

G QUADRUPLEX STABILITY



FERNALDO R. WINNERDY

DEPARTMENT OF PHYSICS
NATIONAL UNIVERSITY OF SINGAPORE

2014

G QUADRUPLEX STABILITY



FERNALDO R. WINNERDY

*A thesis submitted in partial fulfilment for the degree of
Bachelor of Science with Honours in Physics*

Supervisor

Assoc Professor Yan Jie

Dr. You Huijuan

DEPARTMENT OF PHYSICS
NATIONAL UNIVERSITY OF SINGAPORE

2014

Abstract

G Quadruplex is one of many DNA structures that can be formed under certain condition and has some quite significant functions in genes expression regulation, DNA protection, and ligand binding. The study on G Quadruplex structure has been started since more than a decade ago, however, the researches were usually focused on the telomeric G4 sequence, whereas in this study, we try to focus more on ILPR G4 sequence. With a function of transcription regulation and a possibility of being the receptor of insulin located upstream of the insulin gene promoter, ILPR G4 is observed to have an average unfolding force of $F_{ave} = (34.3 \pm 0.9)pN$. The probability of finding the folded state is about 7 times higher than finding the unfolded state at zero pN. The results are hopefully can be used for further research in folding and unfolding kinetics of G quadruplex structure, and might be helpful in understanding the role of G quadruplex structure on regulation of gene transcription.

Acknowledgements

During my four years in NUS, the last year which I spent for this Final Year Project was the most challenging and breathtaking year for me. Although I have already stepped into a research life since my second year through a research internship with CQT, it is my first experience to plan, manage and be responsible of a research project by myself. It was a year full of busy days yet filled with enjoyment, full of busy nights yet filled with vigor, and now it was ended by this report which I am content and grateful of.

Regardless of how excellent or disappointing people might see this project is, I may not boast of my own self for this achievement. It is not by me, but by God, Jesus Christ who has strengthen me. For in Him all things were created. He is before all things, and in Him all things hold together. To Him be all the praises and glory forever.

In order for this project to reach this point, there is a long list of names should my thanks be dedicated to. First of all I would like to thank my great supervisor **Assoc. Prof. Yan Jie**, for all of his patient, ideas, inspirations and advices which made research felt so interesting and vibrant. Under his supervision, I was

given a freedom to think and challenge myself to ask and solve the right questions for research, yet wasn't leaving me unguided. To be honest I admire the way he run the research group. Secondly, I would like to thank my co-supervisor **Dr. You Huijuan**, for all her time, a whole lot of advices and guidance in this whole project. She was a great mentor which through discussions with her I was helped and inspired. This project will not be done without her.

Thanks also for my beloved parents, who have supported me and restlessly prayed for me regardless of my weakness and limitations. I might have disappointed them for not being a good caring son for them, yet they still love me whoever I am.

I would also like to thank **Mr. David Giovanni** for helping in writing this very report by giving me a tutorial to writing in LaTeX. He was my discussion friend for the past year especially in the mathematical part. Thank you for the fun discussions and enjoyable moments solving equations as well as the spirit boosting moments through the game we played together.

My thanks also addressed to the rest of MBI-Lab members in Yan Jie's group whom I did not mention in name one by one, for the fun and vibrant research environment which they created. It has carved a great impression on me. Because of them I decided to enter deeper into research world.

Thanks to the Indonesian Division of Higher Education for financing my whole study in NUS.

Lastly I would like to thanks all of my friends and mentors who have cared about me, prayed for me, cheered me, supported me and listened to my sharing during my difficult days in this project. Thanks to **Theo Arianto, Theo Ariandi, Kelvin Timotius, Manoel Yohannes, Devina Wongso, Christabella Adine, Vonny Caroline**, etc. which I could not mentioned all of them by name here. I am grateful for having such good friends like them, for the unending supports that they give.

Contents

Abstract	ii
Acknowledgements	iii
List of Tables	viii
List of Figures	ix
1 Introduction	1
1.1 Motivations for Research	1
1.2 Objectives of the Work	3
1.3 Scope of the Work	4
2 Theoretical Background	6
2.1 The G4 Genome	6
2.1.1 G4 Motifs in DNA Replication	7
2.1.2 G4 Motifs in Regulated Recombination	10
2.1.3 G4 Motifs in Genes and Transcriptional sites	11

2.1.4	G4 Motifs Expansion in Neurological Diseases	13
2.2	Magnetic Tweezers	13
2.2.1	Basic Principle of Magnetic Tweezers	14
2.2.2	Force Measurement on Short Tether	19
2.2.3	Loading Rate and DNA Unfolding Force	24
2.3	Reaction Kinetics	26
2.3.1	Two-State Reaction Kinetics	26
3	Experimental Methodology	28
3.1	DNA Preparation	28
3.1.1	PCR on λ DNA	29
3.1.2	Restriction Process	30
3.1.3	ILPR G4 Ligation	32
3.2	Channel Preparation	32
3.2.1	APTES Coated Channel	33
3.2.2	Addition of DNA Tether	35
3.3	Magnetic Tweezers Operation	39
3.3.1	Magnetic Tweezers Set up	40
3.3.2	Modes of Magnet Movement	42
4	Results and Discussion	44
4.1	Preliminary Data	44
4.1.1	Overstretching Transition	44
4.1.2	Typical G4 Unfolding Signal	46
4.2	Data Analysis	47
4.2.1	Force Distribution	48

4.2.2	Refolding Probability	49
4.2.3	Data Fitting	51
5	Conclusion and Future Works	54
5.1	Conclusion	54
5.2	Future Works	55
5.2.1	Completion of Data	55
5.2.2	Free energy difference	56
	Bibliography	57

List of Tables

2.1	Neurological diseases and the corresponding G4 motif repeats	14
3.1	Standard z -distance points used in our laboratory for magnetic tweezers experiment	41
4.1	The velocity constants obtained after fitting	52

List of Figures

1.1	Illustration of G quartet (left), a planar array of guanines stabilized by Hoogsteen base pairing, and G4 DNA (right).	2
2.1	Illustration of telomeric G4 structure protecting the DNA from degradation by exonuclease	8
2.2	Illustration of a normal DNA replication, consisting of leading and lagging strand, with the direction of DNA polymerases and helicase as shown.	9
2.3	The distribution of genomwide average of G4 motifs plotted to standard reference points (TSS, UTRs, Exons, and Introns).	11
2.4	Illustration of the possibility of inhibition and facilitation of DNA transcription by G4 structure.	12
2.5	Schematic view of the setup of magnetic tweezers (not to scale).	15
2.6	Illustration of magnetic field lines, density is proportional to the magnetic field strength (the lines going outward are not displayed for clarity).	15
2.7	Illustration of Brownian motion and the calculation of magnetic force	17
2.8	Illustration of diffraction rings produced in the extension measurement	19

2.9	The force-distance curve of the magnetic tweezers. A) The data obtained using 48,502 bp λ DNA. B) The data is shifted in the y axis to overlap each other.	21
2.10	Illustration of the difference in length of B DNA and S DNA	22
2.11	The force extension curve of dsDNA B to S transition.	23
2.12	A) Force histograms of single biotin-streptavidin bonds with different loading rates. B) Dynamic strength spectra for biotin-streptavidin(circles) and biotin-avidin(triangles)	25
3.1	Schematic illustration of the DNA sample, handles on the left and right, and the ILPR G4 motif ssDNA in the middle	29
3.2	Gel electrophoresis result for, A. PCR product, B. Restriction product, C. Ligation product	30
3.3	A. Picture of the centrifuge machine and buffers used for <i>invitrogen</i> purification process. B. Picture of <i>nanodrop</i> device that detects absorbance and calculates DNA concentration	31
3.4	Views of the glass surfaces inside the jar, and the zig zag pattern of the glass shown in red lines	33
3.5	A. The <i>Boekel Scientific</i> Oven. B. The <i>FEMTO Science</i> plasma cleaner	35
3.6	Illustration of APTES attachment to glass surface and the free amine groups that are exposed after the attachment	35
3.7	Pictures and illustrations of the APTES glass covered by clean coverslip, attached by using parafilm and silicon gel.	36
3.8	Illustration of Sulfo-SMCC as a cross linker between the amine group of APTES glass and the thiol group of the DNA tether	38
3.9	Illustration of the completed channel ready for observation under magnetic tweezers	39

3.10	Screen shot of picture of beads from CCD camera attached to the microscope	41
4.1	Preliminary data on DNA overstretching transition	45
4.2	Preliminary data on G quadruplex unfolding signal under constant loading rate	46
4.3	Single G quadruplex unfolding signal	47
4.4	Output of the <i>MATLAB</i> code that detects the jumps of signal, the highlighted data is chosen by manually observing the data	48
4.5	The unfolding force distribution histogram, each bin has a width of 1 pN, the black curve is the Gaussian fit	49
4.6	The summary data of the Gaussian fit, with the constraints of $y_0 = 0$. .	50
4.7	The data points and the exponential fittings of the refolding probability plotted against holding time, each color represents different holding force	51
4.8	Summarized result of the fittings using the exponential equation, y_0 , A , and R_0 with their respective standard errors	52
4.9	Rough exponential fit of the folding and unfolding constant, more data should be obtained to make the graph valid	53

Introduction

1.1 Motivations for Research

G4 or G quadruplex structure is one of the most important secondary structure of DNA beside the usual double-helix structure inside human body. It has diverse structures which contrast the highly predictable B-form duplex. The formation of the four-stranded (quadruplex) structure requires a certain motif of DNA sequence, i.e. $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$, with N symbolizes any base (A, T, G, or C) [1]. The interactions among the four strands are stabilized by G-quartets [2] forming a layer-like structure [3] illustrated in Figure 1.1. Hoogsteen base pairing is formed between the N7 group of one guanine and extracyclic amino group of its neighbor, with the central channel occupied by a monovalent cation (potassium is shown). Other than that, the hydrogen bonds between guanines and the stacking of the hydrophobic G-quartets planar stabilized the G4 structure [1].

There are numerous G4 motifs in human genome, they are especially abundant in the telomeres, rDNA, immunoglobulin switch region, and some variable number tandem repeats (VNTRs). It is unknown whether each G4 motif forms a

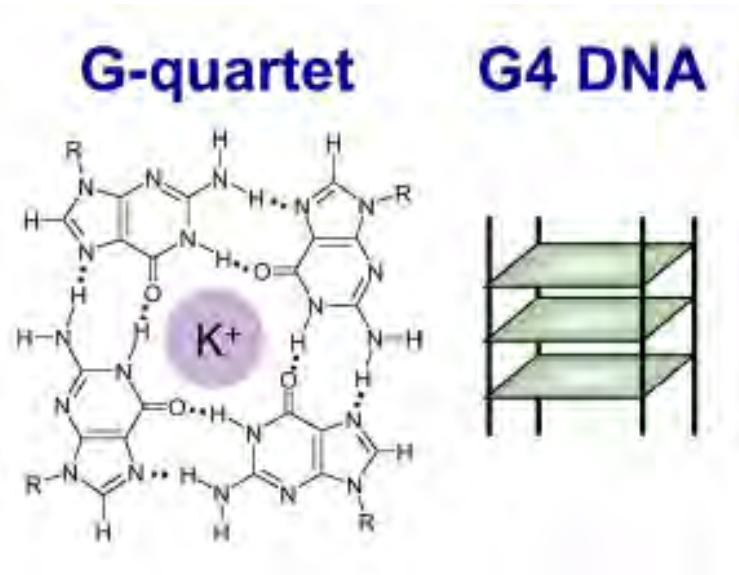


Figure 1.1: Illustration of G quartet (left), a planar array of guanines stabilized by Hoogsteen base pairing, and G4 DNA (right).

quadruplex structure inside the cell, but it provides a potential to form the diverse structures which may correlate to a specific function nonetheless [1]. The G4 motif which is rated by many groups as the most important is the telomeric G4, which consists of thousands of kilobases of duplex DNA repeats ‘TTAGGG’ and a single stranded G-rich 3'-tail. It is known that the formation of G4 structure in the telomeric region inhibits the activity of telomerase enzyme inside the cell, which suggests that G quadruplex structure is a molecular target of telomerase inhibitors [4]. Having said that, since the lack of telomerase activity would result in the shortening of human telomere (which is very much related to cancer and ageing), there are many researches focusing on designing a G4-targeting drug [5].

In this study, however, the focus is on the G4 motif located upstream of the promoter of the insulin gene, which is called Insulin-linked Polymorphism Region (ILPR) G4. In this promoter region, the ILPR has been implicated as a factor of insulin-dependent diabetes mellitus [6]. The region consists of a variable number tandemly repeated variants of the sequence 5'-ACAGGGGTGTGGGG-3', which

is a G4 motif. The variations of sequence and number of repeats of this VNTRs would correlates to the ability of this motif to form G quadruplex structure, which affect the transcriptional activity of the insulin gene [7]. The transcription of this insulin gene is very important to the function of human body especially in the regulation of carbohydrate and fat metabolism [8]. Therefore, it is important to study the kinetics of the formation of the ILPR G4 structure, and the stability of the structure once it has been formed. Those informations are required to find a way to control the G4 formation, which might enable us to induce or inhibit the transcriptional activity of insulin gene. To our hope, this study might provide an insightful perspective to this rarely studied ILPR G quadruplex, which is almost equally important to the telomeric G quadruplex.

The single molecule manipulation method with a magnetic tweezers is used throughout the project. It has many unique advantages in terms of specificity, throughput, and force stability. Our lab had improved the limitation of the normal magnetic tweezers which allow the magnetic tweezers to measure force accurately in short tethers [9]. With the help of this improved high-force magnetic tweezers, we could study the stretching, unfolding, and refolding of short DNA which is required in our proposed study.

1.2 Objectives of the Work

Having such motivation for the research, this thesis contains two main objectives, which are presented as follows:

1. To study the kinetics of the formation of G quadruplex structure. The quantitative results would be in the form of kinetic constants, both $k_{folding}$ and $k_{unfolding}$, which represent the speed of formation. Those results could be used to compute the equilibrium constant between the folded and unfolded state,

and hence the free energy difference between them.

2. To observe the distribution of the force required to unfold the ILPR G4 structure. Previous study by other group (Hanbin Mao) stated that there are two peaks of the force distribution which leads to two different folded states of ILPR G4. We would like to assure this result, since the existence of two peaks is not the case for telomeric G4 structure.

1.3 Scope of the Work

The scope of this study will be limited to observing the probability of the refolding of G4 structure after unfolding it with magnetic tweezers. This probability will depend on the holding time. The probabilities are then used to compute the kinetic constant of the formation of G4 by using a fitting to an exponential function. In addition, in each cycle of DNA pulling, the unfolding force of the DNA was recorded to produce the distribution function.

This thesis is organized as follows:

In Chapter 2, basic theoretical background of the principles and operation of magnetic tweezers will be included, as well as some basic chemistry on reaction kinetics. The purpose of this chapter is to help the reader with limited biochemistry knowledge to understand the essential theoretical framework needed for further discussion in this study.

In Chapter 3, the DNA sample preparation is explained in sufficient detail. The discussion includes not only the technical aspect and procedures of the sample preparation, but also the theory behind them. The sequence of the magnet movements and the reasoning behind them are also explained in this chapter. The purpose of this chapter is to help the reader to appreciate and understand the

experimental technique and instrumentation used in this study.

In Chapter 4, the result and discussion would be presented. It consists of the unfolding force distribution of ILPR G4, the graphs of the probability of refolding as a function of holding time and force, the extrapolation of kinetic constants to zero force, and the equilibrium constant between the folded and unfolded state. The discussion of these results is presented following the subsequent result.

In Chapter 5, the conclusion of the study and further possible studies for more comprehensive result are briefly proposed and discussed.

Theoretical Background

This chapter contains brief introduction on both essential theoretical framework needed and experimental knowledge on the main device used. Comprehension of this chapter will help the reader to understand and appreciate better the meaning and significance of this study.

2.1 The G4 Genome

This section is meant to urge the reader to look at the bigger picture of this study about ILPR G4 sequence. An overall summary of known G4 motifs and their importances are summarized in this section. While the formation of the quadruplex structure is still uncertain, the G4 motif itself has provide a considerable potential to correlate to a specific function. The following are where the G4 motif have well known implications, in DNA replication, in regulated recombination, in genes and transcripts, as well as some neurological diseases relating to G4 sequences.

