Coexistence of an ILPR i-Motif and a Partially Folded Structure with Comparable Mechanical Stability Revealed at the Single-Molecule Level

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · JULY 2010

Impact Factor: 12.11 · DOI: 10.1021/ja100944j · Source: PubMed

6 AUTHORS, INCLUDING:

Joseph D Schonhoft
The Scripps Research Institute
13 PUBLICATIONS **245** CITATIONS
SEE PROFILE

Deepak Koirala
University of Chicago
25 PUBLICATIONS **325** CITATIONS
SEE PROFILE

Zhongbo Yu
Delft University of Technology
20 PUBLICATIONS **334** CITATIONS
SEE PROFILE

Hanbin Mao
Kent State University
52 PUBLICATIONS **1,719** CITATIONS
SEE PROFILE

Published in final edited form as:
J Am Chem Soc. 2010 July 7; 132(26): 8991–8997. doi:10.1021/ja100944j.

Coexistence of an ILPR i-motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level

Soma Dhakal, **Joseph D. Schonhoft**, **Deepak Koirala**, **Zhongbo Yu**, **Soumitra Basu**, and **Hanbin Mao**
Department of Chemistry, Kent State University, Kent, Ohio 44242

Hanbin Mao: hmao@kent.edu

Abstract

Investigation of i-motif is of high importance to fully understand the biological functions of G quadruplexes in the context of double stranded DNA. Whereas single molecule approaches have profiled G quadruplexes from a perspective unavailable by bulk techniques, there is a lack of similar literature on the i-motif in the cytosine (C) rich region complementary to G quadruplex forming sequences. Here, we have used laser tweezers to investigate the structures formed in 5′-$(TGTCCCCACACCCC)_2$, a predominate variant in the insulin linked polymorphic region (ILPR). We have observed two species with the change in contour length (ΔL) of 10.4 (±0.1) and 5.1 (±0.5) nm, respectively. Since ΔL of 10.4 nm is located within the expected range for an i-motif structure, we assign this species to the i-motif. The formation of the i-motif in the same sequence has been corroborated by bulk experiments such as Br_2 footprinting, circular dichroism, and thermal denaturation. The assignment of the i-motif is further confirmed by decreased formation of this structure (23 % to 1.3 %) with pH 5.5 →7.0, which is a well established behavior for i-motifs. In contrast to the i-motif, the formation of the second species with ΔL of 5.1 nm remains unchanged (6.1±1.6 %) in the same pH range, implying that pH sensitive $C:CH^+$ pairs may not contribute to the structure as significantly as those to the i-motif. Compared to the ΔG_{unfold} of i-motif (16.0 ±0.8 kcal/mol), the decreased free energy in the partially folded structure (ΔG_{unfold} 10.4 ± 0.7 kcal/mol) may reflect a weakened structure with reduced $C:CH^+$ pairs. Both ΔL and ΔG_{unfold} argue for the intermediate nature of the partially folded structure in comparison to the i-motif. In line with this argument, we have directly observed the unfolding of i-motif through the partially folded structure. The i-motif and the partially folded structure share similar rupture forces of 22-26 pN, which are higher than those that can stall transcription catalyzed by RNA polymerases. This suggests, from a mechanical perspective alone, that either of the structures can stop RNA transcription.

Introduction

In recent years, non-B DNA structures have attracted intensive research attention due to the likelihood that these structures may be responsible for a variety of human diseases such as Fredrick's ataxia and Huntington disease.[1, 2] Non-B DNA structures[3] include G quadruplex, Z DNA, H DNA, cruciform DNA, and i-motif, to name just a few. Experiments *in vitro* have

Correspondence to: Hanbin Mao, hmao@kent.edu.

Supporting Information **Available:** UV melting curve at pH 5.5; stable structures revealed by mfold program; calculation of nucleotides contained in DNA secondary structure; a representative unfolding curve with sequential unfolding events; and a ΔL histogram of a mutant DNA containing three C4 tracts at pH 5.5. This material is available free of charge via the Internet at http://pubs.acs.org.

suggested that these structures can regulate important biological processes including replication and transcription. Evidence from recent research has also indicated that these non-B DNA structures can cause DNA double strand breakage (DSB)[3], which increases mutations for diseases. Computer based programs have revealed that non-B DNA forming sequences are widely dispersed throughout the human genome. For example, a total of 188,836 G rich sequences have been found to be capable of forming G quadruplexes;[4] whereas a frequency of 1/3,050 bp[5] and 1/49,400 bp[6] exist for potential Z and H DNA in human genome, respectively. Some of these non-canonical DNA structures have been confirmed *in vivo*.[7-9]

Among these non-B DNA structures, G quadruplex and i-motif both contain four DNA strands. Each G-quadruplex forming sequence has a complementary C-rich sequence capable of adopting an i-motif structure. Thus, the prevalence of G quadruplex forming sequence in the human genome also suggests the wide-spread occurrence of i-motif in the genome. In contrast to the G quadruplex, which is independent of pH and readily forms at physiological pH, the formation of i-motif requires hemiprotonated cytosine-cytosine pairs[10] (Figure 1A) and therefore, is pH dependent. The optimal i-motif assembly occurs at pH 5.5, a value close to the pK_a for free cytosines.[11] So far, only scattered evidence[12, 13] has indicated the presence of i-motif structure at neutrality *in vitro*. *In vivo*, however, negative superhelicity[14] and molecular crowding[15, 16] may facilitate i-motif formation. In addition, formation of either an i-motif or a G-quadruplex can leave its complementary DNA as a free strand, thereby facilitates the assembly of the other.

