-peak intensities (glycosidic dihedral angles) and observed phosphorus chemical shifts (backbone dihedral ε and β angles). The planarity restraints were used in agreement with previously observed G-quadruplex X-ray structures. The repulsive restraints were applied to several pairs of well-resolved protons, which did not give observable NOE cross-peaks. Structures were displayed using the PyMOL program.

Data Deposition. The coordinates of ten lowest-energy d[AAGGGTGGGTGTAAGTGTGGGTGGGT] (*26CEB*) G-quadruplex structures have been deposited in the Protein Data Bank (accession code 2LPW).

RESULTS

Features of the Human CEB25 Polymorphism. The hCEB25 locus is localized on chromosome 10q26.3, approximately 1.3 Mb away from the chromosome end, within the intron 9 of the LRRC27 (leucine-rich-repeat containing protein 27) gene of unknown function. The sequence of the originally reported 0.8-kb-long CEB25 allele (accession number AL096806) is shown in Figure S3, Supporting Information. It is composed of 17 motifs of approximately 52 nucleotides long with a 53% GC content and a strong GC asymmetry; 92% of the G nucleotides are on the same strand.[16] To characterize the length polymorphism of the CEB25 locus in individuals of the human population, we examined genomic DNA from 81 individuals from five CEPH families (Figure S4, Supporting Information). Remarkably, the size of the CEB25 locus is extremely variable, ranging from approximately 0.5 to 20 kb, corresponding to ~10−400 motifs, respectively. Genetically, all individuals are heterozygotes carrying two different paternal and maternal size alleles segregating with mendelian inheritance. Along the three generations examined, we found no de novo size variants (0/81). This indicates that CEB25 has a low germline instability, previously estimated to 2%, from the genotyping of approximately 500 children of other CEPH families (G. Vergnaud, personal communication). Recent studies correlated mammalian meiotic hot spot activity to the local presence of PRDM9 sequence-specific binding site,[41−43] only few motifs of the CEB25 alleles have sequences that correspond to a high affinity binding site of the human PRDM9 protein (Figure S3, Supporting Information). Thus, not surprisingly, CEB25 germline instability is not as high as other minisatellites such as CEB1[44] but remains sufficient to generate the diversity of CEB25 alleles in the human population.

To further characterize CEB25, we cloned and sequenced a second 1.7-kb-long hCEB25 allele. From the comparison of the CEB25−0.8 and CEB25−1.7 alleles and motifs (Figure S3, Supporting Information), several conclusions can be drawn: (i) both CEB25 alleles are canonical minisatellites composed of tandem repeats, (ii) the size of the motifs is highly conserved with a large majority of 52-bp-long repeats but few 36-bp-long motifs, and (iii) the sequence of the motifs is slightly polymorphic. Altogether, we identified 21 and 16 polymorphic sites within the CEB25−0.8 and CEB25−1.7 alleles, respectively. They are mostly single-base substitutions. Specifically, the 31 repeats of the CEB25−1.7 allele include 16 different motifs, and the motifs that are not unique are present in two to four copies. Relevant to the potential of CEB25 motifs to form G-quadruplexes, most repeats (81%) contain at least four of their five guanine triplets (Figure S5, Supporting Information). These G-tracts represent approximately 30% of the sequence of the CEB25−1.7 G-strand, and only 10% (15/155) are affected by single-base substitutions suggesting that the G-tracts are conserved under evolutionary constraint.

G-Rich Segment from CEB25 Minisatellite Forms Stable G-Quadruplex in K^+ Solution. The 26-nt G-rich sequence d[AAGGGTGGGTGTAAGTGTGGGTGGGT] (named *26CEB*) harboring four G-tracts was taken from the 52-nt full repeating unit of the CEB25 minisatellite (locus D10S180 or accession number AL096806) (Table 1). In K^+ solution, the proton NMR spectrum of *26CEB* shows 12 well-resolved imino proton peaks at 10−12 ppm (Figure 1A), indicating the formation of a G-quadruplex involving three G-tetrad layers (four imino proton peaks for each G-tetrad

dx.doi.org/10.1021/ja208993r | J. Am. Chem. Soc. 2012, 134, 5807−5816

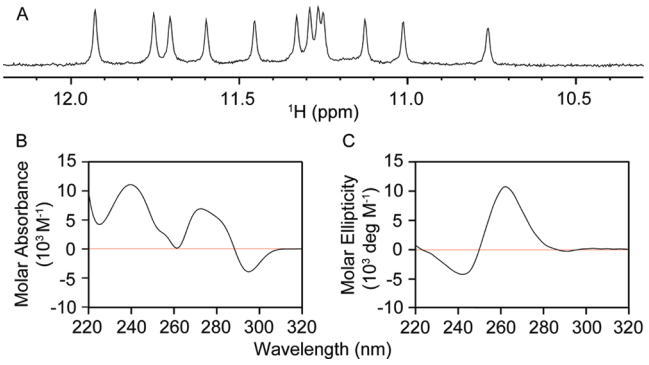


Figure 1. Spectroscopic characteristics of G-quadruplex formation by the *26CEB* G-rich fragment from the CEB25 minisatellite in K^+ solution: (A) imino proton NMR, (B) TDS, and (C) CD spectra of *26CEB*.

layer).[45] The thermal absorption difference spectrum (TDS) of *26CEB* displays typical patterns of a G-quadruplex structure with a negative minimum at 295 nm and two positive maxima at 240 and 275 nm (Figure 1B).[35] The CD profile of *26CEB* is suggestive of a parallel-stranded G-quadruplex with a negative minimum at 240 nm and a positive maximum at 260 nm (Figure 1C).[46] Interestingly, despite the presence of a 9-nt linker between the two central G-tracts (i.e., longer than the maximum loop length of 7 nt generally accepted for bioinformatics studies[47]), the *26CEB* structure is very stable with a melting temperature (T_m) of 76.5 °C in ~100 mM K^+ solution. The independence of the melting temperature on DNA concentration, ranging from 1 to 125 μM (data not shown), implies the monomeric nature of the structure.

