A SINGLE MOLECULE STUDY OF G-QUADRUPLEX AND SHORT DUPLEX DNA STRUCTURES

A thesis submitted
to Kent State University in
fulfillment of the requirements for
the degree of Master of Science

by
William A. Roy, Jr.

August 2016
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ VI

ACKNOWLEDGEMENTS ................................................................................................ VIII

CHAPTER ONE  INTRODUCTION ............................................................................. 1
  1.1 SUMMARY ......................................................................................................... 3
  1.2 WHAT IS G-QUADRUPLEX ........................................................................... 4
    1.2.1 Conformations ......................................................................................... 6
    1.2.2 GQ Stability ............................................................................................. 7
  1.3 WHY STUDY GQ? ............................................................................................ 8

CHAPTER TWO  SMFRET AND TIRF MICROSCOPY ........................................... 9
  2.1 TOTAL INTERNAL REFLECTION FLUORESCENCE MICROSCOPY ............... 9
    2.1.1 Förster Resonance Energy Transfer ....................................................... 10
    2.1.2 Combining the Techniques .................................................................... 11
    2.1.3 FRET Pair Selection ............................................................................... 11
  2.2 EXPERIMENTAL SETUP ............................................................................. 13
    2.2.1 Excitation & Emission Optics ............................................................... 13
    2.2.2 Instrumentation ...................................................................................... 15
  2.3 DATA ACQUISITION AND ANALYSIS ....................................................... 16
    2.3.1 Single Molecule Imaging ....................................................................... 16
    2.3.2 Generating a Mapping Function ............................................................ 17
    2.3.3 Single Molecule Traces ......................................................................... 20
  2.4 THE SAMPLE CHAMBER ............................................................................ 22
    2.4.1 Coating the Surface with PEG ............................................................... 22
    2.4.2 Assembly ................................................................................................ 22
  2.5 SAMPLE PREPARATION ............................................................................ 24
    2.5.1 Design and Annealing .......................................................................... 24
    2.5.2 Creating Biotintylated Construct Binding Sites .................................... 24
CHAPTER THREE  THE MACROCYCLE ........................................... 27

3.1  SUMMARY .................................................................................. 27

3.2  CONSTRUCT DESIGN ................................................................. 27
  3.2.1  The “Macrocycle” ..................................................................... 27
  3.2.2  Development of the Macroduplex ............................................ 29
  3.2.3  Tunable External Force ........................................................... 29

3.3  CONCEPT AND WHY IS A FORCE EXPECTED? ......................... 30

3.4  LIGATION ................................................................................... 31

3.5  SAMPLE PROCEDURE ................................................................. 32
  3.5.1  Immobilization on the surface ................................................ 32
  3.5.2  Bridge Binding Considerations .............................................. 33
    3.5.2.1  Single bridging on opposite end ....................................... 34
    3.5.2.2  “Double bridging” ............................................................ 35
  3.5.3  Macrocycle Formation ............................................................ 37
    3.5.3.1  GQ Forming Bridge .......................................................... 37
    3.5.3.2  Additonal Peak for PolyT Bridge ....................................... 38
    3.5.3.3  A Logical Projection .......................................................... 41

CHAPTER FOUR  DISCERNING THE RELEVANT SIGNAL ..................... 42

4.1  THE PARTIAL DUPLEX CONTROL ............................................... 42
  4.1.1  Design .................................................................................. 42
  4.1.2  Equivalent Identification ......................................................... 43
  4.1.3  Further Validation ................................................................. 45

4.2  THE HUNT FOR A MACROCYCLIC UNFOLDED GQ ................... 45

4.3  ELIMINATION BY QUENCHING ............................................... 47

4.4  UTILIZING THE QUENCHER WITH A MACROCYCLE CONSTRUCT .... 52

CHAPTER FIVE  DISCUSSION ............................................................... 55

5.1  FUTURE WORK & IMPROVEMENTS .......................................... 55

APPENDIX I: SEQUENCES ................................................................ 56
APPENDIX II: MACROCYCLE CONFIGURATIONS ........................................... 59
REFERENCES ........................................................................................................ 65
LIST OF FIGURES

Figure 1.1 G-Quadruplex Constituents ................................................................. 5
Figure 1.2 G-Quadruplex Conformations ................................................................. 7
Figure 2.1 The FRET pair Cy3 and Cy5................................................................. 12
Figure 2.2 The fluorescence excitation and emission spectra for Cy3 and Cy5 ............ 12
Figure 2.3 A plot of $E_{\text{FRET}}$ vs. Interfluorophore Distance .............................. 13
Figure 2.4 A diagram of the prism-type TIRF smFRET microscopy setup ............... 15
Figure 2.5 A single molecule image of a bead sample ........................................... 19
Figure 2.6 A single molecule FRET trace .............................................................. 21
Figure 2.7 A picture of the sample chamber ........................................................... 23
Figure 2.8 A schematic of the DNA bound to the sample chamber surface .................. 26
Figure 3.1 A schematic of the macroduplex with a bound GQ forming bridge strand by
Sticky End 1 ........................................................................................................... 29
Figure 3.2 Macrocycle DNA constructs with a GQ forming bridge ............................ 33
Figure 3.3 Bridge binding at the Sticky End 2 of the macroduplex ............................ 34
Figure 3.4 A schematic of a macroduplex with a bridge bound to each side (double
bridged) ............................................................................................................... 35
Figure 3.5 A smFRET histogram for the macrocycle with a 3Ly1Lp Bridge in 1 M KCl
and 10 mM MgCl$_2$ ............................................................................................... 38
Figure 3.6 A smFRET histogram for the macrocycle with a PolyT bridge in 1 M KCl and
10 mM MgCl$_2$ ....................................................................................................... 40
Figure 4.1 A comparison of the partial duplex construct with that of the macrocycle construct .............................................................. 44

Figure 4.2 A smFRET histogram for the macrocycle with a hGQ bridge in 150 mM LiCl and 5 mM MgCl₂ ................................................................. 47

Figure 4.3 A schematic of the quencher assay for the partial duplex DNA construct … 49

Figure 4.4 A plot of molecule count vs. time using red excitation upon addition of a quencher .............................................................. 50

Figure 4.5 Quencher elimination ........................................................................ 51

Figure 4.6 Imaging the quenching process for the partial duplex control .............. 51

Figure 4.7 A Comparison of the partial duplex and macrocycle constructs after introduction of the quencher .............................................................. 53

Figure 4.8 A plot of molecule count vs. time using red excitation upon addition of a quencher .............................................................. 54
ACKNOWLEDGEMENTS

I’d like to show my great appreciation for those who have supported my education and growth. It is to those whom I attribute my success and development in the field of physics. Specifically, Dr. Mann and Dr. Balci, you two were my strongest supporters and you both have given me the wonderful opportunity to grow and learn under your guidance in your labs. When I worked in each of your labs, I didn’t feel like just another student, I felt like a family member. Of course, like any family, there are the occasional arguments, but there were also a lot of good discussions of both personal and work-related. I think I am going to miss the occasional ramblings like where the good cheese and dates can be found or the cultural customs of another country the most. I truly have learned so much from both of you and I really can’t express how grateful I am.

I’d also like to recognize my fellow students who have helped me during my time in lab and in the courses, specifically Joseph Yarzebinski, Mona Mirheydari, Nabin Thapa, Jagat Budhathoki, and Parastoo Maleki. Without them, I would not have been able to learn the techniques and protocols expected of an advanced scientist in an experimental lab. It was a pleasure learning and working with you all, and I’ll always appreciate the great discussions we’ve had in our group meetings.

William A. Roy, Jr.

August 2016, Kent, Ohio
CHAPTER ONE
INTRODUCTION

Given that certain conditions are met, a single stranded DNA/RNA (ssDNA/RNA) structure called G-quadruplex (GQ) can form in regions throughout the genome, including at the telomeres and internal regions of the chromosomes [1-6]. These structures serve various functions depending on the region in which they form which include protecting the chromosome ends, interfering with telomere elongation in cancer cells, and regulating transcription and translation level gene expression [2, 7-10].

Due to their high stability, various cellular mechanisms, such as GQ destabilizing proteins, are employed to unfold these structures during DNA replication or repair [11-13]. Yet, their distinct layered structure has made GQs an attractive drug target in cancer treatment as GQ stabilizing molecules could inhibit telomerase dependent telomere elongation, a mechanism occurring in the majority of cancer cells to avoid senescence and apoptosis [14, 15]. However, proteins or small molecules interact with GQ that is under the influence of various cellular tension mechanisms, including the tension applied by other nearby molecules or the tension due to DNA structure within the chromatin context [16-18]. Therefore, it is important to characterize the stability of various GQs and their response to interacting molecules when subjected to a tensile force. We employed a novel DNA-based nano tension generator that utilizes the elastic properties of circularized short double-stranded DNA (dsDNA) oligonucleotides to apply tension on the GQ (for schematic of structure, see Appendix II, Configuration K) [19]. Since this is a
completely new approach, the majority of this thesis was dedicated to proof-of-principle studies that demonstrated the feasibility and functionality of the method.
1.1 Summary

In this thesis I will discuss the development of this novel method in which the dynamics of many GQ molecules that are subjected to an adjustable external force can be monitored simultaneously. The investigative work in this thesis mainly focused on studies that were proof-of-concept driven to affirm that this new method could be used to study a large number of constructs and their interactions with proteins or small molecules. Since this project is just now blossoming into an incredibly interesting and useful method, it is my hope that this thesis can help serve as a guide for the next student to take the reins on this wonderful opportunity.

