MECHANICAL STABILITY EVALUATION OF I-MOTIF AND G-QUADRUPLEX STRUCTURES UNDER DIVERSE CIRCUMSTANCES

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

by
Soma Dhakal
May 2013
Dissertation written by

Soma Dhakal

B.S., Tribhuvan University, Nepal, 2001
M.S., Tribhuvan University, Nepal, 2003
Ph. D., Kent State University, USA, 2013

Approved by

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

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

___________________________________, Member, Doctoral Dissertation Committee
Soumitra Basu, Ph. D

___________________________________, Member, Doctoral Dissertation Committee
Chanjoong Kim, Ph. D

Accepted by

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

___________________________________, Associate Dean, College of Arts and Sciences
Raymond A. Craig, Ph. D
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... vi

LIST OF TABLES ........................................................................................................... x

ACKNOWLEDGEMENTS ................................................................................................. xi

Chapter

1  Introduction and background .................................................................................... 1

   1.1 G-quadruplex ................................................................................................. 1

   1.2 i-Motif .......................................................................................................... 3

   1.3 G-quadruplex and i-motif in human ILPR ....................................................... 5

   1.4 G-quadruplex and i-motif in duplex DNA ..................................................... 6

   1.5 Biological roles of G-quadruplex and i-motif ................................................. 7

   1.6 Material applications of G-quadruplex and i-motif ...................................... 9

   1.7 Laser Tweezers for single-molecule manipulation ...................................... 11

2  Materials and methods ............................................................................................. 14

   2.1 Materials and Methods ............................................................................... 14

      2.1.1 DNA Samples ....................................................................................... 14

      2.1.2 DNA construct for ILPR i-motif and telomeric G-quadruplex in ssDNA ........... 15

      2.1.3 DNA construct for i-motif and G-quadruplex in ILPR duplex .............. 18

      2.1.4 Single-Molecule Experiment ............................................................... 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.5</td>
<td>Calculation of change in contour length</td>
<td>21</td>
</tr>
<tr>
<td>2.1.6</td>
<td>Deconvolution of Populations</td>
<td>21</td>
</tr>
<tr>
<td>2.1.7</td>
<td>Control Experiments for Single-Molecule Study</td>
<td>22</td>
</tr>
<tr>
<td>2.1.8</td>
<td>Calculation of Percentage Formation</td>
<td>22</td>
</tr>
<tr>
<td>2.1.9</td>
<td>Calculation of $\Delta G_{\text{unfold}}$</td>
<td>22</td>
</tr>
<tr>
<td>2.1.10</td>
<td>CD Spectroscopy</td>
<td>23</td>
</tr>
<tr>
<td>2.1.11</td>
<td>UV Spectroscopy</td>
<td>24</td>
</tr>
<tr>
<td>2.1.12</td>
<td>Preparation of 5' end radiolabeled DNA</td>
<td>24</td>
</tr>
<tr>
<td>2.1.13</td>
<td>Electrophoretic Mobility Shift Assay (EMSA)</td>
<td>24</td>
</tr>
<tr>
<td>2.1.14</td>
<td>Bromine Footprinting of ssDNA</td>
<td>25</td>
</tr>
<tr>
<td>2.1.15</td>
<td>DMS and Bromine footprinting of dsDNA</td>
<td>26</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>Results and Discussion</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>Conclusions</td>
<td>48</td>
</tr>
</tbody>
</table>

4 Intramolecular folding in human ILPR fragment with three C-rich repeats

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Results</td>
<td>52</td>
</tr>
<tr>
<td>4.3</td>
<td>Discussion</td>
<td>69</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>71</td>
</tr>
</tbody>
</table>
5. G-quadruplex and i-motif are mutually exclusive in double-stranded ILPR DNA.

5.1 Introduction

5.2 Results

5.3 Discussion

5.4 Conclusion

6. CONCLUDING REMARKS

PERMISSION TO REUSE COPYRIGHTED MATERIAL

REFERENCES
# LIST OF FIGURES

**Figure 1.1** Formation of G-quadruplex and its general types based on the strand orientation………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
Figure 3.6 A representative pulling curve with two ~5 nm change in contour length (\(\Delta L\)) unfolding events in the sequence 5'-(TGTC\textsubscript{C}CC\textsubscript{C}ACACCC\textsubscript{C})\textsubscript{2}TGT

Figure 3.7 Histogram of change in contour length (\(\Delta L\)) of the DNA secondary structure formed in 5'-TGTC\textsubscript{C}ACAC\textsubscript{C}TGTC\textsubscript{C}ACA

Figure 3.8 \(\Delta L\) histograms at different pH (23 °C)

Figure 3.9 The percentage formation of the i-motif and the partially folded structure at different pH under 23 °C

Figure 3.10 Rupture force histograms of partially folded structure and i-motif structures different pH under 23 °C

Figure 4.1 CD experiments of ILPR-I3 at different pH and temperature in a 10 mM sodium phosphate buffer with 100 mM KCl and 5 µM DNA concentration

Figure 4.2 Electrophoretic mobility shift assays (EMSA) and thermal melting/reannealing of the ILPR-I3

Figure 4.3 CD spectra of sequences described in Table 4.1

Figure 4.4 A typical force-extension (\(F-X\)) curve obtained from the mechanical unfolding of the secondary structure in the ILPR-I3 at pH 5.5

Figure 4.5 Single-molecule study of the ILPR-I3 at different pH using laser tweezers
Figure 4.6 Four possible structures that employ C:CH\textsuperscript{+} pair stacking in the ILPR-I3 sequence ……………………………………………………………………………60

Figure 4.7 Mutation analysis in a 10 mM sodium phosphate buffer (pH 5.5) with 100 mM KCl ……………………………………………………………………………64

Figure 4.8 Formation of an intermolecular i-motif………………………………………………………66

Figure 4.9 Intensity scan for ILPR-I3 bands and the fold protection for ILPR-I4 and ILPR-I3 for Br\textsubscript{2} footprinting in a 10 mM sodium phosphate buffer at pH 5.5 with 100 mM KCl………………………………………………………………………………………………67

Figure 4.10 Calculation of the unfolding rate constant (k\textsubscript{unfold}) at 0 pN for the intramolecular i-motif and the 45 pN population in the ILPR-I3/ILPR-I1 mixture…………………………………………………………………………………………….68

Figure 5.1 Chemical footprinting of a double-stranded 87-bp ILPR DNA at different pH and salt conditions at 23 °C. ……………………………………………………………….75

Figure 5.2 Single-molecule investigation of G-quadruplex/i-motif in the double-stranded ILPR DNA …………………………………………………………………………………………………79

Figure 5.3 Calculation of the number of nucleotides (N) involved in the tetraplex structures. …………………………………………………………………………………………………81

Figure 5.4 Histogram of change in contour length (∆L) in a pH 7.4 Tris buffer with 100 mM Li\textsuperscript{+} at 23 °C………………………………………………………………………………………………84
Figure 5.5 Kinetic analyses of G-quadruplex and i-motif in the double-stranded ILPR DNA at pH 5.5………………………………………………………………………………87

Figure 5.6 Major species under different pH and ionic conditions……………………90
LIST OF TABLES

Table 3.1 Summary of change in contour length ($\Delta L$), rupture force ($F$), free energy change of unfolding ($\Delta G$), and 286 nm CD melting temperature ($T_m$) at pH 5.5-7.0. .......................................................... 37

Table 4.1 Sequences of wild type ILPR-I4 and ILPR-I3, a scrambled sequence, and the mutants used in this study ................................................................. 53

Table 4.2 Calculation of contour length change ($\Delta L$) of the four possible candidates shown in Figure 4.6 ................................................................. 62

Table 5.1 Change in Contour Length ($\Delta L$), Rupture Force ($F_{\text{unfold}}$), Change in Free Energy of Unfolding ($\Delta G_{\text{unfold}}$), Bias of $\Delta G_{\text{unfold}}$, and Percentage Formation ($\%$ Formation) for fully folded tetraplex structures at different pH and ionic conditions at 23 °C ................................................................. 81

Table 5.2 The percentage of mechanical unfolding events for the control experiments in different buffers at 23 °C. ................................................................. 83
ACKNOWLEDGEMENTS

First and foremost I want to express my deep and sincere gratitude to my advisor Dr. Hanbin Mao for the guidance and support throughout my Ph D study. It is an honor and pleasure to be your first Ph D student. Your intelligence, open-mindedness, passion for Science and deep concern for the students has fascinated me and instilled the love for research in me. You showed me different ways to approach a scientific problem and the need to be persistent to accomplish any goal. Five years of your mentorship has made me a hard worker, confident researcher and a successful person. Thank you very much Dr. Mao.

Apart from my advisor, I would like to thank all of my committee members: Dr. Mietek Jaroniec, Dr. Soumitra Basu, Dr. Chanjoong Kim and previous committee member Dr. Simon Mwongela. Thank you to Dr. Basu for your motivating comments and suggestions on i-motif project. I would like to thank all of the committee members for your continued support throughout the course of my Ph D including your valuable comments and feedbacks on my research proposal and dissertation.

I thank the Department of Chemistry & Biochemistry for providing me an opportunity for the graduate study and allowing me to use the research facilities during the course of my Ph D study.
I thank Joseph Schonhofs from Basu lab. Thank you JD, I enjoyed working with you in the i-motif project. I would like to thank all of the previous and current members of Mao lab. I thank Dr. Mahmoud Chamsaz, Rabindra and Malcom for helping me to get familiarized to Mao lab during my first couple months. I thank Zhongbo, Deepak and Philip for day-to-day help, frequent discussion and valuable contribution to my research projects. I thank Yunxi, Zhe, Chiran, Prakash, Saurabh, and Sangeetha for being so good and helping me on my research projects. I also thank the REU students Ryan and Jay for their contribution to my research projects.

My special thanks go to my family members. I thank my mom, dad, my brothers and sister for their unconditional support throughout my education. I specially thank to my wife for being my motivation. Your willingness to my success helped me work hard and grow professionally. Finally, I thank my kids Sajal and Shayana. You are the heart of my inspiration and joy.

Soma Dhakal

10 January, 2013, Kent, Ohio
CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 G-QUADRUPLEX

DNA, a carrier of genetic information, is a right-handed double helical structure comprised of Watson-Crick base pairing: adenine (A) with thymine (T) and cytosine (C) with guanine (G), under ordinary physiological conditions.\(^1\) Shortly after the discovery of DNA double helix, non-canonical DNA structures consisting of non-Watson-Crick base pairing were evident in repetitive DNA sequences.\(^2\), \(^3\) The potential biological roles of the non-canonical DNA structures attracted tremendous research attention over the past few decades. It has been known that many promoter regions of DNA contain the repeating sequences capable of forming structures other than double helix. The DNA sequences with the contiguous guanines or cytosine repeats have the potential to form G-quadruplex (Figure1.1) and i-motif (Figure 1.2) structures respectively.\(^4\)–\(^7\) The formation of G-quartet, a building block of G-quadruplex structure, was first identified in 1962 on the basis of the aggregation of 5′ monophosphate guanosines to form a guanosine gel.\(^8\) X-ray fiber diffraction revealed the structural basis for the formation of the guanosine gel. This study identified the formation of a tetrameric unit of guanosines interconnected by Hoogsteen hydrogen bond in G-quartet structures. When two or more G-quartets are stacked on top of each other, a tetraplex structure, G-quadruplex, is formed (Figure 1.1). The structure is stabilized by monovalent metal cation, typically K\(^+\) or Na\(^+\), by coordination of the cation with O6 carbonyl of the guanines.\(^9\)
G-quadruplex structures can assume uni-, bi-, or tetramolecular types. Formation of a unimolecular structure requires a sequence with at least four G-rich stretches separated by at least one nucleotide (nt). Upon folding into a G-quadruplex structure, the nucleotides between two tandem G-rich repeats form loops. Based on the strand orientation, a G-quadruplex can be parallel (the same polarity for all strands) or antiparallel (each strand has opposite polarity with respect to the two adjacent strands). Figure 1.1 highlights general types of G-quadruplexes based on the strand orientation.

**Figure 1.1.** (A) Formation of G-quadruplex by stacking of G-quartets. (B) Three general types of G-quadruplexes based on DNA strand orientation.
To date, different G-quadruplex structures have been reported in human genome. G-quadruplex structures are highly polymorphic. A specific conformation depends on the DNA sequence, strand orientation, loop size, and solution factors such as cation species and molecular crowding conditions. In addition, different species can form under different reannealing speed, which reflects the difference between thermodynamic and kinetic stability of G-quadruplex structures.

1.2 I-MOTIF

DNA exists as a double-stranded helix in cell. Cytosine-rich strand complementary to the Guanine-rich strand is known to fold into i-motif structures under certain conditions. The discovery of i-motif with a detailed NMR structure was reported by Gehring et al in 1993. The structure consists of two parallel duplexes held together by hemiprotonated cytosine pairs (C:CH⁺). These C:CH⁺ pairs intercalate into each other with opposite polarity (Figure 1.2). As a requirement for the C:CH⁺ pairs, one of the cytosine must be hemiprotonated at the N3 position.

Until now, i-motif structures are the only known nucleic acid structures stabilized by base intercalation. The structure has two broad (~1.5 nm) and two narrow grooves (~0.7 nm). Depending on the number of DNA strands involved, it has been demonstrated that the DNA sequences with consecutive runs of cytosines can fold into intramolecular or intermolecular i-motif at acidic pH. A DNA strand with one, two
or four cytosine repeats can fold into tetrameric, dimeric, and monomeric (or intramolecular) i-motif structures, respectively. Like the G-quadruplex, i-motif structures also demonstrate structural polymorphism in terms of loop length and number of intercalated base pairs in different gene promoters. The structural polymorphism may offer an important opportunity for small molecule targeting. Besides the fully folded i-motif structures, a recent study reported the formation of intramolecular structure in less than three tandem C-rich repeats highlighting the potential competing structures. Although the formation of i-motif structure is favored at acidic pH, its formation has been reported at neutral and slightly basic pH under molecular crowding conditions, in a DNA template with negative superhelicity, and in the presence of carboxyl modified carbon nanotubes.
1.3 G-QUADRUPLEX AND I-MOTIF IN HUMAN ILPR

The insulin linked polymorphic region (ILPR) is located -363bp upstream of the human insulin gene on chromosome 11. The ILPR is polymorphic in G-repeat sequence and G-repeat length (400 to >8,000 nts), both of which are known to affect insulin expression. Research studies show that long ILPR poses higher transcriptional activity as compared to short ILPR. The ILPR is associated with many human diseases. The tandem G-rich and C-rich sequences in the human ILPR sequences have been studied for their capabilities to fold into stable G-quadruplex and i-motif structures respectively in vitro. The formation of these stable secondary structures in this region may influence the regulation of the insulin gene which ultimately can result in type I diabetes.

There are several reports pertaining to the structural studies of G-quadruplex and i-motif in human ILPR. Most of these studies are focused on the single unit of G-quadruplex or i-motif forming sequences. The presence of tandem G-rich tracts in ILPR certainly increases the opportunity for the formation of multiple G-quadruplex structures. Recent studies supported the high-order quadruplex-quadruplex interaction (QQI) between the G-quadruplex units in the long ILPR repeat (ILPR\textsubscript{n=8}, where ‘n’ is the number of GGGG repeats). The detailed understanding of QQI can assist in targeting these structures with small molecule drugs that recognize specific sites.
1.4 G-QUADRUPLEX AND I-MOTIF IN DUPLEX DNA

In most studies, sequences hosting G-quadruplex or i-motif structures are short ssDNA from promoter or telomere regions. The quadruplex structures in short ssDNA are without the long flanking dsDNA found in cells. Therefore, these in vitro studies may not directly represent the topology and stability of the tetraplexes in vivo. Since the DNA under the physiological condition is mostly present in a duplex form, it must be melted to provide the single-stranded region for the formation of secondary structures including DNA tetraplexes. The single-stranded region can form by local melting within the duplex DNA, or by unwinding of the duplex DNA during processes such as transcription or replication. Formation of secondary structures can compete with the duplex DNA, thus creating equilibrium between the two structures. The equilibrium is expected to shift towards G-quadruplex/i-motif structures if these structures form faster than the reannealing of duplex DNA.

There are ample investigations which have provided evidence of equilibrium between the G-quadruplex/i-motif structures and duplex DNA. It has been demonstrated that the equilibrium between the secondary structures and duplex DNA depends on several environmental factors such as temperature, pH, molar concentration of DNA, template superhelicity, sequence, and salt ions. For example, low pH (≤ 4.9) or high temperature (50–64 °C) favors the formation of G-quadruplex due to the melting of duplex DNA. Low molar concentration of DNA has been shown to favor G-quadruplex formation over DNA duplex even at neutral pH. A transcriptionally active region in a
DNA molecule provides negative and positive superhelicity of the molecule behind and ahead of the transcription bubble, respectively. The negative superhelicity in the G/C-rich region at a close proximity (1.2 kb upstream) of c-myc promoter is known to facilitate the formation of G-quadruplex/i-motif structures.\textsuperscript{32, 54} Although the formation of i-motif requires cytosine protonation, which mainly occurs at acidic pH, its formation at neutral pH has been demonstrated at negative superhelicity in c-myc promoter.\textsuperscript{32} In addition, intracellular proteins upon binding to DNA duplex may play important roles on the ssDNA-dsDNA equilibrium by stabilizing the G-quadruplex/i-motif structures. Although the formation of G-quadruplex or i-motif has been demonstrated in ssDNA, there are only scattered examples to show the folding of G-quadruplex or i-motif in duplex DNA. Depending on the buffer pH and metal ion present, this dissertation discusses the formation of G-quadruplex or i-motif structure in ILPR duplex DNA without the application of negative superhelicity or molecular crowding conditions.

