Published online 8 February 2013 *Nucleic Acids Research, 2013, Vol. 41, No. 6* ***3915–3923***
doi:10.1093/nar/gkt038

Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions

Soma Dhakal, Yunxi Cui, Deepak Koirala, Chiran Ghimire, Saurabh Kushwaha, Zhongbo Yu, Philip M. Yangyuoru and Hanbin Mao*

Department of Chemistry and Biochemistry, Kent State University, Kent, OH 44242, USA

Received November 20, 2012; Revised January 5, 2013; Accepted January 7, 2013

ABSTRACT

Recent experiments provided controversial observations that either parallel or non-parallel G-quadruplex exists in molecularly crowded buffers that mimic cellular environment. Here, we used laser tweezers to mechanically unfold structures in a human telomeric DNA fragment, 5′-(TTAGGG)$_4$TTA, along three different trajectories. After the end-to-end distance of each unfolding geometry was measured, it was compared with PDB structures to identify the best-matching G-quadruplex conformation. This method is well-suited to identify biomolecular structures in complex settings not amenable to conventional approaches, such as in a solution with mixed species or at physiologically significant concentrations. With this approach, we found that parallel G-quadruplex coexists with non-parallel species (1:1 ratio) in crowded buffers with dehydrating cosolutes [40% w/v dimethyl sulfoxide (DMSO) or acetonitrile (ACN)]. In crowded solutions with steric cosolutes [40% w/v bovine serum albumin (BSA)], the parallel G-quadruplex constitutes only 10% of the population. This difference unequivocally supports the notion that dehydration promotes the formation of parallel G-quadruplexes. Compared with DNA hairpins that have decreased unfolding forces in crowded (9 pN) versus diluted (15 pN) buffers, those of G-quadruplexes remain the same (20 pN). Such a result implies that in a cellular environment, DNA G-quadruplexes, instead of hairpins, can stop DNA/RNA polymerases with stall forces often <20 pN.

INTRODUCTION

G-quadruplexes have drawn significant research attention as their *in vivo* existence has been demonstrated and their potential roles in gene regulation have been unveiled (1–4). During the past decade, X-ray crystallography and NMR approaches have shown rather versatile G-quadruplex conformations. For example, human telomeric G-quadruplexes have exhibited at least nine conformations (5–12). In X-ray crystallography, snapshot structures are obtained with limited information on the dynamic interaction between the G-quadruplex and solvent molecules. NMR can probe structures in a solution. However, the relatively high substrate concentration required for NMR may not be fully compatible with that available *in vivo*. In a solution with more than one species, it becomes rather challenging to reveal the conformation of each species using these methods.

Most of the investigations on the G-quadruplex are performed in diluted buffers that have limited biological relevance, as cellular environment is filled with ~40% biomacromolecules such as proteins. To address this issue, molecularly crowded conditions that mimic cellular environment are used (13). In these crowded buffers, NMR revealed parallel conformation of G-quadruplexes (6). However, non-parallel G-quadruplexes were also observed (14,15). The controversial observations may be ascribed to different experimental conditions such as types of cosolutes, concentrations and the sequences of G-rich DNA fragments and re-annealing rates (6,16–18). Recent investigation by HÃnsel *et al.* (15) cautioned the use of polyethylene glycol as a cosolute and suggested that proteins or egg extracts are better cosolutes to mimic cellular conditions. However, the structural complexity of these biomacromolecular cosolutes can interfere with the NMR signal (14,15), which thwarts the efforts to resolve G-quadruplex structures under these conditions.

The force measurement by single-molecule approaches such as AFM or laser tweezers naturally prevents signal interference from the macromolecular cosolutes. Because structural probing is performed one at a time, it is expected that the information can be acquired for each population in a solution mixture. Importantly, the superior sensitivity of single-molecule methods enables structural probing at DNA concentrations close to *in vivo* levels. As only a few DNA copies contain the

*To whom correspondence should be addressed. Tel +1 330 672 9380; Fax: +1 330 672 3816; Email: hmao@kent.edu

sequence of interest in nucleus, this concentration is often in the nanomolar range (see Supplementary Data for detailed calculation).

Force-based single-molecule approach has an additional and unique advantage to reveal mechanical properties, such as unfolding force, F_{unfold}, of a biomolecule. In diluted buffers, it has been discovered that F_{unfold} of G-quadruplexes (19–21) is significantly higher than the stall force of many motor proteins, such as DNA and RNA polymerases (22–24). In contrast, F_{unfold} of hairpins (25), which have a dsDNA stem in the structure, is comparable with the stall forces. This suggests that with respect to hairpins, G-quadruplexes can serve as a stronger road block to the transcription or replication processes. However, there is a lack of reports on the mechanical stabilities of either DNA G-quadruplexes or hairpins under molecularly crowded conditions, which prevents further interpretation of the biological relevance of these secondary structures.

Here, we describe a laser-tweezers-based single-molecule approach to quickly identify each G-quadruplex conformation and reveal its mechanical stability in a molecularly crowded buffer at physiologically relevant concentrations of a human telomere sequence, 5′-$(TTAGGG)_4$TTA. Assisted with a statistical deconvolution method [integrated Population Deconvolution at Nanometer resolution, or iPoDNano, (26)], this approach identifies G-quadruplex species by determining the handle-to-handle distances among three nucleotides arranged in a triangle. To discriminate the dehydration or the excluded-volume effect on the formation of specific G-quadruplex structures, we have analyzed population distribution in different crowded buffers. In dehydration buffers that contain either 40% (w/v) DMSO or 40% (w/v) ACN as cosolutes, parallel G-quadruplex coexists with hybrid 1 or basket G-quadruplex, respectively, with a ratio of 1:1. In buffers that mainly demonstrate steric effects [40% w/v bovine serum albumin (BSA)], we have identified a parallel (10%) and a hybrid 1 G-quadruplex (90%). Previously, the large quantity of BSA in this buffer deteriorated NMR signals, which failed to determine G-quadruplex structures explicitly. Our experiments also reveal for the first time that mechanical stabilities of human telomeric G-quadruplexes remain the same, whereas those of a DNA hairpin significantly decrease, in crowded buffers with respect to diluted solutions.