Recent discovery of proteins interacting with potential i-motif structures suggests that, similar to the G quadruplexes, i-motifs may also have biological functions. Mergny and coworkers observed two nuclear proteins, hnRNP-K, a transcription factor of the *c-myc* gene, and ASF/SF2, a splicing factor, can bind specifically to the C-rich strand of human telomeres.[17] In an ILPR sequence similar to what has been used here, Gupta and coworkers[13] demonstrated that *E. coli* SSB protein binds more efficiently to the C-rich strand compared to other ssDNA regions at neutral pH. In human telomeres, Manzini and colleagues[18] have shown that Hela nuclear extract contains a protein that specifically binds to single stranded CCCTAA repeats. The specific binding of i-motif forming sequences to various proteins implies the potential biological function of the i-motifs that may form in these sequences. Another aspect of biological relevance of the i-motif comes from the potential interaction between this structure and the G quadruplex, the latter of which has shown critical involvement in many biological processes.[14] In accordance with the prospect of biological functions by i-motifs, several laboratories have started to investigate DNA i-motif analogs based on phosphorothioate or peptide backbones for potential pharmaceutical applications.[19]

Since a significant portion of i-motif hosting sequences is located within or downstream of promoter regions,[20, 21] there are ample opportunities for the tetraplex to interact with RNA polymerase (RNAP). As a motor protein, RNAP exerts a maximal load force, beyond which transcription stalls.[22, 23] From mechanical perspective alone, a non-B DNA structure with a mechanical stability higher than this maximal load force can stall RNAP. The mechanical stability of such a structure can be determined by single molecular methods such as laser tweezers.[24]

In this paper, we use laser-tweezers to investigate the structures formed in C-rich regions. To the best of our knowledge, i-motifs have never been investigated at the single molecular level. Compared to the bulk methods for i-motif investigation, such as circular dichroism (CD),[25, 26] gel electrophoresis,[27] ultraviolet (UV) absorbance,[27] X-ray;[28] and NMR,[29-31] single molecular methods can reveal biomolecular structures in a highly dynamic fashion.

The method is very sensitive in identifying small populations, such as intermediates during a folding process.[32] Compared to other single molecule techniques such as fluorescence,[33] the force based approaches do not require bulky fluorophores that may alter native structures. In addition, they can provide mechanical information of structures,[24, 34] which is important not only for transcription, but also for other processes catalyzed by motor proteins such as DNA polymerases.

We choose to investigate non-B DNA structures formed in the most prevalent variant, 5′-$(TGTCCCCACACCCC)_2$, in the insulin linked polymorphic region (ILPR). The region is known to affect the production of Human insulin protein.[35] The C rich DNA sequence used in our experiment is terminated by two dsDNA spacers. Similar to the internal loops in a DNA tetraplex,[36] terminal spacers have also demonstrated their critical role in the tetraplex conformation.[37] Thus, our approach complements very well with bulk methods that have limited sensitivity to long terminal spacers. Since the DNA regions in which tetraplexes are susceptible to form are almost always flanked by double stranded spacers,[4] our method is uniquely equipped to interrogate non-B DNA structure in a situation closer to *in vivo* conditions.

Our mechanical unfolding experiments have revealed two populations, an i-motif and a partially folded structure, in the DNA construct. The population ratio of these two species changes with pH, with partially folded structure predominating at pH 7. Both structures have similar unfolding forces (24.0 ±0.9 pN and 26.1 ±3.0 pN for i-motif and partially folded structures, respectively). These values are higher than the stall forces for known RNAPs, suggesting that the presence of either structure may interfere with the transcription process from the mechanical argument alone.

Materials and Methods

Materials

Olionucleotides were purchased from Integrated DNA Technologies (www.idtdna.com) and purified by denaturing PAGE. Unless specified differently, all of the chemicals were purchased from VWR (West Chester, PA).

Laser Tweezers Instrument

Detailed description of the laser tweezers instrument has been reported elsewhere.[38, 39] Briefly, a diode pumped solid state (DPSS) laser (1064 nm, 4 W, CW mode, BL-106C, Spectra-physics) was used as a trapping laser. *P* and *S* polarized laser light from the same laser source constituted two traps. The *S* polarized light was controlled by a steerable mirror (Nano-MTA, Mad City Laboratories) at a conjugate plane of the back focal plane of a focusing objective (Nikon CFI-Plan-Apochromat 60×, NA 1.2, water immersion, working distance ~320 μm). The exiting *P* and *S* polarized beams were collected by an identical objective and detected by two position sensitive photodetectors (PSD, DL100, Pacific Silicon Sensor) separately.[40] The force of the laser trap was calibrated by the Stokes force and thermal motion measurement. Both methods yielded a similar trap stiffness of ~307 pN/(μm × 100mW) (for 0.97 μm diameter polystyrene beads, Bangs Laboratory, Fishers, IN).

DNA Construct

The DNA construct was comprised of three fragments (Figure 1A, iii): two dsDNA spacers at two termini and one DNA fragment containing an i-motif forming oligonucleotides (bold and underscored in the sequence shown below) in the middle. The 651 bp dsDNA spacer was labeled with biotin, which was introduced through a biotinylated primer (Integrated DNA Technologies, IDT, Coralville, IA) during PCR amplification of pEIB plasmid

template (966bp).[40] This spacer was digested with XbaI restriction enzyme (New England Biolab, NEB). The 2690bp DNA handle was gel purified using a kit (Midsci, St. Louis, MO) after SacI and EagI digestions of a pEGFP plasmid (Clontech, Mountain View, CA). This spacer was subsequently labeled by digoxigenin (Dig) at 3′ end using 18 μM Dig-dUTP (Roche, Indianapolis, IN) and terminal transferase (Fermentas, Glen Burnie, MD). The middle i-motif forming fragment was constructed by annealing an oligonucleotide, 5′-CTAG AC GGTTGAAATACCGCACAGATGCG**TGTCCCCACACCCCTGTCCCCACACCCCT GT**GCCAGCAA GACGTA GCCCA G CGCGTC, with two other oligonucleotides, 5′-CGCATCTGTGCGGTATTTCACACCGT and 5′-GGCCGACGCGCTGGGCTACGTCTTGCTGGC at 97°C for 5 minutes and slowly cooled to room temperature for 6h. This i-motif forming fragment was ligated with the 651 bp DNA handle at one end, followed by a second ligation with the 2690 bp DNA handle using T4 DNA ligase (NEB). The final construct was purified by ethanol precipitation. The DNA pellet was dissolved in water and stored at -80°C. The mutant DNA construct containing **TGTCCCCACACCCCTGTCCCCACA** in place of the i-motif forming sequence was prepared using the same procedure.

