NOVEL APPROACHES TO DNA SEQUENCING

by

Quinn Acelia Spadola

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

May 2008

NOVEL APPROACHES TO DNA SEQUENCING

by

Quinn Acelia Spadola

has been approved

April 2008

Graduate Supervisory Committee:

Stuart M. Lindsay, Chair
Martha McCartney
Joan McGregor
Robert Ros
Hao Yan
Peiming Zhang

ACCEPTED BY THE GRADUATE COLLEGE

ABSTRACT

This dissertation presents work on two different novel DNA sequencing methods. The first was based on using the atomic force microscope (AFM) to detect the changes in force as a nanopore is pulled over surface tethered single stranded DNA. This technique required the synthesis and verification of multiple chemical components. It was determined, both theoretically and experimentally, that the changes in force as purines or pyrimidines pass through the nanopore were too small to be measured using the AFM. However, the method was successfully applied to studying the opening of hairpins in single stranded DNA when a nanopore is forced over them. This geometry is similar to how polymerases approach hairpins while transcribing DNA. It differs from other techniques which pull the ends of the DNA apart in order to study secondary structures. The results show that the nanopore method requires higher forces and less strain when hairpins are opened, as compared to the other methods.

The work studying the translocation of DNA through a nanopore also informed the second sequencing method described. This technique is grounded in work using a scanning tunneling microscope (STM) to determine the conductance between Watson-Crick base pairs of DNA. The full sequencing method consists of a single strand of DNA, under an electrophoretic force, passing through a nanopore to a pair of electrodes functionalized such that the DNA bases form hydrogen bonds to one electrode through bases attached to the electrode, whereas its phosphate backbone forms hydrogen bonds to the other via guanidinium, completing the electrical circuit.

An important aspect of this method is the reversible hydrogen bonding between the phosphate backbone of DNA and guanidinium. Using the AFM, the adhesion of DNA

was studied under various conditions. It was determined that the entropy change as DNA condenses out of solution dominates the DNA-guanidinium interaction. Results also confirmed that the hydrogen bonding is reversible and that the molecular friction of the DNA passing through the functionalized electrodes is strong enough that additional force may have to be applied in order for the DNA to translocate through the nanopore.

For Craig

and

my mom

ACKNOWLEDGEMENTS

I must thank my lab mates, specifically Brian Ashcroft and Shahid Qamar, for their collaboration and commiseration. I also have to thank my advisor, Dr. Stuart Lindsay, for his support and constant optimism and Dr. Peiming Zhang for turning me into a more rigorous organic chemist than most physics graduate students imagine being. Lastly, I am grateful to my friend and mentor, Dr. JoAnn Williams, who always listened and put things in perspective. I was financially supported by the NSF IGERT and the National Institute for Human Genome Research.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

1. Introduction

1.1. Deoxyribonucleic acid (DNA)

A genome is the complete set of hereditary information of an organism.[1] The information is stored in the form of DNA, a polymer composed of a negatively charged phosphate-sugar (deoxyribose) backbone, and nucleobases (bases) (see Figure 1.1). A single human's genome contains 3 billion base pairs. In humans, the genomes of two randomly selected individuals will differ on average by 0.1% or 3 million base pairs.[2] A nucleotide is a unit of DNA that consists of one base, a sugar, and one or multiple phosphate groups. Bases in DNA are adenine (A), thymine (T), cytosine (C), and guanine (G). A and G are purines, consisting of two carbon/nitrogen rings each, and T and C are pyrimidines, consisting of one each. DNA usually exists in its double stranded form; the bases of two different strands hydrogen bond which causes the formation of a double helix structure. One of the features that make DNA such a significant and versatile polymer is its complementarity; A will hydrogen bond with T and G with C, so if one strand of dsDNA is known the second is too. This type of bonding is referred to as Watson-Crick base pairing. The diameter of double stranded DNA (dsDNA) can range from 2.2 to 2.6nm[3, 4], the separation between bases is 0.34nm, and its persistence length (a measure of its stiffness, the length along a polymer beyond which the energetic cost of bending is negligible) is 50nm or about 150 base pairs. Single stranded DNA (ssDNA) has a diameter that is less than 1nm, its persistence length is approximately 0.6nm[5, 6], and, because there is no longer hydrogen bonding and the twisting of the backbone, the distance between bases can be stretched to 0.6nm.[7] It is the sequence of the bases along a DNA strand that make up a gene. A gene determines what proteins a cell assembles,

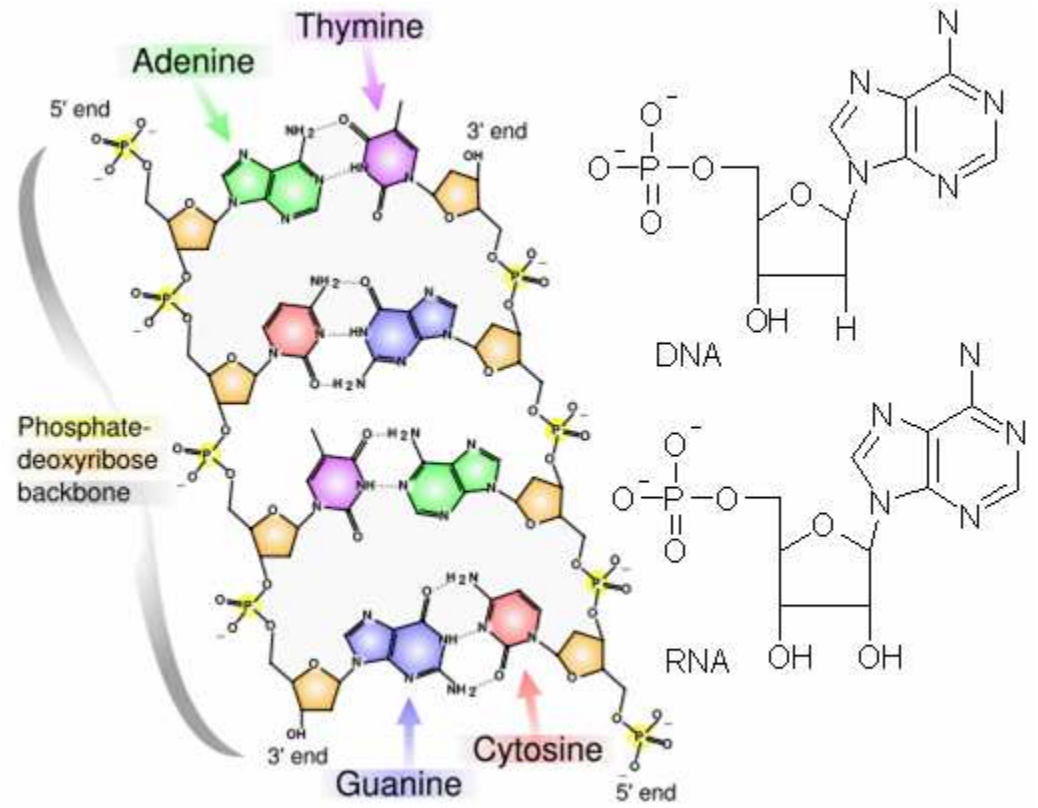


Figure 1.1. On the left is the chemical structure of dsDNA.[8] This image shows two strands under going hydrogen bonding between complimentary nucleobases. On the right are nucleotides of DNA and RNA.

which in turn determines what it's going to do, when it's going to do it and how it will get done. This is the reason why determining the sequence of DNA is important. In chapter two I will address, in more detail, the implications of DNA sequencing.

When a cell requires instructions they are not "read" directly off of DNA, rather the cell makes a complimentary strand of the DNA with ribose nucleic acid (RNA).[1] The major chemical difference between RNA and DNA is the sugar ring. DNA has a hydrogen at the 2' carbon but RNA has a hydroxyl group which leads to its ability to form more complicated structures (see Figure 1.1). RNA also does not use T but replaces

it with uracil (U), which is absent the methyl group. The type of enzyme the cell uses to form the RNA is called a polymerase. Polymerase uses the DNA as a template and builds a complementary strand of RNA as it moves along it. It requires a primer (20 to 30 base single strand of DNA) to hybridize to DNA in order to begin elongating a strand because polymerase needs the 3' hydroxyl group on the sugar of a nucleotide to start its reaction. Exonuclease is an enzyme that can break the bond between nucleotides. Some types of polymerase contain exonuclease and are able to correct any mismatched nucleotides as it moves along a strand of DNA. Later in this chapter I will discuss how scientists borrow these two enzymes from nature in two different DNA sequencing methods.

