Prakash Shrestha, Ph.D., January 2018 Chemistry and Biochemistry

FOLDING DYNAMICS OF G-QUADRUPLEXES DURING TRANSCRIPTION AND IN A NANO-CONFINEMENT

Dissertation Advisor: Dr. Hanbin Mao

The G-quadruplexes are formed by folding of the repetitive sequence of guanine bases in a nucleic acid strand. Bioinformatic analyses have revealed that such G-rich motifs are prevalent inside the genome. Interestingly, these G-quadruplex motifs are more concentrated in the human promoter and telomeric regions. Recent reports have indicated that the DNA and RNA guanine-tracts intermix to form DNA/RNA hybrid G-quadruplex (HQ). The formation of HQs shows that, during the transcription of such G-rich motifs, a complex population dynamics between profoundly different conformations of DNA G-quadruplex (GQ), RNA G-quadruplex (RQ) and DNA/RNA hybrid G-quadruplex (HQ) does exist. The prevalent structure in equilibrium has a significant role in biological processes. Also, such biomolecular structures are highly confined inside the nucleus. Therefore, understanding of the transition dynamics of G-quadruplexes under these cellular conditions is essential. This dissertation has uncovered the previously unknown folding dynamics of G-quadruplexes in two closely cellular states at single molecule level. First, the deconvolution of a complex population equilibrium in the G-rich sequence downstream of transcription-start-sites (TSS) and second, the folding dynamics of a human telomeric DNA G-quadruplex inside a biomimetic confined space created by using DNA origami nanocage.

First, using a T7 transcription model, we found that GQ and HQ form in a natural sequence, $(GGGGA)_4$, downstream of many transcription-start-sites. Using a newly-developed Single Molecule Stalled-transcription Assay, we revealed that RNA transcripts helped to populate quadruplexes at the expense of duplexes. Among G-quadruplexes, HQ predominates GQ in population and mechanical stabilities, suggesting HQ may serve as a better mechanical block during transcription. The fact that HQ and GQ folded within tens of milliseconds in the presence of RNA transcripts justified the co-transcriptional folding of these species. The catalytic role of RNA transcripts in the GQ formation was strongly suggested as the GQ folded >7 times slower without transcription. These results shed light on the possible synergistic effect of GQs and HQs on transcriptional controls.

Next, using DNA origami nanocages, we show the real effect of confined space on the property of individual human telomeric DNA G-quadruplexes. We induce targeted mechanical unfolding of the G-quadruplex while leaving the nanocage unperturbed. We find that the mechanical and thermodynamic stabilities of the G-quadruplex inside the nanocage increase with decreasing cage size. Compared to the case of diluted or molecularly crowded buffer solutions, the G-quadruplex inside the nanocage is significantly more stable, showing 100 times faster folding rate. Our findings suggest the possibility of co-replicational or co-transcriptional folding of G-quadruplex inside the polymerase machinery in cells.

FOLDING DYNAMICS OF G-QUADRUPLEXES DURING TRANSCRIPTION AND IN A NANO-CONFINEMENT

DISSERTATION

Submitted to Kent State University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemistry and Biochemistry

BY

Prakash Shrestha

January 2018

Dissertation written by

Prakash Shrestha

B. Sc., Tribhuvan University, 2006

M. Sc., Tribhuvan University, 2008

Ph.D., Kent State University, 2017

Approved by

Hanbin Mao, Ph.D., Chair, Doctoral Dissertation Committee

Mietek Jaroniec, Ph.D., Member, Doctoral Dissertation Committee

Bansidhar Datta, Ph.D., Member, Doctoral Dissertation Committee

Helen Piontkivska, Ph.D., Member, Doctoral Dissertation Committee

Derek Damron, Ph.D., Member, Doctoral Dissertation Committee

Accepted by

Soumitra Basu, Ph.D., Chair, Department of Chemistry & Biochemistry

James L. Blank, Ph.D., Dean, College of Arts and Sciences

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

At first, I would like to express my sincere gratitude and deepest appreciation to my advisor Dr. Hanbin Mao, who has promptness, persistence, and the intelligence. His watchfulness on the worst possible, promptness in problem-solving, and audacity of the hope in research are always inspiring for me. Without his continuous guidance and motivation, this dissertation would not have been possible. During these five years of my Ph.D. career, his mentorship promoted me to a diligent, confident and a good researcher.

I would like to thank my committee members, Dr. Mietek Jaroniec, Dr. Bansidhar Datta, and Dr. Helen Piontkivska for their continuous support throughout my Ph.D. and valuable suggestions and feedback on my research proposal, prospectus, and dissertation.

I would like to thank all the scientists in Dr. Mao's lab for the great science and society together. I am grateful to the past lab members Dr. Soma Dhakal, Dr. Deepak Koirala, Dr. Philip Yangyuoru, Dr. Yunxi Cui, and Dr. Chiran Ghimire for their help, and valuable suggestions to my research. I learned from Sangeetha Selvam, Jibin Abraham Punnoose, Shankar Mandal, Sagun Jonchhe, Mohammed Enamul Hoque, Mohammad Akter Hossain, and Shankar Pandey. So, thank you all. I also thank REU student, William J. Maximuck for his contribution to my projects during his summer research time.

I would like to acknowledge Professor Hiroshi Sugiyama and Professor Masayuki Endo, Kyoto University, Japan for providing me an excellent opportunity to work in their

labs in a collaborative project. The opportunity to work in their lab was a great opportunity for me to learn DNA origami nanotechnology. I appreciate their comments and suggestions to improve my research skills. I would like to thank all the group members of Professor Sugiyama's lab for their help and support during my stay in Japan. Also, I would like to thank Professor Zheng Tan and Shan Xiao for great efforts to accomplish the project on G-quadruplexes during transcription.

I thank the Department of Chemistry & Biochemistry for providing me a great research platform. It was convenient to use every available research facilities. I would like to thank Dr. Mahinda Gangoda and Larry Maurer for helpful technical advices. Also, I would like to thank Erin Michael, and Janie Viers for their immense help and co-operation.

I would like to thank my parents, Chandraman Pradhan Shrestha and Ram Kumari Shrestha, sisters Sona Shrestha and Pabitra Shrestha and brother Shakti Kumar Shrestha for their love and supports in every step of my life. Special thanks to my wife Niranjana Shrestha for her support, encouragement, and love in every ups and downs. She is an excellent life-partner and past several years would not have been an easy journey without her regarding both academic and personal life. My son Pranidh Shrestha, he is my love and ultimate source of energy to begin my day with an adorable smile.

Prakash Shrestha

January 2, 2018

CHAPTER I

INTRODUCTION AND BACKGROUND

A PART OF THIS CHAPTER HAS BEEN PUBLISHED AS AN ARTICLE IN REVIEWS IN CHEMPHYSCHEM JOURNAL, ChemPhysChem **2015**, *16,* 1829. *ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED TO THE COPYRIGHT PERMISSION FROM WILEY-VCH.*

Besides the well-known canonical double-helix structures of nucleic acids, various tertiary structures of both DNA and RNA such as quadruplexes[1], cruciform[2], A-minor motif[3], ribose zipper[4], etc. are known to have critical biological functions. In recent years, G-quadruplex is one of the highly investigated tertiary structures because of their significant roles in physiological processes such as genetic protection and regulations.[5-9] Recent reports have suggested that DNA and RNA G-quadruplex structures are potential therapeutic targets of cancer and cardiovascular diseases.[10-16] Moreover, because of their highly specific interaction with metal ions and small-molecule compounds, G-quadruplex structures are of high interest in the field of biosensing[17-19], and nanobiotechnology[20-21].

1.1 Types of G-quadruplex structures

In G-rich motifs of DNA and RNA, the guanine bases are assembled by Hoogsteen hydrogen bonding between N_7 and C_6 amino group of the adjacent guanines in a cyclic array to form a square planar structure called Guanine tetrad or G-quartet (Figure 1.1A). In a G-quartet, the O_6 carbonyl of each guanine base coordinates with the monovalent metal cations such as Na^+ or K^+ to further stabilize the structure.[22-25] Two or

more of these planar G-quartets stack on the top of each other to form a more stable tetraplex helical structure called G-quadruplex (Figure 1.1B-I).[22, 26-28]

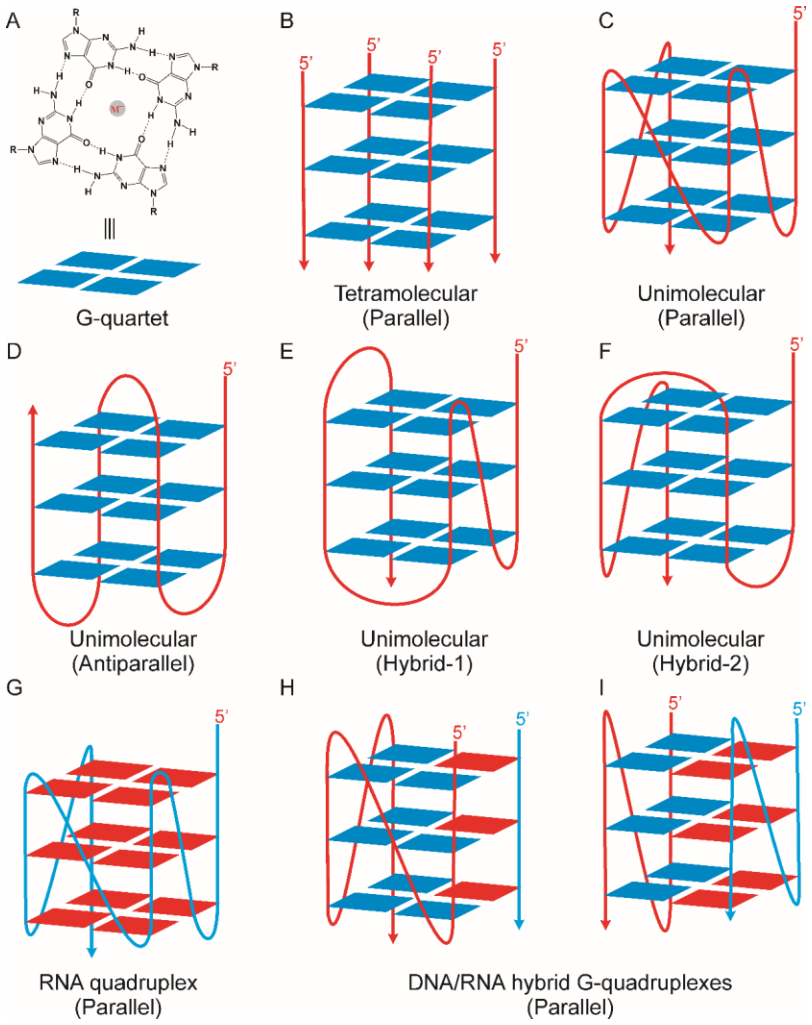


Figure 1.1. Structure and various conformations of G-quadruplexes. Arrows represent the direction of DNA and RNA strands from 5' to 3' end. Blue and red tiles represent the guanine residues of DNA (red lines) and RNA (blue lines) strands, respectively.

There are several types of G-quadruplexes based on the number, orientation, and nature of the nucleic acid strands involved.[29-31] Depending on the number of strands involved, G-quadruplexes may be intermolecular (Figure 1.1B) or intramolecular (Figure 1.1C-I).[29-34] Based on the different orientation of DNA strands involved, there are various conformations of DNA G-quadruplex (GQ),[27,35] basically these are the parallel or antiparallel conformations. All strands are oriented into the same direction in parallel conformation (Figure 1.1B,C,G,H,I) whereas each strand is oriented in opposite orientation to the other two adjacent strands in antiparallel conformation (Figure 1.1D-F). The conformations with the mixed parallel and antiparallel orientation of the strands are also known (Figure 1.1E,F)[36]. The specific conformation of an intramolecular G-quadruplex is governed by the DNA sequence[37], the loop size[38-39], and the ions in buffer[40-41] and molecular crowding conditions[42-44]. Usually, RNA G-quadruplex (RQ) has parallel conformation[45-46]. Recently, it has been reported that the G-rich DNA and RNA strands can intermix to form DNA/RNA hybrid G-quadruplex (HQ)[47-48]. The number of guanine-tracts from DNA and RNA may vary from one to three in such G-quadruplexes[49].

1.2 Biological significance of G-quadruplexes

Repetitive guanine tracts contained nucleic acids are the potential motifs to fold into the G-quadruplex structure.[50] Bioinformatic analysis has revealed the prevalence of likely G-quadruplex forming sequences in both prokaryotic and eukaryotic genomes.[51-55] Several studies have reported that there are about 376000 sites that could simultaneously

form G-quadruplexes in the human genome.[56-57] Interestingly, these sites are spotted in the crucial regions of the human genome such as promoter and telomere regions,[54, 57-60]. Also, these G-quadruplex forming sequences are found to be enriched at the downstream of transcription start sites (TSS).[61]

Human telomere region has the highest prevalence of potential G-quadruplex motifs as there is a long tandem array of TTAGGG repeat sequence in both duplex and single-stranded sections. Recently, Biffi *et al.* visualized the formation of G-quadruplex in human telomere region by using structure-specific antibody.[62] Several reports have suggested that the formation of G-quadruplexes in the human telomere sequence inhibited the expression of the telomerase enzyme.[63] In cancer cells, the telomerase enzyme was found to be overexpressed by more than 85%[64], which has brought the researcher's attention to human telomere G-quadruplexes might be the target of cancer therapeutics. Therefore, the researchers have been investigating to develop cancer drugs that would selectively target the human telomere G-quadruplex and inhibit the overexpression of telomerase.[65-66]

Recent reports on the formation of stable DNA/RNA hybrid G-quadruplexes *in vitro* have suggested that even fewer repetitive guanine bases in the genomic sequences, are also GQ motifs. Bioinformatic analyses have revealed that more than 97% of the human gene contains potential hybrid G-quadruplexes forming motifs and the numbers are correlated with the transcriptomic profiles of human tissues.[61] Also, these G-rich sequences are concentrated in the non-template strands of the duplex DNA downstream

of the TSS, which was found favoring the formation of HQ. Interestingly, comparative genome sequencing analyses of various species showed that HQ forming motifs are found to be evolutionally selected to function in transcription mediated processes in warm-blooded animals.[61] The reports on HQ formation during the transcription of G-rich sequences downstream of the TSS have suggested that GQ and HQ may have a synergistic role to regulate the genetic expression.[49] Also, the evolutionarily conserved HQ forming sequences have indicated that a complex population dynamics between different types of DNA structures including G-quadruplexes does exist to control the biological processes.

Due to the prevalence of G-quadruplex motifs in the crucial regions of the human genome, the G-quadruplexes are considered as critical regulatory units for many biological processes such as replication, transcription, and translation. Several studies have reported that G-quadruplex formation in the promoter regions acts as an on/off switch to modulate the transcription process.[12, 67] This has been directly related to the expression level of related genes. Many promoter sequences of cancer-related genes such as BCl-2, c-Myc, etc. have enrichment of GQ motifs, and hence these G-quadruplexes could be a potential target to cure cancer.[68-69]

1.3 Background of G-quadruplex studies

Given the above discussed biological significances, G-quadruplexes have been investigated by using several biochemical and biophysical methods ranging from ensemble to single molecule analysis. The commonly used ensemble approaches are

circular dichroism (CD) spectroscopy[70], UV-vis spectroscopy[71], fluorescence spectroscopy[72], nuclear magnetic resonance (NMR)[26, 73], X-ray crystallography[74], gel-shift assay and chemical footprinting[75]. The polymorphic nature of G-quadruplex has been confirmed by analyzing their crystal structure using X-ray crystallography and NMR spectroscopy techniques. These methods have reported the binding interactions between G-quadruplex structures and small-molecule or protein ligands. The binding interactions of ligand and G-quadruplex are studied by monitoring its melting temperature with and without ligand by using CD spectroscopy, UV spectroscopy and fluorescence resonance energy transfer (FRET) techniques. Recently, mass spectrometry (MS) with advanced ionization techniques[76-77] and capillary electrophoresis (CE)[78-79] have been used to reveal the molecular insights of the G-quadruplexes and their binding interaction with ligands.

Despite having multiple advantages to use above discussed methods to investigate G-quadruplex and its interaction with ligands, these methods cannot probe the rare sub-populations, which often are important in biological functions. As these techniques acquire the ensemble-averaged signals, they cannot inherently deconvolute the complex population equilibrium between various GQ conformations with their intermediates. For example, it is possible to form several types of G-quadruplexes; such as DNA G-quadruplex, RNA G-quadruplex, and their hybrid G-quadruplexes, during the transcription of a GQ-motif. As a result, there might be a complex equilibrium dynamics among these structures, which cannot be deconvoluted by using the ensemble biochemical methods. However, this can be easily accessed by using single molecule methods such as optical tweezers.[80] Therefore, single-molecule methods not only

complement the ensemble biochemical methods to acquire a complete information of a biological system but also have been advancing their power to visualize the biomolecular interactions at the single molecule level and observe the molecular events in real time.

Novel single molecule methods integrated with fluorescence resonance energy transfer (FRET)[81-87], atomic force microscopy (AFM)[88-89] have been developed to understand the folding dynamics of G-quadruplex in diverse conditions. Alternatively, there are several single molecule methods based on the force measurements. For example, atomic force microscope,[90-91] optical tweezers,[92-94] magnetic tweezers,[95-96] and opto-magnetic tweezers[97] can determine the forces involved in single molecular interactions. The mechanical force is a universal variable that can be used to describe both microscopic and macroscopic processes; many biological reactions can also be characterized regarding mechanical forces.[98] In fact, biochemical processes of several proteins such as polymerases, myosin,[99] kinesins,[93] etc. generate forces of a few picoNewtons. In other words, picoNewton amount of forces have been shown to fold or unfold proteins, DNA, and RNA structures. This dissertation discusses the use of optical tweezers to study the property of G-quadruplexes during transcription[49] and in a nano-confined environment[100].

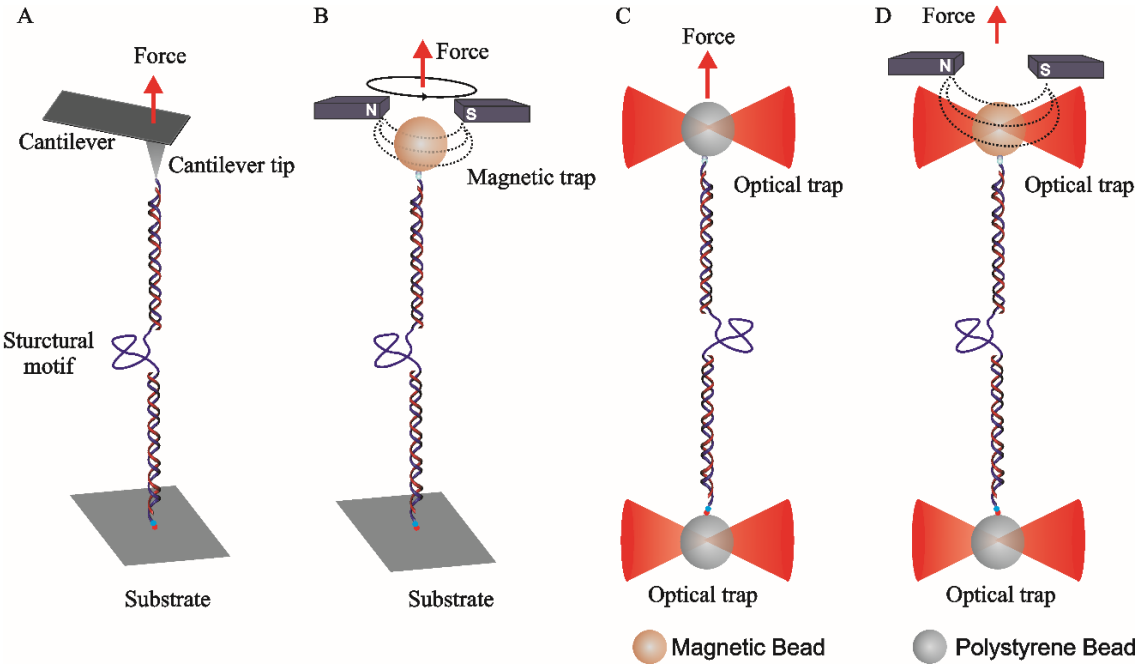


Figure 1.2. Schematic of the commonly used force based single-molecule methods for the unfolding of biomolecular structures. A structural motif (purple) is sandwiched between two long dsDNA handles, which can be tethered by different types of interactions based on the particular method. For example, tethering is done by covalent linkages in atomic force microscope (A), by affinity interactions in magnetic tweezers (B), optical tweezers (C) and opto-magnetic tweezers (D).

In optical tweezers, the macromolecules are tethered by two optically trapped polystyrene beads by non-covalent interactions such as biotin-to-streptavidin and digoxigenin-to-anti-digoxigenin.[101] The levitation of beads isolates the system from the environment, reducing the background noise. Also, as no fluorophore is used, photobleaching is avoided in such a system. As a result, thousands of mechanical cycles can be performed in a span of days without compromising mechanical properties of a

single-molecule template. All these features forecast a highly efficient single-molecule method that can be used in the mechanochemical analysis of a biological system. To unfold the G-quadruplex in optical tweezers, the DNA construct is synthesized by sandwiching the G-quadruplex forming sequence between two long dsDNA handles.[102] The DNA construct is tethered between two optically trapped beads as mentioned above. The tension in the DNA construct can be controlled by moving one of the beads away or closer to the other bead. The unfolding of a G-quadruplex is manifested by a sudden change in the tension recorded in a force-extension (*F*-*X*) curve. The refolding of G-quadruplex can be observed in the same way of sudden change in the tension while letting the molecule to relax. However, the refolding might be indistinguishable in real time if it is very slow. This mechanical unfolding/refolding procedure can be performed in various conditions according to the purpose of the study either by designing such DNA constructs and/or by changing the buffer conditions. Thus, by using optical tweezers, the mechanical stability, thermodynamics and folding-unfolding kinetics of different G-quadruplex structures in various conditions including during transcription and a nano-confined environment, can be studied.[103-107]

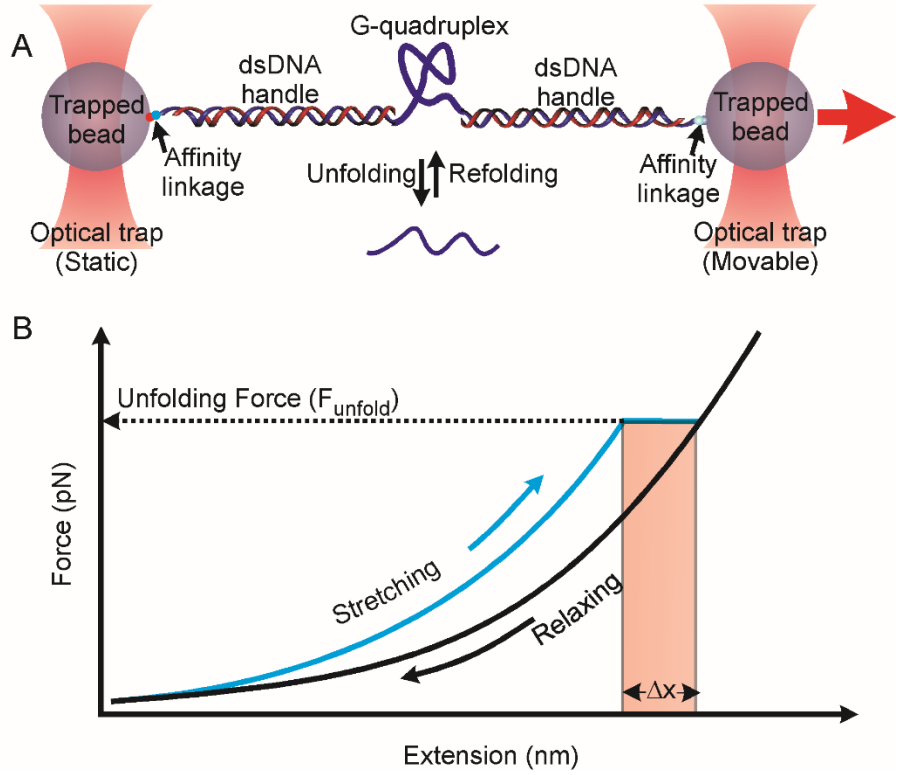


Figure 1.3. (A) Unfolding of a G-quadruplex by using optical tweezers. (B) Typical force-extension (*F*-*X*) curves of unfolding (blue) and refolding (black) of a G-quadruplex. Change in extension (Δx) and the unfolding force (F_{unfold}) of the G-quadruplex is measured directly from the *F*-*X* curves.

1.4 DNA origami nanotechnology to create a biomimetic confined space

The construction of relatively flexible branched junctions and fabrication of more rigid crossover DNA tiles marked the beginning of research in structural DNA nanotechnology.[108] However, the concept of DNA origami is considered as a remarkable breakthrough in structural DNA nanotechnology. Paul in 2006, has demonstrated this concept for the first time by doing self-assembly of a long single-stranded DNA (7329

nts) from M13 phage genome with hundreds of DNA staple oligonucleotides.[109] By taking advantage of sequence specificity and programmability, now one can precisely organize the other materials such as proteins[110], virus capsids[111], and nanoparticles[112-114]. The controlled and programmable arrangements of specific elements have led to the development of unique DNA nanostructures with much improved functional properties. Here, we synthesized DNA origami nanocages to place a human telomere G-quadruplex forming sequence inside. The size and dimensions of these nanocages are selected to mimic the entry/exit channels of polymerase and telomerase enzymes. Then, we can attach the two free ends of the G-quadruplex forming strand to the two long *ds*DNA handles by sequence-specific hybridization.

1.5 Aims and significance of the present study

As discussed in previous sections, single-molecule investigation of the formation of G-quadruplexes in the conditions mimicking inside the cell is highly essential to understand the role of G-quadruplexes in biological processes. This dissertation aims to understand the formation of G-quadruplexes during transcription of a G-rich sequence downstream of the transcription start site in the human genome and a human telomere G-quadruplex in a nano-confined condition. The methods used in this dissertation are mainly based on optical tweezers. The first part of this dissertation includes the deconvolution of a complex population mixture of the duplex and several G-quadruplexes during transcription, which otherwise is not possible from ensemble biochemical methods. The second part includes the application of DNA origami nanotechnology in optical tweezers to understand the folding dynamics of a human

telomere G-quadruplex in a biomimetic nano-confined environment. Altogether, this dissertation aims to mainly address the following aspects which have not been studied before because of the limitation of existing single-molecule methods.

I. It has been reported that DNA and RNA G-rich sequences interact to form a DNA/RNA hybrid G-quadruplex.[47-48, 115] During the transcription of G-rich sequence downstream of the transcription start sites of a human gene, a complex population equilibrium exists because of the presence of G-rich RNA transcript. It is essential to deconvolute such a complex population mixture of various folded structures to understand the role of G-quadruplex species in transcriptional regulation.

II. Polymerase enzymes are the motor proteins and they exert mechanical force when perform their function. It is essential to understand the mechanical stability and kinetics of G-quadruplex structures formed during transcription. Such information of G-quadruplexes during the transcription allows us to estimate which G-quadruplex species may have a predominant role in stalling the polymerases in regulating the genetic processes such as transcription and replication. However, none of the previous studies have reported the mechanical stability of G-quadruplexes formed during transcription.

III. The folding and unfolding of macromolecules in a confined space are prevalent inside cells. Therefore, it is highly significant to understand the folding behavior of biomolecules inside a nano-confinement. Several reports have suggested that the effect of confinement in the folding dynamics of biomolecules with controversy.[116-117] However, no studies have reported the real effect of confined space in folding dynamics of a

biomolecule because of extremely challenging experimental condition to avoid the surface interaction.

IV. Telomerase enzyme extends the length of telomere by adding repetitive GGTTAG sequence by its reverse transcriptase activity.[118-119] Such repetitive sequence is known to form G-quadruplex structure. Telomerase activity involves elongation and translocation of such G-quadruplex forming sequence through a nanometer-sized exit channel. As discussed above, the human telomeric G-quadruplexes are the targets for cancer therapeutics. Therefore, it is essential to understand the transition dynamics of G-quadruplex and its stability in such a confined state.

To accomplish these primary objectives, research was started by designing a single-molecule template to study the interaction of G-rich RNA transcript with the G-rich sequence in non-template strand of a duplex DNA. We developed a single-molecule stalled-transcription assay (SMSA) to meet this objective. The results showed that DNA and RNA G-rich sequences indeed interact to form a folded tetraplex structures called hybrid G-quadruplexes (HQ). Population deconvolution analysis revealed that DNA/RNA hybrid G-quadruplexes are the dominant species over DNA G-quadruplex and G-triplex. These findings depict a complex population dynamics between several types of folded structures in G-rich sequences downstream of the transcription start sites in the human genome. The presence of both HQ and GQ suggested their synergistic roles to regulate the transcription. HQs were found to be more stable than GQs, and this indicated that HQs might be alternative therapeutic targets.

To characterize the folding dynamics of a human telomeric G-quadruplex inside a confined condition, we synthesized DNA origami nanocages by incorporating a G-quadruplex forming sequence inside. The force ramping experiments in optical tweezers revealed that the unfolding of G-quadruplex occurred at a higher force than that without nanocage. This suggested the higher mechanical and thermodynamic stability of GQ inside the nanocage. In addition to that, the refolding kinetics of GQ was dramatically faster than that without nanocage. These results suggested the co-translocational folding of human telomeric G-quadruplex.

CHAPTER II

MATERIALS AND METHODS

2.1 Materials

All the chemicals including KCl, $MgCl_2$, Ethylenediaminetetraacetic acid (EDTA) and Tris-HCl were purchased either from VWR (www.vwr.com) or Nacalai Tesque, Inc., Kyoto, Japan (www.nacalai.com). Bovine Serum Albumin (BSA, biotechnology grade) was purchased from Amresco, Ohio, USA. The single-stranded M13mp18 DNA and all enzymes were obtained from New England Biolabs (www.neb.com). Staple DNA strands and p8064 single-stranded DNA were purchased from Eurofins Genomics (Tokyo, Japan). PEG-modified PCR primers, biotin-modified primer, dig-oxigenin modified primer and photo-cleavable (PC) strands were purchased from Japan Bio Services (Saitama, Japan). The pET-26b (+) plasmid for the PCR template was received from Novagen (Darmstadt, Germany). Biotin and digoxigenin modified PCR primers were purchased from Japan Bio Service (Saitama, Japan). The gel-filtration column and the Sephacryl S-300 were purchased from Bio-Rad Laboratories (California, United States) and GE Healthcare (Buckinghamshire, United Kingdom), respectively. The polystyrene beads coated with streptavidin or anti-digoxigenin for optical tweezers experiments were purchased from Spherotech (Lake Forest, Illinois, United States).

2.2 In vitro transcription for ensemble experiments

A double-stranded DNA (*ds*DNA) construct containing a T7 promoter and a downstream G-core (5'-TT$(GGGGA)_3$GGGGTT-3') was prepared by overlap extension PCR. Transcription was carried out with 0.05 µM dsDNA at 37 °C for 1 hr in 20 µl of transcription buffer of 40 mM Tris-HCl (pH 8.0), 30% (w/v) DMSO, 50 mM KCl, 2 U/µl of T7 RNA polymerase (Thermo Scientific, USA), 8 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 2 mM ATP, UTP and GMP, 1 mM GTP or 7-deaza-GTP, and 0.005 U/µl inorganic pyrophosphatase, (Thermo Scientific, USA). An equal volume of buffer containing 30% DMSO, 50 mM KCl, 1 µM of competitive DNA and the indicated RNases (0.8 µg/µl RNase A, 0.4 U/µl RNase H) was added and the mixture was maintained at 37 °C for 1 hr to terminate the transcription and digest the RNA. DNA samples were then resolved at 4 °C, 8 V/cm on a 10% polyacrylamide gel in 1× Tris-borate-EDTA (TBE) buffer that contained 75 mM KCl[48, 120] and DNA detected by the fluorescence of carboxyfluorescein (FAM) dye labeled at the 5' end of the non-template strand using a Typhoon 9400 phosphor imager (GE Healthcare, USA).

2.3 Analysis of the RNA transcript

Transcription was carried out as aforementioned but with addition of 0.05 mM fluoresein-12-UTP (Roche, Switzerland). The samples were digested with 0.08 U/µl of DNase I (Thermo Scientific, USA) at 37 °C for 30 min. The reaction was stopped by adding equal volume of 40 mM EDTA. The samples were extracted with equal volume of phenol/chloroform (1/1, v/v), dissolved in 50% deionized formamide, and resolved on an 8% denaturing gel.

2.4 DMS footprinting

Transcribed DNA (100 µl) was mixed with an equal volume of Tris−HCl (pH 7.9) buffer containing 30% (w/v) DMSO, 50 mM KCl, 40 mM EDTA and 0.2 % SDS. The DNA was then subjected to footprinting as described[121].

2.5 UV-crosslinking

Transcription was carried out as aforementioned but with the normal UTP being substituted with 1 mM of 4-S-UTP (TriLink BioTechnologies, USA). 100 µl of transcribed and RNase H-digested DNA was subjected to UV-cross-linking as described [48, 120]. The DNA was then treated by RNase A and EDTA as aforementioned, recovered with phenol/chloroform extraction and purified by the TIANquick mini purification kit (Tiangen, Beijing). Primer extension was performed with 0.4 µM 5'-FAM-CCAGCCTGCGGCGAGTG primer, 4 U of Deep VentR DNA polymerase (exo-) (NEB, USA) in a 50 µl volume containing 75 mM Tris−HCl, pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20, 2 mM $MgCl_2$, 0.05 mM dNTP, and 5% (v/v) DMSO. G and T ladders were synthesized in the same way using a non-transcribed and non-cross-linked DNA strand in the presence of Acy-CTP and Acy-ATP (NEB), respectively, in a molar ratio of 1/2 and 1/1 to dCTP and dATP, respectively.