Folding Topology of the *26CEB* G-Quadruplex Determined by NMR. Guanine imino (H1) and aromatic (H8) protons of *26CEB* were unambiguously assigned by site-specific low-enrichment ^{15}N labeling,[36] site-specific ^{2}H labeling,[37] and through-bond correlations at natural abundance ({^{13}C–^{1}H}-HMBC) (Figure 2).[38] These unambiguous assignments were used cooperatively with other through-bond correlation experiments including TOCSY, {^{13}C–^{1}H}-HSQC, and {^{31}P–^{1}H}-HSQC to trace the H8/H6–H1′NOE sequential connectivity and to assign cross-peaks in NOESY spectra (Figures 3 and S6, Supporting Information).

The characteristic imino-H8 proton cyclic NOE connectivity patterns around G-tetrads (Figure 3A,B) established a three-layered parallel-stranded G-quadruplex structure comprising G3·G7·G19·G23, G4·G8·G20·G24, and G5·G9·G21·G25 tetrads. The assignment of the G4·G8·G20·G24 tetrad to the middle layer is consistent with imino protons of the corresponding guanines being the most protected from the exchange with solvent (Figure 2A). Four G-stretches of the

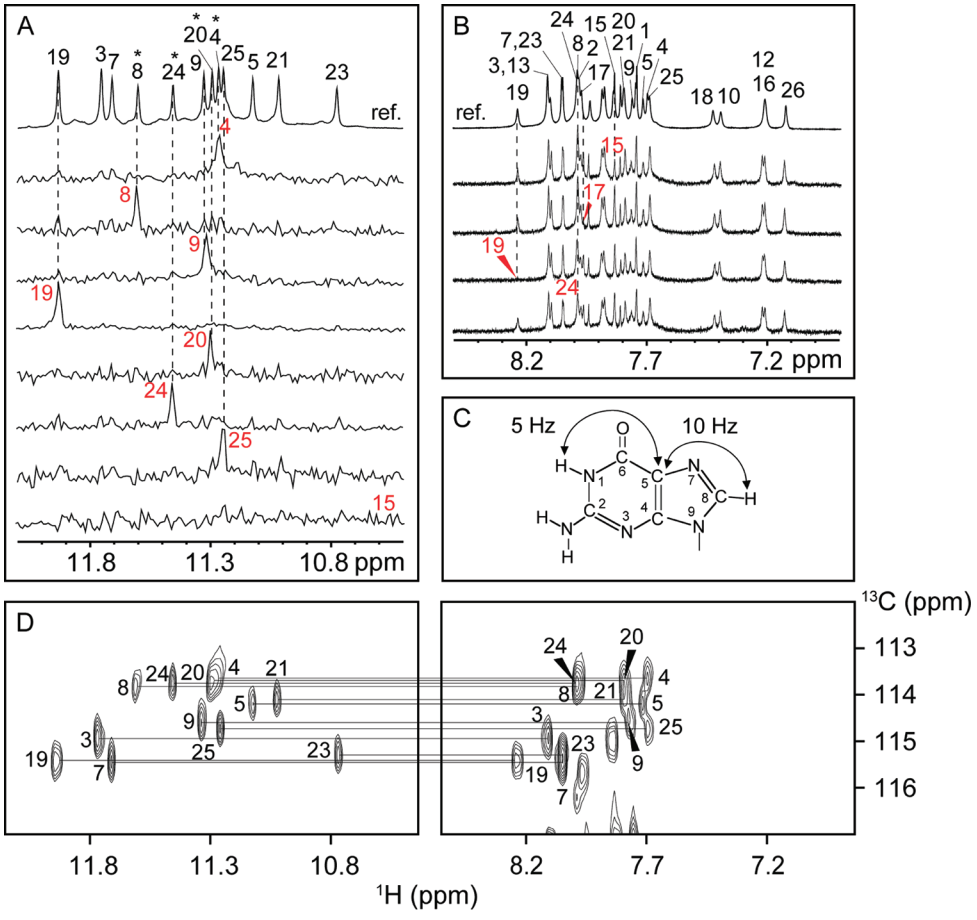


Figure 2. Imino and H8 proton assignments of *26CEB* in K^+ solution. (A) Imino proton assignments from ^{15}N-filtered spectra of samples, 2% ^{15}N-enriched at indicated positions. Peaks marked by asterisks stayed more than 15 h in ^2H_2O solvent at 25 °C. (B) H8 proton assignments by site-specific ^{2}H labeling at indicated positions. The reference spectra (ref.) of imino and aromatic protons are shown at the top. (C) Long-range *J*-couplings between imino/H8 protons and ^{13}C5. (D) Heteronuclear multi-bond correlations (HMBC) spectra at natural abundance. Assignments are labeled with residue numbers.