### 1.5 Biological Roles of G-Quadruplex and I-Motif

Although there is a substantial understanding of G-quadruplex and i-motif structures, their mechanical stability, which potentially plays a key role in RNA transcription and DNA replication, is poorly studied. The repetitive DNA sequences at centromeres and telomeres are known to play an important role in the structure and function of eukaryotic chromosomes. The prevalence of the G/C-rich sequences in or near the gene regulatory regions including those of oncogenes,\textsuperscript{55, 56} highlight the
importance of addressing the biological role of these sequences. For example, the presence of G-/C-rich repeats at close proximity of transcription start sites (TSS) and their ability to form G-quadruplex and i-motif structures signifies the key role of these tetraplex structures in transcriptional regulation. Recent study revealed that within the 1kb upstream of TSS, ~43% of the human genes have a potential to form at least one G-quadruplex structure. With the discovery of proteins that can bind specifically to G-quadruplex and i-motif forming DNA fragments, the biological relevance of these structures have started to emerge in recent years. It becomes important to clarify the biological significance of the interconversion between a DNA duplex and a secondary structure. It has been postulated that such a conversion serves as a molecular switch to modulate transcription.

Located complementary to the G-rich strand, the C-rich strand presumably has an equal prevalence in human genes. To date, stable i-motif structures have been reported in several oncogenes (VEGF, RET, Bcl-2, Rb, c-Myc) and ILPR promoters. Formation of secondary conformations at close proximity of the TSS suggests that the alternative DNA structures serve a regulatory role in transcription. In cells, slow release of superhelicity of a DNA template by topoisomerases, provide an opportunity for the folding of secondary structures. In addition, with the discovery of proteins that can bind specifically to i-motif forming DNA fragments, the biological relevance of this structure has started to emerge in recent years. In addition to their regulatory roles in transcription and replication, the non-B DNA structures have been proposed to induce
genetic instability that may cause human diseases. Therefore, an understanding on the folding mechanism and stability of i-motif structures becomes biologically relevant.

With demonstrated biological roles of G-quadruplexes in the event of gene expressing and processing, it is important to determine the mechanical stabilities of G-quadruplex and the associated role of complementary C-rich strand, especially in the context of naturally occurring dsDNA. For an example, if a DNA polymerase is unable to unfold the G-quadruplex, it could result shortening of the ILPR region due to blockade/spillage during the replication process. This knowledge could be instrumental to evaluate the tetraplexes as therapeutic targets according to a new metrics that exploits the mechanical property of an interaction between secondary DNA structures and motor proteins such as kinesin, myosin, DNA and RNA polymerases. This dissertation focuses on the mechanical stability evaluation of i-motif and a partially folded structure in ssDNA, as well as the evaluation of the G-quadruplex and i-motif formed in dsDNA fragment in human ILPR.

1.6 MATERIAL APPLICATIONS OF G-QUADRUPLEX AND I-MOTIF

Besides their biological roles and potential target of therapeutics, G-quadruplex and i-motif structures have been exploited for a wide range of material applications. Several nanodevices have been built for electronics, biosensors, and biomimetic applications including their applications in supramolecular chemistry. The conformational polymorphism of G-quadruplex structures have been exploited for
developing nanodevices for different applications. For example, parallel G-quadruplex has been used to develop G-wire\textsuperscript{69} and K\textsuperscript{+} ion responsive nanochannels.\textsuperscript{70} Similarly, an antiparallel G-quadruplex has been explored for its catalytic activity for an aldol reaction between a ketone and a porphyrin linked aldehyde.\textsuperscript{71} Furthermore, the G-quadruplexes incorporated in DNA origami has been used for microRNA detection.\textsuperscript{72} Some other examples include the use of G-quadruplex in transmembrane ion transporter in phospholipid bilayer,\textsuperscript{73} and in nanomotors for transporting nanoparticles and microtubules.\textsuperscript{74} Majority of these applications employ the mechanical robustness of G-quadruplex. Therefore, understanding on the mechanical stability of G-quadruplex structures can provide useful insights to design nanomaterial devices.

Unlike the highly polymorphic G-quadruplex, the unique topology of i-motif structure is useful to make a nanodevice of controlled geometry.\textsuperscript{75} Nanodevices have been developed by exploiting the pH responsive folding of C-rich sequences into i-motif structures.\textsuperscript{76} For example, an i-motif based nonodevice has been successfully demonstrated to monitor intracellular pH.\textsuperscript{77} A colorimetric pH meter that works within the physiological pH range (6.0-7.2) has been shown to use i-motif DNA/AuNPs system.\textsuperscript{78} In addition, an i-motif based biosensor has been developed to distinguish single- and multiple-wall carbon nanotubes (CNTs) both in buffers and in cancer cell extracts.\textsuperscript{79} By selective separation of the CNTs, such a biosensor shows a capability to evaluate the toxicological effects of different CNTs. A size tunable pH responsive molecular motor that incorporates an i-motif forming sequence has been developed
recently with its potential application in surface patterning. Since these nanodevices all use i-motif forming sequences in the DNA scaffold, it is necessary to know the mechanical stability of the i-motif structure. The laser tweezers based characterization of i-motif and G-quadruplex structures discussed in this dissertation will provide information of these mechanical properties.

1.7 LASER TWEEZERS FOR SINGLE MOLECULE MANIPULATION

Understanding intracellular processes is dependent on the elucidation of the property of molecules involved in these processes. Since there are only a few copies of biomolecules, such as DNA, inside a cell, it becomes necessary to investigate individual molecules. However, it is difficult to study individual molecules using conventional bulk approaches. The ensemble techniques routinely used for the characterization of G-quadruplex and i-motif structures include NMR, X-ray, fluorescence, chemical footprinting, circular dichroism (CD), and UV-vis spectroscopy. Although used widely for tetraplex studies, these techniques rely on the ensemble-average signal and lack the capability to identify subpopulations present in minority. In addition, the study of the longer G/C-rich sequences that can form multiple units of tetraplex may be difficult to achieve by standard bulk techniques such as NMR and X-ray crystallography due to overlapping signals in NMR and difficulty to achieve a pure crystal for X-ray study. To address these inherent problems of the ensemble average
techniques, new techniques have been developed over time. The most promising technique to become popular in the last two decades is the single-molecule method due to its capability to provide insights into the subpopulations. Compared to the ensemble average methods, investigation of individual molecules allows more direct measurement of microscopic properties of the molecules such as mechanical stability of subpopulations. In addition, measuring the individual events allows determining dynamics of molecules which is obscured in ensemble measurements.

The discovery of highly sensitive fluorescence microscopy in 1990s enabled to track single molecules in their natural environments. Further improvements of the single-molecule technique enabled researchers to address important questions in biology. Although highly sensitive, single-molecule FRET (Fluorescence Resonance Energy Transfer) suffers from background signal and quenching of fluorophores. In addition, the fluorophore labeling not only increases the experimental costs but also can alter the property of the biomolecules. Force based techniques such as atomic force microscopy (AFM) and laser tweezers based single-molecule methods are ideal to overcome the limitations of both bulk techniques and single-molecule FRET. Unlike FRET, the force based techniques are generally not influenced by the solution conditions which make these techniques superior over other bulk techniques.

The laser-based optical trapping was pioneered by Ashkin in the early 1970s. He demonstrated that a laser focus could trap micron-sized dielectric particles in water and air. This discovery led to the development of single-beam gradient force optical trap,
or laser tweezers.\textsuperscript{82, 83} Due to the unique capability of laser tweezers to manipulate single particles, they became a very useful tool in physics and biology for a wide-range of applications, such as in the study of individual bacteria, viruses, proteins, and DNA structures.\textsuperscript{84, 85} Laser tweezers have another ability to apply and measure picoNewton (pN) forces relevant to the load force of motor proteins such as DNA and RNA polymerases. In addition, this tool can detect the displacement of trapped particle at sub-nanometer resolution. As a result of these outstanding properties, they are now widely used in biophysical characterization of polymers, biopolymers, and molecular motors.\textsuperscript{86, 87} This dissertation is focused on laser tweezers based mechanical characterization of G-quadruplex and i-motif structures in human ILPR at the single-molecule level.
2.1 MATERIALS AND METHODS

2.1.1 DNA Samples. Oligonucleotides were purchased from Integrated DNA Technologies (www.idtdna.com) and purified by denaturing PAGE (10%). Concentration of dsDNA samples were calculated using the Uv-vis absorbance at 260 nm and by band-intensity analysis on agarose gel using the Kodak Digital Camera Software (Eastman Kodak Company, Rochester, NY). In case of oligonucleotides, the extinction coefficients were calculated by nearest neighbor approximation. Unless specified differently, all of the chemicals were purchased from VWR.

Laser Tweezers Instrument. The schematic of the laser tweezers instrument is shown in Figure 2.1. Briefly, a diode pumped solid state (DPSS) laser (1064 nm, 4 W, CW mode, BL-106C, Spectra-physics) was used as a trapping laser. $P$ and $S$ polarized laser light from the same laser source constituted two traps. The $S$ polarized light was controlled by a steerable mirror (Nano-MTA, Mad City Laboratories) at a conjugate plane of the back focal plane of a focusing objective (Nikon CFI-Plan-Apochromat 60×, NA 1.2, water immersion, working distance ~320 µm). The exiting $P$ and $S$ polarized beams were collected by an identical objective and detected by two position sensitive photodetectors (PSPD, DL100, Pacific Silicon Sensor) separately. The force of the laser trap was calibrated by the Stokes force and thermal motion measurement. Both
methods yielded a similar trap stiffness of \( \sim 307 \text{ pN/} (\mu\text{m} \times 100\text{mW}) \) (for 0.97 \( \mu \text{m} \) diameter polystyrene beads, Bangs Laboratory, Fishers, IN).

**Figure 2.1** Schematic of the laser tweezers instrument. PBS, polarized beam splitter; OBJ, objective and PSPD, position sensitive photo detector.

**2.1.2 DNA Construct for ILPR i-motif and telomeric G-quadruplex in ssDNA.** The flow chart of making a DNA construct (strategy for making i-motif DNA constructs is shown for example) for single-molecule study is shown below. The DNA construct was comprised of three fragments: two dsDNA spacers at two termini and one DNA fragment containing the sequence of interest (either an ssDNA or dsDNA) in the middle. The 651 bp or 2028 bp dsDNA spacer was labeled with biotin, which was introduced through a
biotinylated primer (Integrated DNA Technologies, IDT, Coralville, IA) during PCR amplification of pEIB (966bp)\textsuperscript{91} or pBR322 plasmid template (NEB) respectively. This spacer was digested with XbaI restriction enzyme (New England Biolab, NEB). The 2690bp DNA handle was gel purified using a kit (Midsci, St. Louis, MO) after SacI and EagI digestions of a pEGFP plasmid (Clontech, Mountain View, CA). This spacer was subsequently labeled by digoxigenin (Dig) at 3′ end using 18 μM Dig-dUTP (Roche, Indianapolis, IN) and terminal transferase (Fermentas, Glen Burnie, MD). For i-motif construct, the middle i-motif forming fragment was constructed by annealing an oligonucleotide, 5’-CTAGACGGTTGAAATACCCAGATGCGTGTCCCCACACCCCTGTCCCTGTCACACACAGCTGTGCGGTATTTCCACCGTGCCAGCAAGAGTCAGCCCAGCAGCTGTC, with two other oligonucleotides, 5’-CGCATCTGTGCGGTATTTCCACCGTGCCAGCAAGAGTCAGCCCAGCAGCTGTC and 5’-GGCCGACCGCGCTGGGCTACGTCTTGGCGTC at 97°C for 5 minutes and slowly cooled to room temperature for 6h. Similarly the underlined sequence above was replaced with 5’-TTAGGGTTAAGGTTAGGGTTAGGTTA for making telomeric G-quadruplex construct. The i-motif or G-quadruplex forming fragment was ligated with the 651 bp or 2028 bp DNA handle at one end, followed by a second ligation with the 2690 bp DNA handle using T4 DNA ligase (NEB). The final construct was purified by ethanol precipitation. The DNA pellet was dissolved in water and stored at -80°C. In
Figure 2.2 Flow chart of making DNA constructs for single-molecule experiment. The length of the DNA fragments is shown in base pair (bp). Plasmids are shown in concentric circles. For clarity, overhang sequences after enzyme digestion are shown.

In the case of i-motif, the mutant DNA construct containing TGTCCCAACACCCCTGTCCCAACA in place of the i-motif forming sequence was
prepared using the same procedure. The same sequence was used to prepare the DNA construct for the formation of secondary structure in three tandem C-rich repeats.

2.1.3 DNA construct for i-motif and G-quadruplex in ILPR duplex. Two DNA oligonucleotides containing ILPR G-quadruplex (5'-CACA GGGG TGT GGGG ACA GGGG TGT GGGG T) and i-motif (5'-CTAGA CCCC ACA CCCC TGT CCCC ACA CCCC TGTGGTAC) forming sequences were annealed with an equimolar ratio to form dsDNA in a 10 mM Tris (pH 8.0) buffer by heating at 97 °C for 10 min, followed by slow cooling to 25 °C in ~6 hours. At each end of the dsDNA, an XbaI or a KpnI restriction site was introduced to allow the cloning of the DNA into the polycloning site of the plasmid pFOXCAT-362hIns (Gift from Michael German Lab).92 The plasmid with the ILPR G-quadruplex/i-motif insert was then transformed to E. coli (GM2163 strain, Fermentas) for amplification. Bacterial cells were collected through centrifugation after overnight shaking in LB growth medium, and the plasmid was extracted using a Maxiprep® kit (Qiagen, Valencia CA). The ILPR G-quadruplex/i-motif insert was validated by DNA sequencing (DNA Sequencing Facility, University of Maine). The plasmid containing the ILPR G-quadruplex/i-motif forming sequence was digested with AflIII (NEB), labeled with biotin-dUTP (eENZYME, Gaithersburg, MD) using a Klenow enzyme (NEB) followed by ethanol precipitation. The DNA was then digested by SacI (NEB) and the larger DNA fragment (~4kb) was purified by an agarose gel. The fragment DNA was labeled with digoxigenin at the SacI end with dig-dUTP (Roche,
Indianapolis, IN) using Terminal Transferase (Fermentas, Glen Burnie, MD). Finally, the sample was purified by ethanol precipitation. To prepare DNA construct without G-quadruplex/i-motif forming sequences, the pFOXCAT-362hIns plasmid without insert was digested with AflIII and labeled with biotin-dUTP, followed by digestion with SacI and labeled with dig-dUTP as described above.

2.1.4 Single-Molecule Experiment. Anti-Dig antibody coated polystyrene beads (diameter: 2.17 \( \mu \text{m} \), Spherotech, Lake Forest, IL) were incubated with diluted DNA constructs obtained above (~0.43ng/uL) in 100 mM salt (KCl or LiCl), 10 mM buffer (Tris, PBS or MES) at a given pH for 1 h at 23 °C to attach the DNA construct via the Dig/anti-Dig complex. Beads coated with streptavidin (diameter: 2.10 \( \mu \text{m} \), Bangs Laboratory) were dispersed into the same buffer before injected into the reaction chamber. These two types of beads were trapped separately using two laser traps. To immobilize the DNA construct between the two beads, the bead already attached with DNA construct was brought close to the bead coated with the anti-Dig antibody by the steerable mirror. Once the DNA tether was trapped between the two beads, the NanoMTA steerable mirror that controls the anti-Dig coated bead was moved away from the streptavidin-coated bead with a loading speed of ~5.5 pN/s. The secondary structure formed in the DNA molecule was unfolded when tension inside the tether was gradually increased. Rupture events with sudden change in the end-to-end distance were observed in the force-extension \((F-X)\) curves due to unfolding the secondary structure. The force-extension \((F-X)\) curves were recorded at 1000 Hz using a Labview\textsuperscript{®} program. The raw
Figure 2.3 Schematic representation of sample preparation and mechanical unfolding of DNA structures in single-molecule experiment.

Data were filtered with a Savitzky-Golay function with a time constant of 10 ms (or 50 ms in molecular crowding experiment) respectively for buffers using a Matlab program (The MathWorks, Natick, MA). Single tether was confirmed by the plateau at ~65 pN or a single breakage event when the DNA was overstretched. Change in contour length ($\Delta L$)
due to the rupture events was calculated by the two data points flanking the rupture events using an extensible Worm Like Chain (WLC) equation\(^9\) (equation 2.1):

\[
\frac{X}{L} = 1 - \frac{1}{2} \left( \frac{k_B T}{FP} \right)^{\frac{1}{2}} + \frac{F}{S} \quad \text{(2.1)}
\]

Here \(X\) is the end-to-end distance, \(k_B\) is the Boltzmann constant, \(T\) is absolute temperature, \(P\) is the persistent length (51.95 nm\(^9\)), \(F\) is force and \(S\) is the elastic stretch modulus (1226 pN\(^9\)).

2.1.5 Calculation of change in contour length. After each \(F-X\) curve was split into the stretching (red) and relaxing (black) traces (Figure 2.4), the change in extension (\(\Delta X\)) was calculated by subtracting the extension of the former trace from the latter at the unfolding force. The resulting \(\Delta X\) was then converted to the change in contour length (\(\Delta L\)) using the Worm-Like-Chain (WLC) model at a particular force (\(F\)) using equation 2.1 above.

