MECHANOCHEMISTRY, TRANSITION DYNAMICS AND LIGAND-INDUCED STABILIZATION OF HUMAN TELOMERIC G QUADRUPLEXES AT SINGLE-MOLECULE LEVEL

A dissertation submitted to Kent State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By
Deepak P. Koirala
May, 2014
Dissertation written by

Deepak P. Koirala

B.S., Tribhuvan University, Nepal, 2002

M.S., Tribhuvan University, Nepal, 2005

Ph.D., Kent State University, USA, 2014

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

Hamza Balci, Ph.D, Member, Doctoral Dissertation Committee

Accepted by

Michael Tubergen, Ph.D, Chair, Dept. of Chemistry & Biochemistry

Janis H. Crowther, Ph.D., Associate Dean, College of Arts and Sciences
# Table of Contents

List of Figures ........................................................................................................ viii

List of Tables ........................................................................................................ x

Acknowledgements .............................................................................................. xii

Summary .............................................................................................................. xvi

Chapter I: Introduction and Background ............................................................. 1

1.1 G-quadruplex DNA ....................................................................................... 2

1.2 DNA G-quadruplex at Human Telomere ..................................................... 5

1.3 Biological Relevance of Human Telomeric G-quadruplex ......................... 8

1.4 Methods of G-quadruplex Study ................................................................. 9

1.4.1 Conventional Methods ........................................................................ 9

1.4.2 Single-Molecule Methods .................................................................. 12
1.4.3 Mechanical Force-Based Methods ................................................................. 13

1.4.4 Manipulation of G-quadruplex Structures by Optical-Tweezers .................... 22

1.5 Significance of Present Study ........................................................................ 24

Chapter II: Materials and Methods ................................................................. 29

2.1 Materials ......................................................................................................... 29

2.2 CD Measurements ........................................................................................... 31

2.3 Molecular Dynamics Simulation ...................................................................... 31

2.4 DMS Footprinting ........................................................................................... 33

2.5 Single-Molecule Measurements ...................................................................... 34

2.5.1 Optical-Tweezers Instrumentation ............................................................... 34

2.5.2 Preparation of Microfluidic Chamber ........................................................... 36

2.5.3 Molecular Biology Synthesis of DNA Constructs ........................................ 39

2.5.4 Single-Molecule Mechanochemical Assays ................................................ 40

2.5.5 Measurements of Unfolding Force and Change in Contour Length ............. 42
Chapter III: Intramolecular Folding in Three and Four Tandem G-Rich Repeats of Human Telomeric DNA: Direct Observation of G-Triplex Species ........................................ 57

3.1 Abstract ........................................................................................................ 57

3.2 Overview ......................................................................................................... 58

3.3 Results and Discussion .................................................................................... 60

3.4 Conclusions ..................................................................................................... 77

4.1 Abstract ........................................................................................................... 78

4.2 Overview ......................................................................................................... 79

4.3 Results and Discussion ..................................................................................... 85

4.3.1 Force-Jump Method and Population Deconvolution at Nanometer Resolution .... 85

4.3.2 Transition Kinetics of Populations in hTel-4 Sequence ............................... 89

4.3.3 Transition Kinetics of Populations in hTel-5 Sequence ............................... 95

4.3.4 Equilibrium Population Dynamics in hTel-4 and hTel-5 Sequences ............ 99

4.3.5 Probability of Misfolded G-quadruplexes in Longer Human Telomeric Sequences ........................................................................................................... 101

4.4 Conclusions ................................................................................................... 104
Chapter V: Interactions of Human Telomeric G-Quadruplex and Small-Molecule Ligands: Mechanochemical Dynamics Revealed at Single-Molecule Level ................................................................. 106

5.1 Abstract ............................................................................................................. 106

5.2 Overview .......................................................................................................... 107

5.3 Results and Discussion .................................................................................... 112

5.3.1 Ligand Binding Increases Mechanical Stability of Human Telomeric G-quadruplex ........................................................................................................ 112

5.3.2 Kinetic Measurements Indicate That Binding of Ligand Facilitates Formation of G-quadruplex ......................................................................................... 117

5.3.3 Determination of Dissociation Constant without Varying Concentration of Either Ligand or Receptor ........................................................................ 120

5.4 Conclusions ..................................................................................................... 125

Chapter VI: Conclusion and Perspective ............................................................. 127

References ............................................................................................................. 130
List of Figures

Figure 1.1: Structure and topology of various G-quadruplex conformations ............................................ 3

Figure 1.2: Physiological relevance of human telomeric G-quadruplexes ................................................. 7

Figure 1.3: Schematic of commonly used single-molecule methods for mechanical unfolding and refolding of biomolecules ........................................................................................................... 15

Figure 1.4: Typical force-extension curve and unfolding energy landscape for mechanical unfolding of biomolecules ........................................................................................................................ 19

Figure 2.1: Optical layout of optical tweezers instrument used for single-molecule mechanochemical measurements of folded DNA structures ............................................................... 35

Figure 2.2: Schematic of a microfluidic chamber used in optical tweezers measurements ...... 37

Figure 2.3: Flowchart of molecular biology approach to synthesize DNA constructs used in single-molecule mechanochemical studies ................................................................. 38

Figure 2.4: Typical force-extension curves and plot of change-in-contour-length observed for unfolding and refolding of G-quadruplex structures .............................................. 41

Figure 3.1: CD analysis of hTelo-3 sequence in 10 mM Tris buffer, pH 7.4 containing 100 mM NaCl ............................................................................................................................................. 60
Figure 3.2: Mechanical unfolding of hTelo-3 in 100 mM Na\(^+\) buffer ................................. 62

Figure 3.3: Results of molecular dynamics simulation of G-triplex structure ....................... 64

Figure 3.4: Mechanical unfolding of hTelo-4 in 100 mM Na\(^+\) buffer ................................. 66

Figure 3.5: Dimethyl sulphate footprinting of hTelo-3, hTelo-3 + hTelo-1 mixture, and hTelo-4 sequences in Na\(^+\) buffer ................................................................. 70

Figure 3.6: CD analysis of 5µM solution of hTelo-3 sequence in 10 mM Tris buffer, pH 7.4 containing 100 mM KCl .......................................................... 71

Figure 3.7: Mechanical unfolding of hTelo-3 and hTelo-4 in a K\(^+\) buffer ......................... 73

Figure 3.8: Dimethyl sulphate footprinting of hTelo-3, hTelo-3 + hTelo-1 mixture, and hTelo-4 in 100 mM K\(^+\) containing buffer ............................................................... 76

Figure 4.1: Schematic structures of regular G-quadruplex, G-triplex and misfolded G-
quadruplex expected to form in hTelo-5 sequence ......................................................... 81

Figure 4.2: Illustration of force-jump method and associated data analysis ....................... 83

Figure 4.3: PoDNano analysis of population distribution in hTelo-4 sequence ................. 86

Figure 4.4: CD measurements of various human telomeric sequences in 10 mM Tris buffer that contains 100 mM KCl or LiCl .............................................................. 88

Figure 4.5: Transition kinetics of hTelo-4 sequence ......................................................... 91
Figure 4.6: Refolding probability of G-quadruplex and G-Triplex formed in hTelo-4 sequence with incubation time ................................................................. 93

Figure 4.7: PoDNano analyses and transition kinetics of hTelo-5 sequence ......................... 97

Figure 4.8: Distribution probability of folded population in human telomeric sequences from hTelo-4 to hTelo-7 ............................................................................................................ 104

Figure 4.9: Typical F-X and ΔL-F plots for mechanical unfolding measurements of hTelo-8 fragment showing two-step unfolding events ......................................................... 104

Figure 5.1: Schematic of experimental setup (not to scale) for force based single-molecule investigation of ligand-quadruplex interaction ................................................................. 109

Figure 5.2: Mechanical unfolding of hTelo-4 with and without pyridostatin (PDS) .......... 111

Figure 5.3: Folding kinetics of human telomeric DNA G-quadruplex structure in presence or absence of pyridostatin ................................................................. 116

Figure 5.4: Proposed pathways for binding of pyridostatin ligand to human telomeric G-quadruplex structure ......................................................................................... 118

Figure 5.5: Measurement of dissociation constant between PDS and human telomeric G-quadruplex structure ......................................................................................... 121

Figure 5.6: Determination of dissociation constant using single ligand concentration ....... 125
List of Tables

Table 2.1: Names and sequences of wild types and mutant human telomeric DNA oligomers used in this study ................................................................. 30

Table 3.1: $\Delta L, F_{\text{unfold}}$ and $\Delta G_{\text{unfold}} (23^\circ C)$ of human telomeric DNA species in 10 mM Tris buffer, pH 7.4 containing 100 mM Na$^+$ ................................................................. 68

Table 3.2: $\Delta L, F_{\text{unfold}}$ and $\Delta G_{\text{unfold}} (23^\circ C)$ of human telomeric DNA species in 10 mM Tris buffer, pH 7.4 with 100 mM K$^+$ ................................................................. 76

Table 4.1: Reaction rate constants ($\text{Avg} \pm \text{SD s}^{-1}$) for transitions in hTelo-4 and hTelo-5 sequences .................................................................................. 101
Acknowledgements

First and foremost, I would like to express my sincere gratitude and cordial admiration to my advisor Dr. Hanbin Mao for his continuous guidance and support throughout my Ph.D. career. I appreciate all his contributions of valuable time, scientific ideas, and funding to make my Ph.D. experience exciting and highly productive. His intelligence, enthusiasm, open-mindedness and passion for the science are always inspiring and motivational for me. His motto “PREPARE FOR THE WORST AND HOPE FOR THE BEST” always strikes my mind and provides a continuous guideline to approach the scientific problems even during the tough times. I learned from him how important it is to be persistent and hardworking to accomplish any goal and become a successful researcher. He was always available to encourage and help the students to solve scientific problems as a professional advisor as well as a guardian. In five years of my Ph.D. career, his mentorship promoted me to a hard working, confident and successful researcher. I do not have enough words to express my thanks to him.

It is a matter of great pleasure to me to thank all of the present and past members of the Mao lab who immeasurably contributed to my personal and professional time at Kent State. The time I spent with them was congenial and full of joy with outstanding environment to share the experiences and cultures from the diverse group of people around the world. The group has not only been a trademark of friendship but also a
forum for excellent teamwork and discussion. Especially, I am grateful to the past lab members Soma Dhakal and Zhongbo Yu for their continuous help, support and valuable contribution in my research. I am also thankful to Malcom Mathis, Philip Yangyuro, Yunxi Cui, Zhe Shi, Chiran Ghimire, Sangeetha Selvam, Prakash Shrestha, Jibin Punnoose and Shankar Mandal for their help and active contribution in my research projects.

I would also like to acknowledge Professor Hiroshi Sugiyama and Professor Masayuki Endo, Kyoto University, Japan for providing me a great opportunity to work in their labs in a collaborative project. This was an exceptional chance for me to learn DNA nanotechnology with the pioneers of the field along with the opportunity to experience the scientific researches in different laboratories around the world. I appreciate their comments and suggestions to improve my research skills and would like to thank them and all of the group members there for their help and support during my stay in Japan. I am also grateful to Professor Shankar Balasubramanian, The University of Cambridge, UK for his insightful comments, inspiring ideas and valuable discussions on a collaborative project. It was a rare chance for me to work with a person who is one of the pioneers in my research field. These collaborative projects allowed me to learn enormous ideas and research skills under the direct supervision of world-class scientists and associated community.

I would like to thank all the members of my doctoral dissertation committee, Dr. Mietek Jaroniec, Dr. Bansidhar Datta, Dr. Hamza Balci and previous committee member
Dr. Chun-Che Tsai for their continuous support throughout my Ph.D. and valuable comments, suggestions and feedbacks on my research proposal, prospectus and dissertation. This dissertation would have not been accomplished as it is without the continuous support from the Department of Chemistry & Biochemistry and its all time permission to use every available research facility. I would like to thank all the faculties and staff members of the Department, especially, the staff members Erin Michael, Janie Viers, Arla McPherson and Larry Maurer for their immense help and co-operation. I am also grateful to my friends, particularly Prabin Rai, Harishchandra Subedi and Prakash Kharel, their families and overall Nepalese community who are the superior and enjoyable part of my life at Kent - thousands of miles away from my home.

At last but not the least, I would like to thank my family members for their unconditional love, encouragement and support throughout my whole academic career. I would like to thank my parents Ek Dev Koirala and Bimakala Koirala, sisters Bindu Bashyal and Anju Koirala and brother Keshab Koirala for their inspiration and love in every steps of my life. I would like to specially thank my wife Sharmila Koirala for her support, encouragement and love in every moment of both good and bad times. She is a true and great life-partner and past several years would have not been an easy journey without her in terms of both academic and personal life. I have no words to convey how much I love her. Finally, I would like to thank and express my love to my daughter Dipisha Koirala. She is my inspiration and ultimate source of energy to begin my day with an adorable smile.
I would like to dedicate this dissertation to my parents and wife for their constant support and unconditional love throughout my academic career and overall life.

Deepak P. Koirala

April 24, 2014
Summary

DNA secondary structures such as G-quadruplexes have attracted extensive research attention in recent years because of their vital biological roles and tremendous applications in therapeutics, biosensing and nanobiotechnology. Human telomeric G-quadruplex structures are of particular interest because human telomeres are closely associated with the genetic integrity, cell proliferation, aging and cancer. Fundamental understanding of the transition dynamics and the equilibrium population distribution of the G-quadruplexes, their intermediates and other alternative structures in the human telomeric DNA sequences and their interactions with small-molecule ligands or proteins are, therefore, crucial. This dissertation has uncovered the previously unknown folding dynamics and population distribution of human telomeric G-quadruplex, intermediate and misfolded G-quadruplex structures as well as their interaction with small-molecule ligands.

Intramolecular folding in three and four tandem guanine repeats of human telomeric DNA have been investigated using Optical-Tweezers, Molecular Dynamics Simulation, Circular Dichroism and Dimethyl Sulphate footprinting. An unprecedented, mechanically and thermodynamically stable, folded species has been observed in the sequence with three tandem guanine repeats which is consistent with a G-triplex conformation. Similar species has also been found to coexist with a prevalent G-
quadruplex structure in a DNA sequence with four tandem guanine repeats. Such observations suggest a complex folding pattern of the human telomeric DNA in which G-triplex could be an intermediate structure to the G-quadruplex.

Due to complex folding dynamics and stunning structural polymorphism, population equilibrium in human telomeric sequence is expected to be intricate and beyond the resolution of ensemble-average techniques, such as CD, NMR, or X-ray crystallography. By combining a force-jump method at the single-molecular level and a statistical population deconvolution at the sub-nanometre resolution, a complex population network with previously unknown transition dynamics in human telomeric sequences that contain four to eight TTA GGG repeats has been revealed. The kinetics data firmly establish that G-triplexes are intermediates to G-quadruplexes while long-loop G-quadruplexes are misfolded population minorities whose formation and disassembly are faster than G-triplexes or regular G-quadruplexes. The existence of misfolded DNA supports the emerging view that structural and kinetic complexities of DNA can rival those of RNA or proteins. While G-quadruplexes are the most prevalent species in all the sequences studied, the abundance of a misfolded G-quadruplex in a particular telomeric sequence decreases with an increase in the loop length or the number of long-loops in the structure. These population patterns support the prediction that in the full-length 3' overhang of human telomeres, G-quadruplexes with shortest TTA loops would be the most dominant species, which justifies the modelling role of regular G-quadruplexes in the investigation of telomeric structures.
Since ligand-induced stabilization of telomeric DNA G-quadruplex has potential in cancer treatment through the inhibition of telomerase, an enzyme overexpressed in many cancer cells, understanding the kinetic, thermodynamic, and mechanical properties of small molecule binding to these structures is important. However, classical ensemble assays are unable to measure these parameters simultaneously in a single platform. The single-molecule platform is anticipated to provide detailed insights into the interactions between human telomeric G-quadruplex structures and small-molecule ligands. The increased mechanical stability of pyridostatin bound G-quadruplex permitted the determination of the dissociation constant ($K_d$) of 490 ± 80 nM. The free energy change of binding obtained from a Hessian-like process provided an identical $K_d$ for pyridostatin and a $K_d$ of 42 ± 3 µM for a weaker ligand RR110. Because of its unique abilities, this single-molecule platform is anticipated to provide detailed insights into the interaction between ligands and macromolecules of biological relevance.
Chapter I

Introduction and Background

A PART OF THIS CHAPTER HAS BEEN PUBLISHED AS AN ARTICLE IN REVIEWS IN ANALYTICAL CHEMISTRY, REV. ANAL. CHEM. 2013, 48, 197-208. ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED WITH THE COPYRIGHT PERMISSION FROM WALTER DE GRUYTER GMBH

DNA is a fundamental building block of life and carrier of genetic information. Naturally occurring canonical B-form of DNA is a right-handed double helix structure, which consists of Watson-Crick base pairing of adenine, A with thiamine, T and guanine, G with cytosine, C. Besides this primary canonical form, DNA has also been found to adopt various secondary conformations such as Z-DNA, H-DNA, cruciform DNA, i-motif and G-quadruplex. These non-canonical DNA structures have attracted tremendous research attention in recent years because of their propensity to intrude vital biological processes, induce genetic instability, and cause variety of diseases. Current advances have revealed that these structures not only play crucial regulatory roles in many biological functions such as gene expression and telomere maintenance but also
render unprecedented therapeutic targets to design efficient and site specific drugs.\textsuperscript{11-16} In addition, DNA secondary structures have high potential in nano-electronic devices, biosensing, and supra-molecular chemistry applications, which further intensify their research consideration over the past few decades.\textsuperscript{17-22} Among variety of non-B DNA structures, G-quadruplex is probably the most extensively studied tetraplex structure so far because of its high physiological relevance and broad application range in many areas such as drug targeting, biosensing and biomolecular nano-machinery.\textsuperscript{11, 14, 15, 17, 19, 23-31}

1.1 G-quadruplex DNA

In 1962, fundamental rationale behind the helical aggregation of guanosine 5'-monophosphate (GMP) to form a guanosine gel led to the discovery of the G-quartet structure.\textsuperscript{32} Unlike Watson-Crick A$\rightarrow$T and G$\rightarrow$C base pairing in a canonical double-helix DNA, G-quartet structure is composed of four guanines interconnected by Hoogsteen hydrogen bonds to form a tetrameric unit (Figure 1.1a). The Hoogsteen base pairing in G-quartet occurs between N\textsubscript{7} and C\textsubscript{6} amino group of the adjacent guanines, and O\textsubscript{6} carbonyl of each guanine coordinates with the monovalent metal cations such as K\textsuperscript{+} or Na\textsuperscript{+} to further stabilize the structure.\textsuperscript{33-35} Stacking of two or more G-quartets on top of other constructs a tetraplex DNA conformation which is commonly known as G4-DNA or G-quadruplex (Figure 1.1b).\textsuperscript{33, 36, 37}
**Figure 1.1:** Structure and topology of various G-quadruplex conformations. Arrows represent the direction of DNA strands from 5' to 3' end.
Based on the number of DNA strands involved in a structure, G-quadruplexes have been categorized into intermolecular or intramolecular types. The tetramolecular and the bi-molecular (collectively intermolecular) structures are formed from four and two separate DNA strands, respectively (Figure 1.1c and d), whereas a unimolecular (intramolecular) structure (Figure 1.1e) is formed by the folding of a DNA sequence having at least four tandem guanine repeats separated by at least one nucleotide. In such case, the nucleotides between two guanine stretches are involved into the loops. The polarity or the orientation of the strands that constitute a G-quadruplex conformation further classify its topological structure as a parallel or an antiparallel type. In a parallel type topology (Figure 1.1c and f), the polarity for all strands is same whereas in an antiparallel type (Figure 1d and e), each strand has opposite orientation with respect to the other two adjacent strands. In many cases, hybrid type conformation of a quadruplex (Figure 1g and h) is also possible due to the mixed parallel and antiparallel orientation of the strands. Almost all intermolecular G-quadruplexes uncovered so far are known to assume a parallel topology. However, the specific conformation of an intramolecular G-quadruplex is governed by the DNA sequence, the loop size, and the solution factors such as cation species and molecular crowding conditions. Since only a few copies of the genomic DNA are present in a cell, intramolecular G-quadruplex structures are physiologically more relevant compared to those of intermolecular structures.
1.2 DNA G-quadruplex and Human Telomere

