Extreme conformational diversity in human telomeric DNA

J. Y. Lee^{*†}, Burak Okumus^{†‡}, D. S. Kim^{*}, and Taekjip Ha^{*§¶||}

^{*}Department of Physics, Seoul National University, Seoul 151-742, Korea; and [†]Center for Biophysics and Computational Biology, [§]Department of Physics, and [¶]Howard Hughes Medical Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Edited by Kiyoshi Mizuuchi, National Institutes of Health, Bethesda, MD, and approved October 28, 2005 (received for review July 20, 2005)

DNA with tandem repeats of guanines folds into G-quadruplexes made of a stack of G-quartets. *In vitro*, G-quadruplex formation inhibits telomere extension, and POT1 binding to the single-stranded telomeric DNA enhances telomerase activity by disrupting the G-quadruplex structure, highlighting the potential importance of the G-quadruplex structure in regulating telomere length *in vivo*. We have used single-molecule spectroscopy to probe the dynamics of human telomeric DNA. Three conformations were observed in potassium solution, one unfolded and two folded, and each conformation could be further divided into two species, long-lived and short-lived, based on lifetimes of minutes vs. seconds. Vesicle encapsulation studies suggest that the total of six states detected here is intrinsic to the DNA. Folding was severely hindered by replacing a single guanine, showing only the short-lived species. The long-lived folded states are dominant in physiologically relevant conditions and probably correspond to the parallel and antiparallel G-quadruplexes seen in high-resolution structural studies. Although rare under these conditions, the short-lived species determine the overall dynamics because they bridge the different long-lived species. We propose that these previously unobserved transient states represent the early and late intermediates toward the formation of stable G-quadruplexes. The major compaction occurs between the early and late intermediates, and it is possible that local rearrangements are sufficient in locking the late intermediates into the stably folded forms. The extremely diverse conformations of the human telomeric DNA may have mechanistic implications for the proteins and drugs that recognize G-rich sequences.

FRET | G-quadruplex | single molecule | telomere | vesicle encapsulation

Eukaryotic chromosomes have a novel nucleoprotein structure at both ends called a telomere. Telomeres are indispensable for protection against genome degradation and have implications in cellular aging and cancer (1–3). In many organisms, a telomeric DNA is composed of tandem repeats of short DNA sequences. *In vitro*, this guanine-rich region forms G-quadruplex (4), which blocks the binding of telomerase (5). The potential importance of G-quadruplex structure was further implied by the recent demonstration *in vitro* that the human POT1, which binds to the single-stranded form of human telomeric DNA (6), facilitates the telomerase activity by disrupting G-quadruplex (7). Human disease helicases, BLM and WRN, also were shown to disrupt G-quadruplexes *in vitro* (8, 9). Because extension of the telomere is critical for generation and growth of cancer cells, the human telomeric G-quadruplex is considered a target for anticancer agents. Structure-based design of chemotherapeutic drugs would require an understanding of a range of conformations that can be adopted by the G-rich sequences, and indeed it was shown to be possible to design molecules that are specific to different types of G-quadruplex structures (10). Although there is no demonstration yet that the G-quadruplex forms in the cell for human telomeric DNA, recent studies of the telomeres of *Stylonychia macronuclei* indicated that G-quadruplexes can indeed form *in vivo* (11), and their formation is controlled by telomere end binding proteins *in vivo* (12). In addition, intracellular formation of the G-quadruplex

was observed in the nontemplate G-rich sequence during transcription (13).