The existence of G4 structure *in vivo* is proven via the ciliated protozoa *Stylonychia*. The existence as well as the deformation of G4 structure is studied very well in this organism especially in the telomeric region [10]. This proof, as well as the

fact that G4 motifs are found in mostly in functional region of genomes, encourage researchers to study the G4 motifs and structures *in vitro* for the potential regulation function of G4 structure *in vivo*.

2.1.1 G4 Motifs in DNA Replication

This subsection is divided into three points, G4 at DNA replication origins, G4 role in telomere replication, and replicative instability at G4 motif.

G4 at DNA replication origins

DNA replication begins at points called origins of replications, where the enzymes and molecules required to do the replication bind to the specific DNA sequence called promoter. Recent studies have revealed, using high-throughput sequencing, that more than half of human DNA replication origins correspond to G4 motifs (67% out of 250,000 origins) [11]. It would be interesting if G4 motifs are recognized by the origin recognizing complex and if different G4 motifs produce different property in the activation DNA replication, although direct proof for this is not yet available.

G4 role in telomere replication

In human DNA, since it is a linear DNA, there exist the end of DNA chromosome called telomere which often has a repeating sequences. Due to the nature of DNA polymerase, human telomere cannot be replicated normally, therefore, there is an enzyme called telomerase who do the work of replicating the telomeric end of the DNA by doing a reverse transcription from its RNA. Human telomerase is inactive in most of human somatic cells, but it is unregulated in cancer cells [12].

G quadruplex structures can influence telomerase activity: intramolecular anti

parallel G4 structure blocked it, while intermolecular parallel G4 structure does not [13–15]. This fact initiated further tests on ligands that bind to G4 structure to look for a specific ligand that stabilize the G4 structure [16].

On the other hand, G4 structure can also protect telomeres. The activity of exonuclease is often misinterpreting the telomeric repeats as damaged DNA sequences, and therefore will try to digest the DNA. To counter that, in a normal condition, there is a telomere capping complex that will stabilize the telomeres and preventing the degradation of those telomeric repeats [17–19]. However, in a condition where the telomere capping complex is depleted, the telomere instability can be countered by G4 stabilizing drugs [20]. This suggests that the G4 DNA structure is very resistant to digestion by exonuclease, meaning that G4 structure could protect the cap-deprived telomere. Figure 2.1 shows an illustration of this G4 property [21].

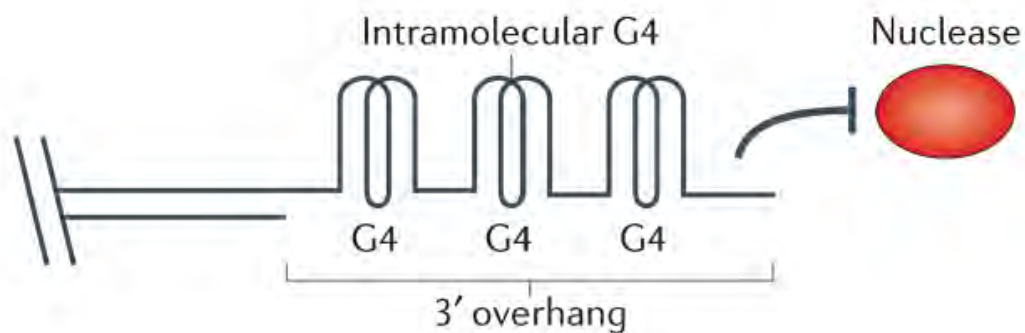


Figure 2.1: Illustration of telomeric G4 structure protecting the DNA from degradation by exonuclease

Replicative instability at G4 motifs

In DNA replication, the two strands of DNA are separated by an enzyme called helicase, both of the strands serve as templates to synthesize the new DNA (semi-conservative property of DNA replication). However, since the DNA polymerase

can only move in one direction during replication, one of the strand is called leading strand, while the other is called lagging strand, with the illustration is shown in Figure 2.2 [22]. For the leading strand, the opening of the helix, the replication, and then the re-closing of helix can be done in a short amount of time, since the DNA polymerase can move directly after the helicase and immediately copy the DNA. However, on the lagging strand, the helicase must open the helix first to create enough space for the DNA polymerase to attach and begin replicating in the opposite direction. This constraint allows enough time for G4 structure to fold while the DNA is in single strand condition without the presence of DNA polymerase yet [21]. To complete the replication, the G4 structure must be resolved since the structure will not allow the DNA polymerase to replicate the DNA, and the enzyme to unfold this G4 structures are another form of helicases called G4 helicase.

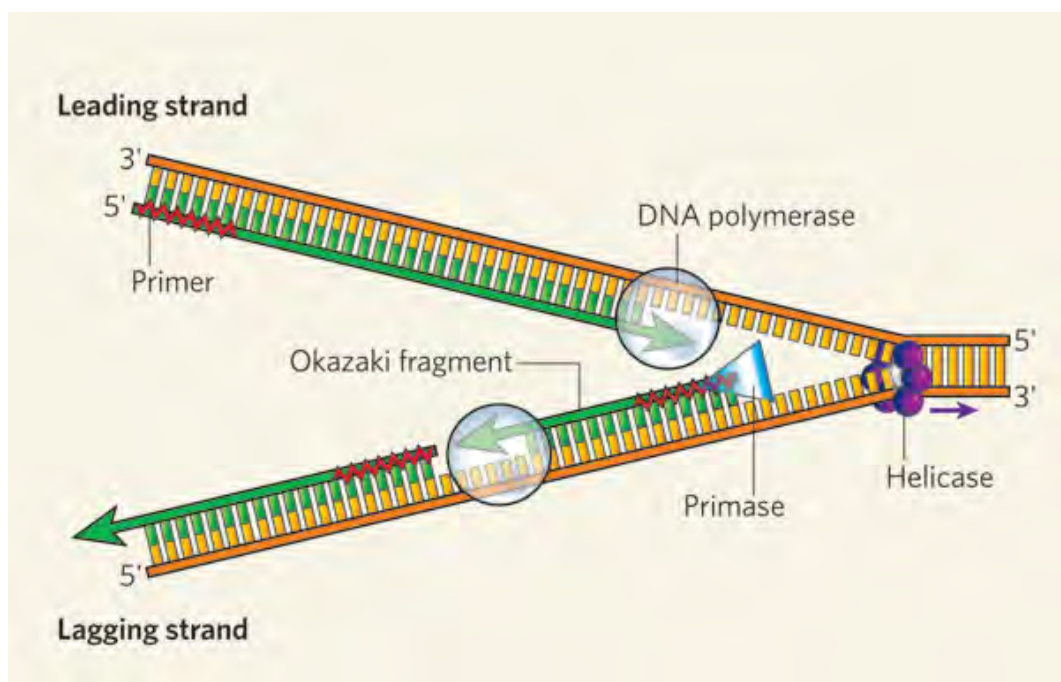


Figure 2.2: Illustration of a normal DNA replication, consisting of leading and lagging strand, with the direction of DNA polymerases and helicase as shown.

G4 helicases must recognize and unwind the G4 structure to maintain the stability of the genes and DNA. Many families of G4 helicases, for instance the RecQ family of G4 helicases called WRN [23] and BLM [24], are having an important role in resolving telomeric quadruplexes. Deficiency in WRN helicase occurs in the case of Werner Syndrome, with the most notable implication is the premature ageing due to depletion of telomeric sequence. While the deficiency in BLM helicase is causing the Bloom Syndrome, an immunodeficiency syndrome due to an impaired recombination at G4-rich immunoglobulin S regions [1].

2.1.2 G4 Motifs in Regulated Recombination

To date, one of the best evidence in the functional role of G4 DNA, and the most detailed examination of the biological function of a specific G4 DNA structure is the study of the pathogenic mechanism of *Neisseria gonorrhoeae* [1]. This particular pathogen uses DNA recombination which provides antigenic variation to escape the host immune surveillance. *N. gonorrhoeae* encodes many pilin genes, which produce hair-like projections called pili on the bacterium surface. These pili act as a surface antigen of this pathogen which should be recognized by the immune system of the host, but the pathogen uses a G4 based system to regulate the expression of the genes.

The gene in *pilE* locus is the only expressed genes that regulates the pili, and it switches by a recombination mechanism with an unexpressed genes from *pilS* [25]. The region upstream of the *pilE* locus contains the 12 bp G-rich element ($G_3TG_3TTG_3TG_3$) [26], which has to form G4 structure, to promote antigenic variation. To prove that G4 structure is required, a further study using a G4 ligand *N-methyl* mesoporphyrin IX affects *pilE* gene conversion events [21].

2.1.3 G4 Motifs in Genes and Transcriptional sites

The distribution of G4 motifs has been characterized in human RefSeq genes (genomic sequences used as reference standards for well-characterized genes [27]). It is summarized in Figure 2.3 [1] below, with the G4 motifs-enriched regions located at the TSS (Transcriptional start site), the 5'-UTR (Untranslated region), and the 5' end of the first intron. As for the coding region, the G4 motif frequency is relatively low.

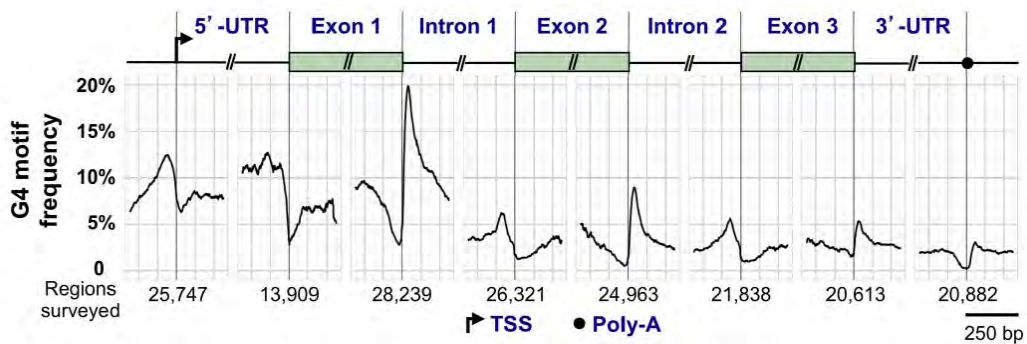


Figure 2.3: The distribution of genomwide average of G4 motifs plotted to standard reference points (TSS, UTRs, Exons, and Introns).

This distribution leads to further studies which results in the fact that one or more G4 motifs are found within 1000 nucleotides upstream of the TSS of 50% of human genes [28]. Moreover, it has been shown by bioinformatics that there are a lot more G4 motifs in the promoters of human oncogenes (gene that has a potential to cause cancer) and regulatory genes than in the promoters of housekeeping and tumour-suppressor genes [28,29]. All these facts suggest a potential function of G4 structure in gene regulation which is related to one of the aims of this project.

Recent study has shown that G4 structures could formed because of the supercoiling-induced stress during transcription [30], and *in vitro* studies show that the structure can compensate for negative supercoiling [31]. These findings result in a possibility that G4 structure can influence the transcription process in both positive and

negative manner. The first illustration in Figure 2.4 [21] shows an inhibition of transcriptions caused by the blocking of polymerase by the G4 structure, while the second illustration shows a facilitation of transcription since the G4 structure formed in the non-template strand induced the forming of single stranded DNA on the template strand.

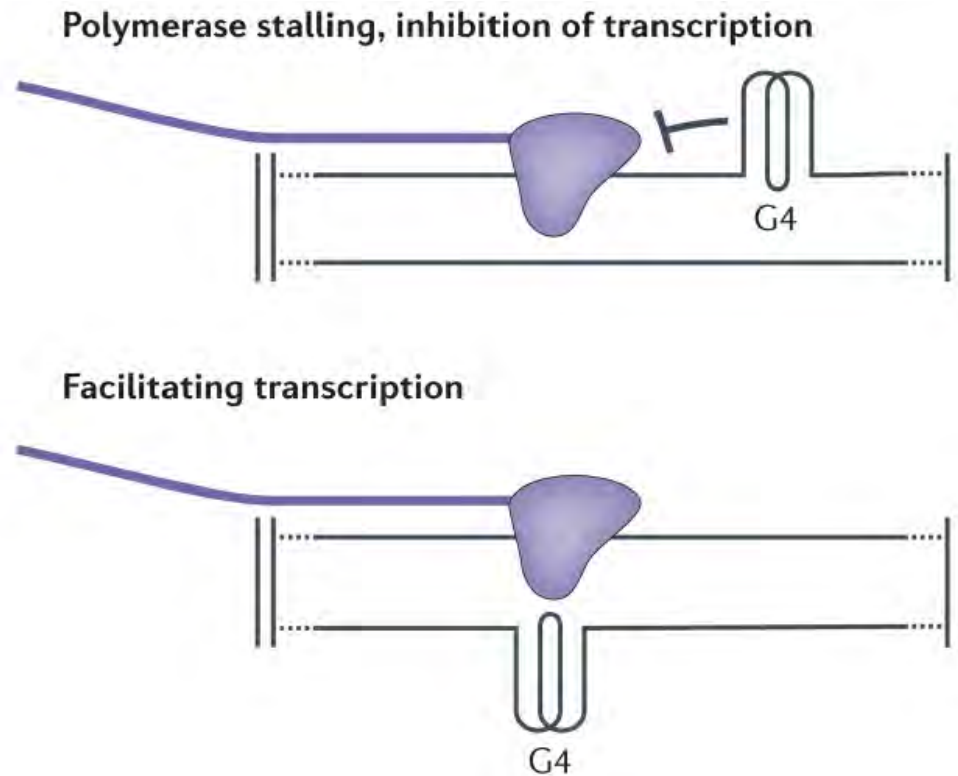


Figure 2.4: Illustration of the possibility of inhibition and facilitation of DNA transcription by G4 structure.

Another possible way to influence the transcription is by a ligand binding to the G quadruplex structure. G4 binding proteins might affect the formation or unfolding of G4 structure which further leads to an inhibition or enhancing of the gene transcription process. One example of *in vitro* study about this is the myosin D (MYoD) family proteins. It is a transcription factor that normally binds to E-boxes in the promoter region, but in the presence of G4 structure, it will bind to it and enhance gene transcription [32, 33]. This study is related to this since we

study the G4 motifs in insulin-linked polymorphic region which is located near the promoter of the insulin gene in human DNA. Having said all the above points, the expectation is to convince the reader that the formation and unfolding mechanics of the G4 structure is important in gene regulation especially in the regulation of insulin gene.

2.1.4 G4 Motifs Expansion in Neurological Diseases

There are some neurological diseases that correspond to the expansions of G4 motifs in several different genes. Impaired mRNA processing caused by the titration of essential RNA binding proteins is one of the abnormalities that occur because of the extended regions of quadruplex structures in the pre-mRNA [1]. The expansions responsible for this is the G4 motif expansions in FMR1 [34], c8orf72 [35–37], and NOP56 [38]. Table 2.1 summarizes the known neurological diseases caused by G4 motif expansions.

Even though the consequences of the G4 motif expansions have already known, the mechanics of the expansion of these G4 motifs however, are yet to be defined.

2.2 Magnetic Tweezers

Single molecule manipulation method was developed and used in studying the force spectra of DNA and protein for more than 10 years. Three of the known techniques or devices are the optical tweezers, magnetic tweezers, and atomic force microscope (AFM). Among these three, magnetic tweezers has many advantages, e.g. it has a higher specificity if compared to AFM, requires no heating as optical tweezers does, and has a high throughput and force stability compared to the other two [9]. In this section, the main machinery in my project, the magnetic tweezers, will be explained in detailed manner. This section will be divided into three parts, the

Repeat	Gene	Disease
CGG	FMR1, 5'-UTR	Fragile X syndrome, fragile X-associated tremor/ataxia syndrome
GGGGCC	C9orf72, intron 1	Amyotrophic lateral sclerosis (ALS4), autosomal dominant frontotemporal dementia
GGCCT	NOP56, intron 1	Spinocerebellar ataxia (SCA36)
CGCGGGGCGGG	CSTB, promoter	Progressive myoclonus epilepsy type 1
CCCCATGGTGGT-GGCTGGGGACAG	PRNP, coding exon	Creutzfeldt-Jakob disease

doi:10.1371/journal.pgen.1003468.t001

Table 2.1: Neurological diseases and the corresponding G4 motif repeats

basic principle of magnetic tweezers, the magnetic tweezers of our laboratory, and lastly the importance of constant loading rate in our experiment.