2.6 Detection of RNase-resistant RNA in transcribed DNA

Transcription was conducted with DNA amplified using a biotinylated downstream primer. 50 µL of streptavidin coated MagneSphere paramagnetic particles (Promega, USA) was washed three times, each with 2× volume of 10 mM Tris-HCl buffer (pH 8.3) containing 30% DMSO, 50 mM KCl, and 10 ng/µl fish sperm DNA. It

was then incubated with 100 μL of transcribed DNA pre-digested with RNase A and H for 15 min at 37 °C. After three times wash each with a 500 μL 10 mM Tris-HCl buffer (pH 8.3) that contains 30% DMSO, 50 mM KCl, 10 ng/μL fish sperm DNA, and 20 mM EDTA, the immobilized DNA was detached by heating at 95 °C for 10 min in 30 μL of 20 mM EDTA and then mixed with an oligomer probe (5'-A$(CCCCT)_3$CCCCA-3') that was labeled with a Cy5 at the 5' end. The samples were cooled down slowly and resolved on 10% native PAGE gel at room temperature.

2.7 CD spectroscopic analysis

DNA:RNA heteroduplex was prepared and subjected to Circular Dichroism (CD) spectroscopic analysis as described,[120] except that 8 mM $MgCl_2$ was used and the PEG was replaced by 30% DMSO. CD spectra were recorded on a Chirascan-plus CD spectrometer (Applied Photophysics, United Kingdom) at 22°C, the path length of cuvette was 0.5 mm.

To study the conformation of a human telomeric G-quadruplex in different buffer conditions, the concentrations of oligonucleotide samples (5 μM) were prepared in a 20 mM Tris buffer (pH 7.8) at various ionic conditions (see Figure 4.5). Before recording CD spectra, the samples were heated at 95°C for 10 min and rapidly cooled in an ice bath followed by incubation at room temperature. CD spectra were recorded in a 1 mm quartz cuvette at room temperature with a Jasco-810 spectropolarimeter (Easton, MD). The spectra were smoothed by using a Savitzky-Golay function in Igor program (WaveMatrics, Portland, OR).

2.8 Single-molecule mechanochemical analysis

2.8.1 Preparation of microfluidic chamber

The procedure to prepare the microfluidic chambers for optical tweezers experiments have been described elsewhere.[122] In brief, the specifically patterned Nesco-film (Azwell, Osaka, Japan) (Figure 2.1) was sandwiched between two glass coverslips (VWR). The desired microfluidic patterns were designed in CorelDraw (Corel Inc., Mountain View, CA) and imprinted into the Nesco-film directly by a laser cutter (VL-200, Universal Laser Systems, Scottsdale, AZ). The two adjacent microfluidic channels were connected by using micro-capillary tubes (ID 20 µm, OD 90 µm). Finally, the patterned and sandwiched Nesco-film between the two coverslips were melted at 160 °C to seal. Then the microfluidic chamber was mounted on a bracket to facilitate the controlled injection of samples and the chamber was placed in between two objective lenses of optical tweezers.

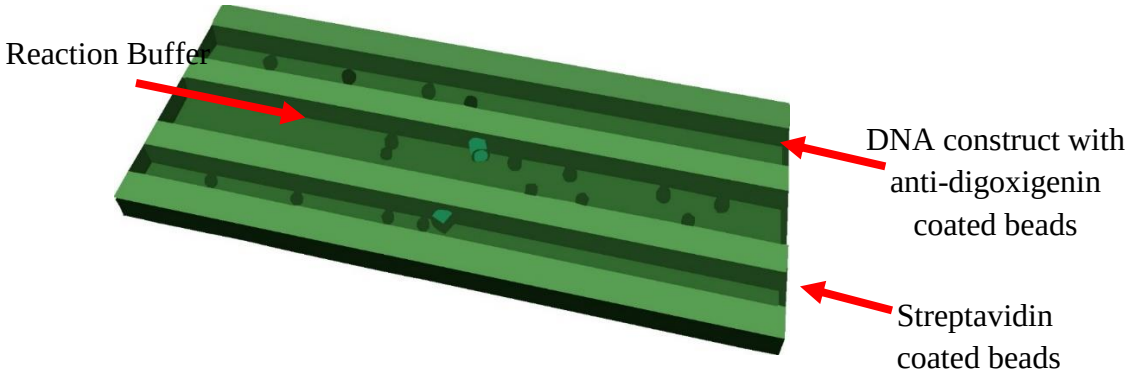


Figure 2.1. Schematic of a microfluidic chamber used in optical-tweezers measurements. The red colored arrows depict the flow directions of the specified solutions.

2.8.2 Synthesis of DNA constructs for Single-Molecule Stalled-transcription Assay (SMSA) of G-quadruplexes

2.8.2.1 Engineering and synthesis of DNA templates for SMSA

DNA constructs for the single-molecule assay was synthesized according to the flow chart described in Figure 2.3. First, *ds*DNA containing a T7-promoter site (5'-TAA TAC GAC TCA CTA TA), a G-core (see Figure 2.2), and a stall site downstream of the G-core was prepared by melting two complementary strands at equimolar ratio at 95 ^{o}C for 10 minutes, followed by slow cooling to 25 ^{o}C in 6 hrs. One end of the *ds*DNA contains an XbaI restriction site and the other an EagI site so that this DNA fragment can be ligated between two long *ds*DNA handles (2028 bp derived from pBR322 and 2690 bp derived from pEGFP) with respective restriction sites. For effective ligation, molar ratio of 4:1:4 (handle: *ds*DNA-Construct: handle) was used. The free ends of the 2028 and 2690 bp handles were labeled with biotin and digoxigenin, respectively. In single-molecule experiments, these two ends were tethered to the two optically trapped beads coated with streptavidin and anti-digoxigenin antibody, respectively. The DNA construct with 2 stall sites (see Figure 2.2B) was prepared similarly.

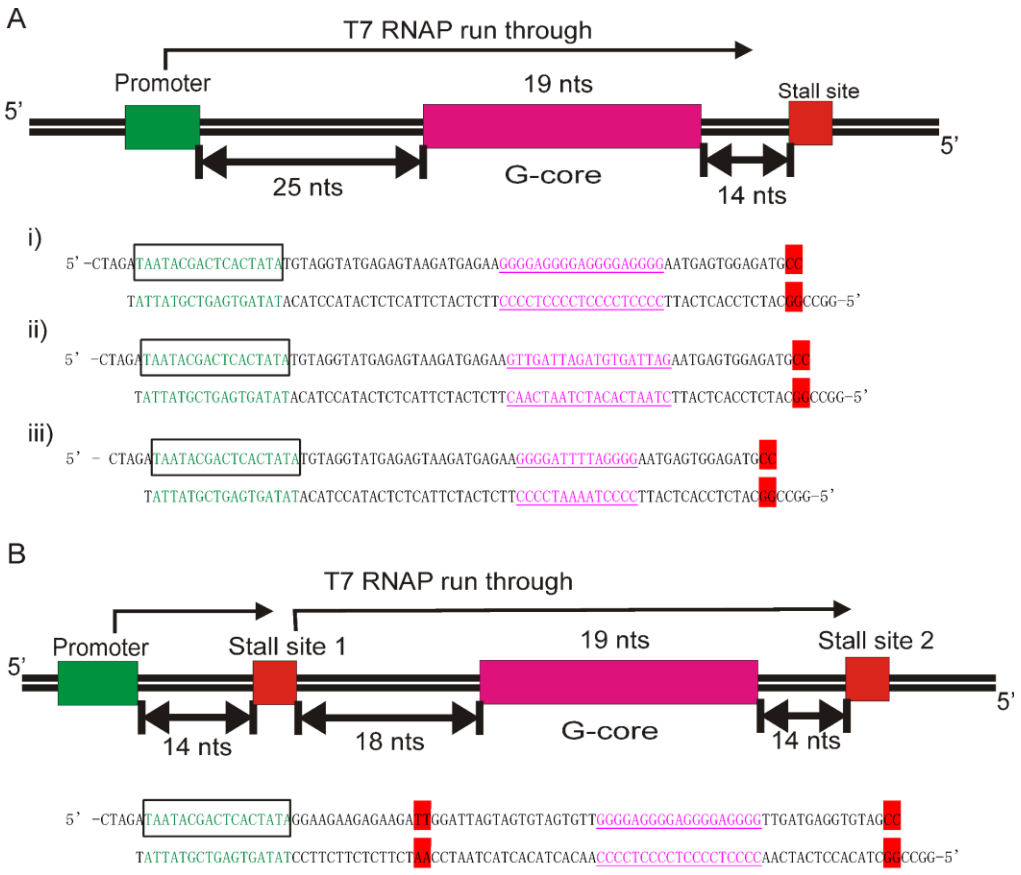


Figure 2.2. Schematics of the DNA construct for Single-Molecule Stalled-transcription Assay (SMSA). (A) DNA construct that contains a single stall site (highlighted with red). (i) GQ-forming construct (GQ-Construct) that contains a wild-type G-core sequence (pink). (ii) Control construct with a scrambled G-core sequence (Mutated-Control). (iii) Loop control construct: one G-tract is mutated to TTTT (bold) between the two G-tracts (Loop-Control). (B) DNA construct that contains two stall-sites (highlighted with red). Boxed green sequences depict the T7 promoter.

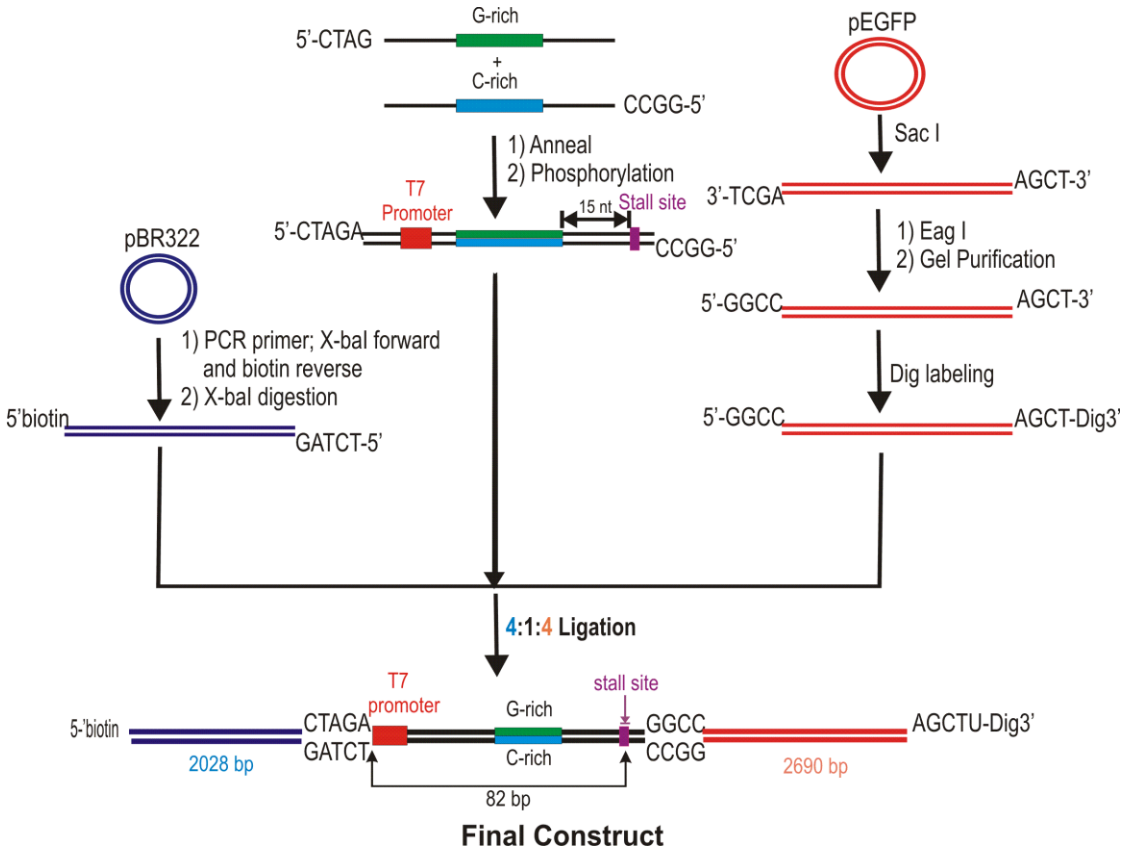


Figure 2.3. Preparation strategy of the DNA construct used for the mechanical unfolding and refolding during transcription. The 2690 bp handle was prepared from pEGFP plasmid by Sac I digestion, followed by Eag I digestion. The digested handle was labeled with digoxigenin at the 3' overhang using terminal transferase (NEB) and purified by agarose gel. The 2028 bp handle was prepared by PCR using a pBR322 template with a biotinylated primer and another primer that contains an XbaI restriction site. The complementary oligos consisting sequence of interest were annealed and phosphorylated. The final construct was obtained by 3-piece (the digoxigenin labeled 2690 bp handle, the phosphorylated double stranded DNA fragment, and the biotinylated 2028 bp handle) ligation with 4:1:4 molar ratio.

2.8.2.2 Preparation of nucleic acid constructs with stalled T7-RNAP transcriptions

Regular transcription was carried out by mixing 1µL T7 RNAP (NEB, 50000 U/ml) with 1µL of the *ds*DNA construct (see the construct for one stall site in Figure 2.2A) prepared above in a pH 7.8 transcription buffer (40 mM Tris-HCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine in 30% DMSO solution). After mixing with three nucleotide triphosphates (0.5 mM each of ATP, GTP, and UTP), 10 µL reaction mixture was incubated at 37 ^oC for 15 minutes to transcribe the DNA and to stall the T7 RNAP at the cytosine site (see Figure 2.2). During the deaza transcription, all conditions were the same as the regular transcription except GTP was replaced by 7-deaza-GTP. During the transcription of the DNA template with the 2-stall sites, T7 RNAP was stalled at the first site (see Figure 2.2B) by supplying ATP and GTP for 10 min. Extra T7 RNAP was dissociated from 1.5 µM DNA with a competitive promoter sequence, 5'-GAA ATT AAT ACG ACT CAC TAT A. A mixture of 0.5 mM each of ATP, GTP, and UTP was then added for 15 minutes to allow the RNAP to progress towards the second stall site. This procedure should allow only one passage of the RNAP over the G-core region.

2.8.3 Synthesis of DNA constructs to unfold human telomeric G-quadruplex inside DNA origami nanocages

The cubic-shaped DNA origami nanocages of different sizes (Figures 4.7A, 4.8A and 4.9A and Tables 4.2-4.4) were designed by using a standard protocol described elsewhere[109, 123]. Briefly, a segment of single-stranded DNA template: scaffold (2385 nts) for medium size (9 nm × 9 nm), scaffold (1841 nts) for small size (6 nm × 6 nm), scaffold

(3624 nts) for large size (15 nm × 15 nm), was obtained by digesting single-stranded circular M13mp18 (for small and medium nanocage) or p8064 (for large nanocage) DNA using appropriate restriction enzymes (see Table 4.5 for the restriction enzymes and respective splint DNA strands). Each scaffold ssDNA was purified by agarose gel and quantified. A mixture of 25 nM scaffold and 0.2 μM of each staple DNA strands (Tables 4.2-4.4) were assembled isothermally at 50 ºC for 1 hour to form an open nanocage. Next, open nanocage was isothermally assembled with 4 equivalents of the 203-nt G-quadruplex (GQ) hosting fragment (Figures 4.7B, 4.8B, 4.9B and 4.10) at 50 ºC for 1 hour. We used guide strands and capture strands to place the G-quadruplex fragment to the inner-walls of the nanocage (see Supporting Information, Figure 4.1B, 4.7-4.9). Short thymine linkages in capture strands and photo-cleavable linkers[124] (Tables 4.2-4.4) in guide strands were introduced to prevent misplacing the G-quadruplex fragment to the outer-walls of the nanocage. This was immediately followed by mixing 4 equivalents of closing strands to form a closed nanocage with the GQ strand contained inside. The mixture was purified by using a hand-packed Sephacryl S-400 gel-filtration column. The purified nanocage with GQ strand was annealed with two double-stranded DNA handles (preparation see below) by slowly cooling the mixture from 40 ºC to 15 ºC at the rate of -1 ºC/min.

2.8.4 Characterization of the single molecular DNA nanocage constructs by AFM

AFM images were obtained using a high-speed AFM system (Nano Live Vision, RIBM, Tsukuba, Japan) with a silicon nitride cantilever (resonant frequency = 1.0 - 2.0

MHz, spring constant = 0.1 - 0.3 N/m, EBD Tip radius <15 nm, Olympus BLAC10EGS-A2). The sample (2 μL) was adsorbed onto a freshly cleaved mica plate [Φ 1.5 mm, pretreated with 0.1% 3-aminopropyl trimethoxysilane (APTES)] for 5 min at room temperature and then washed several times with a buffer containing 20 mM Tris (pH 7.6), 10 mM $MgCl_2$ and 1 mM EDTA). The AFM images were obtained at a scan rate of 0.2 frames per second (fps).

2.8.5 Mechanical unfolding of G-quadruplexes during transcription

The laser-tweezers setup for the mechanical pulling experiments has been described previously.[102, 125] Briefly, a diode pumped solid state (DPSS) laser (1064 nm, 4 W, CW mode, BL-106C, Spectra-physics) was used to form two optical traps. One trap 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 two beams were collected by an identical objective and detected separately by two position sensitive photodetectors (PSD, DL100, Pacific Silicon Sensor).[126]

Unless otherwise specified, all pulling experiments were carried out at 23 °C in the transcription buffer with 30% DMSO as described above. The DNA construct with stalled T7 RNAP prepared above was immobilized onto a 2.10 μm bead (Spherotech) via digoxigenin−anti-digoxigenin antibody interaction. The mixture was then diluted to 1 mL transcription buffer and injected into a microfluidic chamber. To prevent subsequent binding and transcription of other RNAPs, the chamber was filled with the transcription buffer containing 30% DMSO but without T7 RNAP. After trapping the DNA

immobilized bead and the streptavidin coated bead (1.87 μm, Spherotech) by two separate laser foci, the DNA construct was tethered between these two beads. One of the beads was then moved away from the other, increasing the tension inside the DNA construct until structure was unfolded. In a typical force-extension experiment, the tethered DNA was extended below the plateau force (maximum 60 pN) and relaxed to 0 pN with a loading rate of 5.5 pN/s. The kinetic measurement was carried out by the force-pumping force-probing approach (FPP).[80, 127]

2.8.6 Mechanical unfolding of G-quadruplex inside a confined space

The DNA construct consisting of a human telomere G-quadruplex inside the DNA origami nanocage was synthesized as described in the main text (see Supporting Information for DNA sequences and Figure 4.2 for AFM characterizations). To start single-molecule mechanical unfolding experiments, 5 μL of the DNA construct thus prepared (10 nM) was incubated with 1.5 μL of 0.1 % solution (w/v) of antidigoxigenin coated polystyrene beads (diameter 2.10 μm, Spherotech) for 1 hour. This allowed immobilizing the DNA constructs onto the beads via digoxigenin and antidigoxigenin linkage. One of the DNA-attached beads was trapped at the focus of a 1064-nm laser beam in a custom-made dual-trap laser tweezers instrument[128-129]. At the focus of the other 1064-nm laser beam, a streptavidin coated polystyrene bead (diameter 1.87 μm, Spherotech) was trapped. The two beads were brought to contact with each other by a steerable mirror. This allowed the tethering of the other end of the DNA nanocage construct to the streptavidin coated bead through the biotin-streptavidin interaction

(Figure 4.1A). The DNA construct tethered between the two trapped beads was stretched up to 45 pN at the loading rate of 5.5 pN/s using the same steerable mirror. The force vs extension traces during this process were recorded at 1000 Hz using a Labview program and the raw data were filtered with a Savitzky-Golay function at a time constant of 10 ms in a Matlab program. The experiments were performed in a 20 mM Tris buffer (pH 7.8) supplemented by 10 mM $MgCl_2$, 1 mM EDTA, and appropriate monovalent cations (K^+ or Li^+). For molecular crowding investigations, the same buffer was supplemented with 40% (w/v) BSA.

2.8.7 Measurement of unfolding force and change in contour length

Unfolding of a G-quadruplex structures were manifested in *F-X* curves as a sudden drop in force accompanied with an abrupt increase in extension. The *F-X* curves were recorded at 1000 Hz in a LabVIEW program (National Instruments, Austin, TX) and Savitzky-Golay filtered to 100 Hz by MATLAB (The Math Works, Natick, MA), followed by analyses using IGOR programs (WaveMatrics, Portland, OR). The unfolding force of the G-quadruplex was measured directly from this perturbation in *F-X* traces. The change-in-extension (Δx) at unfolding force (F) was calculated as the extension difference between the stretching and the relaxing traces at that force. Then the change-in-contour-length (ΔL) was determined from the resulting Δx by using the Worm-Like-Chain (WLC) model[130] (equation 2.1):

$$\frac{\Delta x}{\Delta L} = 1 - \frac{1}{2}\left(\frac{k_B T}{FP}\right)^{\frac{1}{2}} + \frac{F}{S} \qquad (2.1)$$

where Δx is the end-to-end distance (or extension) between the two optically trapped beads, ΔL is the change-in-contour length, k_B is the Boltzmann constant, T is absolute temperature, P is the persistent length (51.95 nm),[130] and S is the stretching modulus (1226 pN).[130]

2.8.8 Calculation of expected change in contour length

Assuming x is the end-to-end distance (distance between the two termini of a folded structure); the expected change in contour length (ΔL) as a result of its unfolding can be calculated with the following equation,[131-132]

$$\Delta L = N \times L_{bp} - x \qquad (2.2)$$

Where N and L_{bp} are the number of base pairs involved in the structure and the contour length per nucleotide base pair, respectively. Here, L_{bp}= 0.34 nm was used (because the fact that duplex DNA is obtained after unfolding, L_{bp} is equivalent to the contour length per base pair, 0.34 nm).[133] To calculate the expected ΔL of the G-quadruplex structures in a given sequence, a value of x = 1.0 (± 0.1) nm was taken from reported NMR structures of the hybrid-1 type quadruplex.[134]

Four different folded structures equivalent to the following number of nucleotides were observed. For intramolecular G-quadruplex structure (N = 19 nucleotides), equation 2.1 gives an expected ΔL of 5.5 (± 0.1) nm. Similarly, ΔL for all possible structures were calculated according to the number of nucleotides of the template DNA strand that may be involved in the folding. DNA/RNA (3G+1G) hybrid G-quadruplex and intramolecular DNA G-triplex structures with N = 14 give an expected ΔL of 3.8 (± 0.1) nm and DNA/RNA (2G+2G) hybrid G-quadruplex has an expected ΔL of 2.0 (± 0.1) nm.

For unfolding of a human telomeric G-quadruplex, the expected change in contour length (ΔL) as a result of its unfolding can be calculated with the following equation,

$$\Delta L = N \times L_{nt} - x \qquad (2.3)$$

Where, N is the number of nucleotides expected to contain in the structure, L_{nt} (0.44 ± 0.02 nm)[135-136] is the contour length of single nucleotide, and x is the end-to-end distance determined by PDB structures. For the hybrid-1 structure, x is 1.1 nm.[27-28, 134]

2.8.9 Statistical analysis of observed change-in-contour length

For accurate estimation of the change-in-contour length (ΔL) of various folded structures during transcription, the kernel density calculation and bootstrapping analysis of ΔL (PodNano analyses) were carried out as described in literature.[104] The ΔL populations were identified by 3000 times in resampling. Percent formation for different populations was estimated from the *F-X* curves that contain the unfolding of specific ΔL species *vs* the total *F-X* curves. For two populations that are closely located with an intersection region, random deconvolution was performed as described.[137]

2.8.10 Jarzynski's calculation of the change in free energy of unfolding (ΔG_{unfold})

The change in free energy of unfolding (ΔG_{unfold}) for hTel G-quadruplex structure was calculated according to the Jarzynski non-equilibrium equation[138]:

$$\Delta G_{unfold} = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp\left(-\frac{W_i}{k_B T}\right) \qquad (2.4)$$

where k_B is the Boltzmann constant, T is absolute temperature, N is the number of repetitions in the experiment, and W is the non-equilibrium work done to unfold the G quadruplex, which was calculated with the following equation:

$$W = \sum_{i=1}^{N} F_i \Delta x_i \quad (2.5)$$

where F and Δx respectively represent the unfolding force and the change in end-to-end distance due to mechanical unfolding of a structure. To measure the work done to unfold the G-quadruplex, we measured the area under the unfolding region in the *F-X* curve. However, this work includes both the unfolding of the G-quadruplex and stretching of the unfolded ssDNA from 0 pN to the unfolding force. The latter work was corrected by a Freely-Joined-Chain model of stretched ssDNA according to the literature. [92, 132]

The histograms of the work required to unfold the G-quadruplex in the confined nanocage, molecularly crowded buffer, and diluted buffer are shown in Figure 4.17.

Bias of ΔG_{unfold} was calculated according to the literature[139-140]. Briefly, the lower tail of the unfolding work histogram was fitted by the following equation,

$$\text{p(W)} \sim \text{q}\frac{\Omega^{\alpha-1}}{|W-W_c|}\exp\left(-\frac{|W-W_c|^{\delta}}{\Omega^{\delta}}\right) \quad (2.6)$$

where W represents the work, W_c denotes the Gaussian center of the work histogram, q, Ω, α,

and δ are fitting parameters. These parameters were then used to calculate the bias, $<B_N>$.

$$< B_N > = B_{REM} - \lambda^{\frac{1-\delta}{\delta}}\left(\gamma_E + \frac{1-\alpha-\delta}{\delta} loglogN + log\frac{q}{\delta}\right), (for\ \lambda \ll 1) \quad (2.7)$$

$$< B_N >= B_{REM} + \gamma_E - \lambda^{\frac{1-\delta}{\delta}}\left(\gamma_E + \frac{1-\alpha-\delta/2}{\delta} loglogN + \frac{1}{2} log \frac{\pi q^2}{2\delta(\delta-1)} + \theta^2 + logerfc(\theta)\right), (for\ \lambda \leq 1) \quad (2.8)$$

here $\mu \equiv (\delta - 1)(\frac{\Omega}{\delta})^{\frac{\delta}{\delta-1}}$, and $\lambda \equiv (\delta - 1)\frac{logN}{\mu}$, notice N is the number of unfolding work measurements, $\theta = (\lambda^{\frac{1-\delta}{\delta}} - 1)\frac{\sqrt{\delta} logN}{\sqrt{2}(\delta-1)}$, $B_{REM} = D_c + \frac{\mu(\lambda - \delta\lambda^{\frac{1}{\delta}})}{\delta-1}$, where $D_c = \mu$, and γ_E is the Euler-Mascheroni constant.

2.8.11 Bennett acceptance ratio to retrieve change in free energy of unfolding (ΔG_{unfold})

Bennett Acceptance Ratio (BAR) method can be used to estimate the change in free energy due to unfolding of a biomolecular structure if the unfolding and refolding works can be measured[141-143]. This method has been described in the literature[143], here we have discussed this in brief. The ratio of the probability of the unfolding ($P_U(W)$) and refolding ($P_R(-W)$) work distributions is given by,

$$\frac{P_U(W)}{P_R(-W)} = \exp\left(\frac{W-\Delta G}{k_B T}\right) \quad (2.9)$$

where W is work done to unfold the structure, ΔG is the change in free energy due to unfolding, k_B is the Boltzmann constant and T is the temperature.

Equation S8 can be rearranged as,

$$f_x(W)P_U(W)\exp\left(-\frac{W}{k_B T}\right) = \exp\left(-\frac{\Delta G}{k_B T}\right) f_x(W)P_R(-W) \quad (2.10)$$

here $f_x(W) = \frac{exp\left(\frac{x}{2k_B T}\right)}{1+exp\left(\frac{x-W}{k_B T}\right)}$ (2.11)

This arbitrary function minimizes the statistical error involved in experimental measurements.

Integrating the equation between W = -∞ and W = +∞, we obtain,

$$\langle f_x(W) \exp\left(-\frac{W}{k_B T}\right)\rangle_U = \exp\left(-\frac{\Delta G}{k_B T}\right) \langle f_x(W)\rangle_R \quad (2.12)$$

Taking natural logarithm on both sides, the following equation can be obtained,

$$log\, \langle f_x(W) \exp\left(-\frac{W}{k_B T}\right)\rangle_U = log\langle f_x(W)\rangle_R - \frac{\Delta G}{k_B T} \quad (2.13)$$

where $\langle - -\rangle_U \; and \; \langle - -\rangle_R$ represent the average work values taken from the unfolding work distribution $P_U(W)$, and refolding work distribution $P_R(-W)$, respectively.

When we define, $z_U(x) = log\, \langle f_x(W) \exp\left(-\frac{W}{k_B T}\right)\rangle_U \; and \; z_R(x) = log\langle f_x(W)\rangle_R$, we get,

$$z_R(x) - z_U(x) = \frac{\Delta G}{k_B T} \quad (2.14)$$

In our case, we have the same number of repeated observations of unfolding and refolding trajectories, we can evaluate ΔG is from the intersection point of the plot $[z_R(x) - z_U(x)]$ *vs* $\frac{x}{k_B T}$ as derived by Pande and coauthors[144], i.e. The BAR analyses for human telomere GQ inside three different DNA origami nanocages are shown in Figure 4.18.

2.8.12 Calculation of unfolding kinetics of G-quadruplex

The unfolding force histograms of the hTel G-quadruplex inside the DNA nanocage, diluted condition, and molecular crowded condition were fitted by non-equilibrium kinetics model described by Dudko *et al.*[145]:

$$p(F) \propto \frac{k(F)}{r}\exp\left\{\frac{k_{unfold}}{x^{\dagger}r} - \frac{k(F)}{x^{\dagger}r}\left(1 - \frac{x^{\dagger}F}{\Delta G^{\dagger}}v\right)^{1-\frac{1}{v}}\right\} \quad (2.15)$$

where $k(F) = k_{unfold}\left(\left(1 - \frac{x^{\dagger}F}{\Delta G^{\dagger}}v\right)\right)^{\frac{1}{v}-1} exp\left\{\Delta G^{\dagger}\left[1 - \left(1 - \frac{x^{\dagger}F}{\Delta G^{\dagger}}v\right)^{\frac{1}{v}}\right]\right\}$

here $k(F)$ and k_{unfold} are the unfolding rate constants at force F and 0 pN, respectively, $x^{\dagger}$ is the distance to the transition state from the folded state, $\Delta G^{\dagger}$ is the height of energy barrier and v is the parameter to characterize the shape of the energy barrier (v = 1/2 for a sharp, cusp-like barrier and v = 2/3 for a softer, cubic barrier). Here we used both values of 'v' because the shape of the barrier is unknown. The reported values are average from these two shape factors. To estimate uncertainties in all fitted values, we separated data into 3 groups either randomly or according to the chronical order the data were collected. Both approaches gave similar standard deviations.

During the Dudko model fitting of unfolding force histograms, initially guessed values of fitting parameters were selected based on the reports by Zhongbo *et al.*[105], where all the 4 parameters in the Dudko model ($\Delta G^{\dagger}$, k_{unfold}, $x^{\dagger}$, and a proportionality factor (PF)) were set free to fit. In the case this fitting gave unrealistic fitting parameters ($\Delta G^{\dagger}$ in particular), a two-stage fitting was then performed. In the first stage, we held values of k_{unfold} and $x^{\dagger}$ while fitting $\Delta G^{\dagger}$ and PF. This allowed us to obtain a proximate range of PF. In the second stage, we held a specific PF value in the range while releasing all three kinetic parameters ($\Delta G^{\dagger}$, k_{unfold}, and $x^{\dagger}$) to fit. The chi square was monitored while we varied PF values in this fitting stage. The three fitted kinetic parameters were finalized with the least chi square value. The accuracy of the fitting was tested by two

approaches. In the first approach, we found $\Delta G^{\dagger}_{\text{unfold}}$ for the telomeric G-quadruplex in the dilute buffer (15.9 kcal/mol, see Table 4.1) matched well with the value estimated in literature[105] for the same G-quadruplex (16.2 kcal/mol). In the second approach, we used rupture force data for the L1-L3 unfolding geometry of the telomeric G-quadruplex from literature.[105] As reported, this set of data has been successfully fitted by free fitting of the four parameters in the Dudko model. We then used our 2-stage procedure to fit the same set of data. These two methods gave rather similar fitted parameters ($\Delta G^{\dagger}$=20.8 kcal/mol, k_{unfold}=2.8 × 10^{-4} s^{-1} and $x^{\dagger}$ = 0.17 nm for literature report (v=1/2 is used here instead of the average fitting of v=1/2 and v=2/3); they are 21.5 kcal/mol, 2.8 × 10^{-4} s^{-1}, and 0.17 nm, respectively, for our approach (v=1/2)).