dx.doi.org/10.1021/ja208993r | *J. Am. Chem. Soc.* 2012, 134, 5807–5816

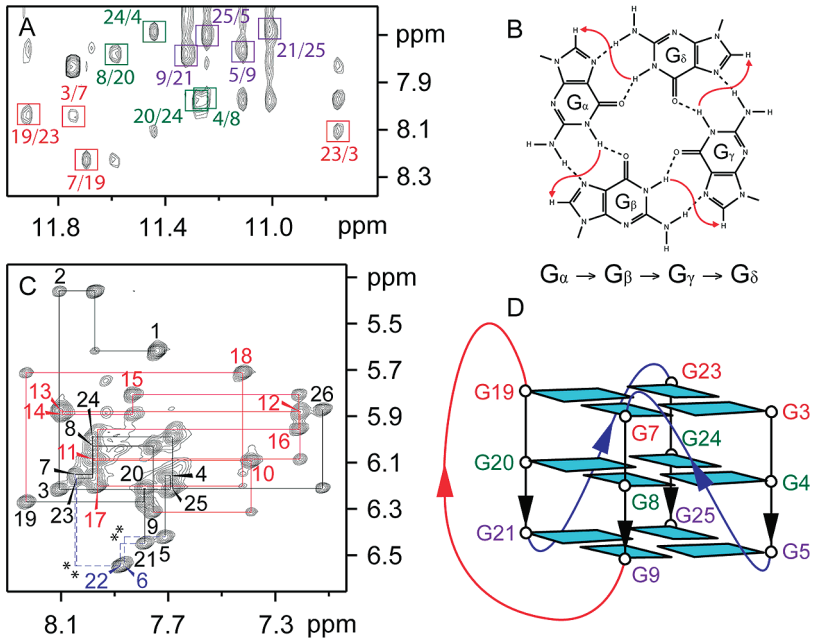


Figure 3. (A) The imino-H8 proton cyclic connectivities on NOESY spectrum (mixing time, 300 ms) of *26CEB*. The arrangements of G-tetrads were identified from framed cross-peaks with the residue number of imino protons labeled in the first position and that of H8 protons in the second position. (B) $G_{\alpha}{\cdot}G_{\beta}{\cdot}G_{\gamma}{\cdot}G_{\delta}$ tetrad showing the proximity of imino and H8 protons (red arrows). (C) The H8/6-H1′ sequential connectivities on NOESY spectrum (mixing time, 350 ms) of *26CEB*. Intraresidue H8/6-H1′ cross-peaks are labeled with residue numbers. Missing sequential connectivities are marked with asterisks. Connectivities through 1-nt and 9-nt loops are colored blue and red, respectively. (D) Schematic structure of the *26CEB* G-quadruplex satisfying the NOE connectivities shown in parts A and C. Guanines in the G-tetrad core are colored cyan. The backbones of the core, 1-nt loops, and 9-nt loop are colored black, blue, and red, respectively.

G-tetrad core, i.e., G3-G4-G5, G7-G8-G9, G19-G20-G21, and G23-G24-G25, are oriented in the same direction defining four medium-size grooves (Figure 3D). The structure has three double-chain-reversal loops: the first and third loops, each consists of a single nucleotide, while the central loop consists of nine nucleotides. The nonparticipation of guanines from the central 9-nt loop in the G-tetrad core formation was supported by NMR data of modified sequences with these residues being substituted by an inosine, in which downfield inosine imino protons could not be observed (Figure S7, Supporting Information). The intensity of intraresidue H8/6-H1′ NOE cross-peaks indicated a *syn* conformation only for the A1 residue (Figure S8, Supporting Information).

A2·T18 Watson–Crick and Potential A1·G17 Base Pairs Anchor 9-nt Double-Chain-Reversal Loop on Top of 5′ G-Tetrad. Previous studies showed the destabilizing effect of long double-chain-reversal loops on a G-quadruplex structure.[48−53] The unexpected high thermal stability (T_m = 76.5 °C) of the *26CEB* G-quadruplex, which contains a 9-nt double-chain reversal loop, suggested the occurrence of extra hydrogen-bond interactions involving this loop. Indeed, the presence of a low-field peak at 13.8 ppm (Figure 4) indicated the formation of a Watson–Crick base pair.[54] Among a number of single-residue modified sequences (Table 1), only mutations affecting A2 and T18, such as *m2T* and *m18C*, resulted in the disappearance of this peak (Figure 4), suggesting the formation of the A2·T18 Watson–Crick base pair. In addition, structure calculation suggested the possibility of pairing alignment between A1 and G17, which is consistent with a number of observed NOEs and the structure destabilization when these bases are mutated (see discussion below). Mutations of residues 10–16 resulted in nearly identical imino proton spectra to that of *26CEB* (Figure S9, Supporting Information),

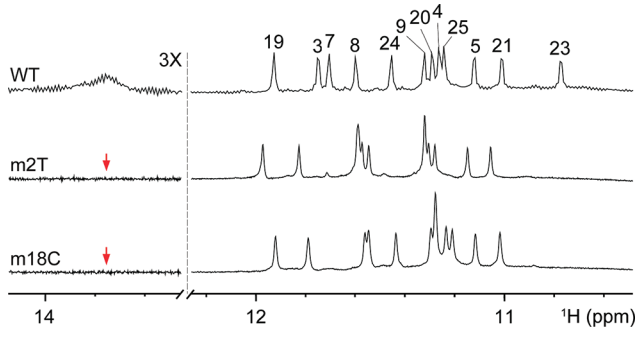


Figure 4. Identification of the A2·T18 Watson–Crick base pair. Imino proton spectra of the natural *26CEB* sequence (WT) and modified sequences, *m2T* and *m18C*. Red arrows highlight the disappearances of the imino proton peak at 13.8 ppm, characteristic of a Watson–Crick base pair.

suggesting the absence of significant interaction between these residues with the G-tetrad core. Nonetheless, the appearance of extra imino proton peaks in the spectrum of *26CEB* at low pH (Figure S10, Supporting Information), suggested that imino protons of the 9-nt loop are somewhat protected from the exchange with solvent due to some secondary structure of this loop.

Solution Structure of the *26CEB* G-Quadruplex. The G-quadruplex structure adopted by the *26CEB* sequence in K^+ solution (presented in stereoview in Figure 5) was calculated on the basis of NMR restraints derived from the buildup of several NOESY spectra (mixing times 100, 150, 200, and 350 ms) and one-bond $^{13}C-^{1}H$ RDCs (see Table 2). The structure of the parallel-stranded three-layer G-tetrad core is well-defined, while that of the long loop is less well converged (Figure 5).