Single Molecular Experiment

Anti-Dig antibody coated polystyrene beads (diameter: 2.17 μm, Spherotech, Lake Forest, IL) were incubated with diluted DNA constructs obtained above (~0.43ng/uL) in 100 mM KCl, 10 mM sodium phosphate buffer (pH 5.5, 6.0, 6.5 and 7.0) for 1 h at 23 °C to attach the DNA construct via the Dig/anti-Dig complex. Beads coated with streptavidin (diameter: 0.97 μm, Bangs Laboratory) were dispersed into the same buffer before injected into the reaction chamber. These two types of beads were trapped separately using two laser traps. To immobilize the DNA construct between the two beads, the bead already attached with DNA construct was brought close to the bead coated with the streptavidin by the steerable mirror. Once the DNA tether was trapped between the two beads, the Nano-MTA steerable mirror that controls the anti-Dig coated bead was moved away from the streptavidin-coated bead with a loading speed of ~5.5 pN/s. The secondary structure formed in the DNA molecule was unfolded when tension inside the tether was gradually increased. Rupture events with sudden change in the end-to-end distance were observed during the process (Figure 1B). Single tether was confirmed by a single breakage event when the DNA was overstretched. Change in contour length (ΔL) due to the rupture events was calculated by the two data points flanking the rupture events using an extensible Worm Like Chain (WLC) equation[41] (eqn1):

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

Here x is the end-to-end distance, k_B is the Boltzmann constant, T is absolute temperature, P is the persistent length (51.95 nm[39]), F is force and S is the elastic stretch modulus (1226 pN[39]). ΔL range of 2.2 to 14 nm was collected to construct the histogram of ΔL.

Calculation of Percentage Formation

The percentage formation of the i-motif and the partially folded structure was calculated based on the number of pulling curves recorded at a given pH. To assign individual pulling curves to a specific folded species, we used two methods. In the first method, we fit the overall ΔL distribution using a two-peak Gaussian function. The cross point between the two Gaussian populations was then used as a threshold to assign pulling curves to the two

species. To account for the stochastic behavior of individual pulling curves, in the second method, we randomly assigned the pulling curves in the intersection of the two Gaussian populations to a specific species according to the ratio determined by the two-peak Gaussian fitting. The percentage formation of an i-motif or a partially folded structure was calculated as the ratio of the pulling curves with the folded structure *vs* total pulling curves. To avoid repetitive counting, the subsequent pulling curves of the same DNA construct were discarded in this calculation. Although both methods yielded similar results, in this paper we adopted the second approach since it better represented the stochastic nature of individual molecules.

Calculation of ΔG_{unfold}

The free energy difference for unfolding of i-motif (ΔG_{unfold}) was calculated according to Jarzynski's equality equation (eqn 2) for non-equilibrium systems.[42, 43]

$$\Delta G = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp\left(-\frac{W_i}{k_B T}\right) \quad (2)$$

Where N is the number of observations in the experiment and W is the non-equilibrium work done during unfolding of the i-motif, which is equivalent of the hysteresis area between stretching and relaxing force-extension curves.[44] A total of 450 and 220 curves were used to calculate the ΔG_{unfold} of i-motif (pH range: 5.5 to 6.5) and partially folded structures (pH range: 5.5 to 7.0), respectively.

CD Spectroscopy

Oligonucleotides samples purchased from Integrated DNA Technologies were purified by denaturing PAGE and prepared in a concentration of 5 μM in 10 mM sodium phosphate buffer at a given pH with 100 mM KCl. The samples were then heated at 97 °C for 10 minutes and immediately cooled using an ice-water bath. The CD spectra were taken in a 1 mm quartz cuvette at room temperature with a Jasco-810 spectropolarimeter (Easton, MD). The reported spectra were the average of three scans with a scan rate of 50 nm/min. The spectrum of each scan was subtracted from a buffer and salt only baseline and smoothed using a Savitzky-Golay function. For CD melting, the signal was measured at 286 nm as the temperature was changed at a rate of 0.5 °C/minute using a Jasco (model PFD-425S) peltier temperature controller. The same rate was used for UV-melting measurement at 295 nm with a Varian Cary 300 spectrophotometer.

Br_2 Footprinting

DNA samples were radiolabeled at the 5′ end by incubating the DNA with T4 polynucleotide kinase (Promega) and [γ-P^{32}] ATP (Perkin Elmer) and purified using denaturing PAGE. To specifically probe cytosine residues in the footprinting experiment, molecular bromine was used as previously described.[45, 46] Solutions containing the 5′-labeled C-rich strand were prepared in 10 mM sodium phosphate buffer (pH 5.5 and 7.0), with the addition of 1 μM unlabeled oligonucleotides. The samples were then heated at 97 °C for 10 minutes and quickly cooled by submersion in an ice-water bath. Then to a 50 μl sample, 1 μl of 20 μM KBr was added, which was immediately followed by the addition of 1 μl of 10 μM $KHSO_5$. The reaction mixture was then incubated for 3 minutes at room temperature. To stop the reaction, 25 μl of the reaction mix was added to 100 μl of stop buffer (1 mg/ml sheared salmon sperm DNA, 300 mM NaCl, and 4 mM HEPES). The DNA was then precipitated using 100% ethanol and the resultant pellet was washed twice with 70% ethanol. The DNA was then cleaved with piperidine and the resulting fragments were

separated on a 10% denaturing polyacrylamide gel (19:1 bis:acrylamide). The gel was dried on Whatman™ paper, exposed to a phosphorimager screen, and scanned with a Typhoon 8600 instrument (Molecular Dynamics). Footprinting gel images were quantified using ImageJ™ software (http://rsbweb.nih.gov/ij/index.html). The intensity of each band was normalized with that of the cytosine in the ACA loop after background correction at pH 5.5 and 7.0 separately. The fold protection of each nucleotide was then calculated as the ratio between the normalized band intensity at pH 7.0 and that of corresponding band at pH 5.5.