Recent advances on genome-wide sequence analysis have discovered the important information on the prevalence and the location of potential G-quadruplex forming sequences in both prokaryotic and eukaryotic genomes. Studies have revealed that the human genome consists of about 376,000 sites potential to form stable G-quadruplex structures. Surprisingly, those sites are found to be located at the crucial regions of the genome such as promoters, oncogenes and telomeres, in which ~43% of the human gene promoters have potential to form at least one G-quadruplex structure. Although the human genomic DNA is a B-form double helix and therefore formation of a G-quadruplex structure requires the breaking of more stable Watson-Crick base-pairing to form the Hoogsten base-pairing, such opportunities could arise during the biological processes such as DNA replication, transcription and recombination. During these processes, when regions of double-stranded DNA become transiently single-stranded, G-quadruplex formation could be favorable. Moreover, negative super-coiling of DNA arising from these processes could further stimulate the G-quadruplex formation in the human genome. On the other hand, 3' end of the human telomeric DNA is naturally a single-stranded overhang that offers an energetically more favorable condition to form G-quadruplex structures in the region.
Human telomeres are the specialized nucleoprotein structures at each terminus of the linear chromosomes, which consist of a long stretch of double-stranded tandem repeat sequence and a short single-stranded G-rich overhang. Telomeres play a vital role in maintaining genetic stability and cell growth by preventing gene erosion, non-homologous end-end fusion of chromosomes, and attack of nuclease. With highly conserved tandem arrays of TTAGGG repeat sequence, telomeres have the highest concentration of potential G-quadruplex forming sequences in the human genome. Indeed, the formation of G-quadruplex structures in the human telomeric DNA sequences containing four tandem TTAGGG repeats have already been demonstrated in several physiologically relevant in vitro studies. The G-quadruplexes formed in those sequences have been observed to be highly polymorphic and dynamic in nature. Particular conformation assumed by a telomeric G-quadruplex has been found to be dependent on variety of factors such as stabilizing ion, flanking sequences and the molecular crowding conditions. For example, same human telomeric DNA sequence prefers to form an antiparallel basket type G-quadruplex in a Na\(^+\) solution whereas a hybrid type G-quadruplex in K\(^+\) solution. However, the reported K\(^+\)-containing crystal structure of human telomeric G-quadruplex is a parallel-stranded propeller type. Although earlier in vitro studies based on bulk experiments under cell-mimicking molecular crowding conditions have proposed that the conformation of a G-quadruplex structure formed inside the cell is parallel, recent in vitro and in vivo studies at the single-molecule level have demonstrated additional possibilities of
assuming multiple conformations by a G-quadruplex in the molecularly crowded solutions or inside the cell.47, 50, 81, 82

**Figure 1.2:** A cartoon depicting the physiological relevance of human telomeric DNA G-quadruplex structures in regulatory biological functions and telomere based cancer therapeutics.
1.3 Biological Significance of Human Telomeric G-quadruplex

Although human telomeric DNA sequences have been shown to readily form G-quadruplexes under physiological concentrations of K\(^+\) and Na\(^+\) or in cell mimicking molecularly crowded solutions \textit{in vitro}, the formation of these structures \textit{in vivo} is expected to be more complicated. However, biochemical and biophysical evidences from many studies suggest that telomeric G-quadruplexes not only exist inside the cell but also involve in various regulatory biological activities.\(^{83-85}\) Recently, direct visualization of DNA G-quadruplex structures in various regions of human chromosomes including telomeres has tremendously amplified their physiological relevance and potential application in therapeutics.\(^{86}\) In fact, human telomeric G-quadruplexes are known as highly potential anti-cancer drug targets (Figure 1.2).\(^{11, 24, 25, 87}\) In normal cells, telomere is shortened each time when chromosome is replicated during cell division that eventually leads to a programmed cell death called apoptosis.\(^{66}\) However, in most of the cancer cells (80-85%), the telomere maintains its length because of upregulated telomerase.\(^{66}\) Telomerase is a reverse transcriptase enzyme which synthesizes and directs the telomeric repeats onto the 3' end of existing telomere using its RNA component as the
Any chemical hindrance or mechanical blockage that interferes with the telomerase could be useful to control its activity and consequently the cancer.

Since the telomerase action requires a single-stranded telomeric overhang, stable G-quadruplex structures can serve as a chemical hindrance by reducing the substrate binding affinity as well as a mechanical barrier by preventing the motory action of the enzyme during the telomere elongation process. Towards such therapeutic applications, several small-molecule ligands have been found to target and stabilize human telomeric G-quadruplex structures \textit{in vitro}. For example, small molecule ligand telomestatin is found to be effective in stabilizing human telomeric G-quadruplex structures \textsuperscript{89, 90} and consequently reducing the polymerase activity of the telomerase.\textsuperscript{15, 89, 90} However, \textit{in vivo} studies on targeting and stabilizing the G-quadruplex structures by small-molecule ligands are the major challenges associated with the real life applications of telomere-based cancer therapy.

1.4 Methods of G-quadruplex Study

1.4.1 Conventional Methods

A wide variety of methods have been used to explore diverse aspects of the G-quadruplex structures formed in various guanine rich DNA or RNA sequences both \textit{in vitro} and \textit{in vivo}. Routinely used ensemble average techniques for the characterization of
G-quadruplex structures include X-ray crystallography, nuclear magnetic resonance (NMR), fluorescence spectroscopy, circular dichroism (CD) spectroscopy, UV-vis spectroscopy and chemical footprinting. Several X-ray crystal structures of the G-quadruplex DNA with or without a bound ligand have been reported in the literature, including that of human telomeric G-quadruplex in Na\(^+\) NMR, circular dichroism, fluorescence, and chemical footprinting methods have been effectively utilized to characterize the conformation of these structures in solutions and probe their structural polymorphism, molecular dynamics and physio-chemical properties. These methods have also been employed to investigate the binding interactions between G-quadruplex structures and small-molecule or protein ligands. In addition, biochemical and biophysical assays based on surface plasmon resonance (SPR) and fluorescence resonance energy transfer (FRET) to characterize the various properties of the quadruplex-ligand binding have also been reported. In recent years, techniques such as mass spectrometry (MS) and capillary electrophoresis (CE) have been used to reveal the molecular insights into the G-quadruplex structures and their binding interactions with variety of ligands.

Although aforementioned methods are commonly used for G-quadruplex associated studies, they suffer from several drawbacks. These techniques rely on the ensemble-averaging of signals, which inherently lack the ability to identify a minor sub-population present in a solution together with a major species. This inherent difficulty of
these methods limit their access to the detailed properties of a specific sub-population, although it is crucial information to understand the molecular population equilibrium and transition dynamics of interrelated species. Especially in the context of a biological function where population-activity relationship play a substantial role, understanding of such equilibrium dynamics among populations become essential. For example, it is of fundamental interest to uncover the transition kinetics and population equilibrium between various conformations of G-quadruplexes to explore their intracellular roles and targeting a particularly desired species with therapeutic drugs. In addition, the folding-unfolding mechanism of the G-quadruplex, which has been proposed to involve intermediate structures, has not been fully understood yet. Furthermore, investigation on the G-quadruplex structures formed in long guanine-rich sequences such as human telomeric DNA by these standard bulk methods is difficult. For example, in such cases, it is complicated to resolve the overlapping NMR signals arising from multiple units of the G-quadruplexes or obtain a pure crystal for X-ray crystallographic studies. In many cases, mutations have been introduced into the molecule to selectively populate the species of interest in a population mixture and thereby resolving the structure. However, such procedure changes the population equilibrium and distorts the transition between different species. Therefore, it becomes necessary to use the single-molecule methods to clarify the population equilibrium and the transition kinetics among different species formed in guanine-rich DNA or RNA sequences. Compared to the bulk methods, capable of probing one molecule at a time, single-molecule methods offer an unprecedented
opportunity to identify conformation and follow transition kinetics of individual species in a population mixture.

1.4.2 Single-Molecule Methods

To overcome the inherent limitations of the ensemble average techniques, new approaches based on the single-molecule methods have been developed in recent years to investigate the G-quadruplex structures and their interactions with the ligands. Single-molecule methods not only provide detailed insights into the minor sub-populations but also allow the direct measurement of microscopic and stochastic properties of the individual molecules such as mechanochemistry and kinetics, which would otherwise not be possible by conventional bulk methods.

Fluorescence microscopy is a highly sensitive single-molecule method which has been used to follow the single-molecules in their natural environments. It is probably the most common method of investigating the biomolecules both in vitro and in vivo including DNA, RNA and proteins.\textsuperscript{108-113} By using single-molecule FRET, dynamic properties of the G-quadruplex folding and its interaction with the small-molecules and protein ligands have been demonstrated recently.\textsuperscript{114-120} Such method has revealed that the G-quadruplex structures are highly polymorphic in solution and their binding interaction with the ligands is highly dynamic.\textsuperscript{118, 121, 122} Although single-molecule FRET is highly sensitive, it suffers from high background signals and easy photo-bleaching of
fluorophores. Moreover, labeling of the fluorophores increases the experimental costs and can change the microscopic properties of the biomolecules under study. In recent years, tapping mode high speed atomic force microscopy has also been successfully employed to directly visualize the assembly and the disassembly of the G-quadruplex structures in real time under physiologically relevant conditions using a DNA origami platform.\textsuperscript{123, 124} However, being highly expensive to make DNA origami, such method could not provide structural details and conformation of the G-quadruplex structure. In addition, this method has only been used to study the inter-molecular G-quadruplexes, which are physiologically not relevant compared to that of intra-molecular structures such as in human telomeric overhang in which single-stranded DNA forms intra-molecularly folded G-quadruplex structures.

1.4.3 Mechanical Force-Based Methods

Mechanical force is a universal parameter that can be used to characterize a wide range of systems ranging from chemical bonds, intermolecular interactions, macromolecular structures, and many biological and biochemical processes. The basic interaction between individual atoms is described in covalent or non-covalent bonds, whose strength can be characterized in force. Change in the conformation of a macromolecule is often accompanied by a variation of the tension sustained by that molecule. Many vital biological processes can be monitored and characterized in terms
of mechanical forces. For example, the biophysical aspects of the attachment and the rolling of leukocytes along the wall of blood capillaries during inflammation involve a complex balance of forces arising from blood-flow-induced hydrodynamic shearing effects and the adhesive forces between the leukocytes and the capillary wall.\textsuperscript{125-127}

Biochemical actions of many proteins, such as DNA\textsuperscript{128} or RNA polymerase,\textsuperscript{122, 123} exonuclease,\textsuperscript{129} myosin,\textsuperscript{130} and kinesin,\textsuperscript{131} generate forces up to tens of picoNewtons. On the other hand, activities of those proteins are often influenced by externally applied force.\textsuperscript{132} In ligand-receptor binding, strength of the interaction between two binding partners can be measured in terms of the mechanical force that is required to destroy the interaction from a specific direction.\textsuperscript{133-135} In such a case, force is applied at a constant rate to overcome the energy barrier. The dynamics of the binding process can be retrieved by measuring the unbinding rate at a constant force or by measuring rupture force as a function of force loading rate.
Figure 1.3: Schematic of the commonly used single-molecule methods for mechanical unfolding and refolding of biomolecules. A biomolecule of interest (yellow) is sandwiched between two spacers, dsDNA for example, which can be tethered between two immobilized surfaces by affinity interactions or covalent linkages in atomic force microscope (a), micro-needle (b), magnetic tweezers (c), or (d) laser tweezers instruments. When a tension is applied to stretch the tethered molecule, the biomolecule of interest is unfolded at a certain force. Reducing the tension allows refolding of the biomolecule. This unfolding-refolding process can be repeated many times until the tether is detached from the surfaces. To observe multiple unfolding and refolding events, the surface attachment should be stronger than the unfolding force ($F_{\text{unfold}}$) of the structure formed in the biomolecule.
Despite its importance, there exists a lack of ensemble-average methods to interrogate force in solution. Hydrodynamic forces from a flowing fluid have been used to monitor the mechanical aspect of biologically relevant processes such as adhesion and rolling of white blood cells inside blood capillaries during inflammation.\textsuperscript{136, 137} However, such methods have difficulties to explore the roles of individual binding units at the molecular level. As a result, only little information has been collected to understand the mechanochemistry,\textsuperscript{138} an emerging field to study the coupling between chemical energy and mechanical energy. Single molecule techniques, however, start to address this problem. Compared to bulk assays where ensemble information is obtained, single molecule experiments produce stochastic and discrete signals that can be statistically analyzed to reveal the property of sub-groups in a population and the energetics of a particular reaction trajectory in a transition process.\textsuperscript{139} Among many tools to manipulate single molecules, force based techniques, such as atomic force microscope,\textsuperscript{140, 141} optical tweezers,\textsuperscript{131, 142, 143} micro-needles,\textsuperscript{144} and magnetic tweezers,\textsuperscript{145, 146} provide a unique capability to investigate the force generated by or applied to individual molecules. Details of each method are reviewed in recent publications.\textsuperscript{147-149} A brief introduction of these methods with their important advantages and disadvantages has been presented below.

Atomic force microscope is commercially available. It uses a micro-fabricated cantilever to attach a biomolecule through one of its ends (Figure 1.3a). The other end of
the biomolecule is immobilized onto a substrate surface. The displacement between the
cantilever and the surface is often controlled by a piezoelectric device. Force signal,
which ranges from $10^{-11}$ to $10^{-7}$ N, in the biomolecule is measured by the deflection of the
cantilever. This method has nanometer spatial resolution and picoNewtons force
resolution. The stiffness of the cantilever ranges from 0.001-100 Nm$^{-1}$. It is noteworthy
that the stiffer the cantilever, the lower the force sensitivity. As dynamic range of the
force measurement is broad, this method has been widely used to investigate the strength
of intramolecular and intermolecular interactions in biomolecules.

Micro-needles are usually 50-500 $\mu$m long and 0.1-1 $\mu$m in diameter. Because of
the lower stiffness ($10^{-6}$ to 1 Nm$^{-1}$), this method has an advantage over AFM to study
delicate systems. Setup is similar to AFM in which a biomolecule is sandwiched
between the tip of a micro-needle and a substrate surface (Figure 1.3b). Force signal can
be measured by observing the displacement of a bendable micro-needle via imaging the
micro-needle itself or by using a chemically etched optical fiber that projects a reflected
light from its tip onto a photodiode. This method can measure force from $10^{-12}$ to $10^{-10}$ N
with a minimum distance of $10^{-9}$ m. The instrument is not commercially available.

Magnetic tweezers use permanent or electro-magnets to trap magnetic micro-
particles (Figure 1.3c). Bio-molecules can be tethered between magnetic particles and a
non-magnetic surface. This method can measure force from $10^{-14}$ to $10^{-11}$ N with a
minimum distance of $10^{-8}$ m. Since it can trap multiple particles, throughput is higher
compared to other force-based single-molecule approaches. It has a unique capability to rotate magnetic particles, thereby allowing torque measurements on torsionally constrained biomolecules. Recently, various types of magnetic tweezers have been developed,\textsuperscript{150, 151} however, the major disadvantage associated with these methods lies in the fact that force is not measured directly.\textsuperscript{147-149}

Optical tweezers can trap dielectric micro-particles at a laser focus (Figure 1.3d). Trap stiffness of a laser trap (10\textsuperscript{-10} to 10\textsuperscript{-3} Nm\textsuperscript{-1}) is much smaller than that of AFM cantilevers, which allows a better force resolution in the range of 10\textsuperscript{-13} to 10\textsuperscript{-10} N. It has a spatial resolution of 10\textsuperscript{-9} - 10\textsuperscript{-10} m, which allows resolving single base pairs in DNA. The force exerted on a micro-particle can be flexibly controlled by parameters such as laser power, particle size, and difference in the refractive index between the particle and the trapping medium. In single-beam optical tweezers, a biomolecule is usually tethered between a trapped particle at the laser focus and a surface of a substrate or a particle surface held at the tip of a micropipette by suction.\textsuperscript{152-155} In dual-beam optical tweezers, the molecule is often tethered between two trapped particles at two separate laser foci.\textsuperscript{152, 156, 157} The major disadvantage of this method is laser induced damage of biomolecules.

Using these single-molecule tools, different approaches can be adopted to investigate the mechanical property of a biomolecule and its interaction with a ligand. In one setup, two binding partners can be immobilized to two surfaces separately. After binding between these two partners is accomplished by bringing two surfaces together,
moving one surface away from another increases the mechanical tension. The increased tension eventually destroys the binding interaction. However, the major challenges in this approach are the precise measurement of the distance change during the unfolding and the reversibility of the process. In another approach, an intramolecularly folded biomolecule is first tethered between two surfaces using affinity or covalent linkages (Figure 1.3). As one surface is moved away from another by a computer-controlled motor, the tension in the biomolecule increases until unfolding occurs, which is manifested by a sudden change in the tension or the end-to-end distance in a force-extension ($F$-$X$) curve recorded for this process. During the opposite process in which the end-to-end distance becomes shorter, the biomolecule refolds. The same processes can be performed in the presence of the ligand to evaluate ligand-receptor interactions.

**Figure 1.4:** Typical force-extension ($F$-$X$) curves of unfolding (red) and refolding (green) of a biomolecule (a). Change in extension ($\Delta x$) and the unfolding force ($F_{\text{unfold}}$)
for the unfolding of the biomolecule is measured directly from the F-X curves. (b) The effect of the applied force on the unfolding-unfolding energy landscape of a biomolecule. The green and red curves represent energy profiles of a biomolecule before and after application of force, respectively.

Mechanical unfolding or refolding processes present a completely different mechanism compared to the temperature or chemical denaturant mediated unfolding or refolding. Whereas the force-induced processes have localized effects on a molecule, the temperature or chemical assisted events are global in nature. The localized force effect suggests that the direction of unfolding is rather critical for the process. On the other hand, it is expected that unfolding/refolding should be an ensemble average of all unfolding directions that are presented in the temperature or the chemical assisted transition processes owing to their global nature. From energy perspective, force ($F$) reduces free energy of a system by a factor of $F \times x$, where $x$ is the unfolding distance of a biomolecule. As shown in Figure 1.4b, such a reduction brings down the energy barrier of a two-state folding/unfolding system by $F \times x^\dagger$, where $x^\dagger$ is the transition distance, or the distance between the folded structure and the transition state. With application of a specific force ($F_{\text{unfold}}$), the energy barrier is reduced to a level within reach of thermal energy of environment ($k_B T$, where $k_B$ is the Boltzmann constant and $T$ is absolute temperature). This leads to unfolding of the structure. Assuming $x^\dagger$ does not change during the application of force, the magnitude of $F_{\text{unfold}}$ is inversely dependent on the
transition distance with a larger distance corresponding to a smaller $F_{\text{unfold}}$ value.\textsuperscript{158} The relationship between $F_{\text{rupture}}$ and $x^\dagger$ can be understood by the elasticity of a molecule. A soft and flexible molecule requires a long distance to be unfolded (larger $x^\dagger$). The force ($F_{\text{unfold}}$) along this long distance is expected to be small. On the other hand, a rigid but fragile molecule only needs a little stretch in its structure (smaller $x^\dagger$) to induce unfolding. However, the force leading to this little perturbation ($F_{\text{unfold}}$) is expected to be significantly higher than the previous case.

