CHARACTERIZATION OF INTER-GENOMIC AND TELOMERIC G-QUADRUPLEX STRUCTURES THROUGH ATOMIC FORCE MICROSCOPY

BY

ALEX RAYMOND KREIG

THESIS

Submitted in partial fulfillment of the requirement
for the degree of Master of Science in Bioengineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Assistant Professor Sua Myong

ABSTRACT

The G-quadruplex DNA structure, a secondary folding motif of DNA, is found throughout the genome. The dynamics of these structures has been seen to affect processes such as telomerase elongation and regulation of gene transcription. Typical investigation techniques require these structures to be present in very high concentrations and in non-biological buffers. Adding to the issues with current methods of investigation, this structure currently relies on bulk measurements, which limit the accuracy with which it can be measured. In this report, we present newly developed atomic force microscopy-based assays for quantifying G-quadruplex folding patterns with sub nanometer accuracy. The assays are based upon a duplex DNA handle ranging from 18 to 525 base pairs in length containing an overhang sequence utilized to attach the DNA to a mica surface. The overhang sequence pairs with a possible G-quadruplex forming sequence and allows for the retention and identification of the folded structure. Using a 12 base DNA probe, we demonstrate that the location of G-quadruplex folding on long strands of DNA are not preferred to occur at the distal end, contrary to DMS foot printing studies. We demonstrate that this method enables detection of the two major folding motifs and is able to distinguish height differences between them. The high resolution and low concentration requirements, combined with the ability to image hundreds of single molecules per study, allow this method to provide quick and accurate characterization of G-quadruplex folding.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Sua Myong for being a great mentor scientifically, academically, and in life. The work that is presented here represents a long process of suggestions, guidance, and innovation from her personally. Her endless motivation and enthusiasm continued the project in spite of several initial set back. Finally her attention to detail proved invaluable in understanding the subtlest details of this work.

I also would like to thank Jennifer Amos who was patient enough to teach me proper atomic force microscopy technique but also her personally finding experts within the field to assist with DNA imaging. Without these contacts, combined with the enormous amount of time that was set aside for my research this project would not be completed.

I would also like to thank the members of the Myong lab who have taught me many of the single molecule and molecule biology techniques presented here. Specifically, I cannot thank enough Helen Hwang for her personal and professional support. Without her initial introduction to the topic covered here my research goal would have taken a vastly different path.

Lastly, I thank my family and Samantha Knoll for their continual encouragement and moral support.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Genetic Basis of Cancer

Cancer is a broad group of related diseases that present with uncontrollable cellular growth resulting in the disruption of regular tissue function. Aggressive dispersal of the primary tumor, known as metastasis, to distal regions of the organism results in a systematic disruption of normal physiological functions and, without intervention, typically leads to death (1). A nondescript classification of cancer by tissue type as a single disease is commonly made. Cancer is thousands of diseases marked by one fundamental feature: a defective genome (2, 3). In cancerous tissue, the genetic instructions for growth regulation as well as other important pathways are permanently altered, resulting in atypical cellular proliferation patterns and response.

The vast web of interconnected genetic networks and enormity of the sample space of the genome suggests that cancer is one of the most complicated diseases. Several attempts have been made to simplify cancer through classification of its characteristics. The most successful approach to classifying cancer resulted in the development of six "hallmarks of cancer" [2, 3]. The five hallmarks of cancer are (1) sustained proliferative signaling, (2) evading growth suppressors, (3) Activating metastasis, (4) Inducing angiogenesis, (5) resisting apoptosis, and (6) enabling limitless replicative potential. By definition, every cancer shows these six hallmarks through some combination of genetic defects. Thus, the removal of just one these hallmarks can result in an effective therapeutic treatment.

1.2 Limitless Replicative Potential

Limitless replicative potential refers to the ability of a cell to function in an immortal phenotype. The number of division iterations a non-cancerous cell can undergo is limited by telomere length. This property is referred to as the Hayflick limit, and is defined as the total number of division cycles a cell can undergo before the division process is permanently stopped. This halt in division is known as senescence.

The oncogenic phenotype of limitless replicative potential allows a cell to bypass the Hayflick limit, resulting in the disappearance of senescence [4]. Limitless replicative potential is a highly targeted cancer hallmark in the development of new therapeutic regimens [5]. This is because the mechanisms governing replicative cellular phenomena are generally understood. There are two ways a cell can bypass the Hayflick limit: telomerase up-regulation and via the alternative lengthening pathway (ALT). Further enhancing scientific interest in targeting this replicative potential hallmark of cancer is the fact that greater than 85% of cancers exhibit up-regulation of telomerase (6). Most importantly, both telomerase up-regulation and the ALT pathway act upon one structure, the telomere.

1.3 Telomere Structure and Function

The processes through which cells respond to breaks in double stranded DNA (dsDNA) are robust, and contribute to maintaining genomic stability and integrity. The linear chromosomes seen in Eukaryotic cells present a problem not present in circular chromosomes of the majority of Prokayotic cells. Each chromosome in a Eukaryotic cell has a terminal beginning and an end, rather than an infinite, circular shape. The problem with this linear chromosome shape is that a lateral dsDNA break and the termination of a chromosome are seen by a cell as chemically identical. As a result, the cellular response to either a dsDNA break or chromosome termination is the same—re-synthesis of the dsDNA by fusing two breaks. The repair process can be detrimental to cell governance as it can alter the internal biological clock of the cell [7]. This response leads to limitless replicative potential, forming conditions favorable for cancer. A biological mechanism that reconciles this issue is the telomere, a specific DNA sequence is found at the ends of each chromosome that signals chromosome termination. This structure is composed of a repeated, specific DNA sequence (TTAGGG in vertebrates). In humans there are approximately 2000 repeats of TTAGGG in dsDNA, followed by 20 to 30 repeats of TTAGGG in a single stranded form.

The large number of repeats found in the telomere is necessary to offset a natural phenomenon known as the "end replication problem". The "end of replication problem" is that DNA polymerase cannot effectively replicate distal regions of the chromosomes during cell division. In cell division, the process in which a parent strand of DNA becomes two daughter strands involves a shortening of the distal region of the telomere due to the directional nature of DNA polymerase (figure 1). In the lagging strand of DNA, a RNA primer sequence is added in order for polymerase to function. Upon reaching the terminal end of the chromosome, there is no space for the primer sequence to form, thus there is no binding site for DNA polymerase resulting in telomere shortening. The typical rate of telomere shortening has been measured to be 3-5 base pairs per end, per division [8]. This problem is further exacerbated by UV/oxidative damage and exonuclease activity.

The length of telomere is utilized as a signal within a cell (figure 2). When the length of a telomere approaches a critical length, a cell responds by either going into senescence (halting cell division) or apoptosis (programmed cell death) [9]. When the telomere length is above its critical level, the cell is free to divide. This shift from an actively replicating cell to a senescent cell prevents limitless replication. Thus, in order for a cancerous cell to replicate immortally, telomere length is maintained.

1.4 DNA G-quadruplexes

As mentioned earlier, the single stranded DNA region that comprises the distal portions of the telomere is repetitive in nature (TTAGGG). This sequence is dominated by guanine bases and is capable of folding into a 3 dimensional structure known as a G-quadruplexes [10]. This prism-like structure is stabilized through Hoogsteen base pairing. Unlike traditional Watson Crick base paring in which A and T bases and G and C bases are stabilized through hydrogen bonding, Hoogsteen base pairing requires four G bases to align in a planar structure [11]. This structure is called a tetrad. A quadruplex forms when three tetrads are stacked upon each other (figure 3) Formation of this structure is only possible when a G-quadruplex-forming run (four repeats of three G bases plus a spacing element between each

repeat) occurs. The final requirement for the formation of G-quadruplexes is the presence of sufficient monovalent cations for stabilization. The size of potassium ions makes them well-suited for stabilization of G-quadruplexes. The physiological environment of a cellular nucleus provides a high concentration of monovalent cations, thus promoting the formation of G-quadruplexes [12].

1.5 DNA G-quadruplexes Folding Properties

Two sets of properties define the folding patterns of G-quadruplexes. The first of these properties is the molecular composition of G-quadruplexes. The G-quadruplex can be composed of a single DNA molecule (intramolecular) or multiple DNA molecules (intermolecular). An intramolecular G-quadruplex is composed of a single oligonucleotide containing all four repeats that comprise a G-quadruplex-forming run. An intermolecular G-quadruplex is composed of two or more DNA oligonucleotides that are combined to form the necessary four repeats found in a G-quadruplex-forming run. Research suggests that the predominant form of G-quadruplex folding that occurs within the nuclear regions of a cell is intramolecular [12]. Scarcity of intermolecular folding in the nucleus is due to the high oligonucleotide concentration required for it to occur. Thus, an emphasis in experimental technique must be placed upon obtaining the desired formation of structures (intramolecular vs. intermolecular) given a specific concentration of oligonucleotides.

The second folding property that affects the formation and structure of G-quadruplex is the folding direction. There are three directional labels: parallel, antiparallel, and a mixed hybrid structure [13]. A parallel structure occurs when the intervening sequence (spacing element) between two runs of three G base (G3) repeats traverse the side of the G-quadruplex structure and terminate at the opposite end of the structure. This results in each G3 repeat having the same 5' to 3' directionality. An antiparallel structure alternates G3 repeats and exhibits opposite directionality amongst adjacent repeats (5' to 3' followed by 3' to 5'). This is achieved through the intervening sequences traversing the top or bottom of each G-

quadruplex, resulting in loops that protrude out of the top or bottom faces of the structure. A visualization of these two structures is seen in figure 4. A hybrid structure is a mixture of both parallel and antiparallel folding schemes within one G-quadruplex. These structures occur in a limited number of very specific sequences [14].