Results and Discussion

Histograms of change in contour length (ΔL) show coexistence of an i-motif and a partially folded structure

To reduce the steric hindrance between the possible folded structure and dsDNA handles, we incorporated a wild type spacer, TGT, at the 3′ end of the most predominant variant sequence in the human ILPR, 5′-$(TGTCCCCACACCCC)_2$. After tethering the DNA construct between the two optically trapped beads *via* biotin/streptavidin and digoxigenin (Dig)/anti-Dig antibody complexes, respectively (Figure 1A, iii), we moved away one of the laser traps and recorded force-extension (F-X) curves of the DNA construct at pH 5.5 (Figure 1B). A sudden drop in force was clearly seen in the force-extension curve, which indicated the unfolding of a structure formed in the DNA construct. The folded structure was confirmed by CD-286 nm (Figure 2A) and UV-295 nm (Supplementary Figure S1) melting experiments, both of which showed a sigmoidal transition with a T_m of ~37 °C at pH 5.5 (see Table 1).

Using a worm-like-chain equation (eqn 1), we could obtain the change in contour length (Δ*L*, see Materials and Methods) from the force extension curves that contain rupture events of folded structures. After correction of end-to-end distance of a folded species, this Δ*L* reflects the contour length of the structure.[24] When we plotted the Δ*L* histogram, to our surprise, two populations were observed (Figure 1B inset). The bigger population has Δ*L* of 10.4 (±0.1) nm (see Figure 1B for a representative F-X curve in green); whereas the smaller population has Δ*L* of 5.1 (±0.5) nm (see Figure 1B for a representative F-X curve in red). The value of 10.4 nm is within the expected range for a folded structure with 25 nucleotides.[24, 47-51] Therefore, we ascribe this population to an i-motif structure. Previous NMR studies[13] have shown that i-motif forms in a similar sequence at pH 5.5 *in vitro*. CD spectra on the same DNA fragment at pH 5.5 clearly revealed a peak at ~286 nm and a trough at ~264 nm (Figure 2B), both of which are characteristic of an i-motif structure.[26, 52-54] In addition, the bromine footprinting data in Figure 2C demonstrated the protection of four C tracts (C4-C7, C11-C14, C18-C21, and C25-C28) from bromine initiated cleavage at pH 5.5. This result is consistent with the formation of an i-motif utilizing the four C tracts in the sequence.[14]

Compared to the Δ*L* of the i-motif, we refer the species with Δ*L* of 5.1 nm to a partially folded structure. Previously, NMR data[13] have suggested the presence of a partially folded structure in a similar C-rich sequence. However, the exact structure is not determined. Free energy calculations using mfold® program (see Supplementary Figure S2) ruled out the hairpin based conformation in this structure. Rather, our data were more complied with a triplex intermediate previously proposed for the formation of i-motif[30] or G quadruplex.[55, 56] Based on the known i-motif structure,[28, 57] our calculation (see Supplementary Materials) showed that the partially folded structure contains 18 (±1) nucleotides. This value was identical to the number of nucleotides (18 nts) contained in the triplex species (Supplementary Figure S3). Our observation that a small fraction of molecules (3.9 %, 26 out of 671 i-motif molecules) showed two sequential rupture events was consistent with the intermediate nature of the partially folded species (see a

representative curve in supplementary Figure S4). To provide further support that partially folded structure adopts a triplex conformation of 18 nts, we designed a DNA mutant that contains only three C4 tracts (see Material and Methods for sequence detail). When we mechanically unfolded this DNA at pH 5.5, the histogram of change in contour length showed only a single population with ΔL of 5.0 (±0.1) nm (supplementary Figure S5). This value was identical within experimental error to the partially folded species (5.1 ±0.5 nm), strongly suggesting the triple stranded nature of the partially folded species. However, to unambiguously determine the structure, other techniques such as NMR or X-ray need to be employed.

The fact that either an i-motif or a partially folded structure was observed in the same DNA construct strongly suggested the two populations were not due to the heterogeneity in sample preparation. Instead, it indicated the coexistence of two structures in the same C-rich DNA sequence.

Effect of pH on the two populations

i-Motif contains a stack of $C:CH^+$ pairs that are pH dependent (Figure 1A i & ii). It is expected that the higher the pH above the pK_a of cytosines (pK_a= 4.3 for free cytosines[11]), the more difficult the formation of the i-motif. This trend was clearly demonstrated in CD spectra at pH 5.5-7.0 (Figure 2B). While CD spectrum showed the existence of i-motif at pH 5.5, it suggested a mixture of unstructured DNA (peak at 277 nm)[45] and i-motif structures (the 286 nm peak) at pH 6.0. At pH 6.5 and 7.0, however, the CD spectra were more indicative of unstructured single stranded DNA. In accordance with the CD spectra, melting was clearly observed at pH 5.5, 6.0 and 6.5 (Figure 2A and Supplementary Figure S1), but not at pH 7.0 (not shown). The same trend was also obvious in Br_2 footprinting results, which strongly suggested the formation of i-motif at pH 5.5 but not at pH 7.0 (Figure 2C). However, caveat should be given that due to reduced sensitivity, the absence of melting in CD/UV measurement, or the absence of protection of C tracts in Br_2 footprinting cannot rule out the existence of a small fraction of i-motif or other folded structures in the sample.

If our assignment of the fully folded structure as an i-motif is correct, we should expect to see similar pH trend in single molecular studies. To this purpose, we mechanically unfolded the DNA construct at pH 5.5-7.0. When ΔL histograms at different pH were compared in Figure 3, it was obvious that two populations coexisted at each pH. ΔL for each species remained unchanged across the pH range (see Table 1), suggesting the structures of i-motif or partially folded species were intact when they formed within this pH range. We then calculated the percentage formation of each species at different pH (see Materials and Methods). When pH was increased from 5.5 to 7.0, we observed decreased formation of the species with the larger ΔL (23 %-1.3 %, Figure 4). Such clear pH dependence confirms our assignment that this species is indeed an i-motif. Interestingly, the formation of the partially folded structure remained constant (6.1±1.6 %) in pH 5.5-7.0, indicating the structure was not as pH sensitive as the i-motif.

The percentage formation of the i-motif was similar between pH 6.5 and 7.0 (2.9 %-1.3 %, Figure 4). However, the formation increased rather significantly below pH 6.5 (from 10.5 % at pH 6.0 to 23 % at pH 5.5). This trend corresponded strikingly well with the CD measurements. Between pH 6.5 and 7.0, the CD spectra indicated the majority of the sample was unstructured. However, below pH 6.5, the CD measurements showed significant increase in the i-motif structure (Figure 2B).