Our goal here is to understand the dynamic structural properties of a G-quadruplex forming DNA using the unique capabilities of single-molecule spectroscopy (14–16) and thereby provide a firm basis for elucidating the interaction between the telomeric DNA and associated proteins. Crystal structures showed that the human telomeric DNA can form a parallel G-quadruplex structure in the presence of K⁺ ions (17). In the parallel structure, all four guanine-contributing edges run in the same direction. In contrast, the previously determined NMR structure in Na⁺ solution was in the antiparallel form (18). More recently, it was shown by NMR studies that parallel and antiparallel forms are in dynamic equilibrium at 100 mM K⁺ (19). Earlier single-molecule fluorescence resonance energy transfer (FRET) measurements of freely diffusing human telomeric DNA molecules detected two populations of differing FRET values (20). The two forms did not interconvert during the time a molecule spends in the excitation volume of a confocal microscope (<1 ms), and both conformations could be made into duplex structures in minutes by means of hybridization to a complementary strand, suggesting that the dynamic interconversion between different conformations may occur in the time window between 1 ms and a few minutes. To probe the dynamic properties of the telomeric DNA, we performed single-molecule FRET experiments on a similar construct (Fig. 1A) but with an added biotin for specific tethering to a quartz surface via biotin–streptavidin linker to extend the observation time. Folding of the DNA into the compact G-quadruplex structure is expected to yield smaller average distance between the donor (tetramethylrhodamine) and the acceptor (Cy5), hence to display higher FRET than the unfolded form (21).

We observed interconversion between three different FRET values, which represent an unfolded and two folded (likely parallel and antiparallel) conformations. Each of the three is further divided into the long- and short-lived species, giving rise to a total of six states. We also observed exchanges between the long- and short-lived species. Kinetic analysis of interconversion pathways, vesicle encapsulation experiments, and mutant studies indicate that the remarkable conformational diversity is intrinsic to the DNA. We propose a model for the identities of these states and their reaction pathways and discuss biological implications on how the functions of telomere-specific proteins and drugs that target telomeric DNA could be modulated by the complex behaviors of the DNA as observed here.

Materials and Methods

DNA Preparation. DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). G-quadruplex

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: ¹U, long-lived unfolded; ¹F₁, long-lived folded 1; ¹F₂, long-lived folded 2; ⁵U, short-lived unfolded; ⁵F₁, short-lived folded 1; ⁵F₂, short-lived folded 2.

[†]J.Y.L. and B.O. contributed equally to this work.

^{||}To whom correspondence should be addressed: E-mail: tjha@uiuc.edu.

© 2005 by The National Academy of Sciences of the USA

Fig. 1. Probing potassium-dependent conformations of human telomeric DNA with FRET. (A) Schematic diagram of DNA construct. The G-quadruplex strand and the complementary stem strand were annealed to form a duplex stem. The quadruplex is depicted in the antiparallel conformation, and the rectangles represent the guanines (yellow in syn and red in anti conformations). Biotin is used to immobilize the DNA to a streptavidin-coated quartz surface, and FRET between tetramethylrhodamine (donor) and Cy5 (acceptor) was measured. (B) Single-molecule FRET histograms as a function of K^+ concentration. (C) An expanded view of FRET histogram at 2 mM K^+ showing three nonzero FRET peaks, named U (unfolded), F_1 , and F_2 (folded). Also shown is the fit by four Gaussians. (D) Single-molecule FRET histograms as a function of K^+ concentration for a mutant sequence as indicated above the graph. (E) Time traces (0.1-s integration time) of donor and acceptor intensities and corresponding FRET from a single molecule at 2 mM K^+ exhibit interconversion between U, F_1 , and F_2 . A transition to zero FRET at 65 s is due to Cy5 blinking. Note that the transition from F_2 to F_1 occurred through U.

strand sequence is 5'-Cy5-(GGGTTA)₃ GGG AGA GGT AAA AGG ATA ATG GCC ACG GTG CG-3'-biotin. For the mutant studies, (GGGTTA)₃ GGG was replaced by (GGGTTA)₂ G TGT-TAGGG, where the underline denotes the change from G to T. Complementary stem strand sequence is 5'-CGC ACC GTG GCC ATT ATC CTT (amino-C6 dT)TA CCT CT-3'.

The G-quadruplex strand includes the human telomeric repeat motif (bold-face type) at the end of which the acceptor dye (Cy5) is attached. Its 3' end is modified by biotin for immobilization. The complementary stem strand is labeled with tetramethylrhodamine internally via amino-modified C6 dT (20). DNA was annealed with 2:1 mixture of the stem and the G-quadruplex strands by heating the mixture to 95°C and then cooling down to room temperature slowly. The overall configuration is shown in Fig. 1A.