Interestingly, these folding properties align with a very important physical property of G-quadruplexes. Melting temperature or stability of the structure has been shown to follow the trends of folding patterns [15]. In the biologically relevant structure of intramolecular G-quadruplexes, higher stability has been seen with structures that fold in a parallel configuration. The telomeric repeat G-quadruplex of $(TTAGGG)_4$ has been shown to fold in a predominantly antiparallel structure in physiologically relevant cation conditions, which matches its melting temperature profile [15, 16].

1.6 G-quadruplexes and Telomere Extension Efficiency

As previously mentioned, the telomere shortens upon replication due to the end replication problem. Telomerase up-regulation and the ALT pathways combats this phenomena. We specifically focus on telomerase as an example of a means to combat end replication due to its higher prevalence (accounts for 85%) [17]. Telomerase is a ribonucleoprotein complex, meaning it is composed of a ribonucleic acid sequence and an enzyme component. This complex is used to synthesize additional nucleotides to the end of the telomere. Its functionality requires that the distal portion of the telomere be accessible [18]. This means that the protein machinery behind the process needs access to this terminal end. As described, the telomere is shortened due to replication and DNA damage. Telomerase is able to bind to the terminal region and thus extend the telomere by adding TTAGGG repeats.

Extension proteins are only able to extend the distal portion of the telomere if the terminal region is accessible. If quadruplex folding incorporates the final repeats of

the single stranded DNA telomeric repeats, this 3-dimensional structure can prevent extensional protein binding [18]. This would prevent elongation in both the possible pathways, through the denial of binding access. Thus if the end of the telomere is not accessible there is a lack of telomere extension [19]. This proves beneficial in cancer treatment research due to the shortening of the telomere with each cellular replication cycle. Thus, the rapid replication of a cancerous cell would swiftly deplete the telomere region resulting in either senescence or apoptosis. Intrinsic cellular processes are capable of unfolding the quadruplex structures thus allowing for binding and extension [20].

External factors are capable of preventing unfolding of the telomere quadruplexes. Specific drugs are capable of binding and stabilizing these structures. With these structures stably formed, lengthening pathways are no longer capable of extending the telomere. Several examples of these compounds are N-Methyl Mesoporphyrin IX, Braco 19, and Telomestatin [21]. These chemical compounds bind to the quadruplex structure, and can prevent their unfolding.

1.7 Genetic Promoter Regions

Unlike the hallmark of cancer, Limitless Replicative Potential, the other five hallmarks of cancer act upon thousands of different genetic targets. This greatly intensifies the difficulty of correcting anyone of the genetic defects causing the phonotypical hallmark expression. Each of these hallmarks relies on a complex interaction between a network of protein products malfunctioning, under-expressing, or overexpressing. These dysfunctions are typically classified in either of two categories: tumor suppressor genes or oncogenes [22]. Tumor suppressor genes are commonly defined by the loss of function phenotype. This means that both copies of an allele for a gene must be either under-expressed or defective. Alteration of only one copy of an oncogene yields a cancer causing effect, meaning either overexpression or a gain of function may result in tumorigenic behavior.

For mRNA production to occur, an enzyme known as RNA polymerase must bind to the DNA directly upstream of a gene. This region of DNA 100-1000 base pairs upstream of the transcription start site (TSS) is referred to as the promoter region. The promoter region has been shown to have regulatory effects in initiation of transcription due to the binding of transcription factors which actively alter RNA polymerase biding affinity [23].

1.8 G-quadruplexes and Regulatory Regions

Increasing amounts of evidence suggest that the G-quadruplex structure may form within the duplex regions of the genome [24]. Further supporting the importance of the G-quadruplex with respect to genetic expression is its prevalence in the promoter regions of genes. G-quadruplexes are commonly located in the promoter regions of oncogenes. Evidence suggests that these structures are capable of down-regulating genetic expression through blocking transcription factors or RNA polymerase function [25]. This matches the function of oncogene control, due to the need to limit high levels of protein expression. The combination of folding identification of G-quadruplex in cellular samples and there suggested effects on transcription control proves this structure an interesting target for cancer therapies.

1.9 Figures

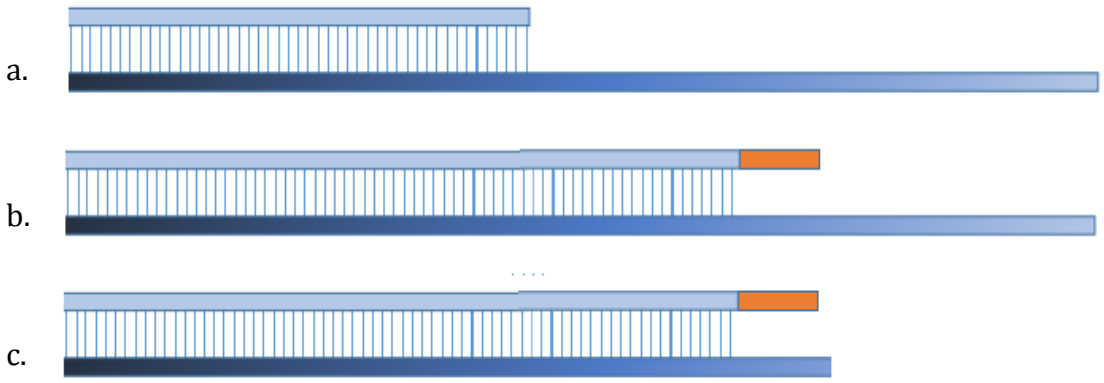


Figure 1 a.) Due to the directional nature of DNA polymerase, the lagging stand of DNA must be synthesized in multiple fragments to be joined together. b.) This process initially starts with the creation of a RNA primer (orange) and synthesis of DNA occurs toward the previously formed section. c.) Due to the inability to attach an RNA primer to the distal portion of a chromosome end, a truncation of the DNA sequence occurs.

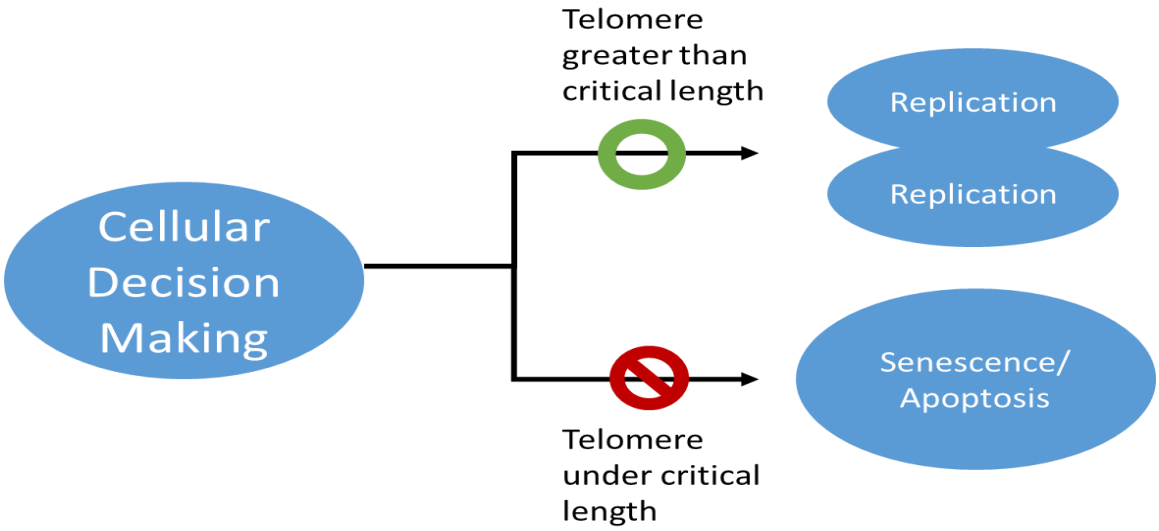


Figure 2 A graphical representation of the genetic sensing pathway that controls the replication of a cell. Note that telomere length can play a role in the decision making process.

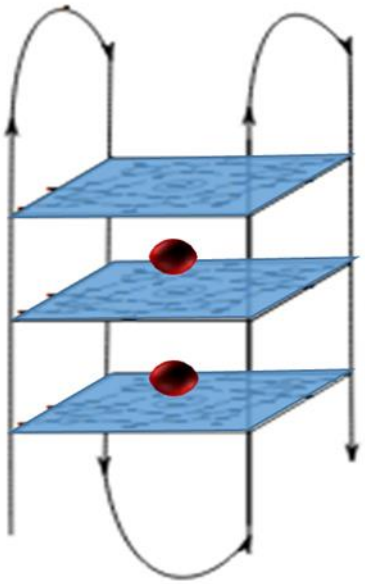


Figure 3 A G-quadruplex structure is shown. The molecule is composed of three G-tetrads (blue) comprising four G run repeats. The molecule is stabilized by monovalent cations located between the tetrads.