1.2. The $1000 genome

The United States government invested three billion dollars and thirteen years, from 1990 to 2003, in the Human Genome Project (HGP).[9] Not only did the project sequence a human genome but it is also credited with catalyzing the United States' public and private multi-billion dollar initiative in innovative biotechnology research. As a part of this initiative the National Institutes of Health through the National Human Genome Research Institute issued a call for grants in 2004 entitled "Revolutionary Genome Sequencing Technologies – The $1000 Genome."[10]

The call for grants states that the drive to push the $1000 Genome ahead is fueled by a belief that "genomic data have the potential to lead to remarkable improvements in many facets of human life and society."[10] Cheaper sequencing will magnify the societal benefits obtained from current sequencing methods and will likely create new ones

because of increased availability. The Human Genome Project listed energy sources, environmental applications, risk assessment, bioarchaeology, anthropology, evolution, human migration, DNA forensics, agriculture, livestock breeding, bioprocessing, and molecular medicine as areas in which genome research is or could be used.[9]

One can get a feeling for the possible impact of cheaper sequencing technology by comparing the number of bases sequenced to the total number in the Earth's biosphere. As of 2004 $3x10^{10}$ bases had been sequenced and stored in databases out of the estimated 10^{38} bases on the Earth (based on estimates of total biomass).[11] The ratio between the surface area of a pinhead and the surface area of the sun is a thousand times smaller than the ratio between the amount of sequenced and unsequenced DNA. The method used to sequence the majority of those 30 billion bases is an automated and much improved version of a method described by Frederick Sanger in 1977.[12]

1.3. The Sanger method

In 1975 Sanger and Coulson published "a simple and rapid method for determining nucleotide sequences."[13] They referred to the method as the "plus and minus system" and it relied on two different types of DNA polymerase (DNA polymerase I and T4 polymerase), limiting amounts of ^{32}P-labeled nucleotides, eight different reaction mixtures (one for each nucleotide for the minus method and for each nucleotide for the plus), and two purification steps. At the end of the article the authors boasted "if successfully carried out, it is possible to deduce a sequence of 50 nucleotides in a few days."[13] Two years later, in 1977, Sanger *et al.* returned with an improved enzymatic approach to sequencing that now relied on nucleotides modified to terminate

oligonucleotide growth.[14] They called this method the "inhibitor method" and it is the precursor to the standard now used for DNA sequencing.

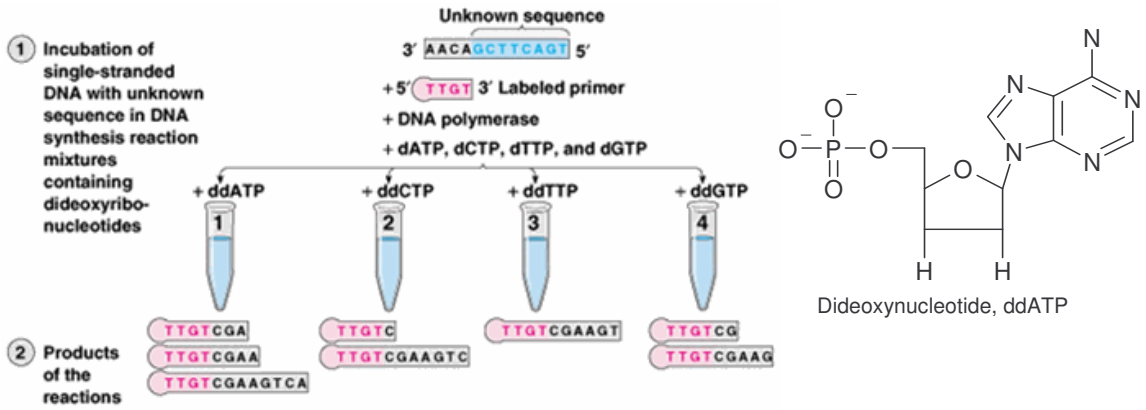


Figure 1.2. An illustration of the Sanger method of sequencing on the left[15] and the structure of a dideoxynucleotide on the right.

The inhibitor method, now known as the Sanger method, relies on dideoxynucleotides that, once incorporated into a DNA strand, terminates said strand because there is no 3'-hydroxyl group on the sugar for the next nucleotide to build from (see Figure 1.2). Basically, the sequence of a short section of the DNA in question has to be known so the appropriate primer can be added, DNA polymerase then starts to build a complementary strand off of the primer, and occasionally the enzyme incorporates a dideoxynucleotide instead of an unmodified deoxynucleotide (a normal DNA nucleotide), terminating elongation for that single strand. A ratio of 1 to 100 of ^{32}P labeled dideoxy to deoxynucleotides is used during strand elongation.[14] Once the reaction is finished there is a mixture of DNA that is complementary to the target DNA differing in length based on the position of the modified nucleotide (see Figure 1.2). Four different mixtures are required to sequence a strand of DNA, each differing by which dideoxynucleotide is

added (ddTTP, ddATP, ddGTP, ddCTP). Gel electrophoresis is then used to separate out the complimentary DNA based on size and the sequence read from bottom to top based on bands in the four lanes (see Figure 1.3).

1.3.1. Improvements to the Sanger Method

This method made it possible to sequence up to 200 nucleotides with one primer.[14] Twenty-five years later, improvements have been made on every step of the inhibitor method, which now allows sequencing of 3 billion base pairs in one month.[16] The preparation of DNA to be sequenced, the polymerase used, the labeling and detection methods, and the separation technique have all been made more efficient and easier to use.

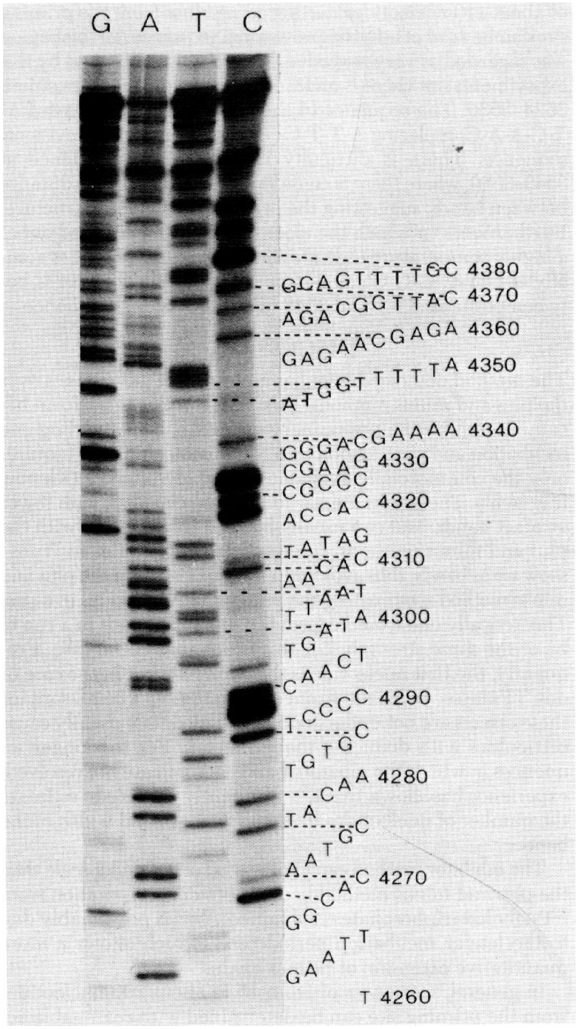


Figure 1.3. Autoradiograph of a gel showing the sequence of a DNA strand determined using the inhibitor method, sequence is read from bottom to top. Each lane is for the four different dideoxynucleotides used.[14]

For example, the DNA polymerase that Sanger worked with was very heat sensitive and required preliminary treatment to remove its exonuclease activity.[12, 13] With the discovery of heat-stable DNA polymerase (Taq polymerase) from *Thermus aquaticus*, a bacteria living in the hot springs of Yellowstone, in 1976[17]

and its subsequent mutation to eliminate a preference for incorporating dideoxynucleotides over deoxynucleotides many problems in sequencing due to the enzymes were eliminated.[12]

After complementary fragments of the target DNA are appropriately extended the mixture must be separated with single base resolution in order for Sanger type sequencing to work. Polyacrylamide gel electrophoresis was originally used for separating sequencing products and its limitations were the greatest barriers to increasing sequencing speed.[12] Gel electrophoresis uses an electric current to drive charged molecules (DNA, RNA) through a cross-linked polymer slab, separating them based on size. The heat created by running the current though the gel limits the amount of current and therefore the speed with which the fragments could be separated.[12] The discovery of capillary electrophoresis in 1983[18], capillary gel electrophoresis in 1990[19], and improvements in gel composition[20] increased separation and readability for sequencing in one run to 1300 bases.[21] Capillary gel electrophoresis is just what it sounds like. A capillary, usually made of fused-silica with an internal diameter of 20-200µm, is filled with polyacrylamide gel and then the sample to be separated is injected into it. An electric field is applied and the sample is driven down the capillary based on size. The larger surface to volume ratio of the capillaries allows for a stronger current to be used which speeds up the separation. It was capillary array electrophoresis[22], however that allowed the human genome to be successfully sequenced within budget and ahead of schedule.[12] An array of capillaries allows multiple samples to be run at the higher speeds at the same time.[22]

Once separated, bands of DNA must be detected in order to read out a sequence. This required new types of labels, labeling techniques, and detection methods. The inhibitor method used ^{32}P as its label and four different samples had to be run on a gel in order to determine a sequence. When fluorescent dye labeling was introduced to sequencing (replacing the radioactive ^{32}P) it became possible to use only one lane to do separation because different fluorescent labels could specify different nucleotides instead of needing four different lanes to do that, allowing the process to become "automated." [12]

