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

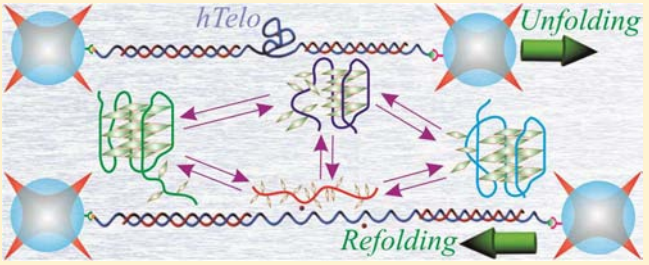
Long-Loop G-Quadruplexes Are Misfolded Population Minorities with Fast Transition Kinetics in Human Telomeric Sequences

Deepak Koirala,[§] Chiran Ghimire,[§] Christopher Bohrer,[#] Yuta Sannohe,[‡] Hiroshi Sugiyama,*[,‡,⊥] and Hanbin Mao*[,§]

[§]Department of Chemistry & Biochemistry and [#]Department of Physics, Kent State University, Kent, Ohio 44242, United States
[‡]Department of Chemistry, Graduate School of Science, and [⊥]Institute for Integrated Cell Material Sciences (iCeMS), Kyoto University, Sakyo-ku, Kyoto, Japan

Supporting Information

ABSTRACT: Single-stranded guanine (G)-rich sequences at the 3′ end of human telomeres provide ample opportunities for physiologically relevant structures, such as G-quadruplexes, to form and interconvert. Population equilibrium in this long sequence is expected to be intricate and beyond the resolution of ensemble-average techniques, such as circular dichroism, NMR, or X-ray crystallography. By combining a force-jump method at the single-molecular level and a statistical population deconvolution at the sub-nanometer resolution, we reveal a complex population network with unprecedented transition dynamics in human telomeric sequences that contain four to eight TTAGGG repeats. Our kinetic data firmly establish that G-triplexes are intermediates to G-quadruplexes while long-loop G-quadruplexes are misfolded population minorities whose formation and disassembly are faster than G-triplexes or regular G-quadruplexes. The existence of misfolded DNA supports the emerging view that structural and kinetic complexities of DNA can rival those of RNA or proteins. While G-quadruplexes are the most prevalent species in all the sequences studied, the abundance of a misfolded G-quadruplex in a particular telomeric sequence decreases with an increase in the loop length or the number of long-loops in the structure. These population patterns support the prediction that in the full-length 3′ overhang of human telomeres, G-quadruplexes with shortest TTA loops would be the most dominant species, which justifies the modeling role of regular G-quadruplexes in the investigation of telomeric structures.



INTRODUCTION

In human cells, telomeres at the end of chromosomes consist of single-stranded 3′ overhang of ∼200 nucleotides with a consensus guanine (G)-rich repeat sequence, 5′-TTAGGG.[1−4] Four such G-rich repeats are known to form a stable DNA secondary structure, G-quadruplex (GQ).[5,6] A G-quadruplex is composed of a stack of G-quartets, each of which is held together by four guanines through Hoogsteen hydrogen bonds and further stabilized by intercalating cations such as K^+ or Na^+.[7,8] Biological investigations suggest that these telomeric DNA secondary structures can regulate the length of telomere either by interfering with telomerase activity or by participating in events such as uncapping of telosomes.[9−11] Since telomere length is closely associated with cellular processes that lead to senescence or cancer, telomeric G-quadruplexes become an attractive target for cancer treatment.[12,13]

Despite their simple repeating sequence, human telomeric G-quadruplexes exhibit a stunning structural polymorphism. At least nine conformations of telomeric G-quadruplex have been revealed in different buffers or in DNA templates that contain four G-rich repeats with varying flanking sequences.[14−21] The observation of partially folded structures either as intermediates to G-quadruplexes or as terminally folded species[22−24] added another level of structural complexity. One rationale for this structural polymorphism is that it presents a flexible regulatory mechanism for cellular processes. Particular biological functions may be regulated by prevailing structures in a population equilibrium that is dependent on cellular conditions such as pH or proteins.

However, ensemble-average techniques, such as circular dichroism (CD), NMR, or X-ray crystallography, have difficulties to deconvolute individual species, especially those with insignificant population fractions or short lifetimes, formed in the same biological molecule. To resolve a structure in such a population mixture, mutations in the biological molecule are often required to selectively populate the species of interest.[20,25,26] Recently, such a practice surprisingly revealed G-quadruplex conformations that harbor $(TTAGGGTTA)_n$ in one of the loops (long-loop GQs) in human telomeric DNA fragments with more than four TTAGGG repeats.[25] This procedure, however, changes population equilibrium and distorts the transition between different species. Due to these difficulties, it is yet to clarify the population equilibrium and the

Received: October 3, 2012
Published: January 17, 2013

   dx.doi.org/10.1021/ja309668t | J. Am. Chem. Soc. 2013, 135, 2235−2241

transition kinetics among different species formed in human telomeric sequences.