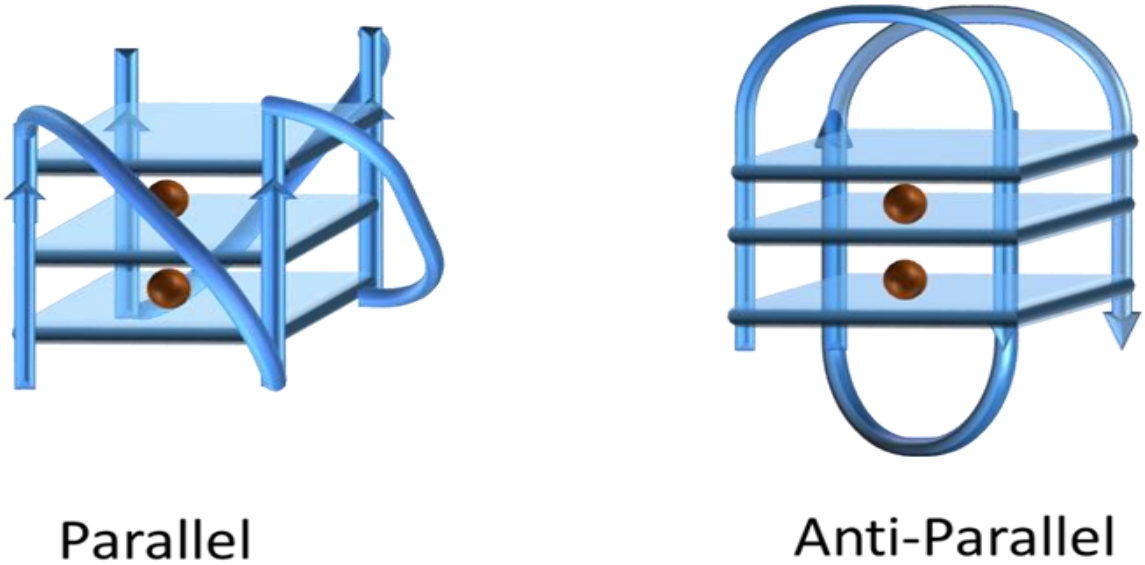


Figure 4 a.) A parallel folding motif of a G-quadruplex is shown. The intervening sequences are seen to transverse the upright sides of the structure with each G run repeat pointed in same direction. b.) The anti-parallel folding motif is shown. The intervening sequence transverse both the top and bottom of the structure. Each G run repeat is seen to alternate direction of travel with its respective neighbors.

CHAPTER 2: ATOMIC FORCE MICROSCOPY (AFM)

2.1 AFM Principles

AFM utilizes Infrared (IR) light to tract the movement of the interaction between a cantilever and a surface. The IR light is reflected off of the back of the cantilever and is collected by a photodiode array. The difference between the voltages produced between two photodiodes is utilized to calculate the relative position of the cantilever in three-dimensional (Z) space. Three piezoelectric components perform the raster scanning of the AFM. This instrument can achieve a lateral resolution of approximately 0.3 nanometers and a height resolution of 100 picometers (or 0.1 nm) [26]. This precision is limited by sensor noise of piezoelectric elements.

AFM was chosen as the imaging platform for investigation of the G-quadruplex structure. It was selected for this study due to functional advantages over other common imaging modalities such as traditional bulk measurements, single molecule fluorescence, electron microscopy, and X-Ray crystallography. Traditional bulk measurements include measurements such as gel shift assays, circular dichroism, and bulk fluorometer readings [27, 28, and 29]. The AFM provides a single molecule reading, while the aforementioned bulk measurement techniques rely on the averaging of many molecules leading to a muddling of the true signals. Single molecule fluorescence has recently been utilized to visualize G-quadruplex folding [30]. This technique relies on a shift in fluorescent output between two fluorophores. Single molecule fluorescence is sufficient for visualizing the dynamics of the quadruplex system, but fails to provide structural data. Because AFM measurements are based on mechanical interaction of a silicon tip and the G-quadruplex, it provides structural information such as width and height. Both electron microscopy and X-Ray crystallography provide similar structural measurements to AFM, but neither technique allows for the imaging of these structures to take place in biologically relevant environment conditions [31]. With AFM, this is not an issue because we have developed the imaging method in a sub micromolar DNA concentration in which the G-quadruplexes are suspended in

biologically similar buffer. Thus, measuring G-Quadruplex structures through an interaction between a nanometer sized AFM tip and surface, can provide 3-dimensional structural data in a biological relevant preparation.

2.2 Tapping AFM

In order to preserve delicate samples such as DNA molecules, the traditional method of direct contact AFM cannot be utilized. To minimize breakage of the DNA molecule, a technique known as tapping AFM was utilized [32]. In the tapping mode, a cantilever is oscillated just below its resonant frequency, typically controlled by a small piezoelectric oscillator. As the cantilever encounters the surface of a sample, the oscillation of the cantilever are dampened. The AFM will correct for the dampening in order to maintain constant amplitude of oscillation, and this movement is translated into a height measurement for a given point [33]. A topographical image can be produced by raster scanning the nanometer sized tip across the surface of the sample.

2.3 DNA Design and Sample Preparation

An 18 base pair double stranded region (18mer) was initially provided as an anchoring structure to the mica surface used in imaging. A 3' single stranded DNA overhang from the duplexed 18mer was designed to form a G-quadruplex upon correct conditions. An additional monitoring DNA element known as C2 was utilized to visualize the extension. This single stranded sequence was composed of 2 complimentary repeats of the G-quadruplex sequence.

Later studies utilized a duplexed DNA handle 525 base pairs in length. This sequence contained an 18 base single stranded overhang that complimented the initial 18 base pair handle mentioned previously. This allowed for the attachment of single stranded G-quadruplex sequences to the terminal end of the structure. The formulation process (an asymmetric polymerase chain reaction) for this molecule will be discussed later.

2.4 Surface Preparation for AFM

An atomically flat surface is required in order to image a molecule on a sub-nanometer scale. The standard surface utilized for AFM imaging of DNA samples is muscovite, a type of mica. There are two material properties that render mica an ideal substrate upon which to image. The first is the ease with which mica can be cleaved, providing a clean and atomically flat surface. The roughness of a new cleavage surface has been measured as less than 0.1 nanometers over distances longer than 1 micron and was confirmed through control measurements. This low roughness scale enables precise imaging of small molecules such as DNA. Upon its cleavage, mica possesses a uniformly negative surface charge. This allows for the attachment of negatively charged DNA through the simple cation bridge [33] Nickel Chloride ($NiCl_2$) and Magnesium Chloride ($MgCl_2$) can be used interchangeably as the divalent cations in order to image DNA. $MgCl_2$ is a more biologically relevant ion, while $NiCl_2$ provides a much stronger binding potential.

CHAPTER 3: OPTIMIZATION OF G-QUADRUPLEXED DNA HANDLE RETENTION IN PHYSIOLOGICAL BUFFER CONDITIONS

3.1 Introduction

As mentioned previously, a potential G-quadruplex forming sequence requires the presence of a monovalent cation in order to stabilize quadruplex formation. This requirement is at odds with the conditions required for secure attachment of DNA to the mica surface. The monovalent cations present for quadruplex formation directly compete with the divalent cations necessary for secure attachment. Thus, in order to image G-quadruplexes through AFM, a proper balance between both cations must be maintained. Due to the limited size of the 18mer handle of the DNA constructs, $NiCl_2$ was chosen as the cation for its superior attachment properties. Optimization parameters for the preferred cation ratio relied upon retention of a maximum number of 525mer handles that contained a folded quadruplex and maximum number of 18mer folded quadruplexes.

3.2 Experiment

Previous studies have shown that potassium chloride (KCl) provides a superior stabilization effect compared to $NiCl_2$, thus a titration of KCl against $NiCl_2$ was performed [34]. $NiCl_2$ ion levels were maintained at 5 mM, while potassium chloride levels were titrated from 10 to 200 mM. DNA concentration was maintained at 1 nanogram per microliter, and the DNA solution was incubated for 5 minutes on freshly cleaved mica. Compressed air was utilized to dry the mica disks after rinsing with 200 μL of Milliq water.

3.3 18mer and 525mer Handle Quadruplex Molecules

Oligonucleotides required to construct the DNA duplex substrates were purchased from IDT DNA. Each 18mer construct relied upon one unique sequence and one 18mer complement in this study. The 18mer handle utilized as a control sequence contained the standard 18mer handle followed by a string of 20 T bases. This

sequence will not display any secondary structure under any the given ionic conditions. The sequences required to construct this handle molecule are listed below with the quadruplex region highlighted in red (Table 1).

In order to create a handle structure longer than 100 base pairs in length, a Polymerase Chain Reaction was used to produce the desired DNA product. Primers utilized in the creation of the 525 base pair handle are listed in Table 1. The host genome utilized in the PCR reaction was lambda phage DNA. Typical PCR reactions result in either a poly A tail overhang or blunt end comprised of duplex DNA molecules. These proved useless for the attachment of the 18mer-G-Quadruplex oligonucleotide mentioned in the previous paragraph. By combining an overhang (18mer compliment) PCR primer with Asymmetric PCR, a single stranded overhang was produced from the PCR products.

Unlike typical PCR reactions, Asymmetric PCR produces one single stranded DNA product. Through production of the top and bottom strands of the DNA handle in separate reactions, the products can be annealed in equal molar ratios to result in one duplexed DNA product. The Asymmetric PCR required 40 cycles of a standard PCR reaction, and the resulting products were purified through extraction from a 1% agarose DNA gel.

3.4 Results

Initial control experiments utilized both the control 18mer handle and the 525mer handle. Prior to analysis, each image is thresholded in order to establish areas of interest. The images in figure 5 show a clear demarcation of the boundaries of the DNA on the mica surface. Ten images were recorded for each experimental condition and analyzed to establish average molecule length. An average length of 168 ±13.5 nanometers was measured which corresponds to the theoretical length of approximately 180 nanometers for the longer handle. Convolution of the quadruplex by the AFM tip resulted in the average width of the 18mer being 23 nanometers. It should be noted that the height of the DNA molecules are lower than

what is expected from DNA geometry. This compaction of the height is due to the interaction of the DNA with the mica surface and divalent cation. Over 100 525mer and 18mer molecules were measured to establish the length of the molecule on the imaging surface.

A comparison of maximum quadruplex folding to total molecule retention are shown in figure 6. The optimum concentration for both the 18mer handle and 525mer handle were 10 mM $NiCl_2$ and 100 mM KCl. This combination resulted in maximum levels of quadruplex formation while maximizing the number of molecules present within a 1 μm^2 imaging area of each sample. The level of potassium ions required for maximal folding (100 mM) was much higher than established values from bulk and single molecule assays (10 mM) [35]. This is attributed to the need for a rinsing and drying step within procedure in order to maintain proper imaging conditions.