Our single molecular method can readily identify populations as low as ~2 % (Figure 4). This fact strongly demonstrates the superior sensitivity of our method over bulk approaches. As a result of this sensitivity, we were able to identify the formation of i-motif even at

neutral pH at room temperature. Although the formation of i-motif under neutral pH has been recently reported at 4°C,[58] its formation at room temperature and neutrality has not been observed without the introduction of template superhelicity,[14] the employment of molecular crowding conditions,[15] or the adoption of chemical modifications[59].

Since the pH dependence of the i-motif is originated from hemiprotonated cytosine-cytosine pairs, the reduced pH dependency observed here suggests that $C:CH^+$ pairs may not contribute to the partially folded structure as significantly as those in i-motif. This observation was consistent with the triplex model (see Supplementary Figure S3) which consists only half of the $C:CH^+$ pairs.

i-Motif and partially folded structures show similar unfolding forces higher than the stall force of RNAP

The single molecular nature of the laser tweezers method allows us to simultaneously survey the mechanical and thermodynamic stabilities of i-motif and partially folded structures. To this end, we separated the i-motif and the partially folded structures (see Materials and Methods) and investigated their respective stabilities.

When we plotted out the histogram of unfolding force for the i-motif and the partially folded structure at different pH (Figure 5 A&B and Table 1), we observed similar rupture force range for these two species (Figure 5C; i-motif: 21.9-26.2 pN; partially folded structure: 25.5-26.5pN). Majority of the rupture events were abrupt, suggesting cooperative unfolding for each species. The unfolding forces for partially folded structures were rather constant over pH, whereas those for i-motif decreased with pH. When we performed the 286 nm CD melting at different pH, we found melting temperature decreased monotonically with pH (see Figure 2A and Table 1). This observation is consistent with the trend of the rupture force for i-motif (Figure 5C), suggesting it is the i-motif structure that predominates the CD signal at 286 nm.

We then used Jarzynski's theorem[42, 43] to calculate the thermodynamic stability (ΔG_{unfold}) of the i-motif and partially folded structures (see Materials and Methods). The free energy change of i-motif (ΔG_{unfold} 16.0 (±0.8) kcal/mol) was close to that predicted from literature (14.5 kcal/mol, calculated according to eight $C:CH^+$ pairs[60]). This result further verified our assignment of the i-motif structure. The calculation also yielded a lowered free energy change for partially folded structure (10.4 (±0.7) kcal/mol). Such a result is consistent with our previous finding that $C:CH^+$ pairs have less contributions to the partially folded species than to the i-motif (see Figure 4 and related text). We surmise the decreased contribution from $C:CH^+$ pairs can lead to reduced H-bonds and $C:CH^+/C:CH^+$ stacking, either of which can reduce ΔG_{unfold} in the partially folded structure.

The unfolding forces are higher than the stall force of RNA polymerases, which ranges from < 20 pN for bacterial RNAP[22] to < 10 pN for Pol II.[23] This suggests, solely from a mechanical aspect, that either i-motif or partially folded structure can stall RNAP. Due to the fact that the partially folded species has a higher population than i-motif at neutral pH, the former species may play a more important role in the regulation of RNAP. Previously, we have observed the unfolding forces of 23 and 37 pN for parallel and antiparallel ILPR G quadruplexes, respectively.[24] Therefore, the mechanical stabilities of ILPR tetraplex DNA structures all seem to be higher than the stall force of RNAP. It remains to be seen whether high mechanical stability is a common theme for other G quadruplexes or i-motifs. For those tetraplexes formed downstream of promoter regions,[21] such a mechanical property can present a novel regulatory opportunity for transcription control through DNA secondary structure itself.

Conclusions

Using laser tweezers, we have demonstrated that an i-motif and a partially folded structure coexisted in the C-rich human ILPR oligonucleotides. The formation of i-motif is decreased with increasing pH, while that of partially folded structure is pH independent. Both the i-motif and the partially folded structure have unfolding forces higher than the stall forces of RNA polymerases, suggesting either of the structure can stop transcription from a mechanical perspective alone. To the best of our knowledge, single molecular investigation on i-motif structures has not been reported before. The methodology described here offers a novel tool to interrogate i-motif structures from a unique mechanical perspective at the single-molecule level.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

H.M. thanks the New Faculty Award Program at Camille and Henry Dreyfus Foundation, a KSU startup, and Ohio Board of Regents for financial support. S.B. acknowledges a KSU startup fund and an Ohio Board of Regents fund. This work is partially supported by NIH R15 DK081191-01 (to H.M. and S.B.).

References

1. Mirkin SM. Curr Opin Struct Biol. 2006; 16:351–358. [PubMed: 16713248]
2. Wells, RD.; Ashizawa, T. Genetic Instabilities and Neurological Diseases. 2nd. Elsevier; Burlington, MA: 2006.
3. Wells RD, Dere R, Hebert ML, Napierala M, Son LS. Nucleic Acids Res. 2005; 33:3785–3798. [PubMed: 16006624]
4. Huppert JL, Balasubramanian S. Nucleic Acids Res. 2005; 33:2908–2916. [PubMed: 15914667]
5. Schroth GP, Chou PJ, Ho PS. J Biol Chem. 1992; 267:11846–11855. [PubMed: 1601856]
6. Schroth GP, Ho PS. Nucleic Acids Res. 1995; 23:1977–1983. [PubMed: 7596826]
7. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A. Proc Natl Acad Sci USA. 2001; 98:8572–8577. [PubMed: 11438689]
8. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Genes Dev. 2004; 18:1618–1629. [PubMed: 15231739]
9. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Nat Struct Mol Biol. 2005; 12:847–854. [PubMed: 16142245]
10. Gehring K, Leroy JL, Gueron M. Nature. 1993; 363:561–564. [PubMed: 8389423]
11. Nonin S, Leroy JL. J Mol Biol. 1996; 261:399–414. [PubMed: 8780782]
12. Han X, Leroy JL, Gueron M. J Mol Biol. 1998; 278:949–965. [PubMed: 9600855]
13. Catasti P, Chen X, Deaven LL, Moyzis RK, M BE, Gupta G. J Mol Biol. 1997; 272:369–382. [PubMed: 9325097]
14. Sun D, Hurley LH. J Med Chem. 2009; 52:2863–2874. [PubMed: 19385599]
15. Xue Y, Kan Z, Wang Q, Yao Y, Liu J, Hao Y, Tan Z. J Am Chem Soc. 2007; 129:11185–11191. [PubMed: 17705383]
16. Miyoshi D, Matsumura S, Nakano S, Sugimoto N. J Am Chem Soc. 2003; 126:165–169. [PubMed: 14709080]
17. Lacroix L, Lienard H, Labourier E, Djavaheri-Mergny M, Lacoster J, Leffers H, Tazi J, Helene C, Mergny JL. Nucleic Acids Res. 2000; 28:1564–1575. [PubMed: 10710423]
18. Marsich E, Piccini A, Xodo LE, Manzini G. Nucleic Acids Res. 1996; 24:4029–4033. [PubMed: 8918808]
19. Diederichsen U. Angew Chem Int Ed. 1998; 37:2273–2276.