Here, in this chapter, I will present key concepts and supporting background necessary to fully understand the advanced method we have developed. I hope to make it both interesting and useful to the reader. I will first discuss what a GQ is, how it forms, the different conformations it can take, the factors that contribute to the stability of GQ, and perhaps most importantly, I want to demonstrate the importance of studying these structures. My desire is to make the reader fully aware of how this work applies to the problems of the real world.

After a brief introduction, I will go into the details of the experimental setup including the physical phenomena we are taking advantage of in our measurements, the ability to measure single molecule interactions, the optics of the instrument, and the analysis of the obtained data. I will then introduce the concept of the new method we have tested, give motivation for creating this new method, discuss the logic behind the design and briefly address the problems we came across in the development and the
solutions to those problems. Then I will give a concise summary and leave the reader with a few of the vast possibilities this project has to offer in future work. The sequences of all constructs used in the study are provided in Appendix I. Furthermore, a schematic of all constructs and their possible configurations are provided in Appendix II. Even though I tried to keep the terminology of referring to constructs and their configurations uniform throughout the thesis, the large number of possibilities might make it a bit difficult to keep track. In such cases, please refer to Appendix II for a visual aid.

1.2 What is G-Quadruplex

The deoxyribonucleic acids (DNA) consist of 4 nucleobases: Adenine (A), Cytosine (C), Thymine (T), and Guanine (G) [20]. Under the physiological conditions, a G-rich single-stranded segment of DNA (or RNA) with the appropriate sequence could form a GQ structure which is significantly different from the well-known double helical DNA [21]. Rather than the typical Watson-Crick hydrogen bond formations between complementary bases, Hoogsteen hydrogen bonding between Gs dominates in GQ [22]. Four Gs orient themselves and bind in the structure that we call a “G-tetrad” as shown in Figure 1.1. When these G-tetrads stack, they form a GQ. A GQ typically has 2-5 stacked tetrads. Another structural feature of GQ is the loop sequences which are not part of the G-tetrad but connect the different parts of the structure. The loops can be as short as one nucleotid (nt) and as long as 10 nt or more [23-26]. There are several hundred thousand potential quadruplex-forming sequences throughout the human genome [27, 28]. The most studied GQ is the human telomeric GQ (hGQ), named for prevalently occurring in the telomeres, and has a sequence of GGG TTA repeats [29-31].
Figure 1.1: G-Quadruplex Constituents. (a) The molecular structure of guanine. (b) Four guanines form a “G-tetrad” via Hoogsteen hydrogen bonding. (c) Two or more G-tetrads stack to form a G-quadruplex (in this case three). Three loops connect the G-tetrads. Since the GQ is formed by a single DNA strand, it is called intra-molecular GQ. (d) Crystal structure of a GQ viewed from the top [32].
1.2.1 Conformations

GQ structures are described with different conformations depending on the polarities of the loops [33]. For example, tracing from the 5’ end, a parallel conformation would be one where the loops have the same 5’ to 3’ polarity. Figure 1.2 is a demonstration of the different conformations. In our measurements, we can see differences in the conformations as additional population distributions in the single molecule Förster resonance energy transfer (smFRET) data. With appropriately placed donor-acceptor fluorophores, smFRET results in different peak positions depending on the conformation, which result in different end-to-end separations [34, 35]. Stated more simply: one conformation results in one peak, if there’s two or more conformations, we see at least two peaks. We can also identify the number of conformations a sequence yields by using circular dichroism (CD) measurements [36]. To follow a systematic and simpler approach, we initially performed measurements with a GQ construct that is well known to fold into a single parallel conformation. After establishing key aspects of the protocol, we then proceeded to work with the more complex yet physiologically very significant human telomeric GQ (hGQ) that can fold into multiple conformations.
Figure 1.2: G-Quadruplex Conformations. The conformations are defined by the polarity of the DNA backbone along the G-tetrads. (a) A parallel conformation. (b) An anti-parallel conformation. (c) A hybrid conformation.

1.2.2 GQ Stability

GQs demonstrate large variations in their stability, due primarily to the loop lengths and the number of stacked tetrads [37, 38]. Another experimentally-tunable factor for modulating GQ stability is the monovalent cation concentration [39]. Not only can the GQ be stabilized by stacking interactions, but also by a chelating effect of an intercalated or planar ion [29]. This effect depends on the ionic radius of the cation: GQs are in general more stable in K⁺ compared to Na⁺ which in turn results in higher stability compared to Li⁺ [36]. That being said, some GQ sequences don’t need any ions to fold. The “3-layer-1-loop” GQ sequence (3Ly1Lp) we use in many of our studies, which has thee G-tetrad layers and 1 nt long loops, is incredibly stable and readily folds without cations.
1.3 Why Study GQ?

A GQ formation can be either detrimental or beneficial to the cell depending where and when it forms. In cells with deficiencies in GQ destabilizing helicases, GQs can interrupt cellular processes such as transcription, replication, or repair which in turn can cause DNA breaks and genomic instability [39]. Bloom and Werner syndromes are examples of such problems [40-42]. A GQ can act as a gene promoter by recruiting transcription factors, stimulating/facilitating transcription [43]. Alternatively, it can inhibit transcription and it can act as a gene inhibitor by recruiting repressors or by preventing the binding of transcription factors [2, 5, 6, 8]. GQ has been shown to be relevant in the field of cancer research [14, 44-48]. GQs that form in telomeres have been shown to inhibit telomerase activity of telomere elongation, which is utilized in about 85% of all cancers to maintain telomere length and avoid cellular senescence and apoptosis [1]. In addition, GQ can act as an oncogene activator or a repressor of a tumor suppression gene [7].
CHAPTER TWO

SMFRET AND TIRF MICROSCOPY

To study the interactions of GQ and small molecules or proteins, we use a combination of total internal reflection fluorescence (TIRF) microscopy and Förster Resonance Energy Transfer (FRET) [49-51]. It is through blending these techniques that we can observe single molecule dynamics at high time resolutions [34, 52-55]. I will first discuss TIRF and the benefits we gain from using it.

2.1 Total Internal Reflection Fluorescence Microscopy

Fluorescence is achieved when a fluorophore, a molecule that can fluoresce, is excited with the appropriate wavelength light. The excited electron loses energy due to various mechanisms and emits a photon of lower energy (longer wavelength) when it returns back to the ground state [56]. The difference in energy of the absorbed and emitted photons forms the foundation of fluorescence microscopy and enables a large number of applications. Fluorescence measurements can be performed in bulk where a solution in a cuvette is excited and the average emission is measured. However, TIRF microscopy has enabled extending these measurements into the single molecule regime [35, 57-69]. In total internal reflection (TIR), an evanescent wave which excites the fluorophores that are immobilized in the vicinity of the TIR interface is created. This wave has a penetration depth of about 100-200 nm [49]. This technique has some benefits that enable single molecule detection. The inelastic Raman scattering of light, which is not possible to distinguish from the fluorescence signal due to similar energy loss in both processes, by the bulk solution, mostly water, is largely eliminated [70]. This greatly
reduces the background signal as incident light does not penetrate beyond a thin layer at the interface. Of course with these benefits come some extra precautions that have to be taken. There are considerably more optical components and alignment that have to be done when using TIRF compared to bulk measurements.

2.1.1 Förster Resonance Energy Transfer

FRET is the energy transfer from an excited donor molecule to an acceptor molecule through a non-radiative dipole-dipole interaction. Since it’s governed by a dipole-dipole interaction, the energy transfer has $1/r^6$ type distance dependence, where $r$ is the separation between the donor and acceptor [71]. The range of distances that FRET is most sensitive depends on the overlap of the spectra of donor-acceptor fluorophores [51].