Capable of probing one molecule at a time, single-molecule methods offer an unprecedented opportunity to identify conformation and follow transition kinetics of individual species in a population mixture. Assisted by statistical analyses on structures with different changes in contour length (ΔL) upon unfolding,[27] we were able to distinguish populations with ΔL up to tens of nanometers at a baseline resolution of 0.5 nm.[28] This method significantly expands the measurement range with respect to single-molecule FRET, which works in the range of 2−9 nm.[29] With this approach, here, we reveal a surprisingly complex population network that contains G-triplexes, G-quadruplexes, and misfolded long-loop G-quadruplexes in human telomeric sequences with four to eight TTAGGG repeats. By a force-jump method with a temporal resolution of 100 ms,[30] the transition kinetics among different species is found to follow a three-state or a four-state model in the telomeric sequences with four or five TTAGGG repeats, respectively. Our results reveal that G-quadruplexes are the most prevalent species and that long-loop G-quadruplexes are misfolded population minorities whose formation and disassembly are faster than G-quadruplexes or G-triplexes. Population analyses further indicate that misfolded G-quadruplexes in a specific telomeric sequence decrease their abundance with increasing loop sizes or number of long-loops. These results lead us to predict that G-quadruplexes with shortest TTA loops would be the most stable population in the full-length 3′ overhang of human telomeres. That DNA can form misfolded conformation rivals proteins or RNA in the complexity of structure and transition kinetics.

MATERIALS AND METHODS

DNA oligomers were purchased from Integrated DNA Technologies (www.idtdna.com) and further purified by denaturing PAGE gel and stored at −20 °C. The polystyrene beads coated with streptavidin or anti-digoxigenin for the single-molecule experiments were purchased from Spherotech (Lake Forest, IL).

Single-Molecule Assays. DNA constructs for the single-molecule assay was synthesized using a protocol described elsewhere[30−32] (also see Supporting Information Figure S1). Briefly, a DNA construct containing a single-stranded DNA oligomer with a sequence of 5′-(TTAGGG)$_4$ TTA (hTelo-4), 5′-(TTAGGG)$_5$ TTA (hTelo-5), 5′-(TTAGGG)$_2$ TTAAAA (TTAGGG)$_2$ TTA (hTelo-5-Mut), 5′-(TTAGGG)$_6$ TTA (hTelo-6), 5′-(TTAGGG)$_7$) TTA (hTelo-7), or 5′-(TTAGGG)$_8$ TTA (hTelo-8) was sandwiched between 2028 and 2690 bp dsDNA handles. One end of these two dsDNA handles was labeled by either biotin or digoxigenin. These terminal modifications allow tethering of the DNA molecule between the streptavidin and anti-digoxigenin coated beads.

Detailed description of the laser-tweezers setup used for the single-molecule assay has been reported previously.[33] All the experiments, unless specified otherwise, were carried out in a 10 mM Tris buffer at pH 7.4 with 100 mM KCl at 23 °C. To start a single-molecule experiment, a DNA construct was first immobilized onto a 2.10 μm bead via the digoxigenin−anti-digoxigenin−antibody interaction. The DNA immobilized bead and the streptavidin coated bead were trapped by two separate laser foci and the DNA construct was tethered between these two beads. In a typical force−extension experiment, the tethered DNA was extended below the plateau force (maximum 60 pN) and relaxed to 0 pN by moving one of the trapped beads with a loading rate of 5.5 pN/s. For the kinetic measurements using the force-jump assay, the "jump-up" or "jump-down" of the force was achieved by a sudden movement of a steerable mirror within 100 ms time resolution. The force−extension (F−X) curves were recorded at 1000 Hz using a LabView program (National Instruments Corp., Austin, TX). These raw data were filtered with a Savitzky−Golay function with a time constant of 10 ms using a Matlab program (The Math Works, Natick, MA).

Data Analyses. The change in extension (Δx) at a particular force (F) was calculated as the extension difference between the stretching and the relaxing traces at that force. The resulting Δx at this force was then converted to the change in contour length (ΔL) using the worm-like-chain (WLC) model[34] (eq 1):

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

where x is the end-to-end distance (or extension) between the two optically trapped beads, L is the contour length, k_B is the Boltzmann constant, T is absolute temperature, P is the persistent length (51.95 nm),[34] and S is the stretching modulus (1226 pN).[34]

The kernel density treatment and bootstrapping analysis on the ΔL were performed as described in the literature.[27] Briefly, each ΔL transition was expanded by a Gaussian kernel[35] with its width determined from the average of the standard errors in the regions flanking the transition. A probability histogram was obtained after combining Gaussian kernel distributions from all transitions. From each kernel density plot, multiple (2−6) peaks were identified by Igor (WaveMatrics, Portland, OR) program. A total of 3000 random re-sampling was performed to construct a bootstrapping histogram of the selected peaks. When more than one population was observed in the kernel density distribution, the folding probability of each population was estimated from the area under the curve fitted with a Gaussian. The folding probability of each population in the bootstrapping histogram was normalized to that determined in the kernel density distribution. Details of the approaches to justify the bootstrapping analysis are described in Supporting Information.

The folding probability of a population during a particular incubation time was calculated by the ratio of the unfolding events observed for the species in the subsequent unfolding curves to the total subsequent pulling curves. Specific populations were determined based on the change in contour length due to the unfolding event. The probability of unfolded species was calculated by subtracting the sum of the probabilities of all folded species from unity.

RESULTS AND DISCUSSION

Force-Jump Method and Integrated Population Deconvolution at Nanometer Resolution (iPoDNano). Our single-molecule assay (Figure 1) allows mechanical unfolding and refolding of a DNA structure. In a typical force-jump experiment, a DNA construct is extended until the structure formed within is unfolded by the tension accumulated in the construct (red curves in Figure 1B,C). The unfolding is indicated by a sudden change in the extension or the force in the force−extension (F−X) curve. Once unfolded, the tension in the DNA construct is relaxed to 0 pN in 100 ms (green curves in Figure 1B,C) to allow the refolding of the structure in a specific time (the "force-pump" procedure). The refolding is evaluated by next round of force-ramping starting from the 7 pN by a second force-jump (Figure 1B) (the "force-probe" step). If a structure is refolded during the "force-pump" step, unfolding event will be observed in the "force-probe" step. The purpose of the two force-jumps is to avoid refolding during the slow processes of force ramping at entropic force regions (<7 pN, see ref 30). Mimicry of temperature jumps, this force-pump and force-probe (or FP^2) process[30] has demonstrated its capability in the kinetic measurement in which force can flexibly serve to denature or refold a biological macromolecule dependent on the force level.