![Figure 2.4](image)

**Figure 2.4** A typical force-extension (\(F-X\)) curve for the unfolding of i-motif structure at pH 5.5.

2.1.6 Deconvolution of Populations. In a histogram with two populations, the overall population was fit with a two-peak Gaussian function. To account for the stochastic
behavior of individual molecules, the population under the intersection region was randomly assigned to one of the populations based on the ratio determined by the Gaussian fitting. The change in contour length ($\Delta L$) and unfolding force histograms were plotted separately for individual populations according to this random assignment.

2.1.7 Control Experiments for Single-Molecule Study. A number of control experiments were performed at the single-molecule level to validate that the unfolding events observed were due to the unfolding of G-quadruplex or i-motif structures. These experiments were performed by choosing either a muted sequence or a buffer condition in which the formation of G-quadruplex or i-motif is discouraged. In addition, mechanical unfolding was performed on the DNA construct containing dsDNA handles only.

2.1.8 Calculation of Percentage Formation. The percentage formation of the secondary structures were calculated based on the number of pulling curves recorded at a given buffer condition. The percentage formation the folded structure was calculated as the ratio of the pulling curves with the folded structure vs total pulling curves. To avoid repetitive counting, the subsequent pulling curves of the same DNA construct were counted only once in this calculation.

2.1.9 Calculation of $\Delta G_{\text{unfold}}$. The free energy difference for unfolding of a secondary structure ($\Delta G_{\text{unfold}}$) was calculated according to Jarzynski’s equality equation (equation 2.2) for non-equilibrium systems.\textsuperscript{94, 95}
\[
\Delta G = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp \left( - \frac{W_i}{k_B T} \right) \tag{2.2}
\]

where \( N \) is the number of observations in the experiment and \( W \) is the non-equilibrium work done during unfolding of the i-motif, which is equivalent of the hysteresis area between stretching and relaxing force-extension (\( F-X \)) curves. In the case where two populations were observed in \( \Delta L \) or unfolding force histograms, the \( \Delta G_{\text{unfold}} \) was calculated using equation 2.2 for each population after the population deconvolution described above. The bias of the \( \Delta G_{\text{unfold}} \) was calculated from the unfolding work histograms as described in the literature.

2.1.10 CD Spectroscopy. Oligonucleotides samples purchased from Integrated DNA Technologies were purified by denaturing PAGE and prepared in a concentration of 5 µM in appropriate buffer at a given pH with 100 mM KCl or LiCl. The samples were then heated at 97 °C for 10 minutes and immediately cooled using an ice-water bath. The CD spectra were taken in a 1 mm quartz cuvette at room temperature with a Jasco-810 spectropolarimeter (Easton, MD). The reported spectra are the average of three scans with a scan rate of 50 nm/ min. The spectrum of each scan was subtracted from a buffer and salt only baseline and smoothed using a Savitzky-Golay function. For CD melting, the signal was measured at 286 nm for i-motif and 285 nm for ILPR-I3 as the temperature was changed at a rate of 0.5 °C/minute using a Jasco (model PFD-425S) peltier temperature controller.
2.1.11 UV Spectroscopy. The basic procedure of sample preparation in UV experiments was the same as the CD experiment described above. All UV experiments were performed in 10 mM sodium phosphate buffer with 100 mM KCl at pH 5.5. In the thermal analysis of the ILPR-I3, UV measurements were performed over the concentration range of 5 to 100 µM. In the thermal analysis of the mutants, 10 µM DNA was used. The UV-melting experiments were performed at 295 nm with a heating rate of 0.5 °C/min in a quartz cuvette (1 cm optical path length) using a Varian Cary 300 spectrophotometer. To correct for buffer signals, the absorbance of the buffer-only solution was subtracted from the UV melting curves. All melting curves were baseline corrected, normalized and plotted as the fraction folded vs temperature and the transition temperatures ($T_{1/2}$) were calculated for melting and reannealing processes as described elsewhere.98-100

2.2.12 Preparation of 5' end radiolabeled DNA. DNA samples were radiolabeled at the 5' end by incubating the DNA with T4 polynucleotide kinase (Promega) and [γ-P32] ATP (Perkin Elmer) and purified either using denaturing PAGE or G-25 columns (GE Healthcare, Buckinghamshire, UK).

2.1.13 Electrophoretic Mobility Shift Assay (EMSA). The radiolabeled oligos (1 µM final concentration) were prepared in a 10 mM sodium phosphate buffer (pH 5.5) with 100 mM KCl. The oligo samples were heated to 95 °C for 10 min and transferred to an ice-water bath for fast cooling and analyzed in 10% native PAGE gels. The sodium phosphate running buffer with 100 mM KCl was changed with fresh buffer every 45 min
to maintain the salt concentration during the electrophoresis. A control experiment was performed under the denaturing condition (7M urea, 10% denaturing PAGE).

2.1.14 Bromine Footprinting of ssDNA. To specifically probe cytosine residues in the footprinting experiment, bromine footprinting was was carried out, in accordance with the published procedure.\textsuperscript{101,102} Solutions containing the 5′-labeled C-rich strand (5′-TGTCCCCACACCC C TGTCCCCACACCCCTGT were prepared in 10 mM sodium phosphate buffer (pH 5.5), with the addition of 1 µM unlabeled oligonucleotides. The samples were then heated at 97 °C for 10 minutes and quickly cooled by submersion in an ice-water bath. Then to a 50 µl sample, 1 µl of 20 mM KBr was added, which was immediately followed by the addition of 1 µl of 10 mM KHSO\textsubscript{5}. The reaction mixture was then incubated for 3 minutes at room temperature. To stop the reaction, 25 µl of the reaction mix was added to 100 µl of stop buffer (1 mg/ml sheared salmon sperm DNA (VWR), 300 mM NaCl, and 4 mM HEPES). The DNA was then precipitated using 100% ethanol and the resultant pellet was washed twice with 70% ethanol. The DNA was then cleaved with piperidine and the resulting fragments were separated on a 10% denaturing polyacrylamide gel (1:19 bis:acrylamide). The gel was dried on Whatman\textsuperscript{TM} paper, exposed to a phosphorimager screen, and scanned in a Typhoon 8600 instrument (Molecular Dynamics). Footprinting gel images were quantified using ImageJ\textsuperscript{TM} software (http://rsbweb.nih.gov/ij/index.html). The intensity of each band was normalized with that of the cytosine in the ACA loop after background correction at pH 5.5 and 7.0 separately. The fold protection of each nucleotide was then calculated as the ratio
between the normalized band intensity at pH 7.0 and that of corresponding band at pH 5.5.

2.1.15 DMS and Bromine footprinting of dsDNA. First, a 87-mer ssDNA strand containing an ILPR G-quadruplex forming sequence (underlined), 5'- CTA GAC GGT GTG AAA TAC CGC ACA GAT GCG ACA GGGG TGT GGGG ACA GCC AGC AAG ACG TAG CCC AGC GCG TC (G-rich strand) or an ILPR i-motif forming sequence (underlined), 5'- GAC GCG CTG GGC TAC GTC TTG CTG GC TGT CCCC ACA CCCC TGT CCCC ACA CCCC TGT CGC ATC TGT GCG GTA TTT CAC ACC GTC TAG (C-rich strand), was radiolabeled at the 5' end purified as described above.

For the DMS footprinting, the radio-labeled G-rich strand was mixed with its unlabeled complementary C-rich strand in equimolar ratio (1 µM each in 30 µL) in a 10 mM Tris (pH 7.4) or in a 10 mM MES (pH 5.5) buffer with 100 mM KCl, 100 mM LiCl, or without salt. The mixtures were heated at 95 °C for 10 minutes and quickly quenched by submersion in an ice-water bath or in a water bath set at 30 °C. Then 0.3 µL DMS was added to the above sample and incubated for 1.5 min at room temperature. The reaction was stopped by adding 300 µL of a stop buffer (250 µg/mL salmon sperm DNA, 30% β-mercaptoethanol, and 300 mM sodium acetate) followed by addition of 750 µL of absolute ethanol. For bromine footprinting, the radio-labeled C-rich strand and unlabeled complementary G-rich strand were prepared separately in a 10 mM sodium phosphate buffer (pH 5.5 or 7.0) supplemented with 100 mM KCl or LiCl. The samples were
heated at 95 °C for 10 min followed by rapid cooling in a water bath set at 30 °C. This was followed by mixing of the complementary strand at equimolar concentration (1 µM) in a 30 µL reaction volume. The samples were incubated at the same temperature for a specific time (15 min to 5 days). The cytosine-specific cleavage was performed by incubating the DNA samples with molecular bromine generated in situ from the reaction between KBr (0.6 µL, 20 mM) and KHSO₅ (0.6 µL, 10 mM). The reactions were performed for 10 min at room temperature, terminated by adding a stop buffer (250 µg/mL sheared salmon sperm DNA, 300 mM NaAc, and 4 mM HEPES) followed by addition of 750 µL absolute ethanol. The subsequent steps (ethanol precipitation, piperidine cleavage, vacuum drying, and band intensity analysis) were performed as described in the section ‘bromine footprinting of ssDNA’ above with the difference that the bands were resolved on a 6 % denaturing PAGE.

To prepare the GA ladder, radio-labeled 87 mer G-rich strand was mixed with salmon sperm DNA (0.1µg/µL) in 10 µL reaction volume. One microliter of 1 M formic acid (pH 2.0) was added to the above reaction mixture and incubated at 37 °C for 30 min, followed by piperidine cleavage (150 µL, 10% piperidine) at 90 °C for 30 min. The sample was cooled in an ice-water bath for 5 min and precipitated with n-butanol. The DNA pellet was resuspended in 1% SDS (150 µL), followed by another precipitation with n-butanol. Finally, the DNA pellet was resuspended in a loading buffer for PAGE analysis.
To calculate the fold protection for both DMS and Br₂ footprinting experiments, net intensity of the bands in the four G4/C4-tracts or the two cytosine bands close to the C4 regions in each lane was measured, normalized against the intensity of the guanine band in the TGT loop in DMS footprinting or the cytosine band in the ACA loop in Br₂ footprinting. Fold protection was obtained by dividing the band intensity of corresponding G/C in the control lane (Li⁺ lane for the DMS footprinting at the same pH or the pH 7.4 lane for the Br₂ footprinting) by the lane of interest. The fold protection for a given time in kinetic experiments was averaged from the four G4/C4 tracts.
CHAPTER 3

COEXISTENCE OF AN ILPR I-MOTIF AND A PARTIALLY FOLDED STRUCTURE WITH COMPARABLE MECHANICAL STABILITY REVEALED AT THE SINGLE MOLECULAR LEVEL

3.1 INTRODUCTION

In recent years, non-B DNA structures have attracted intensive research attention due to the likelihood that these structures may be responsible for a variety of human diseases such as Fredrick’s ataxia and Huntington disease. Non-B DNA structures include G quadruplex, Z DNA, H DNA, cruciform DNA, and i-motif, to name just a few. Experiments in vitro have suggested that these structures can regulate important biological processes including replication and transcription. Evidence from recent research has also indicated that these non-B DNA structures can cause DNA double strand breakage (DSB), which increases mutations for diseases. Computer based programs have revealed that non-B DNA forming sequences are widely dispersed throughout the human genome. For example, a total of 188,836 G rich sequences have been found to be capable of forming G quadruplexes, whereas a frequency of 1/3,050 bp and 1/49,400 bp exist for potential Z and H DNA in human genome, respectively. Some of these non-canonical DNA structures have been confirmed in vivo.

Among these non-B DNA structures, G quadruplex and i-motif both contain four DNA strands. Each G-quadruplex forming sequence has a complementary C-rich
sequence capable of adopting an i-motif structure. Thus, the prevalence of G quadruplex forming sequence in the human genome also suggests the wide-spread occurrence of i-motif in the genome. In contrast to the G quadruplex, which is independent of pH and readily forms at physiological pH, the formation of i-motif requires hemiprotonated cytosine-cytosine pairs\(^{112}\) (Figure 3.1A) and therefore, is pH dependent. The optimal i-motif assembly occurs at pH 5.5, a value close to the pK\(_a\) for free cytosines.\(^{113}\) So far, only scattered evidence\(^{42, 114}\) has indicated the presence of i-motif structure at neutrality in vitro. In vivo, however, negative superhelicity\(^{32}\) and molecular crowding\(^{115, 116}\) may facilitate i-motif formation. In addition, formation of either an i-motif or a G-quadruplex can leave its complementary DNA as a free strand, thereby facilitates the assembly of the other.

Recent discovery of proteins interacting with potential i-motif structures suggests that, similar to the G-quadruplexes, i-motifs may also have biological functions. Mergny and coworkers observed two nuclear proteins, hnRNP-K, a transcription factor of the \(c\)-\(myc\) gene, and ASF/SF2, a splicing factor, can bind specifically to the C-rich strand of human telomeres.\(^{57}\) In an ILPR sequence similar to what has been used here, Gupta and coworkers\(^{42}\) demonstrated that \(E.\) \(coli\) SSB protein binds more efficiently to the C-rich strand compared to other ssDNA regions at neutral pH. In human telomeres, Manzini and colleagues\(^{58}\) have shown that Hela nuclear extract contains a protein that specifically binds to single stranded CCCTAA repeats. The specific binding of i-motif forming sequences to various proteins implies the potential biological function of the i-motifs that
may form in these sequences. Another aspect of biological relevance of the i-motif comes from the potential interaction between this structure and the G quadruplex, the latter of which has shown critical involvement in many biological processes.\textsuperscript{32} In accordance with the prospect of biological functions by i-motifs, several laboratories have started to investigate DNA i-motif analogs based on phosphorothioate or peptide backbones for potential pharmaceutical applications.\textsuperscript{117}

Since a significant portion of i-motif hosting sequences is located within or downstream of promoter regions,\textsuperscript{55, 118} there are ample opportunities for the tetraplex to interact with RNA polymerase (RNAP). As a motor protein, RNAP exerts a maximal load force, beyond which transcription stalls.\textsuperscript{119, 120} From mechanical perspective alone, a non-B DNA structure with a mechanical stability higher than this maximal load force can stall RNAP. The mechanical stability of such a structure can be determined by single molecular methods such as laser tweezers.\textsuperscript{121}

Therefore, we use laser-tweezers to investigate the structures formed in C-rich regions. To the best of our knowledge, i-motifs have never been investigated at the single molecular level. Compared to the bulk methods for i-motif investigation, such as circular dichroism (CD),\textsuperscript{122, 123} gel electrophoresis,\textsuperscript{124} ultraviolet (UV) absorbance,\textsuperscript{124} X-ray,\textsuperscript{125} and NMR,\textsuperscript{126-128} single molecular methods can reveal biomolecular structures in a highly dynamic fashion. The method is very sensitive in identifying small populations, such as intermediates during a folding process.\textsuperscript{129} Compared to other single molecule techniques such as fluorescence,\textsuperscript{130} the force based approaches do not require bulky
fluorophores that may alter native structures. In addition, they can provide mechanical information of structures,\textsuperscript{45, 121} which is important not only for transcription, but also for other processes catalyzed by motor proteins such as DNA polymerases.

We choose to investigate non-B DNA structures formed in the most prevalent variant, 5'-\((\text{TGTCCCAACACCCC})_2\), in the insulin linked polymorphic region (ILPR). The region is known to affect the production of Human insulin protein.\textsuperscript{131} The C rich DNA sequence used in our experiment is terminated by two dsDNA spacers. Similar to the internal loops in a DNA tetraplex,\textsuperscript{50} terminal spacers have also demonstrated their critical role in the tetraplex conformation.\textsuperscript{132} Thus, our approach complements very well with bulk methods that have limited sensitivity to long terminal spacers. Since the DNA regions in which tetraplexes are susceptible to form are almost always flanked by double stranded spacers,\textsuperscript{106} we expect our method is uniquely equipped to interrogate non-B DNA structure in a situation closer to \textit{in vivo} conditions.

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

3.2.1 Histograms of change in contour length ($\Delta L$) show coexistence of an i-motif and a partially folded structure.

To reduce the steric hindrance between the possible folded structure and dsDNA handles, we incorporated a wild type spacer, TGT, at the 3' end of the most predominant variant sequence in the human ILPR, 5'- (TGTCACACACC)2. After tethering the DNA construct between the two optically trapped beads via biotin/streptavidin and digoxigenin (Dig)/anti-Dig antibody complexes, respectively (Figure 3.1A, iii), we moved away one of the laser traps and recorded force-extension ($FgX$) curves of the DNA construct at pH 5.5 (Figure 3.1B). Using a worm-like-chain equation (Equation 2.1), we could obtain the change in contour length ($\Delta L$, see Materials and Methods) from the force extension curves that contain rupture events of folded structures. After correction of end-to-end distance of a folded species, this $\Delta L$ reflects the contour length of the structure.\textsuperscript{121} When we plotted the $\Delta L$ histogram, to our surprise, two populations were observed (Figure 3.1B inset). The bigger population has $\Delta L$ of 10.4 ($\pm 0.1$) nm (see Figure 3.1B for a representative $FgX$ curve in green); whereas the smaller population has $\Delta L$ of 5.1 ($\pm 0.5$) nm (see Figure 3.1B for a representative $FgX$ curve in red). The value of 10.4 nm is within the expected range for a folded structure with 25 nucleotides.\textsuperscript{121,133-137} Therefore, we ascribe this population to an i-motif structure. Previous NMR studies\textsuperscript{42} have shown
Figure 3.1 (A) Schematic of (i) i-motif, (ii) chemical structure of a hemiprotonated cytosine-cytosine pair, and (iii) experimental set up. (B) Typical force extension curves from the same ILPR sequence, 5’-(TGTCGCCACACCCCC)2-TGT, at 23 °C and pH 5.5. Green (the right curve) and red (the left curve) curves represent unfolding of two populations with change in contour length (ΔL) of 10.4 and 5.1 nm, respectively. The sudden drop in force (around 27.5 and 26.0 pN, indicated by dashed circles) in the extending curves indicates unfolding events. Black line is the WLC fitting. Curves are shifted in x axis for clarity. Inset is the ΔL histogram fitted by a two-peak Gaussian.