Two different methods were designed using fluorescent dyes. One method for sequencing still required four different reaction chambers because the dye was bound to the primer and each dideoxynucleotide mixture used a different dye. After polymerization finished, the four mixtures were combined and run on a single lane and the fluorescence read off the gel giving the sequence.[23] The second method only required a single reaction chamber because different fluorescent dyes, each with a different emission peak, were attached to each of the different dideoxynucleotides.[24] Fluorescent dyes used are constantly being improved in order to eliminate any emission overlap, simplify excitation by requiring only one wavelength, and correct for different run times for different dyes during DNA migration through the gel.[12, 16] The above mentioned primer labeled method can also take advantage of fluorescence energy transfer by incorporating both a donor and acceptor into the primer.[25] A common donor can be used with four different acceptor dyes to further simplify the process.[26] However, synthesizing and doubly labeling a primer for sequencing can add to cost so labeling dideoxynucleotides is preferable.[12]

1.4. Overview of single-molecule methods

An automated and improved version of Sanger's method for sequencing allowed for the Human Genome Project to complete its goal. However, completely new single molecule techniques are being developed that may be able to drastically cut time, reduce reagent consumption, extend read length (the length of DNA that can be accurately sequenced at a time), and increase sensitivity.[27] The goal of these methods is to efficiently and reliably read off the sequence of a single DNA molecule while eliminating many of the wasteful steps of today's sequencing methods; amplifying up the strand in question, synthesizing primers, labeling, and separating fragments.[28] Single molecule fluorescence measurements, atomic force microscope (AFM) force measurements and topographic maps, and nanopore ionic current measurements are all techniques under investigation.[27] The fluorescence technique uses exonuclease to sequentially cleave single fluorescently labeled nucleotides from a tethered complementary strand of the target DNA. The released nucleotide flows into a low volume (picoliters) chamber where it is excited and then its emission detected (see Figure 1.4).[29] The rate of the enzyme and the need to repeat the measurement enough to reach 99.9% accuracy have kept this technology from passing capillary array electrophoresis techniques in speed of read.[27]

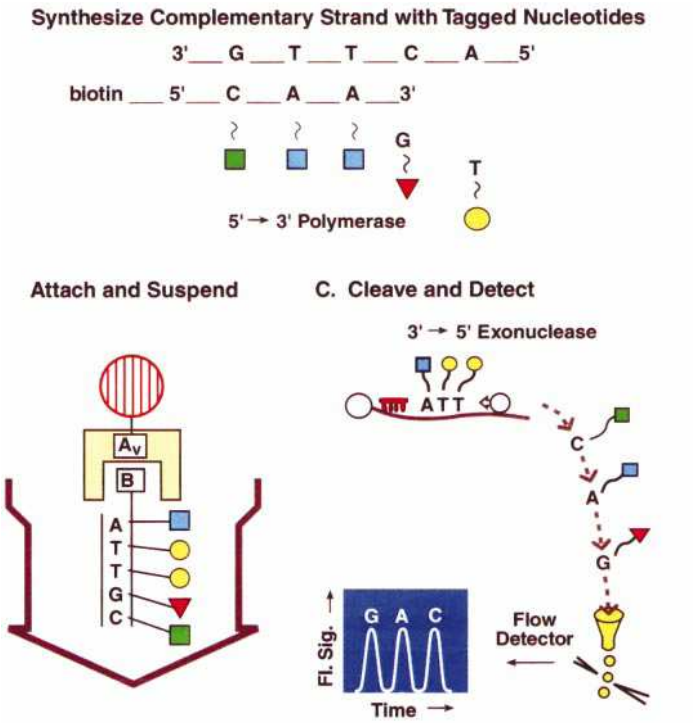


Figure 1.4. Schematic of exonuclease method of sequencing.[12]

In 1995 a topographical scan of flat lying DNA using AFM (in later chapters I will discuss the AFM in greater detail) with a Si tip showed the helical structure of double stranded DNA (dsDNA).[30] The use of carbon nanotubes as tips improved the resolution of the AFM scan by 70%, enough to distinguish between a biotin-streptavidin or fluorophore label on DNA.[31, 32] However, this technique has yet to directly scan a strand of DNA in order to determine its sequence.[27] The AFM (which is capable of measuring a few pN of force[33]) is also being used to measure the force required to pull apart dsDNA. The hope is to eventually be able to determine a sequence by unzipping DNA and measuring the breaking of hydrogen bonds between individual base pairs.[34] The best base resolution to be reached through unzipping has been 500 bases, quite a bit

more than the necessary single base needed.[27] The use of nanopores to determine DNA sequences appears to be the most promising of the single molecule techniques and I will now review the advances that have been made and the obstacles still blocking its progress.

1.4.1. Nanopore sequencing methods

In 1970 Hladky and Haydon measured the change in ionic current across a lipid membrane in the presence of the antibiotics nonactin, nystatin, and gramicidin. They saw an increase in conductance across the membrane that they thought could have been caused by ions being carried across the membrane by nonactin or due to the formation of a pore in the membrane by nystatin and gramicidin.[35] In the 1990s researchers started considering the use of nanopores in DNA sequencing.[36, 37] The idea is to cause DNA to thread through the pore using an electric field (similar to how DNA is driven through a gel) which will cause a drop in current (because fewer ions can flow through when the DNA occupies the pore) that could be correlated to sequence information. In order for the technique to work a pore has to be chosen that is wide enough to accommodate a single strand of DNA but narrow enough that the strand can not form secondary structures; it also needs to be stable in the same pH range as DNA and at large ion concentrations. Biological pores were first investigated and α-hemolysin from *Staphylococcus aureus* (the bacteria used the protein to create pores in targeted cells, killing the cell) fit those requirements (see Figure 1.5).[38] The next step was to see if single stranded DNA (ssDNA) could flow through the pore and if its passage could be detected.

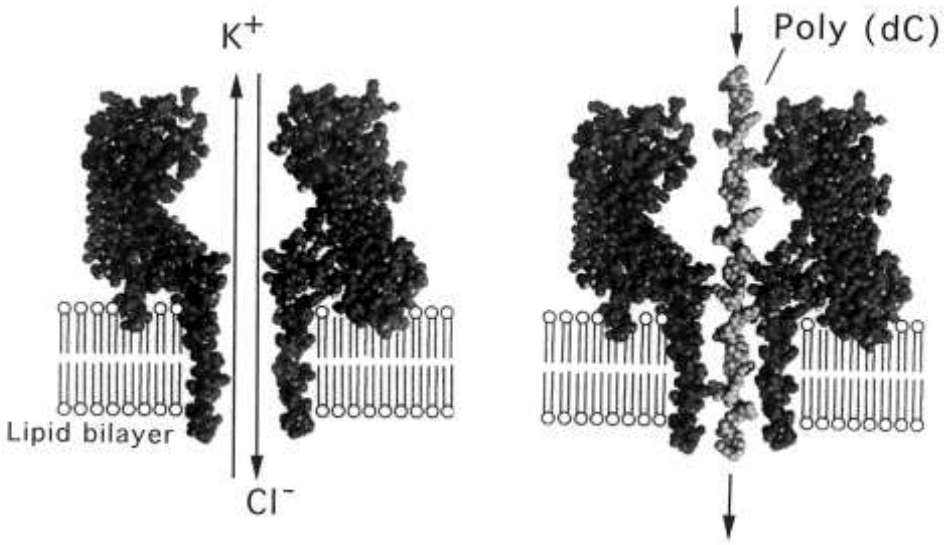


Figure 1.5. On the left is an image of α-hemolysin embedded in a lipid bilayer with ions flowing through the pore, on the right is DNA flowing through it.[36] The narrowest point in the pore has a diameter of 1.5nm, the length embedded in the bilayer is 5nm.[36] The top is called the cis side and the bottom the trans.

Kasianowicz *et al.* performed a number of experiments with α-hemolysin that proved ssDNA could translocate the pore and its movement was detectable.[37] They used short (75 to 430 nucleotides) single stranded homopolymers and heteropolymers, 100-mer dsDNA, and DNA with a mixture of single and double stranded regions. The solution on the trans side (the side the DNA is driven to) of the pore was collected after experiments and PCR followed by gel separation and staining was used as independent proof of pore translocation. Not only did they prove that ssDNA could pass through the pore, thereby momentarily partially blocking ionic flow and causing the current to drop, they also were able to show that dsDNA could not follow it. They also determined that the number of blockages detected was proportional to the concentration of DNA used, the length of the blockage was proportional to the length of the strand and inversely proportional to the

applied potential, and that the direction the DNA passed through the pore (5' to 3' or 3' to 5') also affected blockade lifetime.[37]

Once it was shown that ssDNA could indeed flow through α-hemolysin and be detected doing so, runs of different homopolymers of RNA were used to see if the pore could distinguish between purines (A and G) and pyrimidines (T, U, and C).[36] There was a detectible difference in blockade current (the current from the decreased amount of ions that are still able to flow through the pore with the DNA) between 100 to 500 nucleotide length polyA and polyC (~85% blockage for poly A, ~95% for polyC).[39] Additionally, there was a lifetime difference between polyU and polyA even though blockage current was similar, possibly because polyA will form a helical structure but polyU does not, translocating the pore in an extended state.[39] When a single strand of DNA with stretches of A and stretches of C ($A_{30}C_{70}$) was forced through an α-hemolysin pore the blockade current initially dropped to 95% (polyC section) and then jumped up to 85% blockage of current (poly A section) (see Figure 1.6).[39]