3.5 Discussion

The 525mer binding data indicates that the optimal folding to attachment ratio occurs at 100 mM KCl. While the maximum majority of molecules display a folded quadruplex at this experimental condition, it is seen that the total molecules that are attached to the surface has dropped significantly. This can be attributed to the competition between Nickel and Potassium ions in solution. This effect is greatly enhanced when utilizing only the 18mer handle, due to the decreased number of possible duplex base pairs to stabilize the molecules on the surface. Thus, it has been established that the G-quadruplex structure can be investigated through AFM when formed in a biologically relevant buffer condition.

3.6 Tables and Figures

Table 1

Sequence Name	Sequence
18mer-GQ3TTAG	TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTAG
18mer-GQ4	TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTAGGG
18mer-GQ6	TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
18mer-GQ8	TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
18mer-GQ12	TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
18mer Comp	GCCTCGCTGCCGTCGCCA
F-Primer 18mer-Over	GCCTCGCTGCCGTCGCCAGGCCTCTGTCGTTTCCTTTCTCTG
R-Primer Lambda	AACATGTCGGTTTTCCAGTTCCGG

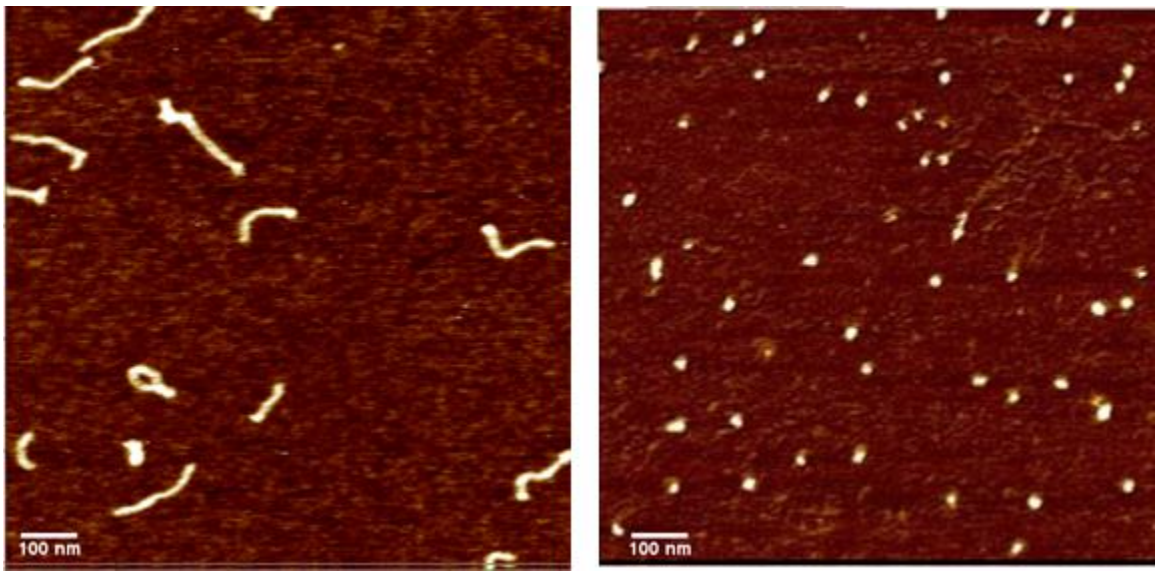


Figure 5 a.) Clear demarcation can be seen between each of the 525 base pair handle presented in this image. Several of the molecules are seen to be wrap around themselves. b.) The 18mer handle is seen a just a circular element as the quadruplex height results in the convolution loss of the duplex area.

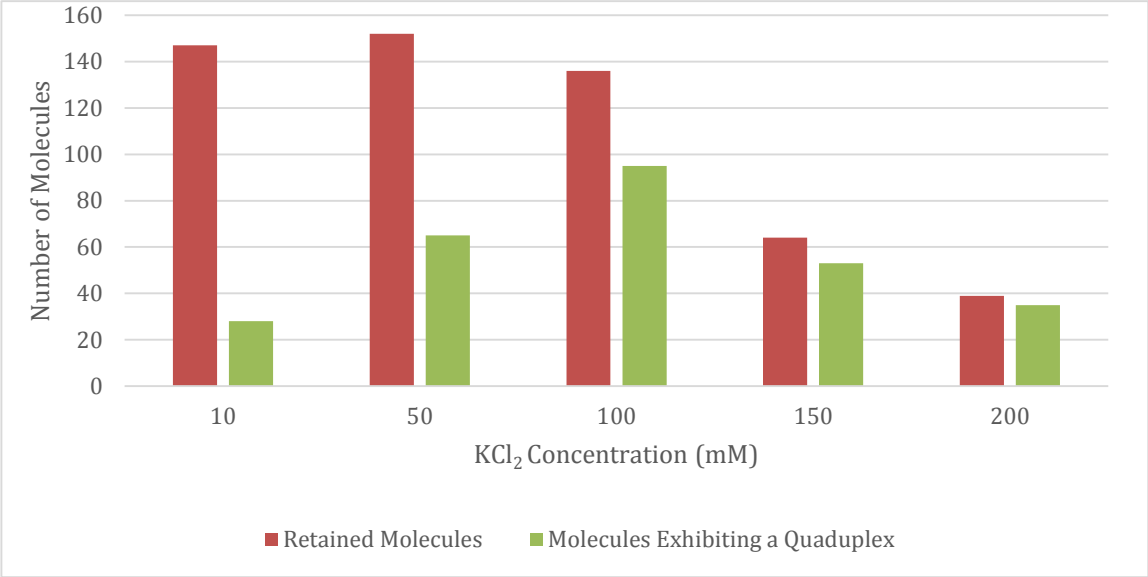


Figure 6 A concentration of 100 mM KCl_2 resulted in the highest level of retention of the molecules while seeing over 50 percent formation of G-quadruplexes. Increased KCl_2 concentration levels resulted in detachment of the handle from the surface. Decreased KCl_2 concentration levels resulted in poor G-quadruplex formation.

CHAPTER 4: TELOMERIC STRUCTURE AND ACCESSIBILITY EXHIBITED BY SMALL OLIGONUCLEOTIDE BINDING

4.1 Introduction

As mentioned previously, the 50 to 200 nucleotide single stranded G-rich terminal region of the telomere can fold into G-quadruplexes. Thus, this region has been suggested to be important for the protection of the terminal end of chromosomes. Nevertheless, this folding presents a problem for both types of telomere elongation that combat the end replication problem, as protein loading to a single stranded overhang repeat is required.

We chose to investigate a single stranded C-circle mimic of complementary DNA (C2) to the telomeric repeat and compared this DNA fragment to a binding of POT1, a telomere-associated protein that contains a similar DNA binding footprint. The data suggests that sequences with a multiple of 4 repeats, $(TTAGGG)_{4,\ 8,\ or\ 12}$ will fold into one, two, or three repeats respectively and thus limit accessibility to the C2 molecule.

4.2 Experiment

Initial experiments focused on establishing the profiles and lengths of the standard 18mer attached to either $(TTAGGG)_{4,\ 6,\ 8,\ or\ 12}$. A sequence of $(TTAGGG)_3TTAG$ was utilized as a non-folding control sequence (table 1). Solutions of DNA were incubated in binding buffer for 10 minutes prior to deposition on the mica imaging surface. The potential G-quadruplex forming sequences were chosen to demonstrate the difference between complete folding and folding of the G-quadruplex with a portion of single stranded DNA overhang. A schematic of this difference is seen in figure 7. Sequences with only one extra telomeric repeat were avoided due to the low melting temperature of the probe attachment and thus transient binding of a 6 nucleotide structure to strand of interest.

After characterization of each quadruplex was complete, the additional reporter probe, 12 base pair complement to the telomeric sequence was added to each reaction. The stable duplexing of this oligonucleotide to the quadruplex sequence suggests that an overhang is present in the specific quadruplex. Solutions of DNA were incubated in binding buffer with the addition of 5 fold excess molar concentration of C2 DNA for 10 minutes prior to deposition on the mica imaging surface.

4.3 Results

Control experiments show a slight indication of increased quadruplex molecule width with increasing number of repeats present. This increase in width is most likely attributed to the increasing number of G-quadruplexes that are present with each additional four repeats that are added. The signature folding pattern that is seen is a Gaussian-like peak with heights ranging from approximately 1 to 1.7 nanometers in height. As seen in figure 8 the length of the molecule increases with an increase in the number of possible G-quadruplexes that can fold. This is demonstrated by the length of G4 (24 nm) compared to G12 (48 nm). There is a lack of intervening single stranded profile between peaks as seen in several other studies [36]. This can be attributed to the increased concentration of potassium ion present in the binding buffer. This difference suggests that the maximal number of G-quadruplexes folding does occur.

After characterization of the varying telomeric lengths, a DNA probe was utilized to mimic the binding pattern of several telomeric associated proteins discussed earlier. The DNA probe consisted of two repeats of the compliment to the telomeric repeat sequence of TTAGGG. Upon binding to two sequential telomeric repeats, this sequence should elongate the single stranded repeats, resulting in a change in dimensional measurements detectable by AFM. The duplex formation of the DNA probe and telomeric repeats should only occur if the telomeric repeats are not incorporated into a G-Quadruplex. A diagram of this folding was seen in figure 9. As suggested by the initial control measurements, sequences containing multiple

four repeats resulted in no change in dimension of the G-quadruplex molecule. This was expected, as the initial imaging of these G-quadruplexes suggested that the maximal number of G-quadruplexes will form. This resulted in no free single stranded repeats being present for the DNA probe to form a duplex. These results were compared to a previous study done utilizing a shift in FRET values to signify DNA probe binding and duplex elongation [37]. In both studies, multiples of four repeats resulted in no apparent binding of the DNA probe.