2.8.13 Estimation of the change in entropy of the system during folding/unfolding of G-quadruplex

The change in free energy of unfolding (ΔG_d) for the GQ in diluted buffer (without cage) can be calculated by eqn 1,

$$\Delta G_d = \Delta H_d - T\Delta S_d \qquad (2.16)$$

for the confined system, the change in free energy can be calculated by eqn 2,

$$\Delta G_c = \Delta H_c - T\Delta S_c \qquad (2.17)$$

Assuming similar enthalpic change for the unfolding of G-quadruplex within or without the nanocage ($\Delta H_d = \Delta H_c$), we can relate the contribution of entropy change to the free energy change by subtracting eqn (S15) from eqn (S16),

$$\Delta G_c - \Delta G_d = T(\Delta S_d - \Delta S_c) \qquad (2.18)$$

Substituting the values of the change in free energy (ΔG_d= 7.1 kcal/mol, ΔG_c= 14.2 kcal/mol, see Table 4.1) obtained by the experiment at T=298 K, we can obtain the contribution of entropy to the stabilizing effect as follow,

$$(14.2 - 7.1) \text{ kcal/mol} = 298\text{K } (\Delta S_d - \Delta S_c) \qquad (2.19)$$

$$\Delta S_d - \Delta S_c = 0.024 \text{ kcal/mol.K}$$

CHAPTER III

SINGLE-MOLECULE STALLED-TRANSCRIPTION ASSAY OF G-QUADRUPLEXES FORMED DURING TRANSCRIPTION

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED IN THE NUCLEIC ACIDS RESEARCH JOURNAL, Nucleic Acids Research, **42**, 11, 7236–7246. *ALL THE MATERIALS OF THE PAPER HAVE BEEN REPRODUCED WITH THE PERMISSION OF THE OXFORD ACADEMIC.*

3.1 ABSTRACT

Recent discovery of the RNA/DNA hybrid Gquadruplexes (HQs) and their potential wide-spread occurrence in human genome during transcription have suggested a new and generic transcriptional control mechanism. The G-rich sequence in which HQ may form can coincide with that for DNA Gquadruplexes (GQs), which are well known to modulate transcriptions. Understanding the molecular interaction between HQ and GQ is, therefore, of pivotal importance to dissect the new mechanism for transcriptional regulation. Using a T7 transcription model, herein we found that GQ and HQ form in a natural sequence, $(GGGGA)_4$, downstream of many transcription start sites. Using a newly-developed singlemolecular stalled-transcription assay, we revealed that RNA transcripts helped to populate quadruplexes at the expense of duplexes. Among quadruplexes, HQ predominates GQ in population and mechanical stabilities, suggesting HQ may serve as a better mechanical block during transcription. The fact that HQ and

GQ folded within tens of milliseconds in the presence of RNA transcripts provided justification for the co-transcriptional folding of these species. The catalytic role of RNA transcripts in the GQ formation was strongly suggested as the GQ folded >7 times slower without transcription. These results shed light on the possible synergistic effect of GQs and HQs on transcriptional controls.

3.2 OVERVIEW

Different DNA species, such as various non-B DNA species and duplex DNA, form, dissolve, and interconvert in the same genetic location. This highly dynamic population equilibrium closely resemble the population dynamics widely used in ecology to describe the change in the population of biological species due to processes such as birth, death, immigration, and emigration.[146] Inspired by this concept, recently, we developed a molecular population dynamics approach to decipher the equilibrium of non-B DNA and duplex DNA species.[147] Dynamic pattern in population distribution offers flexibility for non-B DNA species to modulate cellular functions during different stages of a cell cycle. For example, it has been demonstrated that populations of DNA G-quadruplexes show a periodic pattern, which correlates well with the activity of helicase known to dissolve these structures.[62, 148]

A DNA G-quadruplex (GQ) is a four-stranded structure formed by a stacking of planar G-quartets, which are linked through Hoogsteen hydrogen bonds. GQs have already been found in telomere and promoter regions *in vivo*[62] with proven biological functions such as transcription inhibitions.[67] GQ-hosting sequences contain at least four

guanine (G)-rich repeats. For GQ-hosting sequences with more than four G-rich repeats, different GQ units can form due to permutations of four G-rich repeats. Even for a sequence that hosts only one GQ unit, the conformation of the GQ can be surprisingly versatile.[149-150] These scenarios bring complexity in the population equilibrium of GQ species. As prevailing species with long lifetime likely have significant biological roles, it becomes necessary to evaluate population profiles of GQs during various cellular processes.

Such an evaluation requires accurate determination of the population of individual species at equilibrium. In addition, as formation or dissolution of a species affects the overall equilibrium, evaluation of the transition kinetics among different species (molecular population dynamics[80]) becomes necessary. In *ss*DNA regions such as telomere overhangs, the profiling of the molecular population dynamics is relatively simple. In *ds*DNA regions, however, non-B DNA structures in the complementary strand may participate in the dynamic equilibrium. During transcription, nascent RNA may contribute to the population dynamics of these non-B DNA structures as well. RNA is well known to fold into secondary or tertiary structures on its own or to form a hybrid DNA/RNA duplex in the presence of a complementary DNA strand.[151-153] Recently, it has been suggested that the G-rich RNA transcript of the mitochondria CSB II motif can intermix with non-template DNA strand during transcription to form a new species, hybrid DNA/RNA G-quadruplex (HQ).[47] HQ formation was found to be a general phenomenon in transcription of G-rich DNA duplex. [48, 61] Sequences with the potential to form HQ are found to be evolutionally selected in the genomes of warm-blooded animals.

They are preferentially located in the non-template DNA strand downstream of transcription start sites in > 97% protein-encoding genes with an average of >73 sites per gene. The correlation of HQs with the transcriptional activity in host genes in human tissues has suggested their important roles in biological processes.[48, 61] Therefore, elucidating the effect of nascent RNA transcripts on the population dynamics that involve duplex DNA, various non-B DNA structures, and DNA/RNA hybrid structures provides a new perspective, at the molecular level, to understand the biological functions of these species, especially for their roles in the transcriptional control.

However, the complex molecular population dynamics during transcription is rather difficult to investigate by ensemble methods, which can only resolve the structures and kinetics of pure species, but not the mixed populations. In ensemble methods, mutations are often required to populate one specific species over others to reveal its properties. This, however, disrupts the original population dynamics.[150, 154] Single molecular techniques have an inherent capability to deconvolute individual species one molecule at a time and therefore, present a superior solution to probe complex population dynamics. Among different single-molecule tools, force based techniques, such as optical tweezers, can reveal spatial as well as mechanical information of DNA, RNA, or protein structures.[102, 155-156]

In the research reported here, we evaluated the molecular interactions of GQ and HQ in the same G-rich location for the first time. We first employed ensemble approaches including DMS footprinting, UV-crosslinking, and gel shift to investigate the

formation of DNA GQ and DNA/RNA hybrid GQs in a naturally occurring G-rich region, 5'-$(GGGGA)_4$, using a T7 transcription model. To delineate the kinetic and thermodynamic information of these quadruplex species, we introduced a single-molecule stalled-transcription assay (SMSA) in laser tweezers to unravel the details of molecular population dynamics hitherto unavailable to ensemble techniques. We observed a population equilibrium consisting of DNA GQ, partially folded DNA species, and DNA/RNA HQs during transcription. HQ species consisting of nucleic acid strands with two G-rich repeats predominate among various populations. While HQ folds within 30 ms, the folding kinetics of DNA GQ is beyond the resolution of our force-jump approach (τ < 20 ms). The fast formation kinetics of quadruplexes suggests that co-transcriptional folding of GQ and HQ is possible. The folding of GQ is at least 7 times slower (τ~150 ms) in the absence of the RNA transcripts, indicating the catalytic role of RNA to the formation of GQs. Not only HQs have higher population than GQs, they are stronger in mechanical stability as well (31 vs 25 pN), suggesting that HQs can potentially serve as stronger mechanical blocks to transcription. In addition, there is significant increase in the overall quadruplex population during transcription. All these findings suggest a synergistic effect of GQ and HQ to modulate transcription at the molecular level.

3.3 RESULTS AND DISCUSSION

3.3.1 Ensemble experiments show the formation of HQ species as a result of transcription

We chose a natural sequence, 5'-$(GGGGA)_4$ (G-core), as a model system to probe the complex equilibrium of G-quadruplexes during the transcription catalyzed by T7 RNAP. By genomic analyses, this sequence has been found in the non-template strand of 158 genes (Table 3.2 shows 36 genes that contain the G-core within 10000 bp downstream of the transcription start site (TSS)). Since this sequence is located downstream of the TSS, RNA strands containing the same G-core sequence will be produced during transcription. These RNA strands may participate in the equilibrium in which DNA G-quadruplex could form.

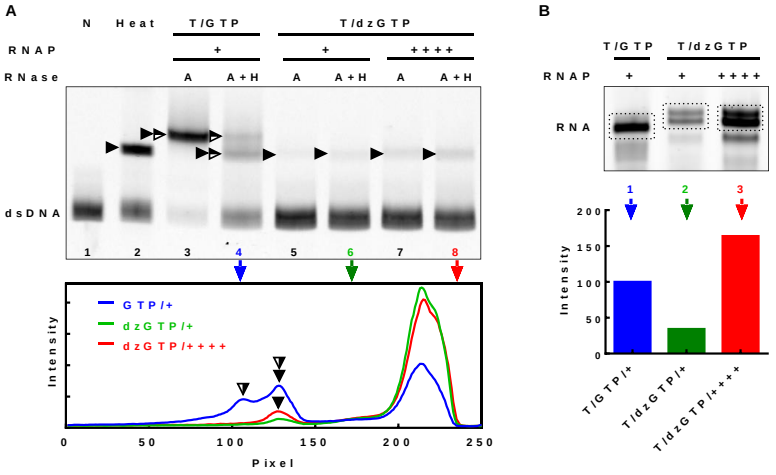


Figure 3.1. (A) Detection of G-quadruplex formation by native gel electrophoresis in G-rich *ds*DNA with GTP or 7-deaza-GTP (dzGTP) and two other NTPs followed by digestion with RNase A or RNase A+H. RNase A can digest ssRNA, RNase (A+H) can digest R-loop, while HQ is resistant to either RNase A or H. Bottom panel shows intensity scan of gel lanes 4, 6, and 8, respectively. Filled arrowheads indicate intramolecular DNA G-quadruplex and half-filled arrowheads indicate HQ. (B)

Transcription efficiency in the presence of normal GTP or dzGTP. RNA product was resolved on a denaturing gel without RNase digestion. Bottom graph shows the quantification of the RNA (dotted box) in the three lanes of the gel. 'N' stands for no transcription; 'Heat' for denaturation/renaturation; 'T/GTP' and 'T/dzGTP' stand for transcription with GTP and dzGTP, respectively. ++++ indicates four folds of RNA polymerase (RNAP) added comparing to +.

Formation of the DNA G-quadruplex was first detected by native gel electrophoresis. It has been shown that DNA containing G-quadruplexes should have a slower migration than the *ds*DNA counterpart.[48, 61] Therefore, the slower moving band in the gel (Figure 3.1A, lane 2, filled arrowhead) suggested the formation of an intramolecular G-quadruplex in the DNA that contains four G-tracts after the molecule was subjected to a heat denaturation/renaturation cycle. To detect G-quadruplex formations in the transcribed DNA, samples were treated with RNases A and H either separately or in combination. RNA in HQ is resistant to both RNases.[48, 120] The DNA transcribed in a K^+ buffer with regular GTP showed an extra band after digestion with RNase A (Figure 3.1A, lane 3, full and half-filled arrowhead), which was more retarded than the DNA carrying an intramolecular G-quadruplex (Figure 3.1A, lane 2). Transcription of G-rich DNA results in the formation of R-loops in which a nascent RNA transcript is base-paired with the template DNA at the guanine-rich (G-rich) region.[157] Because RNase A only cleaves at the 3' end of cytidines and uridines in single-stranded RNA (*ss*RNA) or RNA in the DNA/RNA duplex at low salt concentrations (0-100 mM),[48] the G-core region of the RNA (either in hybrid G-quadruplexes or RNA loops) is

resistant to RNase A. Therefore, the slower migration of this extra band can be explained by a structure containing G-quadruplex, R-loop, or both. To validate this assignment, transcribed product was treated with RNases A and H to cleave both *ss*RNA and R-loop (Figure 3.1A, Lane 4).[120] It was found that the amount of the fully-annealed DNA was increased (Figure 3.1A, lane 4, bottom band), suggesting the presence of the R-loop structure in the slow-migrating band in lane 3. The migration of a smaller portion of the DNA remained unchanged (lane 4, half-filled arrowhead), suggesting a fragment of RNA was protected from the digestion of RNases A and H, probably due to the formation of HQ structures. When the RNA was prevented from forming G-quadruplex in transcription by substituting regular GTP with dzGTP which lacks the 7'-nitrogen for G-quartet assembly, this band disappeared (lane 5). This result suggests that the slowest band represented HQ structure. On the other hand, a significant portion of the slow-migrating DNA in lane 3 restored the mobility (lane 4, middle band) to the same level as the DNA containing the intramolecular G-quadruplex in the heated sample (lane 2). We assume that this band contained a different species of HQ as well as GQ. The former structure seemed to be the major structure since this band was significantly reduced when the transcription was carried out with dzGTP, in which only GQ could be produced (lanes 6&8).

dzGTP reduces the incorporation at the initial nucleotide[158], which can be restored by addition of GMP [48, 120, 158]. To avoid this problem, therefore, all our transcriptions with dzGTP were supplied with GMP. However, the transcript analysis in Figure 1B indicated that the dzGTP still showed a significant reduction in the transcription efficiency (lane 2

versus 1). To ensure the changes observed with the GTP substitution was not due to reduced transcription efficiency, we increased the amount of T7 polymerase by four-folds, which elevated the transcription efficiency to 160% (lane 3 versus 1). This treatment, however, did not alter the result that HQ did not form in the presence of dzGTP (Figure 3.1A, bottom graph). These results confirmed that the HQ was RNA-dependent.

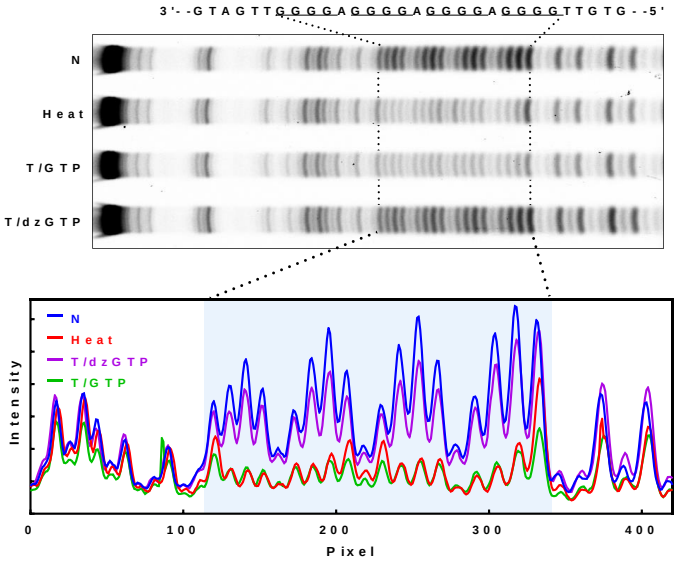


Figure 3.2. Participation of the non-template DNA strand in HQ formation detected by DMS footprinting. DNA was not transcribed (NT), heated (Heat) without transcription, and was transcribed with GTP (T/GTP) or dzGTP (T/dzGTP). DNA cleavage fragments were resolved on a denaturing gel.

To confirm that the G-core fragment can form HQ with an RNA strand of the same sequence, we hybridized a RNA oligonucleotide with a DNA fragment (DNA1) to form a partial duplex construct (Figure 3.10A). Both oligomers carried an overhanging

G-core that could possibly form an HQ. Another DNA:RNA partial duplex construct consisting of a mutated G-core DNA (DNA2) was used as reference. The HQ formation was detected by native gel electrophoresis after the RNA component in the duplex stem region was digested by RNase A or H. The much slower migration of the bands in RNA/DNA1 with respect to RNA/DNA2 can be ascribed to the association of the RNA G-core with the DNA1 via the HQ formation. The formation of HQ in the RNA and DNA1 was further confirmed by CD spectroscopy. A negative peak near 245 nm and a positive peak near 265 nm (Figure 3.10B) are characteristic of a parallel G-quadruplex and similar to those of the HQ we recently reported [120].

Next, we examined the participation of DNA in the G-quadruplexes by DMS footprinting in which guanine residues assembled into a G-quadruplex are protected from chemical cleavage.[75] In Figure 3.2, formation of G-quadruplex in the heated and transcribed DNA led to obvious protection of the guanines in the G-core. In the transcribed DNA, all the four G-tracts were similarly protected as those in the lane with heating. This could imply a similar probability for each G-tract to participate in the HQ assembly. In contrast, the DNA transcribed with the 7-deaza-GTP showed much reduced protection, further supporting the formation of HQ in transcription.

In our previous work, we used UV-crosslinking to verify the participation of RNA in the HQ.[48, 120] For the DNA used here in which the G-core is able to form an intramolecular DNA G-quadruplex, it is especially important to prove the involvement of the RNA transcript. To this purpose, we introduced a di-thymidine (TT) sequence at both 5' and 3' sides of the G-core region in the non-template DNA strand. During

transcription, we used 4-thio-UTP instead of UTP to incorporate modified uridines near the G-core. Due to the possible HQ assembly, the thio-modified uridines would covalently cross-link with nearby nucleotides in the non-template DNA strand upon UV-irradiation. In Figure 3.3A, crosslinking was detected at the two G-tracts from the 3'-end of the G-core (GTP lane). The involvement of the hybrid G-quadruplex structure to the crosslinking was obvious as the crosslinked bands disappeared when transcription was carried out using 7-deaza-GTP (dzG lane) instead of GTP.

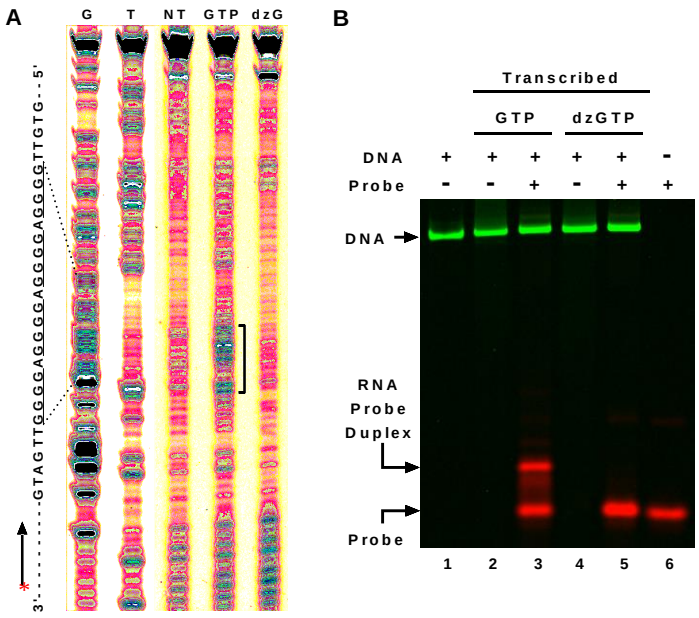


Figure 3.3. (A) UV-crosslinking to show the participation of G-tracts of RNA in G-quadruplex formation. 4-thio-UTP was incorporated into the RNA transcript and subjected to a post-transcription UV-irradiation. Crosslinking between RNA and DNA was detected by primer extension on the non-template DNA strand. Lanes from left to right represent G ladder, T ladder, DNA without transcription, DNA transcribed with GTP or dzGTP, respectively. (B) Detection of the RNase-resistant RNA fragments in

HQ. DNA transcribed with GTP or dzGTP was immobilized on magnetic beads and treated with RNase A and H. After washing, the RNA G-core was released and hybridized with a Cy5-labeled oligomer probe and resolved on a native gel.

To further verify the participation of RNA in the HQ assembly, we immobilized transcribed DNA modified with biotin on streptavidin-coated magnetic beads after it was treated with RNases A and H. As described previously, these two enzymes can digest the RNA in the single stranded form or hybridized with the template DNA, but not the RNA in the hybrid G-quadruplex. After washing, the RNA was released from the beads and probed with a Cy5-labeled DNA oligomer complementary to the RNA G-core (see Materials and Methods). In Figure 3.3B, RNA G-core was detected as indicated by an extra probe band above the original probe (lane 3). When the transcription was conducted with 7-deaza-GTP to prevent HQ formation, this extra band almost disappeared, strongly suggesting that the RNA in the band was associated with the HQ (lane 5).

Taken together, the gel-shift, DMS footprinting, and photo-crosslinking experiments provided convincing evidence for the formation of R-loop, DNA G-quadruplexes and DNA/RNA hybrid G-quadruplexes during transcription.

3.3.2 Single-Molecule Stalled-Transcription Assay (SMSA) confirms transcription induced DNA/RNA hybrid G-quadruplex Species

When the two beads were moved apart, the tension in the nucleic acid construct increased (Figure 3.4), which led to the unfolding of the structures as indicated by a sudden change in force or extension in a force-extension (*F*-*X*) curve (see Figure 3.5A). After each unfolding, tension in the DNA construct can be brought to zero at which the unfolded species starts to refold while both RNA transcript and RNAP stay in close proximity. Therefore, the setup closely mimics co-transcriptional folding process. As repetitive unfolding and refolding of nucleic acid structures can be carried out readily without requirement of new passages of RNAP, such a setup significantly increases the throughput of single-molecular experiments.

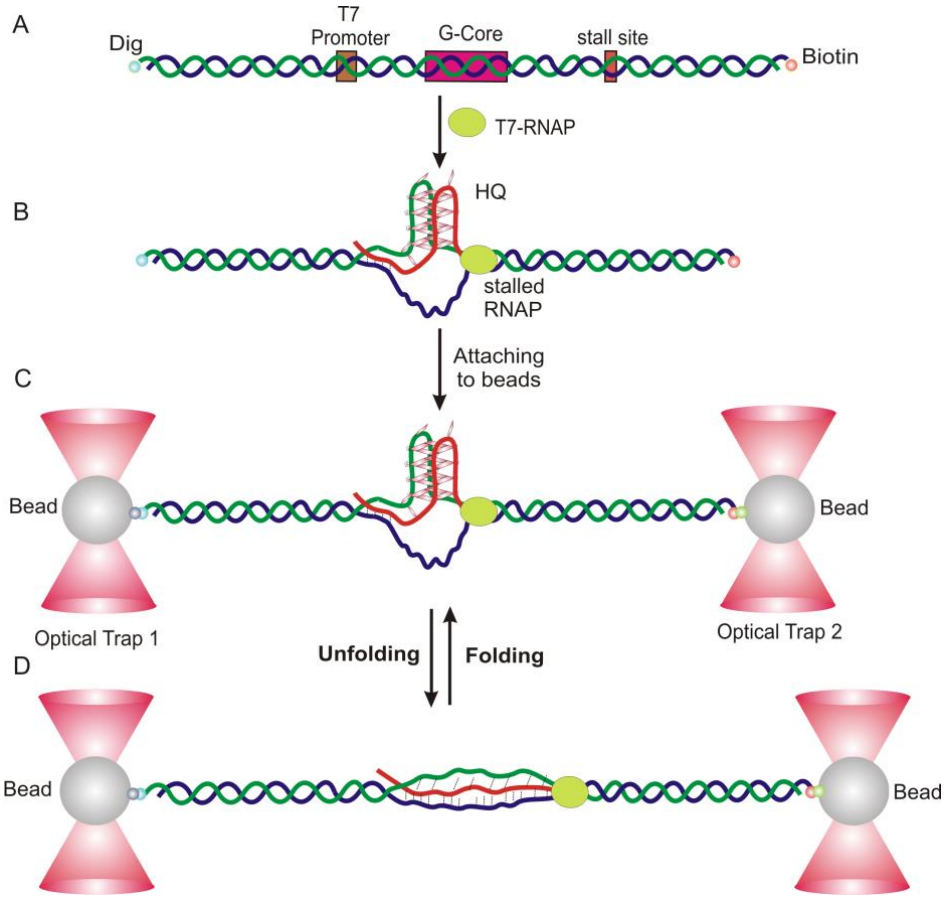


Figure 3.4. Schematic of the single-molecule stalled-transcription assay in optical tweezers. (A) The DNA construct. G-core has a sequence of 5'-$(GGGGA)_4$. (B) The T7 RNAP is stalled during a transcription carried out in 30% DMSO (pH 7.4). (C) DNA construct with stalled RNAP is tethered between two optically trapped beads. (D) Mechanical unfolding of structures formed in the G-core during transcription.

Typical *F-X* curves during regular stalled transcription showed three different types of unfolding events measured by the change-in-contour-length (ΔL) (Figure 3.5A). In the ΔL histograms shown in Figure 3.5B, three populations with ΔL= 1.8 ± 0.2, 3.8 ± 0.5, and 6.0 ± 0.6 nm are obvious. Population Deconvolution at Nanometer-resolution (PoDNano)[104] using resampling and bootstrap strategy confirmed the presence of the three populations (Figure 3.5C). Calculation revealed that these three structures contain 9, 14 and 19 nucleotides (nts), respectively (see Materials and Methods for calculation). While the 19-nt population corresponds to a fragment that contains four GGGGA tracts, the 14- and 9-nt species contain 3 and 2 GGGGA tracts, respectively. Inspired by the ensemble assays in which hybrid DNA/RNA GQs (HQ) were observed, we propose that populations with 9 and 14 nucleotides (1.8 and 3.8 nm in ΔL, respectively) are likely HQ species that consist of two and three G-tracts from the non-template DNA in combination with two and one G-tract of RNA transcripts, respectively (2G-HQ and 3G-HQ).

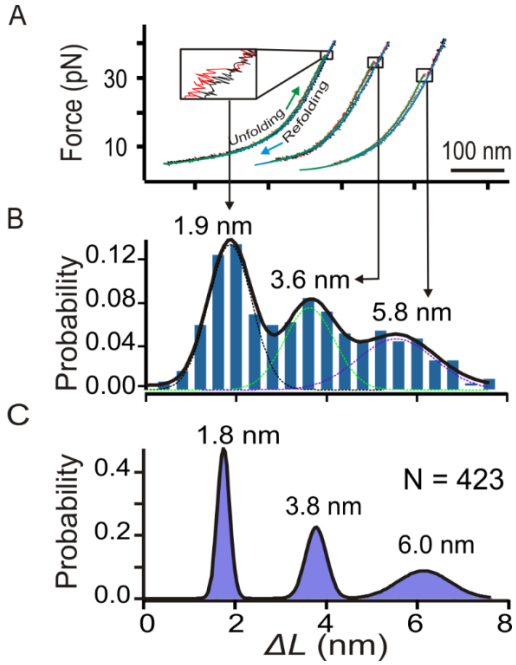


Figure 3.5. (A) Typical F-X curves show the unfolding transitions of three different species with ΔL ~1.9 nm, 3.6 nm and 6.0 nm during regular stalled transcription. Each curve is fit by the Worm-Like-Chain model for stretching (green) and relaxing (blue) processes. (B) Histogram of the change in contour length (ΔL) measured at the rupture force. Solid curves depict 3-peak Gaussian fitting and dotted curves represent Gaussian fittings for individual populations. (C) Histogram of ΔL populations after the PoDNano treatment. Black curves depict Gaussian fittings.

To confirm these assignments, four control experiments were performed. In the first control, we mutated guanine bases in the G-core to obtain a sequence, 5′- GTT GAT TAG ATG TGA TTA G. Since this sequence lacks consecutive guanine residues, neither GQ nor HQ is expected to form. When the sequence is subject to the mechanical unfolding with or without transcription, indeed, we observed very few unfolding events

with ΔL ranging from 2-8 nm (Figure 3.11). This result indicates that all three major populations observed in Figure 3.5A are structures associated with G-tracts. Next, we performed mechanical unfolding of the 5′-$(GGGGA)_4$ in the presence of 100 mM Li^+ after T7 RNAP was stalled during transcription. As expected, we observed only few folded structures (6.7 %, Figure 3.12). It has been shown that the formation of G-quadruplexes is much inhibited in Li^+,[159] therefore, the result in Figures 3.11 and 3.12 supports our assignment of GQ and HQ species observed in Figure 3.5.

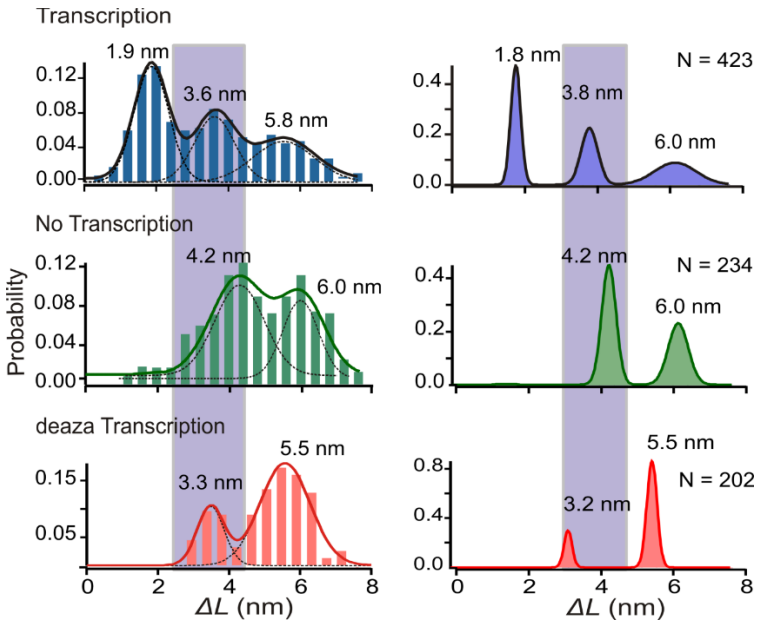


Figure 3.6. Population distribution of species with regular transcription (top panel), without transcription (middle panel), and with transcription in presence of 7-deaza-GTP (bottom panel). Left panels depict regular ΔL histograms measured at the rupture force while right panels show ΔL histograms after the PoDNano treatment. Curves represent Gaussian fittings.

In the third control, mechanical unfolding of the 5′-$(GGGGA)_4$ DNA construct was performed without transcription. As clearly shown in Figure 3.6 (middle panel), the smallest Δ*L* population disappeared while the remaining two populations showed slightly larger Δ*L* values (Δ*L* ~ 4.2 and 6.0 nm). While the largest Δ*L* population represents GQ, the ~ 4.2 nm Δ*L* population may represent partially folded structures, such as G-triplex (GT), as observed previously in other G-quadruplex forming sequences.[160] This control indicated that the ~2.0 nm Δ*L* species observed in Figure 3.5 B&C should involve RNA strands while the 3.6 nm Δ*L* (Figure 3.5) population can be a GT or a HQ. In the last control, we transcribed the same DNA construct by using 7-deaza-GTP instead of GTP. The 7-deaza guanine in RNA is known to compromise Hoogsteen hydrogen bonding and therefore HQ is not expected to form.[48] The Δ*L* histogram clearly revealed a major population at 5.5 ± 0.2 nm, a minor population at 3.3 ± 0.4 nm, and the ~2.0 nm species was again absent (Figure 3.6, bottom panel). Similar to the third control, this result indicated that the ~2.0 nm population in Figure 3.5 B&C should be an HQ species while the 3.3 nm species could be a partially folded DNA structure such as GT.

Table 3.1. Percent formation of different populations during transcription, transcription with 7-deaza-GTP, and no transcription.

Δ*L* (nm)	% Formation		
	Transcription	deaza-GTP Transcription	No Transcription
1.8 (2G HQ)	17.3	0	0
3.6 (3G HQ/GT)	13.2 (9.4/3.8)	8.1 (0/8.1)	14.7 (0/14.7)
5.5 (GQ)	12.2	25.9	9.6

0 (*ds*DNA or R-loop)	57.3	66	75.7

Since the ~ 2 nm species only involves 9 DNA nucleotides that span two GGGG tracts, it is clear that this species is an HQ that employs two DNA G-tracts and two RNA G-tracts (or 2G-HQ). However, for the ~4 nm and ~6 nm species in Figure 3.5 B&C, they may represent HQs that employ at least a total of three DNA G-tracts with at least one GGGG tract serving as an internal loop. To test whether non-tandem G-tracts can be employed in HQ, we prepared a construct that consists of two G-tracts interspersed by a 6-nt spacer, ATTTTA. Mechanical unfolding experiments under the stalled transcription condition revealed only little formation (7.9%) of structures with 2-8 nm in ΔL (Figure 3.13), strongly suggesting that consecutive G-tracts are required to form DNA/RNA hybrid G-quadruplexes.

Such a result indicates that the ~6 nm species observed during the regular transcription can be a GQ or two neighboring 2G-HQs, while the ~4 nm population can be a partially folded DNA G-triplex or a HQ that employs three consecutive DNA G-tracts with one RNA G-tract (3G-HQ). During the unfolding, however, we did not observe two consecutive events that correspond to the unfolding of two tandem 2G-HQ species. This observation led us to propose that the ~6 nm species should be a DNA GQ. To estimate the ratio of GT and HQ in the ~4 nm population, we first calculated the ratio of GT and GQ in the 7-deaza-GTP transcription in which no HQ should be present. We rationalized that GT and GQ maintain the same ratio based on the fact that both 7-deaza-GTP and regular transcriptions were carried out under the same set of conditions. Using

this ratio, we can then estimate the amount of GT in the ~4 nm population by measuring the GQ population in the ~6 nm peak (Figure 3.6, top panel). This algorithm gave the percentage formation of 9.4 and 3.8 % for HQ and GT species, respectively, in the ~4 nm population (Figure 3.6, top panel, see Materials and Methods for calculation).