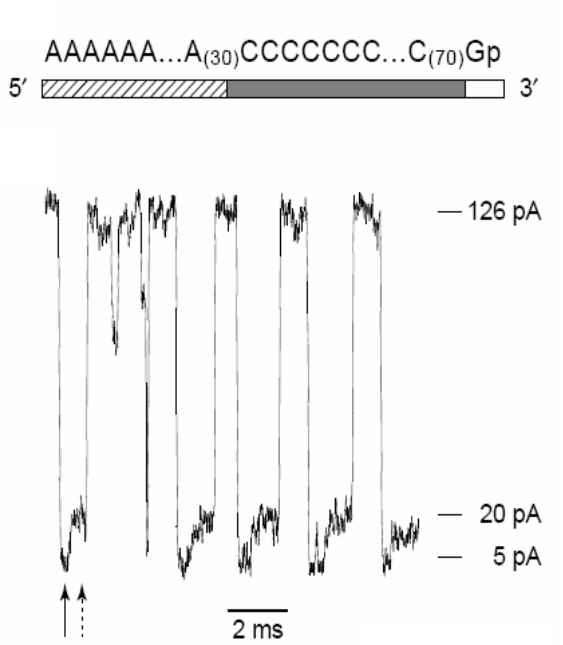


Figure 1.6. Blockage current resulting from the heteropolymer $A_{30}C_{70}$ translocating α-hemolysin. The solid arrow points to the 95% drop in current caused by the poly C section, the dashed arrow indicates the current blockage from poly A.[27]

All of these results are encouraging for nanopore sequencing but none show that α-hemolysin is capable of distinguishing between single nucleotides in a DNA strand while it is moving through the pore. Two major problems are the translocation speed and the width of the pore. When DNA enters the pore it keeps the ions in solution from passing through resulting in the drop in current. There is only a 100 ion difference between a purine or pyrimidine as they translocate.[36] If speed of translocation could be slowed, by decreasing the electrophortic current (but if it is not strong enough DNA could reverse direction ruining the measurement) or placing "brakes" in the form of short strands of DNA the target could hybridize to, then it may be possible for that small difference to be lifted out of the noise. Unfortunately, because of its length α-hemolysin does not accommodate one nucleotide at a time, but 10 to 15 which means even with a slower speed multiple nucleotides would contribute to the blocked current, not the single nucleotide needed for sequencing.[27] Work is being done to improve the pores used for sequencing by modifying proteins (by mutagenesis or chemical synthesis) and by building nanopores in solid state membranes.[27]

Solid state nanopores have many advantages over the ones nature provides. As fabrication techniques have improved it has been possible to build nanopores with diameters as small as 1nm.[40] These pores can operate in a larger pH range, under stronger electric fields, are chemically more stable than the ones made of protein, withstand greater mechanical strain, and can be integrated into devices.[40, 41] The first steps in solid-state nanopore sequencing were to repeat the work that was done with α-hemolysin. Work being done in 2001 proved that DNA could move through a

manufactured pore but the pores were too big to discriminate between ssDNA and dsDNA.[42] By 2006 it became possible to make pores small enough (~2nm) to block dsDNA while allowing ssDNA to translocate the nanopore.[43] It was also found that blockade lifetime is related to the length of DNA translocating the pore.[44] Experiments using long (6kbp to 10kbp, 2μm to 3.4μm) dsDNA and a nanopore with a 10nm diameter and 20nm length determined that the relationship is not linear. The effect of drag on the untranslocated DNA (a random polymer coil that can be approximated as a sphere with a radius equal to the radius of gyration for dsDNA) dominates the electrophoretic force driving the DNA through the pore resulting in the lifetime (τ) increasing with length (L) by $\tau = L^{\alpha}$, where α is experimentally found to be 1.27.[45]

However, unlike α-hemolysin, experiments in 2004 showed an increase in current when dsDNA translocated a nanopore instead of the decrease expected when DNA blocks ion flow.[46] The authors used dsDNA with a length similar to the length of the pore and 0.1MKCl salt concentration (less than the typical 1M used in nanopore experiments[40]). The material used to make the nanopore has a negative surface charge at the pH conditions used (pH 7) which resulted in K^+ ions gathering at the surface interface. When dsDNA enters the pore the bulk flow of K^+ and Cl^- ions is blocked resulting in a drop in current but, at the same time, because of DNA's negatively charged backbone, the number of interfacial K^+ ions increases as they are carried in surrounding the DNA. This increase results in an overall increase in current flow when the DNA is in the pore which then drops back down once the DNA exits on the trans side (see Figure 1.7).[46] Later experiments showed that the concentration of KCl affects the type of

response DNA translocation gives and that 0.4M is the point where the interfacial current increase is completely cancelled by the bulk current decrease, resulting in an overall ionic current drop.[41] Another set of experiments, in 2006, determined the amount of force that acts on DNA as it is being forced through a pore by an electric field. By tethering one end of DNA to a bead in an optical trap and then applying voltage in order to thread a nanopore with the other end it was determined, through the bead displacement, that the DNA is acted upon by 0.23pN/mV.[47]

Despite advances in solid state nanopore production allowing for 1nm sized pores and increasing control over length, single nucleotide resolution has yet to be achieved.[40] The biggest problem is still translocation speed, the DNA is moving to quickly through the pore to pick up changes in ionic current due to individual nucleotides. Similar to biological pores, groups are working on modifying solid-state nanopores in order to overcome that problem. Possibilities once again include adding short strands of DNA to the pore and relying on hybridization to slow down the DNA and using optical tweezers (similar to the set up used in determining the applied force on DNA) to oppose the applied voltage in order to control the rate of translocation.[40] Integrating nanoelectrodes into the pore or changing the material the pore is made from in order to measure changes in capacitance across the pore as DNA move through it are also possibilities being examined to make solid-state nanopore sequencing devices.[40]

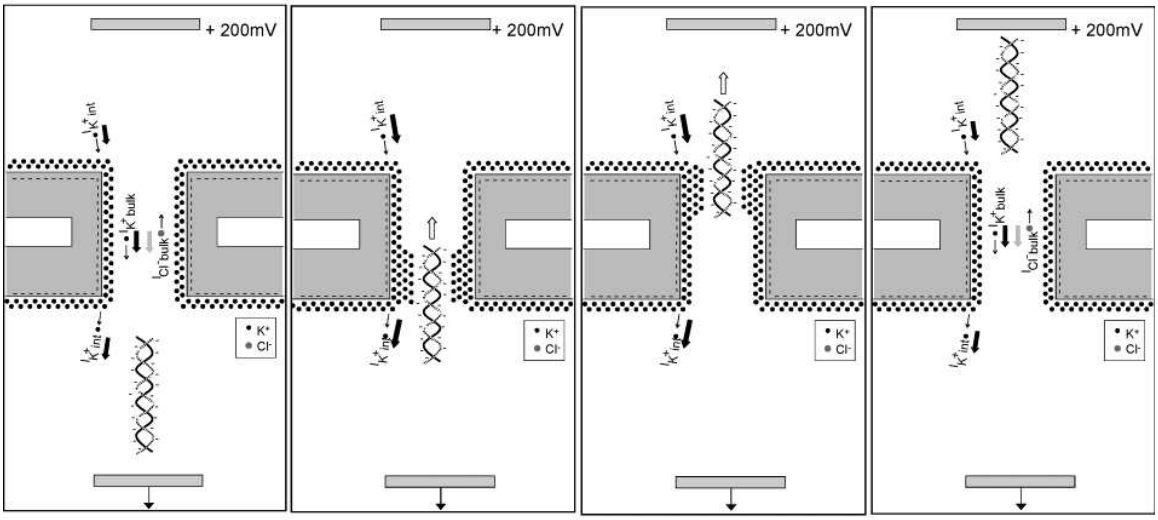


Figure 1.7. Schematic of KCl ions and DNA movement through a solid-state nanopore.[46]

1.4.2. Nanopore sequencing with AFM force spectroscopy

The technique we developed uses a nanopore combined with the low force sensitivity of AFM in order to try to sterically read off the sequence of a single strand of DNA. The general idea is to covalently bind one end of a single strand of DNA with a nanopore (β-cyclodextrin) already threaded on it to a self-assembled monolayer designed to minimize non-specific adhesion, grasp the nanopore with a functionalized AFM tip, and pull it up over the DNA while measuring the deflection of the tip as the pore clears individual bases (see Figure 1.8). Similar to other single-molecule, nanopore sequencing methods this technique has the advantage of reducing time, increasing read length, it does not require amplifying the sample, there is minimal sample preparation, and no need to incorporate labels. It differs from the above mentioned techniques in that the DNA is held in place while the nanopore is moved by the AFM, reducing the chance of DNA reversing direction which is a concern with voltage driven translocation. This means the speed of sequencing and the length to be sequenced is dependent on the range of motion