The addition of the DNA probe to a six repeat strand of the telomeric repeat sequence shows a change in approximately one third of the population dimensional shape. In addition to the typical peak structure attributed to a folded G-quadruplex, a duplex DNA "handle" can be visualized (figure 9). This handle can be attributed as the extension of the initial 18mer by the 12 base pairs of the DNA probe. The length handle with the addition of the DNA probe is approximately 45±5.7 nanometers in length. If the 12 base pair duplex extension is not located proximal to the 18mer handle the AFM tip convolution limits the resolution of this structure, thus no extension is visualized. This accounts for the only one third of the population demonstrating the visualization of the handle. From this formation, we can conclude that the G-quadruplex formation does not selectively limit the accessibility of the DNA probe to single stranded regions of the telomeric repeat. This suggests that the binding site of several proteins similar to the DNA probe utilized in this study may be accessible in the dynamic environments of the single stranded telomeric overhang region.

4.4 Discussion

Several studies have shown that the formation of a G-quadruplex at the terminal regions of the single stranded telomeric overhang can greatly affect the access of this region to specific associated proteins [19.37]. Contrary to this study, results from a DMS foot printing experiment suggest that a G-quadruplex will favorably form at the final possible G-quadruplex forming region [19]. This formation at the terminal position would greatly affect the efficacy of many known proteins that

interact with the telomeric overhang. This observation held true for sequences containing a multiple of four repeats. In these sequences the final four repeats were capable of folding into the G-quadruplex structure and thus effectively capping the DNA sequence. Although, it is seen that only 1/3 of the molecules in a 6 telomeric repeat sequence form a handle structure. This suggests that the G-quadruplex is located on the four terminal repeats in only 1/3 of the total G-quadruplexes observed. This result is independent of DNA probe concentration. This suggests that multiple folding populations are present in the 6 repeat sample, this contradicts the findings of DMS foot printing experiment. A schematic of the possible folding populations are seen below (figure 10).

4.5 Figures

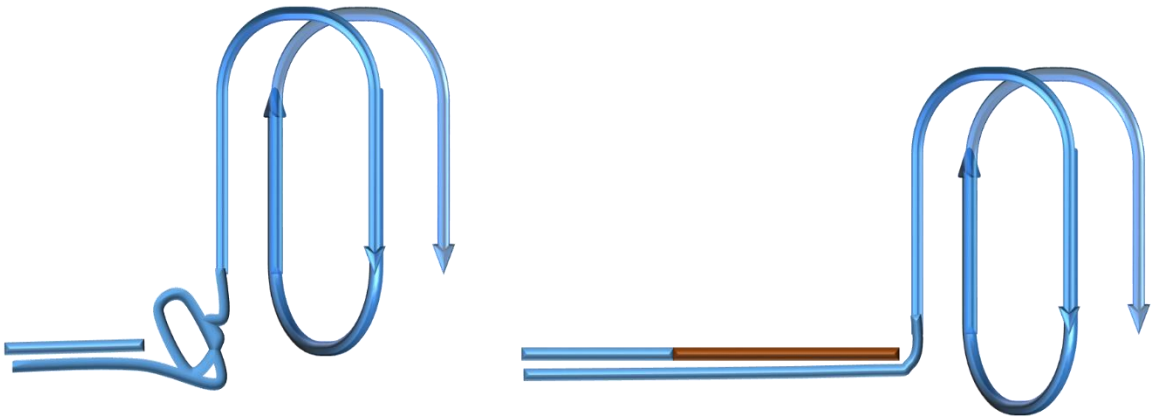


Figure 7 a). A 6 telomeric repeat sequence may form the G-quadruplex at the terminal end (3'). This leaves 12 bases in the single stranded form. b). Addition of a 12 base single stranded complement probe (red) can result in the extension of the original DNA handle.

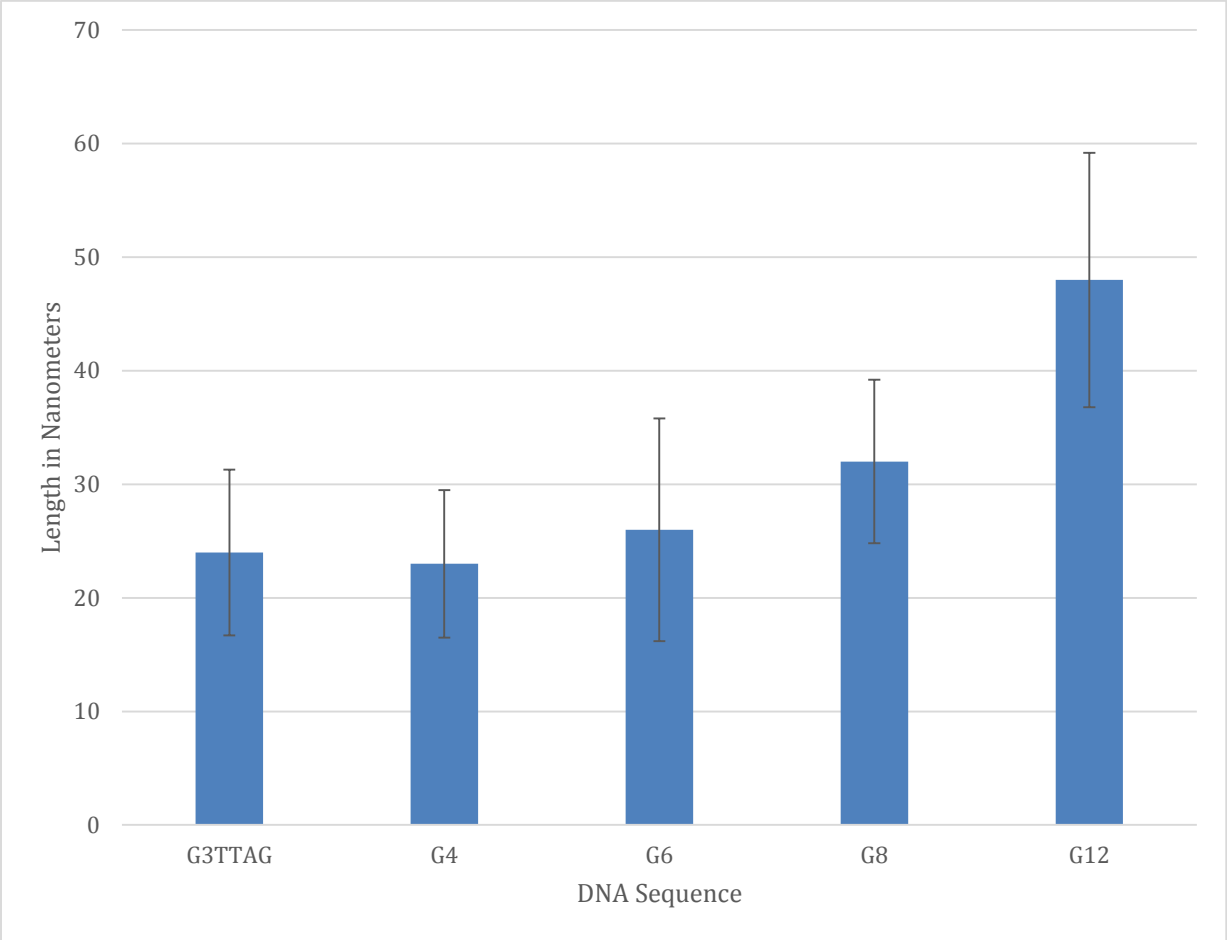


Figure 8 Increasing numbers of telomeric repeats result in larger width of each molecule. There are two transitions seen above. G4 and G6 are only capable of folding one G-quadruplex, while G8 and G12 are capable of folding 2 and 3 G-quadruplexes respectively.