MATERIALS AND METHODS

DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Enzymes were purchased from New England Biolabs (NEB, Ipswich, MA), and all the chemicals (>99% in purity) were purchased from VWR (West Chester, PA). The surface functionalized beads for the single-molecule experiments were obtained from Spherotech (Lake Forest, IL).

DNA constructs

All the DNA sequences (except the dsDNA handles) including azide or terminal alkyne groups at specific nucleotides are shown in Supplementary Table S1. The [5′–3′] (U1) DNA construct for the mechanical unfolding of human telomeric G-quadruplex in the sequence, 5′-TTA $G^1G^2G^3$ $T^4T^5A^6$ $G^7G^8G^9$ $T^{10}T^{11}A^{12}$ $G^{13}G^{14}G^{15}$ $T^{16}T^{17}A^{18}$ $G^{19}G^{20}G^{21}$ TTA-3′, was sandwiched between two double-stranded DNA (dsDNA) handles using procedures described previously (19). Briefly, one of the dsDNA handles 2690 bp in length was prepared from the pEGFP vector (Clontech, Mountain View, CA) by digestion with SacI and EagI restriction enzymes. The SacI end was functionalized with digoxigenin (Dig-dUTP, Roche) through terminal deoxynucleotidyl transferase (Fermentas). The other dsDNA handle of 2028 bp in length was obtained from polymerase chain reaction of a pBR322 template using a biotinylated primer and a primer containing either an XbaI restriction site or a click chemistry functional group (Supplementary Table S1). Literature procedures were followed to prepare the [5′-L2] (U2) and [3′-L2] (U3) constructs (27). Briefly, an EagI overhang was introduced at the 5′ dsDNA terminal (for U2) or the 3′ dsDNA terminal (for U3) in a dsDNA–single-stranded DNA (ssDNA) hybrid that contains the hTel4G sequence. The respective dsDNA termini were then ligated to the 2690 bp dsDNA handle using T4 DNA ligase. The T11 at the central loop (L2) of the hTel4G fragment is replaced with an azide-modified uridine for U2 and U3 constructs (Supplementary Table S1). The 2028 bp dsDNA handle was then attached to the azide group through the copper-catalyzed azide-alkyne cycloaddition reaction (28). The resultant DNA was purified by ethanol precipitation. This DNA construct contains a single-stranded hTel4G sequence linked to one dsDNA handle through a phosphodiester bond and another handle through the click chemistry connection. By replacing the hTel4G sequence with the sequence, 5′-$GC(T)_{19}GCTTTTGC(A)_{19}GC$, a construct that contains a DNA hairpin was similarly prepared.

Single-molecule experiments

The mechanical unfolding of G-quadruplex in hTel4G was performed in a home-built dual-trap 1064 nm laser tweezers (29) at room temperature under different buffer conditions (10 mM Tris buffer that contains 100 mM of either NaCl or KCl species and 40% w/v DMSO, ACN or BSA at pH 7.4). Before the unfolding, the DNA construct was immobilized onto a 2.10 μm bead through the digoxigenin–anti-digoxigenin–antibody complex. The DNA-immobilized bead and streptavidin-coated bead were trapped by two laser foci separately. Next, we brought one of the laser traps close to the other through a motorized steerable mirror to tether the DNA construct between the beads. The tethered DNA was extended below the denaturation plateau force (see later, we set the maximum force <60 pN in diluted buffers and ~50 pN in DMSO-, ACN- or BSA-containing buffers), and relaxed to 0 pN in ~10 s by moving one of the trapped beads with a loading rate of 5.5 pN/s. The force-extension (*F-X*) curves were recorded at 1000 Hz using a Labview® program. The raw data were filtered with a Savitzky–Golay function, with a time

constant of 10 and 50 ms, respectively, for buffers without and with cosolutes using a Matlab program (The MathWorks, Natick, MA).

After each pair of beads is trapped, the trap stiffness of each laser focus is calculated based on the thermal motion of the trapped bead using a reported method (29). The trap stiffness information in diluted and molecularly crowded buffers is summarized in Supplementary Table S2. The accuracy of this instrument on the force measurement was confirmed by the observation of the 65 pN plateau that indicates the denauration of single dsDNA molecules in a 10 mM Tris buffer (pH 7.4) with 100 mM KCl (30). The accuracy of the instrument on the distance measurement was confirmed by the determination of the contour length per DNA nucleotide using the method described in literature (25). The value, 0.44 ± 0.02 nm per nucleotide, is identical with the literature (25) and within the range measured by other approaches (31–34).

Plot of change-in-contour-length versus force

After each *F-X* curve was split into the stretching (red) and relaxing (black) traces (Figure 1B), the change in extension (Δx) was calculated by subtracting the former trace from the latter at a particular force (F). The resulting Δx at the particular force was then converted to the change-in-contour-length (ΔL) using the following equation based on the Worm-Like-Chain (WLC) (35) model when the contour length of the secondary structure is much less than that of the dsDNA handles (<5%) (27,35):

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

where ΔL reflects the change in the apparent contour length at $F = 0$ pN between the two pulling handles before and after a structure is unfolded, k_B is the Boltzmann constant, T is absolute temperature, F is the force, P is the persistent length and S is the elastic stretch modulus. To determine the P and S under different buffer conditions, we first fit *F-X* curves using a sequential WLC model that accounts for both the dsDNA handles and the secondary structure tethered in between (36) (see Supplementary Figure S1 for *F-X* curves and fittings). The results of these fitted parameters (Supplementary Table S2) are then used here to obtain the ΔL values in a specific buffer. The ΔL values from this approach are identical within experimental error with those obtained from the WLC sequential fitting described previously (Supplementary Table S2).

Kernel density treatment and bootstrap analysis

The kernel density treatment and bootstrap analysis on ΔL populations were performed as described in the literature (26,37). Briefly, for kernel density treatment, the probability density of each transition between a folded structure and an unfolded ssDNA was estimated according to a Gaussian kernel (38). A kernel density histogram was obtained by adding the probability density for each transition (see Figure 1D for example). From each kernel density plot, the highest two peaks were identified by Igor (WaveMatrics, Portland, OR) program. A total of 3000 random re-sampling was performed to construct a bootstrapping histogram of selected peaks (see Figure 2A for example). When more than one population was observed in kernel density histograms, each population was fitted with a Gaussian to estimate the population from the area under the curve. The probability of each population in the bootstrapping histogram was normalized to that in the kernel density histogram.