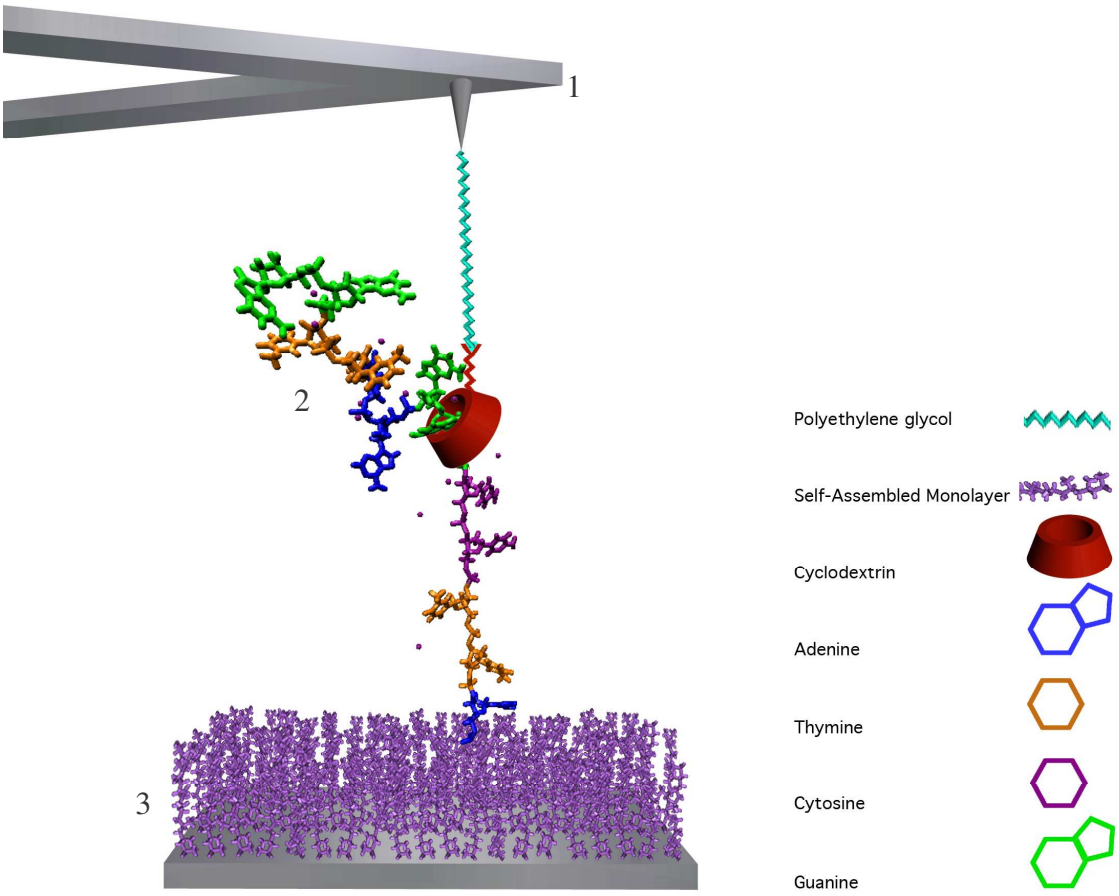


Figure 1.8. Schematic of our DNA sequencing set-up. Cyclodextrin, bound to an AFM tip (1), is being pulled over DNA (2) that was first tethered to a surface (3).

of the AFM cantilever and the speed it can pull the pore. Another advantage is that during sample preparation the nanopore is threaded around the DNA which eliminates the need to wait for the DNA to find the pore and for the pore to capture the DNA. The AFM also allows one to measure both the force being exerted on and the location of the nanopore at the same time. The nanopore we chose has a smaller limiting aperture than

both α-hemolysin and any of the solid-state pores mentioned earlier and, most importantly, a length on the order of the separation between bases of ssDNA.

This technique required the development of an appropriate nanopore (modified β-cyclodextrin), a method of threading the pore onto the target DNA, a surface that allows for tethering of one end of the DNA while preventing adhesion of the entire strand, modified AFM tips designed to bind to the cyclodextrin, and an AFM system sensitive enough to detect minor changes in tip deflection. Each of these requirements will be discussed in detail in later chapters.

2. Implications of and Preparation for the $1000 Genome

The goal of nanopore sequencing using the AFM, like all $1000 Genome projects, is to drastically cut the cost and time needed to sequence a genome. The implications of a method that could accomplish sequencing a genome for $1000 range from beneficial personalized medicine to undesirable genetic discrimination. In this chapter, I intend to address some of the positive and negative consequences of a $1000 genome technology (which my own research is working towards) and how prepared this country is for it.

As I stated in Chapter 1, the U. S. government invested three billion dollars and thirteen years, from 1990 to 2003, in the Human Genome Project (HGP).[9] Additionally, the HGP aimed to determine the 20,000 – 25,000 genes, the biologically functional modular regions of the DNA sequence, contained in that information.[9] The average cost of the project was $1 per base pair. This cost included not only the sequencing of DNA, but also included studies of human diseases and experimental organisms (bacteria, yeast, worms, flies, and mice), development of new technologies for biological and medical research, development of computational methods to analyze genomes, and investigation into the ethical, legal, and social issues related to genomics. In 2001, as infrastructure for sequencing was built and improvements were made to the techniques used, the cost was reduced to a nickel per base pair, or $150 million per genome.[9] The company 454 Life Sciences (Branford, CT) claimed, in 2007, that they could sequence an entire human genome in two months for $1 million.[48] As the push for cheaper sequencing continues, the NIH introduced the challenge for the "$1000 Genome." The federal government committed $6 million in 2004, $5 million in 2005, and $2 million in 2006 and 2007 and has pledged $5million in 2008 to the granting effort.[10] The term "$1000 Genome" is a

useful, catchy phrase to capture the goal of the proposal - reducing the costs of sequencing by at least four orders of magnitude within ten years.[10] Therefore it is the issue of cost that separates today's sequencing methods from the ones that the government hopes will be inspired by the push for the $1000 Genome.

Before the cheap sequencing can become a reality the fields of comparative and functional genomics must be further extended to utilize existing sequence data and current sequencing techniques.[49] Comparative genomics looks at the genomes of different species in an attempt to understand evolution at the genetic level. It is also used to help identify genes.[50] Once the genes are identified, functional genomics uses a number of techniques to determine the biosynthetic or regulatory role of a particular gene.[51] Without a better idea of where genes are located and what they do, the benefits promised by the $1000 Genome technology will be difficult to realize.[49] Cheap sequencing for any of the purposes I mention would be useless without the ability to translate the string of bases into their biological function.

2.1. Benefits of Sequencing

Cheaper sequencing will increase the fields of study where sequencing can be utilized and expedite research in fields where DNA sequencing is currently being used. I will touch on the areas which will be improved by the $1000 Genome, but am most interested in the qualitatively new issues that are unique to cheaper sequencing, rather than just improvements on what is already present. As I mentioned in Chapter 1, HGP website lists a number of areas where in genomic research is or could become valuable. These areas include energy sources, environmental applications, risk assessment,

bioarchaeology, anthropology, evolution, human migration, DNA forensics, agriculture, livestock breeding, bioprocessing, and molecular medicine.[9] The term risk assessment as used above applies not only to preventative medicine based on genome information, but also to the genetic variability in the human population which can be correlated with susceptibility to radiation and toxic compound exposure. The $1000 Genome will help in these areas by simply allowing for more sequencing data to be collected more quickly.

Some of the benefits listed above are not directly related to the sequencing of the human genome but rather to the use of the technology on other organisms. The U.S. Department of Energy's Microbial Genome Program either has sequenced or is working on sequencing well over one hundred microbes.[52] The Genomes Online Database (GOLD), an online database of current and finished sequencing projects around the world has information on 1,951 diverse genome projects.[52] The DOE claims that bacterial genomes will be useful in "energy production, environmental remediation, toxic waste reduction, and industrial processing."[52] Current sequencing methods are essential in studying these fields, but the $1000 Genome will more easily allow for larger numbers of subjects to sequence and study. This is also true for the fields of bioarchaeology, anthropology, evolution, and human migration.

2.1.1. Benefits of Cheap Sequencing

Beyond the general benefits of sequencing listed above cheap sequencing will open up new possibilities for research, medicine, and consumer products and services. The largest effect on the public may be the ability for individuals to easily obtain and store their entire personal genome sequence information.[49] For example, parents could have

their baby's DNA isolated, sequenced, and stored at birth or adults could make the decision for themselves. This stored information could then be easily accessed for testing as more genes are found within sequences. Additionally, a doctor could prescribe preventative medicine or measures based on a person's genetic susceptibility. The doctor could access the stored sequencing data, have a program scan the information, and then act based on the patient's disease probabilities. This may also become a very empowering technology. Individuals may be able to do their own basic analysis of their own genomes. There already exists web-pages that allow people to trace their ancestry genetically.[53] Individuals could download sections of genomes for online paternity tests, dating services could determine a couple's genetic compatibility, and cosmetics and perfume companies could create tailor-made products based on their client's DNA.

Another major medical benefit of fast, cheap sequencing would be in the field of pharmacogenetics. Pharmacogenetics is defined as the heritable component of variation among individuals with respect to positive response or adverse reaction to drugs.[11] The availability of personalized genomic information could allow production of custom made drugs based on an individuals' unique biology.

Currently only a few diseases, such as hereditary nonpolyposis colorectal cancer, cystic fibrosis**,** Tay-Sachs disease, spinal atrophy, and muscular dystrophy, are readily diagnosed through sequencing. The difficulty of diagnosis is related to the number of genes involved in the disease, how far apart mutations are in the succession of base pairs, and how easy it is to link the phenotype or physical appearance and constitution of a genetically inherited feature with the sequence region responsible for it, the genotype.

Disease genes are easier to find if the disease results from a single gene mutation than if multiple genes are involved. Finding all the genes in a genome that are responsible for a disease is like trying to find a collection of needles in a haystack without knowing in advance how many needles there are. It is believed that by sequencing entire genomes from multiple people in an affected population, possible disease-linked genes will be easier to discover.[11] A similar approach can be used in cancer research; comparing the genomes of healthy and cancerous cells will allow researchers to better characterize the molecular-level effects of cancer-inducing mutations.