A sudden drop in force was clearly seen in the force-extension curve, which indicated the unfolding of a structure formed in the DNA construct. The folded structure was confirmed by CD-286 nm (Figure 3.2A) and UV-295 nm (Figure 3.3) melting experiments, both of which showed a sigmoidal transition with a \( T_m \) of ~37 °C at pH 5.5 (see Table 3.1).

That i-motif forms in a similar sequence at pH 5.5 in vitro. CD spectra on the same DNA fragment at pH 5.5 clearly revealed a peak at ~286 nm and a trough at ~264 nm (Figure
3.2B), both of which are characteristic of an i-motif structure. In addition, the bromine footprinting data in Figure 3.2C demonstrated the protection of four C tracts (C4-C7, C11-C14, C18-C21, and C25-C28) from bromine initiated cleavage at pH 5.5. This result is consistent with the formation of an i-motif utilizing the four C tracts in the sequence.

**Figure 3.2** CD experiments and Br₂ footprinting of the ILPR C-rich sequence, 5’-(TGTCCCCACACCCC)₂TGT. (A) CD (286 nm) melting curves of the i-motif forming sequence from pH 5.5 to 6.5. The spectra were baseline corrected and normalized as the fraction of folded form. At pH 7.0 no clear transition was observed (not shown). (B) CD spectra acquired from pH 5.5 to 7.0 at 23 °C. CD experiments were performed with 5 μM oligonucleotides in 10 mM sodium phosphate buffer (pH 5.5, 6.0, 6.5, or 7.0) and 100 mM KCl. (C) Br₂ footprinting at pH 5.5 and 7.0. The trace on the left side of the gel depicts band quantitation. The direction of the arrow at the bottom of the gel indicates increased intensity of the bands. The plot on the right shows the fold protection of
individual bands at pH 5.5 compared to those at pH 7.0. The band intensities at the 3’ end (see *) are not reliable since they are close to the uncut oligonucleotides.

Figure 3.3 UV (295 nm) melting for the i-motif forming sequence 5’-(TGTCCCCACACCCC)₂TGT. Experiment was performed in 10 mM sodium phosphate buffer pH 5.5, 100 mM KCl with an oligonucleotides concentration of 5 µM. The melting curve was corrected for the sloping baselines and the y-axis was normalized as the fraction in the folded form. For comparison, CD (286 nm) melting performed under identical pH, buffer condition and oligonucleotide concentration was plotted together with UV (295nm) melting. Both CD and UV melting showed similar transition temperature ($T_m \sim 37^\circ C$).
Table 3.1 Summary of change in contour length (ΔL), rupture force (F), free energy change of unfolding (ΔG), and 286 nm CD melting temperature (Tm) at pH 5.5-7.0. The values in paranthesis are standard deviations.

<table>
<thead>
<tr>
<th>pH</th>
<th>ΔL Part. Fold. (nm)</th>
<th>F Part. Fold (pN)</th>
<th>ΔG Part. fold. (kcal/mol)</th>
<th>ΔL i-motif (nm)</th>
<th>F i-motif (pN)</th>
<th>ΔG i-motif (kcal/mol)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>5.1 (±0.5)</td>
<td>26.5 (± 0.5)</td>
<td>10.5 (± 0.1)</td>
<td>10.4 (±0.1)</td>
<td>26.2 (± 0.3)</td>
<td>16.6 (± 0.7)</td>
<td>37.0 (±1.3)</td>
</tr>
<tr>
<td>6.0</td>
<td>4.9 (±0.3)</td>
<td>25.5 (± 1.3)</td>
<td>10.7 (± 0.1)</td>
<td>10.2 (±0.1)</td>
<td>23.9 (± 0.7)</td>
<td>16.2 (± 0.3)</td>
<td>33.7 (±0.5)</td>
</tr>
<tr>
<td>6.5</td>
<td>5.2 (±0.3)</td>
<td>26.5 (± 0.9)</td>
<td>10.4 (± 0.6)</td>
<td>9.9 (±0.2)</td>
<td>21.9 (± 0.4)</td>
<td>15.1 (± 0.1)</td>
<td>23.4 (±0.7)</td>
</tr>
<tr>
<td>7.0</td>
<td>4.7 (±0.3)</td>
<td>25.9 (± 2.6)</td>
<td>9.9 (± 0.3)</td>
<td>Not enough data</td>
<td>Not enough data</td>
<td>Not enough data</td>
<td>No melting</td>
</tr>
</tbody>
</table>

Structure 1:
ΔG°=0.10 kcal/mol
Tm=23.1°C

5’-G-T-C-C-C-C-C-Å-T-C-C-C-A-C-C-A-C-C-C-C-T-G-T3’

Structure 2:
ΔG°=0.10 kcal/mol
Tm=23.0°C

5’-G-G-T-C-C-C-C-C-A-C-C-A-C-C-C-C-T-G-T-C-C-C-C-Å-T3’

Structure 3:
ΔG°=0.15 kcal/mol
Tm=22.7°C

5’-G-G-C-C-C-C-C-T-G-T-C-C-C-C-A-C-C-C-C-T-G-T3’

Structure 4:
ΔG°=0.15 kcal/mol
Tm=22.7°C

5’-G-G-C-C-C-C-C-C-T-G-T-C-C-C-C-C-T-G-T3’

Figure 3.4 All of the stable structures in the sequence 5’-(TGTCCTCCCACAC CCC)2TGT at 25 °C. The four structures were obtained using Mfold program with [Na+] =100
mM (Na\(^+\) and K\(^+\) are considered equivalent in hairpin folding\(^{140}\)). Free energy and melting temperature are indicated for each structure. Single dash indicates phosphate backbones while double dashes indicate the hydrogen bonds in the hairpin stem.

Compared to the $\Delta L$ of the i-motif, we refer the species with $\Delta L$ of 5.1 nm to a partially folded structure. Previously, NMR data\(^{42}\) have suggested the presence of a partially folded structure in a similar C-rich sequence. However, the exact structure is not determined. Free energy calculations using mfold® program (see Figure 3.4) ruled out the hairpin based conformation in this structure. Rather, our data were more complied with a triplex intermediate previously proposed for the formation of i-motif\(^{127}\) or G quadruplex.\(^{143, 144}\) Based on the known i-motif structure,\(^{125, 145}\) we calculated the single nucleotide distance followed by the number of nucleotides involved in partially folded structure (Figure 3.5) as described below:

**Calculation of single nucleotide (nt) distance**

Number of nucleotides ($N$) * single nt distance ($L_{\text{single nucleotide}}, \text{nm}$) – end-to-end distance ($X, \text{nm}$) = change in counter length ($\Delta L, \text{nm}$)……………………………………(3.1)

\[
25 \times L_{\text{single nucleotide}} - 0.7 = 9.9
\]

\[
L_{\text{single nucleotide}} = 0.43 \text{ nm}
\]

**Calculation of number of nucleotides in the partially folded structure at pH 6.5**

\[
N \times 0.43 - 2.7 = 5.2
\]

\[
N = 18.3 \text{ nts} \sim 18 \text{ nts}
\]
Our calculation (see above) showed that the partially folded structure contains 18 (±1) nucleotides. This value was identical to the number of nucleotides (18 nts) contained in the triplex species (Figure 3.5). Our observation that a small fraction of molecules (3.9 %, 26 out of 671 i-motif molecules) showed two sequential rupture events was consistent with the intermediate nature of the partially folded species (see a representative curve in Figure 3.6). To provide further support that partially folded

Figure 3.5 Calculation of number of nucleotides contained in a secondary structure in the sequence 5'-(TGTCGCCACACCCC)₂TGT. Typical inter-phosphate distances across the wide and narrow grooves are 1.5 and 0.7 nm, respectively. The height of the i-motif structure was estimated using the C:CH⁺ stacking interval of 0.32 nm determined by Esmaili et al. To calculate the number of nucleotides (N) in the secondary structure, the single nucleotide distance was first obtained from the 25mer i-motif structure at given pH using the end-to-end distance (X) determined in the top panel of this Figure (see equation 3.1). The number of nucleotides involved in the partially folded structure at the pH was then calculated separately using the end-to-end distance determined in the bottom panel of this Figure. The number of nucleotides in the partially folded structure was finally averaged at different pH with a value of 18 (±1) nts.
structure adopts a triplex conformation of 18 nts, we designed a DNA mutant that contains only three C4 tracts (see Material and Methods for sequence detail). When we mechanically unfolded this

![Figure 3.6](image)

**Figure 3.6** A representative pulling curve (red) with two ~5 nm change in contour length ($\Delta L$) unfolding events (dashed circles) in the sequence 5'-(TGTCCCCACACC CC)$_2$TGT. Such an unfolding pattern strongly supports the intermediate nature of the partially folded structure. Blue line is the WLC fitting of the relaxing curve.

DNA at pH 5.5, the histogram of change in contour length showed only a single population with $\Delta L$ of 5.0 ($\pm 0.1$) nm (Figure 3.7). This value was identical within experimental error to the partially folded species (5.1 $\pm 0.5$ nm), strongly suggesting the triple stranded nature of the partially folded species. However, to unambiguously determine the structure, other techniques such as NMR or X-ray need to be employed.
Figure 3.7 Histogram of change in contour length (ΔL) of the DNA secondary structure formed in 5′-TGTC₄ACAC₄TGTC₄ACA. Compared to the i-motif forming sequence, this DNA contains only three C₄ tracts at the 5′ end. The DNA construct was unfolded in 10 mM sodium phosphate buffer pH 5.5, 100 mM KCl at 23 °C. The Gaussian peak for ΔL is centered at 5.0 (± 0.1) nm.

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

3.2.2 Effect of pH on the two populations.

i-Motif contains a stack of C:CH⁺ pairs that are pH dependent (Figure 3.1A i & ii). It is expected that the higher the pH above the pKₐ of cytosines (pKₐ= 4.3 for free cytosines¹¹³), the more difficult the formation of the i-motif. This trend was clearly demonstrated in CD spectra at pH 5.5-7.0 (Figure 3.2B). While CD spectrum showed the
existence of i-motif at pH 5.5, it suggested a mixture of unstructured DNA (peak at 275 nm)\textsuperscript{101} and i-motif structures (the 286 nm peak) at pH 6.0. At pH 6.5 and 7.0, however, the CD spectra were more indicative of unstructured single stranded DNA. In accordance with the CD spectra, melting was clearly observed at pH 5.5, 6.0 and 6.5 (Figure 3.2A and Figure 3.3), but not at pH 7.0 (not shown). The same trend was also obvious in Br\textsubscript{2} footprinting results, which strongly suggested the formation of i-motif at pH 5.5 but not at pH 7.0 (Figure 3.2C). However, caveat should be given that due to reduced sensitivity, the absence of melting in CD/UV measurement, or the absence of protection of C tracts in Br\textsubscript{2} footprinting cannot rule out the existence of a small fraction of i-motif or other folded structures in the sample.

If our assignment of the fully folded structure as an i-motif is correct, we should expect to see similar pH trend in single molecular studies. To this purpose, we mechanically unfolded the DNA construct at pH 5.5-7.0. When $\Delta L$ histograms at different pH were compared in Figure 3.8, it was obvious that two populations coexisted at each pH. $\Delta L$ for each species remained unchanged across the pH range (see Table 3.1), suggesting the structures of i-motif or partially folded species were intact after they formed within this pH range. We then calculated the percentage formation of each species at different pH (see Materials and Methods). When pH was increased from 5.5 to 7.0, we observed decreased formation of the species with the larger $\Delta L$ (23 \%-1.3 \%, Figure 3.9). Such clear pH dependence confirms our assignment that this species is indeed an i-motif. Interestingly, the formation of the partially folded structure remained
**Figure 3.8** $\Delta L$ histograms at different pH (23 °C). Two populations shown from pH 5.5 to 7.0 are fitted with two-peak Gaussian (solid curves). The fitting for the second peak (see *) at pH 7.0 is not reliable due to insignificant formation of the i-motif at this pH (see Figure 3.9). N depicts number of experiments at each pH.

constant (6.1±1.6 %) in pH 5.5-7.0, indicating the structure was not as pH sensitive as the i-motif.

The percentage formation of the i-motif was similar between pH 6.5 and 7.0 (2.9 %-1.3 %, Figure 3.9). However, the formation increased rather significantly below pH 6.5 (from 10.5 % at pH 6.0 to 23 % at pH 5.5). This trend corresponded strikingly well
with the CD measurements. Between pH 6.5 and 7.0, the CD spectra indicated the majority of the sample was unstructured. However, below pH 6.5, the CD measurements showed significant increase in the i-motif structure (Figure 3.2B).

![Graph showing pH dependence of i-motif formation](image)

**Figure 3.9** The percentage formation of the i-motif (dotted line) and the partially folded structure (solid line) at different pH under 23 °C.

Our single molecular method can readily identify populations as low as ~2 % (Figure 3.9). This fact strongly demonstrates the superior sensitivity of our method over bulk approaches. As a result of this sensitivity, we were able to identify the formation of i-motif even at neutral pH at room temperature. Although the formation of i-motif under neutral pH has been recently reported at 4°C,147 its formation at room temperature and neutrality has not been observed without the introduction of template superhelicity,32 the employment of molecular crowding conditions,115 or the adoption of chemical modifications.148

Since the pH dependence of the i-motif is originated from hemiprotonated cytosine-cytosine pairs, the reduced pH dependency observed here suggests that C:CH⁺
pairs may not contribute to the partially folded structure as significantly as those in i-motif. This observation is consistent with the triplex model (see Figure 3.5) which consists only half of the C:CH$^+$ pairs.

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

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

When we plotted out the histogram of unfolding force for the i-motif and the partially folded structure at different pH (Figure 3.10 A & B and Table 3.1), we observed similar rupture force range for these two species (Figure 3.10C; i-motif: 21.9-26.2 pN; partially folded structure: 25.5-26.5pN). Majority of the rupture events were abrupt, suggesting cooperative unfolding for each species. The unfolding forces for partially folded structures were rather constant over pH, whereas those for i-motif decreased with pH. When we performed the 286 nm CD melting at different pH, we found melting temperature decreased monotonically with pH (see Figure 3.2A and Table 3.1). This observation is consistent with the trend of the rupture force for i-motif (Figure 3.10C), suggesting it is the i-motif structure that predominates the CD signal at 286 nm.
Figure 3.10 Rupture force histograms of (A) partially folded structure and (B) i-motif at different pH under 23 °C. The solid lines in (A) and the dotted lines in (B) are Gaussian fits. The rupture force histogram of i-motif at pH 7.0 is not shown due to insignificant formation of the structure at this pH. (C) The rupture forces (obtained from Gaussian peaks in (A) and (B)) vs pH for the i-motif (green dotted line) and partially folded structure (red solid line).

We then used Jarzynski’s theorem\textsuperscript{94, 95} to calculate the thermodynamic stability ($\Delta G_{\text{unfold}}$) of the i-motif and partially folded structures (see Materials and Methods). The free energy change of i-motif ($\Delta G_{\text{unfold}}$ 16.0 (±0.8) kcal/mol) was close to that predicted
from literature (14.5 kcal/mol, calculated according to eight C:CH\textsuperscript{+} pairs\textsuperscript{5}). This result further verifies our assignment of the i-motif structure. The calculation also yielded a lowered free energy change for partially folded structure (10.4 (±0.7) kcal/mol). Such a result is consistent with our previous finding that C:CH\textsuperscript{+} pairs have less contributions to the partially folded species than to the i-motif (see Figure 3.9 and related text). We surmise the decreased contribution from C:CH\textsuperscript{+} pairs can lead to reduced H-bonds and C:CH\textsuperscript{+}/C:CH\textsuperscript{+} stacking, either of which can reduce $\Delta G_{\text{unfold}}$ in the partially folded structure.

