Supplementary Information

Intramolecular Folding in Three Tandem Guanine Repeats of Human Telomeric DNA

Deepak Koirala[a], Tomoko Mashimo[b], Yuta Sannohe[b], Zhongbo Yu[a], Hanbin Mao*[a], and Hiroshi Sugiyama*[b, c, d]

[a] *Department of Chemistry & Biochemistry, Kent State University, Kent, OH 44242, USA.*
Email: hmao@kent.edu
[b] *Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan.*
[c] *Institute for Integrated Cell Material Sciences (iCeMS), Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan.*
[d] *CREST, Japan Science and Technology Corporation (JST), Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan.*
Email: hs@kuchem.kyoto-u.ac.jp

* To whom correspondence should be addressed.

CD Measurements: Desired concentrations of oligonucleotides samples were prepared in 10 mM Tris buffer (pH 7.4) containing 100 mM NaCl. The samples were then heated at 95°C for 10 min, rapidly cooled in an ice bath and incubated at room temperature for up to four hours. The fast cooling was performed to allow the comparison of the results with the fast mechanical unfolding-refolding processes in the single-molecule assay. CD spectra were taken in a 1 mm quartz cuvette at room temperature with a Jasco-810 spectropolarimeter (Easton, MD). The reported spectra were the average of three scans with a scanning rate of 100 nm / min. Spectrum of the corresponding buffer was subtracted from each scan of the sample for baseline correction and smoothed using a Savitzky-Golay function. The CD spectra of the sample, where appropriate, were also taken at different temperatures ranging from 25°C to 85°C by controlling the temperature using a Jasco (model PFD-425S) peltier temperature controller.

DNA Construct Preparation: DNA construct for single-molecule mechanical unfolding studies was prepared by incorporating Tel-3G (or Tel-4G) fragment between two dsDNA handles. One of the DNA handles (2028 bp) was labeled with biotin, which was introduced through a biotinylated primer (Integrated DNA Technologies, IDT, Coralville, IA) during PCR amplification using pBR322 plasmid as the template, followed by digestion with XbaI restriction enzyme (New England Biolab, NEB). Another handle (2690 bp) was obtained by SacI and EagI digestions of a pEGFP plasmid (Clontech, Mountain View, CA) and gel purified using a kit (Midsci, St. Louis, MO). This handle was subsequently labeled by digoxigenin (Dig) at the 3′ end using 18 μM Dig-dUTP (Roche, Indianapolis, IN) and terminal transferase (Fermentas, Glen Burnie, MD). The middle fragment containing Tel-3G (or Tel-4G) was constructed by annealing an oligonucleotide, 5'-CTA GAC GGT GTG AAA TAC CGC ACA GAT GCG **TTA GGG TTA GGG TTA GGG TTA** GCC AGC AAG ACG TAG CCC AGC GCG TC (or 5'-CTA GAC GGT GTG AAA TAC CGC ACAGAT GCG **TTA GGG TTA GGG TTA GGG TTA GGG TTA** GCC AGC AAG ACG TAG CCC AGC GCG TC) with two other oligonucleotides, 5′-CGC ATC TGT GCG GTA TTT CAC ACC GT and 5′-GGC CGA CGC GCT GGG CTA CGT CTT GCT GGC at 95 °C for 5 min and slowly cooled to room temperature for 6 h. This fragment (or the fragment containing Tel-4G) was ligated with the 2028 bp DNA handle at one end, followed by a second ligation with the 2690 bp DNA handle using T4 DNA ligase (NEB). The final construct was purified by ethanol precipitation. The DNA pellet was dissolved in water and stored at -80°C.

Optical Tweezers Single-Molecule Assay : The detailed description of the laser tweezers instrument has been reported elsewhere.[1] For the single-molecule experiment, anti-Dig antibody-coated polystyrene beads (diameter: 2.17 μm, Spherotech, Lake Forest, IL) were incubated with diluted DNA construct obtained above (~1 ng/μL) in the 10 mM Tris buffer (pH 7.4) containing 100 mM NaCl for 1 h at 23 °C to attach the DNA construct via the Dig/anti-Dig linkage. Beads coated with streptavidin (diameter: 0.97 μm, Bangs Laboratories Inc., Fishers, IN) were dispersed into the same buffer before being injected into the microfluidic chamber. These two types of beads were trapped separately using two laser traps. To immobilize the DNA construct between the two beads, the bead already attached with DNA construct was brought close to the bead coated with streptavidin by a Nano-MTA steerable mirror (Madcity Labs, Inc., Madison, WI) in the laser tweezers instrument.[1] Once the DNA tether was trapped between the two beads, the steerable mirror that controls the anti-Dig-coated bead was moved away from the streptavidin-coated bead with a load rate of ~5.5 pN/s. The secondary structure formed in the DNA molecule was unfolded when tension inside the tether was gradually increased. Unfolding