2.2.1 Basic Principle of Magnetic Tweezers

General magnetic tweezers set up is shown in the Figure 2.5 [39] below, it consists of a pair of permanent magnets, the sample which contain the DNA tether and the paramagnetic bead, and the objective lens, all connected to a computer for the resulting output.

The Production of Force

The magnet in the magnetic tweezers produces a gradient of magnetic field as pictured in the Figure 2.6 above, where the magnetic field lines near the magnet is

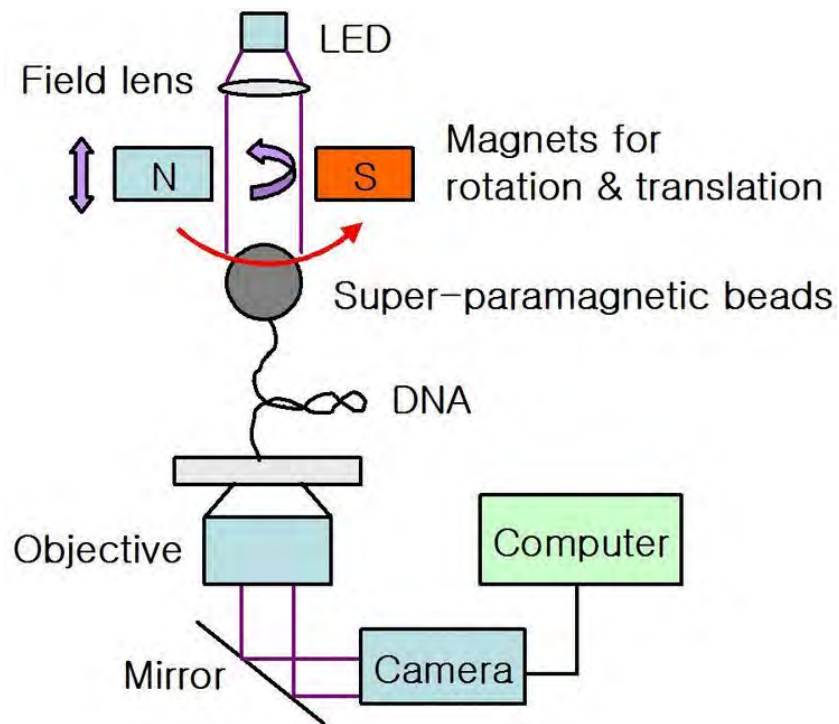


Figure 2.5: Schematic view of the setup of magnetic tweezers (not to scale).

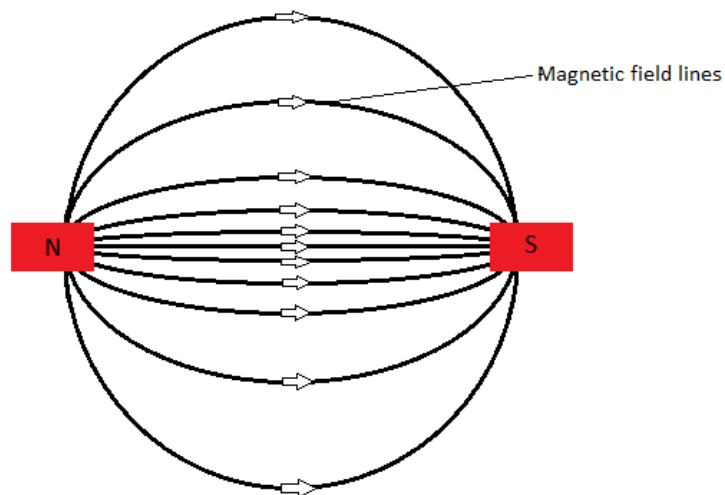


Figure 2.6: Illustration of magnetic field lines, density is proportional to the magnetic field strength (the lines going outward are not displayed for clarity).

denser than the ones further away from the magnet. Due to the magnetic field, induction happens inside the super-paramagnetic (paramagnetic material with extremely high magnetic susceptibility value) bead in the sample. The magnetic

moment produced due to the induction is given by:

$$\vec{m}(\vec{B}) = \frac{V\chi\vec{B}}{\mu_0} \quad (2.1)$$

where V is the volume of the bead, χ is the magnetic susceptibility, μ_0 is the vacuum permeability, and \vec{B} is the magnetic field vector. However, in the case of magnetic tweezers, the distance between the magnet and the sample to produce the desired force of 1-100 pN is in the range of 0.3-5.5 mm. In this case, the magnetic field generated by the magnet is estimated to be >0.2 T, which for a Dynal M-280 bead, this value is will cause the magnetic moment inside the bead to be saturated (maximum value of magnetic moment inside the bead has been reached, therefore is independent of the magnetic field strength) [40–42].

The magnetic potential energy formed due to the magnetic field and the magnetic moment is given by the following equation,

$$U = -\frac{1}{2}\vec{m}_{sat} \cdot \vec{B} \quad (2.2)$$

which will produce the force proportional to the gradient of the magnetic field, therefore it is a pulling force (gradient of the magnetic field is positive to the direction of the magnet),

$$\begin{aligned} \vec{F} &= -\vec{\nabla}U \\ &= \frac{1}{2}\vec{\nabla}(\vec{m}_{sat} \cdot \vec{B}) \\ &= \frac{1}{2}m_{sat} \frac{dB}{dz} \hat{z} \end{aligned} \quad (2.3)$$

with z-direction is going upward towards the magnet. This force is the force responsible for the pulling of the bead-tethered molecule.

Force Measurement

Next thing to do is to measure the force created by the gradient of magnetic field. This measurement can be done by observing the Brownian motion in the y -direction as illustrated in the Figure 2.7 [43].

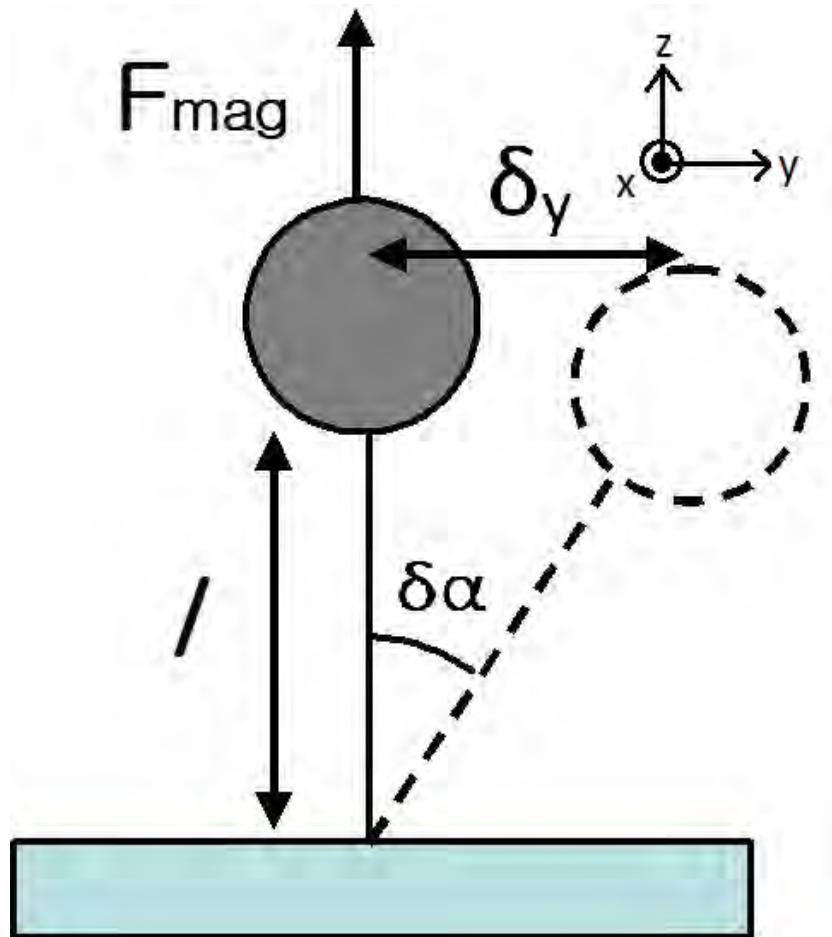


Figure 2.7: Illustration of Brownian motion and the calculation of magnetic force

Note that we cannot use the x -direction as a free Brownian motion, because a movement in x direction will misalign the magnetic moment, which produces torque that tries to realigned the magnetic moment to the magnetic field. The torque produced,

$$\vec{\tau} = \vec{m}_{sat} \times \vec{B} \quad (2.4)$$

By the equipartition theorem, the Brownian motion in the y-direction can be written as,

$$\frac{1}{2}k \langle \delta_y^2 \rangle = \frac{1}{2}k_B T \quad (2.5)$$

where k is the "spring" constant of the tethered DNA, $\langle \delta_y^2 \rangle$ is the time average of the square of y-movement of the bead, k_B is the Boltzmann constant, and T is the temperature.

The force in y-direction of the bead is treated as a spring force,

$$F_y = -k\delta_y \quad (2.6)$$

Looking at the tangent value of the angle $\delta\alpha$ in Figure 2.7, we have another equation,

$$\tan \delta\alpha = \frac{\delta_y}{l} = \frac{F_y}{F_{mag}} \quad (2.7)$$

Therefore, using equations 2.5 and 2.6 to eliminate k and F_y , we can get the magnetic force,

$$F_{mag} = \frac{k_B T l}{\langle \delta_y^2 \rangle} \quad (2.8)$$

Now, we can precisely measure the force in real time, if we can acquire the value of the temperature T , the extension l (for a short tether, the radius of the bead is included in the extension), and the variance of the Brownian motion $\langle \delta_y^2 \rangle$.

Extension Measurement

The measurement of temperature is trivial, and the measurement of the variance of Brownian motion $\langle \delta_y^2 \rangle$ can be done with the help of the image data through the camera and computer. However, this measurement has its own limitation, which will be presented later when I describe the short tether limitation of the magnetic tweezers.

The measurement of the extension of DNA tether was done by doing a calibration around the focal point of the plane of the bead. The bead positioned out of focus makes a diffraction ring around the image as shown in Figure 2.8. At each given time, the current bead image is compared to the calibration image which was constructed previously using the attached bead [44]. The observation of the diffraction rings will tell us the length of extension of the DNA tether in real-time.

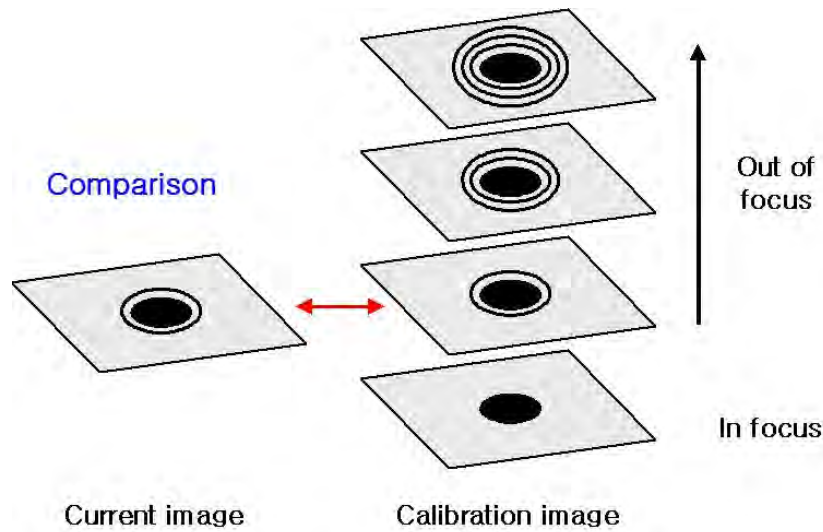


Figure 2.8: Illustration of diffraction rings produced in the extension measurement

2.2.2 Force Measurement on Short Tether

As mentioned in the above paragraph, the measurement of the variance of the y-axis Brownian motion have a limitation. For an accurate measurement to be done,

the sampling rate, f must be substantially faster than a certain frequency called Lorentzian corner frequency [9]. The formula for the frequency is given by,

$$f_c = \frac{F}{2\pi z\gamma} \quad (2.9)$$

where F is the magnetic force exerted, z is the extension of DNA tether, and γ is the drag coefficient of the bead. In a generic magnetic tweezers, the sampling rate of image capturing is around 60-100 Hz, for a long tether $10 \mu m$, the force can be measured up to 100 pN using a $3 \mu m$ bead. However, if the DNA tether is short ($<100 \text{ nm}$), the maximum force that can be calculated is only 1 pN [9].

To be able to measure the exerted magnetic force on a short tether, people must overcome this limitation. One way to do it is presented below, first we need to produce a force - distance graph for long tethers, and then we need to use the overstretching condition of DNA as a calibration standard to measure the actual force [9].

Force - distance calibration curve

For long tethers of DNA, we could produce a force - distance graph since we can directly measure the magnetic force for every magnet position. These force - distance curves are similar from one another, but there is a difference in their vertical positions if the curve is plotted on the logarithmic scale as illustrated in Figure 2.9 A [9]. To see clearer, Figure 2.9 B shows that the curve overlaps each other when the vertical axis of the curves are shifted.

This difference leads to the understanding that the equation for magnetic force as a function of magnet distance from the sample is an exponential function, with a constant in front, which differs for different beads used. Furthermore, from Figure 2.9 B, we can see that for a high force region (small distance), the exponential

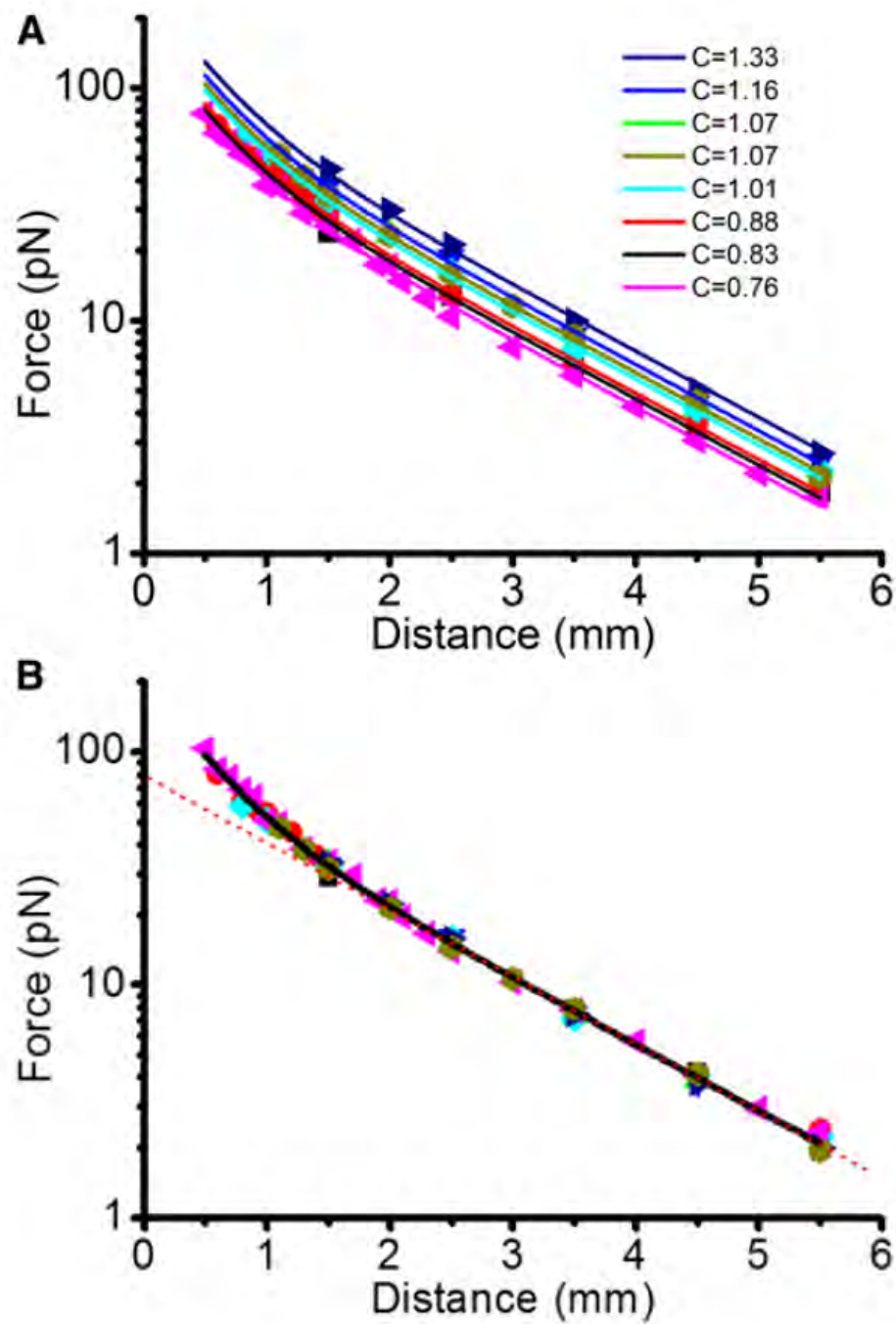


Figure 2.9: The force-distance curve of the magnetic tweezers. A) The data obtained using 48,502 bp λ DNA. B) The data is shifted in the y axis to overlap each other.

function is not as accurate. Therefore, the used of a double-exponential empirical formula to fit the curve is appropriate. In our lab, the final fitting equation for the

force-distance curve is the following,

$$F = C \left(\alpha_1 \exp \left(\frac{-d}{\gamma_1} \right) + \alpha_2 \exp \left(\frac{-d}{\gamma_2} \right) \right) \quad (2.10)$$

where α_1 , α_2 , γ_1 , and γ_2 are empirical constants to fit the curve, and C is the constant that differs for every bead used [9]. For the magnetic tweezers in our lab, using the average of a number of fitting, the values of those four constants are already known to be $\alpha_1 = 76$, $\alpha_2 = 140$, $\gamma_1 = 1.53$, and $\gamma_2 = 0.42$ [9].