Root mean square deviation calculation

The G-quadruplex formed in the hTel4G was identified by comparing the handle-to-handle distances (x) measured in the laser-tweezers experiment with those obtained from the NMR or X-ray G-quadruplex structures (Supplementary Table S3). Using ΔL obtained from the laser-tweezers measurements, we first calculated the x value according to the expression, $x = L - \Delta L = n \times L_{nt} - \Delta L$ (equation 2) (27,39), where L and L_{nt} are the contour lengths for a folded structure and single nucleotide, respectively, and n is the number of nucleotides in the folded structure. All x values were evaluated against corresponding distances measured from NMR or X-ray structures using the root mean square deviation (RMSD). We performed two rounds of RMSD calculations. In the first round, we calculated the RMSD using an average L_{nt} (0.44 nm) obtained from all available G-quadruplex conformations based on the ΔL obtained from the 5′–3′ geometry in our laser-tweezers measurement. In the second round, we used the L_{nt} of the conformation that showed the least RMSD in the first round to recalculate RMSD. These RMSD comparisons have been presented in Figures 2, 3 and 4. The visual comparison of the x values for laser-tweezers measurements and those from NMR or X-ray structures has been shown in Supplementary Figure S2.

Deconvolution of populations

For the deconvolution of two populations in a ΔL histogram, the overall population was fit with a two-peak Gaussian function. Based on the ratio determined by this fitting, the data under the intersection region were randomly assigned to one of the populations. After this random assignment, we plotted the unfolding force histograms in Supplementary Figures S3, S4 and S5.

Calculation of change in free energy of unfolding

The change in free energy of unfolding ($\Delta G_{\mathrm{unfold}}$) for the DNA hairpin was calculated from the Jarzynski non-equilibrium theorem (40) (see equation 3). The $\Delta G_{\mathrm{unfold}}$ for each G-quadruplex population was calculated using the same equation after the population deconvolution described above.

$$\Delta G_{\mathrm{unfold}} = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp\left(-\frac{W_i}{k_B T}\right) \quad (3)$$

where N is the number of observations, and W is the non-equilibrium work done to unfold the G-quadruplex structure(s), which is equivalent to the hysteresis area

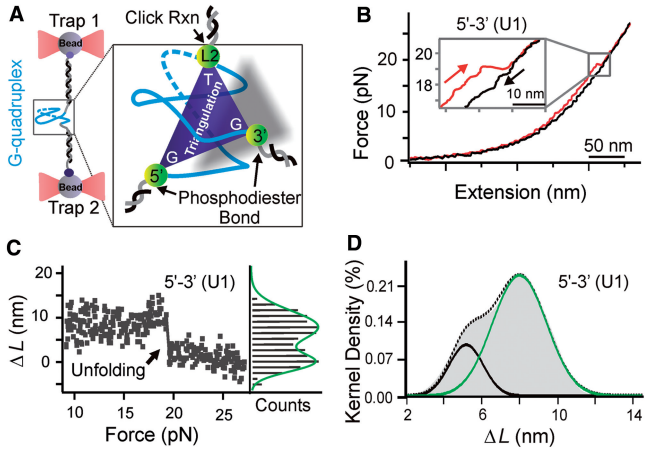

Figure 1. Mechanical unfolding of human telomeric G-quadruplex in 100 mM Na^+ buffer (10 mM Tris, pH 7.4). (**A**) Schematic of the laser-tweezers setup and the triangulation structural probing approach. Three unfolding geometries, 5′–3′ (U1), 5′-L2 (U2) and 3′-L2 (U3), are highlighted in a dark-blue triangle. (**B**) A typical force-extension (*F-X*) curve containing unfolding (red) and refolding (black) of the U1 trajectory. Inset shows the unfolding event. (**C**) Plot of the change in contour length (ΔL) versus force (9–27 pN) based on the *F-X* curve in B. Histogram shown to the right depicts folded ($\Delta L \sim 8.0$ nm) and unfolded ($\Delta L = 0$ nm) populations. (**D**) Kernel density plot for the U1 geometry. The black dotted curve represents a two-peak Gaussian fit, whereas the black and green curves are Gaussian fits for the minor and major populations, respectively.

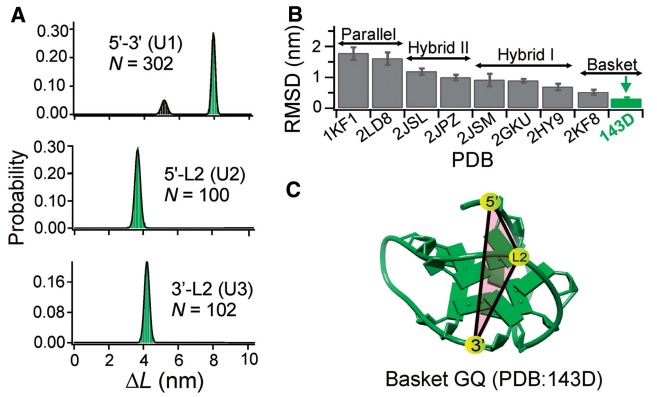

Figure 2. G-quadruplex structural identification in a 100 mM Na^+ buffer (10 mM Tris, pH 7.4) using bootstrap and RMSD analyses. (**A**) Bootstrapping ΔL analysis in the geometries of U1 (top), U2 (middle) and U3 (bottom). The partially and the fully folded species in U1 are indicated with black and green colors, respectively. Each histogram is fitted with a Gaussian. *N* depicts the number of experiments. (**B**) RMSD comparison between the handle-to-handle distance from the laser-tweezers measurements and those from PDB structures. The least RMSD value is depicted by a green arrow. (**C**) The best-matching G-quadruplex (PDB: 143D) with a highlighted triangle.

between the stretching and relaxing *F-X* curves. The bias of the ΔG_{unfold} was calculated from the unfolding work histograms, as described in the literature (41). Due to the presence of the G-triplexes, which may display the same ΔL as that of the G-quadruplexes in the U2 or U3 trajectory (Supplementary Figure S6), the observed *W* contains contributions from G-triplexes. To estimate ΔG_{unfold} in these cases, we used the U1 unfolding in the same buffer as a reference. As U1 provides distinct G-triplex and G-quadruplex populations (Figures 2A, 3A and 4A), we obtained ΔG_{unfold} for each species using equation 3. Next, assuming inseparable G-triplexes and G-quadruplexes, we combined the unfolding work for both species and calculated an apparent ΔG_{unfold} for the U1 geometry. The difference between this value and the ΔG_{unfold} for the G-quadruplexes serves as the correction factor to retrieve the ΔG_{unfold} for the fully folded G-quadruplexes in the U2 or U3 geometry (Table 1).