A major requirement for fluorophore selection is that the emission spectrum of the donor should have a significant overlap with the excitation spectrum of the acceptor. The FRET efficiency ($E_{\text{FRET}}$) can be measured by comparing the donor and acceptor intensities, $I_D$ and $I_A$, respectively:

$$E_{\text{FRET}} = \frac{I_A}{I_A + I_D}$$

When $E_{\text{FRET}} = 1$, there is 100% energy transfer from the donor to the acceptor yielding only acceptor emission and when $E_{\text{FRET}} = 0$, there is no transfer yielding only donor emission.
2.1.2 Combining the Techniques

We excite the donor fluorophores with the appropriate wavelength laser, which is blocked in the emission path by a filter. Then the excited donor fluorophores transfer their energy to the acceptors, which get excited to higher energy levels before returning to the ground state by emitting a fluorescent photon with a much longer wavelength. The amount of energy transferred, or in other words the intensity of acceptor emission can be related to the proximity of the fluorophore pair. This allows us to distinguish between states in which the fluorophores are close or far apart.

2.1.3 FRET Pair Selection

Cyanine 3 (Cy3) is used as the donor and Cyanine 5 (Cy5) is used as the acceptor fluorophore in our assays. These are aptly named based on the number of carbons linking their two cyclic domains [72]. Figure 2.1 shows the molecular diagram of the pair. The excitation and emission profiles of these fluorophores are given in Figure 2.2. The Förster radius (separation where $E_{\text{FRET}} = 0.50$) of Cy3-Cy5 pair is in the 5.5-6.0 nm range, resulting in 3-7 nm to be the most sensitive distance range where changes in distance yield to greater change in $E_{\text{FRET}}$, as shown in Figure 2.3 [73]. This pair has a maximum interaction distance of about 10 nm, beyond which there is no FRET.
Figure 2.1 The FRET pair Cyanine 3 (Cy3) and Cyanine 5 (Cy5). Cy3 is the donor fluorophore and Cy5 is the acceptor fluorophore.

Figure 2.2 The fluorescence excitation and emission spectra for Cy3 and Cy5. The excitation wavelength (532 nm), the wavelength (635 nm) at which the dichroic splits the donor and acceptor emissions are indicated. The significant overlap of the Cy3 emission profile with the Cy5 excitation profile is an important requirement for FRET to take place.
Figure 2.3  A plot of $E_{\text{FRET}}$ vs. Interfluorophore Distance. In the background is a gradient of the emission color where on the left side at low $E_{\text{FRET}}$ one observes high red emission and low green emission (designating acceptor dominated emission). In the middle both emissions are moderate correlating to a moderate $E_{\text{FRET}}$ value and on the right is mostly green emission with little red emission signifying a low $E_{\text{FRET}}$ value (designating donor dominated emission). This curve also demonstrates the more sensitive range of $E_{\text{FRET}}$ for this Cy3-Cy5 fluorophore pair to be between 3-7 nm, where the slope is the steepest.

2.2  Experimental Setup

2.2.1  Excitation & Emission Optics

The method used most in this study is prism-type TIRF with green-excitation, of which the optics is shown in Figure 2.4 [73].
Since TIRF is being used in combination with FRET, we have to ensure we have proper excitation optics and alignment. The way our optical table is set-up, both green and red excitation are possible as well as objective-type and prism-type TIRF. In this study, objective type was not used so I will not discuss it in great detail. Obviously, the first and most important optical component needed is a laser with the appropriate wavelength. One laser is a helium-neon laser of wavelength 632.8 nm (red) and the other is a solid-state laser of wavelength 532 nm (green). We have the lasers mounted such that they are at the same height from the optical table. It is the 532 nm green laser that is used to excite the donor, Cy3. The green laser passes through a couple of filter wheels so the power can be decreased as needed. With the help of some broadband dielectric mirrors and a long-pass dichroic mirror, the red and green lasers can be perfectly overlapped. A long-pass dichroic mirror is an optical component that allows light of wavelengths higher than a specific value to pass while lower wavelengths get reflected. Once the beams are overlapped, then it’s a matter of using mirrors to bring them to the appropriate position on the optical table where they hit a lens that focuses them at the prism-sample chamber interface.
Figure 2.4 A diagram of the prism-type TIRF smFRET microscopy setup. A green laser is used for excitation. Not indicated in this figure are the prism-type red excitation and the objective-type excitation optics. These were left out because the majority of the measurements were made with prism-type TIRF and green excitation.

2.2.2 Instrumentation

The Olympus IX71 inverted microscope has a selection of objectives and filters that can be used to increase the magnification and to block any laser excitation to reach the emission signal. After exiting the microscope, the light is focused at a point about 4 inches away from the microscope. It is at this point that a slit is placed to cut the profile of the beam to make it more manageable and bring it at the correct size for the CCD screen. We use a lens pair to carry the image to a conjugate plane so that we can fit the necessary optical components within the light path. After the light passes through the slit it hits a collimating lens. After that it hits a dichroic which allows the light of wavelengths higher than 635 nm to pass and lower wavelengths to reflect. This cutoff
wavelength is selected based on the emission profiles of the donor and acceptor fluorophores. The dichroic is aptly named since it splits the emission beam into two based on color/wavelength. Then using a set of mirrors, the light that passed the dichroic is reflected such that it intersects the light reflected off the dichroic in the middle of a focusing lens. Then a mirror is used to place the focus of these two beams on two halves of an Andor Ixon EM-CCD camera. Only part of the field of view can be projected onto the CCD screen since the donor and acceptor emissions are separated and projected onto only half of the screen. Recall that a comparison of the intensities of emitted light from the donor and acceptor fluorophores needs to be made simultaneously to calculate $E_{\text{FRET}}$, so they must be separated.

2.3 Data Acquisition and Analysis

2.3.1 Single Molecule Imaging

With this optical setup we have the ability to observe single molecules, by this I mean that we can observe and trace the intensities of both the Cy3 and Cy5 attached to our individual constructs that are immobilized on the imaging surface. Since we split the emission beam into two by using a dichroic and then focus the two parts on two different halves of a CCD, we create parallel images of the same location in the sample chamber, one with wavelengths greater than 635 nm and another with lower wavelengths. The CCD is 512 pixels by 512 pixels, so each “channel” that we’ve created is 512 pixels tall by 256 pixels wide. These channels are identified by the fluorophore from which the signal was collected. The signal of lower wavelength was collected from the emission of the donor fluorophore and is thus called the donor channel and similarly for the acceptor fluorophore.
channel. The pixel dimensions are 16 µm by 16 µm. By using a water objective that has a 60x magnification, each pixel on the CCD spans 267 nm by 267 nm of the imaging surface. This is a good enough spatial resolution that allows is to identify individual molecules [74, 75].

2.3.2 Generating a Mapping Function

Even though the donor-acceptor images are created from the same spot, there still may be some fine adjustments needed to match each molecule between the two channels beyond a simple translation. To ensure we have the correct molecules matched, we need to collect an image from a “bead sample” first. This bead sample is a sealed sample chamber that has multicolor microbeads, approximately 200 nm in size, immobilized on the surface. These beads are very easy to observe in our CCD camera since they are fairly large and bright compared to the typical molecules for which the camera is designed. The bead sample is used to create a “mapping” that correlates molecules in one channel to those in the other channel. We do this by generating a 16 term polynomial function that maps each point in the donor channel to a point in the acceptor channel based on the peaks and their positions that are identified in an earlier step of the analysis. Once this mapping function has been created, it can be used in the analysis of the collected sample emission. Of course, there are special parameters that are used, like collection radius and peak threshold, to identify the peaks and their values can be tuned. Other tune-able parameters used in the collection and analysis include the time resolution for photon collection, data scaler, gain, and peak radius. Figure 2.5 shows an image of the two channels in false color with a scale bar. A collection of molecules is indicated with a
white circle to demonstrate the patterns are the same in both channels. A smaller white circle indicates a single molecule, of which the time trajectory (called trace) can be monitored.
Figure 2.5 A single molecule image of a bead sample. The left is the “donor channel” and the right is the “acceptor channel”. The donor and acceptor channels are shown in false color. As indicated, the scale bar is 13.3 µm which covers a length of 50 pixels on the CCD, with each pixel having an actual size of 16 x 16 µm. The larger white circle demonstrates that the images in each channel have the same patterns and thus represent the same region within the sample chamber. The smaller white circle is a single molecule, for which we collect the emission signal over time.
2.3.3 Single Molecule Traces

By collecting these images over time, traces of the donor and acceptor intensities and $E_{\text{FRET}}$ can be made for each molecule, as shown in Figure 2.6. Then all of the single-molecule $E_{\text{FRET}}$ traces can be binned in a histogram to give a population percentage of molecules in different $E_{\text{FRET}}$ states during the collection time. Once this is generated, then Gaussian functions can be fitted to identify peak positions and further analysis can be made.
Figure 2.6 A single molecule FRET trace. The upper plot is a measure of intensity over time, tracking the intensity in both the donor and acceptor channel for a single molecule. Using the values of these intensities, a trace of the $E_{\text{FRET}}$ can be made as shown in the lower plot. Two well defined and long lived FRET levels (higher FRET being folded and lower FRET being unfolded states) in addition to intermediate states with a shorter dwell time are visible in the bottom trace. This trace demonstrates that the dynamics of the system can be monitored in this assay.
2.4 The Sample Chamber

Given the experimental setup and measuring in the single molecule regime, a specially designed sample chamber must be made. This chamber needs to be thoroughly cleaned and then coated with an isotropic and homogenized monolayer on the surface, which also contains molecules that serve as binding sites for the DNA construct that is to be examined (see this earlier dissertation from our group for a detailed description of the slide preparation methodology) [73]. Since the manner of construct deposition is aqueous in nature, the chamber needs to be able maintain a certain volume of relevant buffers. This volume does have its own criteria for reasons I will explain later.