The unfolding forces are higher than the stall force of RNA polymerases, which ranges from < 20 pN for bacterial RNAP\textsuperscript{119} to < 10 pN for Pol II\textsuperscript{120}. This suggests, solely from a mechanical aspect, that either i-motif or partially folded structure can stall RNAP. Due to the fact that the partially folded species has a higher population than i-motif at neutral pH, the former species may play a more important role in the regulation of RNAP. Previously, we have observed the unfolding forces of 23 and 37 pN for parallel and antiparallel ILPR G quadruplexes, respectively\textsuperscript{121}. Therefore, the mechanical stabilities of ILPR tetraplex DNA structures all seem to be higher than the stall force of RNAP\textsuperscript{119, 120}. It remains to be seen whether high mechanical stability is a common theme for other G quadruplexes or i-motifs. For those tetraplexes formed downstream of promoter regions\textsuperscript{118}, such a mechanical property can present a novel regulatory opportunity for transcription control through DNA secondary structure itself.
3.3 CONCLUSIONS

Using laser tweezers, we have demonstrated that an i-motif and a partially folded structure coexisted in the C-rich human ILPR oligonucleotides. The formation of i-motif is decreased with increasing pH, while that of partially folded structure is pH independent. Both the i-motif and the partially folded structure have unfolding forces higher than the stall forces of RNA polymerases, suggesting either of the structure can stop transcription from a mechanical perspective alone. To the best of our knowledge, single molecular investigation on i-motif structures has not been reported before. The methodology described here offers a novel tool to interrogate i-motif structures from a unique mechanical perspective at the single molecular level.
CHAPTER 4

INTRAMOLECULAR FOLDING IN HUMAN ILPR FRAGMENT WITH THREE C-RICH REPEATS

4.1 INTRODUCTION

Tandem repeats of DNA residues are abundant throughout human genome. Non-B DNA structures, such as G-quadruplex, i-motif, DNA cruciform, and H-DNA, can form in these tandem repeats. Under physiological conditions, these non-B DNA structures can prevail over DNA duplex to regulate DNA processing or gene expression. From this perspective, the DNA sequence does not merely serve as a genetic code; it can also form a structure to interfere with replication or other biological processes. During DNA replication for instance, it has been proposed that the presence of non-B DNA structures can cause deletion or addition of repeated sequences.

In human promoter regions, enrichment of cytosine and guanine repeats has been observed. Four tandem repeats of cytosine and guanine rich sequences can host i-motif and G-quadruplex structures, respectively. The fact that these structures can form in promoter regions has led to the hypotheses that G-quadruplexes or i-motifs may regulate RNA transcriptions. While various structures of G-quadruplex have been illustrated, much less information is available for the i-motif structure that can exist in the regions complementary to the G-quadruplex forming sequences. I-motif is composed of a stack of hemiprotonated C:CH$^+$ pairs. Parallel orientation exists for C-rich repeats that form the C:CH$^+$ pairing in both intermolecular and intramolecular structures. The essential component of hemiprotonated C:CH$^+$ pairs implies that acidic
condition is necessary for the formation of this structure. However, recent studies started to reveal that the structure can form even at physiological pH, especially under molecular crowding conditions or with negative superhelicity of the DNA template. The physiological significance of the i-motif structure is implied in recent discoveries that many proteins, such as single-stranded DNA binding proteins (SSB), helicases and other motor proteins, can specifically recognize C-rich sequences. In addition to Human genome, C-rich repeats are also found in Drosophila virilis DNA and some cardiovirus RNAs. These discoveries expose the i-motif to a broad range of host species for potential biological roles.

To function as a regulatory element in vivo, folding and unfolding of an i-motif structure are equally important. It has been proposed that formation or dissolution of an i-motif undergoes either a strand-by-strand or a duplex-by-duplex pathway. Observation of hitherto evasive intermediates helps to identify a specific pathway. However, due to the small quantity of these intermediates, they are highly difficult to investigate by ensemble techniques such as CD, NMR, X-ray crystallography. Single-molecule approaches offer unique capability to illustrate these structures. For example, our recent mechanical unfolding experiments on i-motif structures in the Insulin Linked Polymorphic Region (ILPR) has suggested that partially folded C-rich structures may adopt a triplex-like DNA structure employing three neighboring C-rich repeats among four available C-tracts. Based on this, it is reasonable to assume that similar structures may also exist in a sequence with only three available cytosine-rich repeats. Compared
to four or more C-rich tandem repeats, three neighboring C-rich repeats are expected to occur more frequently in human genome. This increases the opportunity for the C-rich structures to form in vivo, which sets a premise for the structures to play functional roles in biological processes. Since current algorithm only searches for regions that have more than three C-repeats for possible C-rich structures, formation of a stable structure in a sequence of three C-repeats will transform the searching algorithm. Recently, reports have shown a stable intramolecular structure in three tandem G-rich repeats.\textsuperscript{155, 156} However, stand-alone structures in three tandem C-rich repeats have not been reported.

Here, we have used a DNA fragment with three C-rich tandem repeats, 5'-TGTC\_4ACAC\_4TGTC\_4ACA-3', derived from the ILPR region, to investigate the possible secondary structure formed in this sequence. We have observed a stable structure during mechanical unfolding experiments, CD and UV melting measurements, as well as native gel shift assays. The C:CH\textsuperscript{+} pair stacking in the structure has been suggested by pH titration during mechanical unfolding and CD measurements. The participating C4 tracts in the C:CH\textsuperscript{+} pairing have been identified with mutational analysis. The contour length and unfolding force measurement on the mechanical unfolding experiments have supported the parallel strand orientation in the folded structure. Based on these, we propose a possible structure linchpined by C:CH\textsuperscript{+} pair stacking. Furthermore, laser-tweezers and bromine footprinting experiments have shown that this structure can serve as a building block for i-motif structures. We anticipate the existence of a stable species
in three tandem C-rich repeats not only adds a new possibility for gene regulation, but also generates a novel thought on the DNA based nanomaterials and biosensors.

4.2 RESULTS

4.2.1 CD spectroscopy demonstrated that structures in the ILPR-I3 sequence contain hemiprotonated cytosine pair stackings

First, we performed CD measurements to determine whether stable structures exist in the wild-type ILPR sequence, 5'-TGTCACACATGTCA-3' (ILPR-I3, Table 4.1). The CD spectra at pH < 6 showed a broad positive band at ~285 nm and a negative band at ~260 nm (Figure 4.1A). These CD signatures have been demonstrated previously for either intercalative or non-intercalative C:CH+ pair stackings (see Discussion). When pH increased towards neutrality, the band evolved towards 277 nm, a signal characteristic of a random coil conformation. This pH dependency is similar to those observed for the i-motif structure in the ILPR-I4 sequence, 5'-(TGTCACACAC)2TGTA-3'.

Figure 4.1B shows the direct comparison between these two species. Although both species demonstrated the sigmoidal dependency on pH, the structure in the ILPR-I3 showed a shallower transition compared to the i-motif formed in ILPR-I4. This observation reveals sluggish response of the ILPR-I3 to the pH, possibly due to less C:CH+ stacking involved in the structure. The decreased pH dependency of the structure in the ILPR-I3 is in agreement with that for a partially folded structure revealed by the single-molecule study on the ILPR-I4 sequence.
Table 4.1 Sequences of wild type ILPR-I4 and ILPR-I3, a scrambled sequence, and the mutants used in this study.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild Type (ILPR-I4)</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACCCCTGT-3'</td>
</tr>
<tr>
<td>Wild Type (ILPR-I3)</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Scrambled Seq. (ILPR-S3)</td>
<td>5'-CCTCGCTCACACTCCGCTCACCA-3'</td>
</tr>
<tr>
<td>Wild Type (ILPR-I1)</td>
<td>5'-CCCCTGT-3'</td>
</tr>
<tr>
<td>Mut-C4T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C5T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C6T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C7T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C12T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C11-14T</td>
<td>5'-TGTCCCCACATTTTTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C18T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C19T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C20T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
<tr>
<td>Mut-C21T</td>
<td>5'-TGTCCCCACACCCCTGTCCCCACACA-3'</td>
</tr>
</tbody>
</table>
To provide further evidence that ILPR-I3 fragment folds into a secondary structure, we performed CD melting measurements. With increasing temperature, the 285 nm CD band became blue shifted to 277 nm, a signal indicative of non-structured DNA\textsuperscript{101} (Figure 4.1C). In line with this observation, a melting transition temperature of 44±0.7 °C was observed (Figure 4.1D). These CD measurements demonstrate that ILPR-I3 can fold into a thermally stable structure containing a stack of C:CH\textsuperscript{+} pairs.

4.2.2 Gel shift assay and thermal denaturation analyses suggested that the folded structure is intramolecular
To determine whether the ILPR-I3 folds intramolecularly, we performed thermal denaturation (UV) and gel shift assays. Due to compact conformations, intramolecularly folded DNA structures are expected to migrate faster in gel electrophoresis than unstructured DNA of the same length. Indeed, we observed that a fraction of the ILPR-I3 showed faster electrophoretic mobility (notice the front tail in lane 2, Figure 4.2A, left panel) compared to an unstructured ILPR-S3 sequence (lane 1, Figure 4.2A, left panel, see Table 4.1 for the sequence) in the native electrophoretic mobility shift assay (EMSA) at pH 5.5. To ensure that the front tail on the ILPR-I3 (lane 2 of the native gel in Figure 4.2A) was not due to artifacts, three native gel experiments were performed. In all three experiments, the front tail was consistently observed in the ILPR-I3, but it was not shown in the scrambled sequence, neither was it observed in the denaturing gel.

![Figure 4.2](image)

**Figure 4.2** Electrophoretic mobility shift assays (EMSA) and thermal melting/reannealing of the ILPR-I3. (A) EMSA of the ILPR-S3 (scrambled DNA, lane 1) and the ILPR-I3 (lane 2) at 1 µM strand concentration. Lane 3 shows the DNA marker
(M). Left panel, a native gel at pH 5.5. Right panel, a denatured gel (10% PAGE, 7M urea). (B) Melting ($T_{1/2\text{-melting}}$), reannealing ($T_{1/2\text{-reannealing}}$), and equilibrium ($T_m$) temperatures of the structure in the ILPR-I3 with 5 to 100 µM concentration. The melting and reannealing temperatures were determined by 295 nm UV-melting and UV-reannealing, respectively, in a 10 mM sodium phosphate buffer (pH 5.5) with 100 mM KCl. The equilibrium melting temperature ($T_m$) of the ILPR-I3 was determined based on the non-equilibrated melting and reannealing curves (see Materials and Methods).

The unstructured ILPR-S3 was confirmed by the 277 nm CD band (Figure 4.3) and lack of UV melting transition (data not shown). In addition, the smeared band observed for ILPR-I3 (Lane 2, Figure 4.2A, left panel) under native condition suggests that a fraction of ILPR-I3 was folded intramolecularly. As a control, both ILPR-I3 and ILPR-S3 showed similar electrophoretic mobilities in a denatured PAGE gel (Figure 4.2A, right panel). These results suggest that an intramolecularly folded structure forms in the ILPR-I3 fragment.

![Figure 4.3](image)

**Figure 4.3** CD spectra of sequences described in Table 4.1. CD spectra of the mutants with mutation sites in each of the three C4 tracts in the ILPR-I3 are plotted in A), B), and C), respectively. The spectra of ILPR-I3 (red) and the scrambled sequence (black) are also included in each Figure for direct comparison. These CD experiments were performed at 5 µM oligonucleotide concentration in a 10 mM sodium phosphate buffer (pH 5.5) with 100 mM KCl at 23 °C.

Previous investigations have shown that, in contrast to intermolecular DNA structures, $T_m$ of an intramolecular structure is independent of DNA concentrations.$^{101,159}$
We measured the melting ($T_{1/2}$-melting), reannealing ($T_{1/2}$-reannealing), and equilibrium ($T_m$) temperatures according to the procedures described in literature$^{98-100}$ while varying the concentration of the ILPR-I3 from 5 to 100 µM at pH 5.5 (Figure 4.2B). As shown in Figure 4.2B, we found that $T_{1/2}$-melting, $T_{1/2}$-reannealing and equilibrium melting temperature ($T_m$) remained unchanged over a 20-fold variation in strand concentration, confirming the intramolecular nature of the structure in the ILPR-I3.

4.2.3 Single-molecule study provided the direct evidence for intramolecular folding of the ILPR-I3 from acidic to neutral pH

The decisive evidence for the intramolecular nature of the structure in the ILPR-I3 came from mechanical unfolding studies using laser tweezers. To this end, the ILPR-I3 fragment was sandwiched between two dsDNA handles. One of the handles was labeled with biotin and the other with digoxygenin (Dig) at the free end. This setup allowed the DNA construct to be tethered between the two beads functionalized with streptavidin and anti-Dig antibody, respectively (see Materials and Methods and Figure 4.4 for details). The native flanking sequences, TGT and ACA, were incorporated at the two ends of the ILPR-I3 to reduce the steric hindrance between the folded structure and the dsDNA handles. To unfold the possible secondary structure in the ILPR-I3, the DNA construct tethered between the two optically trapped beads was stretched by moving the two beads apart. In the ensuing force-extension ($F$-$X$) curves, we observed a sudden drop in the force, which indicated the unfolding of a DNA secondary structure (Figure 4.4). Since
the experiment was performed on individual molecules, the rupture event must represent the unfolding of an intramolecular structure, whose size is quantified by the change in contour length ($\Delta L$, Figure 4.5A, left panel). The average unfolding forces for these species were $31\pm1$ pN at pH 5.5 and $30\pm1$ pN at pH 7.0 (Figure 4.5A, right panel), demonstrating that folded structures were mechanically robust.
Figure 4.5 Single-molecule study of the ILPR-I3 at different pH using laser tweezers. (A) Change in contour length ($\Delta L$) (left panel) and rupture force histograms (right panel) of the ILPR-I3 and the Mut-C19T at different pH (23 °C). The histograms are fitted with Gaussians (solid and dotted curves). $N$ depicts the number of experiments. (B) Percentage formation of the ILPR-I3 at pH 5.5 and 7.0 (“pH 5.5” and “pH 7.0”), the Mut-C19T at pH 5.5 (“Mut-pH 5.5”), and a DNA construct that does not contain the C-rich sequence at 23 °C (“Control”). (C) Comparison of $\Delta L$ (triangles linked by the red dotted line at the bottom) and rupture force (filled circles linked by the black dotted line at the top) between the ILPR-I3 and the Mut-C19T fragments. Notations are the same as described in (B).

It is expected that the structure employing hemiprotonated cytosine pair stacking is favored under acidic conditions. When we analyzed the percentage formation at pH 5.5 and 7.0 (see Materials and Methods), indeed, we observed a decreased percentage formation at pH 7.0 (19±2 %) compared to pH 5.5 (33±1 %) (Figure 4.5B). In comparison, a control DNA construct containing only the double-stranded DNA handles (see Materials and Methods) demonstrated negligible unfolding events of 3±1 % (3 out of 105 curves). The substantial formation of the structure at neutrality confirmed the pH
titration results that the ILPR-I3 structure was less pH dependent compared to the i-motif (Figure 4.1B). This fact adds new evidence that C-rich structures containing C:CH\textsuperscript{+} pairs can form at pH 7.0,\textsuperscript{32, 154, 160} which sets a premise for biological roles of these structures.

**Figure 4.6** Four possible structures that employ C:CH\textsuperscript{+} pair stacking in the ILPR-I3 sequence. The three C4 tracts are shown in blue and other regions are shown in red for structures in (A)-(D). The unfolding direction for each structure is shown by black arrows labeled with “F”. Notice the structures with free C4 tracts at the 5'-end yield $\Delta L$ values identical to those of B and D, and therefore, they are not shown here.

To illustrate the structure that employs C:CH\textsuperscript{+} pair stacking, we evaluated four most probable candidates in Figure 4.6. First, we determined the end-to-end distance ($X$) of each structure from coordinates of similar PDB structures.\textsuperscript{145, 161} This value was then
used to derive $\Delta L$ from the equation, $L = \Delta L + X$,\textsuperscript{28} where the contour length of each structure was calculated by using $L_{\text{single nucleotide}} = 0.43$ nm\textsuperscript{134-136} in equation 3.1. The $X$ remains the same for the structures A and B, as well as the structures C and D for the mechanical unfolding experiments. The end-to-end distance ($X$) for structures C&D is 1.5 nm, which is the average inter-phosphate distance obtained from the literature.\textsuperscript{23, 162}

The $X$ for structures A&B was estimated as the hypotenuse (see the green triangles in the top panel) to the rise of the four stacking C:CH$^+$ pairs (the opposite side) and the inter-phosphate distance between the two C4 strands (1.5 nm, the adjacent side). Since the C:CH$^+$ stacking resembles double stranded DNA (dsDNA), we set the lower limit of the rise per C:CH$^+$ as 0.34 nm (single base pair rise in dsDNA).\textsuperscript{134} The upper limit of the rise per C:CH$^+$ is set at 0.66 nm, which is the average rise between the two intercalative C:CH$^+$ stacking pairs determined from the known i-motif structures (PDB Codes; 1YBL, 1G22, 1EL2 and 1CNO).\textsuperscript{126, 145, 161, 163} This calculation yielded the rise of the four C:CH$^+$ pairing between 1.0 and 2.0 nm (shown in the left triangle) for structures A and B. Based on this, the $X$ was calculated as 1.8-2.5 nm for these two structures. Using the calculation described above, these values yielded $\Delta L$ of 5.3-5.9 nm and 2.3-2.9 nm for structures A and B, respectively (summarized in the Table 4.2). As shown in Table 4.2, the observed $\Delta L$ (5.0 ± 0.1 nm at pH 5.5 and 5.2 ± 0.4 nm at pH 7.0) matched with the expected range of $\Delta L$ (5.3-5.9 nm) for structure A only. Among the four species shown in Figure 4.6, only the structure utilizing the first and the third C4 tracts in a parallel orientation
Table 4.2 Calculation of contour length change ($\Delta L$) of the four possible candidates shown in Figure 4.6.