Single-Molecule and Bulk Experiments. Total internal reflection fluorescence microscopy was used for single-molecule FRET imaging (22, 23) with 532-nm excitation and a back-illuminated electron-multiplying charge-coupled device camera (iXON, Andor Technology, South Windsor, CT). The quartz surface of the sample cell was successively treated with biotinylated BSA and streptavidin, and 50–200 pM of annealed DNA molecules was added. The same protocol has been used for single-molecule FRET studies of RNA and DNA without any surface-induced perturbations (22, 24). Two sets of data were obtained, one with 0.1-s integration time and 160-s duration and the other with 0.9-s integration time and 900-s duration. Imaging buffer [10 mM Tris-HCl, pH 7.5/0.4% (wt/wt) glucose/1% 2-mercaptoethanol/0.1 g/ml glucose oxidase (Sigma)/0.02 mg/ml catalase (Sigma)] was used as the base buffer, and specified concentrations of KCl or NaCl were added. Sample temperature was controlled by using a water-circulating temperature controller (Neslab Instruments, Portsmouth, NH) and measured by using a thermocouple with an uncertainty of $\approx 1^\circ\text{C}$ (25). Measurements without the temperature control were performed at

room temperature (22°C). Single-molecule FRET efficiency was approximated by using $1/(1 + I_D/I_A)$ where I_A and I_D are the fluorescence intensities of the acceptor and the donor, respectively, after background subtraction and crosstalk correction (26). The negative values of the FRET efficiency are due to background subtraction from very low signal in the acceptor channel, giving rise to negative acceptor intensity. FRET efficiency in the single-molecule histograms is obtained by averaging the first 10 data points of each molecule's time trace. For the bulk kinetic data, we used the Cary Eclipse fluorometer (Varian) and excitation wavelength of 532 nm, and K^+ concentration was increased by manually adding K^+ .

Results

One Unfolded and Two Folded Conformations. Single-molecule FRET efficiency histograms are shown in Fig. 1B for various K^+ concentrations. The zero FRET peak is due to inactive or missing acceptors and henceforth is ignored. At 2 mM K^+ , three additional peaks are apparent (Fig. 1C), one at a high FRET efficiency (≈ 0.8), another at an intermediate FRET efficiency (≈ 0.6), and the third at a low FRET efficiency (≈ 0.3). As K^+ concentration increases, the low FRET state decreases in population. Depletion of the low FRET state also was observed with addition of Na^+ only, but higher ion concentrations were required ($\approx 50\%$ in the low FRET state at 30 mM Na^+ vs. 2 mM for K^+ ; see Fig. 6A, which is published as supporting information on the PNAS web site). Because monovalent cations are needed to induce the stable formation of G-quadruplex with K^+ being more effective than Na^+ , we assign the low FRET state to the unfolded conformation (disordered single-stranded overhang) and the other two states to folded conformations (G-quadruplex structures). Only the low FRET peak was apparent without monovalent ions, consistent with this assignment. Because the buffering agent, Tris, may act as a cation and the low FRET peak could still be due to a particular folded conformation,

Fig. 2. Extreme conformational diversity revealed by single-molecule dynamics. (A) Representative FRET time traces of single molecules at 2 mM K^+ and room temperature (0.9-s integration time). Although some molecules show a single conformation for hundreds of seconds (bottom two traces, acceptor photobleaching events at 810 and 850 s), others show rapid fluctuations between U, F_1 , and F_2 (third trace from the top). The top two traces show a mixed behavior in which switching events between the long- and short-lived species are observed. Segments of long- and short-lived species are marked by solid and dotted lines, respectively. (B) Fraction remaining in F_2 after a given time. The result is fitted by a double-exponential decay. (C) An example FRET time trace (0.1-s integration time) of a mutant sequence as indicated above the graph. The acceptor was photobleached at 115 s. (D) Bulk FRET efficiency, calculated as the ratio of the fluorescence intensity at the acceptor emission peak wavelength divided by the sum of the fluorescence intensities of the donor and the acceptor at their emission peak wavelengths, was measured as a function of time after adding 2 mM K^+ at $t = 0$. The result is fitted by a double exponential.

we tested the neutral Hepes buffer (10 mM, pH 7.4) but observed only the low FRET state in the absence of cations (data not shown). We will use U to denote the unfolded conformation and F_1 and F_2 to denote the folded conformations with intermediate and high FRET values, respectively.