Since $F_{\text{unfold}}$ alone does not fully account for the mechanical perturbation defined by $F \times x^\dagger$, there is a caveat to use this parameter to compare the mechanical stability of folded molecules. Comparison of $x^\dagger$ among DNA secondary species, such as hairpins\textsuperscript{159}, DNA G-quadruplexes,\textsuperscript{157, 160, 161} DNA i-motifs,\textsuperscript{162} and their intermediates,\textsuperscript{163} has suggested that $x^\dagger$ varies with different molecules. Within the same molecule, $x^\dagger$ is anisotropic among different unfolding trajectories.\textsuperscript{161} Therefore, mechanical aspects of unfolding can be described accurately only when both $F_{\text{unfold}}$ and $x^\dagger$ are taken into account. The other caution to use $F_{\text{unfold}}$ for the comparison is that magnitude of $F_{\text{unfold}}$ is dependent on the force loading rate. The dependence is rather pronounced for processes with slow unfolding rates.\textsuperscript{164, 165} In such a case, $F_{\text{unfold}}$ for reversible folding and unfolding processes occurs in a timescale often not reachable in an experiment.
1.4.4 Manipulation of G-quadruplex by Optical-Tweezers

Three force-based single-molecule techniques, force mode AFM, magnetic tweezers and optical tweezers have been developed in recent years to manipulate the biomolecules.\textsuperscript{148, 149, 152} The force based approaches are generally not influenced by the solution conditions which make these techniques superior over ensemble average techniques. Additionally, these methods significantly expand the measurement range with respect to single-molecule FRET, which works in the range of 2-9 nm.\textsuperscript{166} Nevertheless, method of choice for the force-based G-quadruplex associated studies is determined by several factors such as resolution and dynamic range of force and extension measurements. Because the atomic force microscope has comparatively low resolution of force measurements, it is not an appropriate method for the G-quadruplex associated studies in which structure does not require high force to unfold, although it has been broadly applied to manipulate other biomolecules such as proteins. Compared to other force-based single-molecule approaches, magnetic tweezers has higher throughput and a unique capability to rotate magnetic particles thereby allowing torque measurements on torsionally constrained biomolecules.\textsuperscript{149, 167, 168} It has, therefore, high potential to study the G-quadruplex structures under physiologically more relevant conditions such as superhelicity. However, the major disadvantage associated with this method lies in the fact that force is not measured directly. In recent studies, magnetic tweezers have been
utilized to retrieve the folding-unfolding dynamics and energy landscapes of the human telomeric G-quadruplex structures.\textsuperscript{169, 170}

Currently available optical tweezers are the outcome of the great discovery of Ashkin in early 1970s,\textsuperscript{171} in which a focused laser beam is used to trap and manipulate the dielectric micro-particles.\textsuperscript{143, 172} The force exerted on a micron sized particles can be flexibly controlled by parameters such as laser power, particle size, and difference in the refractive index between the particle and the trapping medium. Because of the better force resolution compared to atomic force microscope and sub-nanometer spatial resolution, optical tweezers have been extensively used to explore a wide-range of biophysical phenomena. They have been exploited to manipulate and characterize individual cells, bacteria and viruses. Being able to resolve single base pairs in DNA, this method have also been implemented to study the individual proteins and nucleic acids structures.\textsuperscript{173, 174-176} Moreover, biophysical characterization of molecular motors such as myosin, kinesin and DNA or RNA polymerases have also been well explored using optical tweezers.\textsuperscript{147, 155, 177} Despite these unique advantages and abilities, however, optical tweezers had not been employed to investigate the G-quadruplex structures until Mao lab in 2009 demonstrated the first example.\textsuperscript{157} Currently, several studies have been reported on different aspects of these structures and their interaction with ligands based on optical tweezers measurements. Using this method, the mechanical stability, thermodynamics and folding-unfolding kinetics of various G-quadruplex structures in
single- as well as double-stranded DNA have already been revealed at the molecular and sub-molecular level.\textsuperscript{81, 160, 161, 178, 179}

1.5 Relevance of Present Study

As described in previous sections, understanding the structure, stability and dynamics of the G-quadruplex structures formed in the human genome are crucial to realize the biological roles of such structures and their possible therapeutic applications. This dissertation aims to describe the \textit{in vitro} biophysical studies on the G-quadruplex structures, particularly those formed in the human telomeric DNA sequences using optical tweezers and other complementary methods. Human telomeres are highly conserved region of the human genome that play vital regulatory functions in the region and are highly promising anti-cancer drug targets as well. To the date, there are several studies reported to reveal the details of the structural polymorphism, thermodynamics and kinetics of G-quadruplex structures formed in this region.\textsuperscript{63, 114, 117, 119, 180-189} Nevertheless, this dissertation aims to particularly address the following aspects which have not been explored well because of the limited access of the conventional methods to those facets.

I. Being a polymerase enzyme, telomerase resembles motor proteins. The inhibition of the telomerase by G-quadruplex structures requires that the mechanical stability of these structures must be higher than the stall force of the enzyme. To evaluate mechanochemistry and dynamic interactions between G-quadruplex and telomerase, it is
crucial to understand the mechanical characteristics of the telomeric G-quadruplex itself under physiologically relevant conditions. However, none of the previous studies have provided the mechanochemical prospective of the human telomeric G-quadruplexes.

II. Several studies have proposed that the intermediate structures are involved in the folding-unfolding pathways of the human telomeric G-quadruplex.\textsuperscript{181, 187, 190} To explore the regulatory mechanism of G-quadruplex structures in cellular processes and targeting them for therapeutic applications, it is fundamentally necessary to investigate the folding-unfolding mechanism and associated thermodynamics and kinetics of these intermediates with respect to the fully folded G-quadruplexes. However, there are no concrete experimental evidences on the existence of such intermediates and thus their thermodynamic, kinetic and mechanochemical properties.

III. Although human telomere consists of a DNA sequence with hundreds of TTAGGG repeats capable of forming multiple G-quadruplex units, most of the previous studies have focused on the monomeric G-quadruplex unit formed by four TTAGGG repeats. However, possibilities of alternative structures in the shorter or longer telomeric sequences have not been explored well. Particular biological functions may be associated with the structural complexity and such functions could be regulated by different structures in a population equilibrium that is dependent on cellular conditions such as pH or proteins. To understand the detailed insight on these aspects, dynamics of such population distribution needs to be discovered.
IV. Since small-molecule ligands that stabilize the human telomeric G-quadruplexes have potential as anti-cancer drugs, understanding the mechanochemical behavior of small-molecule binding to G-quadruplex structures is highly important but such studies have not been reported yet. Although there are some studies that have investigated the structure, kinetics and thermodynamics of ligand-quadruplex interaction, there is a need of an approach that provides the mechanical, thermodynamical and kinetic properties of liganded G-quadruplex structures in a single platform.

To accomplish the goals of the project, research was started with the investigation on folded structures formed in human telomeric sequences with three G-rich repeats, TTA(TTAGGG)$_3$ (hTelo-3), and four G-rich repeats, TTA(TTAGGG)$_4$ (hTelo-4). In these cases, results have demonstrated that G-quadruplex is a major folded population in hTelo-4 and there exists a mechanically and thermodynamically stable species different from G-quadruplex in both hTelo-3 and hTelo-4. In addition, this study revealed that such intramolecularly folded species in the hTelo-3 or hTelo-4 are consistent with a triplex conformation and likely to be an intermediate in the folding-unfolding mechanism of the G-quadruplex structures. These findings depict a complex folding mechanism of the structures in the human telomeres and such a mechanism could be associated with the various regulatory biological processes involving telomeres and telomerase.
Upon characterizing the folding patterns of the hTe-lo-3 and hTe-lo-4 sequences, the research was further extended to follow the transition kinetics and the dynamic population equilibrium of human telomeric sequences containing four to eight TTAGGG repeats. In these sequences, a highly complex population equilibrium that contains G-triplex, misfolded GQs, and predominant GQ species were observed. The misfolded GQs harbor one or more G-rich repeats in the loop(s) and the population of these species in a particular sequence decreases with increase in the loop length or the number of long-loops. The presence of the misfolded species testifies the structural complexity of telomeric DNA. The transition kinetics of hTe-lo-4 and hTe-lo-5, TTA (TTAGGG)$_5$ sequences further reflected the complexity of the systems, in which three-state and four-state kinetic models were followed, respectively. The observed population dynamics of telomeric species in the sequences TTA(TTAGGG)$_n$, where n=4 to 8, indicated that in the full-length 3’ end overhang of human telomere, G-quadruplex units with shortest possible TTA loops would be the most prevalent species. This finding justified the modeling role of monomeric unit of a G-quadruplex in the study of human telomeric structures.

Based on the population dynamics and the transition kinetics of longer telomeric sequences, it was confirmed that long telomeric overhang is composed of many monomeric G-quadruplexes in a row similar to the beads-on-a-string.$^{195, 196}$ Therefore, as a model system, hTe-lo-4 sequence was used to investigate the binding interactions
between human telomeric G-quadruplex and the small-molecule ligand. The results demonstrated that a small-molecule, pyridostatin (PDS)\textsuperscript{197, 198} binds to the human telomeric G-quadruplex, increases the stability and promotes the folding of the structure. By a process analogous to the Hess cycle, the dissociation constant of pyridostatin has determined which has not been accurately determined in ensemble assays due to solubility issues. This method simplifies the dissociation constant assay without the requirement for ligand or receptor titration and offers a general platform that can be applied to other biologically relevant ligand-receptor systems. Specifically, this study highlights that G-quadruplexes are important dynamic structures that could play a regulatory role in the telomerase mediated mechanism of telomere elongation. Furthermore, the mechanical information acquired by this system could provide novel perspectives for drug testing and design in the future. In addition, the extra forces required to unfold ligand bound structures shows that in practice, the G-quadruplex structures stabilized by small-molecule ligands such as pyridostatin could well interfere with RNA and DNA polymerases during the processes of transcription and replication \textit{in vivo}. 

Chapter II

Materials and Methods

2.1 Materials

All chemicals, unless specified otherwise, were purchased from VWR (https://www.vwr.com) and Sigma Aldrich (http://www.sigmaaldrich.com) and used without further purification. The DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) (http://www.idtdna.com) and Gene Link (http://www.genelink.com). The oligonucleotides were further purified by denaturing polyacrylamide gel and stored at -20°C. Listed in the Table 2.1 are the wild types or the mutant human telomeric DNA oligomers used in the study. The polystyrene beads coated with streptavidin or anti-digoxigenin for the single-molecule optical-tweezers experiments were purchased from Spherotech (http://www.spherotech.com). Unless specified otherwise, all the enzymes and doxycyclinucleotides were purchased from New England Biolabs (NEB) (https://www.neb.com). Small-molecule ligands, Pyridostatin (PDS)\textsuperscript{198, 199} and RR110,\textsuperscript{200} were received from Prof. Shankar Balasubramanian’s research laboratory (Department of Chemistry, University of Cambridge, United Kingdom). Upon receiving these compounds, aqueous stock solution (10 mM) was stored
at -80°C and 1 mM working aliquots at -20°C. The synthesis and characterization details of these compounds have already been described elsewhere in the literature. The [γ-32P] ATP for the Dimethyl Sulphate (DMS) footprinting was purchased from Perkin Elmer (http://www.perkinelmer.com).

**Table 2.1:** Names and the sequences of the wild types and mutant human telomeric DNA oligomers used in the study. G-tracts in the sequences are underlined.

<table>
<thead>
<tr>
<th>Name of the DNA oligomers</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTel-1</td>
<td>5'-GGG TTA</td>
</tr>
<tr>
<td>hTel-3</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-4</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-5</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-5-Mut</td>
<td>5'-TTA (GGG TTA) TTA AAA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-6</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-7</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
<tr>
<td>hTel-8</td>
<td>5'-TTA (GGG TTA)</td>
</tr>
</tbody>
</table>

### 2.2 CD Measurements
Desired concentrations of oligonucleotide samples were prepared in a 10 mM Tris buffer (pH 7.4) containing 100 mM KCl or NaCl or LiCl or no added salt. The samples were then heated at 95°C for 10 min and rapidly cooled in an ice bath and incubated at room temperature for up to four hours. The fast cooling was performed to allow the comparison of the results with the fast mechanical unfolding-refolding processes in the single-molecule assays. CD spectra were taken in a 1 mm quartz cuvette at room temperature with a Jasco-810 spectropolarimeter (Easton, MD). The reported spectra were the average of three scans with a scanning rate of 100 nm min\(^{-1}\). Spectrum of the corresponding buffer was subtracted from each scan of the sample for baseline correction and smoothed using a Savitzky-Golay function. The CD spectra of the sample, where appropriate, were also taken at different temperatures ranging from 15°C to 85°C by controlling the temperature using a Jasco (model PFD-425S) peltier temperature controller.

### 2.3 Molecular Dynamics Simulation

The MD simulation was carried out using the sequence, 5′-A GGG TTA GGG TTA GGG TTA-3′. The initial atomic coordinates of triplexes were obtained from a basket type G-quadruplex structure (PDB 143D). Triplex models were constructed by deletion of the fourth strand from this antiparallel G-quadruplex structure. The triplex sequence is 5′-A\(^1\)G\(^2\)G\(^3\)T\(^4\)T\(^5\)A\(^7\)G\(^8\)G\(^9\)G\(^10\)T\(^11\)T\(^12\)A\(^13\)G\(^14\)G\(^15\)G\(^16\)T\(^17\)-3′. The
systems were neutralized with 16 Na\(^+\) counter ions (represented by blue asterisks in Figure 3b). The central sodium ion was positioned at 5′-G\(^3\)(syn)G\(^4\)(anti)-3′/5′-G\(^8\)(syn) G\(^9\) (anti)-3′/5′-G(syn)\(^{15}\)G(anti)\(^{16}\)-3′. Classical MD simulations were performed using AMBER version 8 and 9 from the MD-GRAPE system. All nonbonded interactions, van der Waals and Coulomb forces, and energies were calculated using the MD-Grape-3. An AMBER parm99 was applied as a force field for DNA. The time step was set to be 1 fs. The systems were surrounded spherically by TIP3P water molecules. The circle dimensions were chosen to achieve a minimum distance of 38 Å from G3, resulting a typical fundamental cell of 38 × 38 × 38 Å\(^3\) with about 6849 water molecules. The systems were equilibrated for 300 ps with gradual removal of positional restraints on the DNA with the following protocol for the triplex: (i) minimization of water and counter ions for the triplex; (ii) minimization of DNA with the restrain energy of 50 kcal mol\(^{-1}\) Å\(^{-1}\) for the triplex; (iii) minimization of DNA with the restrain energy of 10 kcal mol\(^{-1}\) Å\(^{-1}\); (iv) minimization of DNA with the restrain energy of 5 kcal mol\(^{-1}\) Å\(^{-1}\); (v) 150 ps MD (T = 300 K) holding DNA fixed (5000 kcal mol\(^{-1}\) Å\(^{-1}\)); (vi) 50 ps MD (T = 300K) holding DNA fixed (50 kcal mol\(^{-1}\) Å\(^{-1}\)); (vii) 50 ps MD (T=300K) holding DNA fixed (10 kcal mol\(^{-1}\) Å\(^{-1}\)); (viii) 50 ps MD (T = 300K) holding DNA fixed (5 kcal mol\(^{-1}\) Å\(^{-1}\)); and (ix) MD simulation for 3 ns. We examined the average structure during equilibrium calculation between 2500-3000 ps. The distance between O of 5′ end A\(^{1}\)G\(^{2}\) and O of 3′ end G\(^{16}\)T\(^{17}\) backbone (sugar and phosphate) was found to be 1.5 ± 0.1 nm. This distance
was used as the end-to-end distance of the folded structure to calculate the expected change in contour length for unfolding of hTelo-3 sequence.

2.4 DMS Footprinting

Details of the Dimethyl Sulphate (DMS) footprinting method has been reported elsewhere. To describe briefly, hTelo-3 or hTelo-4 oligonucleotides were radio-labeled at the 5' end by incubating the DNA with T4 polynucleotide kinase (NEB) and [γ-32P] ATP (Perkin Elmer). The labeled products were then purified using MicroSpin™ G-25 Columns (GE, Piscataway, NJ). Radio-labeled oligonucleotide samples were mixed with cold oligonucleotides to obtain a final concentration of 1 µM in 10 mM Tris–HCl buffer, pH 7.4 containing 100 mM NaCl or KCl or no added salt. An aliquot of 100 µL mixed samples were then heated at 97 °C for 10 min and cooled to 25 °C quickly using an ice bath. The samples were then treated with DMS with a final concentration of 1% for 45 seconds. The reaction was stopped with 1.1 mL of stop buffer (2 M β-mercaptoethanol, 300 mM sodium-acetate, 250 mg/ml salmon sperm DNA) and immediately ethanol precipitated followed by washing with 70% ethanol. The DNA pellet was then dried and cleaved using piperidine followed by separation of the fragments on a 10% denaturing acrylamide (19:1 acrylamide: bisacrylamide) gel. The gel was dried on Whatman paper, exposed to a phosphor-imager screen, and scanned with a Typhoon 8600 instrument (Molecular Dynamics). DMS footprinting gel images were quantified using BioRad
QuantityOne™ software. Each guanine band was normalized by comparing the intensity of each band to the total intensity of the corresponding lane. Then the fold protection of each guanine residues was calculated as the ratio of the normalized intensity of a particular band in the presence of salt (100 mM Na⁺ or 100 mM K⁺) to the corresponding band in the absence of the salt.

2.5 Single-Molecule Measurements

2.5.1 Optical-Tweezers Instrumentation

The schematic of the optical-tweezers used in single-molecule mechanochemical studies is shown in Figure 2.1 and the detailed description of the setup has been reported elsewhere. To describe briefly, a diode pumped solid state (DPSS) laser (1064 nm, 4 W, CW mode, BL-106C, Spectra-physics) was used as a trapping laser. P and S polarized laser light from the same laser source constituted the two laser traps. The S polarized light was controlled by a steerable mirror (Madcity Labs Inc., Madison, WI) at a conjugate plane of the back focal plane of a focusing objective (Nikon CFI-Plan-Apochromat 60×, NA 1.2, water immersion, working distance (~320 µm). The exiting P and S polarized beams were collected by an identical objective and detected by two position-sensitive photodetectors (PSDs) (DL100, Pacific Silicon Sensor) separately.
Figure 2.1: Optical layout of the laser-tweezers instrument used for the single-molecule mechanochemical measurements of the folded structures in human telomeric DNA oligomers.

The force of the laser trap was calibrated by the Stokes force and thermal motion measurement. Both methods yielded a similar trap stiffness of $\sim 307 \text{ pN/ (\mu m \times 100mW)}$ (for 0.97 $\mu$m diameter polystyrene beads, Bangs Laboratory, Fishers, IN). Signals from the PSDs were recorded at a frequency of 1000 Hz using the LabView® 8 program (National Instruments Corporation, Austin, TX). Raw data were treated using a Savitzky-
Golay filter with a time constant of 10 ms and converted to force using a Matlab program (The Math Works, Natick, MA).

2.5.2 Preparation of Microfluidic Chamber

The details of the procedure to prepare the microfluidic chambers for optical tweezers experiments have been reported elsewhere. To describe briefly, two methods were used to construct the microfluidic platform as shown in Figure 2.2. The first method was based on soft-photolithography. In this method, the master features similar to those shown in Figure 2.2 were fabricated by etching a negative photoresist, SU-8 2050 (MicroChem Inc., Newton, MA), coated on a glass substrate or silicon wafer. Coating of the photoresist at 1600 rpm spinning rate and etching with a SU-8 developer created a thickness of 100 ± 5 µm. The pattern was then used to prepare films of polydimethylsiloxane (PDMS) using precursor sylgard-184 silicon elastomer base and sylgard-184 silicon elastomer curing agent (Dow Corning Corporation, Midland, MI) with a ratio of 10:1 under a spin rate of 1000 rpm. The PDMS film was cured at 70 °C for 2 hrs (or 55°C overnight). This generated a PDMS film with 140 µm thickness. The injection ports for each channel were prepared by poking the film using syringe needles (Gauge 16G3/2, Becton Dickinson and Company, Franklin Lakes, NJ). The PDMS film was then peeled off, treated in oxygen plasma (plasma cleaner PDC-32G, Harrick
Plasma, Ithaca, NY) for 1 min, and brought into contact with an oxygen plasma treated borosilicate coverslip.

**Figure 2.2:** Schematic of a microfluidic chamber used in optical-tweezers measurements. The colored arrows depict the flow directions of the specified solutions.

In the second method, we prepared the microfluidic chamber by sandwiching a patterned Nesco-film (Azwell, Osaka, Japan) between two glass coverslips (VWR). The microfluidic patterns as shown in Figure 2.2 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 patterned Nesco-film and the two coverslips were thermally sealed at 160 °C. The thickness of the film thus treated (100 ± 10 µm) determined the channel thickness. Samples were injected into microfluidic channels through the holes in one of the coverslips prepared by the same laser cutter. To
transport the beads attached with DNA samples and the streptavidin coated beads into reaction channel, micro-capillary tubes (ID 20 µm, OD 90 µm) were used. For most of the experiments, this method was used to prepare the chambers.

**Figure 2.3:** Flowchart of the molecular biology approach to synthesize the DNA construct used in single-molecule mechanochernical studies. Plasmid DNAs are shown as concentric circles. The overhang sequences after restriction enzyme digestion are also
shown for clarity. In the final construct, single stranded human telomeric DNA fragment is sandwiched between two end-modified double stranded DNA handles.