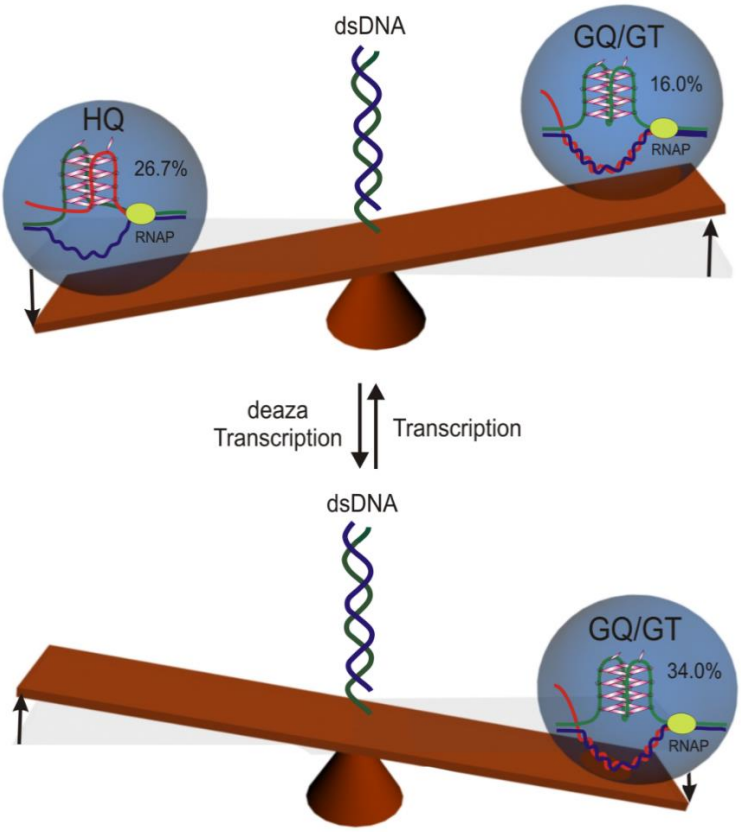


Figure 3.7. Schematic representation of population dynamics of DNA G-quadruplex (GQ), partially folded DNA species (GT), and DNA/RNA hybrid G-quadruplex (HQ) formed during the regular transcription (top) and transcription in the presence of 7-deaza-GTP (deaza transcription, bottom).

3.3.3 Population dynamics of the HQ and GQ species

Population distribution of the GQ and HQ species.

With a clear assignment of individual species, next, we compared population percentages of GQ, HQ, and *ds*DNA species in different transcription experiments (Table 3.1). First, we found that transcription almost doubled the combined population of HQ and GQ/GT with respect to that without transcription (43% vs 24%, see Table 3.1). This surprising result indicated the importance of RNA transcripts for the formation of HQ and GQ species. Population comparison during regular transcription also revealed that the HQ species predominated over the GQ/GT species (26.7% *vs* 16.0 %, see Figure 3.7 and Table 3.1), suggesting a more significant role for the HQ in gene expressions. Notably, the population estimation in single-molecule experiments (57.3% for *ds*DNA, Figure 3.7 and Table 3.1) matched closely with that obtained from ensemble gel shift assays (62.7% in Figure 3.1A, lane 4), which demonstrates the accuracy of the single-molecule measurements.

Mechanical and thermodynamic stabilities of the GQ and HQ species.

Recent evidence in our lab has suggested that the HQ species formed during the first few rounds of transcription could stall subsequent transcription processes.[120] Since RNAP that catalyzes transcription is a motor protein which possesses certain load force, the unfolding force of HQs could be used to evaluate whether HQ can serve as a mechanical block to the RNAP by comparing with the stall force of RNAPs.[161-162] The force-based single-molecule experiment employed here has a unique ability to determine the mechanical stability of folded species.[98] The histograms of the unfolding force

measured by the single-molecule transcription assay are summarized in Figure 3.8. While the unfolding force of the species formed during transcription is centered at ~31 ± 1 pN (it is ~32 pN for 2G-HQ, see dotted histogram in Figure 3.8, top panel), the populations formed without transcription or with deaza transcription (GQ/GT) show rupture forces of ~25 ± 2 pN (see dotted histograms for deconvoluted GQ populations). This difference is statistically significant at a confidence level of 95% by the ANOVA analyses, firmly demonstrating that HQ species are mechanically more stable than the GQ species. As the unfolding forces of both GQ and HQ are higher than the stall force of known RNAP species,[161-163] the increased mechanical stability in HQ suggests it may serve as a more effective mechanical block to the RNAPs.

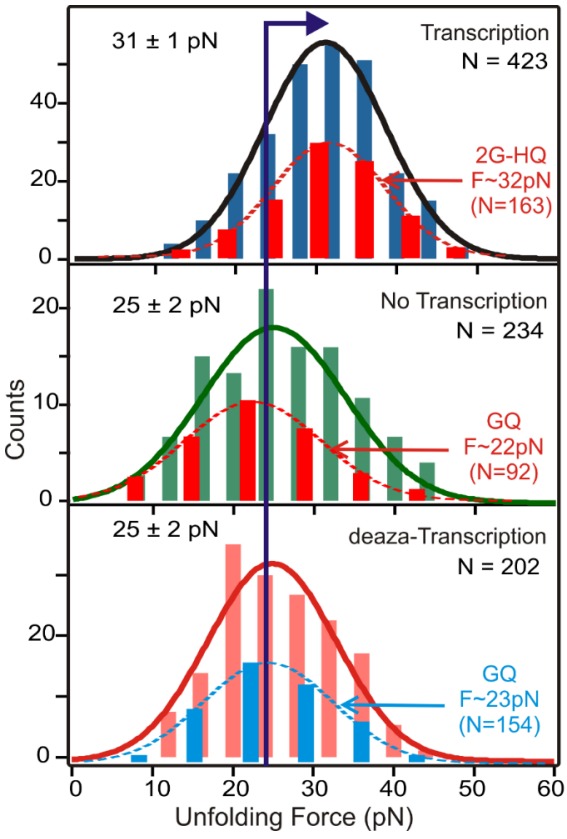


Figure 3.8. Rupture force histograms observed during regular transcription (top panel), no transcription (middle panel) and transcription with 7-deaza GTP (bottom panel). The dotted histogram in each panel depicts the rupture force of HQ or GQ population after deconvolution (see text). For clarity, these dotted histograms are plotted with reduced size.

Using Jarzynski non-equilibrium theorem, we evaluated the thermodynamic stability of HQ *vs* GQ species based on the unfolding work of each species.[102, 107] After deconvoluting 2G-HQ, 3G-HQ, and GQ species as described in literature[137] (see dotted histograms in Figure 3.8), we retrieved their respective change in free energy of unfolding (ΔG_{unfold}) as 5.5 ± 0.1(0.4), 6.8 ± 0.3(-0.6), and 9.5 ± 0.3(0.8) kcal/mol (values in parentheses show the bias in the ΔG_{unfold} estimation[140]). It is interesting that the GQ species show higher thermodynamic stabilities than HQ, which could be attributed to the increased entropic penalty for the formation of intermolecular structures in HQ with respect to intramolecular structures in GQ.

Transition kinetics of the HQ and GQ species.

To investigate the temporal effect on the population dynamics of the GQ and HQ species, we used force pumping and probing (FPP) method.[80, 127] First, we unfolded all structures by mechanically stretching the nucleic acid construct up to 45 pN, followed by a quick relaxation to 0 pN within 20 ms. After a specific incubation time, we probed the folded structures by next round of force ramping that started at 10 pN by another force jump. Folding of a species was revealed by the rupture event while the identity of the

species was determined by the change in contour length as a result of its unfolding (see Figure 3.15). After plotting the major species of GQ and HQ (2G-HQ and 3G-HQ) versus the incubation time at 0 pN, we found that the percentage population of DNA GQ (~10%) reached to a steady state within 20 ms while the formation of HQs followed a single exponential kinetics with a time constant of 30 ± 20 ms (Figure 3.9A). Comparison of the transition kinetics among HQ, GQ, and *ds*DNA species (Figure 3.9C) reveals that the formation of the HQ is at the expense of *ds*DNA until a steady state is reached at 100 ms. This suggested that the presence of a proximal RNA strand helps to convert the G/C-rich duplex DNA to the HQ within 100 ms. Such a fast kinetics suggested that co-transcriptional folding of HQ and GQ is feasible even without the pausing of RNAP during transcription. With a transcription rate of ~10 bp/s,[164] the folding and equilibrium of HQ and GQ could be accomplished as soon as the quadruplex hosting sequence is transcribed given that participating DNA and RNA strands remain single-stranded.

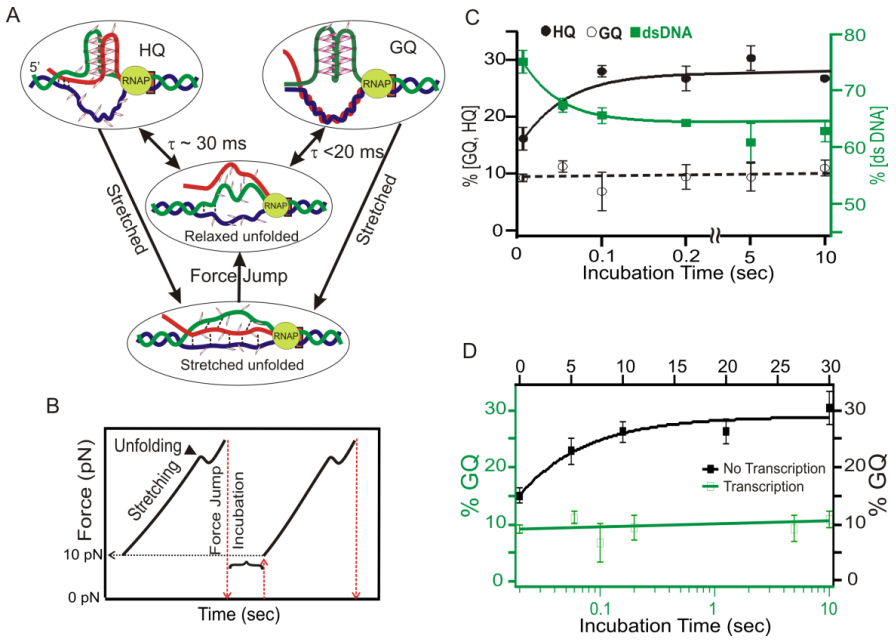


Figure 3.9. (A) Schematic diagram for different G-rich structures during the force jump experiment. Dotted lines represent possible WC or Hoogsteen bonding. (B) Profile of force change during the force pumping and probing experiment. (C) The percent formation of HQ, GQ+GT, and dsDNA for different incubation period during stalled transcription. (D) Comparison of the percent formation of the GQ+GT species with (green) and without (black) transcription. For clarity, the kinetics data for regular transcription are displayed with a semi log plot.

Interestingly, the folding of DNA GQ in the absence of transcription was much slower (150 ± 60 ms, Figure 3.9D). It is possible that negative superhelicity in the wake of a transcription bubble may speed up the DNA GQ formation. However, since the superhelicity of the DNA template quickly reaches equilibrium by the free rotation of the stalled RNAP in our experiments, the superhelicity effect is not likely. An alternative explanation is that the presence of RNAP may facilitate the folding of DNA GQ. Given

the fact that the stalled RNAP is located away from the G-core by ~5 nm, this scenario is not likely either. Instead, we propose a catalytic role of nascent RNA for the GQ formation. We surmise that RNA could serve as a template to bring DNA guanine residues together, which reduces the entropic penalty to form the DNA G-quadruplexes. Another scenario could be that in the presence of a RNA transcript, the hybridization of complementary DNA strands is inhibited due to the formation of the R-loop, which leaves the single-stranded G-rich strand more available for the DNA GQ to fold. We have shown here that mechanical stability of either GQ or HQ can withstand the stall force of known RNAPs (Figure 3.8). This implies a possible synergistic effect for GQ and HQ to serve as mechanical blockers to transcription. The synergistic effect is supported at the molecular level that nascent RNA transcripts serve as catalysts for the formation of GQ (Figure 3.9) while converting more *ds*DNA into HQ and GQ species (Table 3.1). Therefore, the potential biological functions of HQ lie in two folds. First, HQ may itself stall the transcription. Second, HQ could kinetically or thermodynamically change the population of GQ, which is a well-known modulator for various biological processes.

3.3.4 Comparison between the SMSA and ensemble experiments

Our recent ensemble experiments have shown that during the first round of transcription, R-loop, instead of HQ, is formed.[120] The formation of R-loop may be accompanied by a GQ folded in the non-template (G-rich) strand. To test this hypothesis using mechanical unfolding experiments, we prepared a construct that contained a stall site upstream of the G-core. Only adenine and guanine bases were present between the

transcription start site and the stall site (see construct 2 in Figure 2.2). The T7 RNAP was initially stalled by a supply of those two nucleotide triphosphates (NTPs). The construct also contained a second stall site at the 15th nucleotide downstream of the G-core. Resuming the transcription of the RNAP stalled at the first site using 3 NTPs (ATP, GTP and UTP) would allow only one RNAP to pass the G-core and stall at the second site. Mechanical unfolding of this construct showed folded species were identical (Figure 3.14) with those in Figure 3.6. Interestingly, during the first few *F-X* curves, the ~ 6 nm species (DNA GQ) was observed most often, which is consistent with the observation that HQ was not formed during the first round of transcription.[120]

Our ensemble data also showed that during the second round of transcription, not only DNA GQ was unfolded by RNAP, the RNA-DNA hybrid duplex (R-loop) was also displaced by the polymerase so that both single-stranded non-template (G-rich) DNA and previous RNA transcript were available for the HQ formation.[120] This situation was well mimicked in our SMSA method in which all folded structures were first unfolded by mechanical stretching. In addition, mechanical tension is expected to weaken the Watson-Crick pairs[122] in the DNA or the hybrid DNA/RNA duplex. Therefore, the refolding of the species at the lower force in the single-molecule approach resembled the second pass of RNAP during ensemble experiments. The observation of HQ formation during subsequent *F-X* curves again supported the results from our ensemble experiments.

Subtle difference existed between the single molecule assay and the ensemble experiments. In the latter case, the DNA template strand was no longer available for population equilibrium as it was hybridized with the new RNA transcript during the second round of transcription. However, in the mechanical unfolding experiments, this strand was available for the population dynamics under mechanical tension. Such a subtle difference implied that amount of HQ formation might be underestimated in single-molecule experiments (26.7%) compared to ensemble assays.

However, limitations exist for this single-molecule stalled-transcription assay (SMSA). As T7 RNAP is stalled downstream of the G-core, it may have different conformation compared to the active enzyme during transcription, which may lead to different formation kinetics of GQ/HQ. In addition, small uncertainty exists in the estimation of folded population as it is difficult to determine the exact position of the stalled RNAP.

3.4 CONCLUSIONS

Using ensemble experiments and an SMSA, we identified a population mixture of GQ and HQ in a natural G-rich sequence downstream of many TSSs in a T7 transcription model. We revealed that HQ predominated over GQ in population as well as in mechanical stability, which implies that HQ can serve as a more effective mechanical block for transcription. The fact that RNA transcripts catalyzed the folding of GQ while converting more dsDNA into quadruplex species starts to suggest a synergistic effect of HQ and GQ on the transcriptional control. As our SMSA approach closely mimics the

transcription processes, we anticipate that the method can be applied to investigate complex molecular population dynamics that involve non-canonical DNA and RNA structures during transcription.

3.5 SUPPORTING INFORMATION

3.5.1 Detection of HQ formation in synthetic oligonucleotides

DNA:RNA heteroduplex was prepared and subjected to native gel electrophoresis or Circular Dichroism (CD) spectroscopic analysis as described,[120] except that 8 mM $MgCl_2$ was used and the PEG was replaced by 30% DMSO.

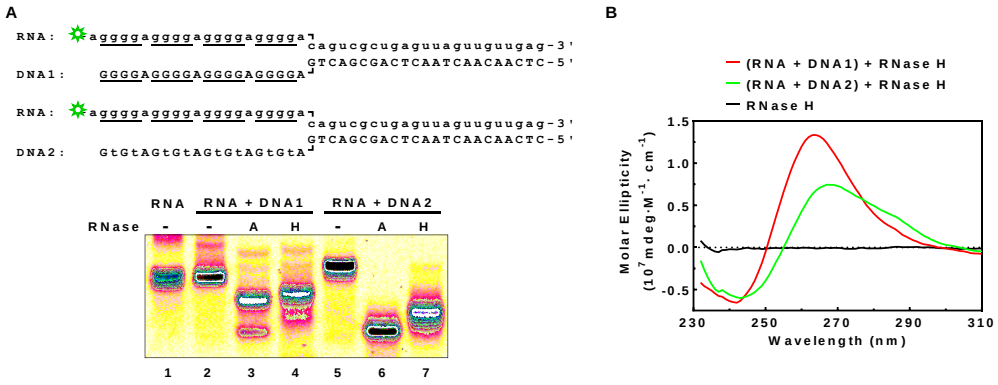


Figure 3.10. HQ formation between the $(G_4A)_4$ G-core in synthetic DNA and RNA oligonucleotides. (A) The RNA annealed with DNA1 or DNA2 was subjected to a digestion with RNase A or H before being resolved on a native gel. A retarded migration of the RNAs in lanes 3 and 4 in comparison with that of the RNA in lanes 6 and 7 in indicates a formation of HQ. (B) CD spectra of RNase H digested RNA + DNA1, RNA + DNA2, and RNase H.

3.5.2 Analysis of percentage of unfolding events

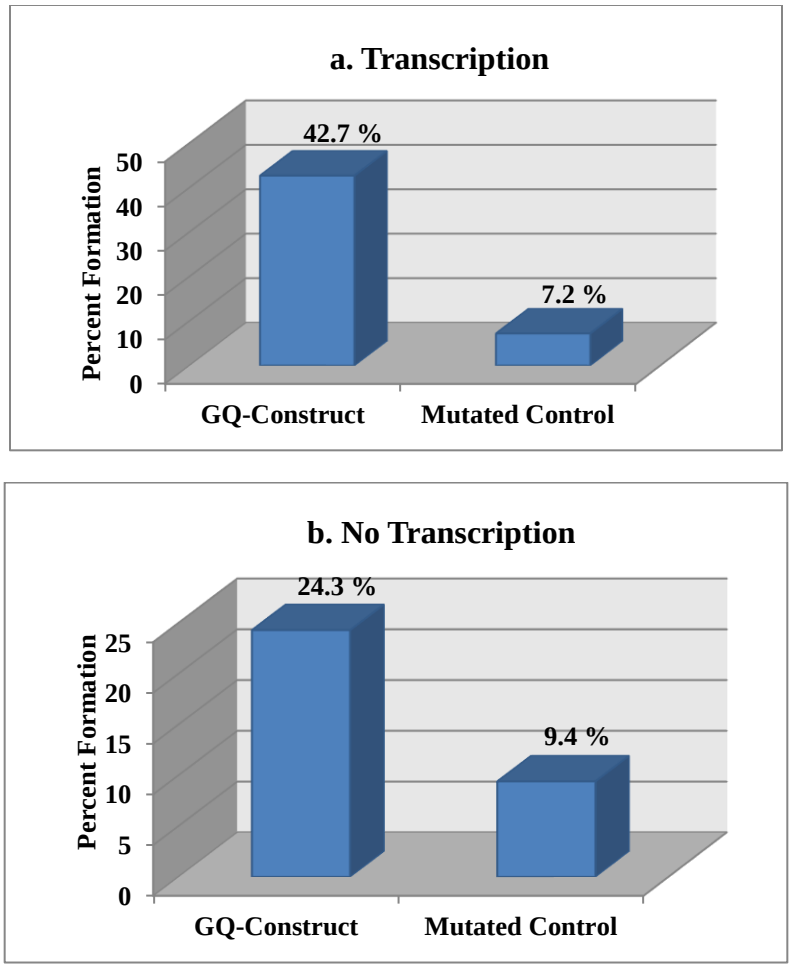


Figure 3.11. Control experiments to confirm the formation of GQ species in the GQ forming construct (GQ-Construct). Percent formation of structures (featured curves) during transcription (a) and without transcription (b) in the single-molecule mechanical unfolding and refolding experiments in a 100 mM KCl buffer.

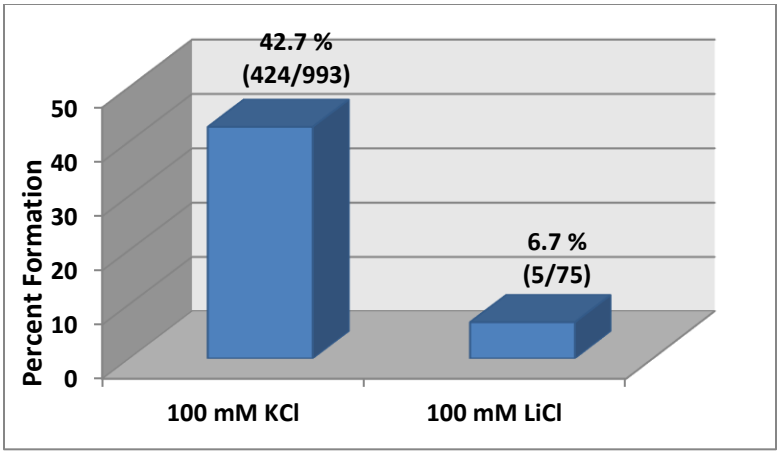


Figure 3.12. Comparison of the percent formation of GQ and HQ species in the stalled transcription assay in 100 mM KCl and 100 mM LiCl. The significantly reduced percent formation of structures in 100 mM LiCl (6.7 %) with respect to that in 100 mM KCl (42.7 %) supports that folded structures are GQ and HQ species. Numbers in parentheses indicate the F-X curves that contain folded structure vs total F-X curves.

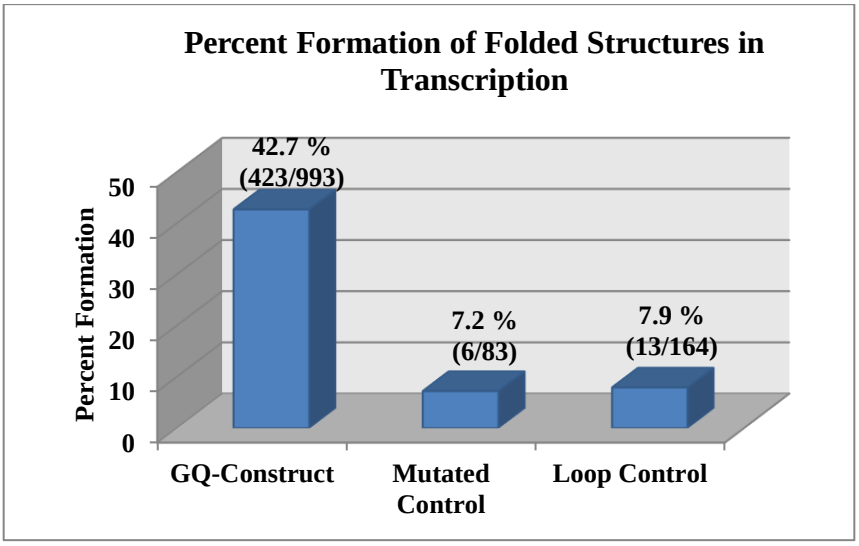


Figure 3.13. Control experiments to show that consecutive G-tracts are necessary for HQ formation. See Figure 2.2 for the sequence of GQ-Construct, Mutated-Control, and Loop-

Control. Rupture events with ΔL ranging from ~2-8 nm were counted. Numbers in parentheses indicate the F-X curves that contain folded structure vs total F-X curves.

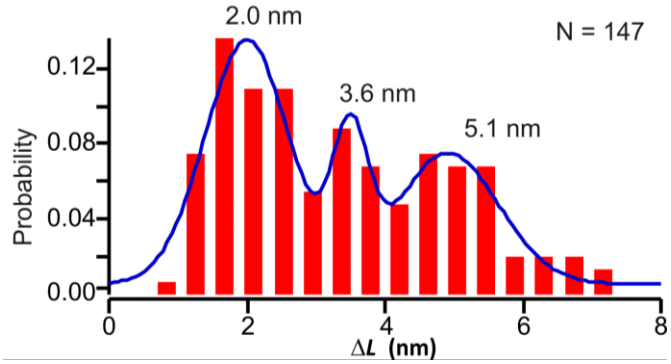


Figure 3.14. Histogram of the change in contour length (*ΔL*) obtained from the single-molecule stalled-transcription assay on the DNA construct that contains two stall sites (see Figure 2.2, 2). The two stall-sites design avoids transcription due to multiple T7-RNAPs (see main text).

3.5.3 Calculation of percent formation of G-triplex, G-quadruplex, and DNA/RNA hybrid G-quadruplex structures

To estimate the percent formation of different species in the control experiments (Figure 3.11-3.13), we took the ratio of *F-X* curves that contain 2-6 nm Δ*L* features *vs* all *F-X* curves. The overall *F-X* curves under specific conditions are shown at the top of the bar diagrams (Figures 3.12 and 3.13).

The population at Δ*L* ~ 4 nm corresponds to the involvement of 14 nucleotides (see equation 2.2) in a structure, which could be either partially folded intramolecular DNA G-triplex or three G tracts participated in the DNA/RNA hybrid G-quadruplex (HQ). To find the percentage population of each structure, we first determined G-triplex

population using the population ratio of G-quadruplex and G-triplex structures observed in the deaza-transcription, in which HQ is not expected to form. The HQ population was then estimated from the rest of the 4 nm ΔL population.

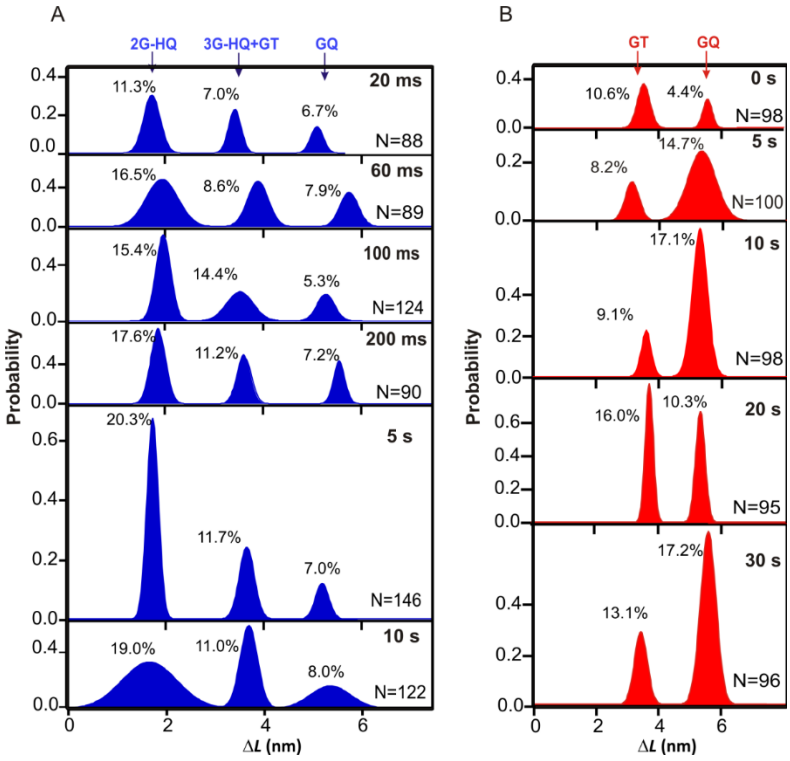


Figure 3.15. Histograms of the change in contour length (ΔL) observed in (A) transcription and (B) without transcription at different incubation time. The histograms were plotted after the bootstrap analysis of ΔL measured from the force-extension curves (see Materials and Methods). The percentages shown are the absolute percentage of the corresponding populations (see Materials and Methods).

Table 3.2. Genes that contain the G-core sequence (5'-$(G_4A)_3G_4$) downstream (within 10000 bp) of the transcription start site. Sequences of human genes were downloaded in the fasta format along with their IDs from the Ensemble genes database (release 74, GRCh37.p13) via the BioMart interface (http://www.ensembl.org/) by selecting known genes in the Gene type filter and Unspliced (Gene) in the Attribute/Sequences panel. The G-core was identified using the ZT-Find-PHQS-GUI.exe software supplied.[120]

S.N.	Ensemble Gene ID	Associated Gene Name	Distance from TSS (bp)	S.N.	Ensemble Gene ID	Associated Gene Name	Distance from TSS (bp)
1	ENSG00000166341	DCHS1	10	19	ENSG00000204220	PFDN6	1554
2	ENSG00000101439	CST3	140	20	ENSG00000206283	PFDN6	1554
3	ENSG00000204540	PSORS1C1	181	21	ENSG00000224782	PFDN6	1554
4	ENSG00000231094	PSORS1C1	181	22	ENSG00000235692	PFDN6	1554
5	ENSG00000233439	PSORS1C1	181	23	ENSG00000237335	PFDN6	1554
6	ENSG00000233734	PSORS1C1	181	24	ENSG00000185087	FAM169B	2903
7	ENSG00000235487	PSORS1C1	181	25	ENSG00000144230	GPR17	3086
8	ENSG00000137094	DNAJB5	388	26	ENSG00000213199	ASIC3	3282
9	ENSG00000019485	PRDM11	580	27	ENSG00000141524	TMC6	4056
10	ENSG00000186471	AKAP14	582	28	LRG_118	TMC6	4056
11	ENSG00000167693	NXN	913	29	ENSG00000233493	TMEM238	4307
12	ENSG00000110887	DAO	994	30	ENSG00000106258	CYP3A5	4330
13	ENSG00000174403	C20orf166-AS1	996	31	ENSG00000166510	CCDC68	5516
14	ENSG00000259988	C20orf166-AS1	996	32	ENSG00000105642	KCNN1	5921
15	ENSG00000110887	DAO	1059	33	ENSG00000163082	SGPP2	8580
16	ENSG00000237345	RP11-344N17.11	1332	34	ENSG00000154153	FAM134B	9199
17	ENSG00000269683	RP11-344N17.11	1332	35	LRG_363	FAM134B	9199
18	ENSG00000173467	AGR3	1446	36	ENSG00000220291	RP3-455E7.1	9425

CHAPTER IV

FOLDING DYNAMICS OF A HUMAN TELOMERE G-QUADRUPLEX INSIDE A NANO-CONFINEMENT

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED AS AN ARTICLE IN NATURE ANOTECHNOLOGY JOURNAL, *Nature Nanotechnology* **12**, 582–588 (2017). *ALL THE MATERIALS OF THE ARTICLE HAVE BEEN REPRINTED WITH THE COPYRIGHT PERMISSION OF THE NATURE.*

4.1 ABSTRACT

Molecular simulations suggest that the stability of a folded macromolecule increases in a confined space due to entropic effects. However, due to the interactions between the confined molecular structure and the walls of the container, clear-cut experimental evidence for this prediction is lacking. Here, using DNA origami nanocages, we show the pure effect of confined space on the property of individual human telomeric DNA G-quadruplexes. We induce targeted mechanical unfolding of the G-quadruplex while leaving the nanocage unperturbed. We find that the mechanical and thermodynamic stabilities of the G-quadruplex inside the nanocage increase with decreasing cage size. Compared to the case of diluted or molecularly crowded buffer solutions, the G-quadruplex inside the nanocage is significantly more stable, showing a 100 times faster folding rate. Our findings suggest the possibility of co-replicational or co-transcriptional folding of G-quadruplex inside the polymerase machinery in cells.

4.2 OVERVIEW

The folding and unfolding of macromolecules in confined space are prevalent inside cells. Notable examples include those in the chaperone or proteasome machinery, as well as in the lumen of translocons or entry/exit channels of ribosomes and polymerases. In an elegant experiment using GroEL-GroES chaperonins, Hartl and coworkers have demonstrated that protein folding is facilitated by nanocage confinement.[165-166] These observations have been ascribed to the confined volume effect.[167-168] In a confined volume, the entropy of unfolded biopolymer is decreased relative to the folded conformation, leading to a more stabilized folded state.[169-171] However, due to the extremely challenging experimental conditions that must prevent direct interaction between the nanocage and the confined biomolecule, and that must selectively unfold protein inside a cage, rather than the cage itself, this prediction has not been tested directly in nanocages made of biomacromolecules. In inorganic sol-gel matrices, some embedded proteins have shown increased thermal stability while others exhibit an opposite trend.[116] Controversial results were also observed for proteins contained inside nanometer-sized reversed micelles.[117]

Molecular crowding, another commonly occurring cellular condition,[172] is also known to stabilize a folded structure. It has been predicted that molecularly crowded condition is less stringent than confined volume for the stabilization effect.[170] Unlike the confined volume that isolates a molecule from surroundings, in molecularly crowded environment, an unfolded state is expected to accommodate more conformational variations since the molecule can survey the continuous space between neighboring

crowding agents. Again, due to the lack of experimental approaches, direct comparison between these two effects has not been carried out.

During telomere DNA elongation, the 5′-GGTTAG repeats are synthesized in a space with a cross section on the order of nanometers[118-119] before exiting telomerase. This spatial confinement raises the question whether a particular G-quadruplex is able to fold inside the telomerase during elongation. Even if the space is too small to accommodate a fully folded G-quadruplex, partial DNA folding[160, 173] may occur in the enzyme, which then folds into a specific G-quadruplex conformation as soon as the whole fragment exits the telomerase. Similar to telomerase, confined spaces regularly exist in exit channels of many polymerases for nucleic acids syntheses. To fully understand the newly discovered biological roles of G-quadruplexes inside cells[174-175], it becomes necessary to elucidate the transition kinetics and thermodynamic properties of G-quadruplexes in confined volume.