RESULTS AND DISCUSSION

The triangulation structural identification method

By attaching an ssDNA fragment, 5′-$(TTAGGG)_4$TTA-3′ (or hTel4G), between two optically trapped beads (Figure 1A), we can unfold the structure formed in the sequence from the 5′ to 3′ direction after moving the two beads apart (19,42). Change-in-contour-length (ΔL) due to the structural unfolding can be retrieved by comparing the *F-X* curves before and after the unfolding transition (Figure 1B). We obtain the distance (x) between the two dsDNA handles that hold the structure using equation 2 (see Materials and Methods for details). This handle-to-handle distance can be compared with that measured from NMR or X-ray crystallography to determine the best-matching conformation. Previously, we inferred G-quadruplex structures in the insulin-linked polymorphic region on the basis of this measurement in a one-dimensional space (19). By using dsDNA handles through different residues in a biomolecule, in principle, multidimensional distance measurements can be obtained to reconstruct a complete topology of the structure formed in the biomolecule. Nevertheless, such a complete structural profiling is time-consuming and redundant if the task is to identify the best-matching conformation only (27,39). With a balance of measurement time and accuracy of structural information, we determined three inter-handle distances arranged in a triangle (Figure 1A) to identify G-quadruplex conformation in the hTel4G.

To better represent the entire structure, we chose three well-separated nucleotides from which handles are attached: two terminal guanines and the middle thymine in the central loop (L2) (Figure 1A and Supplementary Figure S6). Whereas the two terminal handles were introduced through phosphodiester bonds by enzymatic ligations (27), the middle thymidine in the central loop was substituted with an azide-modified uridine to allow the click chemistry attachment (see Materials and Methods) (27). By grabbing two handles among the three nucleotides, a total of three different unfolding geometries can be obtained. We designate 5′ to 3′ geometry as the U1, 5′ to L2 as the U2 and 3′ to L2 as the U3 trajectory.

The reliability of the method is dependent on the accurate distance or ΔL measurement. The accuracy of our laser-tweezers measurement was confirmed by unfolding a DNA hairpin using a reported procedure (25), which revealed an identical contour length per nucleotide (see Materials and Methods for details). The capability of

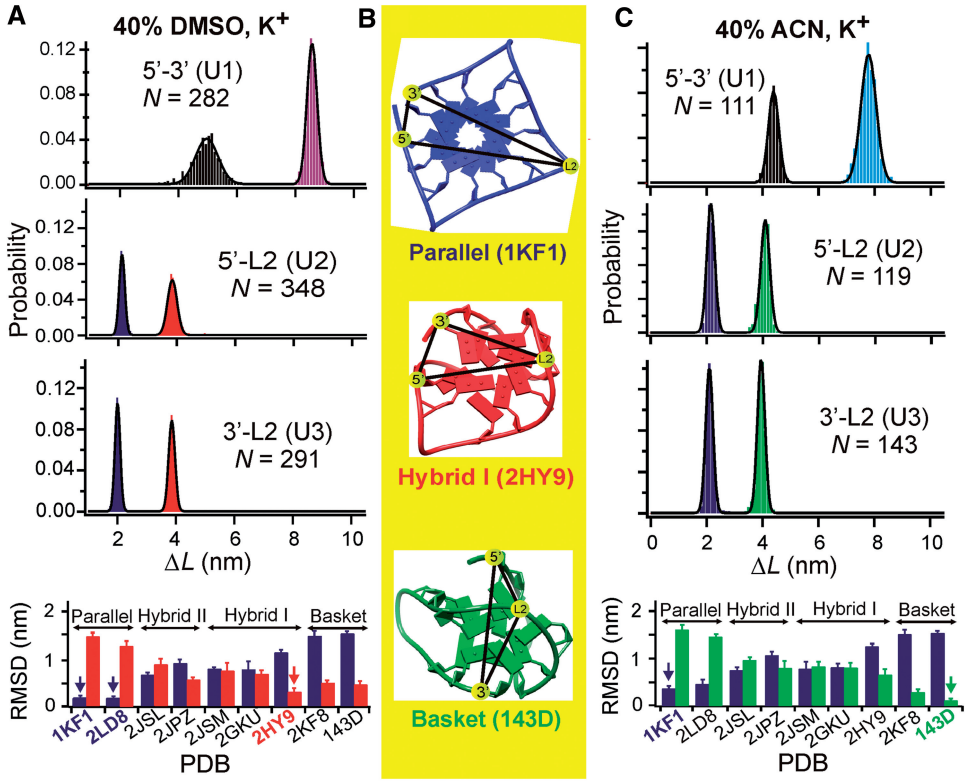

Figure 3. G-quadruplex structural analysis in crowded solutions (pH 7.4, 100 mM K^+) with small-molecule cosolutes (40% w/v). (**A**) Population (top three panels) and RMSD (bottom panel) analyses of folded species in 40% DMSO. Blue bars represent the RMSD for the major species (ΔL = 8.5 nm) in U1 and the ~2 nm populations in U2 and U3. Red bars represent the major species (ΔL = 8.5 nm) in U1 and the ~3.8 nm populations in U2 and U3. *N* depicts the number of experiments. (**B**) The best-matching G-quadruplex structures labeled with triangulation. (**C**) Population (top three panels) and RMSD (bottom panel) analyses of folded species in the same buffer with 40% (w/v) ACN. The *y*-axes of the top three panels are the same as (A). *N* depicts the number of experiments. Blue bars represent the RMSD for the major species (ΔL = 7.9 nm) in U1 and the ~2 nm populations in U2 and U3. Green bars represent the major species (ΔL = 7.9 nm) in U1 and the ~4.0 nm populations in U2 and U3. The least RMSD values are indicated by arrows in (A) and (C).