dx.doi.org/10.1021/ja208993r | J. Am. Chem. Soc. 2012, 134, 5807–5816

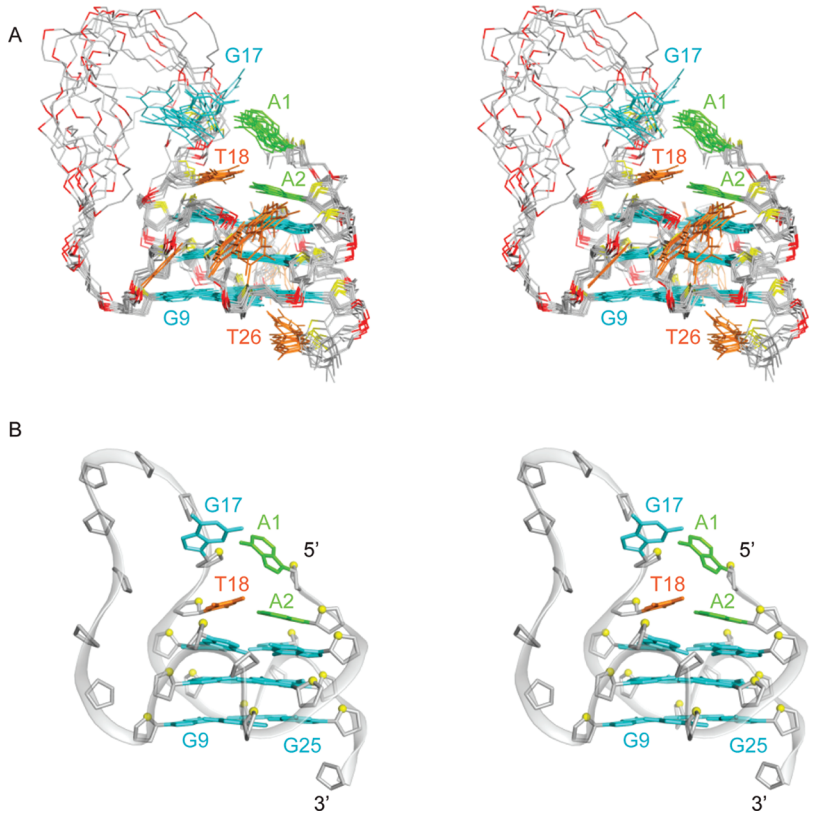


Figure 5. Stereoviews of the parallel-stranded G-quadruplex structure of *26CEB* in K^+ solution. (A) Ten superimposed structures. (B) Ribbon view of a representative structure. Guanines are colored cyan; adenines, green; thymines, orange; backbone and sugar, gray; O4′ atoms, yellow; phosphorus atoms, red. Bases of non-interacting loop residues are not shown for clarity.

Table 2. Statistics of the Computed Structures of the CEB25 Minisatellite Sequence d[AAGGGTGGGTGTAAGTGTGG-GTGGGT]

(A) NMR Restraints		
distance restraints	^2H_2O	H_2O
intraresidue	521	2
sequential (i, i + 1)	289	16
long-range (i, $\geq i$ + 2)	43	57
other restraints		
hydrogen bond	52	
dihedral angle	32	
repulsive	11	
RDC	56	
(B) Structure Statistics for 10 Molecules Following Molecular Dynamics Refinement		
NOE violations		
number (>0.2 Å)	0.2	
maximum violation (Å)	0.151 ± 0.077	
rmsd of violations (Å)	0.012 ± 0.002	
rmsd from RDC restraints (Hz)	3.345 ± 0.195	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.691 ± 0.009	
impropers (deg)	0.385 ± 0.015	
pairwise all heavy atom rmsd values (Å)		
G-tetrad core	0.68 ± 0.11	
G-tetrad core and A1, A2, G17, and T18	1.22 ± 0.30	
all residues	3.10 ± 0.83	

At the 3′ end, the terminal base T26 stacks and caps under the bottom G-tetrad. At the 5′ end, A2 and T18 form a Watson–Crick base pair, which stacks on top of the G-tetrad core (Figures 5 and 6), consistent with the observation of the T18 imino proton and NOEs between A2(H2) and T18(H6/CH_3) (Figure S11, Supporting Information), A2(H2/H8) and G7/G23(H1), as well as T18(H6/CH_3) and G3/G7/G19(H1) (data not shown). The base of A1, which is in the syn conformation, is positioned over the center of the A2·T18 pair, while its sugar is located directly above the A2 five-membered ring, in agreement with numerous NOEs from A2(H2/H8) to A1(H1′/H2′/H2″/H3′) (Figure S11, Supporting Information). Constrained by NOEs observed between G17(H8) and T18(H6/CH_3) (Figure S11, Supporting Information), G17 is staggered above T18. The proximity of A1 and G17 was further supported by the observation of a weak NOE from A1(H2) to G17(H1′) and A1(H8) to T18(CH_3) (Figure S11, Supporting Information). In some structures, A1 and G17 are in such spatial positions to form a noncanonical base pair, in which amino protons of A1 is hydrogen-bonded to G17(N3) (Figure 6A,B).

Aside from G17 and T18, little interaction was detected for other residues in the central 9-nt loop (Figure 6A; see thermodynamic analysis below). T10 points toward the G-quadruplex core, while the remaining residues are less well-converged, with G11, T12, A13, A14, G15, and T16 bases sampling multiple conformations characterized by roughly similar backbone progression geometries (Figure 5A).

Bases T6 and T22 of the first and third single-nucleotide double-chain-reversal loops protrude outward from the G-tetrad core (Figure 6D). T6 and T22 show similar NMR patterns with broken H8/H6–H1′ NOE sequential connectivities and nearly degenerated chemical shifts (Figures 3C and S6, Supporting Information).

dx.doi.org/10.1021/ja208993r | *J. Am. Chem. Soc.* 2012, 134, 5807–5816

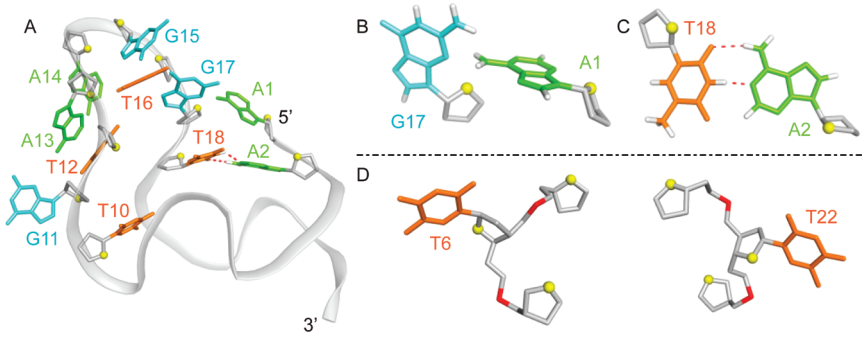


Figure 6. Structural details of the *26CEB* G-quadruplex: (A) 9-nt loop anchored by two base pairs at the 5′ end, (B) potential A1·G17 noncanonical base pair, (C) trans A2·T18 Watson–Crick base pair, and (D) single-residue double-chain-reversal loops T6 and T22.