DNA origami techniques[123] have made it possible to introduce a well-defined 3D bio-nanostructure not hitherto available. In this work, we set out to resolve these fundamental questions by placing a G-quadruplex-hosting human telomeric fragment, 5′-$(TTAGGG)_4TTA$, inside DNA origami nanocages of different sizes (Figure 4.1). The repulsion due to the negative charges from the DNA origami and the telomeric sequence prevents the direct interaction between the two DNA, allowing clear-cut evaluation on the spatial confinement. To solely target the folded structure inside the cage, we used optical tweezers to selectively investigate the stability and transition kinetics of the G-quadruplex with respect to those in a crowded or a diluted buffer. Unlike thermal melting

which has a global effect on all species under study, mechanical unfolding[176] exerts localized effect on a species of interest without affecting other structures including the origami nanocage. This technique is therefore ideal to selectively investigate the stability and transition kinetics of biomacromolecules in confined space.

4.3 RESULTS AND DISCUSSION

4.3.1 G-quadruplex is formed inside DNA nanocages

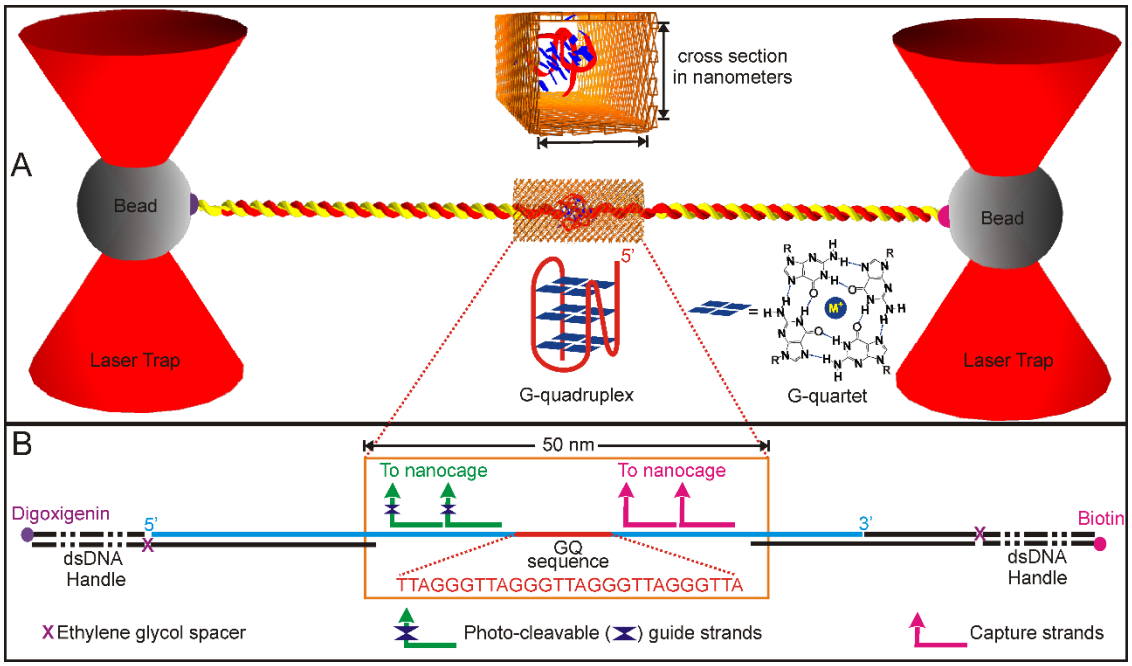


Figure 4.1. Experimental strategy to unfold human telomere G-quadruplex in a DNA nanocage. (A) A single molecular DNA construct is tethered between two optically trapped beads. G-quadruplex (GQ) forming sequence is contained inside the cuboid DNA nanocage. Top: side view of the nanocage that contains a G-quadruplex (PDB # 2HY9)[134]. Bottom: schematic structures of a telomere G-quadruplex and a G-quartet. M+ depicts a monovalent cation. (B) The DNA strand that contains the G-quadruplex

forming sequence and the two dsDNA handles labeled with biotin and digoxigenin at the two ends respectively for affinity attachments to optically trapped beads. See methods for details.

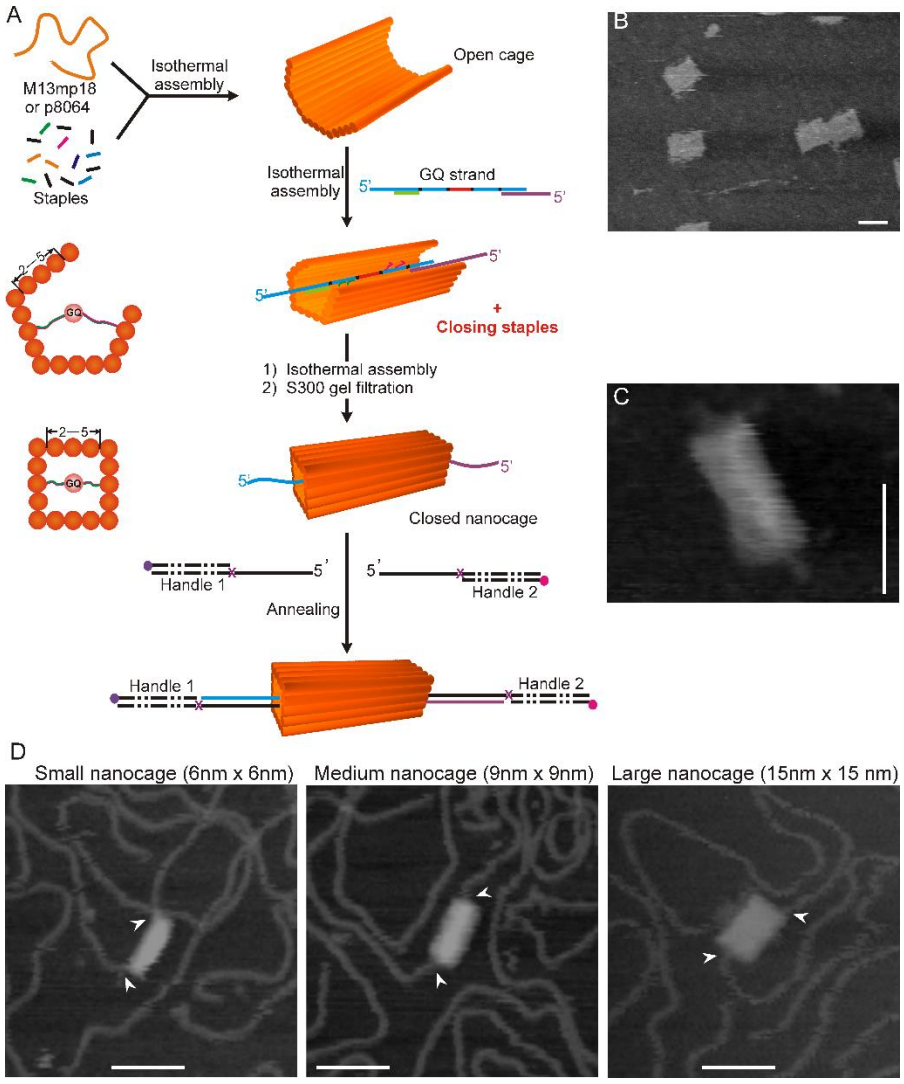


Figure 4.2. Synthesis and characterization of individual DNA origami constructs. (A) Schematics for the preparation of the single-molecule DNA construct with G-quadruplex hosting strand inside the DNA nanocage. Side views of the nanocage are shown to the left (each filled circle depicts a DNA helix in the nanocage). AFM images of an open

nanocage (B), a closed nanocage that contains the G-quadruplex hosting DNA fragment (C), and closed nanocages after annealing with *ds*DNA handles (D). Arrowheads indicate the *ds*DNA handles attached to the DNA nanocage. Scale bars represent 50 nm.

Since a typical G-quadruplex can be approximated by an 8-nm^3 cube, we designed DNA origami nanocages with inner diameters sufficient to accommodate a folded G-quadruplex (Figures 4.7-4.9). To restrict the conformation of the unfolded state, the cross section of the nanocage (~6-15 nm) was set smaller or comparable to the contour length of the telomere fragment. To allow the handles that transduce force from optically trapped beads to a tethered G-quadruplex, two ends of the nanocage were left open (Figure 4.1a).

DNA nanocages that contain G-quadruplex forming sequences were synthesized by an evolved strategy for DNA origami synthesis[109, 123] (see Materials and Methods) and characterized by AFM imaging (Figure 4.2). Each sample was subjected to UV (365 nm, ~0.22 W, 10 min) illumination to photo-cleave the guide strands, thereby releasing the tethering strain between the G-quadruplex hosting DNA and the nanocage due to the short guide/capture strands.

To test whether G-quadruplexes can fold inside nanocages thus prepared, we performed force ramping experiments on a medium nanocage (9 nm × 9 nm in cross section) using laser tweezers (see Materials and Methods). Two transition types were observed in force vs extension traces. In the first type, unfolding at high force (~40 pN) was followed by fast refolding (the left two traces in Figure 4.3a). In the second type, the unfolding occurred at lower force (~21 pN) was followed by slow refolding (the

rightmost trace in Figure 4.3a). We measured the change in contour length (ΔL) during the unfolding to identify species responsible for these features. The overall Gaussian average of ΔL was 8.2 ± 0.2 nm (Figure 4.3b, i), which is identical to that expected for unfolding a telomeric G-quadruplex (8.4 nm, see Materials and Methods for calculation). When we plotted the unfolding force histogram, two populations were identified at 38.6 ± 0.2 pN (major) and 21.0 ± 1.8 pN (minor) (Figure 4.3b, ii). While the 21 pN species is probably a free G-quadruplex without nanocage due to its identical mechanical stability with a reported quadruplex structure[127], the 38.6 pN species is likely a G-quadruplex in the nanocage.

To confirm this assignment, we performed three control experiments. In the first control, we enclosed a scrambled sequence, 5′-GTA GTG TGA TGA GTG TAG TGT GTA GTG, in the nanocage. The mechanical unfolding of this construct did not yield any unfolding events (see Figure 4.11 for population analysis). This confirmed that the unfolding events observed in Figure 4.3a were indeed due to those of G-quadruplexes. In the second control, we prepared the same G-quadruplex hosting sequence without nanocage. In this construct, we found a species with ΔL (8.3 ± 0.1 nm, Figure 4.3c (i)) and unfolding force (20.7 ± 1.4 pN, Figure 4.3c (ii)) identical to those of the minor population in the medium nanocage. However, the 38.6 pN species was not observed. These confirmed that the 21 pN and 38.6 pN species were indeed G-quadruplexes without and within medium nanocage, respectively. During the origami preparation, it is possible that a minor population may exist in such a way that a G-quadruplex is attached to a compromised nanocage, which is expected to show similar properties (slow refolding

kinetics (see below) and decreased unfolding force) as those without nanocage. We further evaluated the effect of Li^+, which is known to disfavour G-quadruplex formation,[177] on the medium nanocage construct. Since the origami buffer requires Mg^{2+} ions, we compared 100 mM KCl and 100 mM LiCl in the same Tris buffer that contains 10 mM $MgCl_2$. As expected, we found Li^+ reduced the formation of G-quadruplex in the medium nanocage to the same degree as that without nanocage (Figure 4.11). It is noteworthy that partially folded structures, such as G-triplexes,[160, 173] were not observed in these experiments. This may be attributed to the Mg^{2+} ions. Indeed, when we mechanically unfolded G-quadruplex without nanocage in the buffer with the same Mg^{2+} concentration, we did not observe partially folded structures (Figure 4.3c (ii)).

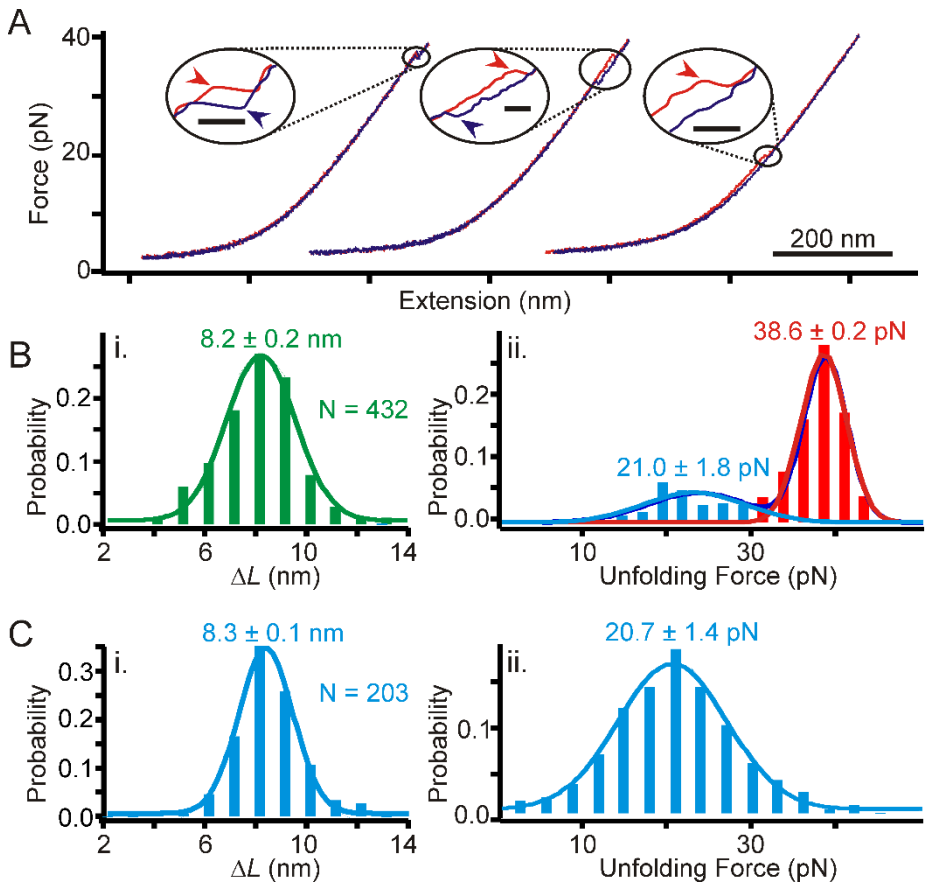


Figure 4.3. Mechanical unfolding of human telomeric G-quadruplexes at room temperature. (A) Typical stretching (black) and relaxing (gray) force vs extension curves of a G-quadruplex hosting sequence inside the medium DNA nanocage. Arrows in blowups depict unfolding or folding transitions. Scale bars represent 8.0 nm. (B) Histograms of (i) the change-in-contour-length (ΔL) and (ii) the unfolding force of the G-quadruplex inside the medium nanocage. (C) Histograms of (i) the change-in-contour-length (ΔL) and (ii) the unfolding force of the G-quadruplex without nanocage. Solid curves represent Gaussian fittings. Number of features is denoted by N.

In addition, force-ramping experiments were performed on the small (6 nm × 6 nm) and large (15 nm × 15 nm) DNA nanocages to further confirm the formation of G-quadruplex inside nanocages. Similar to those observed in the medium nanocage, unfolding forces in the small (Figure 4.12) and large (Figure 4.13) cages show two populations. While the populations with reduced force (~20 pN) depict free G-quadruplexes, the high force populations are likely G-quadruplexes inside nanocages.

4.3.2 G-quadruplex conformation varies with the size of nanocages

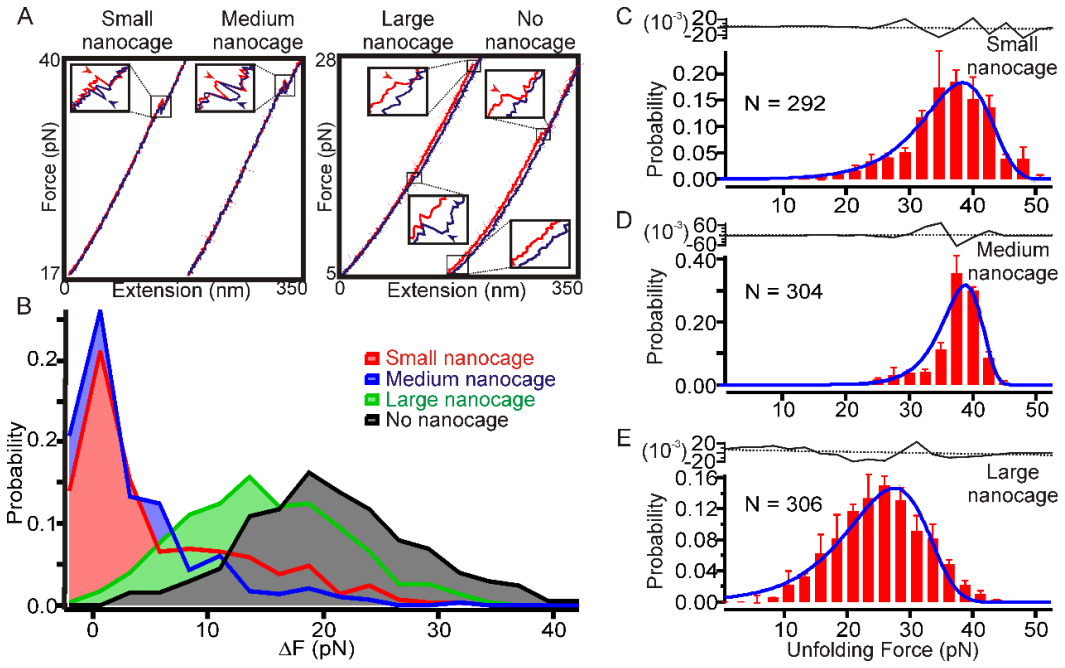


Figure 4.4. Refolding kinetics of G-quadruplexes in DNA nanocages. (A) Sections of force vs extension curves that contain unfolding/refolding features. Red and purple curves indicate the stretching and relaxing trajectories respectively. Zoomed unfolding/refolding transitions are shown in the insets. Each inset has 20 nm in x-axis. Arrowheads depict the unfolding (red) and refolding (purple) transitions. Refolding

transitions become slower when the confined space is larger. (B) Histograms of the difference in the force between unfolding and refolding transitions (ΔF) show increased hysteresis (or slower refolding kinetics) with nanocage size. (C-E), unfolding force histograms of the G-quadruplex in the small, medium, and large nanocages, respectively. Error bar in each histogram represents the standard deviation from three populations. Blue curves represent the fitting based on the equation proposed by Dudko. Black solid and dotted curves on top of each histogram represent the residual analysis and the linear fitting, respectively.

Using similar deconvolution methods (see Materials and Methods and Figure 4.4 C&E), we analyzed the high force populations only. In the small nanocage, the caged population was unfolded at a force (~36 pN) slightly lower than that of the medium nanocage (~39 pN). Interestingly, the hysteresis area between the stretching and relaxing force vs extension traces of the small nanocage is similar to that of the medium cage (Figure 4.4A, left panel and Figure 4.4B). Given the similar unfolding forces, the hysteresis area correlates with the refolding rate of the G-quadruplex. Therefore, the refolding rates seem to be close between the small and the medium nanocages. These similarities strongly support that G-quadruplexes with increased mechanical stability and faster refolding rate are also present inside small nanocages.

On the other hand, in the large nanocage (15 nm × 15 nm cross section, Figure 4.13), both the unfolding force (~26 pN) and the refolding rate (Figure 4.4A right panel and Figure 4.4B) are reduced compared to the medium nanocage. In fact, these values are located between the medium nanocage and G-quadruplexes without nanocage (see

Figures 4.4A&B), which indicate that both the mechanical stability and refolding kinetics of G-quadruplex increase with decreasing confined space (from 15 nm × 15 nm to 9 nm × 9 nm cross section). This trend, however, is no longer obvious after comparing the medium and the small nanocages (Figure 4.4). Since fully unfolded G-quadruplex has a contour length (~8 nm) comparable to the cross section of the two smaller nanocages, we speculate further reduction in the confined space may not significantly promote the folding of the G-quadruplex anymore. In addition, in such a confined volume, the conformation of folded structure may change to better accommodate the constrained space, which is supported by the smaller ΔL value observed in the small nanocage (6.9 ± 0.1 nm) (Figure 4.12) with respect to the other two nanocages (both 8.2 ± 0.2 nm).

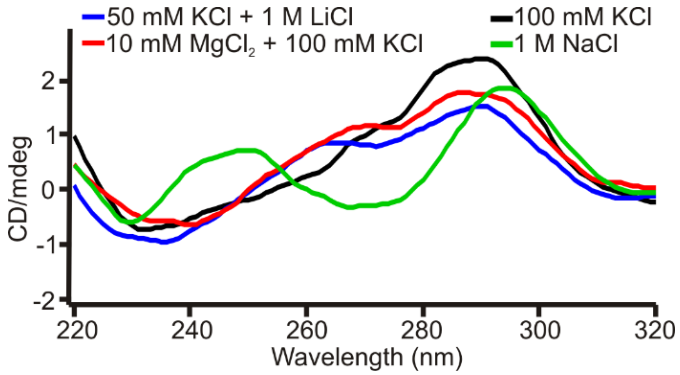


Figure 4.5. CD spectra of the human telomere G-quadruplex in different buffer conditions. Black (100 mM KCl), red (10 mM $MgCl_2$ and 100 mM KCl), and blue (50 mM KCl and 1 M LiCl) curves all show hybrid G-quadruplex conformation, whereas the green curve (1 M NaCl) shows basket G-quadruplex conformation.

The 8.2 nm ΔL population (Figure 4.13) is consistent with that expected for hybrid-1 G-quadruplex conformation[105] (8.4 nm, calculation see Materials and Methods).

A CD peak at 290 nm, a shoulder at 270 nm, and a trough at ~240 nm supported hybrid conformation (Figure 4.5). Further support for this conformation came from the ΔL measurement of the G-quadruplex inside the medium nanocage in a buffer known to promote the hybrid G-quadruplex in a protein nanopore[178]. We retrieved rather similar ΔL (7.8 ± 0.2 nm) and rupture forces (20.2 ± 2.8 pN and 37.2 ± 1.2 pN) (Figure 4.14) as those (8.2 ± 0.2 nm, 21.0 ± 1.8 pN and 38.6 ± 0.2 pN, Figure 4.3) obtained in the origami buffer used above, confirming the hybrid G-quadruplex conformation in the medium nanocage. Compared to expected ΔL values from all known telomeric G-quadruplex structures[105], we surmised that the 6.9 ± 0.1 nm population in the small nanocage is a basket G-quadruplex. To support this, the same telomeric fragment without nanocage was unfolded in a 20 mM Tris buffer (with 1 M NaCl at pH 7.9) that favors basket G-quadruplex conformation [178] (see CD spectra in Figure 4.5, green curve). The G-quadruplex was unfolded at ~15 pN with ΔL of 7.0 nm (Figure 4.15), which are expected values for the basket telomeric G-quadruplex[160]. Therefore, the matching ΔL values within (6.9 ± 0.1 nm, Table 4.6) and without (7.0 ± 0.2 nm, Table 4.6) the small nanocage support a basket G-quadruplex conformation inside the small nanocage. Interestingly, the same basket telomeric G-quadruplex topology was also observed in 6-10 nm reversed micelles.[179] Further verification of the conformation, however, requires unfolding of the structure from different directions.[105]

4.3.3 DNA nanocages increase the stability of G-quadruplex

Our results show that mechanical stability of G-quadruplex within nanocages (35.9-38.6 pN) is about twice than that without nanocage (~20 pN, see Table 4.6). Due to the negative charges in the quadruplex-hosting DNA and the DNA nanocage, the G-quadruplex is not expected to directly interact with the nanocage. Such experimental design allows us to probe the net effect of the confined space,[171] which has not been achieved previously in nanocages due to the possible interaction between the cage surface and the biomolecules contained inside. To quantify this effect, we calculated the change in free energy of unfolding G-quadruplex (ΔG_{unfold} = 13.4 ± 1.4, 14.2 ± 1.0, and 8.2 ± 0.4 kcal/mol, respectively, for the small, medium, and large nanocages, see Table 4.1) using Jarzynski equality equation[139] (see Materials and Methods and Figure 4.17 for the unfolding work histograms). These free energy calculations were verified by Bennett Acceptance Ratio[141-143], which gave almost identical values (see Figure 4.18 and Table 4.1). It is noteworthy that ΔG_{unfold} in the small nanocage is significantly different from that in the medium nanocage at the confidence level of 99.9% (student t-test), which is consistent with the finding that different conformations likely exist in these two nanocages as discussed above.

Compared to the free G-quadruplex in a diluted or a molecularly crowded (40% BSA) solution (ΔG_{unfold} of 7.1 ± 0.2 and 7.4 ± 0.4 kcal/mol, respectively,[107] see Table 4.1 and Figure 4.17), ΔG_{unfold} values for G-quadruplexes inside the small (13.4 ± 1.4 kcal/mol) and medium (14.2 ± 1.0 kcal/mol) nanocages are significantly higher. Given that hybrid-1 conformation is the most likely G-quadruplex structure formed in this DNA

sequence in the diluted (see Figures 4.5&4.14 discussed above) and the crowded (40% BSA, Figure 4.16) solutions, as well as in the medium and the large nanocages (due to similar ΔL ~ 8 nm, see Figure 4.3c (i) and Figure 4.13), such ΔG_{unfold} comparison provides a direct evidence that in confined space, thermodynamic stability of biomacromolecule is increased. This conclusion was further supported by the ΔG_{unfold} observed for a possible basket G-quadruplex in the small nanocage (13.4 ± 1.4 kcal/mol), which is significantly larger than the same conformation in 1 M NaCl solution (6.6 ± 0.3 kcal/mol, see Table 4.1). Based on the observation that G-quadruplexes likely share the same conformation in the large and the medium nanocages and that ΔG_{unfold} in the former cage is smaller than the latter (Table 4.1 and Figure 4.17A), we conclude that the confined space effect inversely scales with the size of the nano-confinement.

4.3.4 Confined space prevails over molecular crowding condition

Interestingly, the high-force unfolding of G-quadruplex in nanocage was always followed by fast refolding that occurred either reversibly or at slightly decreased forces. By contrast, for the G-quadruplex without nanocage, the refolding occurred in the force region < 5 pN, resulting in significant hysteresis between stretching and relaxing curves (Figure 4.4). To quantify the transition kinetics of the G-quadruplex inside nanocages, we applied the equation proposed by Dudko (see Materials and Methods) to fit the unfolding force histograms (Figure 4.4 c-e).[132, 145] The unfolding rate constant of the G-quadruplex in the medium nanocage (k_{unfold}) was found to be 4.6×10^{-6} s^{-1}, the distance from the folded state to the transition state ($x^{\dagger}$) was 0.34 ± 0.01 nm, and the change in

free energy of activation ($\Delta G^{\dagger}$) was 22.9 ± 1.0 kcal/mol (Table 4.1). The same analyses led to relevant kinetic information in the small and large nanocages (Table 4.1). Together with the ΔG_{unfold} values obtained above (Table 4.1), we constructed unfolding free energy profiles in Figure 4.6A. While G-quadruplexes in the medium and large nanocages share similar energy diagrams, the small nanocage shows a drastically different profile. This result is consistent with the fact that the G-quadruplex in the small nanocage likely assumes a basket conformation which is rather different from the hybrid topology in the larger nanocages (see Figure 4.15 discussed above).

Table 4.1. Thermodynamics and transition kinetics of human telomere G-quadruplexes

	ΔG_{unfold} (kcal/mol)	$\Delta G_{unfold}^{\dagger}$ (kcal/mol)	$x^{\dagger}$ (nm)	k_{unfold} (s^{-1})	k_{fold} (s^{-1})
Small nanocage	13.4±1.4 (0.2*)	22.3±0.6	0.20±0.01	$6.6\pm0.5 \times 10^{-4}$	$4.8\pm1.2 \times 10^{6}$
Medium nanocage	14.2±1.0(-0.7*)	22.9±1.0	0.34±0.01	$4.6\pm0.8 \times 10^{-6}$	$1.3\pm0.2 \times 10^{5}$
Large nanocage	8.2 ± 0.4 (0.1*)	16.9±0.1	0.16±0.02	$1.5\pm0.4 \times 10^{-2}$	$1.6\pm0.4 \times 10^{4}$
40% BSA	7.4± 0.4 (-0.5*)	16.3±0.6	0.17±0.01	$2.6\pm0.7 \times 10^{-2}$	$7.2\pm1.4 \times 10^{3}$
Diluted Buffer	7.1± 0.2 (-0.7*)	15.9±0.5	0.17±0.01	$2.8\pm1.0 \times 10^{-2}$	$4.7\pm0.6 \times 10^{3}$
1M NaCl	6.6± 0.3 (-0.4*)	14.7±0.6	0.36±0.04	$7.4\pm1.8 \times 10^{-3}$	$5.3\pm1.7 \times 10^{2}$

* Bias for each ΔG_{unfold} (see Materials and Methods for calculation).

To directly compare the confined space effect and the crowding effect on the G-quadruplex, we performed mechanical unfolding of the same G-quadruplex hosting sequence in a crowded (40% BSA)[107] and a diluted buffer. Changes in the contour length of the G-quadruplex in the crowded (ΔL = 7.9 ± 0.4 nm, Figure 4.16) and the diluted (ΔL

= 8.3 ± 0.1 nm, Figure 4.3c (i)) buffers suggest the formation of hybrid-1 G-quadruplex in both solutions after comparison with known telomeric G-quadruplex structures[105]. These assignments are fully consistent with those determined by more rigorous single-molecule finger-printing approaches[105, 107]. Dudko fitting on the unfolding force histograms (Figure 4.6B&C) and Jarzynski calculation on the unfolding work histograms (Figure 4.17) allowed us to obtain kinetic and thermodynamic information respectively, which were used to construct unfolding free energy profiles of the hybrid-1 G-quadruplex in the diluted, the crowded, and the medium-caged environments (Figure 4.6D). Compared to the unfolding energetic profiles of the G-quadruplex in the diluted (dotted) and crowded (40% BSA, solid) solutions, it is clear that due to the significantly higher activation ($G^{\dagger}_{unfold}$) and unfolding (G_{unfold}) free energies, G-quadruplex in the medium nanocage (dashed) is kinetically and thermodynamically more stable. The G-quadruplex in the medium nanocage has the largest $x^{\dagger}$ among the three conditions, implying a more elastic nature of the G-quadruplex in the nanocage. However, since the difference of the $\Delta G^{\dagger}_{unfold}$ between the medium nanocage and the crowded/diluted buffer predominates over the $x^{\dagger}$ difference, mechanical stability is still strongest for the caged G-quadruplex.

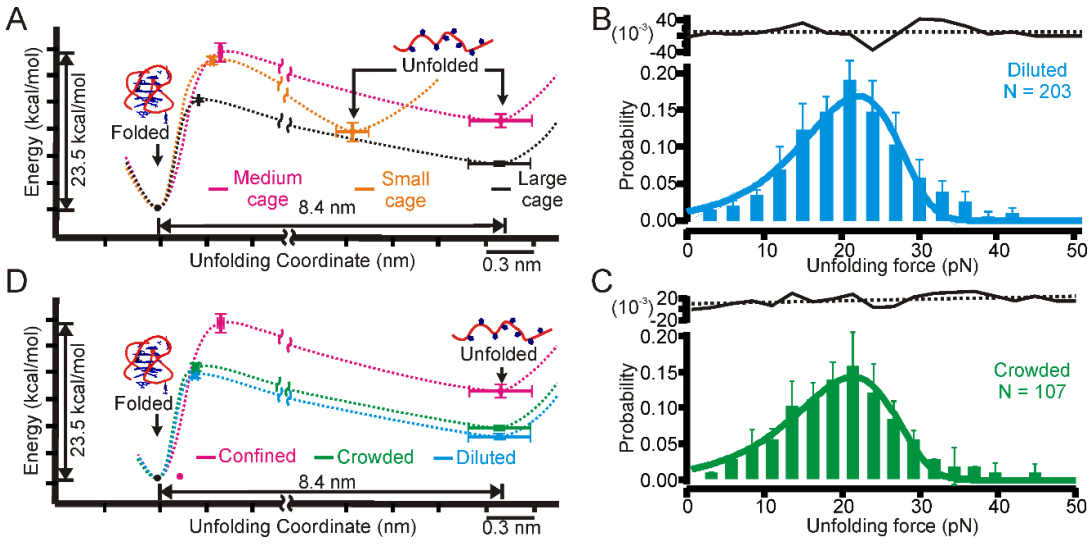


Figure 4.6. Transition kinetics and free energy diagrams of telomeric G-quadruplex in nanocages. (A) Free energy transition diagrams for G-quadruplexes in the small (dotted organge), medium (dashed pink), and large (solid black) DNA nanocages. Dudko fitting (solid curves) of the unfolding force histograms for the G-quadruplex in the diluted buffer (B), and 40% BSA buffer, (C). Error bar in each histogram represents the standard deviation from three populations. Black and dotted curves on top of each force histogram represent the residual analysis and the linear fitting, respectively. N depicts number of features. (D), Free energy transition diagrams for G-quadruplexes in medium DNA nanocage (dashed pink), in crowded buffer (solid green), and in diluted buffer (dotted blue). Arrows in (A) and (D) depict the free energy states of the folded and unfolded G-quadruplex.