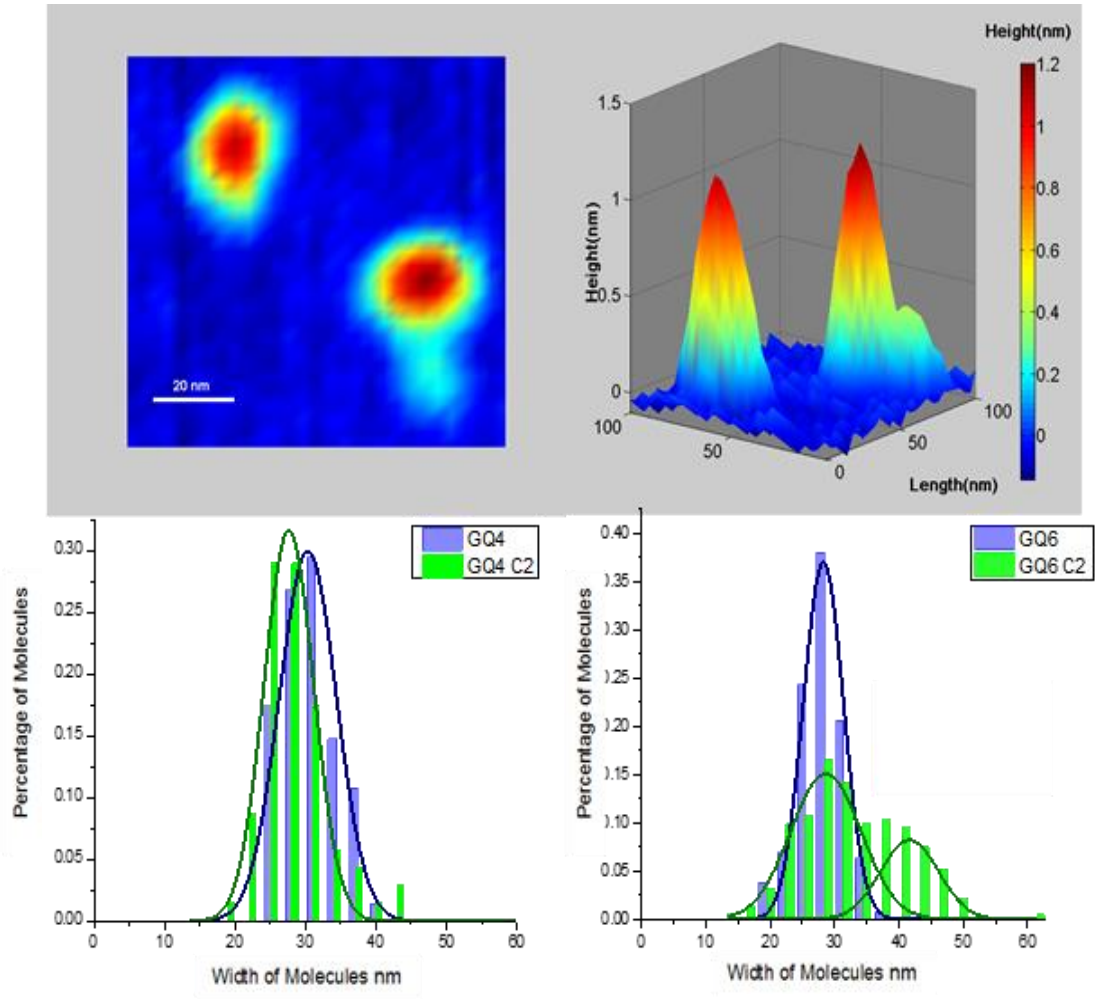


Figure 9 a.) A topographical map of two G6 molecules are seen in the top image. Each represents a single population. The extension seen in the lower molecule is representative of the addition of the C2 probe near the 18mer handle. b.) No population shift in width is seen in the G4 population upon the addition of the C2 probe. Approximately one third of the population of the G6 sequence is seen to shift to width 42 nanometers upon the addition of the C2 probe.

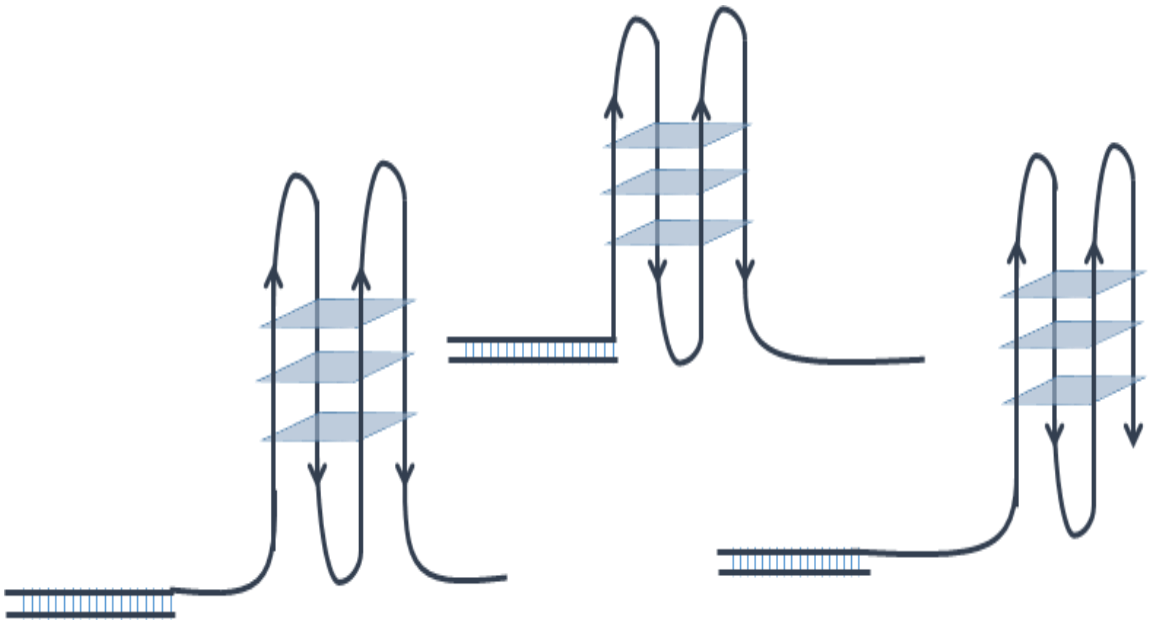


Figure 10 The three possible folding configurations of G6 are seen above. The single stranded DNA regions differ in respect to the location of the G-quadruplex.

CHAPTER 5: IDENTIFICATION OF G-QUADRUPLEX FOLDING MOTIF THROUGH ATOMIC FORCE MICROSCOPY

5.1 Introduction

G-quadruplex repeats are not strictly limited to the single stranded region of the telomere. Recent studies have not only shown the prevalence of G-quadruplexes within the duplex areas of the genome, but have also shown their possible significant effect on genetic expression [38]. Computational studies suggest that G-quadruplexes are most likely found in genetic control regions of oncogenes [39]. This represents an interesting target for cancer therapy development, as oncogenes become tumorigenic in nature due to the increased production or gain of function of the coded protein. This, coupled with the observation that G-quadruplexes limit genetic expression and thus reduce protein expression level, suggests stabilization of G-quadruplexes directly upstream of a protein coding region could remedy the activation of said oncogene.

As mentioned previously, intramolecular G-quadruplexes typically fold into either one of two folding motifs (parallel or antiparallel) based upon tetrad run direction. The direction of folding has been shown to provide specificity to associated protein binding and affects the binding efficiency of G-quadruplex interacting small molecules. Previous work has suggested that intervening sequence length between G-tetrad runs relates to the folding motif [16]. Specifically, longer intervening sequences result in a high prevalence of anti-parallel folding patterns.

Current methodology relies upon circular dichroism (CD) experimental techniques to establish predominant folding motifs of a given sequence. This method has several problems, the first being the reliance on a large sample of high concentration DNA. At sufficiently high DNA concentrations there is a high prevalence of intermolecular folded G-quadruplexes, this sub-population may altering measurement accuracy. Additionally, circular dichroism fails to produce distinguishing results in mixed populations of parallel and anti-parallel G-

quadruplexes. Thus, this measurement method only provides three possible readings: completely parallel, completely anti-parallel, and mixed population. It is easy to recognize the shortcomings of such limited data acquisition. Through the use of AFM, a heterogeneous population of G-quadruplex molecules can be investigated in a single molecule manner, resulting in a much richer picture of the nature of a specific DNA sequence.

5.2 Experiment

To determine if the specific folding motifs could be distinguished through AFM, two known sequences were utilized as references. First, the telomeric quadruplex, which has been shown through CD and crystallography studies to form into a mixture of parallel and anti-parallel configurations, was utilized. The second sequence source was MYC, a human transcription factor that contains a G-quadruplex in the upstream regulatory region. This structure has been confirmed to be parallel in folding nature. Finally G3TTAG, the non-folding control sequence utilized throughout this study, provided the negative control.

Asymmetric Polymerase Chain Reaction was once again utilized to create a duplex DNA handle of approximately 525 base pairs in length. In order to avoid inducing intermolecular G-quadruplex formation, DNA concentrations of both the handle and the potential G-quadruplex forming region were kept below 140 nM for initial G-quadruplex attachment. After initial verification of this method through AFM, the sequences investigated were expanded to include variations upon both the MYC forming parallel sequences and the telomeric anti-parallel sequences. A list of these sequences is provided in table 2.

5.3 Results

A significant height difference can be seen between the folding configurations. An increase in height should be apparent in an anti-parallel folding motif, due to the intervening sequence traversing the top and bottom of the structure. A parallel structure is represented by an increase in width of the molecule due to the

intervening sequences traversing the faces of the molecules. This increase in width was expected to be difficult to visualize due to the convolution of the 8 nanometer AFM tip mentioned in the previous section. Utilizing estimates taken from crystallography studies, an initial height estimate of the 1.2 nm was to be used to indicate parallel folding [40]. Anti-parallel folding was expected to vary in size based upon total length and composition of the intervening sequence, with initial estimates around 1.6 nm in height. Due to the compaction of the molecule as a result of its attraction to the mica surface, as seen in duplex DNA, these values proved to be several hundred picometers larger than was measured.

Population measurements of telomeric DNA overhang heights suggested three populations existed: 1.1, 1.5, and 1.9 nm in height. An example of each folding motif is shown in figure 11. The third population (1.9 nm) can be attributed to intermolecular G-quadruplexes due to the identification of two 525 base pair handles protruding from the G-quadruplex or visualization of the G-quadruplex resting upon the duplex DNA handle (figure 11). Thus, the 1.1 nm population is assumed to represent the parallel population and the 1.5 nm population would indicate anti-parallel folding.

To verify these observations, the MYC quadruplex sequence was utilized to show a strictly parallel population as demonstrated by circular dichroism measurements. Two populations were seen in this sequence: 1.1 and approximately 2 nm in height. As previously stated, this increased height population was attributed to intermolecular folding, and was seen at a much lower frequency percentage (less than 5 percent) as compared to the telomeric sequence and thus was ignored. These results reflect the observation of solely parallel folding occurring in the MYC seen in circular dichroism (figure 12). This led to the identification of the 1.1 nm population as the parallel population and the ~1.5 nm population as anti-parallel-in-nature population.

With the two previous height populations classified in both the telomeric repeat and the MYC G-quadruplex sequences, a larger population representing derivatives of these sequences were tested to determine the distribution of parallel verses anti-parallel. These sequences are seen in table 2. Specifically, sequences 1-3-3, 1-4-4, and 1-5-5 show a similar folding pattern to that of the MYC G-quadruplex sequence (figure 13). All folded quadruplexes are located in the 1.1 nm population suggesting these sequences all fold in a parallel folding motif. Upon a transition to the sequence of 2-3-3, an initial small population is seen to appear in the 1.5 nm population. This signifies the development of an anti-parallel folding motif within the 2-3-3 folding possibilities (figure 13). The trend of the increasing anti-parallel population with an increasing intervening length continues with the sequences 3-3-3, 4-3-3, and 5-3-3(figure 13). The 3-3-3 sequence approaches an equal distribution between the two populations. The addition of one more base to the intervening sequence (4-3-3) results in a dramatic shift to a nearly 100 percent anti-parallel population. This phenomenon is also seen with the sequence 5-3-3.