Table 3. Thermodynamic Analysis by UV Melting and Differential Scanning Calorimetry[a]

sequence name	T_m (°C)[b]	T_m^{cal} (°C)[c]	ΔH_{cal} (kcal/mol)[d]	ΔS_{cal} (kcal/mol/K)[d]	ΔG_{cal} (37 °C) (kcal/mol)[e]
26CEB	76.5	78.9	51.7	0.147	6.13
m1T	74.0	76.5	45.5	0.130	5.20
m2T	73.5	75.5	43.6	0.122	5.80
m10C	76.0	79.2	50.7	0.144	6.06
m11T	76.8	77.5	53.5	0.156	5.14
m12A	76.5	78.8	50.2	0.147	5.87
m13T	76.7	77.4	52.6	0.149	6.41
m14T	76.5	76.6	50.0	0.142	6.00
m15T	75.9	78.3	49.9	0.141	6.20
m16A	76.3	78.6	49.7	0.141	6.00
m17T	75.0	78.8	41.1	0.117	4.83
m18C	72.0	73.5	34.1	0.097	3.80
m(1,2)T	71.5	71.5	35.3	0.102	3.50
m(11,13–15)T	76.5	78.5	49.7	0.141	6.00
m(11,13–15,17)T	75.0	77.8	46.5	0.132	5.42
m(1,2,11,13–15,17)T	70.0	70.6	33.6	0.098	3.22
m1Δ	74.0	74.1	43.4	0.120	6.25
m(1,2,26)Δ	69.5	71.3	33.0	0.095	3.55
m1Δ2T	71.5	71.5	33.6	0.097	3.53
52CEB(5′)	73.5	nd	nd	nd	nd

[a]For all parameters listed, the presumed direction is the G-quadruplex to single-strand transition (unfolding event). [b]Melting temperature deduced from the UV-melting curve. [c]Melting temperature deduced from the DSC profile. [d]Thermodynamic parameters for G-quadruplex denaturation deduced from the DSC profiles (for ΔH and ΔS, the highest relative error is 3%.). [e]$\Delta G°$ extrapolated at 310 K, from the relation $\Delta G°(T) = \Delta H° - T\Delta S°$, assuming temperature-independent enthalpies.

Thermodynamic Analysis by DSC and UV-Melting Experiments. To gain insight about energetic contributions of tetrads and base pairs on the *26CEB* structure, the melting temperature (T_m) and unfolding enthalpy (ΔH_{cal}) of *26CEB* and various mutants were measured by means of UV and DSC spectroscopy (Table 3 and Figure 7). All structural transitions were reversible throughout the experiments for scanning speeds of 1 °C/min or lower, indicating that the denaturation curves correspond to true equilibrium processes. For all mutants studied, NMR spectra (e.g., see Figure S9, Supporting Information) verified that the general folding of the structure remained unchanged following the mutations, ensuring that the folding topology is robust and the observed effect is specific to the altered region. The natural sequence *26CEB*, characterized by a T_m of 76.5 °C and ΔH_{cal} of 51.7 kcal/mol, served as the reference.

Mutations that affect specifically the A1·G17 base pair were found to destabilize the structure. The removal of the first residue A1 in mutant *m1Δ* decreased T_m to 74 °C ($\Delta T_m = -2.5$ °C) and ΔH_{cal} to 43.4 kcal/mol ($\Delta\Delta H_{cal} = -8.3$ kcal/mol). Likewise, the A-to-T modification of this residue in *m1T* lowered T_m and ΔH_{cal} to 74 °C ($\Delta T_m = -2.5$ °C) and 45.5 kcal/mol ($\Delta\Delta H_{cal} = -6.2$ kcal/mol), respectively. A reduction of T_m and ΔH_{cal} to 75 °C ($\Delta T_m = -1.5$ °C) and 41.1 kcal/mol ($\Delta\Delta H_{cal} = -10.6$ kcal/mol), respectively, was also recorded for *m17T*, indicating unfavorable pairing between A1 and T17 due to the spatial arrangement of both residues (Figure 6B). Mutations that disrupt the A2·T18 base pair (Figure 4) also destabilize the structure: *m2T* decreased T_m to 73.5 °C ($\Delta T_m = -3$ °C) and ΔH_{cal} to 43.6 kcal/mol ($\Delta\Delta H_{cal} = -8.1$ kcal/mol), whereas *m18C* produced a remarkable decrease of T_m to 72 °C ($\Delta T_m = -4.5$ °C) and ΔH_{cal} to 34.1 kcal/mol ($\Delta\Delta H_{cal} = -17.6$ kcal/mol). Mutations disrupting both base pairs, *m(1,2)T* and *m1Δ2T*, produced T_m of 71.5 °C ($\Delta T_m = -5$ °C) and comparable ΔH_{cal} of 35.3 kcal/mol ($\Delta\Delta H_{cal} = -16.4$ kcal/mol) and 33.6 kcal/mol ($\Delta\Delta H_{cal} = -18.1$ kcal/mol), respectively. These decreases are relatively close to those for *m1T* and *m2T* combined, i.e. $\Delta T_m = -5.5$ °C and $\Delta\Delta H_{cal} = -14.3$ kcal/mol.