Using a simple two-state model ($\Delta G_{unfold} = -RT\ln(k_{unfold}/k_{fold})$, where R is the gas constant, T is temperature and k_{fold} is the folding rate constant), the folding kinetics is estimated as $k_{fold} = 1.3 \times 10^5$ s^{-1} for the hybrid G-quadruplex inside medium nanocage.

This rate constant is about two orders of magnitude faster than that in the diluted or crowded buffer in which the same hybrid G-quadruplex is expected to form (Table 4.1). For possible basket G-quadruplex inside the small nanocage (100 mM KCl), the increase in the rate constant is even more pronounced than the free solution with different ionic conditions (1 M NaCl, see Table 4.1). These results suggest the possibility for the G-quadruplex to fold inside confined space of telomerase during the elongation of each G-quadruplex hosting sequence (~10 sec)[118, 180-181]. G-quadruplex in crowded solution only showed small increase in stability and folding kinetics than the diluted buffer (Table 4.1). These results are consistent with the finding that bivalent cations can destabilize telomeric G-quadruplex in a crowded condition, which otherwise is known to stabilize G-quadruplex.[107, 182] From the comparison of stabilities and transition kinetics of G-quadruplexes under different conditions, our finding directly supported the hypothesis[170] that confined space has a greater effect than molecular crowding to favour folded biomolecular structures. By assuming similar change in the unfolding enthalpy (ΔH_{unfold}) of the G-quadruplexes within and without medium nanocage, we estimated the entropic contribution (ΔS_{unfold}) of the confined space effect is on the order 0.024 kcal/mol.K (based on the equation, $\Delta S_{unfold} = (\Delta H_{unfold} - \Delta G_{unfold})/T$), see Materials and Methods). During the G-quadruplex folding without nanocage, ~70 water molecules[182] are lost to the less ordered bulk solution, providing an entropic driving force for the folding. However, the nanometer space in a cage and polyelectrolyte nature of the dsDNA strand and origami walls render water molecules inside the cage more ordered than the bulk solution (see Supporting Information for details). Therefore, compared to the bulk

solution, we expect a reduced contribution of water molecules to the folding/unfolding of biomolecules inside confined space.

4.4 CONCLUSIONS

In summary, human telomeric G-quadruplex was mechanically unfolded inside DNA origami nanocages that mimic exit channels in telomerase or polymerases. In confined space, G-quadruplex structures display fast folding kinetics with increased mechanical and thermodynamic stabilities, which suggests the possibility of simultaneous folding of G-quadruplex inside telomerase or other motor proteins during their respective enzymatic processes. These confined space effects inversely scale with the nanocage size and are stronger than molecularly crowded conditions.

4.5 SUPPORTING INFORMATION

caDNAno design of DNA origami nanocages

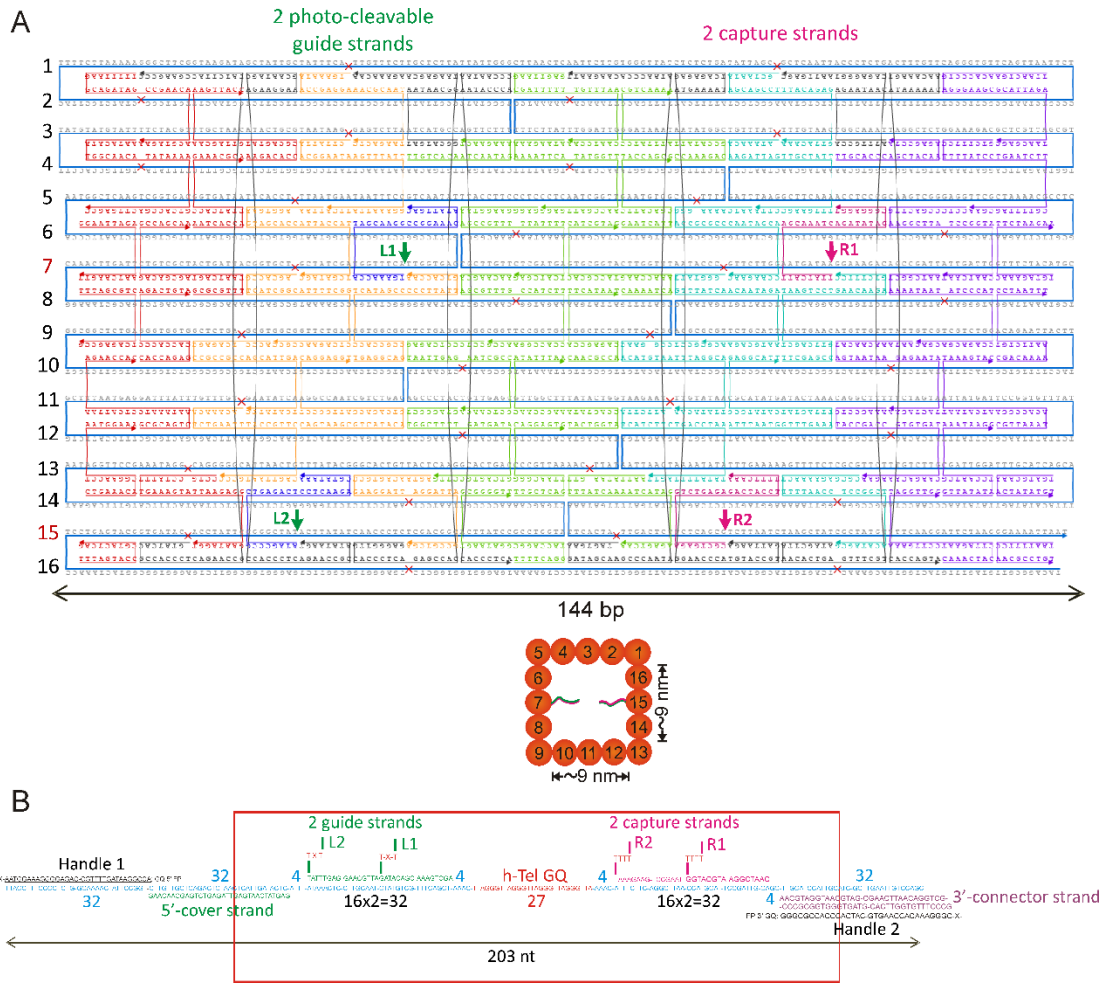
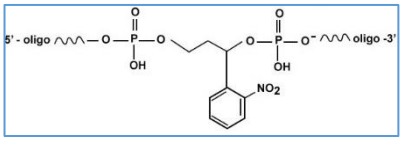


Figure 4.7. (A) caDNAno design of the medium (9 nm × 9 nm) DNA origami nanocage to incorporate the GQ hosting DNA strand inside the nanocage. L1 and L2 represent the connection sites of photo-cleavable guide strands and R1 and R2 represent the connection sites of capture strands to the GQ hosting strand from the wall of the nanocage. A fragment of M13mp18 (2385 nts) was used as a template strand. Sequences of the DNA staples are described in Table 4.2. Diagram shown to the bottom is the side view of the

DNA nanocage. (B) Detailed sequence of the 203-nt GQ hosting strand. The tethering sites to the DNA nanocage and the attachment sites to the dsDNA handles are depicted. Each number in (B) represents the length of a specific DNA fragment.

Table 4.2. Sequences of the DNA staples used to synthesize the medium (9 nm × 9 nm) DNA origami nanocage. Staples highlighted in green and pink contain 16-nt overhangs used to hybridize with guide and capture strands, respectively. Underlined sequences in the guide and capture strands represent the respective staple sequences that hybridize with the overhangs mentioned above. The red "TXT" and "TTTT" sequences in the respective guide and capture strands represent the linkers between the nanocage and the GQ strand, here 'X' in the guide strands depicts a photo-cleavable group[124],

X = 

Staple name	Sequence	Length
5X5 NC 01	TTTTTAAGGCAGATAGCCGAACAACGTAGAA	31
5X5 NC 02	TGAAATAACCGAGGAAACGCAATTAAGACTC	31
5X5 NC 03	GAGTTAAGCGATTTTTTGTTTAACAATCCAA	31
5X5 NC 04	GCTAATAAGCAGCCTTTACAGAGGTTACAAA	31
5X5 NC 05	TTAACTGAACACCCTGAGGGAAGCGCATTAGATAACGAGC	40
5X5 NC 06 C	AGAAGGAAGCAATAGCCCACCCTCAGAACCGCATATAAGT	40
5X5 NC 07 C	ATGAAAATTCAGAGAGGGAACCCATGTACCGTGATTAAGA	40
5X5 NC 08	AATACATATGGCAACATATAAAAGTCACCGACTTGAGCC	39
5X5 NC 09	CTTATTAACGGAATAAGTTTATTGGAAATTATTCA	35
5X5 NC 10 C	GGCATGATAATAACGGAATACCCACCCAATAACACCCTCA	40
5X5 NC 11	ATAAGAAAAAAATTCATATGGTTAACCGATTGAGG	35
5X5 NC 12	ATAAACAAAGATTAGTTGCTATTTTTGAAGCCTTA	35
5X5 NC 13 C	ATTTGCCAAGAATAACATAAAAACAACAAAGTCACCAGTA	40
5X5 NC 14	GTCTTTCCTTTTATCCTGAATCTTCGTTTTAGCGAA	36

5X5 NC 15	AAGACACCCGCAGTATGTTAGCAAAAGTTACC	32
5X5 NC 16	TTGTCACAATCAATAGAAAGAACT	24
5X5 NC 17	CCAAAGACGCCATATTATTTATCCCGTCAAAA	32
5X5 NC 18	TTGCACCCAGCTACAAAGAGCCTA	24
5X5 NC 19	TTAAAGGTGAAGTAGCACCATTACCATTCGATAGC	35
5X5 NC 20	GAGGGAAGGTAAAGCCGTTTTTATTTTTCATCGAG	35
5X5 NC 21	AATCAAAAGGGACCGCGCCCAATAGCACCAAGAAC	35
5X5 NC 22	CCTCCCGACTTGAAGGCTTATCCGGTACAATCAAT	35
5X5 NC 23	GGAATTAGAAATCAAGTT	18
5X5 NC 24	AATCACCATTATCACCGAAACGCA	24
5X5 NC 25	GAATCATTCGACATTCTACCAGCG	24
5x5 NC 26	AGCACCGTCATCGGCATTTTCGGCTCAGAACCGCC	35
5X5 NC 27 L1	TGAAACCATAGCAAGGCCGGAAACATATTGAC	32
5X5 NC 28	AACAAGCAAGCGTTTGCCATCTTCCACCGGAACCG	35
5X5 NC 29	GGGTATTGTTTATCAACAATAGAAATGCAGAACGC	35
5X5 NC 30 R1	TTATCATTAGCAAATCAGATATAGCGGGAGGT	32
5X5 NC 31	AATCGGCTAAAATAATATCCCATCCAGACGACGAC	35
5X5 NC 32	TGTAGAAACTTCTAAGAA	18
5X5 NC 33	TTTAGCGTCTCAGAGCCG	18
5X5 NC 34	GCGCGTTTTAATCAGTAGCGACAGGCCAGCAA	32
5X5 NC 35	CAAAATCAAAACCAAGTACCGCACCATCGTAG	32
5X5 NC 36	ACCCTCAGAGCCCGCCGCCAGCATTGATGATATTC	35
5X5 NC 37	CCTCCCTCAGAGTAATTGAGAATCGCCGCTCAACA	35
5X5 NC 38	GCCTCCGGAACACATGTAATTTAGGCATTATACAA	35
5X5 NC 39	AATAAACAACATAGTAATAAGAGAATATACTAGAA	35
5X5 NC 40	CACCAGAGCACCACCCAGACTGTA	24
5X5 NC 41	TGAGGCAGCCGCCACCTCATAGCCCCCTTATTGTCACCAA	40
5X5 NC 42	CAACGCCACAGAGCCATTCATAAT	24
5X5 NC 43	TTCGAGCCGTTCAGCTTAAGTCCTGAACAAGAGTCTTTCC	40
5X5 NC 44	CGACAAAAAATTCTGTCCTAATTT	24
5X5 NC 45	ACAAACATCTGAATTTACCGTTCCAGTTAATGCCC	35
5X5 NC 46	GTAGGGCTTGGCTTTTGATGATAGTCAGTGCCTTG	35
5X5 NC 47	ATTCTTATCATCTTCTGACCTAATATATTTTAGTT	35
5X5 NC 48	AAAGCCTGTACCGACCGTGTGATCAAGACAAAGAA	35
5X5 NC 49	GCGCAGTCAATAAATCCTCATTAAAGAACCAC	32
5X5 NC 50	CAGTAAGCGTCATACAGTCAGACGATTGGCCTCAGGAGGT	40
5X5 NC 51	TACTGGTACCAGTATAAAGCCAACATATTTAA	32
5X5 NC 52	ATTTAATGGTTTGAAATTTAGTATCATATGCGGAGGCATT	40
5X5 NC 53	CGTTAAATACACCGGAATCATAATTAAAGTAC	32
5X5 NC 54	CCTGCCTATTTTGAAAGTATTAAGAGGGAATAGGT	35

5X5 NC 55	AGTAACAGTGCCGAAGGATTAGGATTAGCCGTCGA	35
5X5 NC 56	AATTATAAGTTTTTATCAAAATCATAGAGAGTCAA	35
5X5 NC 57	CGCGAGAAAACTTTTTAACCTCCGGCTACATAGCG	35
5X5 NC 58	TGATGCAAAACTATATGT	18
5X5 NC 59	CTGAAACACGGAACCTAATGGAAA	24
5X5 NC 60	TTGCTCAGTTAACGGGCAGGAGTG	24
5X5 NC 61	GTTATATATCCAATCGAAATAAGG	24
5X5 NC 62 C	GTATCACGCCACCCTCAGAACCGTATCTTACCGAAGCCC	39
5X5 NC 63 L2	ATAGCCCGCTGAGACTCCTCAAGACGTATAAA	32
5X5 NC 64 C	GAGGGTTGCACCCTCAGAGCCACTAAGAGCAAGAAACAA	39
5X5 NC 65 C	TAGTGAAGATAGCAAGCCCAATAATAACCCACAAGAATT	39
5X5 NC 66 R2	CGCTGAGAGTCTGAGAGACTACCTTTTTCAAA	32
5X5 NC 67 C	ATAGCTTAAACACTGAGTTTCGTCAGAGGGTAATTGAGC	39
5X5 NC 68	TAATTTTCCAACGCCTGT	18
5X5 NC 69	TTTAGTACCCGTACTCAG	18
5X5 NC 70	TTTTCAGGTACCAGGCGGATAAGTGCGGGGTT	32
5X5 NC 71	CAAACTACCTTAGAATCCTTGAAATAGGTTGG	32
	2 photo-cleavable guide strands	
5X5 NC 27 L1 16nt	AGCTGAAACGACATAC-TXT-TGAAACCATAGCAAGGCCGGAAACATATTGAC	
5X5 NC 63 L2 16nt	ATTGCAAGGAGTTTAT-TXT-ATAGCCCGCTGAGACTCCTCAAGACGTATAAA	
	2 capture strands	
5X5 NC 30 R1 16nt	CAATCGGAATGCATGG-TTTT-TTATCATTAGCAAATCAGATATAGCGGGAGGT	
5X5 NC 66 R2 16nt	TTAAGCCTGAAGAAAT-TTTT-CGCTGAGAGTCTGAGAGACTACCTTTTTCAAA	

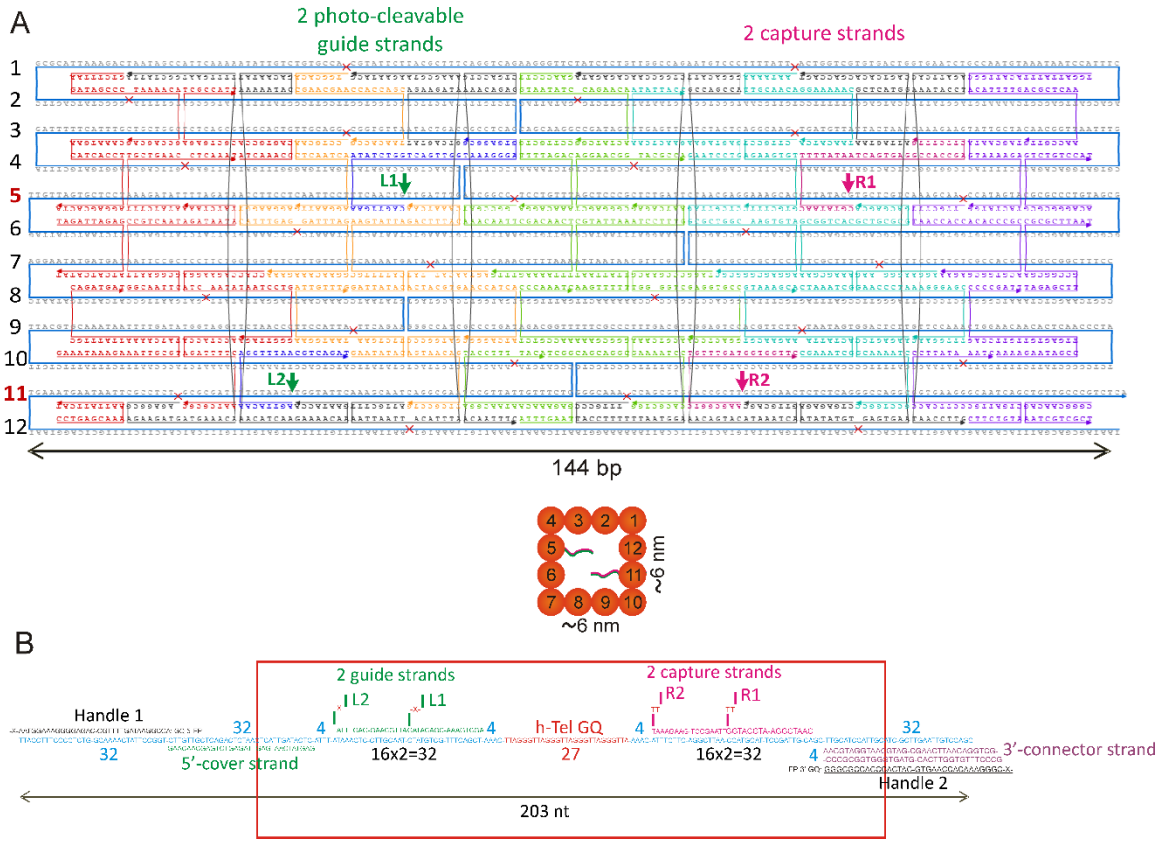
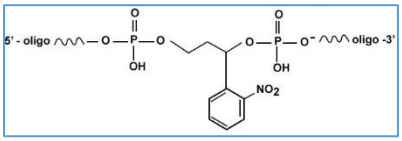


Figure 4.8. (A) caDNAno design of the small (6 nm × 6 nm) DNA origami nanocage to incorporate the GQ hosting DNA strand inside the nanocage. L1 and L2 represent the connection sites of photo-cleavable guide strands and R1 and R2 represent the connection sites of capture strands to the GQ hosting strand from the wall of the nanocage. A fragment of M13mp18 (1939 nts) was used as a template strand. Sequences of the DNA staples are described in Table 4.3. Diagram shown to the bottom is the side view of the DNA nanocage. (B) Detailed sequence of the 203-nt GQ hosting strand. The tethering sites to the DNA nanocage and the attachment sites to the dsDNA handles are depicted. Each number in (B) represents the length of a specific DNA fragment.

Table 4.3. Sequences of the DNA staples used to synthesize the small (6 nm × 6 nm) DNA origami nanocage. Staples highlighted in green and pink contain 16-nt overhangs used to hybridize with guide and capture strands, respectively. Underlined sequences in the guide and capture strands represent the respective staple sequences that hybridize with the overhangs mentioned above. The red "X" and "TT" sequences in the respective guide and capture strands represent the linkers between the nanocage and the GQ strand, here 'X' in the guide strands depicts a photo-cleavable group[124],

X = 

Staple Name	Sequence	Length
4X4 NC 01	GTCTTTAAGATAGCCCTAAAACAGCCAGCAG	31
4X4 NC 02	TGGCACACGAACGAACCACCAGCTAACACCG	31
4X4 NC 03	GAACCCTTGTAATATCCAGAACAACTATCGG	31
4X4 NC 04	AATAAAATTGCAACAGGAAAAACTAACATCA	31
4X4 NC 05	TGGATTATTTACATTGGCACATTTTGACGCTCAAAACCGTTG	42
4X4 NC 06 C	AAAAATACGACAATATAACATCAAGAAAACAAGAATACCA	40
4X4 NC 07 C	GCCAGCCAGGGACATTAACAGTACATAAATCAGTTGCAGC	40
4X4 NC 08	CAAATGAACATCACCTAGGAGCACTA	26
4X4 NC 09	CCTGCAACTCAATCAAGGAATTG	23
4X4 NC 10 C	GTCAGTATAGAAGATAAAACAGAGCTGACCTGACAATTTC	40
4X4 NC 11	CCTTGCTGTTTTAGACAAACAGGA	24
4X4 NC 12	CTTGCCTAGAAGTGTGTGCTTTC	23
4X4 NC 13 C	ATTAGTAAGCTCATGGAAATACCTAGATTCACTAACCTTG	40
4X4 NC 14	TAGCAATAGTAAAAGATACTATGG	24
4X4 NC 15	ATCAAACCCAGTGCCACGCTGAGATCGCCATT	32
4X4 NC 16	GAATCCTGGAGTAGAAGAACTCAAATATTACC	32
4X4 NC 17	ATATCTTTTGCTGAACCTCAAAT	23
4X4 NC 18	AGGAAGGTCATTTGAGGATTTAGGGAACAAAGAAA	35
4X4 NC 19 L1	CAGTTGAAATATCTGGTCAGTTGGTAAAGGGAGTGAGGCG	40
4X4 NC 20	GGCCGATAAACAATTCGACAACTTTTAAAAGTTTG	35

4X4 NC 21	CGGGAGCTAGGAACGGTACGCCA	23
4X4 NC 22	CTCGTTAGGCGCTGGCAAGTGTAAGGAGCGGGCGC	35
4X4 NC 23 R1	CGTATAACTTTTATAATCAGTGAGGCCACCGACTTCTTTG	40
4X4 NC 24	TTGCTTTTAACCACCACACCCGCACGTGGCGAGAA	35
4X4 NC 25	ACAGGGCGCGGTCTGTCCAT	20
4X4 NC 26	TAGATTAGAGATCATCATAT	20
4X4 NC 27	CCACCAGAAGGAATCAATATAATCCTGGGGTTAGA	35
4X4 NC 28	AGTAACATTATACTACGTGAACCATCAATCAGGGC	35
4X4 NC 29	TAGGCCCGAACGTGGGGTCGAGGTGCCGGACTCCA	35
4X4 NC 30	AGGAAGGGAAGGAACCCTAAAGGGAGCTTTGGAAC	35
4X4 NC 31	GGCAATTCGCGGAATTCCGTCAATAGATAATATATCTAAA	40
4X4 NC 32	GATTATACCATTTTGCAAGTATTAGACTTTACCAAATCAA	40
4X4 NC 33	AAGTTTTTTTATTAATCGTATTAAATCCTTTGAATCAGAG	40
4X4 NC 34	CTAAATCGAAAGCGAAGCGGTCACGCTGCGCGGACGAGCA	40
4X4 NC 35	TAGAGCTTGCCGGCGACGCGCTTAAT	26
4X4 NC 36	ACCTACCAAGATTTTCAATCGCGC	24
4X4 NC 37	GATGGCCCAGTAACAGTTCGCCTG	24
4X4 NC 38	ACGTCAAAGAAAATCCCACGCTGG	24
4X4 NC 39	AAGAGTCCGCAAAATCGCCTGGCC	24
4X4 NC 40	GAAATAAAGAAATTGCGTTATCAAAATTATTTGCCAGATGAT	42
4X4 NC 41	GAATATACTTCTGAATAATGGAAATTGTTTG	31
4X4 NC 42	TCGGCAGCAGGCGGGCGAAAAACCGTCTCCCAAATC	36
4X4 NC 43	CGAAATCGACTATTAAAGAACGTGTAAAGCA	31
4X4 NC 44	AAAAGAATAGCCTTGAGTGTTGTTCCAGCCCCGATT	36
4X4 NC 45 C	AGAGGCGAGAAGATGATGAAACATTTTGAATGGCTATTA	39
4X4 NC 46 L2	AGTTACAAAGGTTTAACGTCAGAT	24
4X4 NC 47 C	ATTGCTTTAATTAATTACATTTAAAAGCGTAAGAATACG	39
4X4 NC 48 C	TTTGCCCTACCTTTTTTAATGGACTGGCCAACAGAGATA	39
4X4 NC 49 R2	AAGCGGTCTGTTTGATGGTGGTTC	24
4X4 NC 50 C	CTGAGAGAATATATGTGAGTGAACAGTCACACGACCAGT	39
4X4 NC 51	CGGGCAACAGAATCGTCGCT	20
4X4 NC 52	CCTGAGCAAAAATTATTCAT	20
4X4 NC 53	ATTTGAATGAGAAACAATAACGGATACCTTTTACA	35
4X4 NC 54	CTTCTGTACTGATTGCCCTTCACCCCTTATAAATC	35
	2 photo-cleavable guide strands	
4X4 NC 19 L1 16nt	AGCTGAAACGACATAC-X-CAGTTGAAATATCTGGTCAGTTGGTAAAGGGAGTGAGGCG	
4X4 NC 46 L2 16nt	ATTGCAAGGAGTTTAT-X-AGTTACAAAGGTTTAACGTCAGAT	
	2 capture strands	

4X4 NC 23 R1 16nt	CAATCGGAATGCATGG-TT- CGTATAACTTTTATAATCAGTGAGGCCACCGACTTCTTTG	
4X4 NC 49 R2 16nt	TTAAGCCTGAAGAAAT-TT- AAGCGGTCTGTTTGATGGTGGTTC	

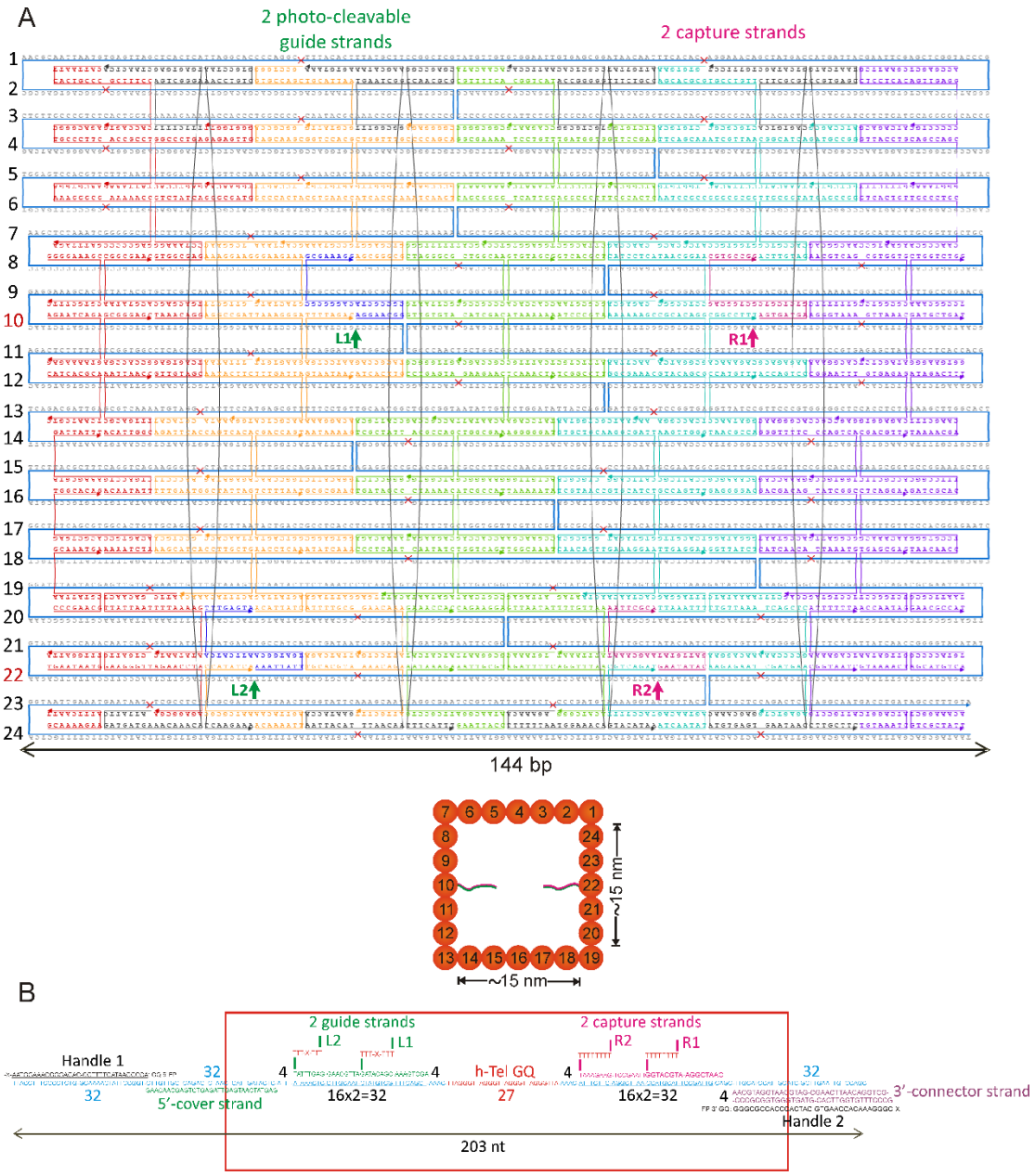
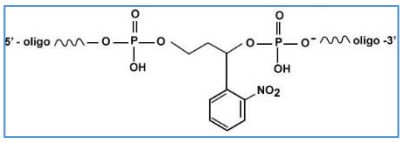


Figure 4.9. (A) caDNAno design of the large (15 nm × 15 nm) DNA origami nanocage to incorporate the GQ hosting DNA strand inside the nanocage. L1 and L2 represent the

connection sites of photo-cleavable guide strands and R1 and R2 represent the connection sites of capture strands to the GQ hosting strand from the wall of the nanocage. A fragment of p8064 (3624 nts) was used as a template strand. Sequences of the DNA staples are described in Table 4.4. Diagram shown to the bottom is the side view of the DNA nanocage. (B) Detailed sequence of the 203-nt GQ hosting strand. The tethering sites to the DNA nanocage and the attachment sites to the dsDNA handles are depicted. Each number in (B) represents the length of a specific DNA fragment.