2.5.3 Molecular Biology Synthesis of DNA Constructs

The flowchart of molecular biology strategy to prepare the DNA constructs for single-molecule mechanochemical studies is shown in Figure 2.3. This procedure to synthesize the DNA construct is similar to the protocol described elsewhere in the literature. Briefly, the human telomeric DNA sequence under investigation was sandwiched between two dsDNA spacers, 2028 bp and 2690 bp. The 2028 bp spacer was labeled at one end with biotin using a 5' end biotinylated primer, 5'-GCA TTA GGA AGC AGC CCA GTA GTA GG, (Integrated DNA Technologies, Coralville, IA) during the PCR amplification of a pBR322 plasmid (New England Biolabs, Ipswich, MA). The PCR product was digested by the XbaI restriction enzyme (New England Biolabs). This biotinylated handle was then ligated with a dsDNA-ssDNA hybrid, 5'-CTA GAC GGT GTG AAA TAC CGC ACAGAT GCG-\textbf{X}-GCC AGC AAG ACG TAG CCC AGC GCG TC, where \textbf{-X-} represents the ssDNA section. This hybrid consists of an XbaI overhang at the 5' end and an EagI overhang at the 3' end. The ssDNA section (\textbf{-X-}) is either a wild type human telomeric or a mutated, or a control sequence or contains no nucleotides depending on the construct used for the specific experiment. The second spacer (2690 bp) was constructed by the ScaI (New England Biolabs) and the EagI (New England Biolabs)
digestions of a pEGFP plasmid (Clontech, Mountain View, CA) and subsequently labeled with digoxigenin (Dig) at the SacI end using 18 µM Dig-dUTP (Roche, Indianapolis, IN) and terminal transferase (Fermentas, Glen Burnie, MD). It was finally ligated with the biotinylated DNA spacer prepared above through the EagI site using T4 DNA ligase (New England Biolabs). The final construct was ethanol precipitated, dissolved in water, and stored at -80°C.

2.5.4 Single-Molecule Mechanochemical Assays

The DNA construct was first immobilized on the surface of the anti-Dig antibody-coated polystyrene beads (diameter: 2.10µm, Spherotech) via the Dig/anti-Dig linkage in 10 mM Tris buffer (pH 7.4 with 100 mM KCl or NaCl or LiCl) at 23 °C. For the ligand binding assays, same buffer containing the ligand in a desired concentration was used. Beads coated with streptavidin (diameter: 0.97 µm, Bangs Laboratory) were also dispersed into the same buffer and injected into the reaction chamber. After separate trapping of these two types of beads by the two laser traps, DNA construct was tethered between the two beads. Once the DNA was tethered, the bead controlled by the steerable mirror was moved away at a constant load speed of ~5.5 pN/s. This rate allowed us to collect significant amount of data in a reasonable time scale at the condition close to the unfolding equilibrium of a DNA secondary structure. This process allowed the gradual increase in the tension of the DNA construct, which unfolded the structures formed in the
human telomeric DNA sequence under study. An unfolding event was identified as the sudden change in the end-to-end distance during the process. Single tether was confirmed by the extension at 31 pN, where the extension the contour length of DNA construct are equal, and plateau at ~65 pN or a single breakage event when the DNA was overstretched. For the kinetic measurements using the force-jump assay, the “jump-up” or “jump-down” of the force was achieved by a sudden movement of a steerable mirror within 100 ms time resolution. The force-extension ($F$-$X$) curves were recorded at 1000 Hz using a LabView® program. These raw data were filtered with a Savitzky-Golay function with a time constant of 10 ms (unless specified otherwise) using a Matlab program.

**Figure 2.4:** Typical force-extension ($F$-$X$) curves observed for the unfolding and refolding of the structures formed in the human telomeric DNA oligomers. (a) Many individual curves (gray) were overlapped and fitted with a sequential WLC model. Red
curve depicts the fitting of the stretching F-X curve whereas green curve represents the fitting of the relaxing F-X curve. (b) Plot of change-in-contour-length ($\Delta L$) vs force (red circles). $\Delta L$ at a particular force was converted from the difference in extension between the red and the green F-X traces in a (see text). The black curve depicts the Gaussian fitting for the histogram of $\Delta L$ (red circles) and pink arrowhead in (a) and (b) shows the unfolding event.

2.5.5 Measurement of Unfolding Force and Change in Contour Length

The unfolding force of the structure was directly measured from the $F$-$X$ curves (Figure 2.4). The change in extension ($\Delta x$) at a particular force ($F$) was calculated as the extension difference between the stretching and the relaxing traces at that force. The resulting $\Delta x$ at this force was then converted to the change in contour length ($\Delta L$) using the Worm-Like-Chain (WLC) model\textsuperscript{205} (equation 2.1):

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

where $x$ is the end-to-end distance (or extension) between the two optically trapped beads, $L$ is the contour length, $k_B$ is the Boltzmann constant, $T$ is absolute temperature, $P$ is the persistent length (51.95 nm),\textsuperscript{205} and $S$ is the stretching modulus (1226 pN).\textsuperscript{205} At
least 400 different molecules were used for the typical unfolding and refolding measurements. Whenever applicable, multiple $F$-$X$ curves from the same molecule were also recorded.

2.5.6 Statistical Analysis of Observed Change in Contour Length

The kernel density treatment and bootstrapping analysis on the $\Delta L$ were performed as described in the literature.\textsuperscript{160} Briefly, each $\Delta L$ transition was expanded by a Gaussian kernel\textsuperscript{206} with its width determined from the average of the standard errors in the regions flanking the transition. A probability histogram was obtained after combining Gaussian kernel distributions from all transitions. From each kernel density plot, Gaussian peaks were identified by Igor (WaveMatrics, Portland, OR) program. A total of 3,000 random re-sampling was performed to construct a bootstrapping histogram of the selected peaks. When more than one population was observed in the kernel density distribution, the folding probability of each population was estimated from the area under the curve fitted with a Gaussian. The folding probability of each population in the bootstrapping histogram was normalized to that determined in the kernel density distribution. Details of the approaches to justify the bootstrapping analysis and population deconvolution using these methods are described in Chapter IV.
2.5.7 Calculation of Expected Change in Contour Length

If \( x \) is the distance between the two termini of a folded structure, the expected change in contour length (\( \Delta L \)) as a result of its unfolding can be calculated with the following equation,

\[
\Delta L = N \times L_{nt} - x \quad \text{(2.2)}
\]

Where \( N \) and \( L_{nt} \) are the number of nucleotides involved in the structure and the contour length per nucleotide, respectively.

In the K\(^+\) buffer, expected \( \Delta L \) of the G-quadruplex structure in the hTelo-4 sequence was calculated by using \( x = 1.1 \) (± 0.1) and \( L_{nt} = 0.48 \) nm.\(^{161}\) The value of \( x \) was taken from the reported NMR structures of hybrid-1 G-quadruplex structure (PDB codes 2HY9,\(^{75}\) 2GKU,\(^{72}\) and 2JPZ\(^{75}\)). For the hybrid-1 G-quadruplex structure involving 21 nucleotides (\( N = 21 \)), equation 2.2 gave an expected \( \Delta L = 9.0 \) (± 0.1) nm. For the G-quadruplex and the misfolded G-quadruplex structures in the hTelo-5, hTelo-6, and hTelo-7 sequences, we have used the same \( x \) value and calculated the expected \( \Delta L \) according to equation 2.2. Furthermore, we measured the distance between the first guanine from the 5’ end to the last guanine in the third stem of the hybrid-1 G-quadruplex to obtain \( x = 2.0 \) (± 0.1) nm. With this value and \( N = 15 \) nts, the expected \( \Delta L \) from
equation 2.2 has been calculated to be $5.2 \pm 0.1$ nm, which is consistent with the partially folded G-triplex species.

For the human telomeric structures in Na$^+$ buffer, the expected $\Delta L$ was calculated by using $x = 1.8 \pm 0.1$ nm obtained from the NMR structure of a basket type quadruplex (PDB 143D)$^{69}$. With $L_{nt} = 0.44$ nm and $N = 21$ nucleotides, equation 2.2 gave $\Delta L = 7.4 \pm 0.1$ nm. To calculate $\Delta L$ of the simulated G-triplex, the end-end distance from the simulated structure was used ($1.5 \pm 0.1$ nm).

2.5.8 Determination of Absolute Percentage Formation

Wherever applicable, the absolute percentage formation of the folded species was calculated as the ratio of the number of the pulling curves with unfolding events to the total pulling curves recorded. To avoid repetitive counting, the subsequent pulling curves from the same DNA construct were discarded in this calculation.

2.5.9 Deconvolution of Multiple Populations

When multiple populations were observed in the $\Delta L$ or unfolding force histograms, the populations were deconvoluted based on the Gaussian probability of the species. To assign individual pulling curves to a specific species, the overall unfolding force histogram was fit using a multi-peak Gaussian function. To account for the
stochastic behavior of individual unfolding events, the pulling curves in the intersection of the two or more Gaussian populations were randomly assigned to a specific group according to the ratio determined by the multi-peak Gaussian fitting. Uncertainty of such a deconvolution rises for populations with significant overlap, but it applies well to distinct populations. For example, the fractions of the ligand-bound and ligand-free populations were calculated from the area under each population in the two-peak Gaussian fitting. Such deconvolution of populations was dramatically improved by using statistical approaches (see Section 2.5.6).

2.5.10 Calculation of Change in Free Energy of Unfolding

The free energy difference for the unfolding of a secondary structure ($\Delta G_{\text{unfold}}$) was calculated according to Jarzynski’s equality equation (equation 2.3) for the non-equilibrium systems,\textsuperscript{207, 208}

$$
\Delta G_{\text{unfold}} = -k_B T \ln \sum_{i=1}^N \frac{1}{N} \exp\left(-\frac{W_i}{k_B T}\right) \quad \text{(2.3)}
$$

where $N$ is the number of observations in the experiment and $W$ is the non-equilibrium work done during unfolding of the structure, which is equivalent of the hysteresis area between stretching and relaxing force-extension ($F$-$X$) curves.\textsuperscript{209} In the case where two populations were observed in $\Delta L$ or unfolding force histograms, the $\Delta G_{\text{unfold}}$ was
calculated, using equation 2.3, for each population after the population deconvolution as described above. The bias of the $\Delta G_{\text{unfold}}$ was calculated from the unfolding work histograms as described in the literature.\textsuperscript{210}

### 2.5.11 Calculation of Refolding Probability

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

### 2.5.12 Analytical Solutions to Kinetics Models

The analytical solutions for kinetics models were obtained by Laplace Transform method. Although the classical approaches has been commonly used to solve two- or three-state kinetics models, Laplace Transform is a general and convenient approach for solving complicated kinetic equations.\textsuperscript{211}
2.5.12.1 Two States Kinetics Model

The rate equation for unfolded \((U)\) or folded \((F)\) species at time \(t\) can be expressed as,

\[
\frac{d[U]}{dt} = k_1[U]_t - k_2[F]_t
\]

\[
\frac{d[F]}{dt} = k_2[F]_t - k_1[U]_t
\]

where \([U]_t\) and \([F]_t\) are the concentrations of the unfolded and folded forms, respectively. Considering that under the initial condition \((t = 0)\), \([U] = [U]_0\) and \([F] = 0\), the Laplace transformation of the above equations yields,

\[
(s + k_1)\mathcal{L}[U]_t - k_2\mathcal{L}[F]_t = [U]_0 \quad \cdots \quad \cdots \quad (2.4)
\]

\[
-k_1\mathcal{L}[U]_t + (s + k_2)\mathcal{L}[F]_t = 0 \quad \cdots \quad \cdots \quad (2.5)
\]

Solving these two equations using Cramer’s rule for linear equations,

\[
\mathcal{L}[U]_t = [U]_0 \left[ \frac{1}{s + (k_1 + k_2)} + \frac{k_2}{s + s(k_1 + k_2)} \right] \quad \cdots \quad \cdots \quad (2.6)
\]
\[ \mathcal{L}[F]_t = [U]_0 \left( \frac{k_1}{s + (k_1 + k_2)} \right) \] \hspace{1cm} (2.7)

Performing inverse Laplace transformation, the kinetic equations for unfolded and folded species are given as,

\[ [U]_t = \frac{[U]_0}{k_1 + k_2} \left[ k_2 + k_1 e^{-(k_1 + k_2)t} \right] \] \hspace{1cm} (2.8)

\[ [F]_t = \frac{k_1 [U]_0}{k_1 + k_2} \left[ 1 - e^{-(k_1 + k_2)t} \right] \] \hspace{1cm} (2.9)

The probability of unfolded and folded species at time \( t \) can be expressed as,

\[ P_{U,t} = \frac{[U]_t}{[U]_t + [F]_t} \]

\[ P_{F,t} = \frac{[F]_t}{[U]_t + [F]_t} \]

Since \( [U]_t + [F]_t = [U]_0 \), equations 2.8 and 2.9 become,

\[ P_{U,t} = \frac{1}{k_1 + k_2} \left[ k_2 + k_1 e^{-(k_1 + k_2)t} \right] \] \hspace{1cm} (2.10)

\[ P_{F,t} = \frac{k_1}{k_1 + k_2} \left[ 1 - e^{-(k_1 + k_2)t} \right] \] \hspace{1cm} (2.11)

The equations 2.10 and 2.11 were used to fit the experimental data of the probability of unfolded and folded species, respectively, over different incubation times.
2.5.12.2 Three States Triangular Kinetics Model

In this model, the rate equations for the formation of each species at time \( t \) can be written as,

\[
\frac{d[U]}{dt} = k_4[T]_t + k_2[Q]_t - (k_1 + k_5)[U]_t
\]

\[
\frac{d[T]}{dt} = k_3[U]_t + k_6[Q]_t - (k_4 + k_5)[T]_t
\]

\[
\frac{d[Q]}{dt} = k_1[U]_t + k_5[T]_t - (k_2 + k_6)[Q]_t
\]

where \([U]_t\), \([T]_t\), and \([Q]_t\) are the concentrations of the unfolded, G-triplex, and G-quadruplex species at time \( t \), respectively.

Using the boundary condition, \([U] = [U]_0\), \([T] = 0\), and \([Q] = 0\) at \( t = 0\), the expressions for the unfolded, the triplex, and the quadruplex species can be obtained as follows.

\[
[U]_t = [U]_0 \left[ \frac{\beta}{\gamma_1\gamma_2} + \frac{(\alpha\gamma_1^2 - \gamma_1^2 - \alpha) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{(\gamma_2^2 - \alpha\gamma_2 + \beta) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \ldots \ldots \ldots \ldots (2.12)
\]
\[ [T]_t = [U]_0 \left[ \frac{\varepsilon}{\gamma_1 \gamma_2} + \frac{(k_3 \gamma_1 - \varepsilon) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{(\beta - k_3 \gamma_2) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \quad (2.13) \]

\[ [Q]_t = [U]_0 \left[ \frac{\delta}{\gamma_1 \gamma_2} + \frac{(k_3 \gamma_1 - \delta) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{(\delta - k_3 \gamma_2) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \quad (2.14) \]

where \( \alpha = k_2 + k_4 + k_5 + k_6 \quad \beta = k_2 k_4 + k_2 k_5 + k_4 k_6 \)

\( \varepsilon = k_1 k_6 + k_2 k_3 + k_3 k_6 \quad \delta = k_1 k_4 + k_1 k_5 + k_3 k_5 \)

\( \gamma_1 \gamma_2 = \beta + \varepsilon + \delta \quad \gamma_1 + \gamma_2 = \alpha + k_1 + k_3 \)

The probability for each species at time \( t \) can be expressed as,

\[ P_{U,t} = \frac{[U]_t}{[U]_t + [T]_t + [Q]_t} \]

\[ P_{T,t} = \frac{[T]_t}{[U]_t + [T]_t + [Q]_t} \]

\[ P_{Q,t} = \frac{[Q]_t}{[U]_t + [T]_t + [Q]_t} \]

Since \( [U]_t + [T]_t + [Q]_t = [U]_0 \), equations 2.12, 2.13 and 2.14 become,

\[ P_{U,t} = \left[ \frac{\beta}{\gamma_1 \gamma_2} + \frac{(a \gamma_1 - \gamma_1^2 - \beta) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{y_2^2 - a y_2 + \beta) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \quad (2.15) \]

\[ P_{T,t} = \left[ \frac{\varepsilon}{\gamma_1 \gamma_2} + \frac{(k_3 \gamma_1 - \varepsilon) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{(\beta - k_3 \gamma_2) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \quad (2.16) \]

\[ P_{Q,t} = \left[ \frac{\delta}{\gamma_1 \gamma_2} + \frac{(k_3 \gamma_1 - \delta) e^{-\gamma_1 t}}{\gamma_2 - \gamma_1} + \frac{(\delta - k_3 \gamma_2) e^{-\gamma_2 t}}{\gamma_2 - \gamma_1} \right] \quad (2.17) \]
The equations 2.15, 2.16 and 2.17 were used to fit the probabilities of the unfolded, the G-triplex, and the G-quadruplex species, respectively, at different incubation time.

The analytical solutions for the linear three-state model are the same set of the equations (2.15-2.17) with $k_1$ and $k_2$ values set to be zero.

2.5.12.3 Four States Kinetics Model

In this model, it is assumed that there is no direct conversion between misfolded species and the G-quadruplex. As depicted by the above model, the formation rate equations for each species at time $t$ can be written as,

\[
\frac{d[U]}{dt} = k_4[T]_t + k_2[Q]_t + k_8[M]_t - (k_1 + k_3 + k_7)[U]_t
\]

\[
\frac{d[T]}{dt} = k_3[U]_t + k_6[Q]_t + k_9[M]_t - (k_4 + k_5 + k_{10})[T]_t
\]

\[
\frac{d[Q]}{dt} = k_1[U]_t + k_5[T]_t - (k_2 + k_6)[Q]_t
\]
\[
\frac{d[M]}{dt} = k_7[U]_t + k_{10}[T]_t - (k_8 + k_9)[M]_t
\]

where \([U]_t\), \([T]_t\), \([Q]_t\), and \([M]_t\) are the concentrations of the unfolded, G-triplex, G-quadruplex, and the misfolded species at time \(t\), respectively.