20. Huppert JL, Balasubramanian S. Nucleic Acids Res. 2007; 35:406–413. [PubMed: 17169996]
21. Kikin O, Zappala Z, D'Antonio L, Bagga PS. Nucleic Acids Res. 2008; 36:D141–D148. [PubMed: 18045785]
22. Mejia YX, Mao H, Forde NR, Bustamante C. J Mol Biol. 2008; 382:628–637. [PubMed: 18647607]
23. Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C. Nature. 2007; 446:820–823. [PubMed: 17361130]
24. Yu Z, Schonhoft JD, Dhakal S, Bajracharya R, Hegde R, Basu S, Mao H. J Am Chem Soc. 2009; 131:1876–1882. [PubMed: 19154151]
25. Bucek P, Jaumot J, Avino A, Eritja R, Gargallo R. Chem Eur J. 2009; 15:12663–12671.
26. Manzini G, Yathindra N, Xodo LE. Nucleic Acids Res. 1994; 22:4634–4640. [PubMed: 7984411]
27. Li W, Miyoshi D, Nakano S, Sugimoto N. Biochemistry. 2003; 42:11736–11744. [PubMed: 14529284]
28. Cai L, Chen L, Raghavan S, Ratliff R, Myozis R, Rich A. Nucleic Acids Res. 1998; 26:4696–4705. [PubMed: 9753739]
29. Phan AT, Gueron M, Leroy JL. J Mol Biol. 2000; 299:123–144. [PubMed: 10860727]
30. Leroy JL. Nucleic Acids Res. 2009; 37:4127–4134. [PubMed: 19433505]
31. Dai J, Ambrus A, Hurley LH, Yang D. J Am Chem Soc. 2009; 131:6102–6104. [PubMed: 19400591]
32. Cecconi C, Shank EA, Bustamante C, Marqusee S. Science. 2005; 309:2057–2060. [PubMed: 16179479]
33. Shirude PS, Okumus B, Ying L, Ha T, Balasubramanian S. J Am Chem Soc. 2007; 129:7484–7485. [PubMed: 17523641]
34. Schonhoft JD, Bajracharya R, Dhakal S, Yu Z, Mao H, Basu S. Nucleic Acids Res. 2009; 37:3310–3320. [PubMed: 19324891]
35. Bell GI, Selby MJ, Rutter WJ. Nature. 1982; 295:31–35. [PubMed: 7035959]
36. Kumar N, Sahoo B, Varun KAS, Maiti S, Maiti S. Nucleic Acids Res. 2008; 36:4433–4442. [PubMed: 18599514]
37. Zhang Z, Dai J, Valiath E, Jones RA, Yang D. Nucleic Acids Res. 200910.1093/nar/gkp1029
38. Luchette P, Abiy N, Mao H. Sens Actuat B. 2007; 128:154–160.
39. Mao H, Luchette P. Sens Actuat B. 2008; 129:764–771.
40. Mao H, Arias-Gonzalez JR, Smith B, Tinoco I Jr, Bustamante C. Biophys J. 2005; 89:1308–1316. [PubMed: 15923237]
41. Baumann CG, Smith SB, Bloomfield VA, Bustamante C. Proc Natl Acad Sci USA. 1997; 94:6185–6190. [PubMed: 9177192]
42. Jarzynski C. Phys Rev Lett. 1997; 78:2690–2693.
43. Liphardt J, Dumont S, Smith SB, Tinoco I Jr, Bustamante C. Science. 2002; 296:1832–1835. [PubMed: 12052949]
44. Greenleaf WJ, Frieda KL, Foster DAN, Woodside MT, Block SM. Science. 2008; 319:630–633. [PubMed: 18174398]
45. Kendrick S, Akiyama Y, Hecht SM, Hurley LH. J Am Chem Soc. 2009; 131:17667–17676. [PubMed: 19908860]
46. Ross SA, Burrows CJ. Nucleic Acids Res. 1996; 24:5062–5063. [PubMed: 9016685]
47. Mills JB, Vacano E, Hagerman PJ. J Mol Biol. 1999; 285:245–257. [PubMed: 9878403]
48. Tinland B, Pluen A, Sturm J, Weill G. Macromolecules. 1997; 30:5763–5765.
49. Record MTJ, Anderson CF, Lohman TM. Quart Rev Biophys. 1978; 11:103–178.
50. Laurence TA, Kong X, Jager M, Weiss S. Proc Nat Acad Sci USA. 2005; 102:17348–17353. [PubMed: 16287971]
51. Hingerty BE, Broyde SB, Olson WK. Biopolymers. 1982; 21:1167–1188. [PubMed: 6178444]
52. Khan N, Avino A, Taylor R, Gonzalez C, Eritja R, Gargallo R. Biochemie. 2007; 89:1562–1572.
53. Mathur V, Verma A, Maiti S, Chowdhury S. Biochem Bioph Res Co. 2004; 320:1220–1227.