<table>
<thead>
<tr>
<th>Structures</th>
<th># nucleotides involved (N)</th>
<th>End-to-end distance ($X$, nm)</th>
<th>Expected change in contour length ($\Delta L$, nm)</th>
<th>Observed ($\Delta L$, nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig 4.6A</td>
<td>18 nts</td>
<td>1.8-2.5</td>
<td>5.9-5.3</td>
<td>5.0 ± 0.1 (pH 5.5)</td>
</tr>
<tr>
<td>Fig 4.6B</td>
<td>11 nts</td>
<td>1.8-2.5</td>
<td>2.9-2.3</td>
<td>5.2 ± 0.4 (pH 7.0)</td>
</tr>
<tr>
<td>Fig 4.6C</td>
<td>18 nts</td>
<td>1.5</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Fig 4.6D</td>
<td>11 nts</td>
<td>1.5</td>
<td>3.2</td>
<td></td>
</tr>
</tbody>
</table>

for the C:CH$^+$ pair stacking yields a $\Delta L$ of 5.3-5.9 nm, which falls into the observed $\Delta L$ of 5.2 ± 0.4 nm at pH 7.0 and close to the $\Delta L$ of 5.0 ± 0.1 nm at pH 5.5 (Figure 4.5A, left panel). Such a strand orientation is consistent with the finding that only parallel arrangement is possible for the C:CH$^+$ pair stacking, which requires anti glycosidic bonds in cytosines to avoid steric effects.164

If the structure does employ the first and the third C4 tracts with parallel orientation, then the mutation of any cytosine residues in these two C4 tracts is expected to destabilize the structure. To test this, we prepared a DNA construct in which the second cytosine (C19) in the third C4 tract was mutated to T (Mut-C19T in Table 4.1, 5'-TGTC4ACAC4TGTC[\text{C}]CCACA). After we mechanically unfolded the mutant at pH 5.5, we found $\Delta L$ of the mutant (5.1 ± 0.4 nm) was identical with that of the wild type; whereas the rupture force (24.8 ± 0.6 pN) was significantly lower (Figure 4.5 A & C). The former observation indicates that the C19T mutation still allows the folding of a secondary structure, which is consistent with the previous finding that intercalative
C:CH\(^+\) pairs are not necessary for a folded structure.\(^{160}\) The latter observation clearly shows that the structure has weakened stability, which is likely due to the loss of one stack of C:CH\(^+\) pair as a result of the C19T mutation. Compared to the structure in the wild-type ILPR-I3 at pH 5.5, the percentage formation of the structure in the C19T DNA (15±1 %) is decreased by ~18 % (Figure 4.5B). This is consistent with the less stable structure due to reduced C:CH\(^+\) pair stacking in the mutant.

4.2.4 Thermal stability analysis and CD measurements of the mutants suggested cytosine pair stacking between distally located C4 tracts

To further confirm the specific C4 tracts involved in the C:CH\(^+\) stacking, we systematically mutated cytosines to thymines in each C4 tract of the ILPR-I3 and evaluated the thermal stability of these mutants (Table 4.1). When a particular cytosine in the C:CH\(^+\) pairing is mutated to thymine, \(T_{1/2}\text{-melting}\) is expected to be lower than that of the wild type.\(^{101}\) Figure 4.7A shows the results of the thermal stability analysis monitored at 295 nm on 10 µM mutants at pH 5.5. Two populations are strikingly clear in Figure 4.7A. One is clustered around the wild-type ILPR-I3 with similar melting temperatures; while the other with 3-10 °C lower in \(T_{1/2}\text{-melting}\) is located close to the Mut-C19T, a mutant with compromised stability demonstrated by the single-molecule experiments. A close analysis reveals that the mutations involving the second C4 tract lead to the population with \(T_{1/2}\text{-melting}\) comparable to that of the wild-type ILPR-I3; whereas the mutations on the first and the third C4 tracts result in populations with
lowered $T_{1/2}$-melting. Such a distribution clearly indicates that the first and the third C4 tracts participate in the stacking of hemiprotonated cytosine pairs. This conclusion was fully supported by structural analysis on all mutants using CD measurements (Figure 4.3). Compared to the wild-type ILPR-I3 whose 285 nm CD band is indicative of the C:CH$^+$ pair stacking (either intercalative or non-intercalative, see Discussion), all mutants show blue shifted bands in Figures 4.1 and 4.3. In fact, the DNA with a scrambled sequence, ILPR-S3, shows such a complete shift that a broad band centered at 277 nm is observed. This indicates an unstructured conformation for this sequence, a result

![Figure 4.7](image)

**Figure 4.7** Mutation analysis in a 10 mM sodium phosphate buffer (pH 5.5) with 100 mM KCl. (A) 295 nm UV melting curves of the ILPR-I3 (“Wild Type”) and the mutants at 10 µM concentration. (B) Top panel, $T_{1/2}$-melt of the mutants and the ILPR-I3. “W” depicts the wild type ILPR-I3. Bottom panel, CD peak shift of the mutants and the scrambled sequence (ILPR-S3) with respect to the 285 nm peak in the ILPR-I3. The horizontal dotted lines (green) represent the average value for each C4 tract. Statistical treatment is represented by the $P$ values in the bottom panel. Please refer to Table 4.1 for DNA sequences.
consistent with the EMSA observation (Figure 4.2A). The mutants involving the first and the third C4 tracts have significantly larger blue shifts than those involving the second C4 tract (p < 0.0002 for the first C4 tract and p < 0.02 for the third C4 tract). Since the blue shift from 285 nm to 277 nm suggests the switching of the C:CH$^+$ stacking to an unstructured conformation, these results well explain the thermal analysis data from a structural perspective.

The thermal stability experiments and CD analysis described here suggest that the structure in the ILPR-I3 is joined by hemiprototated cytosine pair stacking utilizing the first and the last C4 tracts in the sequence. Such a conformation (Figure 4.8A) is also consistent with the proposed structure from single-molecule investigations.

4.2.5 The intramolecularly folded ILPR-I3 served as a building block for intermolecular i-motif structures

We anticipate that in the presence of the ILPR-I1 fragment, 5'-CCCCTGT, folding of the ILPR-I3 can evolve into an intermolecular i-motif (Figure 4.8A). Bromine footprinting experiments provided direct support to this expectation. It has been shown that cytosines in the C:CH$^+$ pair stacking are more protected against Br$_2$ than those in the single stranded region.$^{102}$ At pH 5.5, the Br$_2$ footprinting on the ILPR-I3/ILPR-I1 mixture (1:1 molar ratio) showed more protection of each C4 tract in the ILPR-I3 fragment compared to that at pH 7.0 (Figure 4.8B, lanes 1 & 2 and the fold protection
Figure 4.8 Formation of an intermolecular i-motif. (A) Schematic of the formation of an intermolecular i-motif. The proposed structure in the ILPR-I3 is shown on the left. Each C:CH\(^+\) pair is represented by two opposite rectangles. (B) PAGE gel image of the Br\(_2\) footprinting experiment. Lane 1, the ILPR-I3/ILPR-I1 (I\(_3\)+I\(_1\)) mixture at pH 7.0. Lane 2, the I\(_3\)+I\(_1\) sample at pH 5.5. Lane 3, the ILPR-I3 (I\(_3\)) at pH 5.5. Lane 4, the ILPR-I4 (I\(_4\)) at pH 5.5. The band intensity for lane 2 is shown to the left of the gel. The fold protection for the I\(_3\)+I\(_1\) sample at pH 5.5 is shown to the right. The dotted vertical lines indicate the average fold protection for each C4 tract. The blue arrows indicate the loop cytosines. Error bar represents the standard deviation of three independent experiments. The blue arrows indicate the cytosines in the ACA section of each fragment. Note that the fold protection for adenines at 3'-end (indicated by asterisk *) is not accurate since they are close to the uncut oligo. (C) Normalized rupture force histogram for the I\(_3\)+I\(_1\) sample at pH 5.5. The solid black curve represents a two-peak Gaussian function. The dotted curve is the Gaussian fit for the rupture force histogram of the ILPR-I3 at pH 5.5.

shown to the right). This result can be well explained by the formation of an intermolecular i-motif between the two oligos at pH 5.5, but not at pH 7.0. In accordance with the i-motif formation at pH 5.5, C4 tracts are more protected than the cytosines in the ACA sections, which remain single-stranded (lane 2 and the fold protection pattern).
Figure 4.9 shows that fold protections of the corresponding C4 tracts in the ILPR-I3/ILPR-I1 mixture and the ILPR-I4 fragment are similar. Since ILPR-I4 is known to form an i-motif at pH 5.5\textsuperscript{28}, this result again supports the formation of an intermolecular i-motif between the ILPR-I3 and the ILPR-I1. It is noteworthy that the C4 tracts in the

\textbf{Figure 4.9} Intensity scan for ILPR-I3 bands (the green trace to the left of the gel) and the fold protection for ILPR-I4 (I\textsubscript{4}, black) and ILPR-I3 (I\textsubscript{3}, green) for Br\textsubscript{2} footprinting in a 10 mM sodium phosphate buffer at pH 5.5 with 100 mM KCl. Note that this gel is identical with that in Figure 4.8B (see Materials and Methods in chapter 2 for fold protection calculation). Samples in different lanes are labeled according to Figure 4.8B. The cytosines in ACA sections are indicated by blue arrows in the intensity scan and blue bars in the fold protection graphs. The C4 tracts in the gel are highlighted with corresponding sequences. Error bars represent the standard deviations calculated from three independent experiments.
in the ILPR-I3 sequence (Figure 4.8B, lane 3 and Figure 4.9) show much less protection against Br₂ compared to either the ILPR-I4 sequence or the ILPR-I3/ILPR-I1 mixture. This may reflect the fact that the C4 tracts of the structure in the ILPR-I3 have less steric hindrance compared to those in i-motifs.

The formation of an intermolecular i-motif was further supported by mechanical unfolding experiments. The rupture force histogram showed two populations when the ILPR-I3 was unfolded in the presence of 10 µM ILPR-I1 (Figure 4.8C). Based on the close values of the rupture force between the two (compare the red and the left black population in Figure 4.8C), the population with 24 ± 2 pN rupture force was assigned to the intramolecular ILPR-I3 structure.

Figure 4.10 Calculation of the unfolding rate constant ($k_{\text{unfold}}$) at 0 pN for the intramolecular i-motif (“ILPR-I4”, calculation based on published data121) and the 45 pN population in the ILPR-I3/ILPR-I1 mixture (“ILPR-I3+ILPR-I1”) from the plot of ln[ln(1/N)] versus rupture force. We used the equation 4.2 to estimate the $k_{\text{unfold}}$.  

$$\ln[\ln(1/N)] = \ln[k_{\text{unfold}}/(X^x_{f\rightarrow u} / k_B T)] + (X^x_{f\rightarrow u} / k_B T)F$$  (4.2)
where \( r \) is the loading rate (5.5 pN/s), \( N(F,r) \) is the fraction of folded molecules at force \( F \) and loading rate \( r \), and \( X^f_{\mu \rightarrow \nu} \) is the distance from the folded state to the transition state along the unfolding coordinate. \( k_{\text{unfold}} \) is obtained from the linear fit (solid black lines) in the graph. This calculation yielded \( k_{\text{unfold}} \) of \( 3.7 \times 10^{-3} \) s\(^{-1}\) and \( 7.2 \times 10^{-5} \) s\(^{-1}\), respectively, for the intramolecular ILPR i-motif and the 45 pN population in the ILPR-I3/ILPR-I1 mixture. Notice a randomized deconvolution of the two populations (24 and 45 pN) in the ILPR-I3/ILPR-I1 mixture (Figure 4.8C, black histogram) was used.\(^{121}\)

The population with increased rupture force of 45±1 pN was likely an intermolecular i-motif. This assignment is based on the fact that structures of intermolecular nature should have smaller unfolding rate constant (\( k_{\text{unfold}} \)) compared to that of intramolecular structures.\(^{166}\) Indeed, the unfolding rate constant for the 45 pN species, \( k_{\text{unfold}} = 7.2 \times 10^{-5} \) s\(^{-1}\), is much smaller than that for an intramolecular i-motif, \( k_{\text{unfold, intramolecular}} = 3.7 \times 10^{-3} \) s\(^{-1}\) (Figure 4.10). Taken together, the mechanical unfolding and footprinting results clearly indicate that the ILPR-I3 structure can serve as a building block for intermolecular i-motif. To the best of our knowledge, the unfolding experiments shown here represent for the first time an intermolecular i-motif has been investigated at the single-molecule level.

4.3 DISCUSSION

CD has been extensively used to characterize the i-motif structures in the C-rich oligos. The CD spectrum with a positive peak at \(~280-288\) nm and a negative trough near 260 nm indicates the formation of i-motif structures. Not only intercalative C:CH\(^+\) stackings in a typical i-motif structure show these CD features, similar CD spectrum has also been observed for the non-intercalative C:CH\(^+\) stacking in duplex DNA.\(^{157}\) The latter observation has broadened the scope of CD to characterize the structures involving
non-intercalative C:CH\(^+\) pair stacking. Our study exploits this capability to characterize secondary structures in the ILPR-I3. Using the DNA concentration (5 μM) that favors the intramolecular folding (Figure 4.2B), we observed characteristic CD peaks for the C:CH\(^+\) stacking (the positive peak near 285 nm and the negative trough near 260 nm in Figures 4.1A and 4.1C). The pH dependency of the 285 nm peak (peak gradually blue shifted with increasing pH (4.5 to 8.0), see Figures 4.1A and 4.1C) further supported the presence of the pH sensitive C:CH\(^+\) stacking in the ILPR-I3.

It is possible that the ILPR-I3 structure may assume either a parallel or antiparallel strand arrangement (Figure 4.6). With the antiparallel orientation, structures resemble a hairpin in which the stem is composed of the C:CH\(^+\) pairs (Figure 4.6 C-D). Thus, the unfolding geometry used in current laser-tweezers experiments is equivalent to unzipping a hairpin, which requires around 15 pN.\(^{167}\) However, the rupture forces observed here, 31 and 30 pN at pH 5.5 and pH 7.0 (Figure 4.5A, right panel), respectively, are significantly larger than this value. Instead, they are within the force range required to “slide” open a duplex DNA,\(^{168}\) which shares the same geometry as the unfolding of the structures with parallel strand orientations (Figure 4.6 A-B).

Compared to the ensemble average measurements such as NMR and CD, the laser-tweezers based single-molecule method is highly sensitive. For example, at pH 7.0, laser tweezers revealed that 19% of the population is folded in the ILPR-I3 (Figure 4.5B). However, under the same condition, the 285 nm CD signal characteristic of the C:CH\(^+\) stacking was masked by the broad band at 277 nm, which is the signature from the main
unfolded population (Figure 4.1A). Therefore, although our laser-tweezers method does not produce structural information at atomic details as revealed by NMR or X-ray measurements, its highly sensitive nature has enabled it to probe the structures for species with minute quantity.

Apart from the sensitivity, laser tweezers have a unique capability to measure the mechanical stability of DNA structures.\textsuperscript{28, 121, 169-171} Motor proteins, such as DNA/RNA polymerases and helicase, generate a load force during their enzymatic cycles.\textsuperscript{120, 172-174} Recent finding has revealed that DNA G-quadruplex and i-motif have mechanical stabilities\textsuperscript{28, 121} comparable to the stall force of polymerases, suggesting they may play significant roles to regulate polymerases from the mechanical perspective alone. Here, the C-rich structure in the ILPR-I3 ($F_{\text{unfold}} = 31$ pN) shows a similar mechanical stability, suggesting a similar capability compared to other non-B DNA structures. That the ILPR-I3/ILPR-I1 mixture can form an intermolecular i-motif implies that the ILPR-I3 structure serves as an important building block\textit{ en route} to the i-motif folding. Previously, 3 + 1 G-quadruplex assemblies (three G-tracts from one strand and one G-tract from another strand) have been reported by laser tweezers, NMR, and AFM studies.\textsuperscript{156, 175, 176} These results provide evidence that folding or unfolding of DNA tetraplexes may follow a strand-by-strand,\textsuperscript{127} instead of duplex-by-duplex pathway.\textsuperscript{22, 154}

\section*{4.4 CONCLUSIONS}

In summary, we have shown the existence of a stable structure in the ILPR sequence with three C-rich repeats. Our results show that folded species in the ILPR-I3
is mechanically and thermodynamically stable. The structure is stabilized by the first and the third C4 tracts via hemiprotonated cytosine pair stacking with a parallel strand arrangement.
CHAPTER 5

G-QUADRUPLE AND I-MOTIF ARE MUTUALLY EXCLUSIVE IN DOUBLE-STRANDED ILPR DNA

5.1 INTRODUCTION

Majority of the investigations on the DNA tetraplexes focus on the property of G-quadruplexes in the ssDNA context. Recent reports have started to show the formation of G-quadruplexes in the context of double-stranded DNA templates. For example, interconversion between duplex and quadruplex structures have been observed either on a DNA hairpin \(^{177}\) or in a DNA fragment without flanking dsDNA handles.\(^{47, 178-181}\) To address the effect of dsDNA junctions on the DNA tetraplex,\(^ {182}\) formation of G-quadruplex has been investigated on double-stranded DNA fragments flanked by long dsDNA handles under specialized conditions, such as in a molecular crowding environment or on a negative supercoiled template.\(^ {32, 44}\) Interestingly, single-molecule FRET analysis on a 96-bp duplex DNA in the c-kit region has shown that G-quadruplex can form in the presence of a complementary strand without these specialized settings.\(^ {130}\) Likewise, formation of G-quadruplexes has been demonstrated in dsDNA by DMS footprinting and stop assay analyses under similar conditions.\(^ {44, 183}\) These observations have suggested that G-quadruplexes can readily compete with the reannealing of dsDNA \textit{in vitro}. 
With the formation of a G-quadruplex, the complementary C-rich strand becomes single-stranded, which should facilitate the folding of i-motif. However, tetraplex investigations in the dsDNA context are often performed at neutrality under which i-motif is difficult to form. As a result, it is still an open question whether an i-motif can fold in the complementary single-stranded region after a G-quadruplex is formed in the opposite strand. The answer to this question has two important implications. First, the formation of an i-motif is expected to change the formation kinetics or thermodynamic stability of a proximal G-quadruplex in the opposing strand. This may influence the biological functions of the G-quadruplex. Second, it is likely that i-motif may have its own biological roles in the gene regulation. Together with G-quadruplex, this can introduce a functional versatility for DNA tetraplexes. Compared to the study of G-quadruplex in the dsDNA context, however, even fewer investigations focus on the formation of i-motif in the dsDNA.\textsuperscript{32,47}

Here, by employing laser-tweezers based single-molecule methods and DMS/bromine footprinting approaches, we provide compelling evidence that either a G-quadruplex or an i-motif, but not both, is formed in a long dsDNA construct that contains an ILPR (insulin linked polymorphic region) fragment, without template superhelicity or molecular crowding environment. We have observed that G-quadruplex is more stable than i-motif in the presence of a respective complementary strand.