We also studied a mutant sequence for which a guanine in the middle quartet is replaced by a thymine. As anticipated, G-quadruplex formation was significantly disrupted. Folded conformations were observed only at K^+ concentration of >100 mM (Fig. 1D), and no folding was observed with Na^+ even at 1 M concentration (Fig. 6B). With Na^+ , the low FRET peak shifts to higher values as the ion concentration increases (Fig. 6) likely due to reduced electrostatic repulsion between phosphate groups of the DNA backbone that allows more compact conformations of unfolded ssDNA (27). A similar effect was observed up to 100 mM K^+ for the mutant (Fig. 1D). The results below will be on the wild-type sequence unless specified otherwise.

Transitions between all three FRET states (U, F_1 , and F_2) could be observed within single molecules, most frequently in 2 mM K^+ (Fig. 1E); 97%, 1,395 of 1,436, of transitions between F_1 and F_2 passed through U. This result is expected because a direct transition between two folded conformations without unfolding is highly unlikely considering the structural complexity of G-quadruplexes. This observation also makes it very unlikely that F_1 and F_2 are different only in the dye properties. Even though fluorescent labeling does have some effects in that it slightly increased the stability of the G-quadruplex in 100 mM K^+ (20), distinct conformations seen here are not likely to be caused by mere changes in dyes because interconversion between them requires G-quadruplex unfolding.

At low cation concentrations, raising temperature would readily

disrupt G-quadruplexes and favor U. At 2 mM K^+ , U indeed becomes dominant as the temperature increases (see Fig. 7A, which is published as supporting information on the PNAS web site). At 100 mM K^+ , U does not appear significantly even at 45°C, but F_1 becomes populated at the expense of F_2 as the temperature increases (Fig. 7B). The dominance of the parallel over antiparallel conformations at elevated temperatures was observed in NMR studies of bimolecular G-quadruplex of human telomeric sequence (19). If the same is true for the unimolecular construct studied here, F_1 would correspond to the parallel conformation and F_2 to the antiparallel conformation. Circular dichroism measurements of the same DNA sequence and labeling also suggested that at room temperature and at 100 mM K^+ the structure is mostly antiparallel (20), further supporting our assignment. Thus, our data suggest that the human telomeric DNA in physiological conditions (37°C, 140 mM K^+) (28) is primarily folded with comparable populations of parallel and antiparallel conformations.

Long- and Short-Lived Species. Fig. 2A shows five representative time traces of FRET efficiency at 2 mM K^+ (0.9-s integration time and lower excitation powers were used to extend observation time). From such long time traces (>15 min), two types of dynamic behavior are observed. Some molecules maintain their particular conformation for a long time, whereas others show rapid fluctuations between different conformations. Switching between these two distinct behaviors also was observed for some molecules. To confirm that this dichotomy is statistically significant, we analyzed the dwell times of each conformation. Fig. 2B shows the fraction of molecules remaining in F_2 after time t because the molecule is observed to be in that conformation. The biphasic behavior is apparent and the data are well fitted by a double exponential decay

Fig. 3. Conformational distribution of various states. The relative populations of the folded (■) vs. unfolded (○) conformations (*Top*), antiparallel (■) vs. parallel (○) conformations (*Middle*), and long-lived (■) vs. short-lived (○) species (*Bottom*). The results are shown for various K^+ concentrations at room temperature (*Left*), and for various temperatures at 2 mM K^+ (*Center*) and 100 mM K^+ (*Right*).

with decay times of $188 (\pm 2)$ and $20 (\pm 0.1)$ s, suggesting at least two different time scales for unfolding of **F**₂. Similar biphasic behavior was observed for **U** and **F**₁ (see Fig. 8, which is published as supporting information on the PNAS web site). Hence, we further classify the molecular conformations according to the dwell times: long-lived species if the dwell time is >100 s and short-lived species if the dwell time is <100 s. In total, we observed six distinct states: long-lived states (^L**U**, ^L**F**₁, and ^L**F**₂) and short-lived states (^S**U**, ^S**F**₁, and ^S**F**₂) denoted by the superscripts. Although we assign the long-lived folded states, ^L**F**₁ and ^L**F**₂, observed in high K⁺ concentrations to the parallel and antiparallel conformations as seen in high-resolution structural studies, the short-lived folded conformations, ^S**F**₁ and ^S**F**₂, are likely to be different in microscopic details from their long-lived counterparts, although their global folds may be similar.