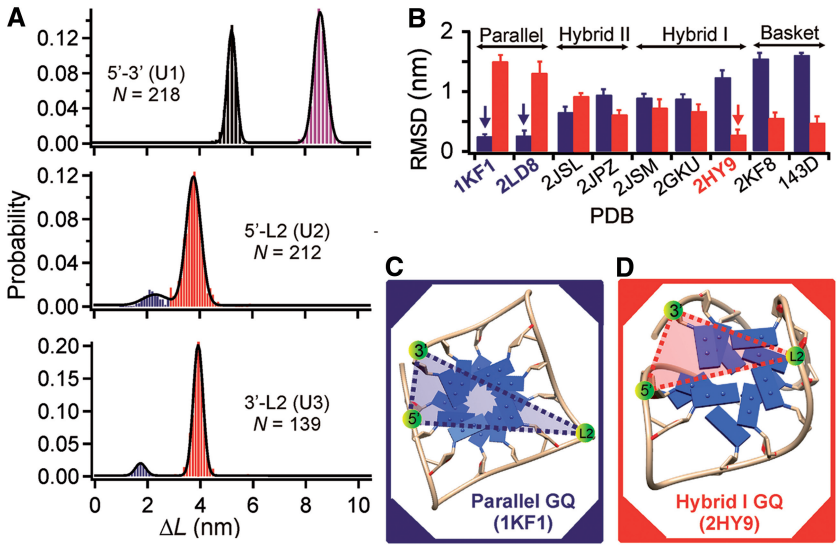

Figure 4. G-quadruplex conformational analysis in crowded solutions (pH 7.4, 100 mM K^+) with biomacromolecular cosolutes (40% w/v BSA). (**A**) Population analysis for the U1 (top), U2 (middle) and U3 (bottom) unfolding trajectories. *N* depicts the number of experiments. (**B**) RMSD analysis for the best-matching G-quadruplex structures (indicated by arrows). Blue bars represent the RMSD for the major species (ΔL = 8.5 nm) in U1 and the ~2 nm populations in U2 and U3. Red bars represent the major species in U1 and the ~3.8 nm populations in U2 and U3. (**C** and **D**) Structures of the best-matching G-quadruplexes.

Table 1. Unfolding force (F_{unfold}, in pN), change in free energy of unfolding (ΔG_{unfold}, in kcal/mol) and % population of G-quadruplexes under different conditions

	5′-3′(U1) (~8 nm)		5′-L2 (U2) (~4 nm) ΔG_{unfold}	3′-L2 (U3) (~4 nm) ΔG_{unfold}	5′-L2 (U2) (~2 nm) ΔG_{unfold}	3′-L2 (U3) (~2 nm) ΔG_{unfold}
	F_{unfold}	ΔG_{unfold}				
100 mM Na^+ (diluted)	16 ± 3	8.2 ± 0.1 (−0.3)	8.1 ± 0.4 (−0.4)	8.2 ± 0.9 (−0.3)	Population not observed	
	Basket (100%)					
100 mM K^+[a] (diluted)	21 ± 1	9.8 ± 0.4 (0.1)	10.5 ± 0.6 (−0.4)	10.0 ± 1.0 (−0.3)	Population not observed	
	Hybrid I (100%)					
100 mM K^+ 40% DMSO (crowded)	19 ± 1	8.4 ± 0.2[b] (−0.6)	8.6 ± 0.2 (−0.1)	9.0 ± 0.3 (−0.4)	7.9 ± 0.4 (−0.3)	8.6 ± 0.4 (0.3)
	Parallel + Hybrid I (100%)		Hybrid I (54 ± 2%)		Parallel (46 ± 2%)	
100 mM K^+ 40% ACN (crowded)	20 ± 1	9.8 ± 1.1[b] (−0.4)	10.2 ± 1.0 (−0.2)	9.8 ± 0.3 (0.7)	9.5 ± 0.9 (−0.2)	8.6 ± 0.5 (−0.1)
	Parallel + Basket (100%)		Basket (52 ± 3%)		Parallel (48 ± 3%)	
100 mM K^+ 40% BSA (crowded)	19 ± 2	8.2 ± 0.1[b] (−0.1)	8.8 ± 0.5 (−0.4)	7.7 ± 0.3 (−0.4)	N/A[c]	N/A[c]
	Parallel + Hybrid I (100%)		Hybrid I (90%)		Parallel (10%)	

Bias for ΔG_{unfold} is shown in parentheses.
[a]Data from (27).
[b]Apparent ΔG_{unfold} from mixed populations (see 'Materials and Methods' section).
[c]Not enough data points for reliable ΔG_{unfold} determination.

this method to differentiate different population is dependent on the resolution in the ΔL measurements, which can be improved by repetitive sampling from many DNA tethers. This practice allows constructing ΔL histograms, from which the most likely ΔL can be obtained at sub-nanometer resolution (Figure 1C, green curve).

We tested this approach in a 10 mM Tris buffer with 100 mM Na^+ (pH 7.4), which is known to populate a basket-type G-quadruplex in the hTel4G (10). The ΔL histogram displays two populations (Figure 1C). To identify each population, we performed an integrated population deconvolution with nanometer or sub-nanometer resolution (or iPoDNano) (26,37), in which a kernel density treatment reveals the probability distribution of all measured ΔL values (Figure 1D), and a bootstrap analysis depicts the most probable ΔL populations in the overall ΔL data set (Figure 2A). This deconvolution clearly shows two populations with ΔL of ~5.1 nm (Figure 2A, top panel, black) and ~8.0 nm (green). Previously, the former species was assigned as a partially folded G-triplex and the latter as a fully folded basket G-quadruplex (21). However, only one species with ΔL of 4.1 or 4.3 nm was observed for the U2 or the U3 unfolding, respectively (Figure 2A). Calculations showed that along these two trajectories, expected ΔL for the G-triplex unfolding is either identical with that for the G-quadruplex, or smaller than 1 nm, which is well below the ΔL values observed previously (see Supplementary Figure S6). Taking into account that G-triplex only constitutes 25% of the overall population (Figure 2A), the majority of the ΔL populations in the U2 or the U3 trajectory were assigned to the G-quadruplexes. Even in the case that G-triplex has a population comparable with G-quadruplex (see later), the ΔL populations in U2 and U3 can still be reliably assigned to G-quadruplexes due to the identical ΔL values between G-triplex and G-quadruplex species (Supplementary Figure S6).