Table 4.4. Sequences of the DNA staples used to synthesize the large (15 nm × 15 nm) DNA origami nanocage. Staples highlighted in green and pink contain 16-nt overhangs used to hybridize with guide and capture strands, respectively. Underlined sequences in the guide and capture strands represent the respective staple sequences that hybridize with the overhangs mentioned above. The red "TTTXTTT" and "TTTTTTTT" sequences in the respective guide and capture strands represent the linkers between the nanocage and the GQ strand, here 'X' in the guide strands depicts a photo-cleavable group[124],

X = 

Staple Name	Sequence	Length
7X7 NC 01	CATTAATTCACTGCCCGCTTTCCCACCAGTG	31
7X7 NC 02	GCCTGGGGTGCCAGCTGCATTAAGCGTATTG	31
7X7 NC 03	ACAACATACGTTTTCACGGTCATCCAGAATG	31
7X7 NC 04	GTGTGAAAGCACGCGTGCCTGTTCTGCGCGC	31
7X7 NC 05	ACCGAGCTCGAATTCGCTCCTCACAGTTGAGGCCCCCTGC	40
7X7 NC 06	AGACGGGCTGCCCTTCACCGCCTTATTAAAG	31
7X7 NC 07 C	TTTCTTTTAGTCGGGAAACCTGTCGTGCCTAATCAAGAAA	40

7X7 NC 08	GGCGCCACAGCAAGCGGTCCACGAGTGTTGT	31
7X7 NC 09 C	GGCGGTTTTGAATCGGCCAACGCGCGAGCCGGTTTCATTT	40
7X7 NC 10	CGGCGGGCGGCGAAAATCCTGTTATAAATCA	31
7X7 NC 11 C	TGCTGCGGACCGGGGGTTTCTGCCATTGTTATGTACATAA	40
7X7 NC 12	CTGTGCATTCAGCAAATCGTTAACCTTACAC	31
7X7 NC 13 C	CAGTGTCACTTCGCGTCCGTGAGCTAATCATGCTTGCTTC	40
7X7 NC 14	ATCAGACGGTTACCTGCAGCCAGCGGGTAAAG	32
7X7 NC 15	AACGTGGAAAAGGGCGAAAAACCGAGCCCCCGATTTAGA	39
7X7 NC 16	GAGTCCACGGCCCTGAGAGAGTTGGGGTGGTT	32
7X7 NC 17	TCCAGTTGCCCACTACGTGAACCGCCGTAAAGCAC	35
7X7 NC 18	TAGGGTTGCTGGTTTGCCCCAGCACGGGGAGA	32
7X7 NC 19	AAAGAATACAGCGGGGTCATTGCTGGAGGTGTCCA	35
7X7 NC 20	AATCCCTTTGATGGTGGTTCCGAACTCTGTGG	32
7X7 NC 21	TGGTGTGAATCCGCCGGGCGCGGGTGGTGCCATCC	35
7X7 NC 22	TAAACATCCGGCATCAGATGCCGGATCCAGCG	32
7X7 NC 23	GTTTCTTTGTCACTGTTGCCCTGCCAACCGCAAGAA	36
7X7 NC 24	TAAATCGGAACAAAGGAAGGGAAGAAAGTACTATG	35
7X7 NC 25	GCATTTTTTGGGCTAGGGCGCTGGCAACCCGCCGC	35
7X7 NC 26	CACGCAACCAGCATCCTCATAACGGAACCTCCGGC	35
7X7 NC 27	TGCCAACGGCAGAACGTCAGCGTGGTGCGCGGTCC	35
7X7 NC 28	GGGGAAAGCCGTGCTTTC	18
7X7 NC 29	GTGGCGAGCCTAAAGGGTCTATCAGGGCGATGTGGAACAA	40
7X7 NC 30	AGCGGGCGGTCGAGGTATCACCCAAATCAAGTGCCCGAGA	40
7X7 NC 31	GTCACGCTCTTACGGCAGGCGCTTTCGCACTCATCGGCAA	40
7X7 NC 32	ACTTGTAGCACCGTCGTTGCGGTATGAGCCGGGCTCGTCA	40
7X7 NC 33	GTTGCTTAGGCCGATTAAAGGGAGAAGTGTT	31
7X7 NC 34	GCTTAATGGTTGTGTACATCGACCGCCAGCA	31
7X7 NC 35	CAGAGCAAAAAAAGCCGCACAGGGCGGATCA	31
7X7 NC 36	GTTTTTTCAGGGTAAAGTTAAACATAGCTCT	31
7X7 NC 37	TCCGGCAAACTGGTCTGG	18
7X7 NC 38	AGAATCAGAGTAAAAGAG	18
7X7 NC 39	TAAACAGGTGACGAGCACGTATAACGGCGAAC	32
7X7 NC 40 L1	AGGAACGGCGCCGCTACAGGGCGCGCGAAAGG	32
7X7 NC 41	ATCCCGTAGCGCGTAACCACCACAGTGTAGCG	32
7X7 NC 42 R1	AGTGATGAGTCTCGTCGCTGGCAGCGTGCCGG	32
7X7 NC 43	TTTATAAAATACTTCTTTGATTAAACGCTCATGGA	35
7X7 NC 44	GTTGGGCGCCTGAGTAGAAGAACACAATATTACCG	35
7X7 NC 45	AACTTAAGGCGAAACGTACAGCGTCACCGGAAAC	34
7X7 NC 46	CACGGAAACGGAATTTGTGAGAGCAGAGGTGGAGC	35
7X7 NC 47	GGTGAAGGGGATGCTGAT	18

7X7 NC 48	CATCACGCATCAATCGTC	18
7X7 NC 49	GTTGTAGCTCAGTGAGGCCACCGAGCGGGAGC	32
7X7 NC 50	ATCACTTGTACGCCAGAATCCTGATTTTAGAC	32
7X7 NC 51	TCGGCCTTATTTCTGCTCATTTGCATAAAAAA	32
7X7 NC 52	ACCAGTCCAAGAGACGCAGAAACACGGCCTTT	32
7X7 NC 53	AATACCTACATAGATTCACCAGTCACAAGAGATAG	35
7X7 NC 54	CCAGCCATTGCATCGCTATTACGCCAGATCGGTGC	35
7X7 NC 55	AATCGCTGGTAAGTGCTGCAAGGCGATTCATTCAGG	36
7X7 NC 56	CGCCACGGGAACGGGTTTTCCCAGTCATGGTGCCG	35
7X7 NC 57	ACATTGGCTTTGACGCAATTAACC	24
7X7 NC 58	AATAAAAGACAGGAAAGTAATAAC	24
7X7 NC 59	AGGGGGATTATCCAGATCAAACTA	24
7X7 NC 60	TAACGCCAGGATAACCCCATGTTT	24
7X7 NC 61	TAAAACGACAAGCTTTATAGACTTT	25
7X7 NC 62	AACCCTTTTTGAATGGCTATTAGTTAACACC	31
7X7 NC 63	GGGCCTCTTGATAGCCCTAAAACCAGAAGAT	31
7X7 NC 64	CTGCGCATCGTAACCGTGCATCTGTGTAGAT	31
7X7 NC 65	GAAACCAGGACGACAGTATCGGCGCGGATTG	31
7X7 NC 66	ACAATATTCTGACCTGAAAGCGTAGATTATTT	32
7X7 NC 67	GCGCGAACGGACATTCTGGCCAACCGACCAGT	32
7X7 NC 68	TAAAAATAACTGTTGGGAAGGGCGCTGGCGAA	32
7X7 NC 69	GAGGGGACGCAAAGCGCCATTCGCAAGTTGGG	32
7X7 NC 70	GATCGCACTTTCCGGCACCGCTTCCGACGTTG	32
7X7 NC 71	GCCTGCAAAGCATCACCTTGCTGACTTTACAAACA	35
7X7 NC 72	AAAACAGACCCTCAATCAATATCAGATAATACATT	35
7X7 NC 73	GGGCGCACAACAGTTGAAAGGAACACTAACAACTA	35
7X7 NC 74	ACCGTAATCATCAACATTAAATGTCCTGTAGCCAG	35
7X7 NC 75	AAAATCTAACAGTGCCACGCTGAGTGGCACAG	32
7X7 NC 76	ATATCAAAGGTGAGGCGGTCAGTATCTTTAAT	32
7X7 NC 77	TGGCAAATCCGAACGAACCACCAGATCGCCAT	32
7X7 NC 78	GGTTATCTGGGATAGGTCACGTTGGCCAGTTT	32
7X7 NC 79	TAACAACCTCCGTGGGAACAAACGCTCAGGAA	32
7X7 NC 80	ATTCGACAACTTTATTAATTTTAAAAGATATAATC	35
7X7 NC 81	TGAGGATTTAGAATTTTGCGGAACAAAATTCCTGA	35
7X7 NC 82	ATAGATTAGAGGTTAATATTTTGTTAAATTTAAAT	35
7X7 NC 83	CTTTAAAATATCTTGTTAAATCAGCTCGCCCCAAA	35
7X7 NC 84	TAATTCGCGGAACGCCAT	18
7X7 NC 85	CCCGAACGCGTATTAAGCAAATGA	24
7X7 NC 86	ACATTATCAGTATTAGAACCTCAA	24
7X7 NC 87	CAGAAGGACCGTCAATTGGTCAGT	24

7X7 NC 88	TTAAATTTTTTAGGAGTTGAGGAA	24
7X7 NC 89	ACCAATAGTCTGGCCTTGAGCGAG	24
7X7 NC 90	CTGATTGGAAGGGTTAGAACCTAAGAGGCGA	31
7X7 NC 91	TTATCAGATGCACGTAAAACAGAATTGCTTT	31
7X7 NC 92	TGTAAACAGATTTTCAGGTTTAAACATCGGG	31
7X7 NC 93	AACAGGAAAAGAGAATCGATGAAAGAGTCTG	31
7X7 NC 94	TGTACCCCGAGCATGTCA	18
7X7 NC 95	TGAATAATGTTTGGATTA	18
7X7 NC 96 L2	AAATTATTTGATGGCAATTCATCATTTGAGTA	32
7X7 NC 97	AATTGCGTGCGGAATTATCATCATGAAACCAC	32
7X7 NC 98 R2	GAATATACGATTGTATAAGCAAATAATTCGCA	32
7X7 NC 99	GTAAAACTGTTGATAATCAGAAAAATTTTTTA	32
7X7 NC 100 C	ATTATTCGATGATGAAACAAACATGAGTGAGCTAACTCA	39
7X7 NC 101 C	GAATACCAAATTACATTTAACAAAAGCATAAAGTGTAAA	39
7X7 NC 102 C	AGAAACATTTTTTAATGGAAACACCGCTCACAATTCCAC	39
7X7 NC 103 C	GAGCAAACATGTGAGTGAATAACGTCATAGCTGTTTCCT	39
7X7 NC 104	ACAAAGGCTGTCGCTATT	18
7X7 NC 105	GCAAAAGAAATTTCAATT	18
7X7 NC 106	ACAAAATTAGTTACAAAATCGCGCCCATATCA	32
7X7 NC 107	GAATTACCATAACGGATTCGCCTGAATAAAGA	32
7X7 NC 108	ATCAATATAGTAACAGTACCTTTTCGTCAGAT	32
7X7 NC 109	TGTAAATCATCAGGTCATTGCCTGCGGTAATC	32
	2 photo-cleavable guide strands	
7X7 NC 40 L1 16nt	AGCTGAAACGACATAC-TTTXTTT- AGGAACGGCGCCGCTACAGGGCGCGCGAAAGG	
7X7 NC 96 L2 16nt	ATTGCAAGGAGTTTAT-TTTXTTT- AAATTATTTGATGGCAATTCATCATTTGAGTA	
	2 capture strands	
7X7 NC 42 R1 16nt	CAATCGGAATGCATGG-TTTTTTTT- AGTGATGAGTCTCGTCGCTGGCAGCGTGCCGG	
7X7 NC 98 R2 16nt	TTAAGCCTGAAGAAAT-TTTTTTTT- GAATATACGATTGTATAAGCAAATAATTCGCA	

Table 4.5. Single-stranded DNA templates for the preparation of nanocages.

Nanocages	Plasmid, Restriction enzymes, Length	Complementary strands
Small nanocage (4HB × 4HB) (6 nm × 6 nm)	M13mp18, BsmBI/PacI, 1841 nts	CTTTTCACCAGTGAGACGGGCAACAGCT (BsmBI) TAAGGGAAAATTAATTAATAGCGACGAT (PacI)
Medium nanocage (5HB × 5HB) (9 nm × 9 nm)	M13mp18, PacI/BsmI, 2385 nts	TAAGGGAAAATTAATTAATAGCGACGAT (PacI) CTGTCTGTGGAATGCTACAGGCGTTG (BsmI)
Large nanocage (7HB × 7HB) (15 nm × 15 nm)	p8064, PacI/Bg1II, 3624 nts	TAAGGGAAAATTAATTAATAGCGACGAT (PacI) TAGCCTTTGTAGATCTCTCAAAAATA (Bg1II)

2.5.3.2 Synthesis of the 203-nt Single-Stranded DNA Containing GQ Forming Sequence

The 203-nt single-stranded DNA containing G-quadruplex motif was synthesized by using splint ligation of the four short DNA oligonucleotides (Figures S1b/S2b/S3b and S4). All the DNA oligonucleotides were first phosphorylated, followed by ligation using the splint DNA oligos. Finally, the splint DNA oligos were removed by denatured PAGE to obtain pure 203-nt single-stranded DNA.

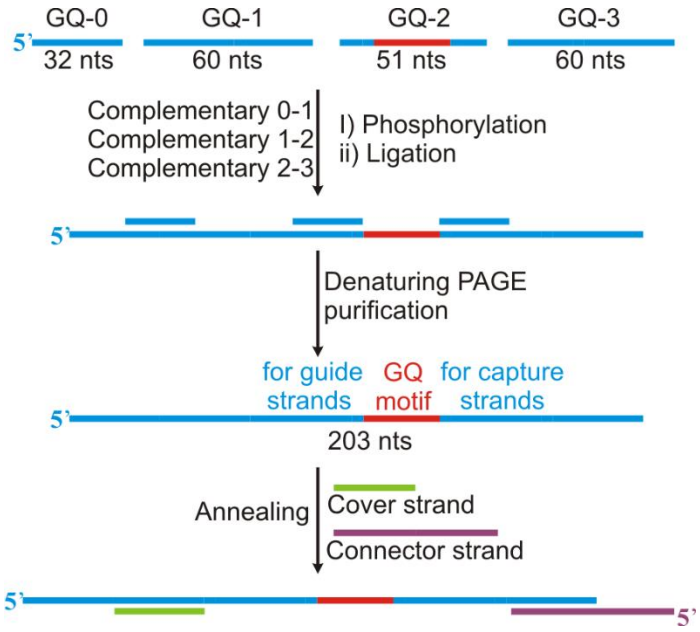


Figure 4.10. Preparation flow chart for the 203-nt DNA fragment (Figure S1b/S2b/S3b) that contains human telomere G-quadruplex motif. This construct passes through the DNA nanocage (see Fig. 1 in the main text).

Sequences of the DNA oligos used in Figure S4 are listed below. The same color codes are used for convenience.

GQ-0: 5'-TTA CCT TTC CCC TCT G-GC AAA ACT ATT CCG GT

GQ-1: 5'-CTT GTT GCT CAG ACT C-TA ACT CAT TGA TAC TC-A TTT- ATA AAC TC-C TTG CAA T-GT ATG TCG

GQ-2: 5'-TTT CAG CT-A AAC TTA GGG TTA GGG TTA GGG TTA GGG TTA AAA C-AT TTC TTC

GQ-3: 5'-AGG CTT AA-C CAT GCA T-TC CGA TTG –CAG C-TT GCA TCC ATT GCA TC-G CTT GAA TTG TCC AGC

Complementary 0-1: 5'-CTG AGC AAC AAG –ACC GGA ATA GTT

Complementary 1-2: 5'-GTT T-AG CTG AAA- CGA CAT AC-A TTG

Complementary 2-3: 5'-ATG G-TT AAG CCT- GAA GAA AT-G TTT

5'-Cover strand: 5'- GAG TAT CAA TGA GTT AGA GTC TGA GCA ACA AG

3'-Connector strand: 5'- GCC CTT TGT GGT TCA C-GT AGT GGG TGG CGC CC-G CTG GAC AAT TCA AGC –GAT GCA ATG GAT GCA A

2.5.3.3 Synthesis of Double Stranded DNA Handles

Two double-stranded 2520 bp DNA handles were prepared by PCR amplification of the pET-26b (+) plasmid. The forward primer of each handle consists of "5'-Connecting (linker) Sequence-O-$(CH_2)_2$-O-$CH_2)_2$-O-$(CH_2)_2$-O-Primer Sequence" (see below). This design introduced a single-stranded 5'-overhang in one end of each handle, which serves as a staple to hybridize with the nanocage. The other end of the handle is labeled with biotin or digoxigenin using 5'-biotin or 5'-digoxigenin modified reverse primers respectively.

Following DNA oligonucleotides were used as primers to prepare the biotin labeled double stranded DNA handle.

Forward Primer: 5'-ACC GGA ATA GTT TTG C-CA GAG GGG AAA GGT AA-**X**-CGC CGA TCA ACT GGG TGC CAG CGT

Reverse Primer: 5'-Biotin-GGG TTC GTG CAC ACA GCC CAG CTT

Following DNA oligonucleotides were used as primers to prepare the digoxigenin labeled double stranded DNA handle.

Forward Primer: 5'- GGG CGC CAC CCA CTA C-GT GAA CCA CAA AGG GC-**X**-CGC CGA TCA ACT GGG TGC CAG CGT

Reverse Primer: 5'-Dig-GGG TTC GTG CAC ACA GCC CAG CTT

X = O-$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$-O

2.5.3.4 Population analysis of the unfolding features

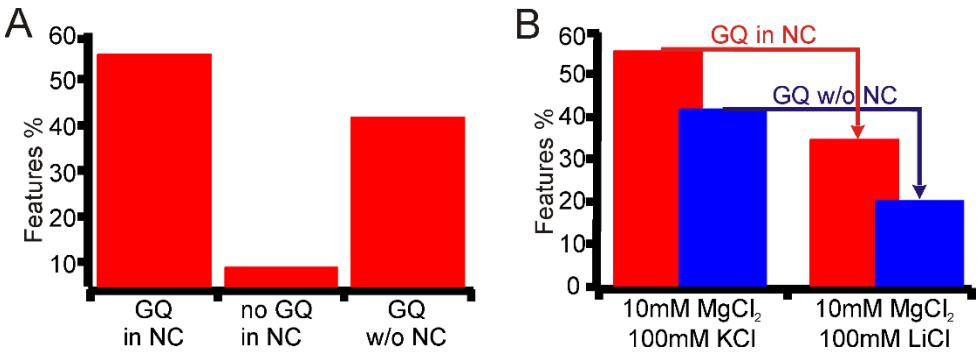


Figure 4.11. Analysis of the percentage of unfolding features. (A) Comparison of the percentage of the unfolding features for G-quadruplex (GQ) (ΔL 8.0 ± 0.4 nm) in F-X curves during the mechanical unfolding experiments for GQ within medium DNA nanocage (GQ in NC), no GQ within medium DNA nanocage (no GQ in NC), and GQ without DNA nanocage (GQ w/o NC). (B) Comparison of the percentage of the unfolding features in an origami buffer (20 mM Tris) that consists of 10 mM $MgCl_2$ and 1 mM EDTA with 100 mM KCl or LiCl (pH =7.8).

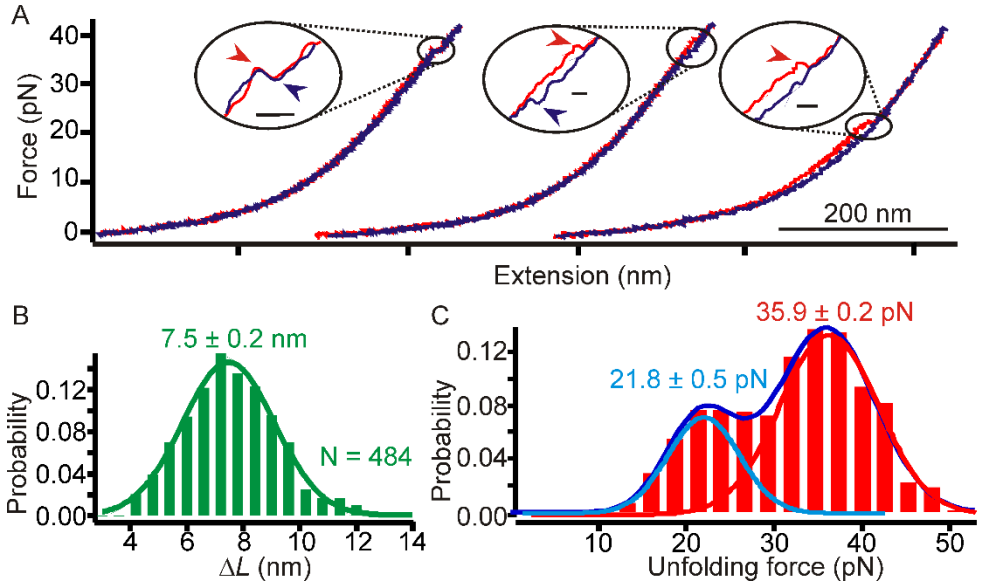


Figure 4.12. Mechanical unfolding of human telomeric G-quadruplexes inside the small (6 nm × 6 nm) DNA nanocage at room temperature in 20 mM Tris buffer (pH 7.8) that contains 10 mM $MgCl_2$, 100 mM KCl and 1 mM EDTA. (A) Typical stretching (red) and relaxing (purple) *F*-*X* curves of a G-quadruplex hosting sequence. Arrowheads in blowups depict unfolding or folding transitions. Scale bars represent 7.0 nm. (B) Histogram of the change-in-contour-length (ΔL) from the G-quadruplex hosting sequence inside the nanocage. (C) Histogram of the unfolding force of G-quadruplex inside the nanocage. Solid curves represent Gaussian fittings. Number of features is denoted by N.

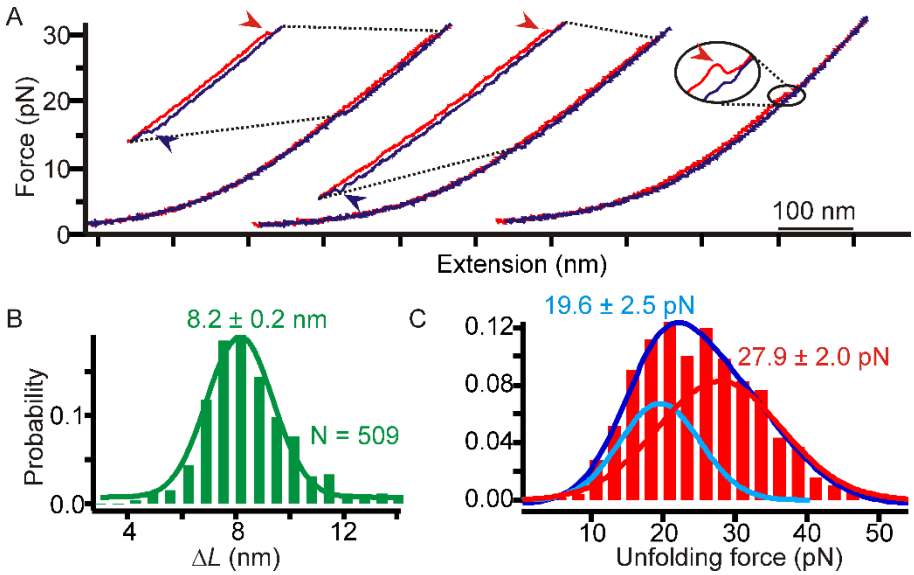


Figure 4.13. Mechanical unfolding of human telomeric G-quadruplexes inside the large (15 nm × 15 nm) DNA nanocage at room temperature in 20 mM Tris buffer (pH 7.8) that contains 10 mM $MgCl_2$, 100 mM KCl and 1 mM EDTA. (A) Typical stretching (red) and relaxing (purple) *F-X* curves of a G-quadruplex hosting sequence. Arrowheads in blowups depict unfolding or folding transitions. (B) Histogram of the change-in-contour-length (ΔL) from the G-quadruplex hosting sequence inside the nanocage. (C) Histogram of the unfolding force of G-quadruplex inside the nanocage. Solid curves represent Gaussian fittings. Number of features is denoted by N.

2.5.3.5 Deconvolution of unfolding force histograms

Since the histogram of the unfolding force of G-quadruplex inside nanocages revealed two major populations, we deconvoluted the two populations using trace-by-trace analysis. Based on the nature of refolding trajectory, all F-X curves could be divided into two types; one type of F-X curves has fast refolding (>5 pN) and the other

with slow refolding (<5 pN). These two pools of F-X curves were used to plot the rupture force histogram and the change in contour length (ΔL) histogram.

2.5.3.6 Mechanical unfolding of a G-quadruplex inside the medium nanocage in a hybrid conformation favoring buffer.

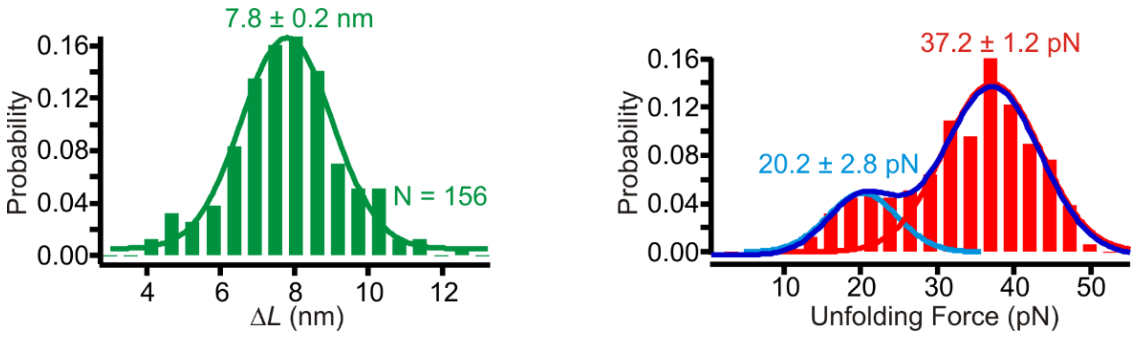


Figure 4.14. Histograms of the change in contour length (ΔL, left panel) and unfolding force (right panel) of the G-quadruplex inside the medium nanocage (9 nm × 9 nm) in the 20 mM Tris buffer (pH=7.9) consisting of 50 mM KCl and 1 M LiCl that favors the hybrid G-quadruplex conformation.

2.5.3.7 Mechanical unfolding of a G-quadruplex without nanocage in a basket conformation favoring buffer.

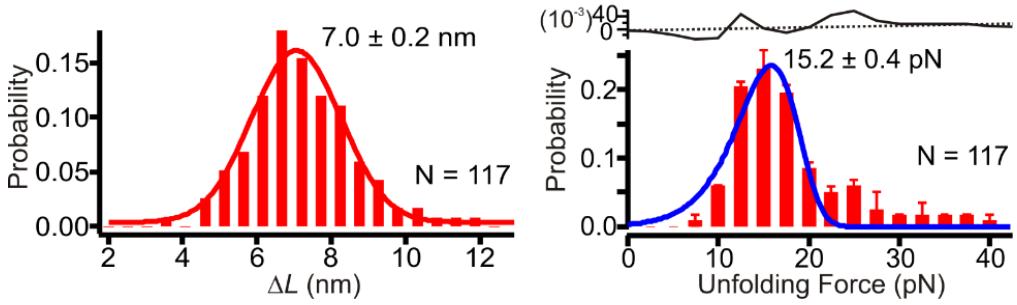


Figure 4.15. Histograms of the change in contour length (left panel) and unfolding force (right panel) during the unfolding of G-quadruplex without nanocage in a 20 mM Tris (pH=7.9) buffer with 1 M NaCl. Blue curve in the right panel depicts fitting by using Dudko model[145]. Black and dotted curves on top of the force histogram represent the residual analysis and the linear fitting, respectively. N depicts number of features.

2.5.3.8 Mechanical unfolding of a G-quadruplex without nanocage in a molecularly crowded buffer.

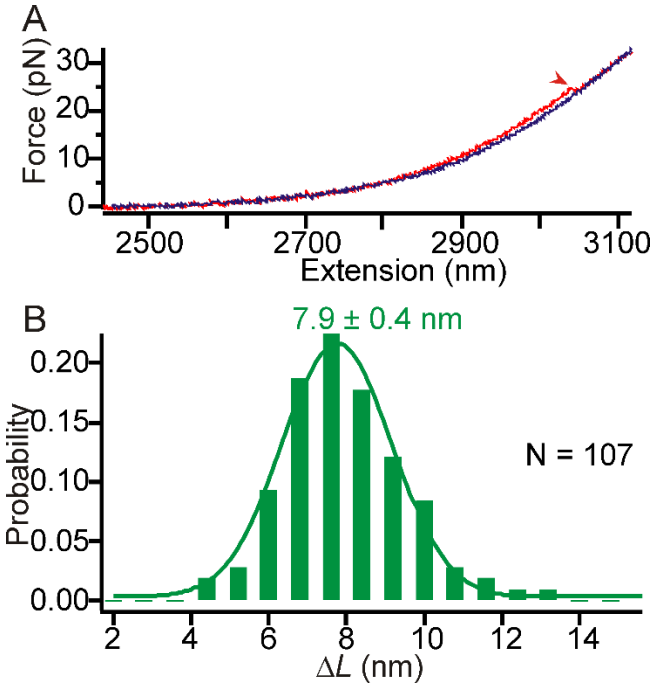


Figure 4.16. (A) A typical *F-X* curve obtained during mechanical unfolding of human telomere G-quadruplex in a 20 mM Tris buffer that contains 40% (w/v) BSA, 100 mM KCl, 1 mM EDTA, and 10 mM MgCl2 at pH 7.8. Red and black traces depict unfolding and refolding traces, respectively. Arrowhead indicates the unfolding transition. (B) Histogram of the change-in-contour-length (ΔL) due to the unfolding transition of the G-quadruplex in the crowded buffer. Green solid curve represents Gaussian fitting.

2.5.3.9 Histograms of unfolding work of G-quadruplexes during different conditions

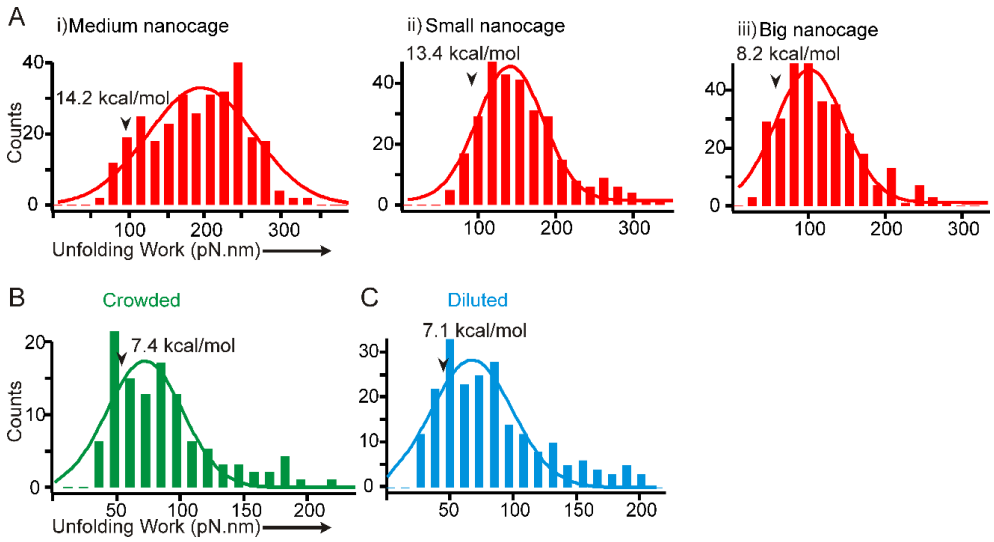


Figure 4.17. (A) Unfolding work histograms of the G-quadruplex inside the i) medium, ii) small, and iii) large nanocages in a 20 mM Tris buffer (pH=7.8, with 100 mM KCl, 10 mM $MgCl_2$ and 1 mM EDTA) with (B) or without (C) 40% (w/v) BSA. Arrowheads depict the work equivalent to the ΔG_{unfold}. Solid curves represent Gaussian fittings.

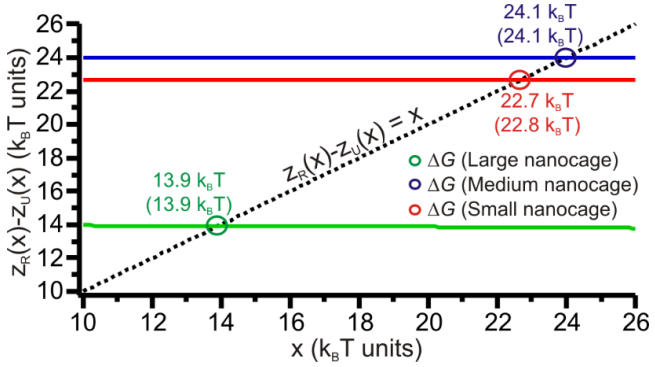


Figure 4.18. Bennett Acceptance Ratio method to estimate ΔG_{unfold} of G-quadruplex inside the DNA nanocage of different sizes. The small circles indicate the intersection point of the function y(x)=zR(x)-zU(x) with y(x)=x kBT at x=ΔG. Green, purple and red curves are the respective plots for the unfolding of GQ in large, medium and small nanocages. Values in parentheses depict those from Jarzynski equality calculation.

Table 4.6. Summary of mechanical properties of human telomeric G-quadruplex (GQ) in various conditions.

Molecular constructs	**Buffer conditions (pH 7.8)**	**Δ*L* (nm)**	**Unfolding Force (pN)**
GQ in Small nanocage	10 mM Mg^{2+} and 100 mM K^+	6.9 ± 0.1	35.9 ± 0.2
GQ w/o nanocage	1M Na^+	7.0 ± 0.2	15.2 ± 0.4
GQ in Medium nanocage	10 mM Mg^{2+} and 100 mM K^+	8.2 ± 0.2	38.6 ± 0.2
GQ in Medium nanocage	50 mM K^+ and 1M Li^+	7.8 ± 0.2	37.2 ± 1.2
GQ in Large nanocage	10 mM Mg^{2+} and 100 mM K^+	8.2 ± 0.2	27.9 ± 2.0
GQ w/o nanocage (diluted)	10 mM Mg^{2+} and 100 mM K^+	8.3 ± 0.1	20.7 ± 1.4
GQ w/o nanocage (crowded)	40% w/v BSA, 10 mM Mg^{2+} and 100 mM K^+	7.9 ± 0.4	19.6 ± 0.9

2.5.3.9 Hydration state of G-quadruplex inside the DNA nanocages

Based on the size of water molecules and the inner size of DNA nanocages, we have calculated that there are 15 layers of water molecules between the G-quadruplex forming DNA and the medium (10 nm × 10 nm) nanocage (for small cage, 6 nm × 6 nm, there are only 7 layers of water). As both G-quadruplex hosting DNA and nanocage DNA walls are multivalently charged (it has been reported that for single duplex DNA strand, hydration shell has 2 layers of waters.[183] The DNA origami walls are expected to have thicker hydration shells due to the intertwining DNA strands), the arrangement of these water molecules between the wall and the G-quadruplex hosting DNA strand should therefore be more ordered than that in the bulk. Folding of the G-quadruplex involves a loss of ~70 water molecules in the immediate vicinity of the G-quadruplex hosting sequence. In the nanocage, these lost water molecules should still stay in the cage (50 nm in length), which has more ordered water layers than the bulk, therefore, the effect of water may not have a predominant effect in our system.

CHAPTER V

CONCLUSIONS AND PERSPECTIVES

Given the prevalence of G-quadruplex motifs in the human genome and their biological significances, it has been essential to study the formation of G-quadruplex during cellular conditions such as during transcription and confined environment. In this regard, by taking the advantages of the methods of molecular biology, and DNA origami nanotechnology in optical tweezers, this dissertation research accomplished two primary objectives. One, to profile the population of different G-quadruplex structures in a sequence downstream of the transcription start site during the transcription and next, to investigate the transition dynamics of a G-quadruplex in a condition mimicking inside the exit channel of the polymerase enzyme. We chose a G-rich sequence in the non-template strand at the downstream of the transcription start site in the human gene as a model system to deconvolute the individual population of G-quadruplexes and investigate their properties during transcription. And, we incorporated a human telomere G-quadruplex sequence inside the DNA origami nanocage to investigate the effect of confined space on its folding behavior.