The effects of larger bases were also investigated by varying the A-base content in the intervening sequences of G-quadruplexes containing an intervening spacing of 3 between each tetrad run. As mentioned above, the 3-3-3 sequence contained a nearly equal distribution between the two population types and a large shift to a predominantly antiparallel folding motif as seen in the telomeric sequence. This trend is seen to continue with the sequence $(TAA)_3$ containing only the antiparallel population as well (figure 13).

5.4 Discussion

Research of G-quadruplex folding sequences has proven to be a difficult task due to the limits of the two major investigational methods currently employed today (circular dichroism and crystallography). As with any bulk measurement modality these techniques prove unsatisfactory in their lack of ability to investigate single molecule events and relying on ensemble analyses. This problem is exacerbated in G-quadruplex studies due to the nature of the molecules to form non-biologically

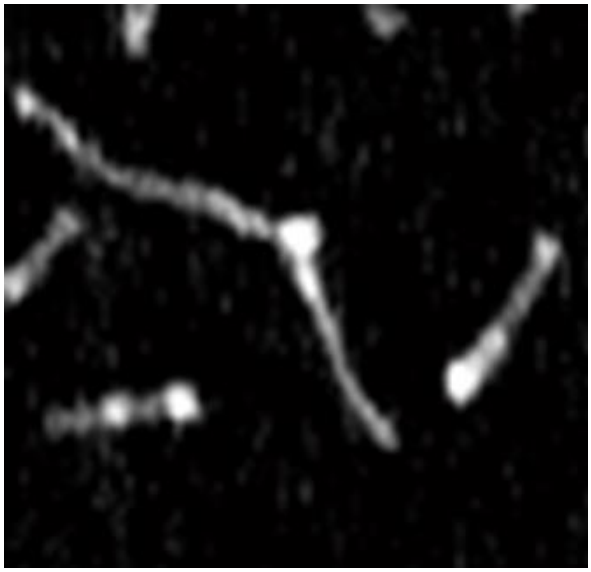
relevant intermolecular structures at the increased concentrations required for bulk measurements. AFM has been show to allow for the identification and classification of G-quadruplex structures in a more natural setting.

As mentioned previously, computational studies have suggested that sequences containing low intervening lengths are preferentially found within the genetic regulatory regions of oncogenes and other interesting DNA sequences. Our data suggests that a clear classification can be made of sequences either falling with a predominantly parallel or antiparallel sequence. This transition occurs between the sequences of 2-3-2 to 3-3-3 to 4-3-3, thus identifying a shift region from almost completely parallel to almost complexly antiparallel over the transition of three sequences. The addition of A base elements to the 3-3-3 initial sequence also results in the same trend as seen in increasing intervening length. The addition of new bases or replacement of bases with bulkier bases (i.e. A's) results in a dramatic shift to the formation of the antiparallel structure. These longer intervening sequences with bulky A bases prove to be a realistic model for the telomeric regions. This suggests that telomeric regions and similar quadruplexes sequences predominantly fold into an antiparallel structure, which is contrary to the sequences commonly found within genetic coding regions.

5.5 Tables and Figures

Table 2

Sequence Name	Sequence
18mer-CYMC	TGGCGACGGCAGCGAGGCGGGTGGGTAGGGTGGG
18mer-(TTA)3	TGGCGACGGCAGCGAGGCGGGTTAGGGTTAGGGTTAGGG
18mer-1-3-3	TGGCGACGGCAGCGAGGCGGGTGGGTTTGGGTTTGGG
18mer-1-4-4	TGGCGACGGCAGCGAGGCGGGTGGGTTTTGGGTTTTGGG
18mer-2-3-3	TGGCGACGGCAGCGAGGCGGGTTGGGTTTGGGTTTGGG
18mer-3-3-3	TGGCGACGGCAGCGAGGCGGGTTTGGGTTTGGGTTTGGG
18mer-4-3-3	TGGCGACGGCAGCGAGGCGGGTTTTGGGTTTGGGTTTGGG
18mer-5-3-3	TGGCGACGGCAGCGAGGCGGGTTTTTGGGTTTGGGTTTGGG
18mer-(TAA)3	TGGCGACGGCAGCGAGGCGGGTAAGGGTAAGGGTAAGGG



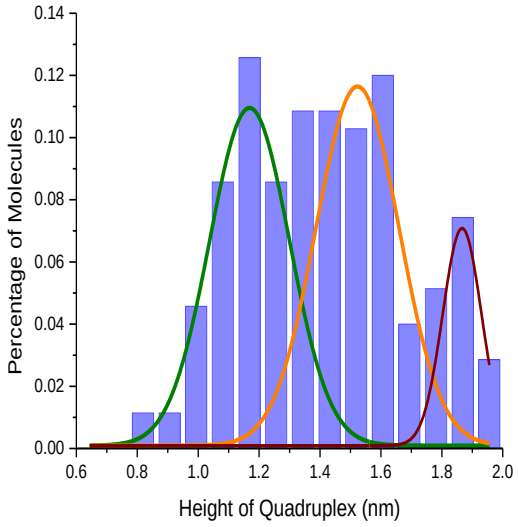


Figure 11 a.) Two handles can be seen on the center G-quadruplex. This suggests that an intermolecular G-Quadruplex is forming. b.) Three populations are seen suggesting that the 1.7 nanometer population is attributed to the intermolecular G-quadruplexes. This leaves the population around 1.1 and 1.5 nanometers to represent parallel and anti-parallel populations.

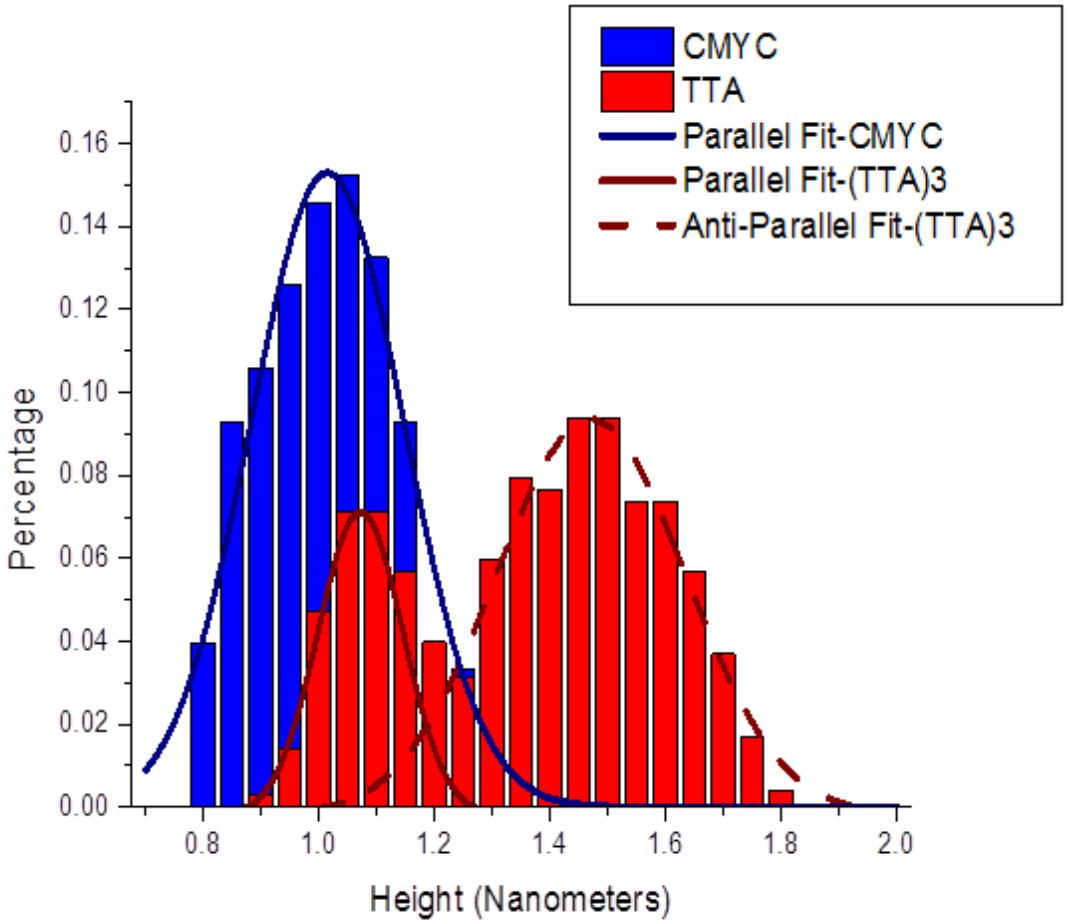


Figure 12 The height distributions of CMYC and $(TTA)_3$ are shown. CD data suggests that CMYC is strictly parallel in population, thus the 1.1 nanometer peak represents this. The $(TTA)_3$ distribution contains two populations that correspond to the 1.1 nanometer parallel and 1.5 nanometer anti-parallel folding motifs.