The handle-to-handle distances calculated from these two ΔL values and that from the 8.0 nm population in the U1 (equation 2) were compared with those of the NMR or X-ray structures (see Materials and Methods) (5–12). Although many PDB structures do not have the same flanking sequence as the hTel4G, they all share the core G-quadruplex forming sequence, $(GGGTTA)_3GGG$, which provides a solid justification for our comparison method. The RMSD between these two data sets (Figure 2B) reveals the best-matching structure is indeed the basket G-quadruplex (PDB: 143D, Figure 2C), which has been shown to populate in the core $(GGGTTA)_3GGG$ sequence in the same buffer by NMR studies (10). Such a result strongly suggests that dsDNA handles introduced by enzymatic ligations or click chemistry reactions do not significantly interfere with the folding of G-quadruplexes. Similar approach with six inter-handle distances confirmed the expected formation of a hybrid 1 G-quadruplex in the hTel4G in a K^+ buffer (27). These findings validated our structural identification method for DNA structures.

Dehydration effect on the G-quadruplex conformation

Next, we set out to probe the G-quadruplex structures in crowded solutions with 40% (w/v) small-molecule cosolutes, such as DMSO or ACN, which are known to contribute to the molecular crowding mainly through dehydration effect (6,14). Similar to the U1 geometry in the dilute solution (Figure 2A), both partially (5.0 nm) and fully folded (8.5 nm) populations were observed in 40% DMSO solution (pH 7.4) with 100 mM K^+ (Figure 3A, top panel). To our surprise, two ΔL species with a

similar population ratio were revealed in either U2 or U3 trajectory (48%:52% for the 2.0:3.9 nm population in U2 and 44%:56% for the 2.1:3.8 nm species in U3, see Table 1, Figure 3A and Supplementary Figure S3). As G-triplexes are population minorities (Figure 3A), the two species are most likely different G-quadruplexes. The similar population ratio suggests that the 2.0 nm population in U2 is the same G-quadruplex as that of the 2.1 nm population in U3, whereas the 3.9 nm G-quadruplex in U2 is identical to that of the 3.8 nm species in U3. This assignment is confirmed by comparing the difference in the handle-to-handle distance between the U2 and U3 trajectories measured from NMR/X-ray structures (Supplementary Table S3, column 5 in top panel) with that determined by mechanical unfolding experiments (Supplementary Table S3, bottom panel). Compared with the assignment of a ~2 nm species and a ~4 nm species to the same population, assignment of the two ~2 nm species as one population and that of the two ~4 nm species as another (Supplementary Table S3, bottom panel) both produce differences in the handle-to-handle distance much closer to those expected from the NMR/X-ray measurement. One exception is for the PDB 2KF8 species, which is a rather unusual two-stack G-quadruplex (11). With this assignment, the RMSD was calculated (Figure 3A, bottom panel), which indicates the ~2.0 nm species as the parallel G-quadruplex (PDB: 1KF1 and 2LD8 in Figure 3B and Supplementary Figure S2) and the ~3.8 nm population as the hybrid 1 G-quadruplex (PDB code: 2HY9, Figure 3B and Supplementary Figure S2). The formation of G-quadruplexes was confirmed by circular dichroism (CD) in the same solution (Supplementary Figure S7), which shows ~265- and ~295-nm peaks indicative of G-quadruplex conformations (43).

Similarly, two populations were also observed in 40% (w/v) ACN solutions (Figure 3C). Using the same strategy described above (Supplementary Table S3), we assigned the 1.9 nm species in U2 and 2.1 nm species in U3 as one population, whereas the 4.0 nm species in U2 and 3.9 nm species in U3 as another. Subsequent RMSD treatments allowed us to determine the best-matching structures as the parallel and the basket G-quadruplexes (Figure 3C and Supplementary Figure S4). It is noteworthy that the smallest RMSD values used for the structure identification in DMSO or ACN were not bigger than those calculated from the assignment of a ~2 nm species and a ~4 nm species as the same population (see above), thereby corroborating current assignment. Previously, either parallel or non-parallel G-quadruplex was observed in this solution (6,14). Our observations that both types of G-quadruplexes exist in the same crowded solution (Table 1) at 23°C run against previous findings obtained at higher DNA concentrations with different flanking sequences (6,14,44,45). As re-annealing time in our mechanical unfolding experiments is much shorter than CD or NMR measurements (<1 min versus tens of minutes), we reasoned that concentration, re-annealing time and flanking sequences of DNA may serve as important factors for the formation of different G-quadruplexes.

Steric effect on the G-quadruplex conformation

DMSO and ACN contribute to the molecular crowding effects mainly through water depletion (6,14). Inside the cell, the excluded-volume effect from biological macromolecules, such as proteins, contributes to the crowding effects (46). However, it is difficult to interpret G-quadruplex conformations amidst background signals from complex protein structures in conventional methods such as NMR (14,15) and CD. The method described here is relied on the distance measurements from the DNA molecules specifically tethered between two trapped beads (Figure 1A), which prevents the inference from cosolutes. To demonstrate this, we used 40% (w/v) BSA as the biomacromolecular cosolute. Similar to the small-molecule cosolutes, parallel G-quadruplex was observed (the ~2 nm species in Figure 4A, middle and bottom panels). However, this conformation only constitutes 10% of the overall population (see Supplementary Figure S5 for population deconvolution), whereas the majority is the hybrid 1 G-quadruplex (the ~4 nm population in Figure 4 and Table 1).