2.4.1 Coating the Surface with PEG

The sample chamber used in this study is created from a quartz microscope slide with 4 or 6 pairs of parallel holes drilled along the length, some double-sided tape, a coverslip and some epoxy. Prior to assembly the slide and coverslip were cleaned thoroughly and coated with a monolayer of 99% polyethylene glycol (PEG) and 1% biotin-PEG. This was done to ensure the surfaces are homogenous and isotropic and to create binding sites. If this protocol wasn’t followed, there would be a lot of not-specific binding to the slide surface that would cause clumping. For further details on the PEGylation protocol, please see this earlier dissertation from our group [73]. Basically, we wouldn’t be in the single-molecule regime anymore.

2.4.2 Assembly

The double sided tape is used to create channels that run parallel to the short width of the slide and also to provide a spacer between the slide and coverslip. These
channels are made primarily so that the same slide can be used for 4 – 6 different experiments, but an additional benefit is that because of the size of the channels, a uniform flow can be achieved. When finished, the volume of each channel is between 10 - 20 µL. A quick-dry epoxy is then used to seal the long edges of the slide and coverslip. A picture of the sample chamber is shown in Figure 2.7.

Figure 2.7 A picture of the sample chamber. Above is an actual picture of the sample chamber. The chamber is illuminated with a green laser which is coupled to the sample via a prism. The bright line in front of the image is an indication of the total internal reflection and the light propagates close to the interface. The epoxy along the length of the slide, the six pairs of parallel holes (used for sample deposition and buffer exchange), the double-sided tape (used to create the individual channels), the clamps used to hold the sample in place), and the prism are visible in the setup. The water-type objective is vaguely visible under the slide. The quartz slide is in contact with the oil/prism and the glass coverslip is in contact with the water/objective.
2.5 Sample Preparation

Before an experiment can be carried out, special considerations must be made when developing the DNA construct to be tested. The construct needs to be carefully designed to ensure the experiment is probing for the appropriate results. Once a solution of the DNA construct is made, it then needs to be immobilized on the surface of the sample chamber.

2.5.1 Design and Annealing

Most of the constructs observed in this lab consist of two ssDNA strands that are then annealed together and hybridized (see Appendix B, Configuration M & N). These constructs generally consist of a fluorophore pair, a biotin, and a segment of duplex DNA. The fluorophore pair serves as a means of measuring the state of the construct since fluorescently unlabeled DNA cannot be imaged in the microscopy methods utilized in our lab. The DNA construct typically has a short duplex stem (typically 18 bp) and a ssDNA overhang that consists of the sequence of interest for the particular study. The duplex stem keeps the sequence of interest away from the surface to avoid unwanted surface interactions, and also provides a scaffold to attach a biotin and a fluorophore. The biotin serves as the linker to immobilize the construct to the surface to avoid non-specific binding and clumping of constructs.

2.5.2 Creating Biotinylated Construct Binding Sites

Once the sample chamber is made and the epoxy has cured, it is ready for use. The channel being used is first hydrated, flushing any dust or other contaminants that appeared during preparation. Then about 50 µL of 0.005 mg/mL Neutravidin is incubated
in the channel. Neutravidin is a protein that non-covalently binds to biotin and creates the binding sites for biotinylated DNA constructs. Once the incubation is complete, the construct of interest is incubated until a decent molecular density is reached on the surface. There are two key details that are important to consider in getting the appropriate density. The first is that we want enough molecules per measurement so that proper statistics can be made during the analysis. The second aspect is that we need a low enough density to prevent overcrowding would prevent single-molecule detection. A proper molecular density in this study is between 250 – 300 molecules per frame. Once a proper density is achieved, the experiment is ready to be started. Figure 2.8 shows a cartoon of a partial duplex construct within the sample chamber in the folded and unfolded configurations.
Figure 2.8 A schematic of the DNA bound to the sample chamber surface. As shown, a monolayer of PEG +biotin-PEG is coated on the clean surface of the slide giving rise to an isotropic and homogeneous surface with binding sites. Neutravidin binds to the biotin and provides further binding sites for biotinylated DNA constructs. These are two “partial duplex” constructs where the left is in the high FRET state since the GQ is folded and the right is in a lower FRET state since the GQ is unfolded.
CHAPTER THREE
THE MACROCYCLE

3.1 Summary

The main goal of this study is to develop a method to study the interactions of GQ with several different agents when the GQ is subjected to an external force. We want to take advantage of the elastic properties of short dsDNA, i.e. shorter than or comparable to the persistence length of dsDNA, as a tool to generate the applied force. Such a scheme would enable us to use smFRET to monitor the dynamics of many DNA molecules simultaneously, unlike many other commonly used force methods such as optical or magnetic tweezers where only a single molecule is studied at a time. So how can this be done? We have some constraints that must be first considered. The sensitive range of smFRET limits how far apart the donor-acceptor fluorophore pair can be, so those must be placed proximal to the GQ sequence. The GQ itself has dimensions that must be considered as well as a few other dimensions I’ll mention shortly.

3.2 Construct Design

3.2.1 The “Macrocycle”

We have designed a construct that we have called a “macrocycle”. It’s a DNA construct consisting of 3 strands. One strand is 120 nt long and has a Cy 3 (Donor) fluorophore attached to the 5’ end via a basic carbon linker. Another strand is 90 nt long, complimentary to the middle 90 nt of the longer strand previously mentioned. This strand has a biotin attached via a long carbon linker about 20 nt from the 5’ end (see Appendix II, Configuration A). The fluorophores and biotin are attached in these locations due to
the PCR strategy we used to create the stock of these strands. Then the two strands are 
hybridized, and we get a 90 bp long duplex strand we call a “macroduplex” which has a 
15 nt long ssDNA overhang we call a “sticky end” on each side (see Appendix II, 
Configuration B). We then add a third strand that we call a “bridge strand” (see Appendix 
II, Configurations C & F). This bridge strand includes a sequence of interest (GQ forming 
sequence or a poly-thymine sequence for the purposes of this thesis) in the middle and 15 
nt complementary “sticky ends” on each side of the sequence of interest. These 
complementary sticky ends hybridize with the sticky ends of the macroduplex, hence 
bridging the ends of the macroduplex (see Appendix II, Configurations K and L). We 
ordered these bridge strands HPLC purified from Integrated DNA Technologies (IDT). 
One sequence of interest we use simply has 20 thymines between the complementary 
sticky ends. The other one is a “three-layer-one-loop” (3Ly1Lp) quadruplex sequence, 
which folds easily, at very little to no salt, and is also fairly stable. As it may be evident 
from the name, this is a quadruplex that consists of three stacked G-tetrads and loops that 
are one nt long. This sequence is known to fold into only a parallel conformation, which I 
admit is actually quite counter-intuitive but CD measurements have proven it as such. 
This parallel conformation is one where the four corner edges all run in the same polarity. 
The bridge strands are internally labeled with a Cy3 (donor) fluorophore between the 
sequence of interest and one of the complementary sticky ends. Please see Appendix I for 
the sequences and Figure 3.1 for a detailed schematic.
Figure 3.1 A schematic of the macroduplex with a bound GQ forming bridge strand by Sticky End 1. This configuration is an example of the linear form of the construct as Sticky End 2 is not hybridized with its complementary to form a circular DNA (the macrocycle). This is a three stranded construct in which the first two hybridize to form the macroduplex and the third is the bridge strand with the sequence of interest, in this case a GQ forming sequence, for the study. Strand 1 is biotinylated (blue circle with a “B”) to allow the construct to bind to the surface of the sample chamber. Strand 1 has a Cy3 (green circle), a 15 nt Sticky End 1 sequence (light blue), the 90 nt complementary to Strand 1 (black), and a 15 nt Sticky End 2 sequence (red). The bridge strand (Strand 3) has the complementary sequence to Sticky End 1 (light blue, 15 nt), the GQ forming sequence in the folded state (purple layered structure), and the complementary sequence to Sticky End 2 (red, 15 nt).

3.2.2 Development of the Macroduplex

We worked with our collaborators in Dr. Basu’s Lab a PCR strategy to create the macroduplex. Using fluorophore-tagged and biotin-tagged primers, we were able to optimize the thermal cycling to create the 120 nt and 90 nt strands. We then purified the strands and annealed them to form our macroduplex stock. One nice aspect of this strategy is that we can use this macroduplex piece for multiple bridge strands, so it is quite useful.