Recently, we applied an Integrated Population Deconvolution at Nanometer resolution (iPoDNano) to investigate G-quadruplex structures in the hTERT promoter sequences.[27]

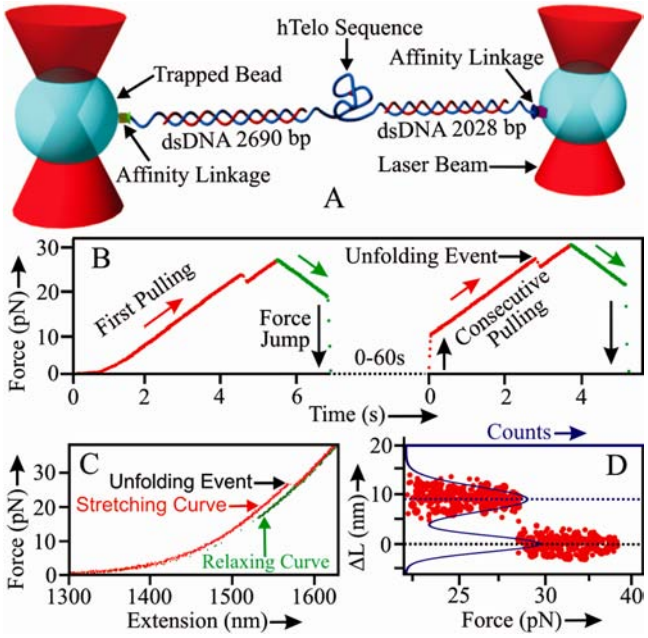


Figure 1. Experimental setup (not to scale). (A) A DNA construct containing a human telomeric sequence (hTelo) is attached to streptavidin and antidigoxigenin coated beads trapped by laser tweezers. (B) The force vs time trace in the force-jump experiment. Force is brought to 0 pN in 100 ms (green dots) after unfolding of the structure in the DNA tether. After incubation (0−60s), force is jumped to 7 pN (red dots) to start subsequent unfolding. (C) A typical force−extension (F−X) curve for the force-jump experiment. Stretching is shown in red; relaxing in green. (D) Plot of change in contour length (ΔL) vs force (filled red circles). ΔL at a particular force is converted from the difference in extension between the red and the green F−X traces in C (see text). Purple curve represent Gaussian fitting.

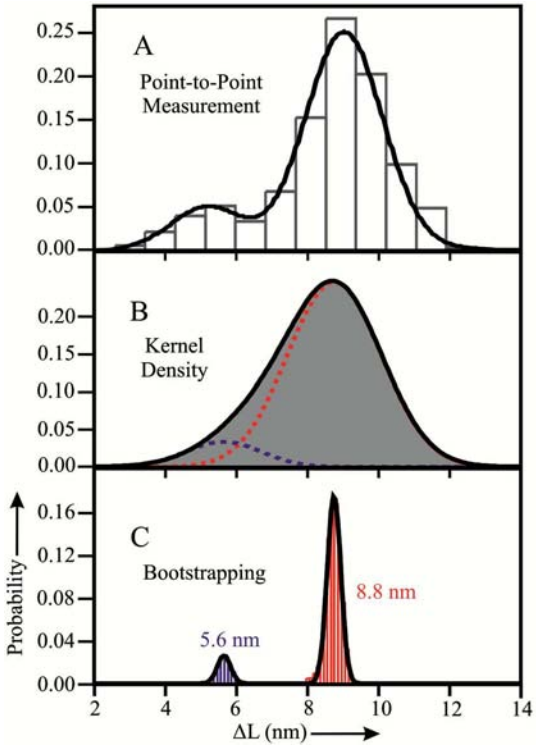


Figure 2. iPoDNano analysis of the population distribution in the hTelo-4 sequence. (A) The ΔL histogram obtained from the point-to-point measurements of the structures unfolded in the hTelo-4 sequence. (B) Kernel density distribution obtained from the trace-by-trace analyses of the same data in A. Populations are fit with Gaussian functions. (C) Bootstrapping analyses deconvolute two populations with a baseline resolution of 1.1 nm.

This method first calculates the change in end-to-end distance (Δx) between the relaxing (green) and the stretching F−X curves (red curve in Figure 1C). The Δx is then converted to the change in contour length (ΔL) by the worm-like-chain (WLC) model[34] in a particular force range (Figure 1D, see Materials and Methods). For each ΔL−F plot, a ΔL histogram is constructed and the ΔL of the folded structure can be retrieved against the background (ΔL = 0, Figure 1D). Further deconvolution using kernel density distribution and bootstrapping analyses (Materials and Methods) allows a spatial resolution of sub-nanometers for ΔL determination.[28] Compared to previous measurements in which each ΔL is estimated from the two data points flanking the unfolding transition (Figure 2A shows a ΔL histogram from this measurement),[31] this method is built on a Gaussian kernel density function that exploits much expanded data sets flanking the transition event. This provides a better signal-to-noise ratio (Figure 2B shows a density distribution of ΔL using a Gaussian kernel function). The resolution is further improved by bootstrapping analyses which filter out noises by selecting the most probable values from the random resampling of Gaussian kernel distributions (Figure 2C). Comparison of the peak centers of the major population identified in Figure 2 revealed that these three methods yielded identical results within experimental errors (9.1 ± 0.1, 8.9 ± 0.2, and 8.8 ± 0.1 nm, respectively, for panels A, B, and C in Figure 2). In addition, the iPoDNano produced ΔL values (8.8 ± 0.1 nm) identical to those obtained from the sequential WLC model fitting[36] method (9.4 ± 1.2 nm, also see Supporting Information).