Conformational Diversity Is Intrinsic. As a further test, we used the vesicle encapsulation technique (29, 30) that confines each molecule in a volume smaller than the diffraction limit. This technique eliminates the possibility of surface artifacts while achieving a long observation time by immobilizing the vesicles. We encapsulated the same G-quadruplex construct, but lacking biotin, in small unilamellar vesicles (200-nm diameter), which in turn were tethered to a supported bilayer (see Fig. 9, which is published as supporting information on the PNAS web site). Control experiments showed that nonspecific binding of DNA to the membrane is undetectable. Vesicle-encapsulated molecules in 2 mM K^+ showed all of the six states as well as interconversions between them (Fig. 9). Therefore, it is highly likely that the extreme conformational diversity observed here is not caused by surface tethering, but is intrinsic to the DNA. As an additional test, we performed bulk kinetic measurements in which 80 nM solution of the DNA was prepared in the imaging buffer without K^+ , and then FRET efficiency was measured as a function of time after addition of 2 mM K^+ (Fig. 2D). Biphasic behavior was observed with lifetimes of 253 (± 20) and 8.8 (± 0.3) s, consistent with the existence of multiple states with markedly different lifetimes. We also note that the long-lived states were not observed for the mutant missing one guanine (Fig. 2C), suggesting that the two distinct lifetimes observed from the wild type are due

to the intrinsic variations in the conformational stability instead of being artifacts caused by fluorescent labeling.

Equilibrium Effect of K^+ and Temperature. Fig. 3 shows how populations shift between different states as K^+ concentration and temperature are varied. The six states are appropriately grouped in this presentation. For example, the folded population includes all four folded states (${}^L\mathbf{F}_1$, ${}^L\mathbf{F}_2$, ${}^S\mathbf{F}_1$, and ${}^S\mathbf{F}_2$). All results were obtained by analyzing the time traces as described (see *Supporting Text*, which is published as supporting information on the PNAS web site).

At room temperature, increasing K^+ concentration induces folding (Fig. 3A). F_1 is maximally populated at 2 mM K^+ and is replaced by F_2 as K^+ concentration is further increased (Fig. 3B). The long-lived species is dominant at both low and high K^+ concentrations and is less occupied at intermediate concentrations because of the appearance of the short-lived species (Fig. 3C). Overall, the DNA is primarily in the long-lived unfolded state ($^L U$) at low K^+ concentrations, fluctuates among the short-lived conformations at intermediate K^+ concentrations, and finally folds primarily into the long-lived antiparallel state ($^L F_2$) when K^+ concentration is further increased. Thus, the short-lived species appears to be a transient form on the pathway from $^L U$ to $^L F_2$.

At 2 mM K^+ , the populations of **U** and short-lived species increase as the temperature increases to 37°C (Fig. 3 *D–F*). Upon further increase in temperature, **¹U** dominates. At 100 mM K^+ , the molecules remain stably folded even at the highest temperature (Fig. 3*G*), but a shift in population from **F₂** to **F₁** was observed (Fig. 3*H*) as has been shown in the histograms (Fig. 7*B*).

Switching Between Long- and Short-Lived Species. Next, we examined the transition rates between various states. Fig. 4 *A* and *B* shows how often the molecule switches between the long-lived (^LU and ^LF) and short-lived (^SX , $\text{X} = \text{U}, \text{F}_1$, or F_2) species. This process is slow, taking hundreds of seconds, but a significant number of molecules showed switching during observation which allowed us to extract the switching rates (see *Supporting Text*). Direct transitions between any two long-lived species are very rare (2%; 12 of 679 traces), whereas much more frequent transitions were observed between the long- and short-lived species (31%; 210 of 679 traces).