The only problem remaining is the determination of C , which will be presented in the next subsection.

DNA Overstretching Condition

One of the nice property of DNA that can be used in order to obtain the C value is called the overstretching condition. This overstretching transition occurs in a well known force of ~ 65 pN [45], with two distinct transition pathways. A slow, hysteretic pull leads to a strand separation [46–48], while a fast, non-hysteretic pull leads to an elongated dsDNA structure referred to as S-DNA. The illustration of B to S DNA transition is in Figure 2.10 [49]. It was reported that pulling any ends of the DNA, either 3'3', 5'5', or 3'5' does the same thing, and the overstretching transition all occur in 65 pN [45, 50].

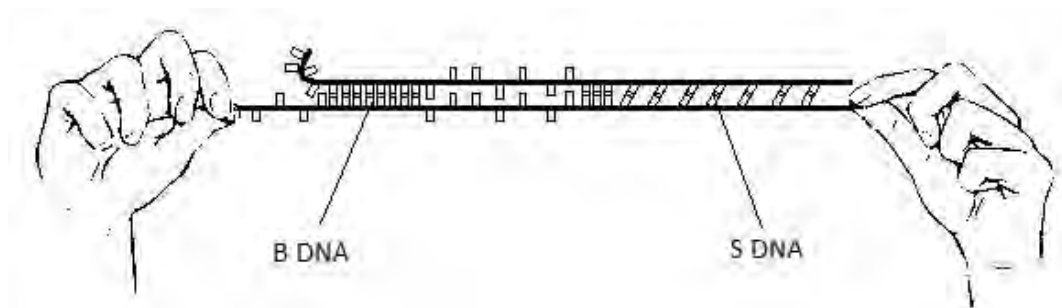


Figure 2.10: Illustration of the difference in length of B DNA and S DNA

Recent research in our laboratory provides the demonstration of DNA B to S transition using a GC-rich dsDNA ($\sim 60\%$ GC) of 586 bp tethered to the glass surface and a streptavidin coated paramagnetic beads under magnetic tweezers. The buffer solution used in the sample is 150 mM NaCl, pH 7.5, at 24°C , the result of the demonstration is in Figure 2.11 [9]. Three samples of dsDNA were used, square data points are forward transition, and circle data points are backward transition, the overlapped forward and backward transitions show that the transition is non-hysteretic and indicate the B to S transition.

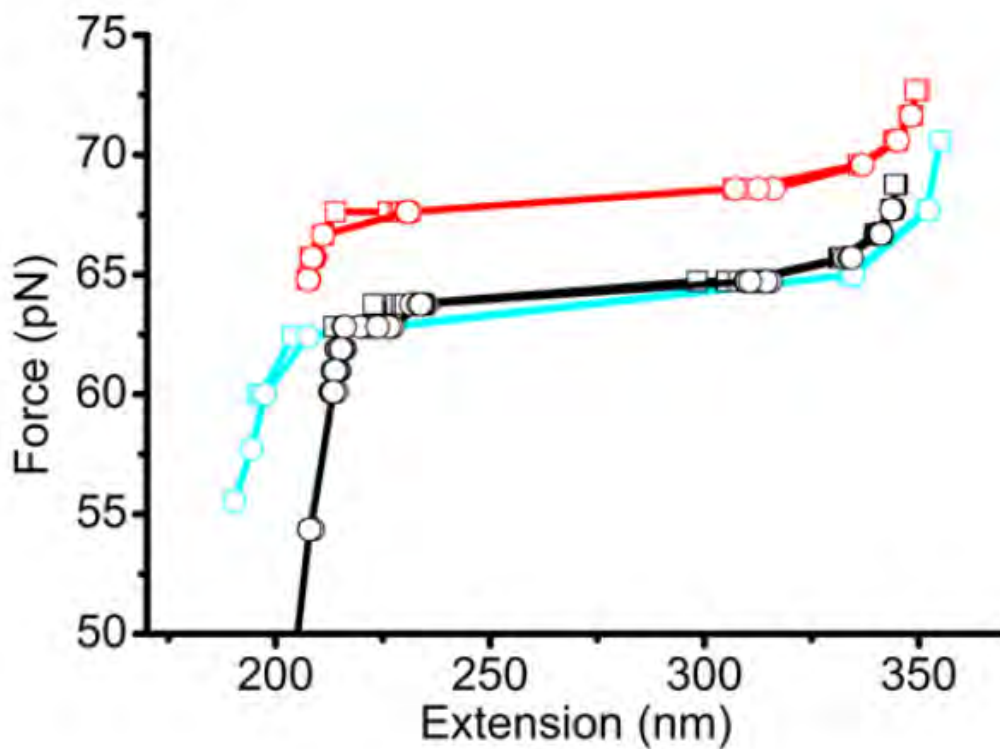


Figure 2.11: The force extension curve of dsDNA B to S transition.

Now that we have the overstretching transition at a determined force of ~ 65 pN, we can use this as a benchmark to calculate the constant C from the previous subsection. The idea is to overstretch the short dsDNA tether that we want to observe, get the distance from the magnet to the sample at the transition, and set

the value of force at that point to be 65 pN. Then, using equation 2.10, we could obtain the value of C of that particular bead, which enables us to calculate the force for each magnet position just as we can do on long DNA tethers.

2.2.3 Loading Rate and DNA Unfolding Force

Studies in the field of ligand bonding and unfolding dynamics has shown that the loading rate of the force (how fast the force is changed) is important factor that determines the rupture force of the bond or structure [51]. The result of previous studies can be summarized in the Figure 2.12 below. Figure 2.12A shows the histogram of rupture force required to break the biotin-streptavidin bond, as seen in the diagram, there is a shift of peaks of the rupture force as well as an increase in the variance value as the loading rate is increased from $100pNs^{-1}$ to $\sim 65,000pNs^{-1}$ [51].

To understand the reasoning behind the dependency of the strength of the bond on the loading rate, it is important to know that the lifetime of a bond or structure sustained by a non-covalent interaction such as hydrogen bond, diminishes rapidly when subjected to a force due to thermal activation [51]. The result in Figure 2.12 A indeed shows that the average rupture force is higher for a higher loading rate, however, the bond survival time, or the time needed to break the bond is decreasing rapidly as the loading rate is increased. Conceptually, the force applied to the bond reduces the activation energy required to break the bond and decreases the likelihood of the bond survival. and speeds up dissociation.

Figure 2.12 B shows the exponential relationship between the rupture force and the loading rate. As shown, there are many different regimes each with a different slope, which are actually exponential constants. These different regimes corresponds to different energy barriers for the biotin-streptavidin bond as well as the biotin-avidin

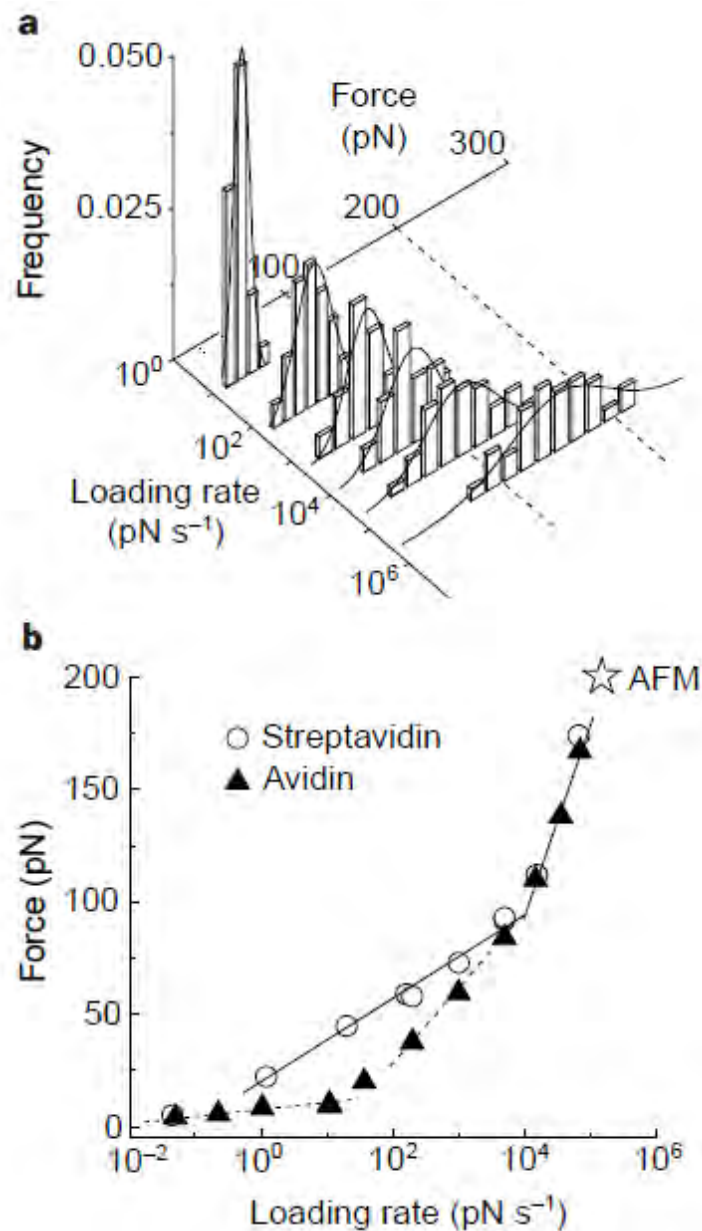


Figure 2.12: A) Force histograms of single biotin-streptavidin bonds with different loading rates. B) Dynamic strength spectra for biotin-streptavidin(circles) and biotin-avidin(triangles)

bond [52,53]. To put it simply, when the force is applied to the bond, it decreases the activation energy of the first (highest) energy barrier, until it reach a condition where the second energy barrier is the highest or the limiting factor in opening the bond, and thus the slope of the graph changed as an increase in force is required

to decrease the activation energy of a different energy barrier.

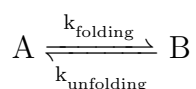
The importance of this finding to our study is that, for us to be able to observe the rupture force of G-quadruplex structure consistently, we need a constant loading rate used throughout the experiment, so that we have the same conditions on the average bond strength and the survival time of the structure.

2.3 Reaction Kinetics

Reaction kinetics is important as this study aims for the quantitative value of the stability of G quadruplex structure i.e. the reaction constant. This section mainly talks about the chemistry of reaction, or in this case transformation from one form to another.

2.3.1 Two-State Reaction Kinetics

Two state reaction is the simplest form of transformation, with the assumption that there is only one unfolded state of G4 motif (ssDNA) and one folded state of G quadruplex structure. Let us denote A as the unfolded state of G4, while B is the folded state, the reaction will look like,



where k_{folding} and $k_{\text{unfolding}}$ are the two reaction constants that we would like to find.

The result of my project will be in the form of probability of finding unfolded structure and the probability to find folded structure. Assuming the initial concentration of the G4 motif is $[M]$, we would like to find the expression of $p(A) = \frac{[A]}{[M]}$ and $p(B) = \frac{[B]}{[M]}$ as a function of time. The two equations that can be used to

get those expressions are the equation 2.11 of first order reaction velocity and the equation 2.12 of the total concentration of the substance (in our case the G4 motif),

$$\frac{d[A]}{dt} = -k_{folding}[A] + k_{unfolding}[B] \quad (2.11)$$

$$[M] = [A] + [B] \quad (2.12)$$

Substituting the expression of $[B]$ from equation 2.12 to equation 2.11 leads into,

$$\begin{aligned} \frac{d[A]}{dt} &= -k_f[A] + k_u([M] - [A]) \\ -\frac{d[A]}{(k_f + k_u)[A] - k_u[M]} &= dt \\ -\int_{[M]}^{[A]} \frac{d[A]}{(k_f + k_u)[A] - k_u[M]} &= \int_0^t dt \\ \ln \left(\frac{(k_f + k_u)[A] - k_u[M]}{k_f[M]} \right) &= -(k_f + k_u)t \\ \left(\frac{k_f + k_u}{k_f} \right) \frac{[A]}{[M]} &= \frac{k_u}{k_f} + e^{-(k_f + k_u)t} \\ p(A) = \frac{[A]}{[M]}(t) &= \frac{k_u}{k_f + k_u} + \left(\frac{k_f}{k_f + k_u} \right) e^{-(k_f + k_u)t} \end{aligned} \quad (2.13)$$

where k_f and k_u are the short form for folding and unfolding reaction constants respectively. Equation 2.13 is an important equation that will be used later to fit the data to obtain the value of the reaction constants.

Experimental Methodology

In this chapter, the step by step methodology is explained, all the experiments in this chapter were done by me with the help of lab members. It consists of the preparation of the ILPR G4 DNA, the preparation of channel, as well as the operation and initial settings of magnetic tweezers before data collection. All these steps are necessary for the any data collection.

3.1 DNA Preparation

The first step is to synthesize the DNA we need for observation. The construction of similar DNA has been done by Yue Xu, one of Post Doctoral student in our lab, the only difference is that I used ILPR G4 sequence instead of i-motif. The DNA constructed contains ILPR G4 single stranded (ss) DNA sequence with flanking oligonucleotides incorporated to two double stranded (ds) DNA handles of 581 base pairs (bp) and 1429 bp resulting in an ssDNA in the middle of the construction (see Figure 3.1) [54]. Below are the step by step methodology to obtain the DNA, it starts with the PCR on the handles, then the restriction, and it finishes with the ligation of the ILPR ssDNA to the handles.

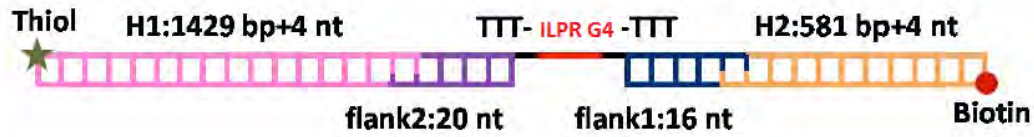


Figure 3.1: Schematic illustration of the DNA sample, handles on the left and right, and the ILPR G4 motif ssDNA in the middle

3.1.1 PCR on λ DNA

The main ingredients for this first step is the DNA handles to the left and to the right of the ssDNA in Figure 3.1. The first handle (H1) is a dsDNA sequence with 1429 bp, labelled with thiol at the end of the sequence, which is useful later to attach the DNA sequence to the coated glass surface. The second handle (H2) is a shorter dsDNA with 581 bp, labelled with biotin at the opposing end of the sequence, useful in attaching the DNA tether to the coated paramagnetic bead later.