By combination of the force-jump approach at the single-molecular level and the population deconvolution at the sub-nanometer spatial resolution, we can resolve multiple species in a population mixture. This allows us to follow the transition kinetics of each species and track their possible interconversions. We anticipate such a method can provide a long-sought solution to investigate population dynamics of folded species in human telomeric DNA, which displays a highly complex network of G-quadruplex conformations with intricate transition kinetics hitherto elusive to ensemble-average approaches.

Transition Kinetics of Populations in the hTelo-4 Sequence. With the iPoDNano approach, first, we analyzed folded structures in the hTelo-4 sequence, $(TTAGGG)_4TTA$, in a 10 mM Tris buffer with 100 mM KCl. The probability distribution based on the Gaussian kernel (Figure 2B) showed a similar pattern compared to that obtained from the point-by-point measurements (Figure 2A). Both histograms display a broad ΔL distribution that prevents accurate resolution of individual species. When we performed bootstrapping analysis (see Materials and Methods), however, two populations with baseline separations were obtained at ΔL = 5.6(±0.1) and 8.8(±0.1) nm, respectively (Figure 2C). Based on the baseline separation with $R \geq 1.5$,[37] we estimated the resolution of 1.1 nm for the two peaks depicted here. The resolution increases to 0.38 nm with an acceptable resolution of $R \geq 0.5$. Since the population with longer ΔL (8.8 ± 0.1 nm) matches with

dx.doi.org/10.1021/ja309668t | J. Am. Chem. Soc. 2013, 135, 2235−2241

expected ΔL (9.0 ± 0.1 nm) for the hybrid-1 GQ (see Supporting Information for details), it was assigned as this conformation.[38] The 5.6 nm species has a size consistent with the partially folded G-triplex species described previously.[23] When we performed similar experiments in a buffer that contains 100 mM LiCl, the yields reduce to 5 ± 2% for unfolding events with $\Delta L \leq 5$ nm and 0% for $\Delta L > 5$ nm. In comparison, yields for G-triplex (~5 nm) and G-quadruplex (~9 nm) are 15 ± 4% and 65 ± 5%, respectively, in the KCl buffer. Since Li^+ is known to disfavor G-quadruplex formation, this control experiment verifies that folded populations observed here are G-quadruplexes, rather than hairpins that are not sensitive to Li^+. Further evidence of the G-quadruplex formation comes from the CD signatures in which ~268 and ~290 nm peaks with a ~238 nm valley are characteristic of hybrid type G-quadruplex conformations (see Figure S4). Using these two ΔL values (8.8 and 5.6 nm) as Gaussian centers, the populations in the kernel density distribution (Figure 2B) were fit with a two-peak Gaussian, which shows 14% for the partially unfolded species and 86% for the hybrid-1 GQ. This result indicates that GQ is a major population in the equilibrium.

Next, we used force-jump method to follow the transition kinetics of the species in the hTelo-4 sequence (Figure 1B). Figure 3A shows ΔL histograms for different species revealed

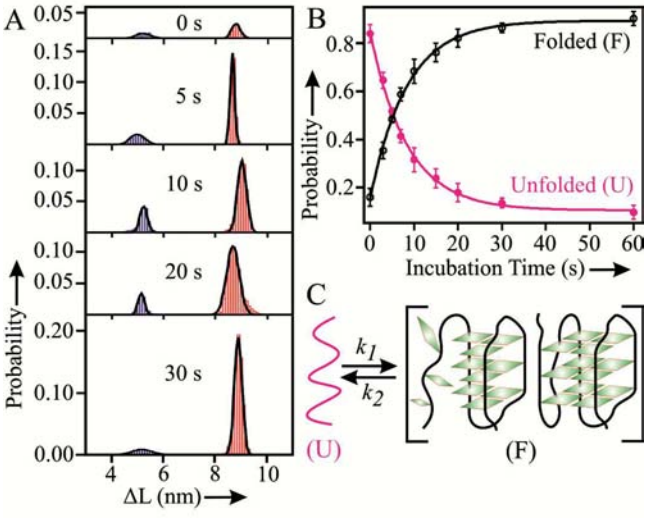


Figure 3. Transition kinetics of the hTelo-4 sequence. (A) iPoDNano analyses reveal ΔL vs incubation time at 0 pN. Each population is fit with a Gaussian. (B) The probability of folded (GQ+G-triplex) and unfolded populations as a function of incubation time. Curves are fitted with the functions solved for the two-state model shown in C (see Supporting Information).

by iPoDNano approach with different incubation time. Each histogram contains two populations that have been assigned as the partially folded G-triplex (ΔL = 5.6 nm) and the fully folded G-quadruplex (ΔL = 8.8 nm) (Figure 2C). Shown in Figure 3B are the probabilities of refolding for each species with incubation time. Here GQ and G-triplex are grouped as a single population of folded species to accommodate the simplest two-state model (Figure 3C, see Supporting Information for fitting equations). The rate constants k_1 and k_2 obtained here serve as starting values for the more complex fitting based on the three-state or four-state model described later.