The excluded-volume effect is expected to accommodate a conformation with smaller size. Based on the fact that parallel G-quadruplex is not the most compact conformation among telomeric G-quadruplexes with known structures (14), the steric effect from the crowding environment was ruled out as the driving force for parallel G-quadruplex formation. Instead, dehydration effect was proposed as a predominant factor, which was supported by the identification of parallel G-quadruplex by NMR in buffers with dehydrating cosolutes such as DMSO or ACN (6). A critical test for this hypothesis is to show that non-parallel G-quadruplexes are major species in a crowded solution that mainly demonstrates excluded-volume effect, such as in a 40% BSA buffer. Due to the signal broadening in NMR, however, limited success achieved so far was only suggestive that non-parallel G-quadruplexes exist in these solutions (14,15), although the exact conformations are yet to be obtained. The coexistence of a hybrid 1 and a parallel G-quadruplex with a 9:1 ratio observed here represents the first definite conformation identification in these solutions. Compared with the fact that ~50% of the population is parallel G-quadruplex in crowded buffers that contain dehydrating cosolutes, the 10% parallel conformation in the BSA buffer provides direct experimental evidence that dehydration is the driving force to induce parallel G-quadruplexes than the excluded-volume effect.

Mechanical and thermodynamic stabilities of the G-quadruplex species

In addition to the structural interpretation, force-based single-molecular approach can provide unique information on the mechanical stability of a folded structure and the ΔG_{unfold} (20,47). Whereas the mechanical stability of dsDNA in a sliding-like geometry (5′–3′) significantly decreases from 65 to 52 pN, at which dsDNA denatures (30), the rupture force (F_{unfold}) of G-quadruplexes remains the same (~20 pN) in a crowded buffer compared with a diluted solution (Table 1 and Supplementary Figures S3, S4

and S5). When a DNA hairpin is unfolded along the same geometry as that of the G-quadruplex, the rupture force also shows a substantial decrease in a crowded solution (9 pN) compared with that in a diluted buffer (15 pN) (Supplementary Figure S8). It is noteworthy that other mechanical properties, such as persistence length (P) and elastic stretch modulus (S) of dsDNA handles and unfolded ssDNA, do not vary significantly between different buffers (Supplementary Table S2). These results indicate that in crowded buffers, F_{unfold} of G-quadruplexes is higher than the stall force of DNA or RNA polymerases (<20 pN), whereas that of hairpins is within the reach of the load force of these polymerases. It is expected, therefore, that G-quadruplex can serve as an effective mechanical impediment to the DNA replication or RNA transcription processes. In contrast, DNA hairpins are likely to be dissembled during these cellular activities.

To estimate the ΔG_{unfold}, we used the Jarzynski non-equilibrium theorem (see equation 3 in Materials and Methods). As shown in Table 1, different unfolding geometries of the same species provide similar ΔG_{unfold}, confirming that free energy is a state function independent of reaction pathways (27). The population ratio of the basket versus parallel G-quadruplex in the ACN buffer and that of the hybrid 1 versus parallel G-quadruplex in the DMSO solution well reflect the thermodynamic equilibrium determined by corresponding ΔG_{unfold} values (Table 1). As the radii of gyration for basket and hybrid 1 G-quadruplexes are among the smallest for G-quadruplexes with known structures (14), the excluded-volume effect may explain their prevalence in crowded buffers. Comparison of ΔG_{unfold} reveals that the basket G-quadruplex displays a 23% increase in stability, whereas the hybrid 1 G-quadruplex shows a 17% decrease, in crowded buffers compared with diluted solutions (Table 1). The increase in the stability of the basket G-quadruplex is expected, as Hoogsteen base pairs are stabilized by molecular crowding effects (48). The 17% decrease in the stability of the hybrid 1 G-quadruplex is surprising. We notice that deconvolution of G-quadruplex from the inseparable G-triplexes may present an uncertainty in ΔG_{unfold} estimation (see 'Materials and Methods' section). It is also possible that cosolutes may have specific recognitions on different G-quadruplex species (15). On the other hand, the 18% decrease in the thermodynamic stability for a DNA hairpin (7.0 → 5.3 kcal/mol, from the Jarzynski calculation), which constitutes a long dsDNA stem (see Materials and Methods for the sequence), fully agrees with the finding that Watson–Crick base pairing has a 14.8% decrease in stability in crowded buffers (48). It is also consistent with the decrease in rupture force observed above.

CONCLUSIONS

In summary, by using a single-molecule structural probing method, we have been able to identify a mixture of parallel and non-parallel G-quadruplexes in the same molecularly crowded solution at physiologically relevant DNA concentrations. Our method has provided direct evidence that dehydration effect is the predominant factor to populate parallel G-quadruplexes. In contrast to the DNA hairpins that show a significant decrease in the mechanical stability in crowded buffers versus diluted solutions, we have revealed that the rupture force of G-quadruplex, which is higher than the stall force of DNA or RNA polymerase, remains the same. This implies that in a cellular environment, G-quadruplex serves as an effective mechanical impediment for replication or transcription processes, whereas DNA hairpins are not likely to stop DNA or RNA polymerases due to their weaker mechanical stability with respect to the stall force of the polymerases. We anticipate that the single-molecule approaches described here can serve as a quick structural identification method for species with known PDB structures, especially for those that exhibit structural polymorphism in the same buffer or display distinct structures under conditions not amenable to conventional structural probing methods.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3 and Supplementary Figures 1–8.

ACKNOWLEDGEMENTS

The authors acknowledge support from National Science Foundation (NSF, CHE-1026532).

FUNDING

Funding for open access charge: National Science Foundation (NSF) [CHE-1026532].

Conflict of interest statement. None declared.

REFERENCES

1. Paeschke,K., Capra,J.A. and Zakian,V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell*, **145**, 678–691.
2. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
3. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
4. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends Cell Biol.*, **19**, 414–422.
5. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
6. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
7. Dai,J., Punchihewa,C., Ambrus,A., Chen,D., Jones,R.A. and Yang,D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.*, **35**, 2440–2450.
8. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
9. Phan,A.T., Kuryavyi,V., Luu,K.N. and Patel,D.J. (2007) Structure of two intramolecular G-quadruplexes formed by