Fig. 4. Conformation transition rates between various states. (A and B) The rates of transition between the long- and short-lived species as a function of K^+ concentration at room temperature (A) and as a function of temperature at 2 mM K^+ (B) (○, transition from $^S F$ to $^L U$; □, transition from $^S U$ to $^L F$; ◇, transition from $^L U$ to $^S F$; and △, transition from $^L F$ to $^S U$). Superscripts S and L denote short- and long-lived respectively. (C and D) The rates of folding (■) and unfolding (○) within the short-lived species as a function of K^+ concentration at room temperature (C) and as a function of temperature at 2 mM K^+ (D).

Thus, it is probable that interconversion between long-lived species require a visit to $^{\text{S}}\text{X}$. Increasing K^+ accelerates transitions from $^{\text{S}}\text{X}$ to $^{\text{L}}\text{F}$ and decelerates transitions from $^{\text{S}}\text{X}$ to $^{\text{L}}\text{U}$ (Fig. 4A). Higher temperature accelerates transitions from $^{\text{S}}\text{X}$ to $^{\text{L}}\text{U}$ and decelerates transitions from $^{\text{S}}\text{X}$ to $^{\text{L}}\text{F}$ (Fig. 4B). In contrast to the large effects on the rates of leaving $^{\text{S}}\text{X}$, neither K^+ concentration nor temperature has much influence on the rates of entering $^{\text{S}}\text{X}$ (Fig. 4A and B).

Folding and Unfolding Rates Within the Short-Lived Species. Data shown above suggest that the short-lived species determine the

overall dynamics because they bridge the long-lived states, prompting us to probe the detailed dynamics within the short-lived species. Transitions within the short-lived species are an order of magnitude faster than those between long- and short-lived species (Fig. 4 *C* and *D*). As K^+ concentration is raised, the unfolding rate decreases relatively slowly, whereas the folding rate increases steeply (Fig. 4*C*). This asymmetry suggests the ion uptake is similar for the folded state and the transition state so that their free energies vary by a similar amount as ion concentration is varied (31).

In contrast, the temperature increase affects the unfolding rate much more than the folding rate (Fig. 4D). Weakly temperature-dependent folding rate indicates that the enthalpic barrier is small and the folding barrier is largely entropic. Therefore, the transition state for folding must be much better ordered than the unfolded state. Overall, this analysis suggests that the folding transition state within the short-lived species is significantly ordered and has similar cation binding characteristics as the folded state.

Discussion

Proposed Model for Conformational Diversity. Our single-molecule measurements revealed the existence of distinct species that differ in kinetic stability and indicated that the human telomeric sequence possesses at least six different conformational states. Below, we propose a model for the identities of the six states and the reaction pathways that connect them (Fig. 5). The long-lived folded states, ${}^{\text{L}}\mathbf{F}_1$ and ${}^{\text{L}}\mathbf{F}_2$, were dominant in high K^+ concentrations and were assigned to the parallel and antiparallel structures, respectively, based on comparison with NMR studies of temperature dependence (19). The crystal structure with K^+ showed that a cation is placed at the cavity between two adjacent planes of G-quartets (17) with a total of two cations observed for the human telomeric sequence composed of three G-quartets. It is likely that ${}^{\text{L}}\mathbf{F}_1$ and ${}^{\text{L}}\mathbf{F}_2$ correspond to such well defined structures with stably bound cations between G-quartets. The long-lived unfolded state, ${}^{\text{L}}\mathbf{U}$, was highly populated at low K^+ concentrations and displayed the characteristic salt dependence of a disordered ssDNA.

Fig. 5. Proposed model for conformational identities and reaction pathways. The cartoon for ${}^{\text{L}}\text{U}$ denotes a disordered conformation of the ssDNA, and ${}^{\text{L}}\text{F}_1$ and ${}^{\text{L}}\text{F}_2$ are depicted as the parallel and antiparallel G-quadruplex structures, respectively, holding K^+ ions between G-quartets. The short-lived species (${}^{\text{S}}\text{X}$) include ${}^{\text{S}}\text{U}$, ${}^{\text{S}}\text{F}_1$, and ${}^{\text{S}}\text{F}_2$. The solid arrows represent the proposed reaction paths, and the dashed arrows point toward the states that are favored upon increasing K^+ concentration or temperature.