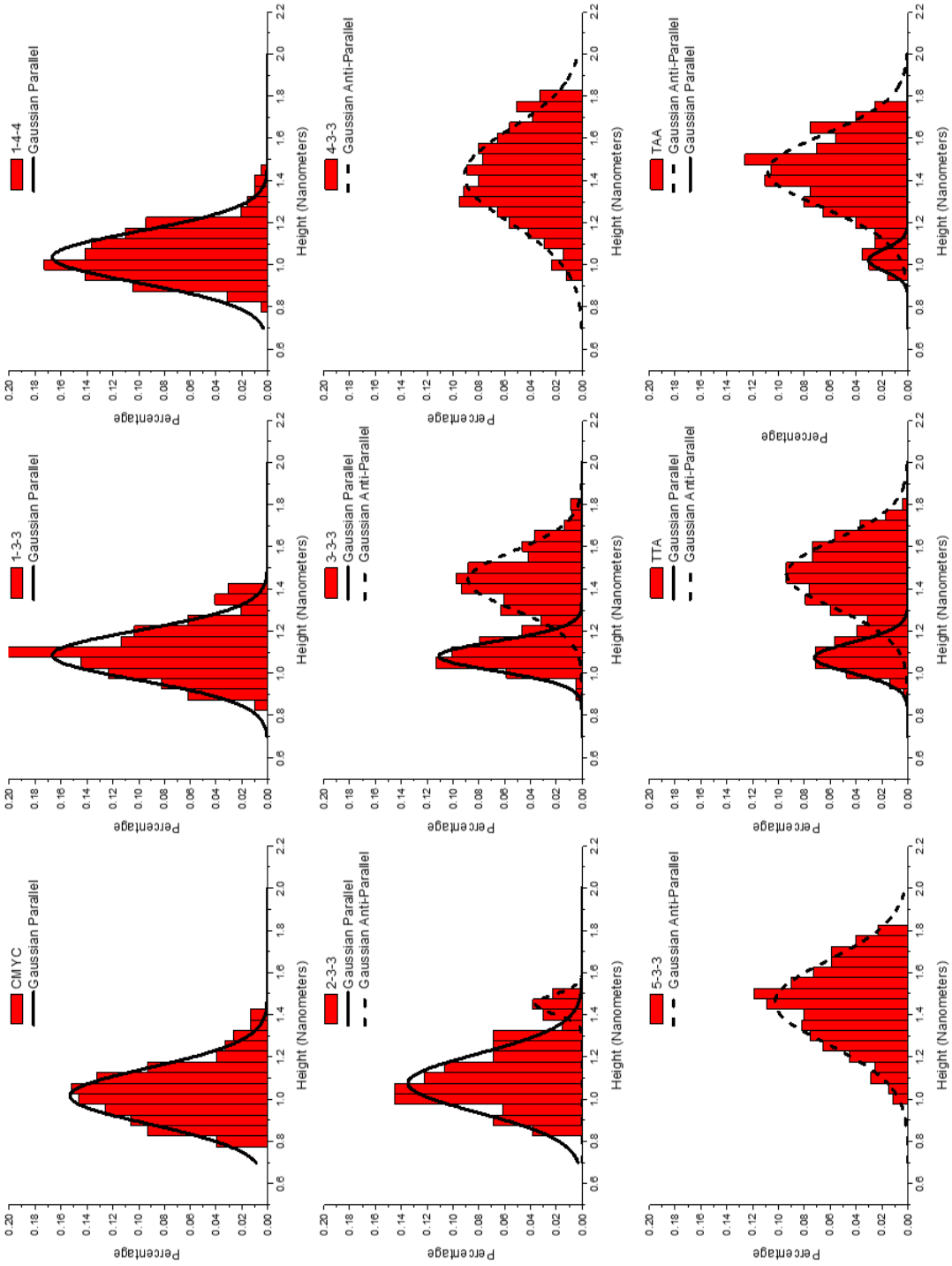


Figure 13 The height profiles of all the tested G-quadruplex forming sequences are shown.

CHAPTER 6: ATOMIC FORCE MICROSCOPY STUDIES VS. TRADITIONAL QUADRUPLEX MEASUREMENTS

G-quadruplexes are involved in a wide range of cellular process. As mentioned, current methods in investigating these structures lack several key properties. The AFM method expands the capacity of investigating these structures in a biologically relevant buffer system and concentration. Most importantly this method allows for a single molecule dimensional analysis of these structures. Through the use of the dimensional analysis properties such as G-quadruplex folding location and distribution of folding motifs can be completed. These properties have not been sufficiently investigated due to the inability and difficulty in examining G-quadruplex structures.

One can also envision utilizing this method to rapidly test pharmaceuticals designed to stabilize G-quadruplexes. Utilizing a native pull-down system or PCR of a target sequence, one may use the methods developed in order to investigate a G-quadruplex in its native DNA sequence. For example, several G-quadruplex forming regions in the human genome have been shown to contain several runs of neighboring G-quadruplex folding sequences. Through the method developed these sequences can be identified and more importantly the location of the G-quadruplex can be pinpointed.

REFERENCES

1. Klein CA. (2008) Cancer. The metastasis cascade. *Science 321*(5897):1785-1787
2. Hanahan D, Weinberg RA. (2000) The Hallmarks of Cancer. *Cell* 100(1):57-70
3. Hanahan D, Weinberg RA. (2000) Hallmarks of Cancer: The Next Generation Cell 144(5):646-674
4. Hayflick L. (1965) The limited in vitro lifetime of human diploid cell strains. *Expiermental Cell Research* 37(3):614-636
5. Kelland L. (2007) Targeting the Limitless Replicative Potential of Cancer: The Telomerase/Telomere Pathway *Clinical Cancer Research* 2007(12):4960-4963
6. Perrem K, Colfin LM, Neumann AA, Yeager TR, & Reddel RR. (2001) Coexistence of alternative lenthing of telomeres and telomerase in hTERT-transfected GM847 cells *Molecular and Cellular Biology 21*(12):3862-3875.
7. Rodriguez-Brenes IA, Peskin CS (2010) Quantitative theory of telomere length regulation and cellular senescence. *Proc Natl Acad Sci U S A* 41 107(12):5387-92
8. Smogorzewska A & de Lange T (2004) Regulation of telomerase by telomeric proteins. *Annual review of biochemistry* 73(1):177-208.
9. Shay JW & Wright WE (2010) Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* 26 (5):867-874
10. Burge S, Parkinson GN, & Hazel P (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research* 34(19):5402-5415
11. Williamson JR. (1994) G-quartet structures in telomeric DNA. *Annual Review of Biophysics and Biomolecular Structures* 23:703-730.
12. Ambrus A, Chen D, Dai JX, et al. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Research* 34(9):2723-2735
13. Campbell, Nancy H.; Neidle, Stephen (2008) Chapter 4. G-Quadruplexes and Metal Ions. *Interplay between Metal Ions and Nucleic Acids. Metal Ions in Life Sciences* 10:119-134
14. Simonsson, T. (2001) G-Quadruplex DNA Structures Variations on a Theme. *Biological Chemistry* 382(4):621-628
15. Lane AN, et al. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* 36(17): 5482-5515
16. Hazel P, Huppert J, & Balsubramanian S.(1994) Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society* 126(50):16405-16415
17. Hiyama E & Hiyama K. (2002) Clinical utility of telomerase in cancer. *Oncogene 21(4):643-649*
18. Zaug AJ, Pdell ER, & Cech TR (2005) Human POT1 disrupts telomeric G---quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 102(31): 10864---10869 .

19. Wang Q*, et al.* (2011) G-quadruplex formation at the 3′ end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Research* 10(1093):45-50
20. Hwange H*, et al.* (2012) POT1-TPP1 regulates telomeric overhang structural dynamics. Structure 20(11):1872-80.
21. Piazza A, et al. (2004) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae *Nucleic Acids Research* 38(13): 4337-4348
22. Chow, A. Y. (2010) Cell Cycle Control by Oncogenes and Tumor Suppressors: Driving the Transformation of Normal Cells into Cancerous Cells. *Nature Education* 3(9):7
23. Carey M & Smale S. (2001) Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques. *CSHL Press* 1. Klein CA. (2008) Cancer. The metastasis cascade. *Science 321*(5897):1785-1787
24. Yi Ni Lam E. (2013) G-quadruplex structures are stable and detectable in human genomic DNA.. *Nature Communications* 4(1796)
25. Smargiasso N, et al (2009) Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes. *BMC Genomics* 10(32)
26. PerremK, Colfin LM, Neumann AA, Yeager TR, & Reddel RR. (2001) Limits of Resolution in Atomic Force Microscopy Images of Graphite *Europhysics Letters 15*(1):49-54
27. Sun D & Hurley LH. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methonds in Molecular Biology* 2010(608):65-79
28. Paramasivan S, Rujan I, Bolton PH. (2010) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* 43(4):324-31
29. Allain C, Monchaud D, Teulade-Fichou MP (2006) FRET templated by G-quadruplex DNA: a specific ternary interaction using an original pair of donor/acceptor partners. *J Am Chem Soc* 128(36):11890-3
30. Okumus B, Ha T. (2010) Real-time observation of G-quadruplex dynamics using single-molecule FRET microscopy. *Methods in Molecular Biology* 608:81-96
31. Yatsunyk LA, et al. (2000) Guided Assembly of Tetramolecular G-Quadruplexes ACS Nano 7(7):5701-10
32. Hansma PK, et al. (1998) Tapping mode atomic force microscopy in liquids. *Applied Physics Letters* 64(13): 1738 - 1740
33. Cleveland JP, et al. (1965) Tapping mode atomic force microscopy in liquids. *Applied Physics Letters* 64(13): 1738 - 1740
34. Schultze P, Hud NV, Smith FW, Feigon J (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). *Nucliec Acids Research* 27(15):3018-28
35. Jaumot J, Gargallo R. (2012) Experimental methods for studying the interactions between G-quadruplex structures and ligands. *Curr Pharm Des 18(14):1900-16*

36. Wang H, Nora GJ, Ghodke H, Opresko PL. (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J Biol Chem* 286(9):7479-89
37. Hwang H, et al(2013) Telomer Overhang Accessibility to Telomerase and Alt Associatied Porteins Depend on Telomeric Repeat Number In preperation
38. Huppert J & Balsubramamian S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research* 35(2): 406–413
39 Brooks TA, Kendrick S, Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS* 277(17):3459-69
40 Wei D, Parkinson GN, Reszka AP, Neidle S. (2012) Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Research* 40(10):4691-700