3.2.3 Tunable External Force

This construct was designed to observe the stability of different quadruplex-forming sequences when subject to an external force. The external force comes from the rigidity of the macroduplex strand, which is shorter than the persistence length of
dsDNA. We can adjust the magnitude of the force exerted by adjusting the concentration of magnesium in the environment, and thus the flexibility of the duplex.

### 3.3 Concept and Why is a Force Expected?

DNA in nature can be found in several forms. In humans it is found densely packed within the cell nucleus in the form of chromosomes. In bacteria it can be found as duplex DNA in a ring called a plasmid. Very short lengths of duplex DNA follow a rigid rod model and longer lengths of DNA follow the worm-like chain model [76]. It has been shown that the flexibility of a segment of DNA increases with increasing cation concentration due to more efficient screening of the negative charges on the DNA backbone [77, 78]. Therefore, the probability that a short segment of duplex DNA with complementary ‘sticky’ ssDNA overhangs to form a circular structure increases with the available cation concentration. Once the duplex DNA forms a circle, the cation concentration could be decreased which would then increase the rigidity of the duplex, generating a force that wants to return the duplex segment to a linear state. The smaller the radius of this circle, the higher the applied force.

One needs to be aware that kinks and bulges or mispairing of duplex DNA as well as nicks in the backbone would greatly change the flexibility of the duplex DNA which could then greatly decrease the force that could be exerted [79]. So the big question is: what is the right radius of the duplex circle to use that would generate a strong enough force to unfold a GQ structure? Previously, an attempt had been made using 70-80 bp long duplex DNA with no success. The construct used in this study is 120 bp long when fully assembled and also includes a sequence of interest within the bridge strand that
varies in length depending on the construct under study. In one unstrained helical turn of
duplex DNA, there are approximately 10 base pairs spanning a length of about 3 nm.
Approximating the sequence of interest to range from 3 nm to 6 nm long, the macrocycle
will have a circumference of about 40 nm, or a radius of about 6 nm. For comparison, the
persistence length of duplex DNA is about 150 bp, or 50 nm; for ssDNA the length is 1
nm [80, 81]. A large persistence length does not mean that the object is not flexible; it
just means that it needs to be quite long for thermal fluctuations to bend it. As the
macrocycle designed for this work is shorter than the persistence length of dsDNA, it
would be expected to be elastically constrained when forced to the circular (macroyclic)
configuration via Watson-Crick pairing of the sticky ends (see Appendix II,
Configuration K). The magnitude of the force this elastic constraint gives rise to is not
clear at this point; however, it could be measured by studying the folding/unfolding
dynamics of a GQ structure for which the force required to unfold the GQ is determined
via another means such as optical tweezers. The design then essentially provides a means
to study the dynamics of GQ structures while a constant force is applied on them [82-84].

3.4 Ligation

Since we create our macroduplex out of multiple strands we observe inherent
nicks in the backbone at the joint points that connect the macroduplex with the duplex
formed by the sticky ends and their complementary strands (see Appendix I,
Configuration K). Such nicks provide increased flexibility for the macrocycle as they
enable the DNA to sharply bend at these points. A ligase treatment can join the two
backbones of the DNA segments and remove the nicks, but was left out of this proof-of-
concept work and will be addressed in future work [85]. Despite the two backbone nicks, we still expect a tensile force to be exerted by the macrocycle.

3.5 Sample Procedure

3.5.1 Immobilization on the surface

Since this macrocycle construct is composed of three ssDNA segments, proper care must be taken in order to ensure a proper density is achieved without the formation of aggregates or other undesired configurations. The first step is to incubate the macroduplex in the channel. Since the macroduplex consists of the two longer segments, one with a biotin modification and the other with a Cy3 modification, we can check if they annealed properly by just checking if we can observe any molecules in the donor channel. The biotin modified strand binds to the surface and if the strands aren’t annealed, the Cy3 modified strand would be washed away. Once the correct number of molecules is immobilized on the surface, the bridge strand is introduced and incubated in the channel. There are precautions taken at this step as well. The incubation concentration of the bridge strand should be low enough so that each macroduplex immobilized on the surface has only one bridge strand bind to it. I will discuss later the treatments needed for the cases in which more than one bridge strand binds (see Appendix II, Configurations E, & H-J). The incubation time needs to be considered because it’s not just a matter of the bridge binding to the macroduplex, but also formation of the macrocycle after bridge binding. We allow the best conditions for this to occur by having high cation concentrations and incubating over a long time period. This of course leads to other questions and problems like how to tell the difference between the circular configuration
and the linear configuration and others I will address later on. This immobilization procedure does have one disadvantage in that the construct starts in a high cation concentration, and special considerations must be made while lowering the cation concentration for GQ forming sequences to ensure the folding pathway is not biased because of incubation in the high cation concentration.

### 3.5.2 Bridge Binding Considerations

Ideally, the bridge strand with two sticky ends should initially bind to one of the sticky ends of the macroduplex as shown in Figure 3.2, and then bind to the other sticky end and form the circular macrocycle configuration. However, several alternatives can happen and must be addressed.

Figure 3.2 Macrocycle DNA constructs with a GQ forming bridge. The schematic on the left illustrates the unfolded configuration where lower FRET is expected and the configuration on the right illustrates the folded configuration where a higher FRET is expected. PEG, biotin, and Neutravidin serve the same purposes as described before.
3.5.2.1 Single bridging on opposite end

Let’s consider the first scenario as demonstrated in Figure 3.3. What happens when the bridge strand binds to the end of the macroduplex in which the sticky end 2 pairs with its complementary sequence (See Appendix II, Configurations D & G)? Before formation of the macrocycle, only one FRET state would be observed, and that is the donor-only state since there will be no energy transfer to the acceptor due to its large distance from the donor. Recall that the maximum distance for observing FRET between a Cy3-Cy5 pair is about 10 nm and in this case the interfluorophore distance is over 30 nm. However, we could get FRET between Cy3-Cy5 if the macrocycle were to form. If the macrocycle doesn’t form, it’s an easy issue to fix. In the analysis, the signal that comes from donor only configurations gets removed and this configuration does not “contaminate” the data.

Figure 3.3 Bridge binding at the Sticky End 2 of the macroduplex. Binding of the bridge to this side results in an interfluorophore distance over 30 nm, at which FRET does not take place. This results in the absence of Cy5 emission while Cy3 emission is maximum (donor only state) as long as the construct remains in this linear state. If it were to transition to the circular macrocyclic configuration, it would yield higher FRET states.
3.5.2.2 “Double bridging”

Another configuration that needs to be addressed is when multiple bridge strands bind to the macroduplex, or rather “double bridging” as shown in Figure 3.4. When this occurs a macrocycle cannot form (see Appendix II, Configurations E, & H-J). As in the previous case, the Cy5 attached to the bridge far from the Cy3 does not contribute and the other bridge can go into two possible FRET states, folded or unfolded (assuming the bridge has a GQ forming sequence, of course). This configuration is locked into the linear configuration. To prevent this, the incubation concentration of the bridge strand is kept low. We used red excitation and counted the number of Cy5 photobleaching steps to see how many acceptor fluorophores and thus bridge strands were bound to each macroduplex. The results of this test are shown in Table 3.1 and demonstrated that at 10 nM bridge strand concentration for a 17 hour incubation, about 80% of our molecules have a single bridge strand. If this configuration is present in the system, it will need to be treated the same way as the case of a single bridge binding without macrocycle formation, which I will discuss later.