To retrieve transition kinetics for individual species, we analyzed GQ, G-triplex, and unfolded states separately. We used a linear (Figure 4A) and a triangular (Figure 4B) three-

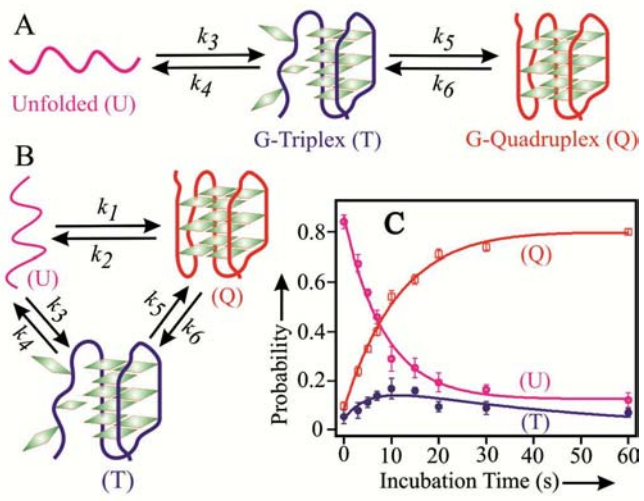


Figure 4. Refolding probability of the GQ and G-triplex with incubation time for the hTelo-4 sequence: (A) linear model and (B) triangular model. (C) Refolding probability data are fit with functions (see Supporting Information) analytically solved from the model shown in B.

state model to represent the in-pathway and off-pathway nature of the G-triplex with respect to the GQ folding, respectively. Both models fit the data equally well by residue analyses (Figure 4C and Figure S3, average κ^2 = 0.0039 vs 0.0040, see Supporting Information for analytically solved equations). When we analyzed the unfolding pattern in each $F-X$ curve, we found it is rare to observe two sequential unfolding or refolding transitions expected for the G-triplex that serves as an in-pathway intermediate. Instead, most often we observed that occurrence of the G-triplex is independent of the GQ, which strongly suggests that the G-triplex is an off-pathway intermediate. The same model has been proposed by Vesnaver and co-workers recently[24] to reflect transition dynamics of different species in a similar telomeric sequence. Fitting of this model to the data allows retrieving six rate constants (Figure 4C and Table 1) for all possible transitions among three species. The transient refolding probability maximum for the partially folded species unequivocally proves the intermediate nature of the G-triplex, since refolding probability would vary monotonically with time as predicted by a two-state, folding-unfolding model if the G-triplex exists independently. Interestingly, direct folding from unfolded state to GQ is the most favorable route with fastest kinetics. The intermediate nature of the G-triplex with a low population fraction well explains the fact that the species is evasive for many ensemble-average techniques.

Transition Kinetics of the Populations in the hTelo-5 Sequence. Next, we proceeded to investigate the transition kinetics for species in a DNA sequence that contains five TTAGGG repeats (hTelo-5). The ΔL histogram shows four populations with 5.0(±0.1), 7.5(±0.1), 8.8(±0.1), and 11.5(±0.2) nm ΔL with formation possibilities of 8%, 8%, 74%, and 10%, respectively (Figure 5A). In control experiments performed in a 100 mM Li^+ buffer, these populations were rarely observed (4 ± 2%, 3 ± 1%, and 0% for species with ΔL

dx.doi.org/10.1021/ja309668t | J. Am. Chem. Soc. 2013, 135, 2235−2241

Table 1. Reaction Rate Constants (Average ± SD, s^{-1}) for Transitions in hTelo-4 and hTelo-5 Sequences

sequence	k_1	k_2	k_3	k_4	k_5
hTelo-4	0.1 ± 0.03	0.009 ± 0.001	0.012 ± 0.004	0.023 ± 0.006	0.013 ± 0.002
hTelo-5	0.083 ± 0.02	0.003 ± 0.0008	0.126 ± 0.02	0.006 ± 0.001	0.036 ± 0.006

sequence	k_6	k_7	k_8	k_9	k_{10}
hTelo-4	0.001 ± 0.0005				
hTelo-5	0.002 ± 0.0009	3.52 ± 0.84	4.32 ± 0.311	0.134 ± 0.09	0.044 ± 0.007

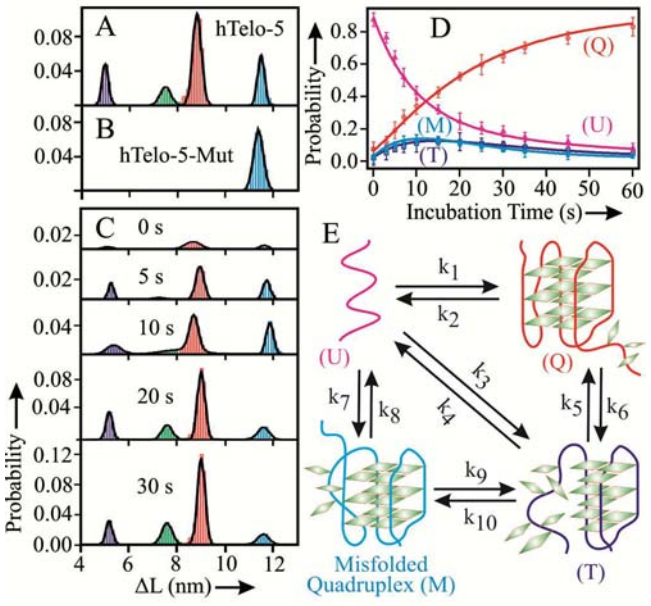


Figure 5. iPoDNano analyses and transition kinetics of the hTelo-5 sequence. (A) ΔL histogram for hTelo-5 obtained by bootstrapping analyses. The purple population depicts G-triplex, the green and red populations represent GQs, and the blue population is a misfolded G-quadruplex. (B) ΔL histogram for the hTelo-5-Mut sequence. (C) ΔL histograms of the hTelo-5 fragment with different incubation time. (D) Refolding probability of individual species over time. Two GQ populations (red and green) are grouped together (orange) in a four-state kinetic model shown in E (see text). Other color codes are described in A. Transition rate expression for each species in the four-state model is solved analytically (see Supporting Information) to fit the data in D.