events with sudden change in the end-to-end distance of the DNA tether were observed during the process. Single tether was confirmed by a single breakage event when the DNA was overstretched. Change in contour length (ΔL) due to the unfolding events was calculated by the two data points flanking the rupture events using an extensible worm-like chain (WLC) model[2] in equation 1:

$$\frac{x}{L} = 1 - \frac{1}{2}\left(\frac{k_B T}{FP}\right)^{\frac{1}{2}} + \frac{F}{S} \quad \text{.....(1)}$$

where x is the end-to-end distance of the DNA tether, k_B is the Boltzmann constant, T is absolute temperature, P is the persistent length (51.95 nm)[2], F is force, and S is the elastic stretch modulus (1226 pN).[2]

Molecular Dynamics Simulation: The MD simulation was carried out using the sequence, 5′-A GGG TTA GGG TTA GGG TTA GGG TTA-3′. The initial atomic coordinates of triplexes were obtained from a basket G-quadruplex structure (PDB 143D).[3] Triplex models were constructed by deletion of the fourth strand from the antiparallel structure. The triplex sequence is 5′-$A^1G^2G^3G^4T^5T^6A^7G^8G^9G^{10}T^{11}T^{12}A^{13}G^{14}G^{15}G^{16}T^{17}$-3′. The systems were neutralized with 16 Na^+ counter ions (represented by blue asterisks in Figure 3b). The central sodium ion (represented by a blue ball in Figure 3b) was positioned at 5′-G^3(syn)G^4(anti)-3′/5′-G^8(syn)G^9(anti)-3′/5′-G(syn)^{15}G(anti)^{16}-3′. Classical MD simulations were performed using AMBER version 8 and 9 for the MD-GRAPE system. All nonbonded interactions, van der Waals and Coulomb forces, and energies were calculated using the MD-Grape-3. An AMBER parm99 was applied as a force field for DNA. The time step was set to be 1 fs. The systems were surrounded spherically by TIP3P water molecules. The circle dimensions were chosen to achieve a minimum distance of 38 Å from G3, resulting a typical fundamental cell of 38 × 38 × 38 $Å^3$ with about 6849 water molecules. The systems were equilibrated for 300 ps with gradual removal of positional restraints on the DNA with the following protocol for the triplex: (i) minimization of water and counter ions for the triplex; (ii) minimization of DNA with the restrain energy of 50 kcal/mol $Å^{-1}$ for the triplex; (iii) minimization of DNA with the restrain energy of 10 kcal/mol $Å^{-1}$; (iv) minimization of DNA with the restrain energy of 5 kcal/mol $Å^{-1}$; (v) 150 ps MD (T = 300 K) holding DNA fixed (5000 kcal/mol $Å^{-1}$); (vi) 50 ps MD (T = 300K) holding DNA fixed (50 kcal/mol $Å^{-1}$); (vii) 50 ps MD (T=300K) holding DNA fixed (10 kcal/mol $Å^{-1}$); (viii) 50 ps MD (T = 300K) holding DNA fixed (5 kcal/mol $Å^{-1}$); and (ix) MD simulation for 3 ns. We examined the average structure during equilibrium calculation (Figure 3) between 2500-3000 ps (Fig. 3a). The distance between O of 5′ end A^1G^2 and O of 3′ end $G^{16}T^{17}$ backbone (sugar and phosphate) was found to be 1.5 ± 0.1 nm. This distance was used as the end-to-end distance of the folded structure in the subsequent calculation of the expected change in contour length of unfolding.

Deconvolution of Populations: In the case of mechanical unfolding of Tel-4G, two types of the folded populations were observed in the histogram of either unfolding force or change in contour length, ΔL (Figure 4). To assign individual force-extension curves to a specific species, we fit the overall distribution using a two-peak Gaussian function. To account for the stochastic behavior of individual molecules, we randomly assigned the population in the intersection of the two Gaussian peaks to a specific species according to the ratio determined by the two peak

Gaussian fitting. After deconvolution, the ΔL and unfolding force for each population can be obtained separately.