Considering that under the initial condition \((t = 0)\), \([U]_0\) is the initial concentration of unfolded population and \([T]_0 = 0\), \([Q]_0 = 0\), and \([M]_0 = 0\), the expressions for the unfolded, the triplex, the quadruplex, and the misfolded species are shown as follows.

\[
[U]_t = [U]_0 \left[ \frac{\eta}{y_1y_2y_3} + \frac{(y_1^3 - \alpha y_1^2 + \beta y_1 - \eta) e^{-\gamma_1 t}}{y_1(y_1 - y_2)(y_1 - y_3)} + \frac{(-\gamma_2^3 + \alpha y_2^2 - \beta y_2 + \eta) e^{-\gamma_2 t}}{y_2(y_1 - y_2)(y_2 - y_3)} + \frac{(-\gamma_3^3 + \alpha y_3^2 - \beta y_3 + \eta) e^{-\gamma_3 t}}{y_3(y_1 - y_3)(y_3 - y_2)} \right] \tag{2.18}
\]

\[
[T]_t = [U]_0 \left[ \frac{\theta}{y_1y_2y_3} + \frac{(-k_3 y_1^2 + \varepsilon y_1 - \theta) e^{-\gamma_5 t}}{y_1(y_1 - y_2)(y_1 - y_3)} + \frac{(k_3 y_2^2 - \varepsilon y_2 + \theta) e^{-\gamma_2 t}}{y_2(y_1 - y_2)(y_2 - y_3)} + \frac{(k_3 y_3^2 - \varepsilon y_3 + \theta) e^{-\gamma_3 t}}{y_3(y_1 - y_3)(y_3 - y_2)} \right] \tag{2.19}
\]

\[
[Q]_t = [U]_0 \left[ \frac{\psi}{y_1y_2y_3} + \frac{(-k_1 y_1^2 + \delta y_1 - \psi) e^{-\gamma_1 t}}{y_1(y_1 - y_2)(y_1 - y_3)} + \frac{(k_1 y_2^2 - \delta y_2 + \psi) e^{-\gamma_2 t}}{y_2(y_1 - y_2)(y_2 - y_3)} + \frac{(-k_1 y_3^2 - \delta y_3 + \psi) e^{-\gamma_3 t}}{y_3(y_1 - y_3)(y_3 - y_2)} \right] \tag{2.20}
\]

\[
[M]_t = [U]_0 \left[ \frac{\omega}{y_1y_2y_3} + \frac{(-k_7 y_1^2 + \phi y_1 - \omega) e^{-\gamma_1 t}}{y_1(y_1 - y_2)(y_1 - y_3)} + \frac{(k_7 y_2^2 - \phi y_2 + \omega) e^{-\gamma_2 t}}{y_2(y_1 - y_2)(y_2 - y_3)} + \frac{(-k_7 y_3^2 + \phi y_3 + \omega) e^{-\gamma_3 t}}{y_3(y_1 - y_3)(y_3 - y_2)} \right] \tag{2.21}
\]
where $\alpha = k_2 + k_4 + k_5 + k_6 + k_8 + k_9 + k_{10}$

\[ \beta = k_2k_4 + k_2k_5 + k_4k_6 + k_5k_8 + k_6k_8 + k_2k_9 + k_4k_9 + k_5k_9 + k_6k_9 + k_{10} + k_9k_{10} \]

$\eta = k_2k_4k_8 + k_2k_5k_8 + k_4k_6k_8 + k_2k_4k_9 + k_2k_5k_9 + k_4k_6k_9 + k_6k_8k_{10} + k_2k_8k_{10}$

$\varepsilon = k_1k_6 + k_2k_3 + k_3k_6 + k_3k_8 + k_3k_9 + k_7k_9$

$\theta = k_1k_6k_8 + k_2k_3k_8 + k_3k_6k_8 + k_1k_6k_9 + k_2k_3k_9 + k_3k_6k_9 + k_6k_7k_9 + k_2k_7k_9$

$\delta = k_1k_4 + k_1k_5 + k_3k_5 + k_1k_8 + k_1k_9 + k_{10}$

$\psi = k_1k_4k_8 + k_1k_5k_8 + k_3k_5k_8 + k_1k_4k_9 + k_1k_5k_9 + k_3k_5k_9 + k_5k_7k_9 + k_1k_8k_{10}$

$\phi = k_4k_7 + k_5k_7 + k_6k_7 + k_2k_7 + k_3k_{10} + k_7k_{10}$

$\omega = k_2k_4k_7 + k_2k_5k_7 + k_4k_6k_7 + k_2k_7k_{10} + k_3k_6k_{10} + k_1k_6k_{10} + k_6k_7k_{10} + k_2k_3k_{10}$

$\gamma_1\gamma_2 + \gamma_2\gamma_3 + \gamma_1\gamma_3 = \beta + \varepsilon + \delta + \phi$ \hspace{1cm} $\gamma_1 + \gamma_2 + \gamma_3 = \alpha + k_1 + k_3 + k_7$

$\gamma_1\gamma_2\gamma_3 = \eta + \theta + \psi + \omega$

Since $[U]_t + [T]_t + [Q]_t + [M]_t = [U]_0$, final expressions for the probabilities of the species at time $t$ is given as,
The equations 2.22, 2.23, 2.24 and 2.25 were used to fit the respective probabilities of unfolded, G-triplex, G-quadruplex, and misfolded species over incubation time.
To determine the folding rate constant, $k_{\text{fold}}$, from the kinetics measurements in the force-jump experiments, a single exponential equation, $P = A - B \times \exp(-k_{\text{fold}} \times t)$, was used to fit the experimental data of refolding probability of the various species at different incubation time. Here, $P$ represents the folding probability, $t$ is the incubation time, and $A$ and $B$ are fitting constants. A similar single exponential equation ($R = A - B \times \exp(-t/\tau)$) was used to fit the experimental data of the amount of ligand bound to the G-quadruplex species at different incubation time. Here, $R$ is the ratio of the ligand bound G-quadruplex over free G-quadruplex, $t$ is the incubation time, $\tau$ is the time constant, and $A$ and $B$ are fitting constants. To determine the dissociation constant ($K_d$), the model of ligand binding, $\Phi = \frac{[L]}{K_d + [L]}$ was used, where $\Phi$ is the fraction of the G-quadruplex tightly bound with a ligand and $[L]$ is the concentration of the ligand.
Chapter III

Intramolecular Folding in Three and Four Tandem G-Rich Repeats of Human Telomeric DNA: Direct Observation of G-triplex Species


3.1 Abstract

Intramolecular folding in three and four tandem guanine repeats of human telomeric DNA have been investigated using optical-tweezers, molecular dynamics simulation and circular dichroism. A mechanically and thermodynamically stable folded species has been observed in the sequence with three tandem guanine repeats which is consistent with a G-triplex conformation. Similar species has also been found to coexist with a prevalent G-quadruplex structure in a DNA sequence with four tandem guanine
repeats. Such observations suggest a complex folding pattern of the human telomeric DNA in which G-triplex could be an intermediate structure to the G-quadruplex.

3.2 Overview

Telomeres of eukaryotic chromosomes play a vital role in maintaining gene integrity and cell growth.\(^{64}\) In normal cells, telomere is shortened each time a chromosome is replicated during cell division, eventually leading to apoptosis.\(^{212}\) However, majority of cancer cells maintain the length of the telomere due to enhanced activity of telomerase.\(^{213}\) Human telomeres have the propensity to form intramolecular G-quadruplex structures in the 3’ overhang of the region.\(^{68}\) A G-quadruplex is a tetraplex structure which is composed of a stack of G-quartets,\(^{32}\) each of which contains a quadrilateral of four guanines held together by Hoogsteen base pairs and stabilized by monovalent cations such as K\(^+\) or Na\(^+\).\(^{34}\) It has been observed that small-molecules such as telomestatin,\(^{90}\) quarfloxin,\(^{214}\) and pyridostatin\(^{198}, 215\) effectively stabilize the G-quadruplex structures and reduce the telomerase activity. This implied a potential regulatory function for G-quadruplex to maintain the length of telomeres.

The structural diversity has been well recorded for G-quadruplexes in four tandem guanine repeats of the human telomeric DNA.\(^{37, 186, 216}\) However, the possibility of forming alternative structures in such sequences has not been explored well. Single-molecule investigations by Balasubramanian\(^{217}\) and Ha\(^{181}\) have revealed unknown species
other than G-quadruplexes in the human telomeric DNA. Independent studies by Yang et al., Sugiyama et al., and Chaires et al. have proposed that these structures could be G-quadruplex intermediates that employ only three tandem guanine repeats to assume a triplex structure. NMR structure of a 3+1 bimolecular G-quadruplex has supported these proposals. Nevertheless, stand-alone G-triplex structures have not been observed directly in three tandem guanine repeats.
Figure 3.1: CD analysis of the hTelo-3 sequence in a 10 mM Tris buffer, pH 7.4 containing 100 mM NaCl. (a) CD spectra of the hTelo-3 (solid curve) and hTelo-4 (dotted curve). (b) CD spectra of the hTelo-3 at 25-85 °C. (c) Intensity of the 257 nm peak with temperature. The dotted curve is the sigmoidal fitting. (d) Melting temperature ($T_m$) of 3-30 µM hTelo-3. The dotted line is a linear fitting for guidance. Errors depict the standard deviation (SD) from three independent measurements.

Using optical tweezers based single-molecule assay in assistance with molecular dynamics simulation and circular dichroism, here, we have investigated the folded structures formed in the human telomeric sequences with three G-rich repeats, TTA (GGG TTA)$_3$ (hTelo-3), and four G-rich repeats, TTA (GGG TTA)$_4$ (hTelo-4). We have observed mechanically and thermodynamically stable species different from the G-quadruplex in both sequences. In addition, we have demonstrated that such intramolecularly folded species in the hTelo-3 or hTelo-4 are consistent with the triplex conformation. These findings depict a complex folding pattern in human telomeres. The mechanical stabilities of these structures revealed by optical tweezers are of unique importance to elaborate their potential interactions with telomerase, a motor protein that generates a load force.

3.3 Results and Discussion
First, we used CD to investigate the folded structures formed in the hTelo-3 and hTelo-4 in a 10 mM Tris buffer (pH 7.4) containing 100 mM Na\(^+\). Selection of buffer was based on the fact that hTelo-4 in this buffer is known to form a single conformation of the G-quadruplex, which reduces the complexity of the system arising from structural polymorphism. As shown in Figure 3.1a, CD spectrum of the hTelo-4 has two peaks at 293 and 250 nm with a valley at 271 nm, which is consistent with previous findings that human telomeric sequence prefers a basket type G-quadruplex in this buffer.\(^{106}\) The CD spectrum of the hTelo-3, however, shows dramatically different features with a pronounced peak at 257 nm and a valley at 238 nm. In addition, a small valley and a peak were observed at 280 nm and 295 nm respectively. These CD features of hTelo-3 are consistent with previous observations on similar sequences that suggested the presence of intramolecularly folded structures other than G-quadruplex.\(^{221, 222}\) When temperature was raised, we observed the dissolution of all these features except the 238 nm valley. At the same time, a broad peak of 275 nm appeared (Figure 3.1b), which is consistent with the spectrum of unstructured ssDNA.\(^{223}\) A plot of the 257 nm peak intensity vs temperature reveals a melting transition with \(T_m\) of 55 ± 1°C at 10 µM sample concentration (Figure 3.1c). This result indicated the formation of a stable structure in the hTelo-3. As demonstrated in Figure 3.1d, \(T_m\) of 3-30 µM of hTelo-3 is constant, indicating that the folded structure in the hTelo-3 is indeed of intramolecular, not intermolecular, nature in this concentration range.
Figure 3.2: Mechanical unfolding of the hTel-3 in the 100 mM Na$^+$ buffer. (a) Schematic of the single-molecule mechanical unfolding assay (not to scale). (b) A typical force-extension (F-X) curve for the hTel-3. The red and green arrows depict the stretching and relaxing F-X curves, respectively. The blue arrowhead shows the unfolding event. Change in contour length ($\Delta L$) (c) and unfolding force (d) histograms for structures in the hTel-3. The solid curves in (c) & (d) are Gaussian fittings.

To confirm the intramolecular nature of the folded species in hTel-3 and to reveal its mechanical and thermodynamic properties, an optical tweezers based single-molecule assay was performed in the same buffer. As shown in Figure 3.2a, the structure
in the hTelo-3 was mechanically unfolded when the DNA construct tethered between two optically trapped beads was stretched with a constant load rate of 5.5 pN s\(^{-1}\). A characteristic unfolding event (indicated by the arrowhead in Figure 3.2b) in the force-extension (F-X) curve was observed during the stretching. Because a single molecule was pulled, the observation directly supported the formation of intramolecularly folded structure in the hTelo-3. The change in contour length (\(\Delta L\)) when the structure was mechanically unfolded was found to be 5.0 ± 0.2 nm (Figure 3.2c & Table 3.1). The unfolding force (\(F_{\text{unfold}}\)) for this structure is 32 ± 3 pN (Figure 3.2d & Table 3.1) demonstrating its strong mechanical stability. The change in free energy of unfolding (\(\Delta G_{\text{unfold}}\)) at 0 pN was found to be 6.3 ± 0.4 kcal/mol at 23°C, confirming the structure is stable at room temperature. Since each unfolding experiment involves a single DNA molecule that contains only three G3 tracts, G-quadruplex is not possible to form. A hairpin structure is also ruled out since the \(F_{\text{unfold}}\) observed here is significantly higher than that to unfold a hairpin.\(^{159,202}\) Instead, we surmise a triplex structure\(^{187}\) may form.

To provide evidence for such a structure and to depict molecular arrangements of the nucleotides in the folded species, we performed classical MD simulation using AMBER version 8 and 9 for the MD-GRAPE system. The MD simulation was carried out using the sequence, 5′-A GGG TTA GGG TTA GGG TTA GGG TTA-3′. The initial atomic coordinates of triplexes were obtained from a basket type G-quadruplex (PDB 143D).\(^{69}\) Triplex models were constructed by deletion of the fourth strand from the
antiparallel structure. The triplex sequence is 5′-A\textsuperscript{1}G\textsuperscript{2}G\textsuperscript{3}G\textsuperscript{4}T\textsuperscript{5}T\textsuperscript{6} A\textsuperscript{7}G\textsuperscript{8}G\textsuperscript{9}G\textsuperscript{10}T\textsuperscript{11}T\textsuperscript{12} A\textsuperscript{13}G\textsuperscript{14}G\textsuperscript{15}G\textsuperscript{16}T\textsuperscript{17}-3′. More details regarding the simulation have been described in Chapter II, Section 2.3.

**Figure 3.3:** Results of MD simulation. (a) The distance between O of 5′ end A\textsuperscript{1}G\textsuperscript{2} and O of 3′ end G\textsuperscript{16}T\textsuperscript{17} backbone (sugar and phosphate) over time. (b) Energy minimized average structure of the triplex. The red ribbon traces the backbone, blue ball represents
the central Na\(^+\) ion and the asterisks represent the surrounding Na\(^+\) ions. (c) The simplified schematic structure of the triplex showing anti (red) and syn (blue) guanines.

When we examined the energy minimized average structure (Figure 3.3b) during equilibrium calculation between 2500-3000 ps (Figure 3.3a), the distance between O of 5' end A\(^1\)G\(^2\) and O of 3' end G\(^{16}\)T\(^{17}\) backbone (sugar and phosphate) demonstrates a value of 1.5 ± 0.1 nm. Furthermore, using this value as the end-to-end distance of the folded structure, predicted $\Delta L$ for the unfolding of simulated structure was found to be 5.1 ± 0.1 nm. This value matches very well with that observed in the single-molecule assay (5.0 ± 0.2 nm, Table 3.1). These results support that the folded species in the hTelo-3 assumes a triplex conformation (Figure 3.3b and 3.3c).

Next, we investigated folded species in the hTelo-4 which contains four G3 tracts. At least two species are expected to form in this sequence: a fully folded G-quadruplex involving four G-tracts with 21 nucleotides\(^{69}\) and a partially folded structure involving three G-tracts with 15 nucleotides (similar to the triplex discussed above). When we performed single-molecule mechanical unfolding on the hTelo-4 in the same buffer (Figure 3.4), indeed, both $\Delta L$ and $F_{\text{unfold}}$ histograms (Figure 3.4c & 3.4d) showed two species with a major population (~45%) having $\Delta L$ of 7.8 ± 0.2 nm and a minor one (~15%) with $\Delta L$ of 5.1 ± 0.3 nm. The former species is most likely a G-quadruplex since the measured $\Delta L$ (7.8 ± 0.2 nm) is consistent with the $\Delta L$ (7.4 ± 0.1 nm) obtained from the NMR structure of the expected G-quadruplex in the same buffer (PDB 143D).\(^{69}\) We
assigned the minor population as a triplex species based on the fact that $\Delta L$ of this species was much smaller than G-quadruplex, but matched well with that of the folded species in hTel-3. This was further supported by the $F_{\text{unfold}}$ measurement, which showed two populations centered at $15.3 \pm 0.9$ pN and $33.1 \pm 0.5$ pN (Fig. 4d).

**Figure 3.4:** Mechanical unfolding of the hTel-4 in the 100 mM Na$^+$ buffer. The magnified F-X regions depict unfolding events with $\Delta L$ of (a) $\sim 8$ nm and (b) $\sim 5$ nm. The red and green colors indicate the stretching and relaxing F-X curves, respectively. (c) $\Delta L$ and (d) $F_{\text{unfold}}$ histograms for the species in the Tel-4G. The histograms are fitted with two-peak Gaussian (black curves). The dotted curves show individual Gaussian
populations depicted by the same color codes. (e) $F_{\text{unfold}}$ histogram of the hTelo-3 (gray bars fitted with black Gaussian) and the hTelo-3/hTelo-1 mixture (red bars fitted with red Gaussian).

To correlate $\Delta L$ and $F_{\text{unfold}}$ for a specific population, we performed trace-by-trace analysis on the F-X curves. We found that the major population ($\Delta L = 7.8$ nm) had a $F_{\text{unfold}}$ of 15.3 pN whereas the minor population ($\Delta L = 5.1$ nm) corresponded to a $F_{\text{unfold}}$ of 33.1 pN (Fig. 4c & d). The latter species matches with that of hTelo-3 (32 ± 3 pN). In addition, $\Delta G_{\text{unfold}}$ for the major population, $8.2 \pm 0.1$ kcal/mol, is comparable to telomeric G-quadruplexes, $^{184, 215}$ but higher than the minor population ($6.3 \pm 0.9$ kcal/mol). This is consistent with the expectation that a G-quadruplex (the major population) is more stable than the triplex species (the minor population) probably due to the greater contribution from stronger base stacking and more Hoogsteen hydrogen bonds involved in the former species. It is noteworthy that $F_{\text{unfold}}$ is a measure of the unfolding kinetics, rather than the thermodynamics ($\Delta G_{\text{unfold}}$) of a structure (see Chapter 1 Section 1.4.3 for details).

The species in the hTelo-3 with possible triplex conformation was further tested in the mechanical unfolding of the hTelo-3 in the presence of 5 µM 5'GGGTTA (hTelo-1) fragment. Although $\Delta L$ of the species in the hTelo-3 + hTelo-1 mixture ($5.2 \pm 0.1$ nm) is similar to that of the hTelo-3, both $F_{\text{unfold}}$ (38 ± 0.5 pN) and $\Delta G_{\text{unfold}}$ (7.7 ± 0.5 kcal/mol) are significantly higher than those in the hTelo-3 alone (Figure 3.4e & Table 3.1). These results are consistent with the formation of a more stable bimolecular 3+1 G-
quadruplex with 3 strands contributed from the hTel-0 and one strand from the Tel-1G. Dimethyl Sulfate (DMS) footprinting (Figure 3.5a and 3.5b) further supports the formation of such a structure. That the protection of the guanines in the hTel-0 is much reduced with respect to G-quadruplexes (Figure 3.5a and 3.5b) suggests that either the formation of the G-triplex species is much less compared to the G-quadruplexes, or the hydrogen bondings associated with N7 guanines in G-triplex are weaker than those in G-quadruplex. These facts are consistent with the low abundance of the G-triplexes in our single-molecule measurements and the simulated G-triplex structure described recently.

Table 3.1: $\Delta L, F_{\text{unfold}}$ and $\Delta G_{\text{unfold}}$ (23°C) of human telomeric DNA species in 10 mM Tris buffer, pH 7.4 containing 100 mM Na$^+$. Errors represent the SD from three independent measurements.

<table>
<thead>
<tr>
<th>Folded Species</th>
<th>$\Delta L$ (nm)</th>
<th>$F_{\text{unfold}}$ (pN)</th>
<th>$\Delta G_{\text{unfold}}$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTel-0 (G-triplex)</td>
<td>5.0 ± 0.2</td>
<td>32.0 ± 3.0</td>
<td>6.3 ± 0.4</td>
</tr>
<tr>
<td>hTel-0 (G-quadruplex)</td>
<td>7.8 ± 0.2</td>
<td>15.3 ± 0.9</td>
<td>8.2 ± 0.1</td>
</tr>
<tr>
<td>hTel-0 (G-triplex)</td>
<td>5.1 ± 0.3</td>
<td>33.1 ± 0.5</td>
<td>6.3 ± 0.9</td>
</tr>
<tr>
<td>hTel-0 + Telo-1</td>
<td>5.2 ± 0.1</td>
<td>38.0 ± 0.5</td>
<td>7.7 ± 0.5</td>
</tr>
<tr>
<td>Triplex (Simulation)</td>
<td>5.1 ± 0.1</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>
We have shown that the intramolecularly folded structure in three tandem guanine repeats of the human telomeric DNA in a Na$^+$ buffer is consistent with a triplex conformation. We further extended our studies on the same hTelo-3 and hTelo-4 sequences in a buffer with 100 mM K$^+$ ion, which is a physiologically more relevant condition. Previously, the G-quadruplex species formed in four tandem repeats of the human telomeric DNA sequence in this buffer have been found to be highly polymorphic. Although, it is beyond the scope of this study to probe the detailed structures of the folded species, our investigations have revealed that hTelo-3 and hTelo-4 can form similar triplex-based folded structures also in K$^+$ buffer. The results of CD, single-molecule optical tweezers and DMS footprinting measurements for hTelo-3 and hTelo-4 in K$^+$ buffer have been summarized below in Figure 3.6, Figure 3.7, Figure 3.8 and Table 3.2, respectively.
Figure 3.5: DMS footprinting of hTel3, hTel-3 + hTel-1 mixture, and hTel-4 in the Na+ buffer. (a) The gel image of the footprinting. The content of each lane has been...
labeled below the image. (c) The fold protection of the guanine residues in the hTelo-3, hTelo-3 + hTelo-1 mixture, and hTelo-4 in buffer containing 100 mM Na\(^+\) with respect to the corresponding species in a buffer without Na\(^+\). Errors represent the SD from three independent measurements.