≤ 5 nm, 5 < ΔL ≤ 10 nm, and ΔL > 10 nm, respectively). These results suggest that the species with ΔL > 5 nm formed in the K^+ buffer are likely to be G-quadruplexes that are sensitive to Li^+ ions. The formation of G-quadruplexes in hTelo-5 is further supported by the observation of ∼268 and ∼290 nm peaks and a ∼238 nm valley in the CD spectrum (Figure S4), which are indicative of G-quadruplex conformations. Based on the matching ΔL values, the 5.0 and 8.8 nm populations are assigned to the partially folded species and the hybrid-1 GQ, respectively. The 7.5 nm species has a ΔL similar to that of the basket type GQ previously observed in a 100 mM Na^+ buffer.[23] Observation of the basket type GQ in the K^+ buffer is possible since GQ conformation is dependent on various factors such as flanking DNA sequences[39] or concentrations of a DNA template.[38,40] However, the 11.5 nm population does not match with any of the known human telomeric GQs. Since hTelo-5 contains five G-rich repeats, it is possible to form a conformation in which a G-rich repeat (TTAGGG) is harbored in one of the internal loops.[25]

To validate this conformation, we performed experiments on a mutant sequence, $(TTAGGG)_2$-TTAAAA-$(TTAGGG)_2$ TTA, or hTelo-5-Mut, which can only form a G-quadruplex with terminally located G-rich repeats while leaving a long loop in the middle. Indeed, CD signatures of this sequence match well with those of a hybrid type GQ (∼268 and ∼292 nm peaks, and a 242 nm valley, see Figure S4). In mechanical unfolding experiments, we observed a single population with ΔL = 11.4(±0.2) nm (Figure 5B), which confirms that the 11.5 nm ΔL population observed above (Figure 5A) is indeed a G-quadruplex species harboring a TTAGGG sequence in a long loop. This observation is consistent with the NMR structure that shows a long-loop GQ in a human telomeric sequence with five TTAGGG repeats in a K^+ buffer.[25] Since it is not possible for the long-loop G-quadruplex to fold into a GQ with regular TTA loops without unfolding first, we surmise it serves as a misfolded species similar to those often observed in protein or RNA species.

To test the misfolded nature of the long-loop GQ, we investigated the dynamic equilibrium among the species formed in the hTelo-5 using the force-jump approach described above. Figure 5C shows ΔL histograms for all four species with various incubation times. The probabilities of refolding for individual species are plotted in Figure 5D. To reduce the complexity of kinetic equations to a manageable level, two conformations of GQs (hybrid-1, ΔL = 8.8 nm; basket type, ΔL = 7.5 nm) were grouped as a single GQ population in a four-state equilibrium model (Figure 5E). In this model, the long-loop GQ cannot convert to a regular GQ without first entering a third state such as unfolded DNA or G-triplex. Such a scenario is a likely result of the topological exclusivity for a G-rich repeat that either serves as one of the four G-stems in the GQ or stays free inside the long-loop GQ. Since current approach cannot distinguish individual G-triplexes or misfolded GQs that are associated with specific GQs probably due to similar sizes among these structures, the grouping of the two GQs becomes necessary to analytically solve the kinetics equations. With these assumptions, we obtained analytical solutions to the four-state model shown in Figure 5E (see Supporting Information). We notice that only few analytical solutions exist for a linear four-state model,[41] which is different from current model. The data can be fit well with the rate equation for each species in the four-state model (Figure 5D, Table 1, and Supporting Information), validating the misfolded nature of the long-loop GQ. Both G-triplex and misfolded-GQ demonstrate a maximal refolding probability during incubation, a feature indicative of the interconversion between these species and other DNA populations. Although misfolded DNA was implied in previous investigations in artificial DNA fragments,[42,43] the long-loop GQ shown here is the first physiologically relevant misfolded species with defined conformations.[25] This direct observation lends strong support to the notion that DNA can be as complex as RNA or proteins from structural and kinetic perspectives.