\subsection*{5.2 RESULTS}
5.2.1 Chemical footprinting shows ILPR G-quadruplex or i-motif is stable in the context of dsDNA

To investigate the effect of the complementary strand on the ILPR tetraplex formation, first, we performed DMS footprinting on a 87-bp DNA, 5'-CTA GAC GGT GTG AAA TAC CGC ACA GAT GCG ACA GGGG TGT GGGG ACA GCC AGC AAG ACG TAG CCC AGC GCG TC, which consists of a 5'-32P labeled G-rich strand and an unlabeled complementary C-rich strand (Figure 5.1A). This fragment contains an ILPR G-quadruplex forming sequence (underlined)36, 44, 121 sandwiched between a 33 base-pair (bp) and a 29-bp sequence.

![Figure 5.1](image)

**Figure 5.1** Chemical footprinting of a double-stranded 87-bp ILPR DNA at different pH and salt conditions at 23 °C. Black vertical lines depict the G4 or C4 tracts involved in the tetraplex formation. The predominant species is shown at the bottom of each lane. The blue dotted arrows indicate the increasing band intensities. (A) DMS footprinting of the 87-bp in a 10 mM MES buffer at pH 5.5 (lanes 1-3) and in a 10 mM Tris buffer at pH
7.4 (lanes 5-7). Lane 4 is the GA ladder for the G-rich strand. The orange dotted and grey traces are line-scans for lanes 1 and 3 respectively. Fold protections at pH 7.4 (lane 3 vs lane 1, blue empty bars) and pH 5.5 (lane 7 vs lane 5, red solid bars) in 100 mM K\(^+\) are shown to the left. The red solid dots to the right of the gel show the guanine residues in the TGT loops. (B) Bromine footprinting of the 87-bp dsDNA in a pH 7.4 or pH 5.5 sodium phosphate buffer with 100 mM ions. The pre-mixed sample was denatured at 95 °C for 10 min and rapidly quenched in a water bath set at 30 °C before Br\(_2\) treatment. The similar band intensities of all cytosines in the i-motif forming region (see the quantitation to the left of the gel) indicate that i-motif was not formed under these experimental conditions. (C) Bromine footprinting of the dsDNA at pH 5.5 (lanes 1 and 2) and pH 7.4 (lane 3). G-rich and C-rich strands were heated separately to 95 °C and rapidly quenched in a water bath set at 30 °C before mixing. The black trace to the left of the gel is the line scan for lane 1. Fold protections at pH 5.5 with 100 mM K\(^+\) (lane 3 vs lane 1, pink solid bars) and 100 mM Li\(^+\) (lane 3 vs lane 2, black empty bars) are shown at the left side. The red solid dots to the right of lane 2 depict cytosines in the ACA loops. The green open circles to the left of lane 1 indicate the cytosines closest to the i-motif forming region. They are used to quantify the hybridization of flanking DNA strands (see text).

After these two strands were mixed, incubated at 95 °C, and rapidly cooled in an ice-water bath (the mix→heat→cool procedure, Figure 5.1A), they were subjected to DMS footprinting (see Materials and Methods in Chapter 2). It is noteworthy that fast temperature quenching was carried out to be consistent with our mechanical unfolding experiment in which the DNA molecules were quickly (~5.5 pN/sec) relaxed to zero force after unfolding (see below). PAGE gel showed less DMS modification for the four G4 tracts, which was manifested by the reduced band intensity due to less piperidine cleavage, in a 10 mM MES buffer (Lane 1 in Figure 5.1A: pH 5.5 with 100 mM K\(^+\)) or in a 10 mM Tris buffer (Lane 5: pH 7.4 with 100 mM K\(^+\), detailed quantitation see Materials and Methods). Compared to the control in which G-quadruplex is not expected to form (Lanes 2 & 6: without salt; Lanes 3 & 7: with 100 mM Li\(^+\)), the increased cleavage
protection in the four G4 tracts suggests the formation of a G-quadruplex employing these G4 tracts. Calculation of the fold protection showed ~2.5 fold higher protection of these G4 tracts in K+ compared to Li+ at the same pH.

Next, we performed bromine (Br₂) footprinting to probe the formation of i-motif in the same dsDNA, which contains a 5′-32P labeled i-motif forming C-rich strand, 28,42,43 and an unlabeled complementary G-rich strand (see sequence above). However, samples treated with the same procedure (mix→heat→cool) as that in the DMS footprinting did not show any formation of i-motif in the pH 5.5 buffer with 100 mM KCl (Figure 5.1B). Whereas lane 2 suggested that the folding of i-motif in the presence of a complementary strand cannot compete with the duplex formation during the mix→heat→cool process; lane 4 implied that with the same process, the co-existence of the i-motif with the G-quadruplex formed in the complementary strand is not favored. This surprising result indicates that in the single-stranded region complementary to the strand where G-quadruplex has already formed (Figure 5.1A), i-motif does not fold even under favorable pH possibly due to steric hindrance (see Discussion).

We further evaluated the stability of a preformed i-motif in the presence of a complementary strand. To this end, the G-rich and C-rich strands were heated at 95 °C and cooled separately before mixing at 30 °C for 15 min (the heat→cool→mix procedure). Notice that this procedure is different from the previous practice in which two complementary strands are mixed before the heating and cooling steps. Compared to the cytosines in the ACA loops, lanes 1 & 2 in Figure 5.1C show significantly reduced
(−9 fold) Br₂ cleavage of cytosines closest to the i-motif formation regions (cytosines highlighted by the green open circles). These observations are in accordance with the literature that shows cytosines in dsDNA are about 10 times more protected than those in ssDNA context. In the i-motif forming region, the protection of the C4 tracts (highlighted with vertical black lines) with respect to the loop cytosines was obvious in the pH 5.5 MES buffer with either 100 mM K⁺ (Lane 1, Figure 5.1C) or Li⁺ (Lane 2), which demonstrates that under the heat→cool→mix procedure, the preformed i-motif is stable in the presence of the complementary strand for at least 15 min. Further evidence showed that such a preformed i-motif was stable in dsDNA even after 5 days (see below). As a control, no protection of cytosines was observed in the pH 7.4 Tris buffer (Lane 3, Figure 5.1C) in which ILPR i-motif was not expected to form. These experiments demonstrate the strong stability of either ILPR G-quadruplex or preformed i-motif in the dsDNA context.

5.2.2 Laser-tweezers experiments confirm the presence of ILPR tetraplexes in a dsDNA template

Formation of folded structures in the duplex ILPR DNA was confirmed by laser-tweezers based single-molecule experiments (Figure 5.2A). The dsDNA construct consisted of the ILPR tetraplex forming sequence, 5'-CACA GGGG TGT GGGG ACA GGGG TGT GGGG T, which was sandwiched between two dsDNA handles with 1800
bp and 2000 bp in length (See Materials and Methods in Chapter 2). This construct was tethered between two optically trapped beads through one of the strands to prevent

**Figure 5.2** Single-molecule investigation of G-quadruplex/i-moif in the double-stranded ILPR DNA. (A) Schematic of the laser-tweezers set up. The inter-conversion between tetraplex and duplex is highlighted in the red dotted box. (B) An overlay of seven force-extension ($F$-$X$) curves from the same molecule at pH 5.5 in a 10 mM MES buffer with 100 mM K$^+$. The stretching and relaxing curves are indicated by blue dotted and red solid arrows, respectively. Top and bottom insets are blowup regions that show the refolding (black) and unfolding (green bold line) events, respectively. (C) Change in contour length ($\Delta L$) histograms. (D) Unfolding force ($F_{\text{unfold}}$) histograms for fully folded species. (E) Unfolding work histograms obtained from the hysteresis area between the stretching and relaxing $F$-$X$ curves. The vertical shaded strips in (E) represent the work (mean ± σ) equivalent to the $\Delta G_{\text{unfold}}$ (dashed green and gray lines in the bottom panel correspond to the blue dotted and red solid histograms, respectively). Experiments were performed in a 10 mM Tris buffer (top panel: pH 7.4, 100 mM K$^+$), or in a 10 mM MES...
buffer (middle panel: pH 5.5, 100 mM Li$^+$; bottom panel: pH 5.5, 100 mM K$^+$). Histograms are fitted with one or two-peak Gaussian functions.

the accumulation of template superhelicity. The two beads were moved apart to stretch the DNA construct until an unfolding event was observed (Figure 5.2B) in a 10 mM Tris (pH 7.4) or in a 10 mM MES (pH 5.5) buffer with different ionic conditions (100 mM K$^+$ or Li$^+$). Histograms of the change in contour length ($\Delta L$) due to the unfolding, the rupture force histograms, and the unfolding work histograms were plotted in Figure 5.2 C-E.

Of these histograms, the $\Delta L$ (the ~7 nm population in Figure 5.2C and Table 5.1) under all three conditions correspond to the unfolding of a structure that contains ~25 nucleotides (nts) (see Figure 5.3 and Equation 3.1 for calculation). Within experimental error, this number is identical with the nucleotides employed in a fully folded G-quadruplex or i-motif structure in ILPR dsDNA. We also observed a partially folded C-rich species at pH 5.5 with 100 mM Li$^+$ ($\Delta L$ ~4.5 nm population in the middle panel of Figure 5.2C), which was consistent with previous findings$^{28}$ under similar conditions.
Table 5.1 Change in Contour Length ($\Delta L$), Rupture Force ($F_{\text{unfold}}$), Change in Free Energy of Unfolding ($\Delta G_{\text{unfold}}$), Bias of $\Delta G_{\text{unfold}}$, and Percentage Formation (%) for fully folded tetraplex structures at different pH and ionic conditions at 23 ºC. The values in parentheses are standard deviations.

<table>
<thead>
<tr>
<th>pH</th>
<th>Salt ions</th>
<th>$\Delta L$ (nm)</th>
<th>$F_{\text{unfold}}$ (pN)</th>
<th>$\Delta G_{\text{unfold}}$ (kcal/mol)</th>
<th>Bias of $\Delta G_{\text{unfold}}$ (kcal/mol)</th>
<th>% Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>K$^+$</td>
<td>6.7 (± 0.2)</td>
<td>24 (± 1)</td>
<td>7.8 (± 1.4)</td>
<td>-0.5</td>
<td>18</td>
</tr>
<tr>
<td>5.5</td>
<td>Li$^+$</td>
<td>7.2 (± 0.1)</td>
<td>28 (± 2)</td>
<td>8.7 (± 1.0)</td>
<td>1.0</td>
<td>34</td>
</tr>
<tr>
<td>5.5</td>
<td>K$^+$</td>
<td>6.7 (± 0.2)</td>
<td>17 (± 1)</td>
<td>7.4 (± 0.5)</td>
<td>1.2</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(left peak)</td>
<td>(left peak)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>36 (± 1)</td>
<td>10.4 (± 1.2)</td>
<td>1.1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(right peak)</td>
<td>(right peak)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.3 Calculation of the number of nucleotides ($N$) involved in the tetraplex structures. The diagram shows that unfolding of a tetraplex structure, a G-quadruplex for example (top, green frames represent the G-quartets), leads to a dsDNA (bottom). The red double-head arrow represents the end-to-end distance ($X$) for the structure. The number of nucleotides ($N$) involved in the structure is calculated using equation 3.1, $^{28, 96}$.
in which the contour length of single nucleotide ($L_{\text{single nucleotide}}$) is substituted with the contour length of each base pair ($L_{\text{bp}}$) in the B form of dsDNA ($L_{\text{bp}} = 0.34$ nm).

With $\Delta L = 6.7 \pm 0.2$ nm (Table 5.1) and an estimation of $X$ from various G-quadruplex structures after counting for different number of G-quartet stacks ($X = 1.5$ nm for the parallel G-quadruplex [PDB code, 1KF1]; $X = 1.6$ nm for the hybrid-1 and hybrid-2 quadruplexes [PDB codes, 2HY9 and 2GKU]; and $X = 2.1$ nm for the basket type quadruplex [PDB code, 143D]), we obtained $N$ as $24 \pm 1$, $25 \pm 1$, and $26 \pm 1$ nts for respective structures. Similarly, with $\Delta L = 7.2 \pm 0.1$ nm (Table 5.1) and an estimation of $X$ from various i-motif structures (PDB codes, 1ELN, 1A83 and 1YBR, $X = 0.8$ nm for average narrow groove distance and $X = 1.3$ nm for average wide groove distance), we obtained $N$ as $24 \pm 1$ and $25 \pm 1$ nts for respective structures. These calculations confirmed the formation of fully folded tetraplex structures in the ILPR duplex.

Unfolding force histograms for the fully folded species showed a single population in a pH 7.4 Tris buffer with 100 mM K$^+$ ($F_{\text{unfold}} = 24 \pm 1$ pN, Figure 5.2D, top) or in a pH 5.5 MES buffer with 100 mM Li$^+$ ($F_{\text{unfold}} = 28 \pm 2$ pN, Figure 5.2D, middle). Since only G-quadruplex can form under the former condition whereas only i-motif can fold in the latter buffer (Figure 5.1 and literature), we assigned the folded species at pH 7.4 as the G-quadruplex while those at pH 5.5 as the i-motif. These assignments were confirmed by control experiments that showed little formation of any fully folded structures (3.2% at pH 7.4 with 100 mM Li$^+$, Figure 5.4 and Table 5.1), which was expected for respective tetraplexes. In comparison, the experiments shown in
Figure 5.2C-D had a percentage formation in the range of 18 - 34% (Table 5.1). It is noteworthy that the unfolding force for these ILPR structures has a broad distribution, which is similar to those of tetraplexes in the ssDNA context.\textsuperscript{28,121} It may reflect the fact that DNA tetraplex structures are rather dynamic.\textsuperscript{45,189} Single-molecule experiments described here corroborate the footprinting findings that either G-quadruplex or i-motif can stably exist in the dsDNA context.

**Table 5.2** The percentage of mechanical unfolding events for the control experiments in different buffers at 23 °C.

<table>
<thead>
<tr>
<th>Buffer/pH</th>
<th>Ion</th>
<th>DNA Construct</th>
<th>Unfolding Events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM Tris/pH 7.4</td>
<td>100 mM K\textsuperscript{+}</td>
<td>dsDNA handles only</td>
<td>1.9 (Fully Folded)</td>
</tr>
<tr>
<td>10 mM MES/pH 5.5</td>
<td>100 mM Li\textsuperscript{+}</td>
<td>dsDNA handles only</td>
<td>1.0 (Part. Folded)</td>
</tr>
<tr>
<td></td>
<td>100 mM Li\textsuperscript{+}</td>
<td>dsDNA with G/C rich sequences</td>
<td>1.6 (Fully Folded)</td>
</tr>
<tr>
<td>10 mM Tris/pH 7.4</td>
<td>100 mM Li\textsuperscript{+}</td>
<td>dsDNA with G/C rich sequences</td>
<td>3.1 (Part. Folded)</td>
</tr>
<tr>
<td></td>
<td>100 mM Li\textsuperscript{+}</td>
<td>dsDNA with G/C rich sequences</td>
<td>3.2 (Fully Folded)</td>
</tr>
</tbody>
</table>

As shown in Table 5.2, we observed rare unfolding events (< 2%) when a DNA construct without G-quadruplex/i-motif forming sequences (dsDNA handles only) was mechanically stretched at pH 5.5 with 100 mM Li\textsuperscript{+} or at pH 7.4 with 100 mM K\textsuperscript{+}. When a DNA construct with ILPR G-quadruplex/i-motif forming sequence was used, ~ 3% unfolding events were observed for partially or fully folded populations (Figure 5.4) at pH 7.4 with 100 mM Li\textsuperscript{+}. These results were consistent with the DMS footprinting of a 87-bp dsDNA in the same buffer where no protection of the G4 tracts (Figure 5.1A, lane 7) or the C4 tracts (Figure 5.1B, lane 1) was observed. As a comparison, 18% unfolding events were observed at pH 7.4 with 100 mM K\textsuperscript{+}, 44% were observed at pH 5.5 with 100
mM Li\(^+\) (10\% partially folded and 34\% fully folded, see Figure 5.2C, middle panel, and Table 5.1); and 33\% were observed at pH 5.5 with 100 mM K\(^+\) (14\% left peak population and 19\% right peak population, see Figure 5.2D, bottom panel, and Table 5.1). These results confirmed the formation of G-quadruplex or i-motif in the DNA construct containing the ILPR G-quadruplex/i-motif forming sequence.