Both of the handles are produced from the cutting of bacteria's λ DNA, that acts as a template in the polymerase chain reaction (PCR). The PCR includes this long λ DNA (48,502 bp, New England Biolabs, NEB), and DNA primers that will specifically replicate the sequence from 4,479 bp to 5,895 bp to produce H1, and the sequence from 5,331 bp to 5,899 bp to produce H2 [54]. Both of these handles are GC-rich (GC content >60 %), to favor the B to S transition of the DNA as a specific control when applied to a certain force [45, 46, 55].

Post PCR treatment of the DNA includes the DNA electrophoresis and the purifying process. DNA electrophoresis was done by using a marker DNA and a little amount ($\sim 5\mu l$) of dyed DNA samples (H1 and H2). The typical result can be seen in Figure 3.2 A below, the leftmost column is the marker DNA, where each band has a 500 bp gap from each other. The bottom band corresponds to the H2 (581

bp) while the middle band corresponds to H1 (1429 bp). To be clear, this DNA electrophoresis is only to make sure that we have the correct result for our sample and not used for further steps.

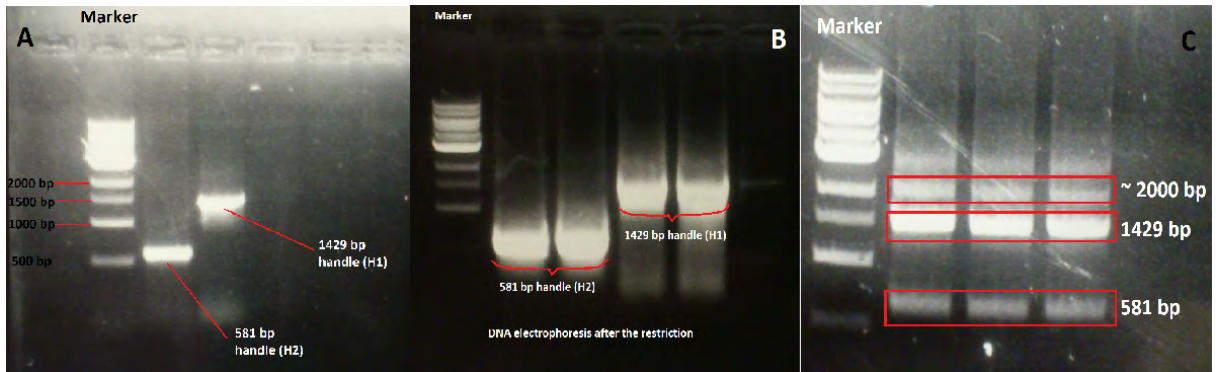


Figure 3.2: Gel electrophoresis result for, A. PCR product, B. Restriction product, C. Ligation product

The remaining DNA samples (typically $\sim 195\mu\text{l}$ each in my experiment) are purified using the *invitrogen* purifying process to get rid of small DNA sequences, the primers, and all other mixtures in the solution. This purifying process includes the addition of binding buffer, washing buffer, and elution buffer, with an incubation and centrifugation in between each addition (see Figure 3.3A below). The results of the purification are then measured with the *nanodrop* software to get their nucleic acid concentrations based on their absorption levels at 260 nm wavelength (Figure 3.3B). In my experiment procedure, the final results would be two columns of $\sim 48\mu\text{l}$ dsDNA, with pure H1 and H2 sequences inside.

3.1.2 Restriction Process

To make the two handles able to attach to the middle part of the desired product, they must be cut at a specific location, using a specific restriction enzyme. The restriction enzyme used is the BstXI restriction enzyme which cut the site 5'CCA NNNN|N TGG from our handles. Incubation in 37°C for two hours is required to

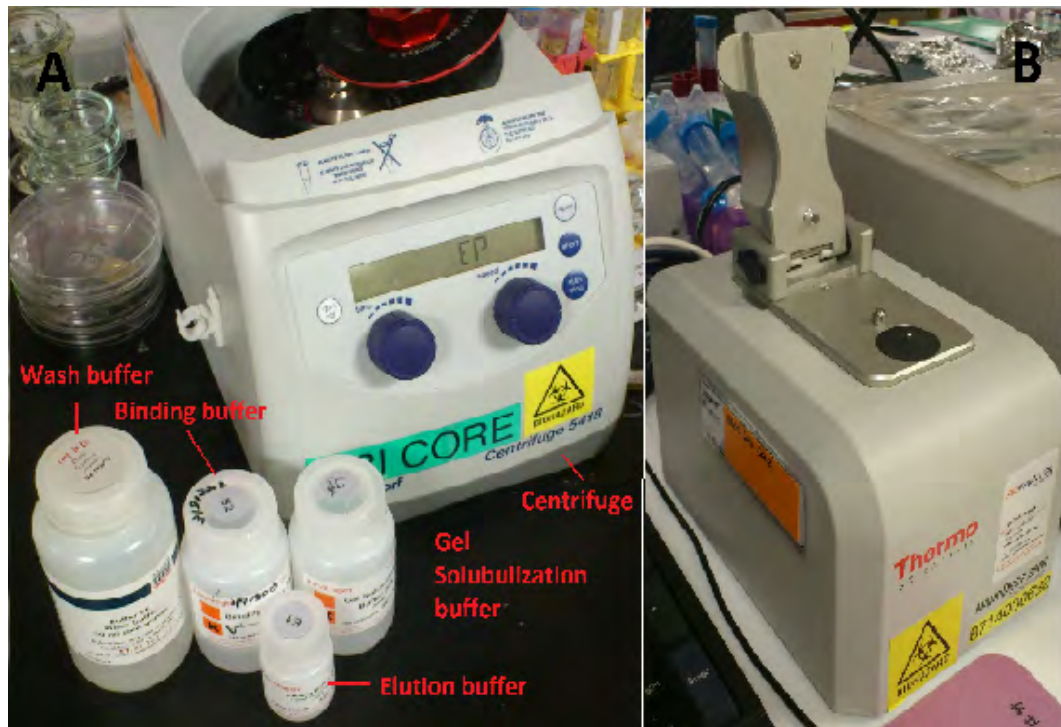


Figure 3.3: A. Picture of the centrifuge machine and buffers used for *invitrogen* purification process. B. Picture of *nanodrop* device that detects absorbance and calculates DNA concentration

allow the restriction process to complete. These cuttings on our handles produce the exact pink coloured H1 and cream coloured H2 from Figure 3.1, where the end part of the DNA is single stranded, which allow the ligation process to occur later on.

After the incubation for 2 hours, a DNA electrophoresis is done to the products to eliminate the small DNA fragments inside the solutions. The DNA electrophoresis result can be seen in Figure 3.2 B above, again, the bottom bands correspond to the shorter handle (H2), while the top bands correspond to the longer handle (H1). Next step is to cut the gel and do a purification process. The purification process is very similar to the previous *invitrogen* purification, but with an addition of gel solubilization buffer instead of a binding buffer, and a longer incubation time is required to solve the gel in 55 ° C. The results of the purification are our final

results in this step, again with the volume of $\sim 48\mu\text{l}$ for each handle, but with generally lower concentration of nucleotide compared to the previous step.

3.1.3 ILPR G4 Ligation

The middle oligonucleotides must be synthesized first by mixing the DNA flanks to the ssDNA with an ILRP G4 sequence. The sequences are as follow,

ssDNA: 5'-phosphate- CTTG TGCA CAGA CTCG **ACA GGGG TGT GGGG**
ACA GGGG TGT GGGG ACA GCA GCC AGG TCA GTA G **CGAC** -3'

flank1 (phosphate - H3, 20 nt): 5'-phosphate-CTA CTG ACC TGG CTG C

flank2 (BstXI downH, 16 nt): 5'- CGA GTC TGT GCA CAA G **GTGC**-3'

Mixing all three of these sequences up and anneal them (...) will result in the middle oligonucleotides needed for the final result.

The final step needed is to ligate the handles to the middle oligonucleotides which is done by mixing the correct amount of H1, H2, middle nucleotides, T4 DNA ligase (NEB) enzyme, and T4 ligase buffer altogether and incubate the mixture in 16°C overnight. The usual gel electrophoresis was done to isolate the desired DNA which is on the size of ~ 2000 bp, Figure 3.2 C above shows the resulting DNA electrophoresis, where there are 581 bp bands, 1429 bp bands, and 2000 bp bands. Next step is to cut the desired part of the gel and do the usual *in vitro* purification process to finally get the complete DNA construct.

3.2 Channel Preparation

This section explains the methods to attach the DNA tether to the glass surface and be ready for observation under the magnetic tweezers. The steps are to clean the glass surface and its cover slip, to coat it with 3-aminopropyl triethoxysilane

(APTES), to add sulfo-SMCC, to tether the DNA inside, and finally to attach the paramagnetic beads to the DNA tether. All those steps are explained in detail in below sections.

3.2.1 APTES Coated Channel

Sterilizing the glass surface and the cover slips is the first thing to do. Both the glass surface and the cover slips have the same cleaning procedure, so I will only explain the procedure of the glass surface. Seven pieces of glass surfaces were put into one jar in a zig zag pattern as you can see in the Figure 3.4 below, this pattern is a default pattern that will make the glass surface not easily fall down when we rinse it with distilled water or any liquid. Next, each jar was washed and rinsed about 6 times with distilled water and then a detergent was used to fill the jar for ultrasonic cleaning.



Figure 3.4: Views of the glass surfaces inside the jar, and the zig zag pattern of the glass shown in red lines

The process of ultrasonic cleaning is also called sonication, where an ultrasonic waves are used to induce cavitation bubbles which agitate the liquid inside the jar [56]. Alternating positive and negative pressure waves in the solution induce

the formation and the growth of the cavitation bubble, with the pressure inside could reach 500 atm. When the bubble reach the surface of the glass, the implosion event occurs, where the bubble transforms into one-tenth of its size and travels with a velocity up to 400 km per hour to the glass surface. These fast bubbles free contaminants from the glass, and for the small size and high energy of the bubbles, they can reach small crevices and remove entrapped soils effectively [57]. This process took place for about 15 minutes in my experiment for complete cleaning of the glass surface and the cover slips.

After the sonication, we need to dry the jars. The jars were washed and rinsed again for at least six times with distilled water, and 100% ethanol was used for the last wash and rinse to dry the glass surfaces more easily. Rinsed, the jars were put into an oven (*Boekel Scientific*, Figure 3.5 A) to dry the glass in approximately 15 minutes.

From now on, the cleaning of cover slips is completed, and the steps below are only for the glass surfaces as we want to coat them with 3-aminopropyl triethoxysilane (APTES). First, we need to do a plasma cleaning, not only for cleaning the glass from any inorganic materials, but also to expose the oxygen group from the glass surface instead of the usual hydroxide group. This plasma cleaning was done using *FEMTO Science* machine shown in Figure 3.5 B, and it happened for about 20 minutes. Next, the solution of 1 % APTES in methanol was prepared in fume hood, and this solution was added to the plasma cleaned jars to allow the APTES be attached to the glass surface. After those steps, more washing and rinsing were to be done, and also the drying using the *Broker Scientific* oven. The result are the APTES coated glass surfaces.

The whole intention of coating the glass surface to APTES is to add an amino group to the glass surface. The silane group of the APTES would have a covalent



Figure 3.5: A. The *Boekel Scientific Oven*. B. The *FEMTO Science* plasma cleaner

bond to the silica on the glass, where the amino group is ready for any ligand or crosslinker (in our case the crosslinker would be sulfo-SMCC, explained in the next section). The structure of APTES and the attachment of the structure to glass is illustrated in the Figure 3.6 below [58].

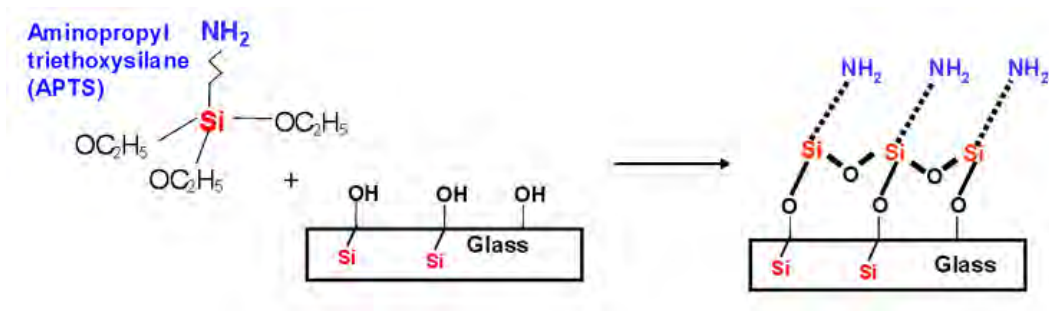


Figure 3.6: Illustration of APTES attachment to glass surface and the free amine groups that are exposed after the attachment

Hence, this section ends with the APTES coated glass surfaces and clean cover slips. These glass surfaces and cover slips can be store in room temperature for at least one month.

3.2.2 Addition of DNA Tether

Next, to prepare the actual channel, we attached the clean cover slip to the APTES glass surface by using sliced parafilms as a glue and silicon gel to strengthen the

attachment and as a guiding surface for the liquid to go in and out of the channel. Figure 3.7 below is the illustration of the channel for easier understanding about what are the roles of the parafilms and the silicon gel.

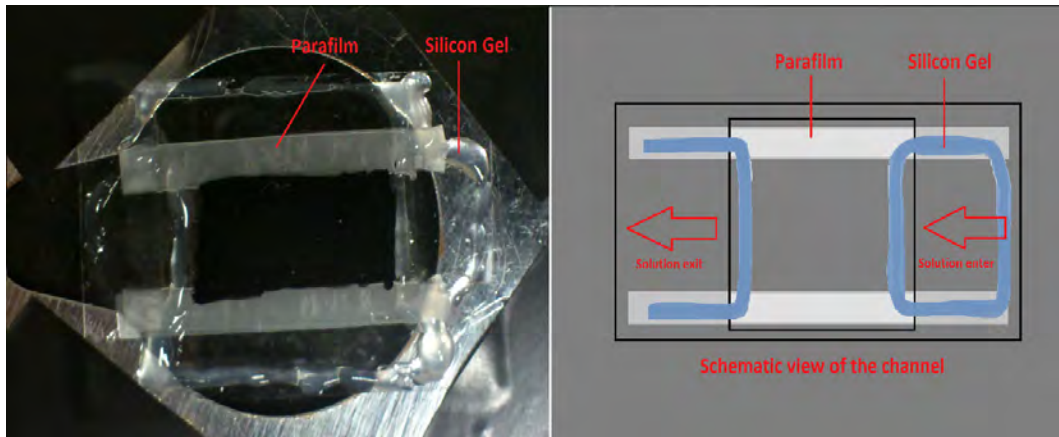


Figure 3.7: Pictures and illustrations of the APTES glass covered by clean coverslip, attached by using parafilm and silicon gel.

Having done with the physical preparation of the channel, the next step is to put the chemicals inside the channel. The first thing added to the channel was the white magnetic beads, $\sim 50\mu\text{l}$ of $3\mu\text{m}$ diameter white beads solution was added to the channel, and a 15 minutes incubation at room temperature follows. To note, the volume of the the channel is $\sim 20 - 25\mu\text{l}$, so an addition of $50\mu\text{l}$ is already filled up all the glass surface below the cover slip. The addition of this white bead has the intention of attaching the white beads to the glass surface to act as a reference bead that later will be used to calibrate the extension of the DNA tether as mentioned in Section 2.2.1.