54. Pataskar SS, Dash D, Brahmachari SK. J Biomol Struct Dyn. 2001; 2:307–313. [PubMed: 11697735]
55. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Nucleic Acids Res. 2006; 34:2723–2735. [PubMed: 16714449]
56. Mashimo T, Sannohe Y, Yagi H, Sugiyama H. Nucleic Acids Symp Ser. 2008; 52:409–410.
57. Esmaili N, Leroy JL. Nucleic Acids Res. 2005; 33:213–224. [PubMed: 15647504]
58. Zhou J, Wei C, Jia G, Wang X, Feng Z, Li C. Mol Biosyst. 2010; 6:580–586. [PubMed: 20174686]
59. Kumar N, Petersen M, Maiti S. Chem Commun. 2009:1532–1534.
60. Leroy JL, Gehring K, Kettani A, Gueron M. Biochemistry. 1993; 32:6019–6031. [PubMed: 8389586]

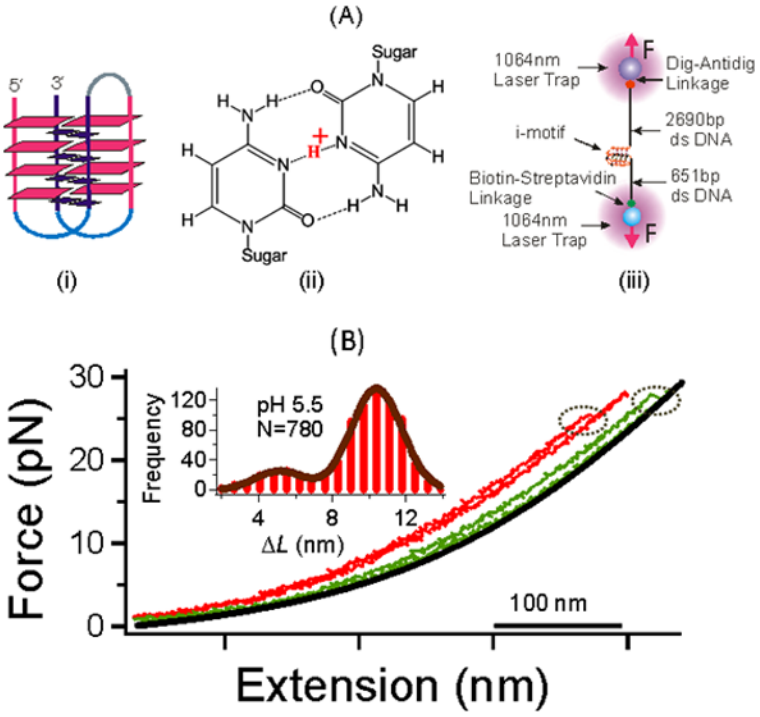

Figure 1.

(A) Schematic of (i) i-motif, (ii) chemical structure of a hemiprotonated cytosine-cytosine pair, and (iii) experimental set up. (B) Typical force extension curves from the same ILPR sequence, 5′-(TGTCCCCACACCCC)$_2$TGT, at 23 °C and pH 5.5. Green (the right curve) and red (the left curve) curves represent unfolding of two populations with change in contour length (ΔL) of 10.4 and 5.1 nm, respectively. The sudden drop in force (around 27.5 and 26.0 pN, indicated by dashed circles) in the extending curves indicates unfolding events. Black line is the WLC fitting. Curves are shifted in *x* axis for clarity. Inset is the ΔL histogram fitted by a two-peak Gaussian.

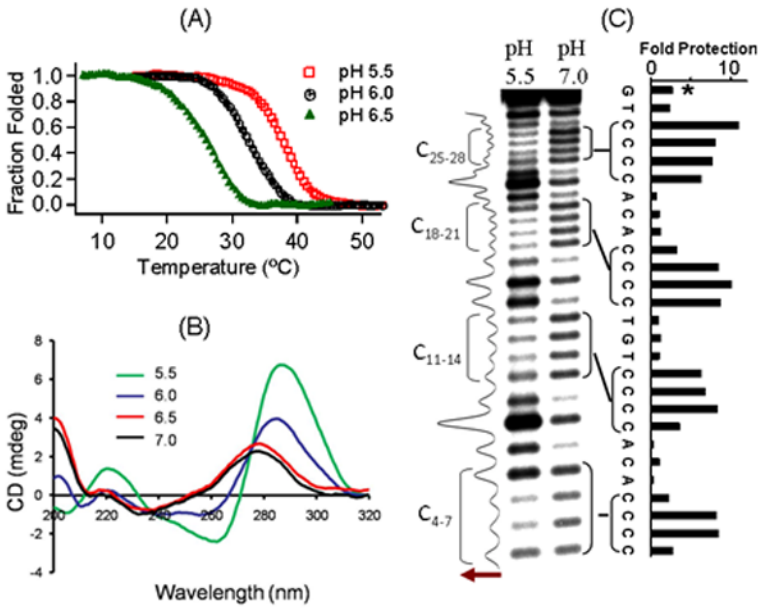

Figure 2.

CD experiments and Br_2 footprinting of the ILPR C-rich sequence, 5′-$(TGTCCCCACACCCC)_2TGT$. (A) CD (286 nm) melting curves of the i-motif forming sequence from pH 5.5 to 6.5. The spectra were baseline corrected and normalized as the fraction of folded form. At pH 7.0 no clear transition was observed (not shown). (B) CD spectra acquired from pH 5.5 to 7.0 at 23 °C. CD experiments were performed with 5 μM oligonucleotides in 10 mM sodium phosphate buffer (pH 5.5, 6.0, 6.5, or 7.0) and 100 mM KCl. (C) Br_2 footprinting at pH 5.5 and 7.0. The trace on the left side of the gel depicts band quantitation. The direction of the arrow at the bottom of the gel indicates increased intensity of the bands. The plot on the right shows the fold protection of individual bands at pH 5.5 compared to those at pH 7.0. The band intensities at the 3′ end (see *) are not reliable since they are close to the uncut oligonucleotides.