The energetic contribution of the tetrads could be best evaluated from the thermodynamic stability of *m(1,2,11,13–15,17)T* whereby all but guanines participating in the G-tetrad core

dx.doi.org/10.1021/ja208993r | J. Am. Chem. Soc. 2012, 134, 5807–5816

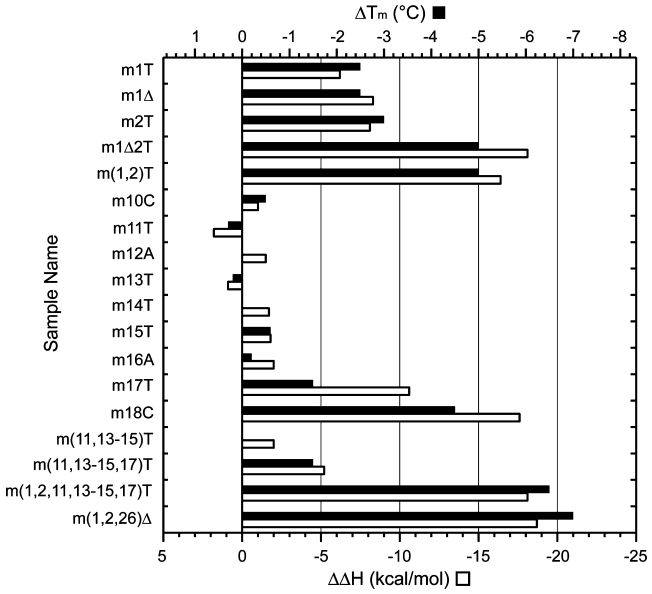


Figure 7. Histogram representation of changes in the melting temperature (filled bars) and unfolding enthalpy (open bars) values observed for various mutants with respect to *26CEB*. Most

formation were replaced by thymines. The mutant was characterized with T_m of 70 °C (ΔT_m = −6.5 °C) and ΔH_{cal} of 33.6 kcal/mol ($\Delta\Delta H_{cal}$ = −18.1 kcal/mol), rather similar values to those of *m(1,2)T* or *m1Δ2T*. Taken together, A2·T18 and A1·G17 stabilize the G-quadruplex by 6 ± 1 °C in T_m and by 17 ± 1 kcal/mol in ΔH_{cal}.

Mutations of other loop residues, including single-residue modifications (*m10C*, *m11T*, *m12A*, *m13T*, *m14T*, *m15T*, and *m16A*) and the multiple-residue modification *m(11,13−15)T* only marginally affect the thermodynamic stability of the structure ($|\Delta T_m| < 0.5$ °C and $|\Delta\Delta H_{cal}| < 2.0$ kcal/mol), while the effect observed for *m(11,13−15,17)T* (ΔT_m = −1.5 °C and $\Delta\Delta H_{cal}$ = −5.2 kcal/mol) could be explained by the disruption of the A1·G17 base pair.

Formation of the *26CEB* G-Quadruplex in Longer Sequence Contexts. To test whether the structure observed for the 26-nt G-rich *26CEB* fragment is formed within longer sequence contexts, we recorded NMR spectra of extended sequences comprising several 52-nt full-repeating units. Imino proton spectra of the 78- and 130-nt sequences *78CEB(5′,3′)* and *130CEB(5′,mid,3′)* (Table 1) displayed a similar pattern to that of *26CEB* (Figure 8A), indicating that the *26CEB* segments in these sequences form the same G-quadruplex structure as that of an isolated *26CEB*, while the remaining parts are essentially unstructured. The spectral broadening and minor shifts of some peaks in *78CEB(5′,3′)* and *130CEB(5′,mid,3′)* as compared to those of *26CEB* are due to larger molecular sizes and different 5′ or 3′ flanking sequences. The imino proton spectrum of the *78CEB(5′,3′)* sequence containing two *26CEB* segments (Figure 8A) closely resembles the calculated weighted sum of the two spectra of the *78CEB(5′mut,3′)* and *78CEB(5′,3′mut)* sequences (Figure 8B) containing mutations that disrupt G-quadruplex formation at the 5′ and 3′ side, respectively, supporting the formation of two G-quadruplex blocks connected by a 26-nt linker in the *78CEB(5′,3′)* sequence. Formation of G-quadruplex blocks at different locations of long CEB25 sequences was also supported by NMR data for sequences

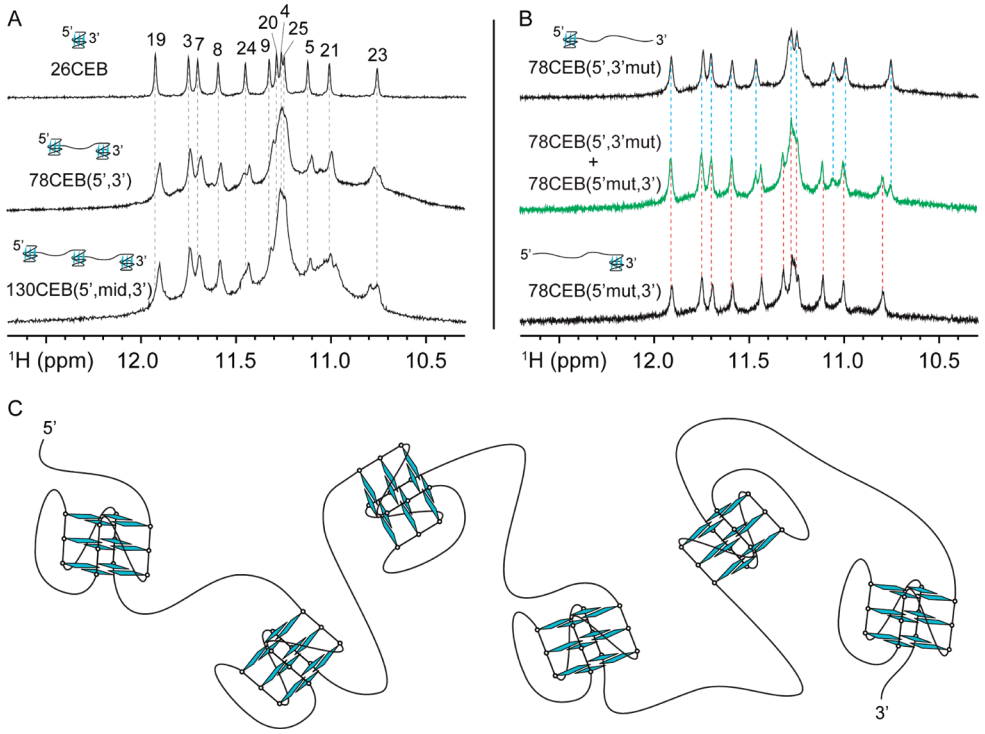


Figure 8. (A) Imino proton spectra of CEB25 minisatellite sequences of various lengths. Proposed folding topologies are shown by the schematics with G-tetrads being highlighted in cyan. (B) Imino proton spectra of 78-nt CEB25 minisatellite sequences containing mutations that disrupt G-quadruplex formation at the 3′ side (top) and 5′ side (bottom), respectively, as well as their weighted sum (middle) that simulates the spectrum of the *78CEB(5′,3′)* sequence containing two G-quadruplex blocks. Note the similarity between the latter and the observed spectrum of *78CEB(5′,3′)* in part A. (C) CEB25 minisatellite pearl-necklace model.

dx.doi.org/10.1021/ja208993r | *J. Am. Chem. Soc.* 2012, 134, 5807−5816

containing a single inosine substitution in one G-quadruplex block (Figure S12, Supporting Information) and other long sequences (Figure S13, Supporting Information). On the basis of these results, a "*pearl-necklace*" model of stable G-quadruplex blocks interconnected by non-quadruplex-forming sequences on a long single-stranded CEB25 repeats is proposed (Figure 8C).