After the incubation, the channel was washed by $\sim 100\mu\text{l}$ of phosphate buffered saline (PBS), a buffer that mimics human ionic condition. The washing is mainly for getting rid of the white beads that were not properly attached to the glass

surface. Next, before the addition of the DNA sample, we need to add a solution of Sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (Sulfo-SMCC), a water soluble, non cleavable, and membrane impermeable crosslinker. We used sulfo-SMCC in PBS with a concentration of $\sim 1\mu\text{g}/\mu\text{l}$. This sulfo-SMCC molecule has two reactive groups, the first one is sulfo-NHS ester, which is reactive towards amino group, while the second one is maleimide, which is reactive towards thiol group (R-SH). The illustration of the bonds formed by sulfo-SMCC crosslinker can be seen in Figure 3.8 below [59]. Thus, we have managed to attach a cross linker between the APTES coated glass surface and the DNA tether with thiol group at its end. An incubation for about 30 minutes to one hour is required to let the sulfo-SMCC attached properly to the APTES glass.

More washing by PBS is required to push away improperly attached sulfo-SMCC molecules, each channel with $\sim 1\text{ml}$ of PBS. Having the sulfo-SMCC, the channel is now ready to accept the DNA sample. The DNA sample is usually has a concentration of about $7 - 10\text{ng}/\mu\text{l}$, and it needs to be diluted around 20 to 100 times to have a suitable concentration. Incubation time of ~ 1 hour is required after the addition of DNA. As shown in the Figure 3.1 on the previous section, we now have a free biotin-end of the DNA tether inside the channel. Washing by 1ml of PBS is required after the incubation with a slower pace since the we do not want to wash away the DNA.

After the DNA addition, the next step is the blocking of the surface. Here, we used 1ml of $1\text{mg}/\text{ml}$ bovine serum albumin (BSA) mixed with $10\mu\text{l}$ of $100\mu\text{M}$ 2-mercaptoethanol (β -Me) as a blocking buffer. These BSA blocked surfaces are negatively charged and will not nonspecifically interact with DNA tethers in neutral pH, which is what we aimed for. The blocking procedure requires 2 - 3 hours incubation at room temperature.

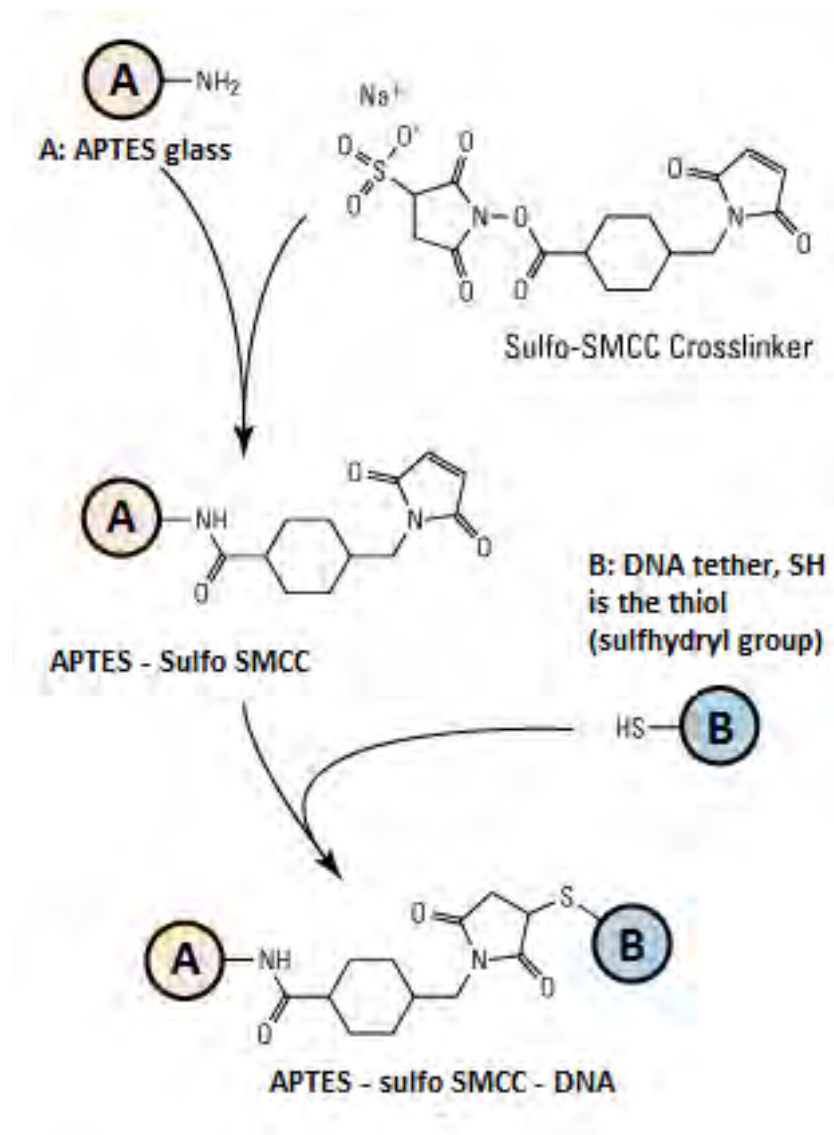


Figure 3.8: Illustration of Sulfo-SMCC as a cross linker between the amine group of APTES glass and the thiol group of the DNA tether

Final step was the addition of the streptavidin coated paramagnetic bead (brown bead). The bond between biotin complex and streptavidin was then made in the channel, it is one of the strongest non-covalent interaction known in nature [60]. An addition of $\sim 100\mu\text{l}$ of brown beads solution was done and an incubation time of 10 minutes was required for the bonds to form.

After that, $\sim 200\mu\text{l}$ of the working buffer, Tris-KCl, was added to the channel.

Thus, the channel was ready to be put under the magnetic tweezers. The final illustration of the channel can be seen in the Figure 3.9 below.

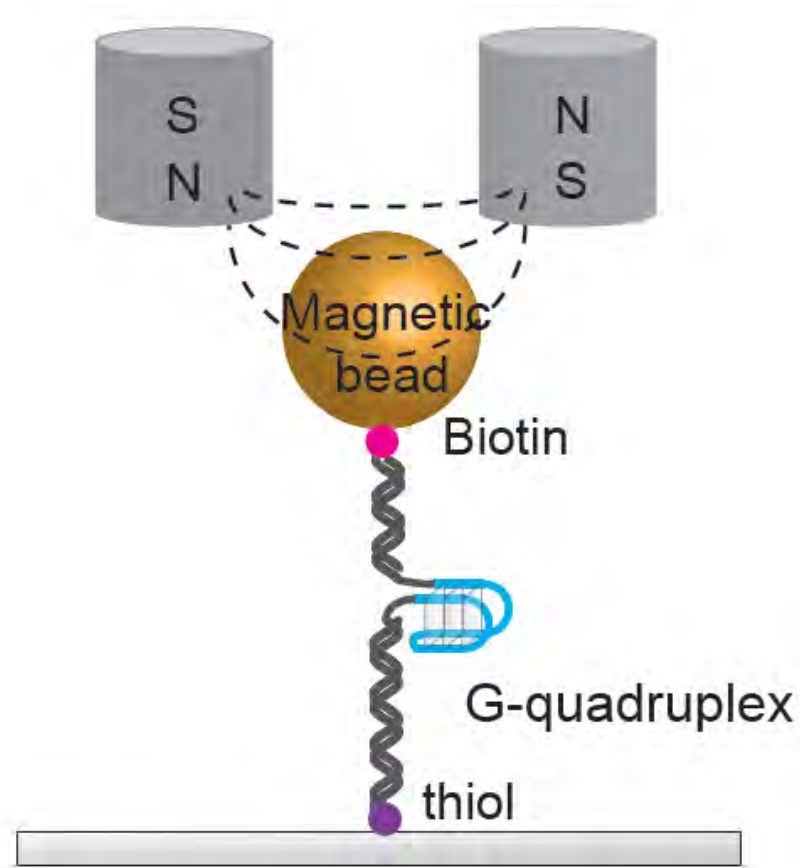


Figure 3.9: Illustration of the completed channel ready for observation under magnetic tweezers

3.3 Magnetic Tweezers Operation

This section mainly consist of the magnetic tweezers set up for my experiment. It consists of positioning the magnet, finding good reference bead and brown bead, and to find an overstretching transition. Post the set up, the data collection could run with two types of magnetic movements which are explained below.

3.3.1 Magnetic Tweezers Set up

These are the initial steps required to start collecting the data using magnetic tweezers,

XY position of the magnet

The objective lens of 4x or 10x is used to observe the tip of the magnet, and there are knobs on the side of the tweezers to displace the x and y position of the magnet pair. The correct set up is where the center of the lens is directly below the center of the pair of magnets. This positioning is important to reduce the torque leading to rotation effect in the observation of the DNA tether later.

Z position of magnet

As a standard point in our laboratory, 13.5 mm is used as the zero magnet position, meaning that if the magnet is moved downwards by 13.5 mm, the tip of the magnet will just hit the sample glass. For the calibration, we used a reference glass, and move the magnet until it hit the glass, and named it as a 13.5 mm point. However, the real sample glass is thicker than our reference glass, since it has a cover slip and a black tape on top of that, which caused the ~ 13 mm point to be the touching point. Other standard points in our lab can be seen in the Table 3.1 below.

Moving bead and reference bead

After the magnet position calibration, the sample was mounted and the objective lens 40x was used (with immersion oil added). Using this set up, at magnet position of 9 mm, we did the searching. The sample image from the CCD camera can be seen in the Figure 3.10 below, with a black and white color. However, under the microscope, the color of the moving bead is yellow (streptavidin coated bead), while the color of the reference bead is white, so that they are easily recognizable. In

Distance (μm)	Standard points
-9000	Mounting and unmounting the sample
9000	Position for finding good vibrating bead (around 2 or 3 pN force)
11000	Position for the calibration of the bead with the reference bead for the extension measurement
13000	Position when the tip of the magnets reach the surface of the sample
13500	Position when the tip of the magnets reach the surface of the reference glass

Table 3.1: Standard z-distance points used in our laboratory for magnetic tweezers experiment

this step, we basically looking for a good moving bead and a good reference bead in the same screen of microscope for the calibration.

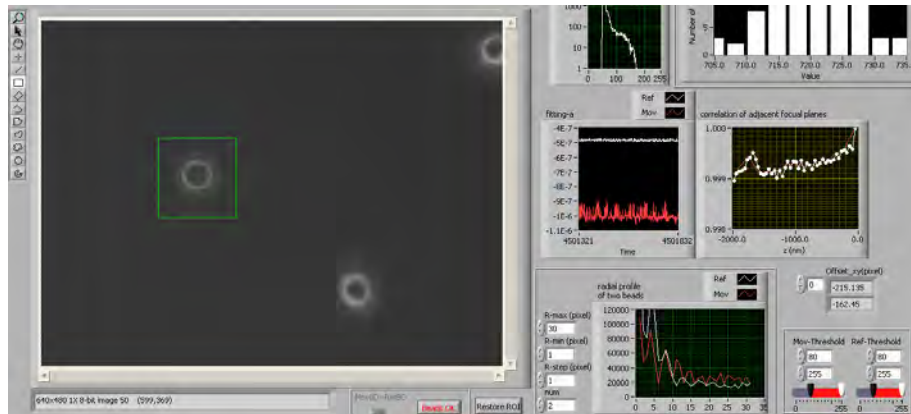


Figure 3.10: Screen shot of picture of beads from CCD camera attached to the microscope

Piezo Calibration

At z position of 11 mm, we did the calibration where the focus point of the reference bead and the moving bead were being compared and we get the extension value of the DNA tether as a result. After the calibration, we can see the correlation curve between the two beads, and the typical good shape of the correlation curve has a single peak curve for both the reference bead and the moving bead.

DNA overstretch

Now, to know the force acting on the DNA at each magnet position, we need to find the overstretching transition and take note of the magnet position when it happens. That particular magnet position was taken as a 65 pN point, and from that reference point, we know the force exerted from the magnet at any other magnet position (section 2.2.2).

3.3.2 Modes of Magnet Movement

After all those set ups, we now know the force exerted by the magnet as a function of the magnet position. Thus, the tweezers could be programmed to follow a certain modes of magnet movement. In our lab, there are two modes of magnet movement,

1. Constant magnet movement. In this mode, the magnet just moves with a constant speed which leads to exponential force increase. We did not use this mode in the study of G quadruplex unfolding
2. Constant loading rate. In this mode, the magnet moves in such a way that it generate constant increasing of force. We used this mode to study G quadruplex unfolding for a reason stated in the previous chapter, section 2.2.3.

Choosing the mode concludes the data taking part, all that is left to do after the data has been collected is to analyse the data which will be explained in the next chapter.

Chapter 4

Results and Discussion

This chapter talks about the data collected at the beginning of the project under the preliminary data section, with the intention of getting used to the experimental set up and to observe the typical unfolding signal of G quadruplex structure, as well as the real data collected at the end of the project under the data analysis section.

4.1 Preliminary Data

This section has the intention of showing the common preliminaries which must be present before every data collection. The overstretching transition must be reached for calibration purpose, and the typical G4 unfolding must be observe to make sure that the DNA pulled is the correct tether.

4.1.1 Overstretching Transition

As mentioned a couple of times in the previous chapters, overstretching transition of our DNA handles occur at ~ 65 pN force, where the DNA underwent the B to S transformation, and the DNA extension was increased by 1.7 fold [54].

Before any data collection, for us to be able to determine the force versus magnet position curve, we need to know what is the position of the magnet when the DNA undergoes an overstretching. Figure 4.1 below shows the overstretching transition of our DNA handle from $\sim 825\mu\text{m}$ to $\sim 1225\mu\text{m}$. From the position of the magnet when this transition happens, we look for the C value of the particular bead we observe according to the following formula,

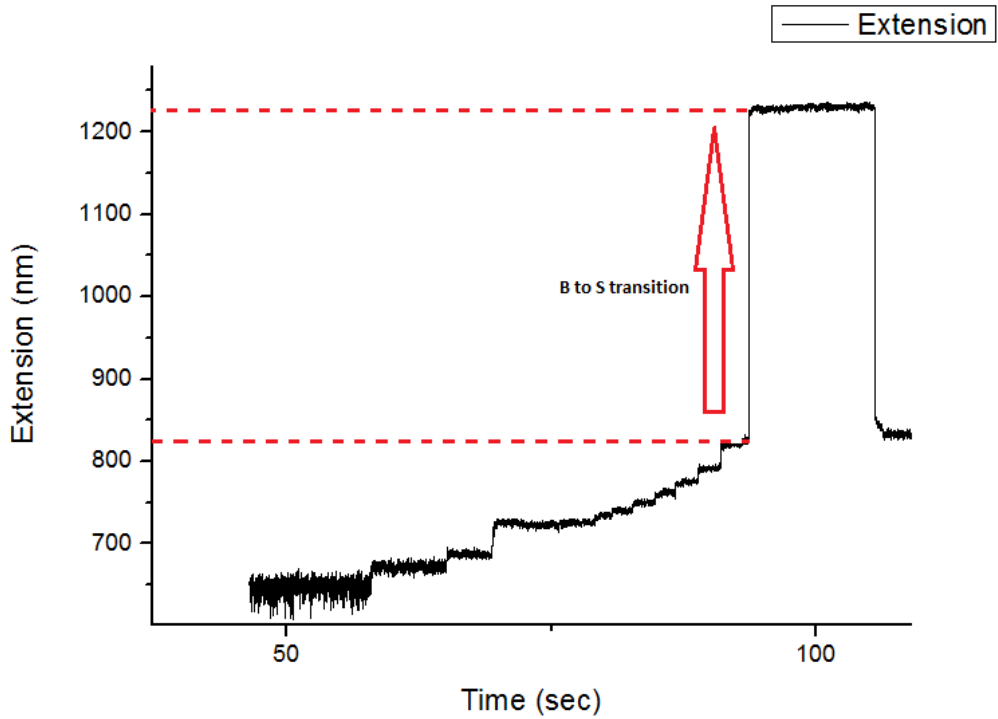


Figure 4.1: Preliminary data on DNA overstretching transition

$$F = C \left(\alpha_1 \exp\left(\frac{-d}{\gamma_1}\right) + \alpha_2 \exp\left(\frac{-d}{\gamma_2}\right) \right) \quad (4.1)$$

inserting the force (F) to be 65 pN, the distance between magnet and sample (d), and the constants ($\alpha_1, \alpha_2, \gamma_1, \text{ and } \gamma_2$) accordingly, we can obtain the value of C for this particular DNA tether.