Calculation of Change in Contour Length (ΔL): The change in contour length (ΔL) for the unfolding of a structure having end-to-end distance of x can be calculated with the following equation,

$$\Delta L = N \times L_{nt} - x \quad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

Where N and L_{nt} are the number of nucleotides involved in the structure and contour length per nucleotide, respectively. We used L_{nt}= 0.44 nm as reported in the literature.[4]

To calculate expected ΔL for the G-quadruplex in the Tel-4G, we used x = 1.8 ± 0.1 nm from the NMR structure of a basket type quadruplex (PDB 143D)[3] and N = 21 nucleotides. Using equation 2, this gave ΔL = 7.4 ± 0.1 nm. This value was similar to the experimentally observed ΔL (7.8 ± 0.2 nm) of the major population in Tel-4G. To calculate ΔL of the simulated triplex, the end-end distance from the simulated structure was used (1.5 ± 0.1 nm). This gave the expected ΔL of 5.1 ± 0.1 nm, which matches exactly with those observed for the Tel-3G and partially folded species in the Tel-4G (Table I).

Calculation of ΔG_{unfold}: The stretching and relaxing force extension curves were used to calculate the change in free energy (ΔG_{unfold}) associated with the unfolding of a species according to Jarzynski's equation (equation 3) for non-equilibrium systems[5],

$$\Delta G = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp\left(-\frac{W_i}{k_B T} \right) \ldots\ldots\ldots\ldots\ldots (3)$$

where N is the number pulling experiments and W is the non-equilibrium work done during unfolding of the folded structure, which equals to the hysteresis area between stretching and relaxing force-extension curves.[6]

Percentage Formation of Folded Species: The percentage formation of folded species was calculated as the ratio of the number of the pulling curves with unfolding events to the total pulling curves recorded. To avoid repetitive counting, the subsequent pulling curves from the same molecule were not considered in this calculation. According to this calculation, the percentage formation of the minor species (Triplex) and the major species (Quadruplex) in Tel-4G was found to be ~15% and ~45% respectively.

DMS Footprinting: Details of the DMS foot-printing method has been reported elsewhere.[7] To describe briefly, Tel-3G or Tel-4G oligonucleotides were radio-labeled at the 5' end by incubating the DNA with T4 polynucleotide kinase (NEB) and [γ-32P] ATP (Perkin Elmer). The labeled products were then purified using MicroSpin™ G-25 Columns (GE, Piscataway, NJ). Radiolabeled oligonucleotide samples were mixed with cold oligonucleotides to obtain a final concentration of 1 µM in 10 mM Tris–HCl buffer, pH 7.4 containing 100 mM NaCl. An aliquot of 100 µl mixed samples were then heated at 97 °C for 10 min and cooled to 25 °C quickly using an ice bath. The samples were then treated with DMS with a final concentration of 1% for 45 seconds. The reaction was stopped with 1.1 ml of stop buffer (2 M β-mercaptoethanol, 300 mM

sodium-acetate, 250 mg/ml salmon sperm DNA) and immediately ethanol precipitated followed by washing with 70% ethanol. The DNA pellet was then dried and cleaved using piperidine followed by separation of the fragments on a 10% denaturing acrylamide (19:1 acrylamide: bisacrylamide) gel. The gel was dried on Whatman paper, exposed to a phosphorimager screen, and scanned with a Typhoon 8600 instrument (Molecular Dynamics). DMS footprinting gel images were quantified using BioRad QuantityOne^{TM} software. Each guanine band was normalized by comparing the intensity of each band to the total intensity of the corresponding lane. Then the fold protection of each guanine residues was calculated as the ratio of the normalized intensity of a particular band in the presence of salt (100 mM Na^{+}) to the corresponding band in the absence of the salt.