DISCUSSION

Over the past two decades, many different G-quadruplexes have been described in telomeres and oncogenic promoters.[2] In this work, we have shown that the G-rich sequence of the CEB25 minisatellite forms a propeller-type parallel-stranded G-quadruplex involving two single-nucleotide and one 9-nt double-chain-reversal loops. There are additional base pairs formed between bases of the long loop and those at the 5′-end, anchoring this loop to the top of the G-tetrad core.

Previously, an inverse correlation between the length of a propeller loop T_n in a G-quadruplex and the stability of the structure has been described: each added base leads to a 2 °C drop in T_m or 0.3 kcal/mol lost in ΔG.[52,53] The G-quadruplex in CEB25 minisatellite, consisting of a 9-nt loop, is more stable than expected from previous studies. Our thermodynamic analysis have shown that the two base pairs between the long loop and the 5′-end sequence contribute to the stability of the overall structure by about 17 kcal/mol in enthalpy or 6 °C in melting temperature, highlighting the importance of the loop and flanking sequences in stabilizing G-quadruplex folds.

The structure of the CEB25 G-quadruplex can be readily compared with the form I of the *c-kit2* promoter sequence.[55] A noncanonical C1·A13 base pair formed in the *c-kit2* structure between the last base of a 5-nt central double-chain-reversal loop and a 5′-end flanking base is comparable to the A2·T18 base pair in the CEB25 structure, suggesting a general rule for the formation of a base pair between the last residues of a double-chain-reversal loop and the 5′-end flanking bases. However, there is no equivalence of the second base pair A1·G17 of the CEB25 structure in the *c-kit2* G-quadruplex. It will be interesting to address the question regarding the context (loop length and sequence) necessary for the formation of two or more base pairs.

In contrast to telomeric G-quadruplexes,[4] the CEB25 G-quadruplex has been observed to be monomorphic and robust. This was evident from the observation of only one major conformation identified following a number of modifications and extensions made on the sequence. It is likely that the first and third single-residue loops drive the whole strand to adopt the propeller-type parallel G-quadruplex independent of the size and sequence of the flanking and the central loop segments, consistent with the previous observation that a single-residue linker between G-tracts favors a double-chain-reversal loop.[47−52]

The "*beads-on-a-string*" formation of G-quadruplexes on a long telomeric sequence has been suggested earlier.[9,56,57] Consecutive G-quadruplexes have been observed to be separated rather irregularly from each other by AFM,[58] and each four-repeat fragment can adopt multiple conformations.[4] On the contrary, the proposed pearl-necklace model for human CEB25 minisatellite repeats is based on the observation of a single G-quadruplex conformation by NMR. Arrays of G-quadruplexes might be formed regularly in CEB25 repeats interlinked by essentially unstructured strings and thus provide higher-ordered structure and biological functions which can be specific to tandem arrays compared to an isolated G-quadruplex. For example, a high local concentration of intramolecular G-quadruplexes might provide additive or synergistic probability of G-quadruplex formation and/or interference in resolution influencing the array stability dependent on replication.[13,59]

CONCLUSION

We have solved the NMR structure of a three-layered G-quadruplex structure formed by the G-rich fragment of each CEB25 minisatellite repeat in K^+ solution. This is a propeller-type parallel-stranded G-quadruplex involving a 9-nt central double-chain-reversal loop. This long loop is anchored to the 5′ end of the sequence by an A·T Watson−Crick base pair and a potential G·A noncanonical base pair. Thermodynamic and mutation analyses have established the role of these base pairs in the stability of the G-quadruplex structure. We show that such a structure can be formed in long sequence contexts, providing evidence for a pearl-necklace of G-quadruplexes formed by single-stranded CEB25 minisatellite.

ASSOCIATED CONTENT

Supporting Information

Additional experimental data (Table S1 and Figures S1−S13) and complete ref 25. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

phantuan@ntu.edu.sg; jean-louis.mergny@inserm.fr

Author Contributions

‖These authors contributed equally to this work

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Mr. Kah Wai Lim and Dr. Serge Bouaziz for their initial help with the structure calculation. The research in ATP group was supported by Singapore Ministry of Education grants (RG62/07 and RG72/10). M.A. is supported by the Yousef Jameel scholarship. J.L.M acknowledges support from the Association pour la recherche sur le cancer (programme libre A.R.C.), the Région Aquitaine, the Fondation pour la Recherche Médicale (F.R.M.), INCa and ANR grants (F-DNA and G4-TOOLBOX). The research in A.N.'s team is supported by grants from the Ligue Nationale contre le Cancer (LNCC; Equipe Labellisée EL2007.LNCC/AN and EL2010.LNCC/AN) and a postdoctoral fellowship to A.S. from the F.R.M.

REFERENCES

(1) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668.
(2) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429.
(3) Neidle, S. *Curr. Opin. Struct. Biol.* **2009**, *19*, 239.
(4) Phan, A. T. *FEBS J.* **2010**, *277*, 1107.
(5) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Natl. Chem. Biol.* **2005**, *1*, 167.
(6) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 4386.
(7) Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824.
(8) Yue, D. J.; Lim, K. W.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 11462.
(9) Yu, H. Q.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2006**, *128*, 15461.
(10) Maizels, N. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055.
(11) Lipps, H. J.; Rhodes, D. *Trends Cell Biol.* **2009**, *19*, 414.