natural human telomere sequences in K+ solution. *Nucleic Acids Res.*, **35**, 6517–6525.
10. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.
11. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.
12. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
13. Ellis,R.J. and Minton,A.P. (2003) Cell biology: join the crowd. *Nature*, **425**, 27–28.
14. Miller,M.C., Buscaglia,R., Chaires,J.B., Lane,A.N. and Trent,J.O. (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.*, **132**, 17105–17107.
15. HÃnsel,R., Löhr,F., Trantírková,S.F., Bamberg,E., Trantírek,L. and Dötsch,V. (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.*, **3**, 1–8.
16. Petraccone,L., Pagano,B. and Giancola,C. (2012) Studying the effect of crowding and dehydration on DNA G-quadruplexes. *Methods*, **57**, 76–83.
17. Fujimoto,T., Nakano,S.-i., Sugimoto,N. and Miyoshi,D. (2012) Thermodynamics-hydration relationships within loops that affect G-quadruplexes under molecular crowding conditions. *J. Phys. Chem. B*, **117**, 963–972.
18. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.A. and Yang,D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
19. Yu,Z., Schonhoft,J.D., Dhakal,S., Bajracharya,R., Hegde,R., Basu,S. and Mao,H. (2009) ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities. *J. Am. Chem. Soc.*, **131**, 1876–1882.
20. Koirala,D., Dhakal,S., Ashbridge,B., Sannohe,Y., Rodriguez,R., Sugiyama,H., Balasubramanian,S. and Mao,H. (2011) A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nat. Chem.*, **3**, 782–787.
21. Koirala,D., Mashimo,T., Sannohe,Y., Yu,Z., Mao,H. and Sugiyama,H. (2012) Intramolecular folding in three tandem guanine repeats of human telomeric DNA. *Chem. Commun.*, **48**, 2006–2008.
22. Mejia,Y.X., Mao,H., Forde,N.R. and Bustamante,C. (2008) Thermal probing of E. coli RNA polymerase off-pathway mechanisms. *J. Mol. Biol.*, **382**, 628–637.
23. Yin,H., Wang,M.D., Svoboda,K., Landick,R., Block,S.M. and Gelles,J. (1995) Transcription against an applied force. *Science*, **270**, 1653–1657.
24. Galburt,E.A., Grill,S.W., Wiedmann,A., Lubkowska,L., Choy,J., Nogales,E., Kashlev,M. and Bustamante,C. (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, **446**, 820–823.
25. Woodside,M.T., Behnke-Parks,W.M., Larizadeh,K., Travers,K., Herschlag,D. and Block,S.M. (2006) Nanomechanical Measurements of the Sequence-Dependent Folding Landscapes of Single Nucleic Acid Hairpins. *Proc. Natl Acad. Sci. USA*, **103**, 6190–6195.
26. Yu,Z. and Mao,H. (2013) Non-B DNA structures show diverse conformations and complex transition kinetics comparable to RNA or proteins-a perspective from mechanical unfolding and refolding experiments. *Chem. Rec.*, **13**, 102–116.
27. Yu,Z., Koirala,D., Cui,Y., Easterling,L.F., Zhao,Y. and Mao,H. (2012) Click chemistry assisted single-molecule fingerprinting reveals a 3D biomolecular folding funnel. *J. Am. Chem. Soc.*, **134**, 12338–12341.
28. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed. Engl.*, **40**, 2004–2021.
29. Mao,H. and Luchette,P. (2008) An integrated laser tweezers instrument for microanalysis of individual protein aggregates. *Sens. Actuat. B*, **129**, 764–771.
30. Smith,S.B., Cui,Y.J. and Bustamante,C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, **271**, 795–799.
31. Mills,J.B., Vacano,E. and Hagerman,P.J. (1999) Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). *J. Mol. Biol.*, **285**, 245–257.
32. Tinland,B., Pluen,A., Sturm,J. and Weill,G. (1997) Persistence length of single-stranded DNA. *Macromolecules*, **30**, 5763–5765.
33. Record,M.T.J., Anderson,C.F. and Lohman,T.M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quart. Rev. Biophys.*, **11**, 103–178.
34. Laurence,T.A., Kong,X., Jager,M. and Weiss,S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl Acad. Sci. USA*, **102**, 17348–17353.
35. Baumann,C.G., Smith,S.B., Bloomfield,V.A. and Bustamante,C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc. Natl Acad. Sci. USA*, **94**, 6185–6190.
36. Greenleaf,W.J., Frieda,K.L., Foster,D.A., Woodside,M.T. and Block,S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, **319**, 630–633.
37. Yu,Z., Gaerig,V., Cui,Y., Kang,H., Gokhale,V., Zhao,Y., Hurley,L.H. and Mao,H. (2012) The tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J. Am. Chem. Soc.*, **134**, 5157–5164.
38. Cheng,W., Arunajadai,S.G., Moffitt,J.R., Tinoco,I. Jr and Bustamante,C. (2011) Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. *Science*, **333**, 1746–1749.
39. Dietz,H. and Rief,M. (2006) Protein structure by mechanical triangulation. *Proc. Natl Acad. Sci. USA*, **103**, 1244–1247.
40. Jarzynski,C. (1997) Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **78**, 2690–2693.
41. Palassini,M. and Ritort,F. (2011) Improving free-energy estimates from unidirectional work measurements: theory and experiment. *Phys. Rev. Lett.*, **107**, 060601.
42. Dhakal,S., Schonhoft,J.D., Koirala,D., Yu,Z., Basu,S. and Mao,H. (2010) Coexistence of an ILPR i-motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level. *J. Am. Chem. Soc.*, **132**, 8991–8997.
43. Kypr,J., Kejnovska,I., Renciuk,D. and Vorlickova,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
44. Xu,L., Feng,S. and Zhou,X. (2011) Human telomeric G-quadruplexes undergo dynamic conversion in a molecular crowding environment. *Chem. Commun.*, **47**, 3517–3519.
45. Xue,Y., Kan,Z., Wang,Q., Yao,Y., Liu,J., Hao,Y. and Tan,Z. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-Quadruplex in K^+ solution under molecular crowding condition. *J. Am. Chem. Soc.*, **129**, 11185–11191.
46. Hall,D. and Minton,A.P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim. Biophys. Acta*, **1649**, 127–139.
47. Liphardt,J., Dumont,S., Smith,S.B., Tinoco,I. Jr and Bustamante,C. (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, **296**, 1832–1835.
48. Miyoshi,D., Nakamura,K., Tateishi-Karimata,H., Ohmichi,T. and Sugimoto,N. (2009) Hydration of Watson−Crick base pairs and dehydration of hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. *J. Am. Chem. Soc.*, **131**, 3522–3531.