Figure 5.4. Histogram of change in contour length (\(\Delta L\)) in a pH 7.4 Tris buffer with 100 mM Li\(^+\) at 23 °C. The histogram was fitted with a two-peak Gaussian function (red solid curve) and further deconvoluted randomly into left (black-dotted curve) and right (blue-dotted curve) populations (see Materials and Methods). These two populations represent a partially folded (left) and a fully folded (right) species.

5.3.3 G-quadruplex and i-motif are mutually exclusive in the double-stranded ILPR fragments

In a pH 5.5 buffer (10 mM MES) with 100 mM K\(^+\), two fully folded populations with unfolding forces of 17 (±1) and 36 (±1) pN were observed (Figure 5.2D, bottom panel & Table 5.1). Either a G-quadruplex, an i-motif, or a G-quadruplex with an i-
motif, can form in the dsDNA in this buffer. To identify the exact species, we compared the change in free energy of unfolding ($\Delta G_{\text{unfold}}$) for structures formed in each buffer. Mechanical unfolding allowed a unique capability to retrieve $\Delta G_{\text{unfold}}$ by application of Jarzynski equation (equation 2.2, see Figure 5.2B as an example and Materials and Methods in Chapter 2 for detailed calculation) for non-equilibrium systems.\textsuperscript{94, 95} This method recovers $\Delta G_{\text{unfold}}$ by using an exponentially weighted algorithm to count for the dissipated work during the non-equilibrated unfolding processes. According to this weighing pattern, smaller unfolding work contributes more to the change in free energy. Therefore, $\Delta G_{\text{unfold}}$ is expected to be smaller than the average unfolding work performed under non-equilibrium conditions. When histograms of this unfolding work were plotted for species in all three buffer conditions, we indeed observed that values of $\Delta G_{\text{unfold}}$ were significantly below the average work (shaded strips in Figure 5.2E). Previously, this method has allowed accurate retrieval of the $\Delta G_{\text{unfold}}$ for hairpins\textsuperscript{95} and tetraplexes in ssDNA.\textsuperscript{28, 121}

As shown in Table 5.1, $\Delta G_{\text{unfold}}$ were similar between a G-quadruplex (7.8 ± 1.4 kcal/mol, pH 7.4 with 100 mM K\textsuperscript{+}) and an i-motif (8.7 ± 1.0 kcal/mol, pH 5.5 with 100 mM Li\textsuperscript{+}). When the two populations at pH 5.5 with 100 mM K\textsuperscript{+} were deconvoluted (Figure 5.2D, bottom panel, see Materials and Methods in chapter 2), similar values of $\Delta G_{\text{unfold}}$ were obtained for these species ($\Delta G_{\text{unfold}}$ for the high-force population, 10.4 ± 1.2 kcal/mol; $\Delta G_{\text{unfold}}$ for the low-force population, 7.4 ± 0.5 kcal/mol, see Figure 5.2E, bottom panel and Table 5.1). The accuracy of these $\Delta G_{\text{unfold}}$ values was reflected by their
small biases (within ±1.2 kcal/mol, see Table 5.1), which were calculated according to literature. In fact, these $\Delta G_{\text{unfold}}$ values are in the same range as that of a stand-alone G-quadruplex or an i-motif in the dsDNA context (see above). They are about half as that expected to unfold a G-quadruplex and an i-motif together (16.5 kcal/mol). These results indicate that only one tetraplex structure (either i-motif or G-quadruplex, but not both) is formed in a given dsDNA molecule. Such a scenario is fully consistent with the footprinting results (Figure 5.1A and Figure 5.1B) that indicate only G-quadruplex, but not i-motif, is formed in the dsDNA when complementary strands are mixed before the heating and cooling steps (see discussion below).

5.2.4 Kinetic experiments reveal that G-quadruplex is more stable than i-motif in dsDNA

To identify each species in the pH 5.5 MES buffer with 100 mM K$^+$ (Figure 5.2D, bottom), we performed kinetic footprinting experiments using the heat→cool→mix procedure. First, two complementary strands used in the chemical footprinting were heated separately at 95 °C for 10 min and quenched rapidly in a water bath set at 30 °C, a temperature significantly below the $T_m$ of the ILPR G-quadruplex (77 °C and 88 °C for parallel and antiparallel G-quadruplexes in ssDNA, respectively) or the $T_m$ of the ILPR i-motif (37 °C) in ssDNA. This facilitates the formation of respective structures in ssDNA. Notice that even for ILPR i-motif, 30 °C is well within the plateau of the melting curve that indicates fully folded structures (data not shown). These two ssDNA strands with preformed tetraplex structures were then mixed at equimolar ratio
and incubated at 30 °C to facilitate the hybridization of flanking dsDNA handles over 5 days. To allow the DMS or Br₂ footprinting on G-quadruplex (Lanes 2-6, Figure 5.5A) or i-motif (Lanes 7-11), respectively, either the G-rich or the C-rich strand in the DNA mixture was 5′-³²p labeled.

**Figure 5.5** Kinetic analyses of G-quadruplex and i-motif in the double-stranded ILPR DNA at pH 5.5. (A) DMS (lanes 1-6) and bromine (lanes 7-12) footprinting of equimolar mixture of G-rich and C-rich strands after the heat→cool→mix procedure (see text). All lanes are from experiments performed in a 10 mM MES buffer at pH 5.5 with 100 mM K⁺, except lane 1 (with 100 mM Li⁺) and lane 12 (pH 7.4). Lanes 2-6 and lanes 7-11 represent footprinting on the dsDNA samples taken from 5 intervals during 5 days. Open circles at the right of lane 11 depict the cytosines in the ACA loops. (B) Fold protection of C4 or G4 tracts over time. DMS (filled circles) and bromine (squares) footprinting results are fitted with a dotted line and an exponential curve, respectively. (C) Rupture force histograms for structures in the double-stranded ILPR DNA by mechanical unfolding experiments performed in the 10 mM MES buffer at pH 5.5 with 100 mM K⁺.
The DNA molecules were incubated at zero force with 1 sec (grey) or 60 sec (black) before unfolding.

Control experiments were performed at pH 5.5 with 100 mM Li\(^{+}\) for DMS footprinting (Lane 1) or at pH 7.4 for Br\(_2\) (Lane 12) footprinting assays. The quantitation (see Materials and Methods) showed that while the G-quadruplex population remained relatively constant over time (dotted line, Figure 5.5B), the i-motif decreased slowly (solid curve) with its presence still significant after five days. In addition, the i-motif population monitored by Br\(_2\) footprinting at pH 5.5 with 100 mM Li\(^{+}\), a condition in which the G-quadruplex is not expected to form, showed similar decrease in i-motif population (data not shown) over 15 min to 3 days. These results suggest that the G-quadruplex is thermodynamically more stable than the i-motif in the presence of a corresponding complementary strand.

To correlate the kinetic property of tetraplex species obtained from footprinting experiments, next, we performed kinetic analysis on the mechanical unfolding of these species in the pH 5.5 MES buffer with 100 mM K\(^{+}\). After unfolding a specific tetraplex structure, we varied incubation time at 0 pN to allow the refolding of the structure. We then carried out next round of stretching procedure, in which the observation of a rupture event would indicate a tetraplex had been folded during the incubation.\(^{190}\) When we analyzed the rupture force populations at two incubation times (1 sec and 60 sec), we found that the high-force species increased its population with respect to the low-force species over time (Figure 5.5C, ratio of high-force to low-force population decreases from 1.6 to 1.1). Since the footprinting assays suggested that the i-motif population
decreases with time, we assigned the high-force species as a G-quadruplex, whereas the decreasing low-force population as an i-motif. This assignment is consistent with the fact that ILPR G-quadruplex ($\Delta G_{\text{unfold}}$, 10.4 kcal/mol, Table 5.1) has higher stability compared to ILPR i-motif ($\Delta G_{\text{unfold}}$, 7.4 kcal/mol, Table 5.1) in the same buffer.\textsuperscript{28, 121}

Compared to the $\Delta G_{\text{unfold}}$ for tetraplexes in the context of ssDNA, changes in the free energy of unfolding for the tetraplexes in dsDNA are significantly smaller. This could be due to the reannealing of the complementary strands after unfolding of the tetraplexes (Figure 5.2A, due to the force-induced melting discussed below, partial, rather than full, reannealing is more likely), which reduces the free energy of the unfolded state with respect to the ssDNA.

Figure 5.6 summarizes our observations. In a pH 7.4 buffer with 100 mM Li$^+$, the duplex DNA was predominant (Figure 5.6, top left), as suggested by few unfolding events (Figure 5.4 & Table 5.2) and lack of G4 or C4 protection in the footprinting experiments (Figures 5.1A and 5.1B). At pH 7.4 with 100 mM K$^+$, G-quadruplex formed predominantly (Figure 5.6, top right). In a pH 5.5 buffer with 100 mM Li$^+$, i-motif prevailed (Figure 5.6, bottom left). Finally, at pH 5.5 buffer with 100 mM K$^+$, either G-quadruplex or i-motif, but not both, formed in each dsDNA (Figure 5.6, bottom right). Whereas the G-quadruplex was stable over time, the i-motif reduced its population during the same period.
Figure 5.6 Major species under different pH and ionic conditions.

5.3 DISCUSSION

That i-motif did not fold in the dsDNA context after thermal melting procedure (Figure 5.1B) was probably due to the competition from the dsDNA formation during the reannealing. Previously, ILPR i-motif in the single-stranded C-rich sequence has shown a melting temperature of 37 °C, which is much lower than the $T_m$ expected for the dsDNA with the same sequence (80.5 °C by Nearest Neighbor calculation). Therefore, there is a temperature range (37-80.5 °C) in which dsDNA prevails over i-motif. At lower temperatures, the formation of i-motif becomes rather difficult due to the insurmountable energy barrier to separate the two complementary strands that have already been hybridized. On the other hand, ILPR G-quadruplex formed in the single-stranded G-rich sequence has $T_m$ of ~77 °C and ~88 °C for parallel and antiparallel quadruplexes respectively. Since these melting temperatures are comparable with the $T_m$ for the dsDNA, a mixture of dsDNA and G-quadruplex is expected to form during the cooling. Once the G-quadruplex is formed (Figure 5.1A), it generates a single-stranded complementary region in which i-motif may form. However, Br$_2$ footprinting showed no
evidence of i-motif formation (Figure 5.1B). This result strongly supported our laser-tweezers finding that only one tetraplex, in this case G-quadruplex, can form in dsDNA (Figure 5.2E).

During mechanical folding/unfolding experiments, we observed that either ILPR G-quadruplex or i-motif formed in the dsDNA construct (Figure 5.2), which had been stretched above 30 pN in the previous pulling cycle at room temperature (23 °C). Such a result was in contrast with footprinting assays, which showed only G-quadruplex, but not i-motif, formed in the dsDNA context. We believe it is the different melting process that leads to the different observations. It is known that dsDNA can melt by the force-induced melting process. 192, 193 As the tension in the dsDNA increases, it generates locally melted regions that can lead to the formation of either a G-quadruplex or an i-motif. In the thermomelting and reannealing processes, however, the competition from the dsDNA hybridization prevents the formation of i-motif while allows G-quadruplex to form (see above). The unprecedented mutual exclusiveness between the G-quadruplex and i-motif is perhaps due to the steric hindrance. It is possible that formation of one tetraplex restricts the degree of freedom for the folding of the other species in the complementary strand. It has interesting biological implications. For example, it provides versatility in the regulation since either i-motif or G-quadruplex may be involved whilst each of the species has a different set of interacting proteins. 42, 57, 58, 111, 194, 195 In addition, the different stability or formation kinetics in the G-quadruplex and i-motif may play an important role for differential regulation of gene expressions.
Compared to biochemical techniques, the determination of the $\Delta G_{\text{unfold}}$ by the laser-tweezers approach provides a unique and straightforward capability to evaluate the formation of tetraplexes in dsDNA. The dsDNA construct used in the mechanical folding/unfolding experiments closely resembles the physiological situation in which tetraplex forming sequences in the promoter are always flanked by double-stranded DNA regions. Likewise, the force-induced melting is physiologically more relevant with respect to the thermal melting or denaturant melting. Whereas there exists little possibility for the latter two processes \textit{in vivo}, numerous events, such as DNA replication, RNA transcription, and cell growth and division, can generate tension in a DNA template, leading to force-induced melting. The other unique aspect of the laser-tweezers approach is that it can evaluate the mechanical stability of the DNA tetraplexes. Here, the ILPR tetraplex structures in the duplex DNA demonstrate a mechanical stability ($F_{\text{rupture}} \geq 17$ pN) larger or comparable to the stall force of polymerases (14-25 pN).\textsuperscript{120, 172-174}

From mechanical perspective alone, this could justify the regulatory role a DNA tetraplex may play in the expression of human insulin inside cell in which dsDNA is the predominate form.

\section*{5.4 CONCLUSION}

By combining two complementary techniques, chemical footprinting and mechanical folding/unfolding, we have provided convincing evidence that either G-quadruplex or i-motif, but not both, forms in the double-stranded ILPR region. Although
experiments were performed at pH 5.5, we anticipate that general aspects of this conclusion can provide insights for the formation of tetraplexes in dsDNA at physiological pH under which recent studies have shown that, in addition to G-quadruplex, i-motif can also fold.\textsuperscript{32,147} It is interesting to see whether this conclusion still holds under \textit{in vivo} conditions such as on a DNA template with negative superhelicity or in a molecular crowding environment.
CHAPTER 6
CONCLUDING REMARKS

Single- and double-stranded DNA fragments from the human ILPR were chosen as model systems for single-molecule manipulation of DNA secondary structures under different buffer conditions. Laser tweezers were used as a primary technique in these studies, which were also investigated by complementary ensemble-average techniques such as CD, UV-Vis, chemical footprinting, and mutational analysis. This dissertation covered the roadmap to determine the mechanical stability of G-quadruplex and i-motif structures first in single-stranded DNA then in double-stranded ILPR DNA fragment.

Chapter 3 describes the studies on the mechanical stability of the structures formed in the cytosine-rich fragment of human ILPR (published in reference 28) using laser tweezers. We found that a partially folded structure co-exists with a fully folded i-motif structure. These structures were otherwise not distinguishable by the bulk techniques. Surprisingly, the partially folded structure demonstrated mechanical stability comparable to that of i-motif. It is noteworthy that the single-molecule approach used in this study allowed to identify as low as ~2% i-motif population near the physiological pH, which was otherwise not observed in CD and chemical footprinting methods. This demonstrates the superior sensitivity of single-molecule method over bulk techniques. Additionally, the fact that both i-motif and partially folded structures having unfolding forces higher than the stall force of RNA polymerase (RNAP) provided the possibility
that the formation of these structures could modulate RNA transcription catalyzed by RNAP.

The follow up study on the truncated ILPR sequence with only three C-rich repeats (Chapter 4, published in reference 30) explored the plausible folding of the partially folded structure co-existed with the ILPR i-motif. This study revealed an unprecedented structure stabilized by the C.CH\(^+\) pairing between distally located C-rich tracts. The fact that this structure demonstrates a lower pH dependency relative to the ILPR i-motif while maintaining a comparable mechanical stability has the following important implication: its folding near physiological pH sets a premise for its biological roles. Furthermore, the formation of intermolecular i-motif when supplemented with a fragment, 5'-CCCCTGT oligo, has provided further insights on the i-motif folding mechanism, which suggests the strand-by-strand folding pathway.

The secondary structures in ILPR dsDNA have been studied in Chapter 5 (published in reference 40). Based on the results from the single-molecule manipulation and chemical footprinting experiments, the mutual exclusivity of i-motif and G-quadruplex structures in the dsDNA context has been proposed. This was the first demonstration that G-quadruplex and i-motif are mutually exclusive in dsDNA. The exclusivity could be the result of steric hindrance as the structures must form at the same location within the dsDNA construct. The molecular details of such a mutual exclusivity are yet to be understood, and therefore, are the subject of further study. As in ssDNA experiments, the strong mechanical stability (rupture force ≥ 17 pN) of the structures near
promoter region under various buffer conditions (pH 5.5-7.4) showed the relevance of these structures when mechanical strength is an issue in the biological processes such as RNA transcription and DNA replication. For example, shortening of ILPR region due to blockade or slippage of a DNA polymerase during replication has been proposed for the down regulation of insulin which can result type I diabetes. Additionally, such an implication could provide guidelines for targeting these structures for therapeutic applications through interference of RNA transcriptions.

In summary, formation and the mechanical stability of the i-motif and G-quadruplex structures in human ILPR are studied under diverse circumstances (ssDNA, dsDNA, different pH, and salt ions). Under near physiological pH and salt ion concentration, ILPR i-motif and G-quadruplex resist unfolding until tension reaches ~20 pN. Specifically, these studies can be helpful to understand the role of i-motif and G-quadruplex in the regulation of insulin gene. I anticipate the methods and the model system evaluated in this dissertation can be used to explore and understand other genes whose promoters contain G-quadruplex/i-motif forming sequences.
APPENDIX A: PERMISSION TO REUSE COPYRIGHTED MATERIAL

Copyright Clearance Center

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

Author: Soma Dhakal, Joseph D. Schonhoff, Deepak Koirala, Zhongbo Yu, Soumitra Basu, and Hanbin Mao

Publication: Journal of the American Chemical Society

Publisher: American Chemical Society

Date: Jul 1, 2010

Copyright © 2010, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK  CLOSE WINDOW

**Copyright**: © 2012 Dhakal et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
REFERENCES