Beyond the diagnosis of diseases which are purely the result of genetic mutations, such as Huntington's disease or sickle cell anemia, physicians would be able to prescribe preventative steps for patients whose genetic predisposition to illness could be mitigated by such steps. In most cases, a specific mutation of a gene only changes a person's probability of contracting an illness. Most people tested today have a family history of a specific disease.[11] For example, before testing for BRCA genes which have been linked to breast cancer, most women in the U.S. undergo genetics counseling and, based on her family history of breast cancer, the counselor decides whether genetic testing is necessary.[54] Rather than sequencing an entire genome, current genetic tests look for a specific mutation by examining, and in some cases sequencing, a small region of a person's genome.

The number of such tests for genetic diseases has been increasing. In 1999 only approximately 600 tests existed either clinically or in a research environment. By 2002, the number had increased to 900.[55] As of 2007, there were over 1000 genetic tests.[41]

The hope is that with the $1000 Genome a patient's entire genome could be examined, probabilities determined, and actions taken to prevent or mitigate disease.

2.2. Challenges of Cheap Sequencing

While there may be many positive effects for society resulting from cheaper sequencing, there are also many ethical issues that must be addressed. There is a fear and a realization among scientists that biotechnology is moving farther ahead of biology, ethics, and common sense.[56] Easy and cheap sequencing may potentially lead to any individual's genetic information being readily available. This will most likely exacerbate existing issues with current sequencing technology such as genetic discrimination and may also create new problems.

How a specific individual's sequencing information is used and by whom is currently a major concern that may become a more serious problem with the $1000 Genome. Genetic discrimination hiring and promotion practices in the workplace is a real possibility. A study performed by the American Management Association in 2001 claimed that 65% of major U.S. firms require medical examinations of new hires and 34% require it of current employees.[55] The survey also showed that employers use the results of medical exams when they make hiring decisions. While in most cases genetic testing is not a standard part of the exams, cheaper sequencing could make genetic tests more attractive. It is not surprising, therefore, that 63% of people surveyed would not undergo genetic testing if it were possible that an employer or insurance company could access the information.[55]

According to a poll by *Time*/CNN 75% of people "would not want their health insurer

to have information about their genetic profile."[55] An example of this fear of discrimination by insurance companies is a policy in many genetics clinics which stipulates that no genetic testing information be kept in a person's medical record. However, as Shobita Parthasarathy points out, a "reluctance to include genomic information . . . could impede continuity of care when primary care physicians lack information about potentially significant genetic test results."[54]

Additionally, how or whether there should be counseling provided when an individual learns their genetic information must be decided. A personal genome sequence can often only give people information on their probability of contracting a gene-linked disease, so the presentation of that information, along with caveats about testing accuracy, must be considered. Similarly to the way Macbeth misinterpreted the witches' prediction for his future, a patient may know that their genome holds information about their possible medical future but they may misinterpret what they hear to their own detriment. Issues may also arise because of all the information contained in a person's genome. Once sequencing is completed the patient may not want to know if he or she will contract a disease that is inevitable and appears later in life. Methods of prevention, such as a restricted diet, may be unacceptable to the patient so they may choose not to know probabilities about specific diseases.[49]

Genetic counselors are people who are trained in genetics and counseling. Their job is to help patients understand the uncertainty in genetic tests, give them information on what their results may mean, and help them to make informed decisions. As the number of genetic tests and the number of people requesting the tests increase, the need for

genetic counselors will only increase. However, Dr. Beverly Yashar, director of the University of Michigan's genetic counseling training program, feels that there are not enough counselors. She worries that, "As the pace at which genetic testing is being developed continues to outstrip the number of genetic counselors who are being trained, how are we going to deal with the fact that more and more genetic tests are going to be ordered by primary care physicians, who may or may not have the time or expertise to deal with the complexity of the tests?"[57]

This issue extends to future parents' knowledge of fetal genetic information and the decisions that they may face. With genetic testing those decisions are becoming less personal. As an example, the American Society of Human Genetics proposed legislation in 1990 concerning restrictions to abortion being considered in many states. The society wanted pregnant women to have the right to terminate a pregnancy if it was likely the fetus had a genetic disorder, such as Down Syndrome, or if they were at a greater risk of having a child with a serious genetic disorder for which precise prenatal diagnosis was not available.[58] In Europe, geneticists "significantly influenced legislators establishing limits within which abortion would be at all permissible."[58] The geneticists wanted to make sure that the time limit for legal abortion was set so that genetic test results from amniocentesis were available.[58]

To borrow a phrase from Lori B. Andrews, there are also ethical concerns centered on the "nonconsensual, undemocratic impacts of these technologies."[2] She was concerned with the future prospect of genetically designing children and the possibility that plants, animals, and human beings may someday all be partly human-made.[2] The technology to

genetically alter a person will likely only be available to the rich at first but eventually everyone will feel pressure to design their children to keep pace with a rapidly changing "normality." This would also be nonconsensual for the future person in the sense that the procedures would be performed at the embryo stage. The possible future human being has no say in whether they are tall enough to play professional basketball or smart enough to be the next Einstein.

Beyond the concerns about future genetic engineering of humans there are serious issues with the availability of sequencing procedures and resultant products, and consent with regards to who, when, and what is sequenced. While $1000 for a complete genome sequence seems relatively cheap by U.S. standards, the World Health Organization reported that, as of 2004, 83% of the 192 countries they have data on spend less than $1000 per person per year on total healthcare expenditures, this includes both government and private expenditures.[59] For the 44 countries that spend less than $100 per person per year, is a $1000 for a personal genome sequence even a possibility?[59] Additionally, genetics research is strongly tied to commercial development. In 1980, the US Supreme Court decided that a "live artificially-engineered microorganism" is patentable.[60] U.S. patent law is also being interpreted such that an isolated version of an existing human gene may be eligible for patenting.[61] Gene patenting does allow for researchers to be rewarded for their work, decreases the chance of redundant effort, and the disclosure required by the patent process ensures access to the information.[62] Unfortunately, patenting of DNA could also cause access to new treatments to be restricted.[63] This concern extends to non-human biotechnology products such as

genetically modified seeds which may be made too expensive for the poor.[63]

A patent submitted in England to genetically engineer mammals to produce drugs in their milk tried to extend its coverage to human women because of the belief that eventually "someone, somewhere may decide that humans are patentable."[2] This kind of thinking could be the beginning of what Abby Lippman calls the "geneticization" of society.[58] As our genetic testing becomes more widespread and our DNA becomes a more important part of our identities how might our society change? The $1000 Genome will make diagnosing a person's genes the easy fix. Does this mean that in the future a woman will know her chances of contracting breast cancer or heart disease but the biggest challenges to women's health - violence and poverty – which may not lend themselves to genetic solutions, will be neglected even more?[58]

The argument of geneticization is also related to accessibility. If the medical industry becomes more reliant on a person's sequence to diagnosis, treat, or prevent disease then cheap sequencing should be provided universally regardless of ability to pay, as is now the case with disease screening for newborns.[49] Accessibility at the prenatal stage carries its own set of issues. The question is not whether women will be able to have the tests performed, but what they will do with the information. The looming specter of eugenics is certainly present, but another possibility is that providing too much information can actually limit women's (or parents') decision-making ability. As an example, it is currently mandatory for hospitals to test for phenylketonuria (PKU), a genetic disorder that can cause mental retardation, after a child is born. The disease can be treated with a special diet but only four states require insurance companies to pay for

the special food required. How much choice will poor women, who know they cannot afford the necessary food, have if PKU testing becomes possible at the prenatal stage?[58] They may feel pressure to terminate the pregnancy rather than risk having a mentally retarded child.

The above listed concerns arising from the $1000 Genome are all issues that will eventually have to be faced. However, I believe that privacy matters, including consent, storage, and the ownership of sequence information or tissue samples, will be the first major problem for the $1000 Genome. This is because, as I will show, very little has been done to protect people from misuses of current sequencing technology. Additionally, the majority of concerns are related to fear about the control of sequencing information.[49] The California HealthCare Foundation believes that people worry that individuals or agencies will obtain a copy of a person's sequence and use it to cause "the loss of insurance or employment, having a mortgage called in or denied, or having genetic information used in child custody disputes or personal injury lawsuits."[55] These are problems to which traditional sequencing is also subject, but it is the future widespread use of cheap sequencing that makes them particularly relevant to the $1000 Genome.

Something as seemingly simple as the storage of information generated from cheap sequencing is more complicated than it appears. In order to sequence a person's genome a tissue sample must first be collected. Once the information is extracted does the tissue sample have to be kept and stored? If further testing is required, the person would probably be able to give another sample. However, problems may arise in situations

where additional collection would be unfeasible. Storage of the physical sample is not the only problem. Another problem is who retains the electronic media on which the genetic information is stored. Does the customer/patient/client/suspect/victim take home the only copy of their sequence or does the institution that did the sequencing have a right to store the information? These issues are tied to people's fear of discrimination and "the ease of clinical testing would have to be weighed against the risks that other persons would obtain unauthorized access to a person's DNA."[49] How does ownership of genetic information relate to the criminal justice system? If an individual had her genome sequenced for medical purposes do law enforcement agencies have the right to use that information rather than collecting their own sample from the suspect? Once a sample is done being processed for criminal justice purposes does the tissue belong to a government? If so, can it be sold to a pharmaceutical company? This highlights another major problem with cheap sequencing. Does the individual have the right to insist that her DNA sample or sequence information be used or manipulated only with her consent?