Figure 3.4 A schematic of a macroduplex with a bridge bound to each side (double bridged). The macrocycle cannot form since the remaining ssDNA overhangs are not complementary. The acceptor on the right doesn’t contribute to FRET since it is beyond the FRET range. The acceptor on the left is in a high FRET state since GQ is folded.
Table 3.1 Red excitation to probe the number of bridge strands attached to quantify the fraction of constructs bound with two bridge strands. A 10 nM bridge strand concentration and an incubation period of 17 hours were used. Since each bridge strand has an acceptor molecule, the number of photobleaching steps under red excitation indicates the number of bridge strands. The N/A column represents the percentage of molecules that did not have any photobleaching steps during the period of observation. The adjusted One Step and Adjusted Two Step columns refer to the statistics when the non-photobleaching molecules (the N/A column) are eliminated from the statistics. Of the molecules that exhibited photobleaching steps, over 75% had a single step which demonstrates this is a sufficient bridge strand incubation period for the given concentration.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Movie</th>
<th>Total Number of Molecules</th>
<th>Frames Collected</th>
<th>Time Resolution</th>
<th>One Step</th>
<th>Two Step</th>
<th>N/A</th>
<th>Adjusted One Step</th>
<th>Adjusted Two Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hel 1</td>
<td>110</td>
<td>1200</td>
<td>500 ms</td>
<td>20.95%</td>
<td>7.62%</td>
<td>71.43%</td>
<td>73.33%</td>
<td>26.67%</td>
</tr>
<tr>
<td>1</td>
<td>hel 2</td>
<td>101</td>
<td>2000</td>
<td>500 ms</td>
<td>28.87%</td>
<td>8.25%</td>
<td>69.89%</td>
<td>77.78%</td>
<td>22.22%</td>
</tr>
<tr>
<td>1</td>
<td>hel 3</td>
<td>100</td>
<td>2000</td>
<td>500 ms</td>
<td>41.24%</td>
<td>9.28%</td>
<td>49.48%</td>
<td>81.63%</td>
<td>18.37%</td>
</tr>
<tr>
<td>1</td>
<td>hel 4</td>
<td>101</td>
<td>2000</td>
<td>500 ms</td>
<td>40.40%</td>
<td>15.15%</td>
<td>44.44%</td>
<td>72.72%</td>
<td>27.27%</td>
</tr>
<tr>
<td>2</td>
<td>hel 1</td>
<td>113</td>
<td>2000</td>
<td>500 ms</td>
<td>52.68%</td>
<td>8.65%</td>
<td>38.65%</td>
<td>85.94%</td>
<td>14.06%</td>
</tr>
<tr>
<td>2</td>
<td>hel 2</td>
<td>94</td>
<td>2000</td>
<td>500 ms</td>
<td>52.27%</td>
<td>14.77%</td>
<td>32.95%</td>
<td>77.97%</td>
<td>22.03%</td>
</tr>
<tr>
<td>2</td>
<td>hel 3</td>
<td>97</td>
<td>2000</td>
<td>500 ms</td>
<td>51.65%</td>
<td>21.98%</td>
<td>25.27%</td>
<td>70.15%</td>
<td>29.85%</td>
</tr>
</tbody>
</table>
3.5.3 Macrocycle Formation

3.5.3.1 GQ Forming Bridge

Once the macroduplex is bound to the surface in an appropriate molecular density, the bridge strand can be incubated. Incubating in high cation conditions ensures the macroduplex is flexible enough to allow macrocycle formation, however high cation concentration also enables GQ to fold. The smFRET population distribution we obtain for the 3Ly1Lp Bridge in 1 M KCl and 10 mM MgCl₂ is given by Figure 3.4. We see two peaks in the population distribution, the lower one related to the donor only emission and the high peak due to the folded GQ, which could be within the linear or circular (macrocyclic) state (see Appendix II, Configuration C & K). Only the macrocyclic state is the desired configuration so it is necessary to remove the linear configuration.
Figure 3.5 A smFRET histogram for the macrocycle with a 3Ly1Lp Bridge at 1 M KCl and 10 mM MgCl₂. This histogram shows two FRET peaks, the lowest one is the donor only peak and is removed in the following analysis steps but is demonstrated here to show the earlier state of the data, and the high FRET peak represents a folded GQ. However, due to the similarity of the folded GQ peak in the linear and circular configurations, it is not possible to determine what fraction of these constructs are in the circular configuration. The data were acquired on 03/03/2016.

3.5.3.2 Additional Peak for PolyT Bridge

As a control and also to probe the system further, the same construct was made with the exception of changing the GQ-forming sequence to a poly-thymine (PolyT) sequence so that no structure can be formed within the bridge strand. At the same salt concentrations, three peaks are observed, the lowest being donor-only and the other two
due to the linear and circular (macrocyclic) configurations (see Appendix II, Configurations F & L) as shown in Figure 3.5. This begs the question, why is it that the peaks representing these different configurations are significantly separated for the PolyT bridge, but overlap for the 3Ly1Lp bridge? It is in part because of the stability of the 3Ly1Lp GQ; the macrocyclic state exerts a force on the bridge strand and stretches the coiled ssDNA of PolyT bridge. However, in the case of GQ such a stretching does not take place since GQ is very stable and remains folded. Therefore, the PolyT bridge demonstrates two peaks, one for the linear configuration and one for the circular configuration (see Appendix II, Configurations F & L), while the GQ-containing bridge demonstrates a single folded GQ peak for both the linear and circular configurations (see Appendix II, Configuration C & K). Since only the circular configuration is desired, it is important to be able to distinguish between the two for the GQ forming sequences, or at least eliminate the molecules that remain in the linear configuration and do not form the macrocycle.

The above conjecture is logically sound, but several things need to be done to justify it. Firstly, the two peaks observed for the PolyT bridge and the states they represent need to be positively identified. After discerning the peaks and the particular state they represent, it becomes necessary to remove the peak that is due to the linear configuration since we aim to study only the circular configuration. The PolyT system serves as an ideal system to test our ideas for eliminating the undesired linear configuration since we can easily determine if the method is working by observing if the peak representing the linear configuration is eliminated while the peak representing the
circular configuration is left unaffected after the prospective treatment. Testing these ideas would not have been possible for the GQ-forming bridge since the FRET peaks for the linear and circular configurations overlap thus making it impossible to identify whether the prospective method eliminates the circular, linear or both configurations.

Figure 3.6 A smFRET histogram for the macrocycle with a PolyT bridge in 1 M KCl and 10 mM MgCl$_2$. This histogram shows two populations (the donor only peak has been removed). A PolyT sequence should yield a singlet for most constructs, but this construct is found to be in two configurations. The linear configuration should yield one peak and the circular configuration should yield a different peak since it is subjected to a force that stretches the ssDNA bridge thus decreasing the E$_{\text{FRET}}$ by increasing the interfluorophore distance. The data were acquired on 06/05/2016.
3.5.3.3 A Logical Projection

Reconsidering the PolyT bridged macrocycle in the high cation condition, we suspect that the higher FRET peak is due to the linear configuration and the lower FRET peak is due to the macrocyclic configuration. The force exerted by the macrocycle stretches the coiled ssDNA bridge thus increasing the interfluorophore distance. On the other hand, in the linear configuration, the bridge strand is unstretched and is just freely hanging from the end of the macroduplex strand. This free end is only subjected to charge screening from the ions in the nearby environment (we call this salt compaction) and thus behaves essentially as a coiled ssDNA. We affirm this by creating a partial duplex control construct which can only be in this coiled (linear) configuration and identify the FRET level representing this state.
CHAPTER FOUR

DISCERNING THE RELEVANT SIGNAL

4.1 The Partial Duplex Control

4.1.1 Design

The partial duplex control must mimic the behavior of the linear configuration but we also wanted it to be cost effective. Like most of the constructs we use, we want the GQ forming sequence or the sequence of interest to be sufficiently far from the surface to avoid undesired interactions. Since the bridge strands already contain the sequences of interest and are readily useable in the macrocycle construct, it makes sense to somehow adapt and use those. We designed a short ssDNA strand that is complementary to sticky end 1 of the bridge strand. This new strand has a biotin modification to anchor it to the surface and a Cy3 label at the opposite end to biotin, which will serve as the donor for the Cy5 that is only on the bridge strand. Annealing this new strand with the bridge strand results in a partial duplex construct that has a 15 bp long duplex stem and a ssDNA overhang that includes the sequence of interest, Cy5, and the complementary to sticky end 2 (see Appendix II, Configurations M & N). The complementary to sticky end 2 is insignificant for this construct and it does not influence the observed FRET as it is not between the donor and acceptor fluorophores. Being only 15 bp long, this duplex stem is very similar to the 18 bp duplex stem we frequently use in the lab and have established its functionality numerous times. This small stem strand is essentially a highly truncated version of the macroduplex in the linear configuration.
4.1.2 Equivalent Identification

The main idea of this partial duplex control is that the sequence of interest freely hangs at the end of the linear macroduplex should behave similarly to the partial duplex construct when subject to the same environmental conditions (see Appendix II, Configurations F & N). Since they behave in the same manner, they should maintain identical interfluorophore distances and thus should result in the same peak positions on E\textsubscript{FRET} histograms. This is indeed the case and is easily visible in the case of a PolyT bridge, shown in Figure 4.1. The E\textsubscript{FRET} = 0.63 is therefore identified as the FRET peak representing the coiled state. In the case of the bridged macroduplex, this FRET peak then must be due to the linear configuration while the lower FRET peak must then represent the circular configuration, as expected since it is under tension due to the elastic properties of the duplex DNA.
Figure 4.1 A comparison of the partial duplex construct with that of the macrocycle construct. 1 M KCl and 10 nM MgCl2 were used in the assays. The top histogram is for a partial duplex construct and shows a donor only peak and a single peak representing the unstrained or coiled DNA population. The bottom histogram is for the macrocycle construct that can be in a linear or circular form, hence two peaks other than donor only peak. We expect the linear form of the bottom histogram to match the partial duplex construct since the ssDNA is in the unstrained coiled configuration in both cases. As expected the higher FRET peak in the bottom histogram matches the peak in the top histogram. Therefore, the lower FRET peak in the bottom histogram must represent the PolyT bridge in a macrocycle. The stretching force results in a lower FRET peak due to increased interfluorophore distance. The data were acquired on 07/03/2016 and 02/09/2016.
4.1.3 Further Validation

A similar partial duplex control with a short duplex stem was also performed for the GQ forming construct to confirm that the observed high FRET peak does represent the folded GQ state (see Appendix II, Configurations C, K, & M). As expected both the partial duplex construct and the 3Ly1Lp GQ bridged macrocycle yield a FRET peak at $E_{\text{FRET}} = 0.85$. These observations further validate the idea of using particular duplex DNA constructs to identify the FRET peaks when the bridged macroduplex is in the linear configuration. They also demonstrate that the PolyT bridge data does indeed show the first reliable evidence for formation of circular DNA and that it can apply force on the bridge strand inserted between the ends of the macroduplex DNA. However, we still need to devise a method to eliminate the linear configuration to ensure that only the data from the circular DNA is identified for further analysis.