dx.doi.org/10.1021/ja309668t | J. Am. Chem. Soc. 2013, 135, 2235−2241

Equilibrium Population Dynamics in the hTelo-4 and the hTelo-5 Sequences. Compared to the kinetics in the hTelo-4, the formation of GQ (Table 1, k_1) remains constant whereas that for the G-triplex (k_3) increases about 8 times in the hTelo-5 (k_3/k_1 of hTelo-4 vs hTelo-5 = 1 vs 12.5). Such a behavior can be ascribed to the fact that there are increased possibilities to form G-triplexes rather than GQs in longer telomeric sequences. If this reasoning is true, then it is expected that the unfolding rates of G-triplexes should not depend on the DNA size, since G-triplexes already exist in these DNA fragments. Indeed, the unfolding kinetics of the G-triplex (k_4) with respect to that of the GQ (k_2) remains the same in both hTelo-4 and hTelo-5 fragments (k_4/k_2: 2.6 vs 2.0 for hTelo-4 vs hTelo-5). However, due to the formation of the misfolded GQ in the hTelo-5, the overall formation probability of the G-triplex reduces to 3.4% ($k_3/(k_1+k_3+k_7)$) from 10.7% in the hTelo-4. After G-triplex is formed, the species has a similar probability to fold into the regular GQ (hTelo-4: hTelo-5 = 36%:42%, calculated by $k_5/(k_4+k_5+k_{10})$). On the other hand, the G-triplex has an increased tendency to unfold in the hTelo-4 (64%, $k_4/(k_4+k_5+k_{10})$) compared to the hTelo-5 (7%). These observations are consistent with the fact that in the hTelo-5, either GQ or misfolded GQ can serve as a destination for G-triplex (Figure 5E), leading to a reduced possibility for the competing unfolding pathway. In the hTelo-4, however, only GQ competes with the unfolding (Figure 4B), which increases the unfolding probability of the G-triplex. Compared to the hTelo-5, such a result suggests that G-triplex in the hTelo-4 has an increased tendency to exist as an independently folded species off-pathway to the GQ. This off-pathway feature in the hTelo-4 is confirmed by the fact that in comparison to the hTelo-5, the unfolding probability of GQ to G-triplex, which is also present in the in-pathway process, is much reduced in the hTelo-4 (10% vs 40%, calculated by $k_6/(k_2+k_6)$). Notably, the long-loop GQ shows the fastest formation (k_7) and disassembly (k_8) kinetics among all species, suggesting that although the sequence misfolds into the long-loop GQ, it is not "trapped" as escape from this state has a comparable rate. Indeed, analyses from Table 1 show 94% ($k_7/(k_1+k_3+k_7)$) folding probability of the misfolded GQ, which is comparable to the unfolding probability ($k_8/(k_8+k_9)$= 97%).

Formation Probability of Misfolded GQs in Longer Human Telomeric Sequences. In previous NMR studies with carefully designed mutants, long-loop GQs (or misfolded-GQs) were observed in human telomeric sequences with 5−7 G-rich repeats.[25] Using iPoDNano, here we observed that under physiologically relevant buffer conditions and DNA concentrations, these species are population minorities. As shown in Figure 6, these misfolded species represent 10%, 11%, and 13% of the total population in hTelo-5, hTelo-6, and hTelo-7 fragments, respectively. When we compared misfolded populations in the same fragment, we found that species with longer loop(s) or more long-loops have smaller populations, suggesting a decreased thermodynamic stability for these misfolded species. In hTelo-6, the populations of folded species with longer loops or more long-loops drop from 8% to 3%; in HTelo-7, these values reduce from 9% to 2%. We argue that entropic factor could be the culprit for the decrease in population of such misfolded GQs.[44,45] As loops become longer or more long-loops occur, the loss of entropy due to the confinement of the loop(s) inside the GQ becomes larger, which reduces the population of misfolded species.

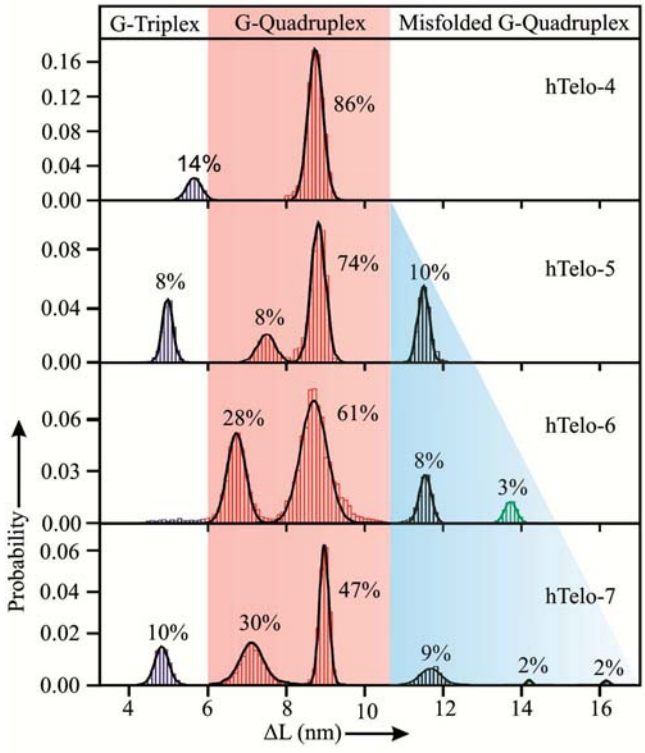


Figure 6. Distribution probability of folded population in human telomeric sequences (hTelo-4 to hTelo-7). The 11.5 nm species is the GQ that contains one G-rich repeat (TTAGGG) in a loop, the 14 nm species contains two G-rich repeats in loop(s), while the 16 nm species has three G-rich repeats in loop(s).

In all DNA constructs investigated here (hTelo-4 to hTelo-7), we observed that regular G-quadruplexes are the most populated species (Figure 5). Such an observation has been supported by CD spectra as discussed above (Figure S4). This population pattern is further supported in a DNA construct that contains eight G-rich repeats (hTelo-8) in which two individual G-quadruplex units can form. Mechanical unfolding experiments on this construct revealed two-step (∼45%) as well as one-step (∼55%) unfolding transitions (see Figure S5 for typical $F-X$ and $\Delta L-F$ curves that contain two-step transitions in hTelo-8). The ΔL of each transition in the two-step unfolding events bears similarity to that observed in the hTelo-4 fragment (G-quadruplex to G-triplex ratio ∼6:1). Consistent with this ratio, majority of the two-step events shows unfolding of two individual GQs (60% out of all two-step transitions). Among 55% one-step transitions, the majority (85% out of all one-step transitions) is the unfolding of G-quadruplex (including simultaneous unfolding of two G-quadruplexes) with 11% of G-triplex and few (∼4%) misfolded GQ species. These observations suggest that regular GQs with the shortest possible TTA loops would be the predominant species in the human telomeric regions. However, further investigations on longer telomeric sequences with multiple runs of 4 G-rich repeats are required to fully support this argument.