Having obtained the C value, the program to produce a constant loading rate can

be executed by the tweezers and we are ready to unfold the G quadruplex structure.

4.1.2 Typical G4 Unfolding Signal

Using a constant loading rate, we were able to observe the G quadruplex structure unfolding, Figure 4.2 below are the results of cycles of repeated increase in force from 1 pN to 50 pN with loading rate of 2 pN / sec,

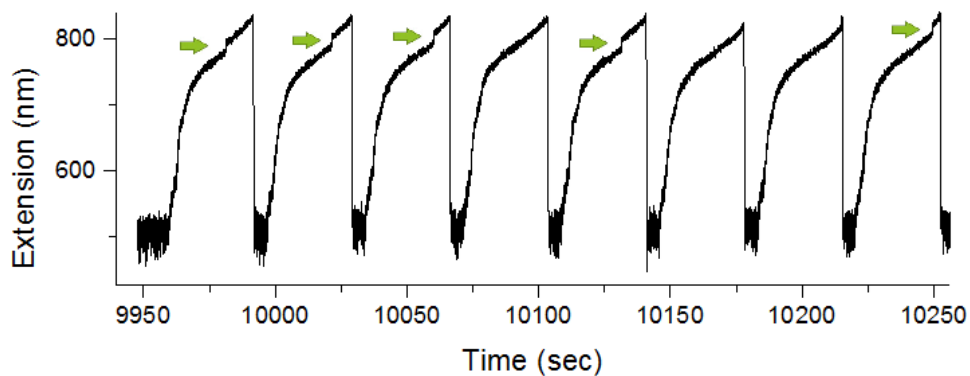


Figure 4.2: Preliminary data on G quadruplex unfolding signal under constant loading rate

the green arrows represent the unfolding signals observed. The typical extension of DNA tether for the unfolding of G quadruplex structure is $10 \mu m$, but it is actually depend on the loading rate used when we stretch the DNA [54]. For a clearer view, Figure 4.3 below shows a single unfolding signal observed,

where the extension increases by about $20 \mu m$. For consistency, the unfolding signal for our data collection is detected by a *MATLAB* program written by our lab member, which will be used throughout the experiment.

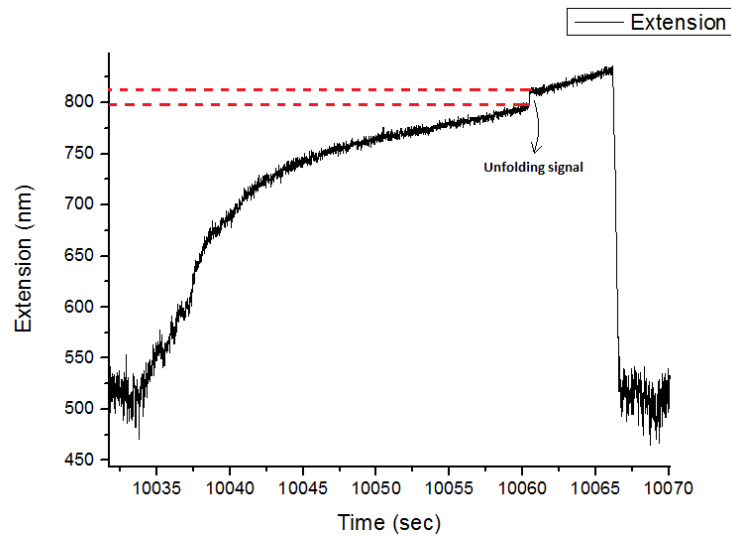


Figure 4.3: Single G quadruplex unfolding signal

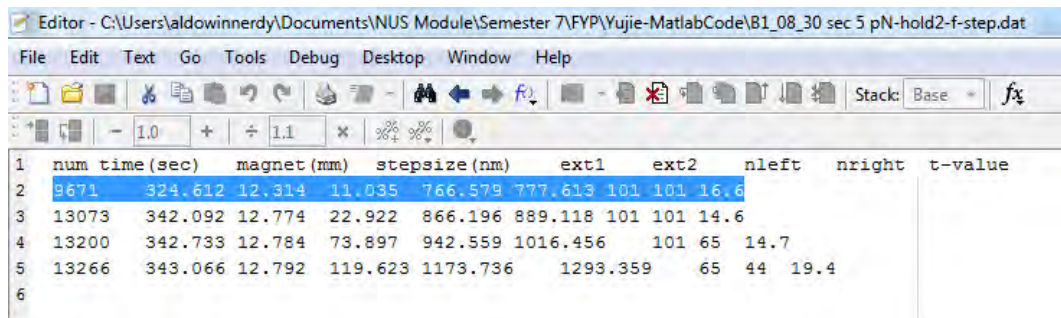
4.2 Data Analysis

The complete set of data consist of at least 32 cycles of overstretching and relaxing back the DNA sample for each condition. The changing variables are as follows,

1. The tethered DNA selected. This variable is normalized by detecting the overstretching transition and adjusting the C value of the calibration equation, so that this variable is actually 'kept constant'.
2. The holding force. This variable is the force applied to the tethered DNA at the beginning of each cycle, it is the period where the DNA is given the chance to refold the G4 structure (holding period). It consists of 1, 2, and 5 pN.
3. The holding time. This variable is the time given for the holding period. It consists of 30, 20, 10, 5, 2, and 1 seconds for each holding force.

4.2.1 Force Distribution

The force needed to unfold the structure can be obtained by observing the magnet position at the time when the unfolding signal occurs. *MATLAB* code written by one of the lab member is used to detect the jumps occurring when the DNA is pulled. Below (Figure 4.4) is the example of the detected jumps, where the highlighted data is the one that corresponds to the G4 unfolding signal.



num	time (sec)	magnet (mm)	stepsize (nm)	ext1	ext2	nleft	nright	t-value
9671	324.612	12.314	11.035	766.579	777.613	101	101	16.6
13073	342.092	12.774	22.922	866.196	889.118	101	101	14.6
13200	342.733	12.784	73.897	942.559	1016.456	101	65	14.7
13266	343.066	12.792	119.623	1173.736	1293.359	65	44	19.4

Figure 4.4: Output of the *MATLAB* code that detects the jumps of signal, the highlighted data is chosen by manually observing the data

Knowing the magnet position, I used the *C* value of the DNA to compute the force exerted by the magnet (with the equation 4.1). Hence, I can get the data of the unfolding force for each cycle of pulling. Compiling all the data, the result of the force distribution can be seen in the Figure 4.5 below.

Looking at the graph, it forms a single peak Gaussian distribution, suggesting that there is only one type of G4 structure observed. This is in conflict with the previous result by Mao Hanbin which showed two peaks on an ILPR G4 unfolding force distribution [61].

Fitting to the normal distribution formula, our results state the the average unfolding force is $(34.3 \pm 0.9)pN$, with the standard deviation of $(13 \pm 3)pN$. The summary of the fitting can be seen in Figure 4.6 below.

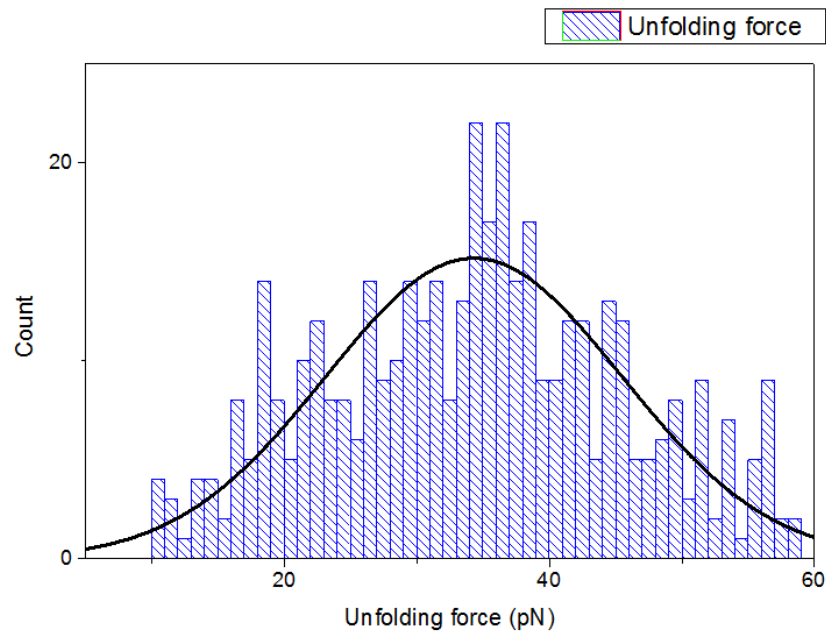


Figure 4.5: The unfolding force distribution histogram, each bin has a width of 1 pN, the black curve is the Gaussian fit

The consequence of this result is that the G4 structure is very stable at least *in vitro*, since a force of more than 30 pN is needed to unfold the structure. In context with the transcription of DNA, we should compare this force to the 'stall' force that is exerted by the RNA polymerase or other enzymes with helicase activities which is below $\sim 20\text{pN}$ [61]. It can be argued then that the force from the enzyme is not enough to unfold the ILPR G4 structure, which supports the hypothesis that the G4 structure might affect the transcription regulation by blocking the transcription process. However, we can only deal with the mechanical property *in vitro*, without analysing too much on the chemical property.

4.2.2 Refolding Probability

There are 18 different conditions that have been observed, 3 different holding forces, with each force having 6 different holding times. From every condition, the number of cycles in which the DNA tether was having an unfolding signal is compared to

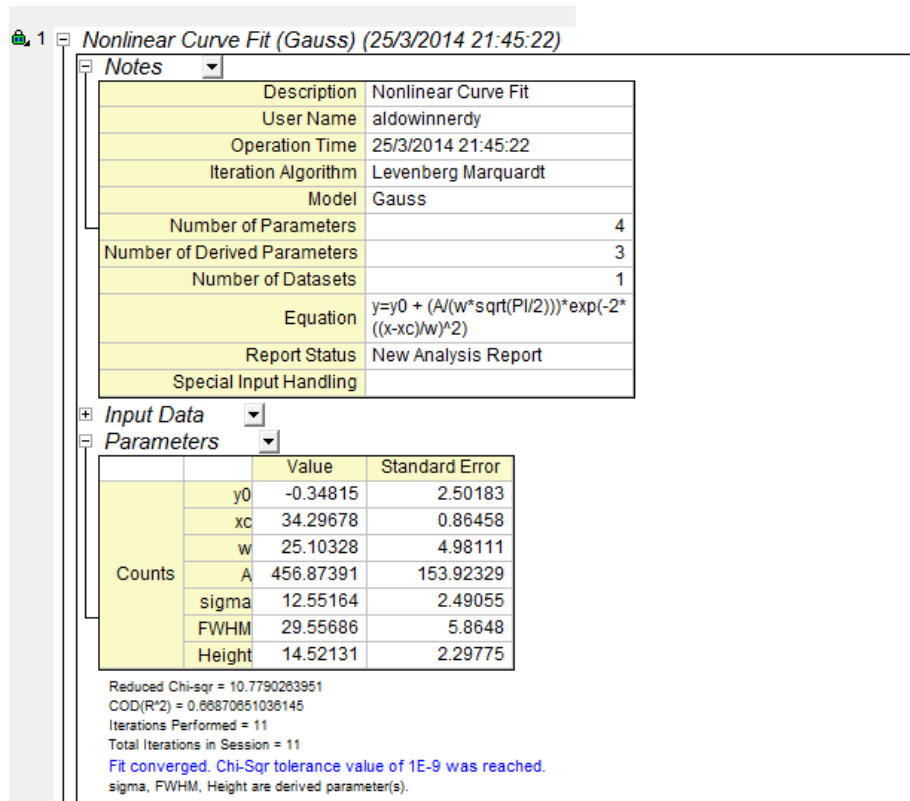


Figure 4.6: The summary data of the Gaussian fit, with the constraints of $y_0 = 0$ the total number of observed cycles, which gives us the refolding probability of the DNA tether from the unfolded state to the folded state.

The overall results of the observation can be seen in the Figure 4.7 above. All three holding forces behave in the same manner where the refolding probability is increasing as the holding time is increased, but with different rate. This data is in agreement with the calculated refolding probability formula,

$$p(\text{refolding}) = \frac{k_f}{k_f + k_u} - \left(\frac{k_f}{k_f + k_u} \right) e^{-(k_f + k_u)t} \quad (4.2)$$

where k_f and k_u are the velocity constants of the folding and unfolding process respectively.

Higher holding forces would favor the unfolded state, therefore there is little point in taking more data in the higher force region (for example, at 7 pN and 30 seconds

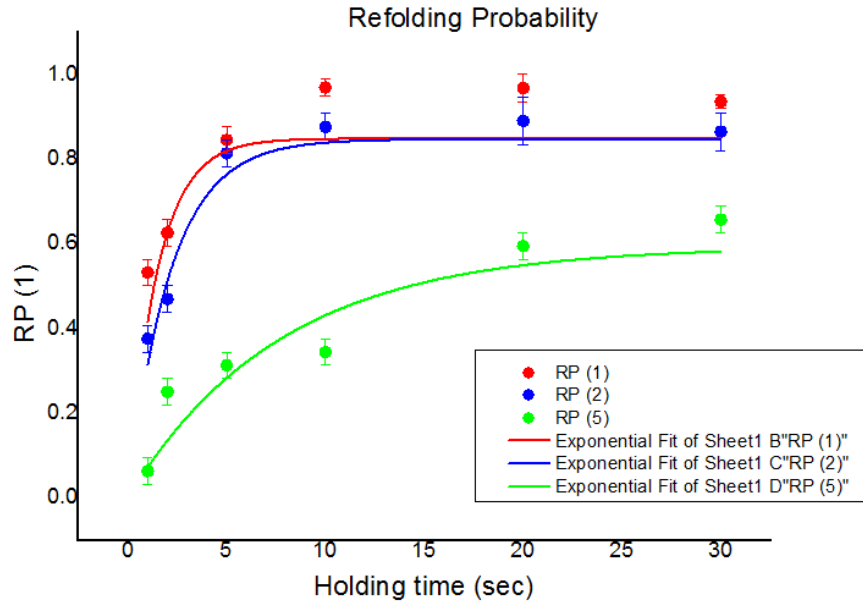


Figure 4.7: The data points and the exponential fittings of the refolding probability plotted against holding time, each color represents different holding force

holding time, we almost cannot see a single DNA unfolding signal). A lot more data can be collected between the force of 2 pN and 7 pN, and also for a longer holding time (60 and 120 seconds), but time constraint was limiting the collection of data.

4.2.3 Data Fitting

The fitting of the refolding probability graph is done using *Origin*, the equation used for the fitting is,

$$y = y_0 + Ae^{R_0 t} \quad (4.3)$$

with a constraint of $y_0 = -A$, to match the previous equation 4.2. Figure 4.8 below is the summary of the fitting results for the three forces used in the experiment,

Using this data and equation 4.2, the value of k_f and k_u can be computed for each

Summary									
	y0		A		R0		Statistics		
	Value	Standard Error	Value	Standard Error	Value	Standard Error	Reduced Chi-Sqr	Adj. R-Square	
RP (1)	0.83267	0.07783	-0.83267	0.75006	-0.70259	0.76114	0.02013	0.43433	

Summary									
	y0		A		R0		Statistics		
	Value	Standard Error	Value	Standard Error	Value	Standard Error	Reduced Chi-Sqr	Adj. R-Square	
RP (2)	0.84821	0.03656	-0.84821	0.18287	-0.46053	0.16397	0.00371	0.9296	

Summary									
	y0		A		R0		Statistics		
	Value	Standard Error	Value	Standard Error	Value	Standard Error	Reduced Chi-Sqr	Adj. R-Square	
RP (5)	0.59338	0.10627	-0.59338	0.12326	-0.1293	0.08527	0.00979	0.80017	

Figure 4.8: Summarized result of the fittings using the exponential equation, y_0 , A , and R_0 with their respective standard errors

holding force, the result is summarized in the following table,

Holding force (pN)	$k_{folding} (s^{-1})$	$k_{unfolding} (s^{-1})$
1	0.585	0.118
2	0.391	0.070
5	0.077	0.053

Table 4.1: The velocity constants obtained after fitting

This result suggest that the value of both the folding and unfolding constants are decreasing when we increase the holding force. However, the value of k_f decreased in a significantly steeper manner if compared to the decreasing value of k_u . This fact gives us the rough idea of the kinetics of the folding process. At low force (1 and 2 pN), the transition between the folded and unfolded states is easier than at high force (5 pN), which can be seen by the value of velocity constants that are higher in the case of low force.