4.2 The Hunt for a Macrocyclic Unfolded GQ

One must ask: if cation concentration increases the stability of the GQ, what happens at low enough cation concentration where the GQ cannot remain folded? Does the GQ unfold and then mimic the stretched PolyT case and thus split the histogram into discernable peaks? Given the high stability of the 3Ly1Lp GQ, we did not observe any unfolding of the GQ in either the macrocycle or the partial duplex constructs, even in the absence of any cations in the environment. This could be due to two reasons, the first as hinted in the previous statement is that the GQ is very stable and simply withstands the force exerted by the macrocycle. The other scenario is that the macrocycle could have too large a radius and thus may not be exerting a significant force on the GQ, even in the no
salt condition. The way to check this is to use a bridge strand with a weaker GQ. Using the hGQ sequence in our bridge strand altered the length of the bridge slightly which lead to a shift in the folded and unfolded peak positions but these states should still be easily discernable from each other. It is with this construct that folding/unfolding dynamics were observed. However, interestingly there was a third unidentified peak in the FRET distribution in addition to the two folded and unfolded GQ peaks, as shown in Figure 4.2. We will discuss the dynamics of the hGQ bridge strand later, but this shows that the 3Ly1Lp sequence is just very stable and the macrocycle is in fact exerting a significant force on the bridge strands.
Figure 4.2 A smFRET histogram for the macrocycle with a hGQ bridge in 150 mM LiCl and 5 mM MgCl₂. The donor only peak has already been removed and the E$_{\text{FRET}}$ peak positions are 0.40, 0.55, and 0.82. The nature of the third peak is still under investigation as are the dynamics between these three populations. The data were acquired on 04/12/2016.

4.3 Elimination by Quenching

The matter of removing the undesired linear configuration is still an issue that needs to be resolved before bridges with a GQ forming sequence can be studied. What is a difference between the linear and macrocyclic configurations that can be exploited to allow us to remove the undesired population? The difference we exploited is the
existence of an open ssDNA overhang, the sticky end 2 complementary sequence of the bridge strand, which is available for binding a new strand only in the linear configuration (see Figure 4.3). Regardless of bridge sequence, this strand is an available binding site in the system for the linear macroduplex but not the circular macrocycle. This sequence is available for binding in the partial duplex control constructs as well, as described earlier. By incubating a 4th strand into the system, we can identify all of the molecules with this open binding site. This additional strand is 15 nt long (the sticky end 2 sequence) and has a Cy5 quencher attached to the 3’ end. When introduced to the system, it will form Watson-Crick pairing to any open sticky end 2 complementary strands. When the strand is paired the quencher will be in close proximity to the Cy5 and will absorb any emission from it. The molecule then essentially becomes a donor-only signal which gets eliminated during the analysis. The molecules that are in circular configurations will be unaffected because the binding site (sticky end 2 complement) for the additional strand with the quencher is blocked since the Watson-Crick pairing has already taken place with the other end of the macroduplex. Without the open binding site, the quencher will not be able to remove the Cy5 signal and FRET will be unaffected. Thus by adding this 4th strand, the undesired linear configuration can be removed for any given bridge strand. Obviously for any partial duplex control, regardless of bridge strand used, all of the signal should be turned into donor-only upon addition of the quencher strand since the binding site is open by design. We did use the partial duplex construct to determine the appropriate concentration and incubation period to achieve proper quenching as shown in Figure 4.4.
Figure 4.3 A schematic of the quencher assay for the partial duplex DNA construct. The ssDNA strand with the quencher (black circle) binds to the available sticky end complement and quenches the fluorescence from Cy5.
Figure 4.4 A plot of molecule count vs. time using red excitation upon addition of a quencher. Red excitation data demonstrating the reduction in the number of fluorescing Cy5 molecules after 10 nM quencher strand is added to the chamber. A partial duplex construct that contains a hGQ sequence was used. As time progresses, more of the quencher strands bind to the available site and result in a reduction in the number of fluorescing Cy5 molecules. The assay was performed in 150 mM KCl and 5 mM MgCl₂. After about two minutes of incubation, almost all available sites are bound by the quencher. The data were acquired on 05/31/2016.
Figure 4.5 Quencher elimination. The smFRET histogram on the left is the macrocycle with a PolyT bridge with the donor only peak removed. This assay was carried out at 1 M KCl and 10 mM MgCl$_2$. The peak positions of the macrocyclic construct are $E_{\text{FRET}} = 0.49$ and $E_{\text{FRET}} = 0.63$. The peak position of the quenched system is $E_{\text{FRET}} = 0.49$, showing that the high population, the linear configuration, was blocked as suspected. The data were acquired on 06/05/2016.

Figure 4.6 Imaging the quenching process for the partial duplex control. (a) An image produced red excitation displaying the population containing fluorescent Cy5. Note there is no donor excitation and thus the channel looks blank. (b) An image produced by red excitation of the same population 10 minutes after the quenching strand was injected. As shown, the population was greatly reduced and the donor channel is blank like before. (c) An image produced by green excitation at a time greater than 10 minutes after introduction of the quencher strand. This gives evidence that the quencher does not affect the emission of the donor fluorophores. This assay was run at 150 mM KCl, 5 mM MgCl$_2$, and 10 nM quencher strand for the partial duplex construct with a hGQ sequence. The data were acquired on 05/31/2016.
4.4 Utilizing the Quencher with a Macrocycle Construct

Having established the functionality of the quencher with a partial duplex construct, we proceeded to apply it to macrocycle construct. We expected the quencher to eliminate the linear configurations, regardless of whether they have a single or double bridge, and not influence the circular configuration in which the binding site for the quencher is not available. A bridge strand with a hGQ sequence was utilized in the macrocycle construct. Figure 4.7 shows a promising difference between the partial duplex and macrocycle constructs in that almost all Cy5 molecules were quenched after incubating 10 nM of quencher strand for 10 minutes for the partial duplex construct, while a significant number of Cy5 molecules continue to fluoresce for the macrocycle construct under the same assay conditions. These results suggest that there are some constructs that are not accessible to the quencher strand, which we consider to be the circular form of the construct.
Figure 4.7 A Comparison of the partial duplex and macrocycle constructs after introduction of the quencher. (a) An image produced by red excitation of the partial duplex construct after 10 minutes of quencher strand incubation. (b) An image produced by red excitation of the macrocycle construct after 10 minutes of quencher strand incubation. This assay was run at 150 mM KCl, 5 mM MgCl₂, and 10 nM quencher strand for both constructs. The partial duplex data were acquired on 05/31/2016 and the macrocyclic data were acquired on 06/01/2016.

With these encouraging results at hand, we proceeded to quantify the number of Cy5 molecules that continue to fluoresce after adding the quencher strand. Figure 4.8 demonstrate these results. Unlike the partial duplex construct where essentially all Cy5 molecules were quenched after 2 minutes of incubation (see Figure 4.4), a significant number of Cy5 molecules continue to fluoresce even after one hour, which we project to be the circularized DNA constructs. The data essentially shows a fairly constant number
of Cy5 molecules after an initial sharp drop, which we believe is due to quenching of the molecules that are in linear configuration.

Figure 4.8 A plot of molecule count vs. time using red excitation upon addition of a quencher. Red excitation data demonstrating the reduction in the number of fluorescing Cy5 molecules after 10 nM quencher strand is added to the chamber. A macrocycle construct that contains a hGQ sequence was used. As time progresses, more of the quencher strands bind to the available site and result in a reduction in the number of fluorescing Cy5 molecules. The assay was performed in 150 mM KCl and 5 mM MgCl₂. Even after an hour, Cy5 signal was still present indicating that those molecules did not have an open binding site for the quenching strand, and thus were in the macrocyclic configuration. The data were acquired on 06/01/2016.
CHAPTER FIVE

DISCUSSION

5.1 Future Work & Improvements

Now that the method has been demonstrated to function as projected, there are several unique and interesting studies that can be done. One could use this assay to measure the binding affinities of small GQ-stabilizing molecules or one could observe helicase activity in a stressed environment. Another path this project could take would be collecting data from different macrocycle radii and thus different ranges of applied force. This study could be expanded even further by substituting other non-canonical DNA structures in the sequence of interest on the bridge strand. The possibilities are endless because there are so many different sequences that could be used in the bridge strand. Ultimately, there is plenty of opportunity for expansion and growth of this project and thus opportunities for other students to become published.