There has already been a legal ruling on informed consent pertaining to the use of cells collected during a medical procedure. After a Washington man named John Moore had his spleen removed as part of a cancer treatment, he discovered that his doctor used cells which had been collected from the spleen to produce a profitable cell line. Moore sued in *Moore vs. Regents of the University of California*, because the doctor had made plans to produce the cell line before the operation had taken place and never informed Moore. The court ruled that doctors need the consent of a patient if they plan to use cells which were originally removed for another reason for research.[11] Although the patients

consent is required before using them, property rights were not extended to the cells, which implies that if a patient gives informed consent for tissue removal they also sign away the "right to information that is derived from the biological material itself."[11] At this point the discussion is about people with the ability to give informed consent. How will the $1000 Genome affect situations in which consent is impossible? For example, there is speculation that sequencing will first occur in utero or shortly after birth and become part of a person's medical records throughout their life.[49]

In 2005 Francis Collins, director of the National Human Genome Research Institute, projected that "a $1000 complete genetic screening for any person remains a realistic goal with the next decade."[64] At its inception, a portion of the HGP budget, 3-5%, was set aside to study the "ethical, legal, and social issues (ELSI) surrounding [the] availability of genetic information."[65] Additionally, a joint NIH-DOE task force was created in 1997 because of concerns over the quality of genetic testing.[66] While there is interest in determining and understanding the impact of sequencing on our society there has been only a scattered attempt at protecting people from its negative affects.

2.3. Protections

Among some of the protections that currently exist is the federal Health Insurance Portability and Accountability Act (HIPAA) of 1996 which attempts to address some issues of genetic privacy. Its purpose is to safeguard health information that is either created by or given to private healthcare providers.[55] The term "health information" does include "genetic information that otherwise meets the statutory definition" as defined by the U.S. Department of Health and Human Services.[55] This protection

means that written consent is necessary before healthcare providers that fall under HIPAA regulation can use or disclose genetic information for treatment, payment, or other health care purposes.[55] Additionally, in 2000 President Clinton signed an executive order "prohibiting every federal department and agency from using genetic information in any hiring or promotion."[67] Beginning in 1990 the U. S. congress has tried to pass genetic nondiscrimination laws.[64] Ten other bills have been introduced of the course of 16 years.[64] They have all been primarily concerned with genetic discrimination in employment or health insurance coverage and limiting the ability of employers or insurers to require genetic testing of their clients or employees.[68]

At the state level, genetic anti-discrimination laws have been passed, but according to the Human Genome Project's own research findings "none of them are comprehensive."[67] Some type of coverage exists in 26 states, and only 14 states "require informed consent before a third party can . . . obtain genetic information."[55] The laws vary in coverage, protections, and methods of enforcing the law.[67]

Reinterpretations of existing laws are also being used to try to protect people from infringements on genetic privacy. The Americans with Disabilities Act (ADA) of 1990 may be interpreted to cover genetic discrimination, but its scope is limited. In 1995 the Equal Employment Opportunity Commission interpreted ADA specifically for the case of genetic discrimination however the HGP feels that the interpretation "is policy guidance that does not have the same legal binding effect on a court as a statute or regulation."[67]

In April 2007 the Genetic Information Nondiscrimination Act of 2007 (GINA) passed in the House with only 3 votes against it.[64] President Bush has openly supported GINA

and will sign it into law, if given the opportunity.[64] The bill, if approved by the Senate, would succeed in granting uniformed protection to all Americans against genetic discrimination by insurers and employers. GINA protects individuals from employers, employment agencies, labor organizations, and job training programs that discriminate in hiring, firing, and other employment decisions based on genetic information.[69] Genetic information extends beyond an individual's genetic test results. It includes family members' results, the occurrence of disease or disorder in the family, and the receiving of genetic counseling by either the individual or family member.[69] Insurance companies would not be able to require genetic tests; GINA makes sure not to limit a "health care professional . . . from notifying an individual about genetic tests or providing information about a genetic test" if it is part of a wellness program.[64] Additionally, insurers could not use genetic information to underwrite insurance policies.[64] Senator Tom Coburn, R-Okla., placed a hold on GINA in August 2007 which has stopped all debate because the bill does not protect embryos and fetuses from discrimination that have undergone genetic testing.[70]

Collins and James Watson have openly advocated for a federal nondiscrimination law. They point out that "all of us carry dozens of glitches in our DNA sequence, yet no one should be denied a job … [or] should be denied health insurance because of predispositions found in their DNA."[71] They make the additional case that, without protection, people will not take part in genetic research out of fear of discrimination which will slow medical advances and scientific research. John A. Robertson, law professor at The University of Texas School of Law at Austin, echoes their sentiment.

He claims that there will have to be "strong protection [of] the rights [of] persons who are genotyped" before individuals are comfortable with sequencing.[49] Even according to HGP, the first major beneficiary in sequencing research, the legal protections listed above are inadequate. There are many ways in which a person's genetic information could be revealed regardless of laws restricting employers and insurance companies from performing the sequencing themselves. For example, individuals could simply be required to give the information or sign a consent form so that a third party could release the information to the company.[55] An individual may also have to disclose some or all of their sequence to receive workers' compensation, disability accommodation requests, paid or unpaid sick leave, or leave due to a family member's medical condition.

Another major problem with existing legislation is a lack of tissue ownership protection. The Health Insurance Portability and Accountability Act offers no protection or regulation of the "actual tissue or blood sample that generated the genetic information."[55] There is a need for "distinguish[ing] between the physical embodiment of the genome in DNA and its informational content."[49] These laws will also have to address DNA discarded by an individual when, for example, they drink from a cup or smoke a cigarette.

Ownership issues are closely related to issues of informed consent. The HIPAA allows companies that collected sequencing information for health purposes to give that information to law enforcement agencies without informing the patient.[55] There are also many institutions that do not have to comply with the Act's privacy provisions. They include employers, pharmaceutical companies, pharmacy benefit managers, workers'

compensation, life, and disability income insurers, and many genetics researchers.[55] It has been suggested that informed consent be required at every stage in acquiring sequencing information. This would include consent at collection of DNA, sequencing it, and testing it for mutations or known genes. A person would have to grant additional consent before the information can be stored, disclosed to another party, or used for purposes beyond those originally agreed upon.[49] This would eliminate the whole genome sequencing of babies, embryos or fetuses because only the individual whose DNA is being sequenced could give consent.

Storage is another issue that could lead to either another level of protection or make genetic information more vulnerable, depending on how storage of the sequence information or the physical sample (tissue, blood, or isolated DNA) is regulated. There are multiple ways in which storage could be controlled. The simplest may be to require the destruction of the physical sample once sequencing is complete. This still leaves all sequencing data, which contains the information people may not want their employers or insurers to see. Robertson believes that while electronic storage of a person's sequence may have many positive affects on their health care, keeping that information with other medical records would be the worst method of protection.[49] Other methods of storage could include a third party entrusted with the information. This could take the form of a data bank with the appropriate security that could only be accessed with the consent of the individual. More security could be added by allowing only the information relevant for the purpose of disclosure to be released.[49]

There are a number of reasons why increased sequencing capacity will improve the US population's quality of life. In the next few chapters I will describe the method we have developed in response to the need for the $1000 Genome. However, it is important to remember that, there are issues that must be faced to ensure that abuses of cheaply acquired sequencing information are limited. The major concern with cheap sequencing is that the large amounts of information it will generate will be used to discriminate against people. There are protections in place to guard against misuse but with the coming of the $1000 Genome they will not be enough. Issues of ownership, consent and storage must all be worked out to ensure that abuses do not occur.

3. Constructing an AFM Nanopore Sequencer

Before assembly of the set-up shown in Figure 1.8 could take place, the individual components had to be chosen, synthesized, optimized, and verified. The critical element of this method is the construction of a surface-bound rotaxane. Rotaxanes resemble a dumbbell shaped structure consisting of a macrocycle, such as a nanopore, encircling an axis, which is referred to as a threading molecule, locked by two bulky groups referred to as stoppers (Figure 3.1). Our rotaxane consists of a surface that can bind the threading molecule and act as one stopper (Figure 3.2a), a nanopore that will spontaneously slide onto the thread and includes a group that can covalently bond to a functionalized AFM tip (Figure 3.2b and d), and another stopper that will keep the nanopore in place but allows it to slide over with the force applied by an AFM and can bind a strand of DNA at its other end (Figure 3.2c). All of this requires that an acceptable nanopore be identified and modified. Then a threading molecule must be tested with the modified nanopore to ensure it will form a rotaxane. The surface not only acts as a stopper; it must also keep DNA from sticking to it in order to eliminate nonspecific adhesion. Lastly, attachment chemistry between the nanopore and AFM tip must be designed that facilitates binding.