Figure 3.6: CD analysis of 5\(\mu\)M solution of hTelo-3 sequence in a 10 mM Tris buffer, pH 7.4 containing 100 mM KCl. (a) CD spectra of the hTelo-3 (solid curve) and hTelo-4 (dotted curve). The CD signature of the hTelo-4 is consistent with that of hybrid (3+1) G-
quadruplex described previously. The CD spectrum of the hTelo-3 indicates the presence of a folded species different from G-quadruplex. (b) CD spectra of the hTelo-3 at 15-75 °C. (c) Intensity of the 290 nm peak with temperature. The dotted curve is the sigmoidal fitting which provided a melting temperature ($T_m$) of $45 \pm 1.5$ °C. (d) Melting temperature of 5-100 µM hTelo-3. The dotted line is a linear fitting for eye guidance and errors depict the SD from three independent measurements.
Figure 3.7: Mechanical unfolding of the hTelo-3 and hTelo-4 in a K⁺ buffer. Change in contour length (ΔL), (a) and $F_{\text{unfold}}$, (b) histograms (blue sticks) for the structures in hTelo-3. The solid curves are the Gaussian fittings which revealed the $\Delta L = 5.0 \pm 0.2$ nm and $F_{\text{unfold}} = 33.4 \pm 0.8$ pN. (c) The $\Delta L$ and (d) $F_{\text{unfold}}$ histograms (blue sticks) for the species in hTelo-4 sequence. The $\Delta L$ histogram is fitted with two-peak Gaussian (black
curve) centered at $\Delta L = 5.6 \pm 0.1 \text{nm}$ and $\Delta L = 8.9 \pm 0.1 \text{nm}$, respectively. The single peak Gaussian fitting (black curve) for the $F_{\text{unfold}}$ histogram (d) indicates the similar mechanical stability of the both species observed in $\Delta L$ histogram. Deconvolution of these populations (see Chapter II Section 2.5.9 for details) based on the $\Delta L$ histogram showed that the G-triplex and the G-quadruplex species have $F_{\text{unfold}}$ of $21.0 \pm 1.2 \text{pN}$ and $19.8 \pm 3.0 \text{pN}$, respectively. (e) The $\Delta L$ histogram $\text{hTel}_{o73} + \text{hTel}_{o71}$ mixture which showed similar $\Delta L$ ($5.1 \pm 0.2 \text{nm}$) to that of $\text{hTel}_{o73}$ only (a). The $F_{\text{unfold}}$ histogram of the mixture (blue bars fitted with red Gaussian), however, showed two populations. The first population ($F_{\text{unfold}} = 30.0 \pm 2.8 \text{pN}$) is similar to that of $\text{hTel}_{o73}$ only (b) whereas second population ($F_{\text{unfold}} = 44.0 \pm 0.4 \text{pN}$) indicates the formation of inter-molecular $3+1$ G-quadruplex between $\text{hTel}_{o73}$ and $\text{hTel}_{o74}$. Such observations were further supported by the calculation of $\Delta G_{\text{unfold}}$ (see Table 3.2). These single-molecule mechanical unfolding measurements revealed that the folded species with $\Delta L \sim 5 \text{nm}$ in $\text{hTel}_{o73}$ and $\text{hTel}_{o74}$ sequences are similar to the G-triplex observed in $\text{Na}^+$ buffer as discussed above. Reported errors in $\Delta L$ and $F_{\text{unfold}}$ represent the SD obtain from the histograms of three independent measurements.
1. hTelo-4 (100 mM K\(^+\))  
2. hTelo-4 (no salt)  
3. hTelo-3 (100 mM K\(^+\))  
4. hTelo-3 (no salt)  
5. hTelo-3G + 1 (100 mM K\(^+\))  
6. hTelo-3G + 1 (no salt)
**Figure 3.8:** DMS footprinting of hTelo-3, hTelo-3 + hTelo-1 mixture, and hTelo-4 in a 100 mM K\(^+\) containing buffer. (a) The gel image of the footprinting. The content of each lane has been labeled below the image. (c) The fold protection of the guanine residues in the hTelo-3, hTelo-3 + hTelo-1 mixture, and hTelo-4 in buffer containing 100 mM Na\(^+\) with respect to the corresponding species in a buffer without Na\(^+\). Errors depict the SD from three independent measurements.

**Table 3.2:** \(\Delta L\), \(F_{\text{unfold}}\) and \(\Delta G_{\text{unfold}}(23\degree C)\) of human telomeric DNA species in a 10 mM Tris buffer, pH 7.4 with 100 mM K\(^+\). The \(\Delta L\) and the \(F_{\text{unfold}}\) in hTelo-3 and hTelo-3 + hTelo-1 mixture are different from those of partially-folded species in hTelo-4 but similar to that observed in Na\(^+\) buffer. This could possibly due to the complex folding patterns of the G-quadruplexes in K\(^+\) buffer compared to that in Na\(^+\) buffer. Reported errors in \(\Delta L\), \(F_{\text{unfold}}\) and \(\Delta G_{\text{unfold}}(23\degree C)\) represent the SDs obtained from the histograms of three independent measurements.

<table>
<thead>
<tr>
<th>Folded Species</th>
<th>(\Delta L) (nm)</th>
<th>(F_{\text{unfold}}) (pN)</th>
<th>(\Delta G_{\text{unfold}}) (kcal mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTelo-3 (G-triplex)</td>
<td>5.0 ± 0.2</td>
<td>33.4 ± 0.8</td>
<td>6.7 ± 0.5</td>
</tr>
<tr>
<td>hTelo-4 (G-quadruplex)</td>
<td>8.9 ± 0.1</td>
<td>21.0 ± 1.2</td>
<td>9.8 ± 0.4</td>
</tr>
<tr>
<td>hTelo-4 (G-triplex)</td>
<td>5.6 ± 0.1</td>
<td>19.8 ± 3.0</td>
<td>5.9 ± 0.5</td>
</tr>
<tr>
<td>hTelo-3 + Telo-1</td>
<td>5.1 ± 0.2</td>
<td>44.0 ± 0.4</td>
<td>7.9 ± 0.5</td>
</tr>
</tbody>
</table>
3.4 Conclusions

In summary, we have confirmed the existence of folded intramolecular structures other than the G-quadruplex in human telomeric sequences under physiologically relevant conditions of Na\(^+\) and K\(^+\). Molecular dynamics simulation, single-molecule optical tweezers, and bulk methods such as CD and DMS footprinting suggest that the folded structure in hTelο3 and the partially folded structure in hTelο-4 assume a triplex conformation. These results reveal an unprecedented and diverse folding pattern in the human telomere region, which is instrumental for the development of new telomere-targeting small-molecule drugs.

This project was financially supported by the NSF CHE-1026532, the New Faculty Award Program at Camille and Henry Dreyfus Foundation, the Ohio Board of Regents to Hanbin Mao and the Core Research for Evolutional Science and Technology (CREST) of JST and a Grant-in-Aid for Science research from the MEXT, Japan to Hiroshi Sugiyama.
Chapter IV


4.1 Abstract

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

4.2 Overview
In human cells, telomeres at the end of chromosomes consist of single-stranded 3' overhang of ~200 nucleotides with a consensus G-rich repeat sequence, 5'-TTAGGG.\textsuperscript{60, 61, 180, 224} Four such G-rich repeats are known to form a stable DNA secondary structure, G-quadruplex.\textsuperscript{38, 63} A G-quadruplex is composed of a stack of G-quartets, each of which is held together by four guanines through Hoogsteen hydrogen bonds and further stabilized by intercalating cations such as K\textsuperscript{+} or Na\textsuperscript{+}.\textsuperscript{34, 35} Biological investigations suggest that these telomeric DNA secondary structures can regulate the length of telomere either by interfering with telomerase activity or by participating in events such as uncapping of telosomes.\textsuperscript{28, 225, 226} Since telomere length is closely associated with cellular processes that lead to senescence or cancer, telomeric G-quadruplexes become an attractive target for cancer treatment.\textsuperscript{15, 65}
Figure 4.1: Schematic structures of the regular G-quadruplex (a), G-triplex (b) and misfolded G-quadruplex (c) expected to form in htelo-5 sequence. Shown in d is the experimental setup (not drawn to scale). A DNA construct containing a single stranded human telomeric sequence (hTel0) is attached to streptavidin and anti-digoxigenin coated beads trapped by laser tweezers. Stretching and relaxing of the tethered DNA molecule allow unfolding and refolding of the structures in the sequence of interest.

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

However, ensemble-average techniques, such as Circular Dichroism (CD), NMR, or X-ray crystallography, have difficulties to deconvolute individual species, especially those with insignificant population fractions or short life times, formed in the same biological molecule. To resolve a structure in such a population mixture, mutations in the biological molecule are often required to selectively populate the species of interest. Recently, such a practice surprisingly revealed G-quadruplex conformations that harbor (TTAGGGT)\textsubscript{n} in one of the loops (Figure 4.1c, long-loop G-quadruplex or misfolded G-quadruplex) in human telomeric DNA fragments with more than four TTAGGG repeats. This procedure, however, changes population equilibrium and distorts the transition between different species. Due to these difficulties, the population equilibrium and the transition kinetics among different species formed in the longer human telomeric sequences are still unclear.
Figure 4.2: Illustration of force-jump method and associated data analysis. (a) The force vs time trace in a force-jump experiment. Force is brought to 0 pN in 100 ms (green dots) after unfolding of the structure in the DNA tether. After incubation (0-60s), force is jumped to 7 pN (red dots) to start subsequent unfolding. (b) A typical force extension (F-X) curve for the force-jump experiment. Stretching is shown in red while relaxing in green. c) Plot of change-in-contour-length ($\Delta L$) vs force (filled red circles). $\Delta L$ at a particular force is converted from the difference in extension between the red and the green F-X traces in b (see text). Purple curves represent Gaussian fitting.
Capable of probing one molecule at a time, single-molecule methods offer an unprecedented opportunity to identify conformation and follow transition kinetics of individual species in a population mixture. Assisted with statistical analyses on structures with different change-in-contour-length ($\Delta L$) upon unfolding,$^{160}$ we were able to distinguish populations with $\Delta L$ up to tens of nanometers at a baseline resolution of 0.5 nanometer.$^{139}$ This method significantly expands the measurement range with respect to single-molecule FRET, which works in the range of 2-9 nm.$^{166}$ With this approach, here, we reveal a surprisingly complex population network that contains G-quadruplexes (Figure 4.1a), G-triplexes (Figure 4.1b), and misfolded long-loop G-quadruplexes (Figure 4.1c) in human telomeric sequences with four to eight TTAGGG repeats. By a force-jump method with a temporal resolution of 100 ms,$^{215}$ the transition kinetics among different species is found to follow a 3-state or a 4-state model in the telomeric sequences with 4 or 5 TTAGGG repeats, respectively. Our results reveal that G-quadruplexes are the most prevalent species and that long-loop G-quadruplexes are misfolded population minorities whose formation and disassembly are faster than G-quadruplexes or G-triplexes. Population analyses further indicate that misfolded G-quadruplexes in a specific telomeric sequence decrease in abundance with increasing loop sizes or number of long-loops. These results lead us to predict that G-quadruplexes with shortest TTA loops would be the most stable population in the full-length 3’ overhang of human telomeres. That DNA can form misfolded conformation rivals proteins or RNA in the complexity of structure and transition kinetics.
4.3 Results and Discussion

4.3.1 Force-Jump Method and Population Deconvolution at Nanometer Resolution (PoDNano)
**Figure 4.3:** PoDNano analysis of the population distribution in the hTel-4 sequence. a) The $\Delta L$ histogram obtained from the point-to-point measurements of the structures unfolded in the hTel-4 sequence. b) Kernel density distribution obtained from the trace-by-trace analyses of the same data in A. Populations are fit with Gaussian functions. c) Bootstrapping analyses deconvolute two populations with a baseline resolution of 1.1 nm.

Our single-molecule assay (Figure 4.1d) allows mechanical unfolding and refolding of a DNA structure. In a typical force-jump experiment, a DNA construct is extended until the structure formed within is unfolded by the tension accumulated in the construct (red curves in Figure 4.2a and 4.2b). The unfolding is indicated by a sudden change in the extension or the force in the force-extension ($F$-$\lambda$) curve. Once unfolded, the tension in the DNA relaxes to 0 pN in 100 ms (green curves in Figure 4.2a and 4.2b) to allow the refolding of the structure in a specific time (the “force-pump” procedure). The refolding is evaluated by next round of force-ramping starting from the 7 pN by a second force-jump (Figure 4.2a) (the “force-probe” step). If a structure is refolded during the force-pump step, unfolding event will be observed in the “force-probe” step. The purpose of the two force-jumps is to avoid refolding during the slow processes of force ramping at entropic force regions (<7 pN).\textsuperscript{215} Mimicry of temperature jumps, this force-pump and probe (or FP2) process\textsuperscript{215} has demonstrated its capability in the kinetic measurement in which force can flexibly serve to denature or refold a biological macromolecule dependent on the force level.\textsuperscript{215}
Recently, we applied an integrated Population Deconvolution at Nanometer resolution (PoDNano) to investigate G-quadruplex structures in the hTERT promoter sequences.\textsuperscript{160} This method first calculates the change in end-to-end distance ($\Delta x$) between the relaxing (green) and the stretching $F$-$X$ curves (red curve in Figure 4.2b). The $\Delta x$ is then converted to the change-in-contour-length ($\Delta L$) by the worm like chain (WLC) model\textsuperscript{205} in a particular force range (Figure 4.2c, see Chapter II Section 2.5.5). For each $\Delta L$-$F$ plot, a $\Delta L$ histogram is constructed and the $\Delta L$ of the folded structure can be retrieved against the background ($\Delta L=0$, Figure 4.2c). Further deconvolution using kernel density distribution and bootstrapping analyses (see Chapter II Section 2.5.6 ) allows a spatial resolution of sub-nanometers for $\Delta L$ determination.\textsuperscript{139} Compared to previous measurements in which each $\Delta L$ is estimated from the two data points flanking the unfolding transition (Figure 4.3a shows a $\Delta L$ histogram from this measurement),\textsuperscript{157} this method is built on a Gaussian kernel density function that exploits much expanded data sets flanking the transition event. This provides a better signal to noise ratio (Figure 4.3b shows a density distribution of $\Delta L$ using a Gaussian kernel function). The resolution is further improved by bootstrapping analyses which filter out noises by selecting the most probable values from the random resampling of Gaussian kernel distributions (Figure 4.3c). Comparison of the peak centers of the major population identified in Figure 2 revealed that these three methods yielded identical results within experimental errors ($9.1 \pm 0.1$ nm, $8.9 \pm 0.2$ nm, and $8.8 \pm 0.1$ nm, respectively, for panel a, b, and c in Figure 4.3). In addition, the PoDNano produced $\Delta L$ values ($8.8 \pm 0.1$ nm) identical to
those obtained from the sequential WLC model fitting\textsuperscript{204} method (9.4 ± 1.2 nm, see Chapter II Section 2.2.5).

\textbf{Figure 4.4:} CD measurements of various human telomeric sequences in a 10 mM Tris buffer that contains 100 mM KCl or LiCl. CD spectrum of each sample in K\textsuperscript{+} buffer shows peaks at ~268 nm and ~290 nm with a valley at ~238 nm. These features are characteristic of a hybrid type G-quadruplex (GQ) as reported previously.\textsuperscript{231} Since Li\textsuperscript{+} is known to disfavor the GQ formation, the CD spectra of hTelo-4 in Li\textsuperscript{+} buffer serve as a control. The results indicate that GQ is the most dominant species in each DNA fragment which is consistent with our single-molecule studies using optical tweezers. It is noteworthy that hTelo-5-Mut, which is potential to form a misfolded-GQ with an internal long-loop with TTA AAA TTA sequence, showed a CD spectrum similar to that of regular-GQ suggesting the formation of GQ associated species.

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

4.3.2 Transition Kinetics of Populations in hTelo-4 Sequence

With the PoDNano approach, first, we analyzed folded structures in the hTelo-4 sequence, (TTA GGG)$_4$TTA, in a 10 mM Tris buffer with 100 mM KCl. The probability distribution based on the Gaussian kernel (Figure 4.3b) showed a similar pattern compared to that obtained from the point-by-point measurements (Figure 4.3a). Both histograms display a broad $\Delta L$ distribution that prevents accurate resolution of individual species. When we performed bootstrapping analysis (see Chapter II Section 2.5.6), however, two populations with baseline separations were obtained at $\Delta L$=5.6 ($\pm$ 0.1) and 8.8 ($\pm$ 0.1) nm respectively (Figure 4.3c). Based on the baseline separation with $R \geq 1.5$, we estimated the resolution of 1.1 nm for the two peaks depicted here. The resolution increases to 0.38 nm with an acceptable resolution of $R \geq 0.5$. Since the population with longer $\Delta L$ (8.8 ± 0.1 nm) matches with expected $\Delta L$ (9.0 ± 0.1 nm) for the hybrid-1 G-quadruplex (see Chapter II Section 2.5.7 for details), it was assigned as
this conformation.\textsuperscript{161} The 5.6 nm species has a size consistent with the partially folded G-triplex species described previously (see Chapter 3). When we performed similar experiments in a buffer that contains 100 mM LiCl, the yields reduce to 5 ± 2 \% for unfolding events with $\Delta L \leq 5$ nm and 0\% for $\Delta L > 5$ nm. In comparison, yields for G-triplex (~5 nm) and G-quadruplex (~9 nm) are 15 ± 4 \% and 65 ± 5 \%, respectively, in the KCl buffer. Since Li$^+$ is known to disfavor G-quadruplex formation, this control experiment verifies that folded populations observed here are G-quadruplexes, rather than hairpins that are not sensitive to Li$^+$. Further evidence of the G-quadruplex formation comes from the CD signatures in which ~268 and ~290 nm peaks with a ~238 nm valley are characteristic of hybrid type G-quadruplex conformations (Figure 4.4). Using these two $\Delta L$ values (8.8 and 5.6 nm) as Gaussian centers, the populations in the kernel density distribution (Figure 4.3b) were fit with a two-peak Gaussian, which shows 14 \% for the partially unfolded species and 86\% for the hybrid-1 G-quadruplex. This result indicates that G-quadruplex is a major population in the equilibrium.
Figure 4.5: Transition kinetics of the hTelo-4 sequence. a) PoDNano analyses reveal $\Delta L$ vs incubation time at 0 pN. Each population is fit with a Gaussian. b) The probability of folded (G-quadruplex + G-triplex) and unfolded populations as a function of incubation time. Curves are fitted with the functions solved for the two-state model shown in c (see Chapter II Section 2.5.12 for details).

Next, we used force-jump method to follow the transition kinetics of the species in the hTelo-4 sequence (Figure 4.2a). Figure 4.5a shows $\Delta L$ histograms for different
species revealed by PoDNano approach with different incubation time. Each histogram contains two populations that have been assigned as the partially folded G-triplex ($\Delta L = 5.6$ nm) and the fully folded G-quadruplex ($\Delta L = 8.8$ nm) (Figure 4.3c). Shown in Figure 4.5b are the probabilities of refolding for each species with incubation time. Here, G-quadruplex and G-triplex are grouped as a single population of folded species to accommodate the simplest two-state model (Figure 4.5c, see Chapter II Section 2.5.12). The rate constants $k_1$ and $k_2$ obtained here serve as starting values for the more complex fitting based on the 3-state or 4-state model described later.
Figure 4.6: Refolding probability of the G-quadruplex and G-Triplex with incubation time for the hTel-4 sequence. a) The folding-unfolding pathways of hTel-4 sequence represented by a linear model. b) Refolding probability data are fit with functions (see Chapter II Section 2.5.12) solved analytically from the linear model. Shown in c is a
To retrieve transition kinetics for individual species, we analyzed G-quadruplex, G-triplex, and unfolded states separately. We used a linear (Figure 4.6a) and a triangular (Figure 4.6c) three-state model to represent the in-pathway and off-pathway nature of the G-triplex with respect to the G-quadruplex folding, respectively. Both models fit the data equally well by residue analyses (Figure 4.6b and 4.6d, average $\kappa^2 = 0.0040$ vs. 0.0039, see Chapter II Section 2.5.12 for analytically solved equations). When we analyzed the unfolding pattern in each $F-X$ curve, we found it is rare to observe two sequential unfolding or refolding transitions expected for the G-triplex that serves as an in-pathway intermediate. Instead, most often we observed that occurrence of the G-triplex is independent of the G-quadruplex, which strongly suggests that the G-triplex is an off-pathway intermediate. The same model has been proposed by Vesnaver and coworkers recently to reflect transition dynamics of different species in a similar telomeric sequence. Fitting of this model to the data allows retrieving six rate constants (Figure 4.6d & Table 4.1) for all possible transitions among three species. The transient refolding probability maximum for the partially folded species unequivocally proves the intermediate nature of the G-triplex, since refolding probability would vary monotonically with time as predicted by a two-state, folding-unfolding model if the G-triplex exists independently. Interestingly, direct folding from unfolded state to G-
quadruplex is the most favorable route with fastest kinetics. The intermediate nature of the G-triplex with a low population fraction well explains the fact that the species is evasive for many ensemble-average techniques.