In the first part (discussed in Chapter III), we used gel-shift assay, DMS footprinting, CD spectroscopy, and UV-crosslinking assay to confirm the participation of the nascent G-rich RNA with G-rich DNA strand to form DNA/RNA hybrid G-quadruplex. Also, these methods revealed the existence of a complex equilibrium

dynamics among several G-quadruplex structures during the transcription, but these ensemble methods did not allow us to study the properties of individual structures. Then, we developed a Single-Molecule Stalled-transcription Assay (SMSA) using a T7 transcription model. The G-rich sequence was transcribed and stalled at its downstream so that the nascent G-rich RNA would be available to interact with the G-rich sequence in the non-template DNA strand. By using SMSA in optical tweezers, we successfully deconvoluted the different G-rich folded structures and found that DNA/RNA hybrid G-quadruplexes are formed with the different number of combinations between the DNA and RNA G-tracts and some minor populations of DNA G-quadruplex and G-triplex. These DNA/RNA hybrid G-quadruplexes have high mechanical and thermodynamic stability comparing to the intramolecular DNA G-quadruplex while the DNA G-quadruplex folds faster during transcription than in no transcription condition. These results implied the two main conclusions. First, the high mechanical stability of DNA/RNA hybrid G-quadruplex suggests that HQ may serve as a stronger mechanical block for transcription. Second, the catalytic role of a nascent G-rich RNA transcript to convert more duplex to quadruplex DNA suggests that both DNA G-quadruplex and DNA/RNA hybrid G-quadruplex play a synergistic role to control the transcription. We anticipate that the method may be generic to apply for the investigation of such complex population dynamics in other biological processes.

Second, the techniques of DNA origami nanotechnology were used to synthesize the molecular construct containing a human telomere G-quadruplex inside the DNA nanocage (discussed in Chapter IV). The DNA nanocage was used to mimic the naturally

existing nano-sized entry/exit channels of polymerase enzymes and ribosomes. By using optical tweezers, a G-quadruplex inside a DNA nanocage could be individually unfolded and folded at a time. This experiment was performed in different buffer conditions, allowing us to determine the properties of G-quadruplex under confinement accurately. This research revealed two critical conclusions. First, optical tweezers permitted us to isolate the system and apply force individually onto the G-quadruplex structure while the DNA nanocage intact. By using this method, we successfully investigated the real confinement effect on the folding process of a G-quadruplex which otherwise is impossible to other existing techniques. Therefore, this technique is anticipated to be instrumental in studying the impact of confined space on the folding/unfolding process of a biomolecule. Second, we unfolded the G-quadruplex inside the different size of nanocages and in various buffer conditions. We confirmed that the G-quadruplex inside the confined space has higher mechanical stability comparing to the free G-quadruplex. Also, both the mechanical and thermodynamic stabilities of G-quadruplex were found to be decreased with the increase in the size of confinement. In addition, the folding kinetics of G-quadruplex was found to be correlated with the size of confinement that the folding becomes faster in optimally small nanocage. The faster folding kinetics suggested the possibility of co-replicational or co-transcriptional or co-translocational folding of G-quadruplex inside cells and the stronger G-quadruplex inside the confined space implies that they may act as mechanical blocks to control the back-tracking of polymerase enzymes during transcription and replication. Together, the faster dynamics and higher

mechanical stability of G-quadruplex inside a confined space may help to explain the roles of a G-quadruplex in several biological processes.

REFERENCES

1. Neidle, S.; Balasubramanian, S., Quadruplex Nucleic Acids. *RSC Biomolecular Sciences* **2006**.
2. Brázda, V.; Laister, R. C.; Jagelská, E. B.; Arrowsmith, C., Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **2011,** *12* (1), 1-16.
3. Doherty, E. A.; Batey, R. T.; Masquida, B.; Doudna, J. A., A universal mode of helix packing in RNA. *Nat Struct Mol Biol* **2001,** *8* (4), 339-343.
4. Tamura, M.; Holbrook, S. R., Sequence and Structural Conservation in RNA Ribose Zippers. *J. Mol. Biol.* **2002,** *320* (3), 455-474.
5. Wells, R. D.; Dere, R.; Hebert, M. L.; Napierala, M.; Son, L. S., Advances in machanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **2005,** *33* (12), 3785-3798.
6. Wells, R. D., Non-B DNA conformations, mutagenesis and disease. *Trends in Biochem. Sci.* **2007,** *32*, 271-278.
7. Rhodes, D.; Lipps, H. J., G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015,** *43* (18), 8627-8637.
8. Cahoon, L. A.; Seifert, H. S., An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* **2009,** *325* (5941), 764-7.
9. Hansel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S., DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* **2017,** *18* (5), 279-284.

10. Patel, D. J.; Phan, A. T.; Kuryavyi, V., Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007,** *35*, 7429-7455.

11. Neidle, S.; Read, M. A., G-quadruplexes as therapeutic targets. *Biopolymers* **2000-2001,** *56* (3), 195-208.

12. Balasubramanian, S.; Hurley, L. H.; Neidle, S., Targeting G-quadruplexes in Gene Promoters: a Novel Anticancer Strategy? *Nat Rev Drug Discov* **2011,** *10* (4), 261-275.

13. Pandey, S.; Agarwala, P.; Maiti, S., Targeting RNA G-Quadruplexes for Potential Therapeutic Applications. In *RNA Therapeutics*, Garner, A. L., Ed. Springer International Publishing: Cham, 2018; pp 177-206.

14. Hurley, L. H.; Wheelhouse, R. T.; Sun, D.; Kerwin, S. M.; Salazar, M.; Fedoroff, O. Y.; Han, F. X.; Han, H.; Izbicka, E.; Von Hoff, D. D., G-quadruplexes as Targets for Drug Design. *Pharmacol Ther* 2000, pp 141-58.

15. Zhou, W.; Brand, N. J.; Ying, L., G-quadruplexes-novel mediators of gene function. *J Cardiovasc Transl Res* **2011,** *4* (3), 256-70.

16. Zhou, W.; Suntharalingam, K.; Brand, N. J.; Barton, P. J. R.; Vilar, R.; Ying, L., Possible Regulatory Roles of Promoter G-Quadruplexes in Cardiac Function-Related Genes – Human TnIc as a Model. *PLoS ONE* **2013,** *8* (1), e53137.

17. Guo, Y.; Yao, W.; Xie, Y.; Zhou, X.; Hu, J.; Pei, R., Logic gates based on G-quadruplexes: principles and sensor applications. *Microchimica Acta* **2016,** *183* (1), 21-34.

18. He, H.-Z.; Chan, D. S.-H.; Leung, C.-H.; Ma, D.-L., G-quadruplexes for luminescent sensing and logic gates. *Nucleic Acids Res.* **2013,** *41* (8), 4345-4359.

19. Engelhard, D. M.; Pievo, R.; Clever, G. H., Reversible Stabilization of Transition-Metal-Binding DNA G-Quadruplexes. *Angew. Chem. Int. Ed.* **2013,** *52* (49), 12843-12847.

20. Alberti, P.; Mergny, J.-L., DNA duplex–quadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100* (4), 1569-1573.

21. Alberti, P.; Bourdoncle, A.; Sacca, B.; Lacroix, L.; Mergny, J.-L., DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem.* **2006,** *4* (18), 3383-3391.

22. Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D., Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. *J Am Chem Soc* **1978,** *100* (11), 3625-3627.

23. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R., Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **1989,** *59*, 871-880.

24. Chowdhurdy, S.; Bansal, M., G-Quadruplex Structure Can Be Stable with Only Some Coordination Sites Being Occupied by Cations: A Six Nanosecond Molecular Dynamics Simulation. *J. Phys. Chem. B* **2001,** *31*, 7572-7578.

25. Campbell, N. H.; Neidle, S., G-quadruplexes and metal ions. *Met Ions Life Sci* **2012,** *10*, 119-34.

26. Wang, Y.; Patel, D. J., Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1993,** *1* (4), 263-282.

27. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006,** *128*, 9963-9970.

28. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res* **2007,** *35* (19), 6517-25.

29. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **2006,** *34*, 5402-5415.

30. Williamson, J. R., G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* 1994, pp 703-30.

31. Sannohe, Y.; Sugiyama, H., Overview of Formation of G-quadruplex Structure. *Curr. Protoc. Nucleic Acid Chem.* **2010**, 17.2.1-17.2.17.

32. Phan, A. T., Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS Journal* **2010,** *277*, 1107-1117.

33. Bochman, M. L.; Paeschke, K.; Zakian, V. A., DNA secondary structures: stability and function of G-quadruplex structures. *Nature Rev Genet* **2012,** *13* (11), 770-780.

34. Risitano, A.; Fox, K. R., Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes. *Biochemistry* **2003,** *42*, 6507-6513.

35. Phan, A. T.; Kuryavyi, V.; Patel, D. J., DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.* **2006,** *16*, 288-298.

36. Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D., An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *J. Am. Chem. Soc.* **2006,** *128*, 1096-1098.

37. Tippana, R.; Xiao, W.; Myong, S., G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **2014,** *42* (12), 8106-8114.

38. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S., Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **2004,** *126* (50), 16405-16415.

39. Risitano, A.; Fox, K. R., Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes. *Nucleic Acids Res.* **2004,** *32*, 2598-2606.

40. Lu, H.; Li, S.; Chen, J.; Xia, J.; Zhang, J.; Huang, Y.; Liu, X.; Wu, H.-c.; Zhao, Y.; Chai, Z.; Hu, Y., Metal ions modulate the conformation and stability of a G-quadruplex with or without a small-molecule ligand. *Metallomics* **2015,** *7* (11), 1508-1514.

41. Largy, E.; Mergny, J. L.; Gabelica, V., Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met Ions Life Sci* **2016,** *16*, 203-58.

42. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O., 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.* **2010,** *132* (48), 17105-17107.

43. Heddi, B.; Phan, A. T., Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* **2011,** *133* (25), 9824-9833.

44. HÃnsel, R.; Löhr, F.; Trantírková, S. F.; Bamberg, E.; Trantírek, L.; Dötsch, V., The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011,** *3*, 1-8.

45. Cheong, C.; Moore, P. B., Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **1992,** *31* (36), 8406-14.

46. Bugaut, A.; Balasubramanian, S., 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **2012,** *40* (11), 4727-4741.

47. Wanrooij, P. H.; Uhler, J. P.; Shi, Y.; Westerlund, F.; Falkenberg, M.; Gustafsson, C. M., A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* **2012,** *40* (20), 10334-10344.

48. Zheng, K.; Xiao, S.; Liu, J.; Zhang, J.; Hao, Y.; Tan, Z., Co-transcriptional formation of DNA: RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* **2013,** *41* (10), 5533-5541.

49. Shrestha, P.; Xiao, S.; Dhakal, S.; Tan, Z.; Mao, H., Nascent RNA transcripts facilitate the formation of G-quadruplexes. *Nucleic Acids Res.* **2014,** *42* (11), 7236-46.

50. Sen, D.; Gilbert, W., Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988,** *334* (6180), 364-66.

51. Todd, A.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *33* (9), 2901–2907.

52. Rawal, P.; Kummarasetti, V.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S.; Chowdhury, S., Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **2006,** *16*, 644-655.

53. Johnson, J. E.; Smith, J. S.; Kozak, M. L.; Johnson, F. B., In vivo veritas: Using yeast to probe the biological functions of G-quadruplexes. *Biochimie* **2008,** *90* (8), 1250-1263.

54. Huppert, J., Hunting G-quadruplexes. *Biochimie* **2008,** *90* (8), 1140-1148.

55. Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L. S.; Johnson, F. B., Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.* **2008,** *36* (1), 144-156.

56. Todd, A. K.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *33* (9), 2901-2907.

57. Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33* (9), 2908-2916.

58. Maizels, N., Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **2006,** *13* (12), 1055-59.

59. Huppert, J. L.; Balasubramanian, S., G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007,** *35*, 406-413.

60. Brooks, T. A.; Kendrick, S.; Hurley, L. H., Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **2010,** *277*, 3459-3469.

61. Xiao, S.; Zhang, J.; Zheng, K.; Hao, Y.; Tan, Z., Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals. *Nucleic Acids Res.* **2013,** *41* (22), 10379-10390.

62. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **2013,** *5*, 182-186.

63. Wang, Q.; Liu, J.-q.; Chen, Z.; Zheng, K.-w.; Chen, C.-y.; Hao, Y.-h.; Tan, Z., G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* **2011,** *39*, 6229-6237.

64. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Peter, L. C. H.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W., Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science* **1994,** *266* (5193), 2011-2015.

65. Cuesta, J.; Read, M. A.; Neidle, S., The Design of G-quadruplex Ligands as Telomerase Inhibitors. *Mini. Rev. Med. Chem.* **2003,** *3*, 11-21.

66. Chen, Z. F.; Qin, Q. P.; Qin, J. L.; Liu, Y. C.; Huang, K. B.; Li, Y. L.; Meng, T.; Zhang, G. H.; Peng, Y.; Luo, X. J.; Liang, H., Stabilization of G-quadruplex DNA, inhibition of telomerase activity, and tumor cell apoptosis by organoplatinum(II) complexes with oxoisoaporphine. *J Med Chem* **2015,** *58* (5), 2159-79.

67. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H., 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. U. S. A.* **2002,** *99*, 11593-11598.

68. Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D., NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* **2006,** *34* (18), 5133-44.

69. Grand, C. L.; Powell, T. J.; Nagle, R. B.; Bearss, D. J.; Tye, D.; Gleason-Guzman, M.; Hurley, L. H., Mutations in the G-quadruplex silencer element and their relationship to c-MYC overexpression, NM23 repression, and therapeutic rescue. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (16), 6140-6145.

70. Paramasivan, S.; Rujan, I.; Bolton, P. H., Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods* **2007,** *43*, 324-331.

71. Karsisiotis, A. I.; Hessari, N. M. a.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M., Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. *Angewandte Chemie* **2011,** *123* (45), 10833-10836.

72. Mergny, J.-L.; Maurizot, J.-C., Fluorescence Resonance Energy Transfer as a Probe for G-Quartet Formation by a Telomeric Repeat. *ChemBioChem* **2001,** *2* (2), 124-132.

73. Adrian, M.; Heddi, B.; Phan, A. T., NMR spectroscopy of G-quadruplexes. *Methods* **2012,** *57* (1), 11-24.

74. Parkinson, G. N.; Lee, M. P.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-80.

75. Sun, D.; Hurley, L. H., Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.* **2010,** *608*, 65-79.

76. Gabelica, V. r.; Shammel Baker, E.; Teulade-Fichou, M.-P.; De Pauw, E.; Bowers, M. T., Stabilization and Structure of Telomeric and c-myc Region Intramolecular G-Quadruplexes:The Role of Central Cations and Small Planar Ligands. *J. Am. Chem. Soc.* **2007,** *129* (4), 895-904.

77. Talib, J.; Green, C.; Davis, K. J.; Urathamakul, T.; Beck, J. L.; Aldrich-Wright, J. R.; Ralph, S. F., A comparison of the binding of metal complexes to duplex and quadruplex DNA. *Dalton Trans.* **2008,** (8), 1018-1026.

78. Huang, C.-C.; Cao, Z.; Chang, H.-T.; Tan, W., Proteinâ''Protein Interaction Studies Based on Molecular Aptamers by Affinity Capillary Electrophoresis. *Anal. Chem.* **2004,** *76* (23), 6973-6981.

79. Xu, Y.; Feng, X.; Du, W.; Liu, X.; Luo, Q.; Liu, B.-F., Kinetic and Thermodynamic Characterization of Telomeric G-Quadruplex by Nonequilibrium Capillary Electrophoresis: Application to G-Quadruplex/Duplex Competition. *Z. Anal. Chem.* **2008,** *80* (18), 6935-6941.

80. Koirala, D.; Ghimire, C.; Bohrer, C.; Sannohe, Y.; Sugiyama, H.; Mao, H., Long-Loop G-Quadruplexes Are Misfolded Population Minorities with Fast Transition Kinetics in Human Telomeric Sequences. *J. Am. Chem. Soc.* **2013,** *135*, 2235-2241.

81. Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S., Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 14629-14634.

82. Green, J. J.; Ladame, S.; Ying, L.; Klenerman, D.; Balasubramanian, S., Investigating a Quadruplex−Ligand Interaction by Unfolding Kinetics. *J. Am. Chem. Soc.* **2006,** *128* (30), 9809-9812.

83. Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S., Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit. *J. Am. Chem. Soc* **2007,** *129*, 7484-7485.

84. Okamoto, K.; Sannohe, Y.; Mashimo, T.; Sugiyama, H.; Terazima, M., G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution. *Bioorg. Med. Chem.* **2008,** *16* (14), 6873-6879.

85. Jena, P. V.; Shirude, P. S.; Okumus, B.; Laxmi-Reddy, K.; Godde, F. d. r.; Huc, I.; Balasubramanian, S.; Ha, T., G-Quadruplex DNA Bound by a Synthetic Ligand is Highly Dynamic. *J. Am. Chem. Soc.* **2009,** *131* (35), 12522-12523.

86. Murat, P.; Singh, Y.; Defrancq, E., Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev.* **2011,** *40* (11), 5293-5307.

87. Ray, S.; Bandaria, J. N.; Qureshi, M. H.; Yildiz, A.; Balci, H., G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc Natl Acad Sci USA* **2014,** *111* (8), 2990-2995.

88. Sannohe, Y.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H., Visualization of Dynamic Conformational Switching of the G-Quadruplex in a DNA Nanostructure. *J. Am. Chem. Soc.* **2010,** *132* (46), 16311-16313.

89. Rajendran, A.; Endo, M.; Sugiyama, H., Single-Molecule Analysis Using DNA Origami. *Angew Chem Int Ed Engl* **2012,** *51*, 874 – 890.

90. Binnig, G.; Quate, C. F.; Gerber, C., Atomic Force Microscope. *Phys. Rev. Lett.* **1986,** *56*, 930-933.

91. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E., Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997,** *276* (5315), 1109-12.

92. Smith, S. B.; Cui, Y. J.; Bustamante, C., Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **1996,** *271* (5250), 795-799.

93. Visscher, K.; Schnitzer, M. J.; Block, S. M., Single kinesin molecules studied with a molecular force clamp. *Nature* **1999,** *400* (6740), 184-189.

94. Ashkin, A.; Dziedzic, J. M.; Bjorkholm, J. E.; Chu, S., Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles. *Optics Lett.* **1986,** *11*, 288.

95. Smith, S. B.; Finzi, L.; Bustamante, C., Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **1992,** *258,* 1122-1126.

96. Strick, T. R.; Allemand, J.-F.; Bensimon, D.; Bensimon, A.; Croquette, V., The Elasticity of a Single Supercoiled DNA Molecule. *Science* **1996,** *271,* 1835-1837.

97. Selvam, S.; Koirala, D.; Yu, Z.; Mao, H., Quantification of Topological Coupling between DNA Superhelicity and G-quadruplex Formation. *J. Am. Chem. Soc.* **2014,** *136,* 13967-13970.

98. Koirala, D.; Yangyuoru, P. M.; Mao, H., Mechanical affinity as a new metrics to evaluate binding events. *Rev Anal Chem* **2013,** *32*, 197-208.

99. Finer, J. T.; Simmons, R. M.; Spudich, J. A., Single myosin molecule mechanics: piconewton forces and nanometer steps. *Nature* **1994,** *368* (6467), 113-119.

100. Shrestha, P.; Jonchhe, S.; Emura, T.; Hidaka, K.; Endo, M.; Sugiyama, H.; Mao, H., Confined space facilitates G-quadruplex formation. *Nat. Nanotechnol.* **2017,** *12* (6), 582-588.

101. Koirala, D.; Yu, Z.; Dhakal, S.; Mao, H., Detection of Single Nucleotide Polymorphism Using Tension-Dependent Stochastic Behavior of a Single-Molecule Template. *J. Am. Chem. Soc.* **2011,** *133* (26), 9988-9991.

102. Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H., ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities. *J. Am. Chem. Soc.* **2009,** *131* (5), 1876-1882.

103. Schonhoft, J. D.; Bajracharya, R.; Dhakal, S.; Yu, Z.; Mao, H.; Basu, S., Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. *Nucleic Acids Res.* **2009,** *37*, 3310-3320.

104. Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H., 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.* **2012,** *134* (11), 5157-5164.

105. Yu, Z.; Koirala, D.; Cui, Y.; Easterling, L. F.; Zhao, Y.; Mao, H., Click Chemistry Assisted Single-Molecule Fingerprinting Reveals a 3D Biomolecular Folding Funnel. *J. Am. Chem. Soc.* **2012,** *134* (30), 12338-41.

106. Dhakal, S.; Yu, Z.; Konik, R.; Cui, Y.; Koirala, D.; Mao, H., G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophys. J.* **2012,** *102* (11), 2575-2584.

107. Dhakal, S.; Cui, Y.; Koirala, D.; Ghimire, C.; Kushwaha, S.; Yu, Z.; Yangyuoru, P. M.; Mao, H., Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions. *Nucleic Acids Res.* **2013,** *41*, 3915-3923.

108. Kallenbach, N. R.; Ma, R.-I.; Seeman, N. C., An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **1983,** *305*, 829.

109. Rothemund, P. W. K., Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006,** *440*, 297-302.

110. Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H., DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* **2003,** *301* (5641), 1882-1884.

111. Stephanopoulos, N.; Liu, M.; Tong, G. J.; Li, Z.; Liu, Y.; Yan, H.; Francis, M. B., Immobilization and One-Dimensional Arrangement of Virus Capsids with Nanoscale Precision Using DNA Origami. *Nano Letters* **2010,** *10* (7), 2714-2720.

112. Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J., A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **1996,** *382*, 607.

113. Aldaye, F. A.; Sleiman, H. F., Sequential Self-Assembly of a DNA Hexagon as a Template for the Organization of Gold Nanoparticles. *Angew. Chem. Int. Ed.* **2006,** *45* (14), 2204-2209.

114. Zhao, Z.; Jacovetty, E. L.; Liu, Y.; Yan, H., Encapsulation of Gold Nanoparticles in a DNA Origami Cage. *Angew. Chem. Int. Ed.* **2011,** *50* (9), 2041-2044.

115. Xu, Y.; Suzuki, Y.; Komiyama, M., Click Chemistry for the Identification of G-Quadruplex Structures: Discovery of a DNA-RNA G-Quadruplex. *Angew. Chem. Int. Ed.* **2009,** *48*, 3281-3284.

116. Eggers, D. K.; Valentine, J. S., Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Science : A Publication of the Protein Society* **2001,** *10* (2), 250-261.

117. Senske, M.; Smith, A. E.; Pielak, G. J., Protein Stability in Reverse Micelles. *Angew. Chem. Int. Ed.* **2016,** *55* (11), 3586-3589.

118. Mitchell, M.; Gillis, A.; Futahashi, M.; Fujiwara, H.; Skordalakes, E., Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **2010,** *17* (4), 513-518.

119. Jansson, L. I.; Akiyama, B. M.; Ooms, A.; Lu, C.; Rubin, S. M.; Stone, M. D., Structural basis of template-boundary definition in Tetrahymena telomerase. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 883-8.

120. Zhang, J.; Zheng, K.; Xiao, S.; Hao, Y.; Tan, Z., Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.* **2014,** *136* (4), 1381–1390.

121. Zheng, K.; Chen, Z.; Hao, Y.; Tan, Z., Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res.* **2010,** *38* (1), 327-338.

122. Koirala, D.; Yu, Z.; Dhakal, S.; Mao, H., Detection of Single Nucleotide Polymorphism Using Tension-Dependent Stochastic Behavior of a Single-Molecule Template. *J. Am. Chem. Soc.* **2011,** *133*, 9988-9991.

123. Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **2009,** *459* (7245), 414-418.

124. Pelliccioli, A. P.; Wirz, J., Photoremovable protecting groups: reaction mechanisms and applications. *Photochem. & Photobio. Sci.* **2002,** *1* (7), 441-458.

125. Mao, H.; Luchette, P., An Integrated Laser Tweezers Instrument for Microanalysis of Individual Protein Aggregates. *Sensors and Actuators B* **2008,** *129*, 764-771.

126. Mao, H.; Arias-Gonzalez, J. R.; Smith, B.; Tinoco, I. J.; Bustamante, C., Temperature Control Methods in a Laser Tweezers System. *Biophys. J.* **2005,** *89*, 1308-1316.

127. Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H., A Single-Molecule Platform for Investigation of Interactions between G-quadruplexes and Small-Molecule Ligands. *Nat. Chem.* **2011,** *3*, 782-787.

128. Luchette, P.; Abiy, N.; Mao, H., Microanalysis of clouding process at the single droplet level. *Sens. Actuators B-Chem.* **2007,** *128* (1), 154-160.

129. Mao, H.; Luchette, P., An integrated laser-tweezers instrument for microanalysis of individual protein aggregates. *Sens Actuators B* **2008,** *129*, 764-771.

130. Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C., Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. USA.* **1997,** *94*, 6185-6190.

131. Dietz, H.; Rief, M., Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Nat. Acad. Sci. USA* **2004,** *101* (46), 16192-16197.

132. Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M., Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **2008,** *319*, 630-633.

133. Sinden, R. R., *DNA structure and function*. Academic Press: San Diego, CA, 1995.

134. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D., Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.* **2007,** *35* (7), 2440-2450.

135. Woodside, M. T.; Behnke-Parks, W. M.; Larizadeh, K.; Travers, K.; Herschlag, D.; Block, S. M., Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. USA.* **2006,** *103*, 6190-6195.

136. Mills, J. B.; Vacano, E.; Hagerman, P. J., Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). *J. Mol. Biol.* **1999,** *285*, 245-257.

137. Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H., 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.* **2010,** *132* (26), 8991–8997.

138. Jarzynski, C., Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997,** *78*, 2690 - 2693.

139. Gore, J.; Ritort, F.; Bustamante, C., Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl Acad. Sci. U S A* **2003,** *100* (22), 12564-9.

140. Palassini, M.; Ritort, F., Improving free-energy estimates from unidirectional work measurements: theory and experiment. *Phys Rev Lett* **2011,** *107* (6), 060601-1-5.

141. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.

142. Crooks, G. E., Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **2000,** *61* (3), 2361-2366.

143. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I. J.; Bustamante, C., Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005,** *437*, 231-234.

144. Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S., Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003,** *91* (14), 140601.

145. Dudko, O. K.; Hummer, G.; Szabo, A., Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (41), 15755–15760.

146. Blasius, B.; Huppert, A.; Stone, L., Complex Dynamics and Phase Synchronization in Spatially Extended Ecological Systems. *Nature* **1999,** *399*, 354-359.

147. Cui, Y.; Koirala, D.; Kang, H.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H., Molecular Population Dynamics of DNA Structures in a Bcl-2 Promoter Sequence is Regulated by Small-molecules and the Transcription Factor hnRNP LL. *Nucleic Acids Res.* **2014,** *42*, 5755-5764.

148. Paeschke, K.; Capra, J. A.; Zakian, V. A., DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **2011,** *145* (5), 678-691.

149. Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T., Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 18938-18943.

150. Lim, K. W.; Ng, V. C. M.; Martín-Pintado, N.; Heddi, B.; Phan, A. T., Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.* **2013,** *41* (22), 10556-62.

151. Onoa, B.; Tinoco Jr, I., RNA folding and unfolding. *Curr. Opin. Struct. Biol.* **2004,** *14* (3), 374-379.

152. Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M., Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 14579-14584.

153. Casey, J.; Davidson, N., Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res.* **1977,** *4* (5), 1539-1552.

154. Lim, K. W.; Lacroix, L.; Yue, D. J.; Lim, J. K.; Lim, J. M.; Phan, A. T., Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J Am Chem Soc* **2010,** *132* (35), 12331-42.

155. Yangyuoru, P. M.; Zhang, A. Y. Q.; Shi, Z.; Koirala, D.; Balasubramanian, S.; Mao, H., Mechanochemical Properties of Individual Human Telomeric RNA G-quadruplexes. *ChemBioChem* **2013,** *14*, 1931-1935.

156. Gao, Y.; Zorman, S.; Gundersen, G.; Xi, Z.; Ma, L.; Sirinakis, G.; Rothman, J. E.; Zhang, Y., Single Reconstituted Neuronal SNARE Complexes Zipper in Three Distinct Stages. *Science* **2012,** *337*, 1340-1343.

157. Li, X.; Manley, J. L., Cotranscriptional processes and their influence on genome stability. *Genes Dev.* **2006,** *20*, 1838-1847.

158. Kuzmine, I.; Gottlieb, P. A.; Martin, C. T., Binding of the Priming Nucleotide in the Initiation of Transcription by T7 RNA Polymerase. *J. Bio. Chem.* **2003,** *278*, 2819-2823.

159. Hardin, C. C., Perry, A.G. and White, K., Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **2000,** *56*, 147-194.

160. Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H., Intramolecular folding in three tandem guanine repeats of human telomeric DNA. *Chem. Commun.* **2012,** *48* (14), 2006-2008.

161. Galburt, E. A.; Grill, S. W.; Wiedmann, A.; Lubkowska, L.; Choy, J.; Nogales, E.; Kashlev, M.; Bustamante, C., Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **2007,** *446*, 820-823.

162. Mejia, Y. X.; Mao, H.; Forde, N. R.; Bustamante, C., Thermal probing of E. coli RNA polymerase off-pathway mechanisms. *J. Mol. Biol.* **2008,** *382* (3), 628-37.

163. Yin, H.; Wang, M. D.; Svoboda, K.; Landick, R.; Block, S. M.; Gelles, J., Transcription against an Applied Force. *Science* **1995,** *270* (5242), 1653-1657.

164. Anand, V. S.; Patel, S. S., Transient State Kinetics of Transcription Elongation by T7 RNA Polymerase. *J. Biol. Chem.* **2006,** *281* (47), 35677-35685.

165. Tang, Y.-C.; Chang, H.-C.; Roeben, A.; Wischnewski, D.; Wischnewski, N.; Kerner, M. J.; Hartl, F. U.; Hayer-Hartl, M., Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein. *Cell* **2006,** *125* (5), 903-914.

166. Brinker, A.; Pfeifer, G.; Kerner, M. J.; Naylor, D. J.; Hartl, F. U.; Hayer-Hartl, M., Dual Function of Protein Confinement in Chaperonin-Assisted Protein Folding. *Cell* **2001,** *107* (2), 223-233.

167. Baumketner, A.; Jewett, A.; Shea, J. E., Effects of Confinement in Chaperonin Assisted Protein Folding: Rate Enhancement by Decreasing the Roughness of the Folding Energy Landscape. *J. Mol. Biol.* **2003,** *332* (3), 701-713.

168. Jewett, A. I.; Baumketner, A.; Shea, J.-E., Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: Creation of an alternate fast folding pathway. *Proc. Natl Acad. Sci. U.S.A* **2004,** *101* (36), 13192-13197.

169. Takagi, F.; Koga, N.; Takada, S., How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proc. Natl Acad. Sci. U.S.A* **2003,** *100* (20), 11367-11372.

170. Zhou, H.-X., Protein folding and binding in confined spaces and in crowded solutions. *J. Mol. Recog.* **2004,** *17* (5), 368-375.

171. Zhou, H.-X.; Dill, K. A., Stabilization of Proteins in Confined Spaces. *Biochemistry* **2001,** *40* (38), 11289-11293.

172. Zhou, H. X.; Rivas, G.; Minton, A. P., Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008,** *37*, 375-397.

173. Li, W.; Hou, X.-M.; Wang, P.-Y.; Xi, X.-G.; Ming Li, M., Direct Measurement of Sequential Folding Pathway and Energy Landscape of Human Telomeric G-quadruplex Structures. *J. Am. Chem. Soc.* **2013,** *135*, 6423–6426.

174. Han, H.; Hurley, L. H., G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol Sci* **2000,** *21* (4), 136-142.

175. Balasubramanian, S.; Neidle, S., G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **2009,** *13* (3), 345-353.

176. Moffitt, J. R.; Chemla, Y. R.; Smith, S. B.; Bustamante, C., Recent advances in optical tweezers. *Annu. Rev. Biochem.* **2008,** *77*, 205-228.

177. Venczel, E. A.; Sen, D., Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence. *Biochemistry* **1993,** *32*, 6220-6228.

178. An, N.; Fleming, A. M.; Middleton, E. G.; Burrows, C. J., Single-molecule investigation of G-quadruplex folds of the human telomere sequence in a protein nanocavity. *Proc. Natl Acad. Sci. USA* **2014,** *111* (40), 14325-14331.

179. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C., Formation and stabilization of G-quadruplex in nanosized water pools. *Chem. Commun.* **2010,** *46* (10), 1700-1702.

180. Xie, M.; Podlevsky, J. D.; Qi, X.; Bley, C. J.; Chen, J. J.-L., A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic Acids Res.* **2010,** *38* (6), 1982-1996.

181. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J., Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat. Struct. Mol. Biol.* **2005,** *12*, 847-854.

182. Miyoshi, D.; Karimata, H.; Sugimoto, N., Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *J. Am. Chem. Soc.* **2006,** *128* (24), 7957-7963.

183. Schneider, B.; Patel, K.; Berman, H. M., Hydration of the phosphate group in double-helical DNA. *Biophys. J.* **1998,** *75* (5), 2422-2434.