Of course, there are improvements that could be made to this assay. Currently, the molecular density yield is low (due to the multi-step construct immobilization and formation) when compared to that of the more typical single-step immobilization partial duplex constructs our lab works with most of the time. That is not a statement of the lack of statistics of the macrocycle assay; it just requires a greater number of collected movies than the partial duplex assay to achieve the same number of single molecule traces that are used to generate the statistics. Another improvement that should be made is increasing the percentage of constructs that achieve the macrocyclic state, or in other words increasing the ratio of circular to linear configurations before the linear ones are eliminated with the quencher assay. And of course, the overall protocol is rather robust so it could be improved with time. I wish I could make these improvements myself but I will have to leave it up to the next student. Thank you for your time and consideration in reading this thesis.
APPENDIX I: SEQUENCES

A: Strand 1 of the “macroduplex”

Developed from pUC18 after PCR and gel purification which were performed in Basu Lab. This strand is 90 nt long and a biotin (dark blue lettering) is linked from the 5’ end. In this and the other sequences in this section, the subscript numbers after some bases designate the nt number from the end, i.e. G_{15} designates the 15\textsuperscript{th} nt in the sequence.

5’)/Biotin/ GCG GTA AGA TCC TTG_{15} AGA GTT TTC GCC CCG_{30} AAG AAC GTT TTC CAA_{45} TGA TGA GCC CTT TTA_{60} AAG TTC TGC TAT GTG_{75} GCG CGG TAT TAT CCC_{90} (3’)

B: Strand 2 of the “Macroduplex”

Developed from pUC18 after PCR and gel purification which were performed in Basu Lab. This strand is 120 nt long and a Cy3 is attached to the 5’ end. The first 15 nt on the 5’ end is called “Sticky End 1” (gold lettering) and the last 15 nt on the 3’ end is called “Sticky End 2” (purple lettering).

5’)/Cy3 GCC CGG CGT CAA TAC_{15} GGG ATA ATA CCG CGC_{30} CAC ATA GCA GAA CTT_{45} TAA AAG TGC TCA TCA_{60} TTG GAA AAC GTT CTT_{75} CGG GGC GAA AAC TCT_{90} CAA GGA TCT TAC CGC_{105} TGT TGA GAT CCA GCG_{120} (3’)


C: Bridge Strand 1

Purchased as HPLC purified from IDT. This strand consists of the complementary sequences of “Sticky End 2” in the first 15nt of the 5’ end (purple lettering), and then a two Thymine spacer (black lettering) followed by a 3Ly1Lp GQ forming sequence (light blue lettering). Then there is another two Thymine spacer (black lettering) followed by an internally labeled Cy5 (red lettering) and another single Thymine spacer (black lettering). The strand is finished in the last 15 nt with a sequence complementary to “Sticky End 1” (gold lettering).

5’)

GTA TTG ACG CCG GGC<sub>15</sub> TTG GGT GGG TGG GTG<sub>30</sub> GGT T/iCy5/T<sub>35</sub> CGC TGG ATC TCA ACA<sub>50</sub> (3’)

D: Bridge Strand 2

Purchased as HPLC purified from IDT. This strand consists of the complementary sequences of “Sticky End 2” in the first 15nt of the 5’ end (purple lettering), then a two Thymine spacer (black lettering) followed by a poly-Thymine sequence the length of the 3Ly1Lp GQ forming sequence (15 nt) (light blue lettering). Then there is another two Thymine spacer (black lettering) followed by an internally labeled Cy5 (red lettering) and another single Thymine spacer (black lettering). The strand is finished in the last 15 nt with a sequence complementary to “Sticky End 1” (gold lettering).

5’)

GTA TTG ACG CCG GGC<sub>15</sub> TTT TTT TTT TTT<sub>30</sub> TTT T/iCy5/T<sub>35</sub> CGC TGG ATC TCA ACA<sub>50</sub> (3’)

57
E: Bridge Strand 3

Purchased as HPLC purified from IDT. This strand consists of the complementary sequences of “Sticky End 2” in the first 15nt of the 5’ end (purple lettering), then a two Thymine spacer (black lettering) followed by a human telomeric GQ forming sequence (light blue lettering). Then there is a single Thymine spacer (black lettering) followed by an internally labeled Cy5 (red lettering) and another single Thymine spacer (black lettering). The strand is finished in the last 15 nt with a sequence complementary to “Sticky End 1” (gold lettering).

5’) GTA TTG ACG CCG GGC<sub>15</sub> TTG GGT TAG GGT TAG<sub>30</sub> GGT TAG GGT /iCy5/T<sub>40</sub> CGC TGG ATC TCA ACA<sub>55</sub> (3’)

F: pd(Sticky15) Stem Strand

Purchased from IDT as PAGE purified. This strand consists of a Cy3 attached to the 5’ end, the “Sticky End 1” sequence (gold lettering), and a biotin (dark blue lettering) linked to the 3’ end. It is essentially a truncated version of strand 2 of the “macroduplex” with an added biotin.

5‘) Cy3 GCC CGG CGT CAA TAC / Biotin / (3’)

G: Quencher Strand

Purchased form IDT. This strand consists of the “Sticky End 2” sequence (purple lettering) and an Iowa Black® RQ Cy5 quencher (black lettering) linked to the 3’end.

5‘) TGT TGA GAT CCA GCG /Quencher/ (3’)

58
APPENDIX II: MACROCYCLE CONFIGURATIONS

The following is a list of all the possible configurations the macroduplex strands and the bridge strand combination could attain along with a short description an accompanying cartoon. I also give a brief statement about the expected FRET signal for the given configuration. The PolyT sequence will yield the same configurations as that of an unfolded GQ.

A. Only Strand 1 bound to the surface due to improper annealing of the macroduplex. Invisible to our setup since no fluorophores are present.

B. Formation of the macroduplex by annealing of strands 1 and 2. Yields a donor-only signal when bound to the surface.

C. GQ forming bridge that contains a folded GQ and is bound to the macroduplex on the Cy3 end, however the macrocycle is not formed. Yields a high FRET state.
D. GQ forming bridge that contains a folded GQ and is bound to the macroduplex on the end opposite the Cy3 end; however the macrocycle is not formed. Yields a donor-only signal due to the fluorophore separation distance being over three times greater than the FRET range for the Cy3-Cy5 pair.

E. Two GQ forming bridges that contain a folded GQ and are bound to the macroduplex. The presence of the second bridge strand prevents the formation of the macrocycle. The Cy5 proximal to the Cy3 will yield a high FRET state since the GQ is folded. The Cy5 from the second strand does not contribute to a FRET signal since it is beyond the range for FRET to occur.

F. GQ forming bridge that contains an unfolded GQ and is bound to the macroduplex proximal to the Cy3, however the macrocycle is not formed.
G. GQ forming bridge that contains an unfolded GQ and is bound to the macroduplex on the end opposite the Cy3 end; however the macrocycle is not formed.

H. Two GQ forming bridges, both unfolded, bound to the macroduplex. The macrocycle cannot be formed. The unfolded GQ near the Cy3 yields a moderate FRET signal and the other bridge is beyond the FRET range.

I. Two GQ forming bridges bound to the macroduplex, the one closest to the Cy3 is unfolded yielding a moderate FRET and the other is folded but beyond the FRET range. The macrocyclic cannot be formed.

J. Two GQ forming bridges bound to the macroduplex, the one closest to the Cy3 is folded yielding a high FRET and the other unfolded bridge is beyond the FRET range. The macrocycle cannot be formed.
K. The macrocycle with a folded GQ, or rather the folded circular configuration. Yields a high FRET state.

L. The macrocycle with an unfolded GQ, the unfolded circular configuration. Yields a moderate FRET, but one that is lower than the moderate FRET achieved by the unfolded linear configuration [F] or the unfolded partial duplex [N]. This is due to the force the macrocycle exerts on the single stranded region. This force stretches the ssDNA thus increasing the interfluorophore distance and lowering the FRET.
M. The partial duplex control with a folded GQ. Yields a high FRET.

N. The partial duplex control with an unfolded GQ. Yields a moderate FRET.
REFERENCES


[58] Kruger, A. C., and Birkedal, V. Single molecule FRET data analysis procedures for FRET efficiency determination: probing the conformations of nucleic acid structures, Methods 64, (2013) 36-42.
[84] Le, T. and Kim, H. Jove, (2014) e51667