4.3.3 Transition Kinetics of Populations in hTelo-5 Sequence

Next, we proceeded to investigate the transition kinetics for species in a DNA sequence that contains five TTAGGG repeats (hTelo-5). The $\Delta L$ histogram shows four populations with $5.0 \pm 0.1$, $7.5 \pm 0.1$, $8.8 \pm 0.1$, and $11.5 \pm 0.2$ nm $\Delta L$ with formation possibilities of $8\%$, $8\%$, $74\%$, and $10\%$, respectively (Figure 4.7a). In control experiments performed in a 100 mM Li$^+$ buffer, these populations were rarely observed ($4 \pm 2 \%$, $3 \pm 1 \%$, and $0 \%$ for species with $\Delta L \leq 5$ nm, $5 < \Delta L \leq 10$ nm, and $\Delta L > 10$ nm, respectively). These results suggest that the species formed in the K$^+$ buffer are likely to be G-quadruplexes that are sensitive to Li$^+$ ions. The peaks at $\sim 268$ and $\sim 290$ nm with the valley at $\sim 238$ nm in CD spectrum (Figure 4.4) are indicative of G-quadruplex conformations, which provide further evidence for the formation of G-quadruplexes in the hTelo-5. Based on the matching $\Delta L$ values, the 5.0 and 8.8 nm populations are assigned to the partially folded species and the hybrid-1 G-quadruplex, respectively. The 7.5 nm species has a $\Delta L$ similar to that of the basket type G-quadruplex previously observed in a 100 mM Na$^+$ buffer.$^{230}$ Observation of the basket type G-quadruplex in the K$^+$ buffer is possible since G-quadruplex conformation is dependent on
various factors such as flanking DNA sequences\textsuperscript{233} or concentrations of a DNA template.\textsuperscript{81,161} However, the 11.5 nm population does not match with any of the known human telomeric G-quadruplexes. Since hTel-5 contains five G-rich repeats, it is possible to form a conformation in which a G-rich repeat (TTAGGG) is harbored in one of the internal loops.\textsuperscript{107}

To validate this conformation, we performed experiments on a mutant sequence, (TTAGGG)\textsubscript{2}-TTAAAA-(TTAGGG)\textsubscript{2}, or hTel-5-Mut, which can only form a G-quadruplex with terminally located G-rich repeats while leaving a long loop in the middle. Indeed, CD signatures of this sequence match well with those of a hybrid type G-quadruplex (~268 nm and ~292 peaks, and a 242 nm valley, Figure 4.4). In mechanical unfolding experiments, we observed a single population with \( \Delta L \) of 11.4 (± 0.2) nm (Figure 4.7b), which confirms that the 11.5 nm \( \Delta L \) population observed above (Figure 4.7a) is indeed a G-quadruplex species harboring a TTAGGG sequence in a long loop. This observation is consistent with the NMR structure that shows a long-loop G-quadruplex in a human telomeric sequence with five TTAGGG repeats in a K\textsuperscript{+} buffer.\textsuperscript{107} Since it is not possible for the long-loop G-quadruplex to fold into a G-quadruplex with regular TTA loops without unfolding first, we surmise it serves as a misfolded species similar to those often observed in protein or RNA species.
Figure 4.7: PoDNano analyses and transition kinetics of the hTelo-5 sequence. a) $\Delta L$ histogram for hTelo-5 obtained by bootstrapping analyses. The purple population depicts G-triplex, the green and red populations represent G-quadruplexes, and the blue population is a misfolded G-quadruplex. b) $\Delta L$ histogram for the hTelo-5-Mut sequence.
c) $\Delta L$ histograms of the hTel-5 fragment with different incubation time. d) Refolding probability of individual species over time. Two G-quadruplex populations (red and green) are grouped together (orange) in a four-state kinetic model shown in e (see text). Other color codes are described in a. Transition rate expression for each species in the four-state model is solved analytically (see Chapter II Section 2.5.12) to fit the data in d.

To test the misfolded nature of the long-loop G-quadruplex, we investigated the dynamic equilibrium among the species formed in the hTel-5 using the force-jump approach described above. Figure 4.7c shows $\Delta L$ histograms for all four species with various incubation times. The probabilities of refolding for individual species are plotted in Figure 4.7d. To reduce the complexity of kinetic equations to a manageable level, two conformations of G-quadruplexes (hybrid-1, $\Delta L=8.8$ nm; and basket type, $\Delta L=7.5$ nm) were grouped as a single G-quadruplex population in a four-state equilibrium model (Figure 4.7e) in which the long-loop G-quadruplex does not interconvert with regular G-quadruplex without first entering a third state such as unfolded DNA or G-triplex. Such a scenario is a likely result of the topological exclusivity for a G-rich repeat that either serves as one of the four G-stems in the G-quadruplex or stays free inside the long-loop G-quadruplex. Since current approach cannot distinguish individual G-triplexes or misfolded G-quadruplexes associated with specific G-quadruplexes probably due to similar sizes among these structures, the grouping of the two G-quadruplexes becomes necessary to analytically solve the kinetics equations. With these assumptions, we
analytically solved the four-state model shown in Figure 4.7e (see Chapter II Section 2.5.12). We notice that only few analytical solutions exist for a linear four-state model, which is different from current model. The data can be fit well with the rate equation for each species in the four-state model (Figure 4.7d, Table 4.1, and Chapter II Section 2.5.12), validating the misfolded nature of the long-loop G-quadruplex. Both G-triplex and misfolded-G-quadruplex demonstrate a maximal refolding probability during incubation, a feature indicative of the interconversion between these species and other DNA populations. Although misfolded DNA was implied in previous investigations in artificial DNA fragments, the long-loop G-quadruplex shown here is the first physiologically relevant misfolded species with defined conformations. This direct observation lends strong support to the notion that DNA can be as complex as RNA or proteins from structural and kinetic perspectives.

4.3.4 Equilibrium Population Dynamics in hTelo-4 and hTelo-5 Sequences

Compared to the kinetics in the hTelo-4, the formation of G-quadruplex (Table 4.1, $k_1$) remains constant whereas that for the G-triplex ($k_3$) increases about 11 times in the hTelo-5 ($k_3/k_1$ of hTelo-4 vs hTelo-5 = 1 vs 12.5). Such a behavior can be ascribed to the fact that there are increased possibilities to form G-triplexes rather than G-quadruplexes in longer telomeric sequences. If this reasoning is true, then it is expected
that the unfolding rates of G-triplexes should not depend on the DNA size, since G-triplexes already exist in these DNA fragments. Indeed, the unfolding kinetics of the G-triplex ($k_4$) with respect to that of the G-quadruplex ($k_2$) remains the same in both hTelo-4 and hTelo-5 fragments ($k_4/k_2$: 2.6 vs 2.0 for hTelo-4 vs hTelo-5). However, due to the formation of the misfolded G-quadruplex in the hTelo-5, the overall formation probability of the G-triplex reduces to 3.4% ($k_3/(k_1+k_3+k_7)$) from 10.7% in the hTelo-4. After G-triplex is formed, the species has a similar probability to fold into the regular G-quadruplex (hTelo-4: hTelo-5 = 36%:42%, calculated by $k_5/(k_4+k_5+k_10)$). On the other hand, the G-triplex has an increased tendency to unfold in the hTelo-4 (64%, $k_4/(k_4+k_5+k_10)$) compared to the hTelo-5 (7%). These observations are consistent with the fact that in the hTelo-5, either G-quadruplex or misfolded G-quadruplex can serve as a destination for G-triplex (Figure 4.7e), leading to a reduced possibility for the competing unfolding pathway. In the hTelo-4, however, only G-quadruplex competes with the unfolding (Figure 4.6d), which increases the unfolding probability of the G-triplex. Compared to the hTelo-5, such a result suggests that G-triplex in the hTelo-4 has an increased tendency to exist as an independently folded species off-pathway to the G-quadruplex. This off-pathway feature in the hTelo-4 is confirmed by the fact that in comparison to the hTelo-5, the unfolding probability of G-quadruplex to G-triplex, which is also present in the in-pathway process, is much reduced in the hTelo-4 (10% vs 40%, calculated by $k_6/(k_2+k_6)$). Notably, the long-loop G-quadruplex shows the fastest formation ($k_7$) and disassembly ($k_8$) kinetics among all species, suggesting that although
the sequence misfolds into the long-loop G-quadruplex, it is not “trapped” as escape from this state has a comparable rate. Indeed, analyses from Table 1 show 94% \((k_7/(k_1+k_3+k_7))\) folding probability of the misfolded G-quadruplex, which is comparable to the unfolding probability \((k_8/(k_8+k_9))=97\%\).

**Table 4.1:** Reaction rate constants (Avg ± SD s\(^{-1}\)) for transitions in hTelo-4 and hTelo-5 sequences.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>(k_1)</th>
<th>(k_2)</th>
<th>(k_3)</th>
<th>(k_4)</th>
<th>(k_5)</th>
<th>(k_6)</th>
<th>(k_7)</th>
<th>(k_8)</th>
<th>(k_9)</th>
<th>(k_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTelo-4</td>
<td>0.10 ± 0.03</td>
<td>0.009 ± 0.001</td>
<td>0.012 ± 0.004</td>
<td>0.023 ± 0.006</td>
<td>0.013 ± 0.002</td>
<td>0.0010 ± 0.0005</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>hTelo-5</td>
<td>0.083 ± 0.020</td>
<td>0.0030 ± 0.0008</td>
<td>0.126 ± 0.020</td>
<td>0.006 ± 0.001</td>
<td>0.036 ± 0.006</td>
<td>0.0020 ± 0.0009</td>
<td>3.52</td>
<td>4.32</td>
<td>0.134</td>
<td>0.044</td>
</tr>
</tbody>
</table>

4.3.5 Probability of Misfolded G-quadruplexes in Longer Human Telomeric Sequences

In previous NMR studies with carefully designed mutants, long-loop G-quadruplexes (or misfolded-G-quadruplexes) were observed in human telomeric sequences with 5-7 G-rich repeats.\(^{107}\) Using PoDNano, here we observed that under physiologically relevant buffer conditions and DNA concentrations, these species are population minorities. As shown in Figure 4.8, these misfolded species represent 10%, 11%, and 13% of the total population in hTelo-5, hTelo-6, and hTelo-7 fragments, respectively. When we compared misfolded populations in the same fragment, we found
that species with longer loop(s) or more long-loops have smaller populations, suggesting a decreased thermodynamic stability for these misfolded species. In hTelo-6, the populations of folded species with longer loops or more long-loops drop from 8% to 3%; in HTelo-7, these values reduce from 9 to 2%. We argue that entropic factor could be the culprit for the decrease in population of such misfolded G-quadruplexes. As loops become longer or more long-loops occur, the loss of entropy due to the confinement of the loop(s) inside the G-quadruplex becomes larger, which reduces the population of misfolded species.

In all DNA constructs investigated here (hTelo-4 to hTelo-7), we observed that regular G-quadruplexes are the most populated species (Figure 4.8). Such an observation has been supported by CD spectra as discussed above (Figure 4.4). This population pattern is further supported in a DNA construct that contains eight G-rich repeats (hTelo-8) in which two individual G-quadruplex units can form. Mechanical unfolding experiments on this construct revealed two-step (~45 %) as well as one-step (~55 %) unfolding transitions (Figure 4.9 for typical $F-X$ and $\Delta L-F$ curves that contain two-step transitions in hTelo-8). The $\Delta L$ of each transition in the two-step unfolding events bears similarity to that observed in the hTelo-4 fragment (G-quadruplex to G-triplex ratio ~ 6:1). Consistent with this ratio, majority of the two-step events shows unfolding of two individual G-quadruplexes (60% out of all two-step transitions). Among 55 % one-step transitions, majority (85% out of all one-step transitions) is the unfolding of G-
quadruplex (including simultaneous unfolding of two G-quadruplexes) with 11% of G-triplex and few (~4%) misfolded G-quadruplex species. These observations suggest that regular G-quadruplexes with shortest possible TTA loops would be the most predominant species in the human telomeric regions. However, further investigations on longer telomeric sequences with multiple runs of four G-rich repeats are required to fully support this argument.
Figure 4.8: Distribution probability of folded population in human telomeric sequences (hTel-4 to hTel-7). The 11.5 nm species is the G-quadruplex that contains one G-rich repeat (TTAGGG) in a loop, the 14 nm species contains two G-rich repeats in loop(s), while the 16 nm species has three G-rich repeats in loop(s).

Figure 4.9: Typical F-X (a) and ΔL-F (b) plots for the mechanical unfolding measurements of the hTel-8 fragment showing two-step unfolding events. The ΔL of each transition in the hTel-8 shows a similar pattern as that of hTel-4 (~9 nm) for the unfolding of individual GQ.

4.4 Conclusions

Using force-jump and population deconvolution at sub-nanometer resolution, we have identified individual populations and followed their kinetics with unprecedented resolutions in human telomeric sequences containing four to eight TTAGGG repeats. We have observed a highly complex population equilibrium that contains G-triplex,
misfolded G-quadruplexes, and predominant G-quadruplex species. The misfolded G-quadruplexes harbor one or more G-rich repeats in the loop(s) and the population of these species in a particular sequence decreases with increase in the loop length or the number of long-loops. The presence of the misfolded species testifies the structural complexity of DNA. The complexity of the system is further reflected by the transition kinetics in the hTelo-4 and hTelo-5 DNA in which three-state and four-state kinetic models are followed, respectively. The population dynamics of telomeric species observed here indicate that in the full-length 3’ end overhang of human telomere, G-quadruplex units with shortest possible TTA loops would be the most prevalent species. This justifies the modeling role of regular G-quadruplexes in the study of human telomeric structures.

This project was supported by NSF CHE-1026532 to Hanbin Mao and ICC program of Japan Society for the Promotion of Science and Core Research for Evolitional Science and Technology (CREST) of JST to Hiroshi Sugiyama.
Chapter V

Interactions of Human Telomeric G-quadruplex
and Small-Molecule Ligands: Mechanochemical
Dynamics Revealed at Single-Molecule Level

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED AS AN ARTICLE IN NATURE CHEMISTRY JOURNAL, NAT. CHEM. 2011, 3, 782-787. ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED WITH THE COPYRIGHT PERMISSION FROM NATURE PUBLISHING GROUP

5.1 Abstract

Ligands that stabilize the formation of telomeric DNA G-quadruplex have potential as cancer treatments because the G-quadruplex structure can inhibit telomerase - an enzyme over expressed in many cancer cells. Understanding the kinetic, thermodynamic, and mechanical properties of small molecule binding to these structures is therefore important, but classical ensemble assays are unable to measure these simultaneously. Here, we employed laser-tweezers method to investigate such interactions. With a force jump approach, we observed that pyridostatin promote the
folding of telomeric G-quadruplexes. The increased mechanical stability of pyridostatin bound G-quadruplex permitted the determination of the dissociation constant \( (K_d) \) of 490 ± 80 nM. The free energy change of binding obtained from a Hess-like process provided an identical \( K_d \) for pyridostatin and a \( K_d \) of 42 ± 3 µM for a weaker ligand RR110. We anticipate this single-molecule platform can provide detailed insights into the mechanical, kinetic, and thermodynamic properties of liganded bio-macromolecules that have biological relevance.

5.2 Overview

G-quadruplexes are four-stranded nucleic acid structures arising from the folding of particular guanine (G) rich DNA and RNA sequences.\(^{32, 237, 238}\) With hundreds of 5'-TTA GGG repeats in its sequence, the human telomere has the propensity to form G-quadruplexes.\(^{60, 68}\) These quadruplexes have been proposed to play an important role in the maintenance of telomere length through inhibition of telomerase.\(^{66, 239}\) Such an effect has profound implications in cancer cells, the majority of which have over-expressed levels of telomerase.\(^{66, 213}\) Small molecules such as telomestatin\(^{89, 90}\) have been demonstrated to bind to telomeric G-quadruplexes and suppress telomerase activity. This discovery has begun the expansion of telomeric G-quadruplexes into potential therapeutic targets for cancer treatment.\(^{15, 25, 87}\)
As a polymerase, telomerase extends the telomere region using an RNA sequence, 3'-CAAUCCCAAUC-5' as a template.\textsuperscript{88} Our previous studies on human ILPR sequence have shown that the mechanical stability of G-quadruplexes\textsuperscript{157} is higher than the stall force of RNA and DNA polymerases.\textsuperscript{240, 241} This suggests that quadruplexes \textit{per se} can stall the replication or transcription processes. Telomeric G-quadruplex structures stabilized by a small molecule may have increased mechanical stability to further thwart the efforts of telomerase to elongate its substrate, the G-overhang. Previously, bulk methods such as X-ray\textsuperscript{91} and NMR\textsuperscript{72, 231} have determined detailed structural information on the telomeric G-quadruplexes. Other ensemble assays such as thermal melting analysis and surface plasmon resonance (SPR) have provided insights into folding and unfolding kinetics.\textsuperscript{242, 243} Fluorescence based single-molecule methods have revealed the dynamic nature of these structures with and without bound ligands.\textsuperscript{114, 118, 181} However, the mechanical information of the quadruplex-ligand complex has not been investigated. Here we present studies that have exploited a single-molecule mechanical assay to provide kinetic, thermodynamic, and mechanical information on quadruplex-ligand interactions.
Figure 5.1: Schematic of the experimental setup (not to scale). Ligands PDS and RR110 are shown in (a) and (b), respectively. (c) The force-based single-molecule platform for the investigation of the ligand-quadruplex interaction. Human telomeric fragment with a G-quadruplex forming sequence (hTelo-4) is sandwiched between two dsDNA spacers. The overall construct is tethered between two optically trapped beads via digoxigenin (dig)-anti-digoxigenin (anti-dig) linkage at one end and biotin-streptavidin linkage at the other end. The mechanical stability of the G-quadruplex structure in the presence or absence of a ligand is then evaluated by mechanical unfolding of the quadruplex structure.

Our assay employs a platform that contains a single telomeric G-quadruplex structure to evaluate its interaction with small-molecule ligands in a laser tweezers instrument. Based on structural similarities, two ligands have been evaluated, pyridostatin...
(PDS)\textsuperscript{198, 199} (Figure 5.1a), a highly potent small molecule for telomeric G-quadruplex binding, and a less potent analogue, RR110 \textsuperscript{200} (Figure 5.1b). PDS has already been shown to induce dysfunctional telomeres in cancer cells with the uncapping of POT1 that results in DNA damage signaling activation.\textsuperscript{198} In contrast, this phenotype has not been observed upon incubation of RR110. Moreover, FRET-melting experiments have shown that in agreement with in cellulo data, the $\Delta T_m$ of human telomeric G-quadruplex induced by RR110 is poor compared to PDS. This can be explained by the absence of a third positively charged side chain, which increases the selectivity of RR110 for G-quadruplex over dsDNA, but decreases its potency in the binding and stabilization of the G-quadruplex.\textsuperscript{200} To investigate the dynamic binding events, we have incorporated a “force jump” approach that allows the observation of the events in a milliseconds time range. We have further revealed that upon ligand binding, human telomeric G-quadruplex increases its mechanical stability. This mechanical information has been exploited to retrieve the thermodynamic dissociation constant between the G-quadruplex and the ligand. Previous efforts to obtain the same information from bulk assays have failed due to solubility issues of the two ligands. Using analysis similar to the Hess cycle\textsuperscript{244}, we have derived change in the free energy of binding ($\Delta G_{\text{binding}}$) and used it to calculate the dissociation constant without varying either quadruplex or ligand concentration. The dissociation constants obtained from these two single-molecule calculations agree very well with each other. The generic nature of this single-molecule platform makes it amenable to investigate other receptor and ligand systems.
Figure 5.2: Mechanical unfolding of hTelo-4 with and without PDS. (a) Typical $F$-$X$ curves showing the unfolding events (marked by arrows) during the stretching for ligand bound (red) and free (blue) G-quadruplex structures. The black trace represents the relaxing $F$-$X$ curve for both cases. Histograms of unfolding forces for the structure in the hTelo-4 in the absence (b) and presence (c) of 0.5 µM PDS. Red curves in (b) and (c) represent Gaussian fittings. The $21 \pm 1$ pN population in (b) or (c) represents free G-
quadruplex whereas the 41±2 pN species represents G quadruplex tightly bound with PDS.