■ CONCLUSIONS

Using force-jump and population deconvolution at sub-nanometer resolution, we have identified individual populations and followed their kinetics with unprecedented resolutions in human telomeric sequences containing four to eight TTAGGG

dx.doi.org/10.1021/ja309668t | J. Am. Chem. Soc. 2013, 135, 2235−2241

repeats. We have observed a highly complex population equilibrium that contains G-triplex, misfolded GQs, and predominant GQ species. The misfolded GQs harbor one or more G-rich repeats in the loop(s) and the population of these species in a particular sequence decreases with increase in the loop length or the number of long-loops. The presence of the misfolded species testifies the structural complexity of DNA. The complexity of the system is further reflected by the transition kinetics in the hTelo-4 and hTelo-5 DNA in which three-state and four-sate kinetic models are followed, respectively. The population dynamics of telomeric species observed here indicate that in the full-length 3′ end overhang of human telomere, G-quadruplex units with shortest possible TTA loops would be the most prevalent species. This justifies the modeling role of regular G-quadruplexes in the study of human telomeric structures.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S5 and associated details, ΔL calculation, justification of iPoDNano, and the analytical solutions of different kinetic models. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

hs@kuchem.kyoto-u.ac.jp; hmao@kent.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

H.M. acknowledges support from NSF CHE-1026532. This work was partly supported by ICC program of Japan Society for the Promotion of Science and Core Research for Evolutional Science and Technology (CREST) of JST to H.S.

■ REFERENCES

(1) Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J.-R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6622−6626.

(2) Blackburn, E. H. *Nature* **1991**, *350*, 569−573.

(3) Balagurumoorthy, P.; Brahmachari, S. K. *J. Biol. Chem.* **1994**, *269*, 21858−21869.

(4) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. *Genes Dev.* **1997**, *11*, 2801−2809.

(5) Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275−283.

(6) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703−730.

(7) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871−880.

(8) Chowdhurdy, S.; Bansal, M. *J. Phys. Chem. B* **2001**, *31*, 7572−7578.

(9) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* **1991**, *350*, 718−720.

(10) Lipps, H. J.; Rhodes, D. *Trends Cell Biol.* **2009**, *19*, 414−422.

(11) Wang, Q.; Liu, J.-q.; Chen, Z.; Zheng, K.-w.; Chen, C.-y.; Hao, Y.-h.; Tan, Z. *Nucleic Acids Res.* **2011**, *39*, 6229−6237.

(12) Balasubramanian, S.; Neidle, S. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345−353.

(13) Cian, A. D.; Lacroix, L.; Douarre, C. l.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Biochimie* **2008**, *90*, 131−155.

(14) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263−282.

(15) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876−880.

(16) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963−9970.

(17) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927−4940.

(18) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 2440−2450.

(19) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517−6525.

(20) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 4301−4309.

(21) Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824−9833.

(22) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938−18943.

(23) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. *Chem. Commun.* **2012**, *48*, 2006−2008.

(24) Bončina, M.; Lah, J.; Prislan, I.; Vesnaver, G. *J. Am. Chem. Soc.* **2012**, *134*, 9657−9663.

(25) Yue, D. J. E.; Lim, K. W.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 11462−11465.

(26) Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2010**, *38*, 1009−1021.

(27) Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 5157−5164.

(28) Yu, Z.; Mao, H. *Chem. Rec.* **2012**, DOI: 10.1002/tcr.201200021.

(29) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer-Verlag: New York, 2006.

(30) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. *Nat. Chem.* **2011**, *3*, 782−787.

(31) Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2009**, *131*, 1876−1882.

(32) Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991−8997.

(33) Mao, H.; Luchette, P. *Sens. Actuators B* **2008**, *129*, 764−771.

(34) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185−6190.

(35) Cheng, W.; Arunajadai, S. G.; Moffitt, J. R.; Tinoco, I., Jr.; Bustamante, C. *Science* **2011**, *333*, 1746−1749.

(36) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630−633.

(37) Harris, D. C. *Quantitative Chemical Analysis*, 8th ed.; W.H. Freeman and Co.: New York, 2010.

(38) Yu, Z.; Koirala, D.; Cui, Y.; Easterling, L. F.; Zhao, Y.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 12338−12341.

(39) Vorlíčková, M.; Chládková, J.; Kejnovská, I.; Fialová, M.; Kypr, J. *Nucleic Acids Res.* **2005**, *33*, 5851−5860.

(40) Dhakal, S.; Cui, Y.; Koirala, D.; Ghimire, C.; Kushwaha, S.; Yu, Z.; Yangyuoru, P. M.; Mao, H. *Nucleic Acids Res.* **2012**, DOI: 10.1093/nar/gkt038.

(41) Goldman, L. *Biophys. J.* **2006**, *91*, 173−178.

(42) Ansari, A.; Shen, Y.; Kuznetsov, S. V. *Phys. Rev. Lett.* **2002**, *88*, 069801.

(43) Nakamura, I.; Shi, A.-C.; Nutiu, R.; Yu, J. M. Y.; Li, Y. *Phys. Rev. E* **2009**, *79*, 031906.

(44) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405−16415.

(45) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J. L. *Nucleic Acids Res.* **2010**, *38*, 7858−7868.