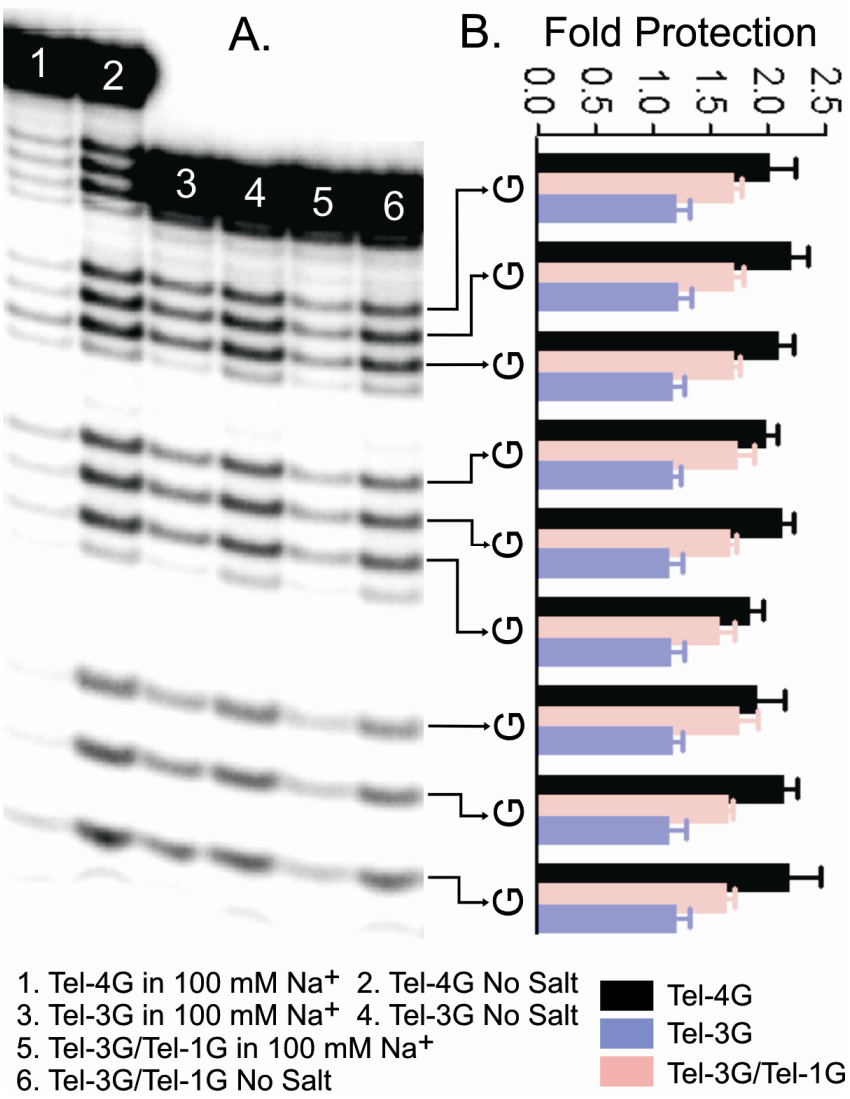

Figure S1: *DMS footprinting of Tel-3G, Tel-3G/Tel-1G mixture, and Tel-4G. (A) The gel image of the footprinting. The content of each lane is labeled below the image. (B) The fold protection of the guanine residues in the Tel-3G, Tel-3G/Tel-1G mixture, and Tel-4G in buffer containing 100 mM Na^{+} with respect to the corresponding species in a pure buffer.*

Supplementary References:

1 H. Mao and P. Luchette, *Sensors and Actuators B,* 2008, **129**, 764-771.

2 C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proc. Natl. Acad. Sci. USA,* 1997, **94**, 6185-6190.

3 Y. Wang and D. J. Patel, *Structure,* 1993, **1**, 263-282.

4 J. B. Mills, E. Vacano and P. J. Hagerman, *J. Mol. Biol.,* 1999, **285**, 245-257; T. A. Laurence, X. Kong, M. Jager and S. Weiss, *Proc. Nat. Acad. Sci. USA,* 2005, **102**, 17348-17353; M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag and S. M. Block, *Proc. Natl. Acad. Sci. USA,* 2006, **103**, 6190-6195; Z. Yu, J. D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu and H. Mao, *J. Am. Chem. Soc.,* 2009, **131**, 1876-1882.

5 C. Jarzynski, *Phys. Rev. Lett.,* 1997, **78**, 2690 - 2693; J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr. and C. Bustamante, *Science,* 2002, **296**, 1832-1835.

6 W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside and S. M. Block, *Science,* 2008, **319**, 630-633.

7 J. D. Schonhoft, R. Bajracharya, S. Dhakal, Z. Yu, H. Mao and S. Basu, *Nucleic Acids Res.,* 2009, **37**, 3310-3320.