(12) Paeschke, K.; Capra, J. A.; Zakian, V. A. *Cell* **2011**, *145*, 678.
(13) Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M. P.; Foiani, M.; Nicolas, A. *EMBO J.* **2011**, *30*, 4033.
(14) Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N. *Genes Dev.* **2004**, *18*, 1618.
(15) Cahoon, L. A.; Seifert, H. S. *Science* **2009**, *325*, 764.
(16) Vergnaud, G.; Denoeud, F. *Genome Res.* **2000**, *10*, 899.
(17) Bois, P.; Jeffreys, A. J. *Cell. Mol. Life Sci.* **1999**, *55*, 1636.
(18) Kashi, Y.; King, D. G. *Trends Genet.* **2006**, *22*, 253.
(19) Verstrepen, K. J.; Jansen, A.; Lewitter, F.; Fink, G. R. *Nat. Genet.* **2005**, *37*, 986.
(20) Lalioti, M. D.; Scott, H. S.; Antonarakis, S. E. *Nat. Genet.* **1997**, *17*, 17.
(21) Benett, I. *Diabet. Med.* **1995**, *12*, 452.
(22) Klenova, E.; Scott, A. C.; Roberts, J.; Shamsuddin, S.; Lovejoy, E. A.; Bergmann, S.; Bubb, V. J.; Royer, H. D.; Quinn, J. P. *J. Neurosci.* **2004**, *24*, 5966.
(23) Phelan, C. M.; Rebbeck, T. R.; Weber, B. L.; Devilee, P.; Ruttledge, M. H.; Lynch, H. T.; Lenoir, G. M.; Stratton, M. R.; Easton, D. F.; Ponder, B. A. J.; CannonAlbright, L.; Larsson, C.; Goldgar, D. E.; Narod, S. A. *Nat. Genet.* **1996**, *12*, 309.
(24) Jeffreys, A. J.; Barber, R.; Bois, P.; Buard, J.; Dubrova, Y. E.; Grant, G.; Hollies, C. R.; May, C. A.; Neumann, R.; Panayi, M.; Ritchie, A. E.; Shone, A. C.; Signer, E.; Stead, J. D.; Tamaki, K. *Electrophoresis* **1999**, *20*, 1665.
(25) Law, M. J.; et al. *Cell* **2010**, *143*, 367.
(26) Panigrahi, G. B.; Lau, R.; Montgomery, S. E.; Leonard, M. R.; Pearson, C. E. *Nat. Struct. Mol. Biol.* **2005**, *12*, 654.
(27) Amrane, S.; Saccà, B.; Mills, M.; Chauhan, M.; Klump, H. H.; Mergny, J. L. *Nucleic Acids Res.* **2005**, *33*, 4065.
(28) Amrane, S.; Mergny, J. L. *Biochimie* **2006**, *88*, 1125.
(29) Yauk, C. L.; Dubrova, Y. E.; Grant, G. R.; Jeffreys, A. J. *Mutat. Res.* **2002**, *500*, 147.
(30) Napierala, M.; Bacolla, A.; Wells, R. D. *J. Biol. Chem.* **2005**, *280*, 37366.
(31) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059.
(32) Jeffreys, A. J.; Wilson, V.; Neumann, R.; Keyte, J. *Nucleic Acids Res.* **1988**, *16*, 10953.
(33) Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74.
(34) Mergny, J. L.; Lacroix, L. *Oligonucleotides* **2003**, *13*, 515.
(35) Mergny, J. L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, e138.
(36) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160.
(37) Huang, X.; Yu, P.; LeProust, E.; Gao, X. *Nucleic Acids Res.* **1997**, *25*, 4758.
(38) Phan, A. T. *J. Biomol. NMR* **2000**, *16*, 175.
(39) Phan, A. T.; Guéron, M.; Leroy, J. L. *Methods Enzymol.* **2001**, *338*, 341.
(40) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
(41) Myers, S.; Bowden, R.; Tumian, A.; Bontrop, R. E.; Freeman, C.; MacFie, T. S.; McVean, G.; Donnelly, P. *Science* **2010**, *327*, 876.
(42) Baudat, F.; Buard, J.; Grey, C.; Fledel-Alon, A.; Ober, C.; Przeworski, M.; Coop, G.; de Massy, B. *Science* **2010**, *327*, 836.
(43) Berg, I. L.; Neumann, R.; Lam, K. W.; Sarbajna, S.; Odenthal-Hesse, L.; May, C. A.; Jeffreys, A. J. *Nat. Genet.* **2010**, *42*, 859.
(44) Buard, J.; Bourdet, A.; Yardley, J.; Dubrova, Y.; Jeffreys, A. J. *EMBO J.* **1998**, *17*, 3495.
(45) Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225.
(46) Gray, D. M.; Wen, J. D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J. *Chirality* **2008**, *20*, 431.
(47) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(48) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710.
(49) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405.
(50) Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2007**, *35*, 4214.
(51) Bugaut, A.; Balasubramanian, S. *Biochemistry* **2008**, *47*, 689.
(52) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J. L. *Biochimie* **2008**, *90*, 686.
(53) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J. L. *Nucleic Acids Res.* **2010**, *38*, 7858.
(54) Patel, D. J.; Tonelli, A. E. *Biopolymers* **1974**, *13*, 1943.
(55) Kuryavyi, V.; Phan, A. T.; Patel, D. J. *Nucleic Acids Res.* **2010**, *38*, 6757.
(56) Xu, Y.; Ishizuka, T.; Kurabayashi, K.; Komiyama, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 7833.
(57) Martadinata, H.; Heddi, B.; Lim, K. W.; Phan, A. T. *Biochemistry* **2011**, *50*, 6455.
(58) Wang, H.; Nora, G. J.; Ghodke, H.; Opresko, P. L. *J. Biol. Chem.* **2011**, *286*, 7479.
(59) Ribeyre, C.; Lopes, J.; Boule, J. B.; Piazza, A.; Guédin, A.; Zakian, V. A.; Mergny, J. L.; Nicolas, A. *PLoS Genet.* **2009**, *5*, e1000475.