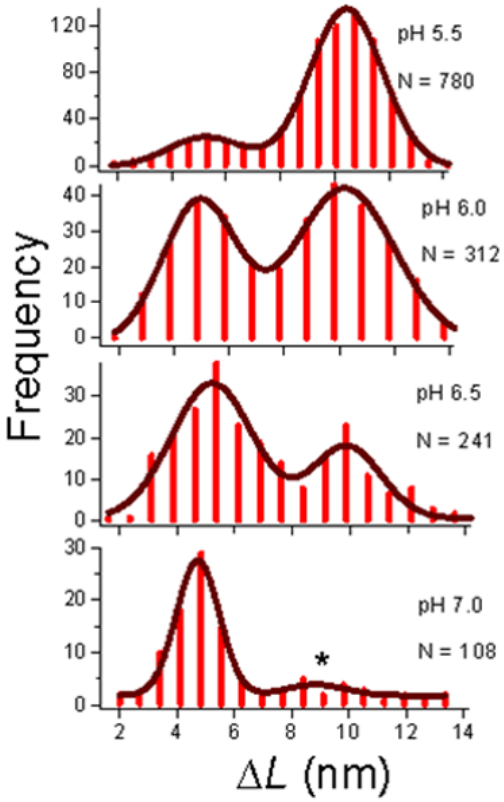

Figure 3.

Δ*L* histograms at different pH (23 °C). Two populations shown from pH 5.5 to 7.0 are fitted with two-peak Gaussian (solid lines). The fitting for the second peak (see *) at pH 7.0 is not reliable due to insignificant formation of the i-motif at this pH (see Figure 4). N depicts number of experiments at each pH.

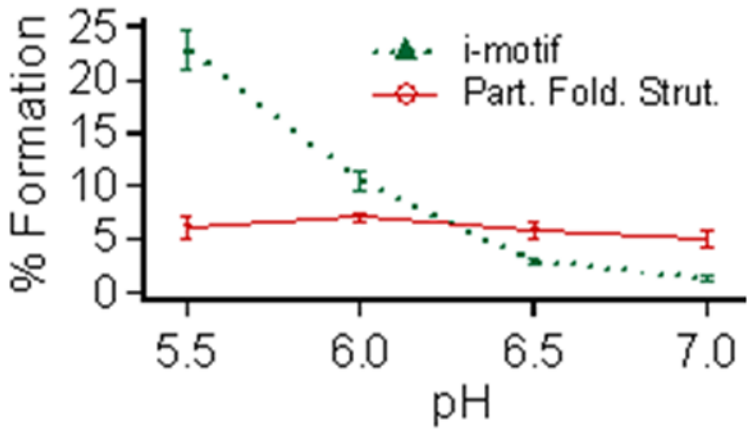

Figure 4.
The percentage formation of the i-motif (dotted line) and the partially folded structure (solid line) at different pH under 23 °C.

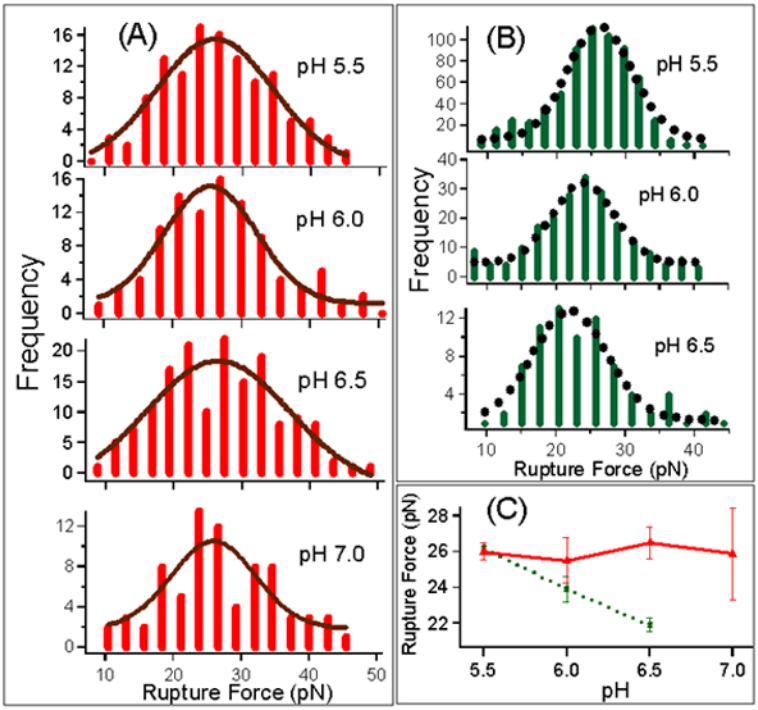

Figure 5.

Rupture force histograms of (A) partially folded structure and (B) i-motif at different pH under 23 °C. The solid lines in (A) and the dotted lines in (B) are Gaussian fits. The rupture force histogram of i-motif at pH 7.0 is not shown due to insignificant formation of the structure at this pH. (C) The rupture forces (obtained from Gaussian peaks in (A) and (B)) *vs* pH for the i-motif (green dotted line) and partially folded structure (red solid line).

Table 1

Summary of change in contour length (ΔL), rupture force (F), free energy change of unfolding (ΔG), and 286 nm CD melting temperature (T_m) at pH 5.5-7.0. The values in paranthesis are standard deviations.

pH	$\Delta L_{\text{Part. Fold.}}$ (nm)	$F_{\text{Part. Fold}}$ (pN)	$\Delta G_{\text{part. fold.}}$ (kcal/mol)	$\Delta L_{\text{i-motif}}$ (nm)	$F_{\text{i-motif}}$ (pN)	$\Delta G_{\text{i-motif}}$ (kcal/mol)	T_m (°C)
5.5	5.1 (±0.5)	26.5 (± 0.5)	10.5 (± 0.1)	10.4 (±0.1)	26.2 (± 0.3)	16.6 (± 0.7)	37.0 (±1.3)
6	4.9 (±0.3)	25.5 (± 1.3)	10.7 (± 0.1)	10.2 (±0.1)	23.9 (± 0.7)	16.2 (± 0.3)	33.7 (±0.5)
6.5	5.2 (±0.3)	26.5 (± 0.9)	10.4 (± 0.6)	9.9 (±0.2)	21.9 (± 0.4)	15.1 (± 0.1)	23.4 (±0.7)
7	4.7 (±0.3)	25.9 (± 2.6)	9.9 (± 0.3)	Not enough data	Not enough data	Not enough data	No melting