Other important values that we could obtain from our data are the extrapolation of k_f and k_u to zero pico Newton force, as well as the holding force required to make the k_f equals to k_u . These two things can be derived from the equations in the Figure 4.9 below,

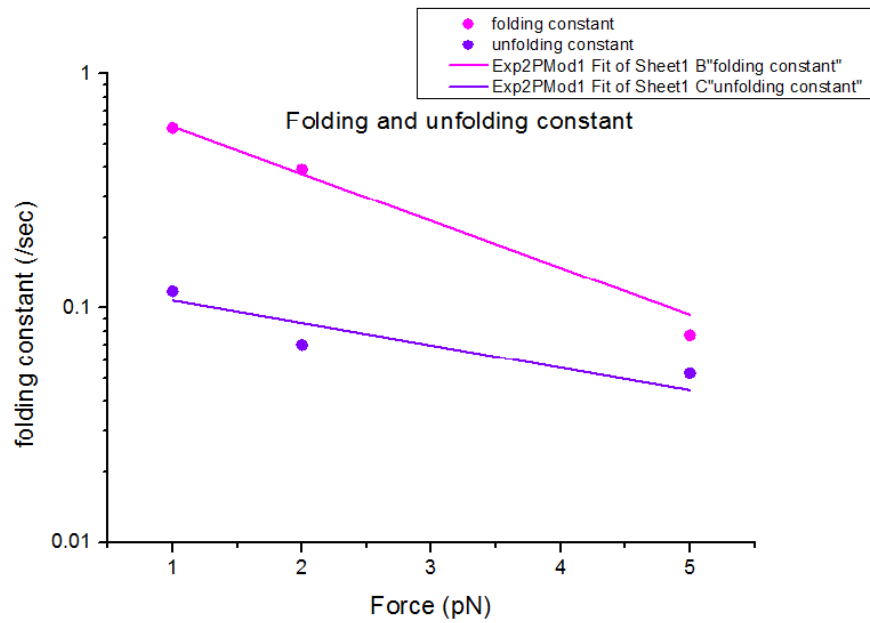


Figure 4.9: Rough exponential fit of the folding and unfolding constant, more data should be obtained to make the graph valid

The extrapolation is mainly done to look at the G4 structure stability *in vivo*, where the force affecting the DNA should be near zero. However, since it is done only with three data, the result of the extrapolation is very rough and has a high data error. The folding constant at zero force is determined to be $k_f = (0.94 \pm 0.08)s^{-1}$, while the unfolding constant at zero force is $k_u = (0.13 \pm 0.03)s^{-1}$. The ratio between these two values is the equilibrium constant between these two states of G4 motif. At zero force, the computed equilibrium constant is therefore $K_{eq} = 7.01 \pm 0.04$.

The intersection of the Figure 4.9 above is the point where $k_f = k_u$, which means that the equilibrium constant is equal to one. The force at the intersection point is $F_{intersect} = 8.05pN$, and at this force, the probability of finding the folded state or unfolded state is exactly half. Above this force, the probability of unfolded state is actually higher than the folded state.

Conclusion and Future Works

5.1 Conclusion

G quadruplex stability especially in the insulin-linked polymorphic region (ILPR) is studied *in vitro*. Magnetic tweezers is used to pull the tethered DNA of about 650 μm , with the unfolding signal of about 10 to 20 μm . The use of magnetic tweezers instead of optical tweezers or atomic force microscopy (AFM) is due to the advantages in terms of specificity, no heating, and force stability. The DNA sample is prepared by polymerase chain reaction, with the handles from bacterial λ DNA, and the human ILPR G4 motifs in the middle of the two handles.

The average force required to unfold the G quadruplex structure is obtained to be $F_{ave} = (34.3 \pm 0.9)pN$, and the distribution of the data is Gaussian with the standard deviation of $(13 \pm 3)pN$.

The refolding probability of the DNA tether is observed for three different holding forces, with each force having 6 different holding times. The results are in the form of velocity constants of folding and unfolding processes. We obtained the value of folding constant extrapolated to 0 pN force to be $k_f = (0.94 \pm 0.08)s^{-1}$, whereas the unfolding constant at 0 pN is determined to be $k_u = (0.13 \pm 0.03)s^{-1}$, which

leads to their ratio, $K_{eq} = 7.01 \pm 0.04$. This result suggest that the G4 motifs has a high probability to form G4 structure *in vivo*. A Force of 8.05 pN is obtained to be the force required to make the K_{eq} equal to 1, this is the condition where the probabilities of finding the unfolded state and the folded state are equal.

Having obtained the value of velocity constants and the unfolding force in a constant loading rate of 2 pN/s, we hope that it will be useful for further research and helpful in studying the kinetics of the unfolding or folding processes. The stretch goal of this study is to provide a possible explanation on how the G4 structure unfolding and folding in our insulin gene promoter can affect the transcription process and therefore the gene expression.

5.2 Future Works

This section is mainly used to list the data that could be collected if there is more time, as well as the relevant subject that can be studied in order to develop and analyse the data obtained to be more relevant for further research in G4 unfolding kinetics.

5.2.1 Completion of Data

A lot more data can be obtained in the same manner as what has been done in this project, varying the conditions such as the holding time of 60 and 120 seconds, as well as the holding force of 3, 4, and 6 pN would be nice. These data might increase the accuracy and the reliability of the result significantly.

Once the data has been collected, we could also do the Bell model analysis, Bell model is a simple two-state transition model that have a force-dependent kinetic constant as follows,

$$k(f) = k_0 \exp\left(\frac{x_u f}{k_B T}\right) \quad (5.1)$$

where x_u is the transition distance to the unfolding barrier along the force direction, and k_0 is the unfolding constant at zero force [54].

With above equation, a fitting could be done to the graph of kinetic constant versus holding force in our experiment, and we might get the transition distance of unfolding and folding transitions, which might be helpful in further research in G4 unfolding and folding kinetics.

5.2.2 Free energy difference

Plotting the graph of the equilibrium constant against the holding force would give us the value of the free energy difference between the folded and unfolded state of the DNA. This value might be important for further research.

Equilibrium constant, K_{eq} can be obtained by looking directly at the ratio between the unfolding and folding kinetic constants, since there are only two states. With about 5 or 6 data on equilibrium constant at various holding forces, the graph could be plotted, and we can fit the graph with the following Nernst equation at equilibrium,

$$\ln K_{eq} = C \Delta G^0 \quad (5.2)$$

where ΔG^0 is the free energy difference between the two states, and C is the constant related to the holding force. Fitting the graph with above equation can give us the result of the free energy difference which might be helpful for further study of G4 unfolding kinetics.

Bibliography

- [1] Nancy Maizels and Lucas T. Gray. The g4 genome. *PLOS Genetics*, 9(4), 2013.
- [2] M. Gellert *et al.* Helix formation by guanylic acid. *PNAS*, (48), 1962.
- [3] D. Sen and W. Gilbert. Formation of parallel four-stranded complexes by guanine rich motifs in dna and its implications for meiosis. *Nature*, (334):364–366, 1988.
- [4] Daekyu Sun *et al.* Inhibition of human telomerase by a g-quadruplex-interactive compound. *J Med Chem*, (40):2113–2116, 1997.
- [5] S.M. Haider *et al.* A structural analysis of g-quadruplex/ligand interactions. *Biochimie*, (93):1239–1251, 2011.
- [6] Adam Siddiqui-Jain *et al.* Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. *PNAS*, 99(18):11593–11598, 2002.

-
- [7] A. Lew *et al.* Unusual dna structure of the diabetes susceptibility locus iddm2 and its effect on transcription by the insulin promoter factor pur-1/maz. *PNAS*, 97(23):12508–12512, 2000.
- [8] Insulin. <http://http://en.wikipedia.org/wiki/Insulin>, 2014.
- [9] Hu Chen *et al.* Improved high-force magnetic tweezers for stretching and refolding of proteins and short dna. *Biophysical Journal*, 100:517–523, 2010.
- [10] H. J. Lipps and D. Rhodes. G-quadruplex structures: *in vivo* evidence and function. *Trends in Biology*, 19(8):416–422, 2009.
- [11] E. Besnard *et al.* Unravelling cell type-specific and reprogrammable human replication origin signatures associated with g-quadruplex consensus motifs. *Nat Struct Mol Bio*, (19):837–844, 2012.
- [12] J. W. Shay and W. E. Wright. Role of telomeres and telomerase in cancer. *Seminars Cancer Biol*, (21):349–353, 2011.
- [13] A. M. Zahler *et al.* Inhibition of telomerase by g-quartet dna structure. *Nature*, (350):718–120, 1991.
- [14] L. Organesian *et al.* Extension of g quadruplex dna by ciliate telomerase. *EMBO J*, (25):1148–1159, 2006.
- [15] L. Organesian *et al.* Telomerase recognizes g quadruplex and linear dna as distinct substrates. *Biochemistry*, (46):11279–11290, 2007.
- [16] S. Neidle. Human telomeric g quadruplex: the current status of telomeric g quadruplexes as therapeutic targets in human cancer. *FEBS J*, (277):1118–1125, 2010.
- [17] T. de Lange. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev*, (19):2100–2110, 2005.

-
- [18] A. Sfeir *et al.* Mammalian telomeres resemble fragile sites and require trf1 for efficient replication. *Cell*, (138):90–103, 2009.
- [19] A. Sfeir *et al.* Removal of shelterin reveals the telomere end-protection problem. *Science*, (336):593–597, 2012.
- [20] J.S. Smith *et al.* Rudimentary g-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Bio*, (18):478–485, 2011.
- [21] M. L. Bochman *et al.* Dna secondary structures: stability and function of g-quadruplex structures. *Nat Rev Genet*, 13(11):770–780, 2012.
- [22] Dna replication process. http://www.nature.com/scitable/content/ne0000/ne0000/ne0000/ne0000/14668888/U2.cp1.1_439542a-f1.2.jpg, 2014.
- [23] L. Crabbe *et al.* Defective telomere lagging strand synthesis in cells lacking wrn helicase activity. *Science*, (306):1951–1953, 2004.
- [24] C. Barefield and J. Karlseder. The blm helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. *Nucleic Acids Res*, (40):7358–7367, 2012.
- [25] H. S. Seifert *et al.* Dna tranformation leads to pilin antigenic variation in *Neisseria gonorrhoeae*. *Nature*, (336):392–395, 1988.
- [26] L. A. Cahoon and H. S. Seifert. An alternative dna structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science*, (325):764–767, 2009.
- [27] Refseq genes, ncbi. <http://www.ncbi.nlm.nih.gov/refseq/rsg/about/>, 2014.

-
- [28] J. L. Huppert and S. Balasubramanian. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, (35):406–413, 2007.
- [29] J. Eddy and N. Maizels. Gene function correlates with potential for g4 dna formation in the human genome. *Nucleic Acids Res*, (34):3887–3896, 2006.
- [30] D. Sun and L. H. Hurley. The importance of negative superhelicity in inducing the formation of g-quadruplex and i-motif structures in the c-myc promoter: implications for drug targeting and control of gene expression. *J Med Chem*, (52):2863–2874, 2009.
- [31] T. A. Brooks *et al.* Making sense of g-quadruplex and i-motif functions in oncogene promoters. *FEBS*, (277):3459–3469, 2010.
- [32] Q. Wei and B. M. Paterson. Regulation of myod function in the dividing myoblast. *FEBS Lett*, (490):171–178, 2001.
- [33] A. Yafe *et al.* Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by myod, mrf4 and myogenin. *Nucleic Acids Res*, (36):3916–3925, 2008.
- [34] M. R. Santoro *et al.* Molecular mechanism of fragile x syndrome: a twenty-year perspective. *Annu Rev Pathol*, (7):219–245, 2012.
- [35] M. DeJesus-Hernandez *et al.* Expanded ggggcc hexanucleotide repeat in non-coding region of c9orf72 causes chromosome 9p-linked ftd and als. *Neuron*, (72):245–256, 2011.
- [36] P. Fratta *et al.* C9orf72 hexonucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms rna g-quadruplexes. *Sci Rep*, (2):1016, 2012.
- [37] A. E. Renton *et al.* A hexanucleotide repeat expansion in c9orf72 is the cause of chromosome 9p21-linked als-fts. *Neuron*, (72):257–268, 2011.

- [38] H. Kobayashi *et al.* Expansion of intronic ggcctg hexanucleotide repeat in nop56 causes sca36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am J Hum Genet*, (89):121–130, 2011.
- [39] Setup of magnetic tweezers. <http://large.stanford.edu/courses/2007/ph210/sung1/images/f1big.jpg>, 2014.
- [40] C. Haber and D. Wirtz. Magnetic tweezers for dna micromanipulation. *Rev Sci Instrum*, (71):4561–4570, 2000.
- [41] B. G. Hosu *et al.* Magnetic tweezers for intracellular applications. *Rev Sci Instrum*, (74):4158–4163, 2003.
- [42] G. Fonnum *et al.* Characterization of dynabeads (r) by magnetization measurements and mossbauer spectroscopy. *J Magn Magn Mater*, (293):41–47, 2005.
- [43] Brownian motion of dna tether. <http://large.stanford.edu/courses/2007/ph210/sung1/images/f2big.jpg>, 2014.
- [44] Jong Min Sung. Single molecule studies using magnetic tweezers. 2007.
- [45] S. B. Smith *et al.* Overstretching b dna: the elastic response of individual double-stranded and single-stranded dna molecules. *Science*, (271):795–799, 1996.
- [46] H. Fu *et al.* Two distinct overstretched dna states. *Nucleic Acids Res*, (38):5594–5600, 2010.
- [47] C. P. Calderon *et al.* Quantifying dna melting transitions using single-molecule force spectroscopy. *J Phys Condens Matter*, (21):34114, 2009.
- [48] M. Rief *et al.* Sequence-dependent mechanics of single dna molecules. *Nat Struct Biol*, (6):346–349, 1999.

- [49] Dna b to s transition. <http://web.mit.edu/smart/research/biosym/DNA%20stretch-PNAS-2.jpg>, 2014.
- [50] C. Danilowicz *et al.* The structure of dna overstretched from the 5'5' ends differs from the structure of dna overstretched from the 3'3' ends. *Proc Natl Acad Sci USA*, (106):13196–13201, 2009.
- [51] R. Merkel *et al.* Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Letters to Nature*, 397:50–53, 1999.
- [52] P. C. Weber *et al.* Structural origins of high-affinity biotin binding to streptavidin. *Science*, 243:85–88, 1989.
- [53] O. Livnah *et al.* Three-dimensional structures of avidin and the avidin-biotin complex. *Proc Natl Acad Sci*, (90):5076–5080, 1993.
- [54] Xu Yue *et al.* Mechano-chemical selections of two competitive unfolding pathways of a single dna i-motif. 2013.
- [55] P. Cluzel *et al.* Dna: An extensible molecule. *Science*, (271):795–799, 1996.
- [56] Ultrasonic cleaning. http://en.wikipedia.org/wiki/Ultrasonic_cleaning, 2014.
- [57] Ultrasonication cleaning 101. <http://bluewaveinc.com/ultrasonic-cleaning-101/>, 2014.
- [58] Surface modification - chemical and biochemical. <http://nanostructure.usc.edu/research/bio1/SAM9.PNG>, 2014.
- [59] Sulfo smcc cross linker. <http://www.piercenet.com/media/22322-001-Sulfo-SMCC-HRP-Ab-312px.jpg>, 2014.
- [60] Biotin streptavidin bond wikipedia. <http://en.wikipedia.org/wiki/Streptavidin>, 2014.

-
- [61] Y. Zhongbo *et al.* Ilpr g quadruplexes formed in seconds demonstrate high mechanical stabilities. *JACS*, (131):1876–1882, 2009.