5.3 Results and Discussion

5.3.1 Ligand Binding Increases Mechanical Stability of Human Telomeric G-quadruplex

The single-molecule platform constituted a single-stranded fragment of human telomeric sequence 5′-TTA (GGG TTA)_4, or hTelo-4, sandwiched between two dsDNA spacers (see Chapter II Section 2.5.3). Two spacers were tethered to two optically trapped beads via digoxigenin-antidigoxigenin and biotin-streptavidin linkages, respectively (Figure 5.1c). To investigate the mechanical stability of the DNA structure formed in the hTelo-4 sequence in 10 mM Tris buffer (pH 7.4) with 100 mM KCl, we stretched the DNA construct by moving one of the optically trapped beads away from the other with a load rate of 5.5 pN/s. A rupture event was observed in the force-extension (F-X) curve (Figure 5.2a, bottom arrow), indicating the unfolding of a secondary structure in the hTelo-4 sequence. The contour length of this structure, 9.3 ± 0.1 nm, indicated a folded structure of 21 nucleotides, which was expected for a telomeric G-quadruplex in this construct. The histogram of the rupture forces (Figure 5.2b) showed a population of 21 ± 1 pN for this structure. This suggests that the telomeric G-quadruplex can withstand the
load force of polymerases.\textsuperscript{240, 241} Using the Jarzynski’s theorem for non-equilibrium systems\textsuperscript{207, 208} (see Chapter II Section 2.5.10), the change in free energy of unfolding the quadruplex, $\Delta G_{\text{unfold}}$, was found to be $9.8 \pm 0.4$ kcal mol\textsuperscript{-1} at 23 °C. This value was consistent with those previously determined for telomeric G quadruplexes.\textsuperscript{180, 184, 188, 245}

The same single molecule setup was employed to evaluate the effect of ligands on the mechanical stability of telomeric G-quadruplex. First, we incubated the hTelo-4 construct in the same buffer with 0.5 µM PDS, a small molecule specific for G-quadruplex structures.\textsuperscript{198, 199} The mechanical stability of the G-quadruplex in the presence of the ligand was revealed by the rupture events in F-X curves (Figure 5.2a, top arrow). After each rupture event, we brought back the DNA construct to zero force, waited at least 30 seconds to allow for ligand binding, and then started the next round of stretching procedure. When the rupture force histogram was analyzed, two populations with $20 \pm 1$ pN and $41 \pm 2$ pN were observed (Figure 5.2c). The population of 20 pN demonstrated the same rupture force as that without ligand (Figure 5.2b). Therefore, it was assigned as a free G-quadruplex population. The species with an increased rupture force of 41 pN may be due to either specific or nonspecific binding of PDS to the DNA construct. To rule out the nonspecific binding of PDS to dsDNA, we mixed 10 µM PDS with a DNA construct that contained only two dsDNA spacers (see Chapter II Section 2.5.3). The percentage of the unfolding events ($\sim 2.5\%$, see Chapter II Section 2.5.8 for details) dramatically reduced compared to that of the rupture events at 41 pN (50\%) for hTelo-4.
with 0.5 µM ligand. This percentage (~2.5%) was identical with that observed for the same construct without ligand. To rule out the binding of the ligand with single stranded sequences incapable of forming any G-quadruplex, further experiments were performed on the DNA construct that contained 5'-TGT (CCC CAC ACC CCT GT)_2 instead of the hTelo-4 (see Chapter II Section 2.5.3). This sequence has been shown to remain single stranded at pH 7. Mechanical unfolding of this construct yielded a negligible percentage of unfolding events (~2% without PDS and ~3% with PDS). These results effectively ruled out the nonspecific binding of the ligand to the DNA construct and were in accord with previously observed specificity of PDS for G-quadruplex structures. Based on this, we assigned the population with increased rupture force as ligand bound telomeric G-quadruplexes.
**Figure 5.3:** Folding kinetics of quadruplex structure in the presence or absence of PDS. (a) The profile of force over time for the force jump method (see text for details). (b) Histograms of the unfolding force of the telomeric G-quadruplex structure with incubation time in a solution with 0.5 μM PDS indicate the fraction of PDS-bound population increases over time. Red curves represent Gaussian fittings. (c) Refolding probability of all quadruplex species vs. incubation time in a solution with (filled circles with a dotted red curve) and without (open circles with a solid green curve) 0.5 μM PDS. Open diamonds with a dotted gray curve (“Ligand Free”) and filled diamonds with a solid gray curve (“Ligand Bound”) are the refolding probabilities for free and ligand-bound G-quadruplexes in a solution containing PDS, respectively. The sum of these two curves is equivalent to that represented by the filled circles with a dotted red curve (“In Ligand”). Curves represent single exponential fittings, which yield reaction rate constants of $0.20\pm0.01$/s, $0.11\pm0.01$/s, $0.48\pm0.06$/s, and $0.13\pm0.02$/s for top to bottom traces, respectively. The black star indicates little formation of the 41 pN population at the 0 s incubation in the solution with PDS, which suggests the presence of free or loosely bound G-quadruplexes before tightly bound with PDS (see text for details). (d) Ratio of ligand-bound quadruplex to free quadruplex with incubation time indicates that the equilibrium has been reached at ~30 s. The solid red curve represents single exponential fitting (see Chapter II Section 2.5.13). Error bars in (c) and (d) are the standard deviations from three independent measurements.
5.3.2 Kinetic Measurements Indicate That Binding of Ligand Facilitates Formation of G-quadruplex

To investigate the dynamic binding of PDS to the telomeric G-quadruplex, we designed a force-jump method (Fig. 5.3a, see Chapter II Section 2.5.4). In this approach, we stretched the DNA construct until G-quadruplex was unfolded. This was followed by a rapid force-jump to 0 pN to allow the refolding of the secondary structure for a specific time interval. The refolding of the structure was indicated by a rupture event in the next round of the stretching procedure that started at 7 pN by another force jump process. These two force jumps prevent the refolding of the structure during the stretching and relaxing of the DNA at the low force region (<10 pN). Events as fast as 100 ms can be readily measured by this procedure. Using this strategy, we varied the incubation time to measure the folding kinetics of telomeric G-quadruplex with and without 0.5 µM PDS. Figure 5.3b shows force histograms with incubation times ranging from 0-30 s. Except for the 0 s incubation, each histogram shows two populations similar to those in Figure 5.2c. Based on this similarity, we assigned the population with 20 pN rupture force as free G-quadruplex, whereas the 41 pN species as the G-quadruplex tightly bound with PDS. At 0 s incubation, the 41 pN population was not observed (Figure 5.3b first histogram and the star in Figure 5.3c), indicating that either the ligand is yet to bind to the quadruplex (see the first step of the pathway B in Figure 5.4), or the quadruplex and the
ligand complex has not transformed into a structure with increased mechanical stability (see the first step of the pathway A in Figure 5.4).

**Figure 5.4:** Proposed pathways for the binding of PDS to the telomeric G-quadruplex. Notice that loosely associated quadruplex-ligand complex has similar mechanical stability compared to the ligand free G-quadruplex.
With this assignment, we calculated and plotted the probability of refolding for each species with a specific incubation time (Figure 5.3c, see Chapter II Section 2.5.11). Single exponential functions provided excellent fitting to the observed data, from which folding rate constants were attained. The single-molecule nature of the method allowed us to simultaneously determine the kinetics of individual populations in a mixture of different species. We observed the overall folding rate constant for the G-quadruplexes (including both 20 and 41 pN populations), $k_{\text{fold, ligand}} = 0.20 \pm 0.01 \text{ s}^{-1}$, in a solution with 0.5 µM PDS. Further deconvolution analysis (see Chapter II Section 2.5.9 for details) of these two populations allowed us to simultaneously determine the folding rate constant of the ligand free or loosely bound quadruplex (the 20 pN population), $k_1 = 0.48 \pm 0.06 \text{ s}^{-1}$, and the quadruplex tightly bound with the ligand (the 41 pN population), $k_2 = 0.13 \pm 0.02 \text{ s}^{-1}$ (see Figure 5.4). Compared to the rate constant for the G-quadruplex in a solution without PDS, $k_{\text{fold, no-ligand}} = 0.11 \pm 0.01 \text{ s}^{-1}$ (see Figure 5.4), the $k_1$ (0.48 ± 0.06 s$^{-1}$) was much faster. This result indicates that PDS increases the formation kinetics of the telomeric G-quadruplex, likely due to the lowered activation energy as PDS participates in the folding of the quadruplex (see the first step of pathway A in Figure 5.4). Such kinetics information also disfavors the scenario that PDS binds to a preformed G-quadruplex (pathway B in Figure 5.4), which predicts the ligand has less effect on the folding kinetics of the free G-quadruplex. The fact that $k_2$ (0.13 ± 0.02 s$^{-1}$) is much slower
than $k_1 (0.48 \pm 0.06 \text{ s}^{-1})$ suggests that the rate limiting step for the PDS binding is the process in which the loosely associated ligand searches for a tightly bound state (the second step in pathway A in Figure 5.4). It is noteworthy that with the continuous flow of 0.5 µM PDS, the ligand concentration is effectively constant during the binding to the single-molecule DNA substrate. Therefore, the first step in pathway A follows a pseudo first order reaction kinetics, which validates our kinetic analysis discussed here.

To reveal the time to reach the equilibrium between free quadruplexes and ligand bound structures in the presence of PDS, we plotted the ratio of these two species with time. Figure 5.3d depicts that the equilibrium is reached within 30 s with a time constant of $\tau = 10$ s (see Chapter II Section 2.5.13). In the following experiments, we kept a 30 s incubation time to ensure the equilibrated binding between ligands and the G-quadruplex.

### 5.3.3 Determination of Dissociation Constant without Varying Concentration of Either Ligand or Receptor

Our finding that PDS substantially increases the mechanical stability of telomeric G-quadruplex (Figure 5.2c) correlates well with the previous finding that melting temperature is increased by $\sim 35$ K for the telomeric G-quadruplex in the presence of PDS.$^{198}$ To quantify the stabilization effect of the ligand, we evaluated the unfolding force histogram of G-quadruplex in the presence of 20 nM to 2 µM PDS (Figure 5.5a-e).
By integrating the total rupture events of each population in the rupture force histogram using a randomized deconvolution method (see Chapter II Section 2.5.9), we analyzed the ratio of free G-quadruplex vs ligand bound structure at each concentration. As shown in Figure 5.5f, the ratio clearly follows a Langmuir isotherm of single binding site (see Chapter II Section 2.5.13). With this model, the dissociation constant ($K_d$) was found to be $490 \pm 80$ nM. This value is similar to the IC50 value of 200 nM previously evaluated for a competition assay between PDS and POT1.

**Figure 5.5:** Measurement of the dissociation constant between PDS and the telomeric G-quadruplex. (a-e) Histograms of unfolding force for the G-quadruplex structure in the presence of various concentrations of PDS. The left Gaussian population in each histogram corresponds to the free quadruplex structure while the right population
represents the ligand bound quadruplex species. (f) The fraction of the G quadruplex bound with PDS vs concentration of the ligand. The solid red curve is the fitting according to the single site binding model (see Chapter II Section 2.5.13 for details), which provides the dissociation constant of 490 ± 90 nM for the PDS binding. Error bars are standard deviations from three independent measurements.

We further evaluated the stabilization effect of the PDS from free energy perspective. After deconvoluting the rupture events associated with the ligand bound or free G-quadruplexes as described above, we combined the rupture events for each species at different concentrations and calculated the change in free energy (ΔG\textsubscript{unfold}) using the Jarzynski’s theorem for non-equilibrium systems (see Chapter II Section 2.5.10).\textsuperscript{207, 208} The ΔG\textsubscript{unfold} for the ligand bound and free quadruplexes were 8.0 ± 0.5 and 16.6 ± 0.5 kcal mol\textsuperscript{-1}, respectively. This result demonstrates that G-quadruplex bound with ligand increases the thermodynamic energy of the structure. As a state function, the change in free energy of binding, ΔG\textsubscript{binding}, could be calculated using a process analogous to the Hess cycle (Figure 5.6a), \( ΔG\text{binding} = ΔG_{Q\text{-unfold}} - ΔG_{Q-L\text{-unfold}} \) (Eqn 1). Here, ΔG\textsubscript{Q\text{-unfold}} is the free energy change associated with unfolding of a free quadruplex and ΔG\textsubscript{Q-L\text{-unfold}} is associated with the unfolding of a ligand bound quadruplex. The dissociation constant was then calculated using \( ΔG\text{binding} = -RT \ln K_d \) (Eqn 2), where R is the gas constant, T is absolute temperature, and ΔG\textsubscript{binding} = -8.6 kcal mol\textsuperscript{-1} was calculated from Eqn 1. This calculation provided the dissociation constant of 460 ± 40 nM, which was identical
within experimental error with the value obtained from the Langmuir isotherm.

Compared to current dissociation constant assays in which experiments must be performed with different ligand or receptor concentrations, this new method dramatically simplifies the process by requiring only one concentration of ligand or receptor. This is especially beneficial for screening purposes in which large quantity of ligands is often not available.

We have applied this new method to obtain the dissociation constant of a related PDS analogue, RR110 (Figure 5.1b). In our estimation, this ligand displays >20 fold weaker binding than PDS. A practical limitation of the weak binding is that it prevents the accurate determination of the dissociation constant due to the solubility problem of this small molecule. We performed the mechanical unfolding of the hTelo-4 in 1µM RR110. The rupture force histogram in Fig. 5b shows a predominant peak at 20 ± 1 pN with a shoulder at 34 ± 1 pN. The 20 pN species was assigned as free G-quadruplex based on the identical rupture force between the two. The population with 34 pN was assigned to the ligand bound quadruplex. To confirm this assignment, we performed the same experiment in the presence of 10 µM RR110. The unfolding force histogram (Figure 5.6c) now shows an increased 34 pN population. This result is expected as the population of the ligand bound quadruplex increases with increasing ligand concentrations. After obtaining $\Delta G_{quadruplex-unfold}$ for free quadruplex (8.7 ± 0.2 kcal mol$^{-1}$) and $\Delta G_{Q-L-unfold}$ for ligand bound structure (14.6 ± 0.3 kcal mol$^{-1}$), we attained
\[ \Delta G_{\text{binding}} \approx -5.9 \pm 0.4 \text{ kcal mol}^{-1} \] using the Hess-like cycle described above (Eqn 1). From Eqn 2, this yielded a dissociation constant of \( 42 \pm 3 \) µM for RR110, which was much lower than that of PDS determined above. It is noteworthy that the mechanical stability of the RR110 bound quadruplex (34 pN) is also significantly lower than that of the PDS bound quadruplex (41 pN).
Figure 5.6: Determination of the dissociation constant from single ligand concentration. (a) Schematic representation of the process analogous to the Hess cycle for the calculation of $\Delta G_{\text{binding}}$. The histograms of the unfolding force of the G-quadruplex in the presence of (b) 1µM and (c) 10 µM of RR110. Each histogram is fit with a two-population Gaussian (black curve). The 20 ± 1 pN population fit with a solid red Gaussian curve represents the ligand free G-quadruplex and the 34 ± 1 pN population fit with a dotted blue Gaussian curve depicts the G-quadruplex bound with a weaker ligand, RR110, which shows a dissociation constant of 42 ± 3 µM. Note that this dissociation constant is ~86 times higher than that of PDS. Reported errors in the unfolding force and dissociation constant represent the standard deviations from three independent measurements.

5.4 Conclusions

Using a force based single-molecule platform, we have demonstrated that pyridostatin binds to human telomeric G-quadruplex and promotes the folding of this G-quadruplex. By a process analogous to the Hess cycle, we have obtained the dissociation constants of pyridostatin and RR110, although neither of which has been accurately determined in ensemble assays due to solubility issues. This method simplifies the dissociation constant assay without the requirement for ligand or receptor titration and offers a general platform that can be applied to other biologically relevant ligand-receptor interactions.
systems. Specifically, this study highlights that G-quadruplexes are important dynamic structures involved in the mechanism of telomere elongation by the action of the enzyme complex telomerase. Furthermore, we consider the mechanical information acquired by this system could provide novel perspectives for drug testing and design in the future. In addition, the extra forces required to unfold ligand bound structures shows that in principle, the small molecule pyridostatin could well interfere with RNA and DNA polymerases during the processes of transcription and replication in vivo.

For this project, Hanbin Mao was supported by the New Faculty Award Program at Camille and Henry Dreyfus Foundation, Ohio Board of Regents, and NSF CHE-1026532, Shankar Balasubramanian by the BBSRC of the UK for a studentship to BA and Cancer Research UK for programme and Hiroshi Sugiyama by the Core Research for Evolutional Science and Technology (CREST) of JST and the Grant-in-Aid for Science research from the MEXT, Japan.
Chapter VI

Conclusion and Perspective

Since human telomeres are associated with genetic integrity, cell proliferation, aging and cancer, fundamental understanding of the equilibrium dynamics and the transition kinetics of G-quadruplex structures, their intermediates and other alternative structures formed in the human telomeric DNA sequences and their interactions with small-molecule ligands or proteins are crucial. This dissertation research has confirmed the existence of folded intramolecular structures other than the G-quadruplex in human telomeric sequences under physiologically relevant conditions of Na$^+$ and K$^+$. Molecular dynamics simulation, single-molecule optical tweezers, and bulk methods such as CD and DMS footprinting strongly support that the folded structure in hTelo-3 and the partially folded structure in hTelo-4 assume a triplex conformation. These results reveal an unprecedented and diverse folding pattern in the human telomere region, which is instrumental for the development of new telomere-targeting small-molecule drugs.

Using force-jump and population deconvolution at sub-nanometer resolution techniques, single-molecule optical-tweezers investigations are able to identify the individual populations associated with human telomeric DNA sequences containing four
to eight TTAGGG repeats and precisely follow their kinetics with unprecedented resolutions. A highly complex population equilibrium that contains G-triplex, misfolded G-quadruplexes, and predominant G-quadruplex species has been observed. The misfolded G-quadruplexes harbor one or more G-rich repeats in the loop(s) and the population of these species in a particular sequence decreases with increase in the loop length or the number of long-loops. The presence of the misfolded species testifies the structural complexity of DNA. The complexity of the system is further reflected by the transition kinetics in the hTel-4 and hTel-5 DNA in which three-state and four-state kinetic models are followed, respectively. The population dynamics of telomeric species observed here indicate that in the full-length 3’ end overhang of human telomere, G-quadruplex units with shortest possible TTA loops would be the most prevalent species. This justifies the modeling role of regular G-quadruplexes in the study of human telomeric structures.

Single-molecule mechanochemical platform discussed in this dissertation discovered that pyridostatin binds to human telomeric G-quadruplex and promotes its folding kinetics. By a process analogous to the Hess cycle, the dissociation constants of pyridostatin and RR110, have been obtained, although neither of which has been accurately determined in ensemble assays due to their water-solubility issues. This method simplifies the dissociation constant assay without the requirement for ligand or receptor titration and offers a general platform that can be applied to other biologically
relevant ligand-receptor systems. Specifically, this study highlights that G-quadruplexes are important dynamic structures involved in the mechanism of telomere elongation by the action of the enzyme complex telomerase. Furthermore, this mechanochemical information acquired by this system could provide the novel perspectives for drug testing and design in the future. In addition, the extra forces required to unfold ligand bound structures reveal that small-molecule stabilized DNA secondary structures could well interfere with RNA and DNA polymerases during the processes of transcription and replication \textit{in vivo}.
References


27. Nagatoishi, S.; Nojima, T.; Galezowska, E.; Juskowiak, B.; Takenaka, S., G quadruplex-based FRET probes with the thrombin-binding aptamer (TBA) sequence


60. Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J.-R., A highly conserved repetitive DNA sequence,


134. Paramanathan, T.; Vladescu, I.; McCauley, M. J.; Rouzina, I.; Williams, M. C., Force spectroscopy reveals the DNA structural dynamics that govern the slow binding of Actinomycin D. *Nucleic Acids Res.* 2012, 40, 4925–4932.