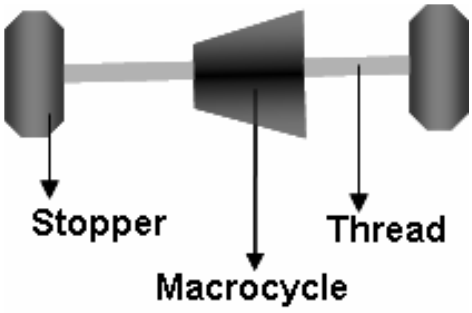


Figure 3.1. Structure of a rotaxane.

3.1. Materials and Methods

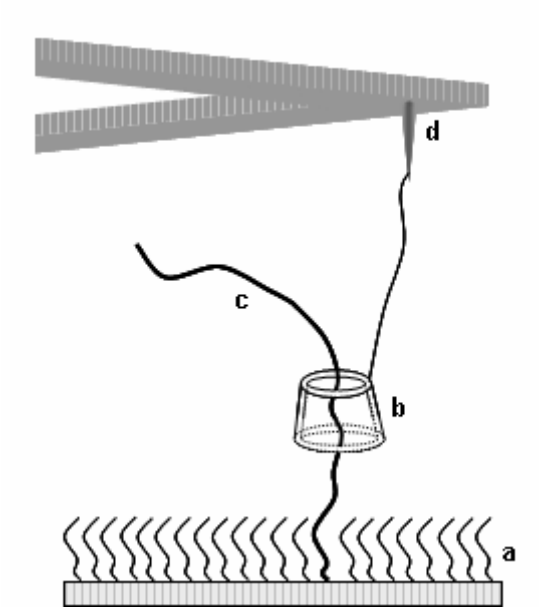


Figure 3.2. Surface rotaxane consisting of two stoppers (a and c) and a nanopore (b) that binds to an AFM tip (d).

Modified DNA was acquired from Integrated DNA Technologies (Skokie, IL) and purified by HPLC. β-Cyclodextrin was generously provided by Cerestar (Mechelen, Belgium). The acetonitrile used for separation was HPLC grade from Pierce (Rockford, IL). PBS (pH 7) buffer was purchased from VWR (West Chester, PA). Silanes were purchased from Gelest (Tullytown, PA). Triethyl ammonium acetate buffer was prepared using freshly distilled triethylamine mixed with HPLC pure water, Pierce (Rockford, IL), and brought to pH 7 with acetic acid. Vinylsulfone-PEG-NHS (MW 3200) was purchased from Nektar Therapeutics (San Carlos, CA). All other chemicals were purchased from Sigma-Aldrich (Milwaukee, WI). Preparative TLC plates (10x20 cm, 250 µm) were from EMD Chemical Inc (Gibbstown, NJ). All water was 18 mΩ from Nanopure Diamond of Barnstead.

All NMR spectra were recorded on Varian Inova 500 NMR spectrometer at 25ºC, ^{1}H spectrum were referenced to D_2O (4.65ppm), ^{13}C was referenced using the gyromagnetic ratios of the nuclei to calculate their frequency based on the locked frequency. Matrix Assisted Laser Desorption Ionization Time-Of-Flight (MALDITOF) mass spectra were recorded on a VG TofSpec spectrometer. 3-Hydroxypicolinic acid was used for the

rotaxane-DNA conjugates, and 4-hydroxybenzylidenemalononitrile for all other samples. An Agilent 1100 series binary pump HPLC system was used for separation of the rotaxane-DNA conjugates. The spectra were collected on a Harrick Scientific GATR accessory (Ge crystal as an internal reflection element with an incident angle of 65°) coupled into a Thermo-Nicolet 6700 FTIR spectrometer equipped with a mercury cadmium tellurium (MCT) detector. The whole FTIR system was protected by nitrogen generated from a flow-controlled liquid nitrogen tank. For ATR analysis the substrates were pressed upside-down against the Ge crystal with pre-set torque limited pressure of ~55 oz-in. All IR spectra were recorded against the ambient air background between 4000 and 650 cm^{-1} over 1000 scans with a spectral resolution of 4 cm^{-1}. Each spectrum was automatically smoothed and background corrected, and peak intensities normalized using OMNIC software. Differential spectra were generated by subtracting a clean silicon substrate spectrum from the sample spectra. Only absorption bands in the 3700 to 1300 cm^{-1} range were considered for surface characterization; bands below this region were disregarded due to the ambiguity caused by the constraints of the Ge crystal and interference from the silicon.[72] An oxidized silicon surface has a very strong absorbance between 1250 and 950 cm^{-1} and additional peaks around 875 cm^{-1} and 3740 cm^{-1}.[73]

3.2. Nanopore

The requirements for the nanopore are that it should accommodate a single strand of DNA, its width should be comparable to the distance between bases in a single strand of DNA, it should be water soluble, and it should be modifiable so it can bind to the AFM tip. These requirements are met by β-cyclodextrin. Cyclodextrins (CD) are a family of

cyclic oligosaccharides consisting of six (α-CD), seven (β-CD), or eight (γ-CD) glucose units linked to form a toroid shape (Figure 3.3).[74] As shown in Figure 3.37, the length of β-CD is 0.78nm which is very close to the stretched distance between bases in single stranded DNA, 0.6nm. The diameter of single stranded DNA is usually given as less than 1nm, the narrowest diameter of β-CD is 0.65nm so it can accommodate the DNA but, we expect, there will be resistance to the cyclodextrin sliding over the DNA that will cause measurable force differences. The interior of the cyclodextrin is relatively hydrophobic and when it is mixed with a linear, organic molecule, the molecule will spontaneously thread the cyclodextrin in aqueous solution.[76] However, the outside has hydroxyl groups that make it water soluble and these groups are easily modified so it is possible to functionalize β-CD so that an AFM tip can grab it.

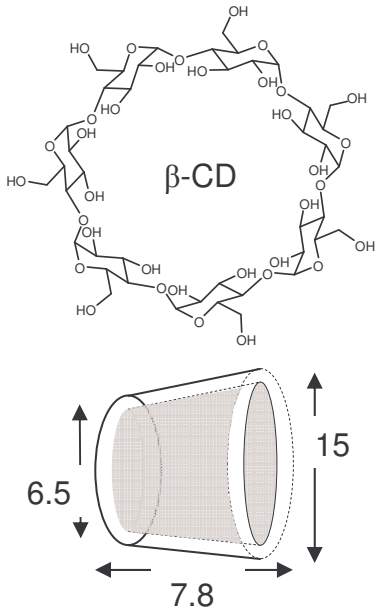


Figure 3.3. Chemical structure and dimensions, in Å, of β-CD.[75]

For the purposes of sequencing, a β-CD with a protected thiol was synthesized so that it can be attached to the AFM tip. The thiol group on the cyclodextrin is the ideal group for the AFM to "fish" for because it is not hydrolyzed in water and can form a stable bond quickly with reactive groups that are commercially available on the end of a PEG tether. Other modified cyclodextrins were also synthesized to study how different groups would interact with DNA during rotaxane formation (Figure 3.4).[77] An amine and

guanidinium group should repel the amine on the end of the threading molecule, but there are strong electrostatic and hydrogen bonding interactions with the phosphate backbone. A pyrene group was expected to base-stack with the exposed bases of single stranded DNA. The thymine may experience hydrogen bonding and also base-stacking with the DNA. A carboxylic acid group should be repelled by the negatively charged backbone but attracted to the protonated amine group at the 5'- end of the DNA conjugate in a buffered solution, (pH 7.5).[77]

3.2.1. Synthesis and characterization of nanopores

N-(mono-6'-deoxy-β-cyclodextrin)-3-(pyridin-2-yldisulfanyl)propanamide (β-CD-SPDP in Fig. 3.4) was prepared with β-CD-NH_2 (0.050 g, 0.044 mmol) which was mixed with adamantane (0.004 g, 0.029 mmol) in a 10:1 mixture of water and methanol and heated until the solution went clear. Once cooled, N-succinimidyl 3-(2-pyridyldithio)propionate (0.055 g, 0.176 mmol) dissolved in the minimum amount of DMSO was added. The reaction was monitored by TLC (5:3:3:3 isopropanol:ethyl acetate:ammonium hydroxide:water, detected with H_2SO_4 in ethanol and UV, $R_f = 0.54$). Preparative TLC (5:3:3:3 isopropanol:ethyl acetate:ammonium hydroxide:water, detected with H_2SO_4 in ethanol) was used to separate the products and to fully remove any DMSO so that the adamantane could be rinsed clean of the cyclodextrin using chloroform. (estimated yield 34%). MALDI-TOF Mass (m/z): 1353 for $[M + Na]^+$ (calculated mass for $C_{50}H_{78}O_{35}N_2S_2$: 1330). ^1H NMR (D_2O) δ 8.48 (d, 1 H), 7.94 (m, 2 H), 7.27 (d, 1 H), 5.23 (s, 1 H), 5.07 (s, 6 H), 3.99-3.63 (m, 42 H), 3.22 (m, 4 H)

3', 3''-Dithiopropionic acid mono(N-mono-6-deoxy-β-cyclodextrin)amide (β-CD-SS-COOH in Figure 3.4) was prepared using β-CD-NH_2 (0.016 g, 0.014 mmol) which was dissolved in anhydrous DMF (10 ml) and added dropwise to a solution of 3,3'-dithiobis(succinimidyl propionate) (0.011 g, 0.028 mmol) in DMF (10 mL) while stirring under N_2.[78] The reaction was monitored by TLC (5:3:3:3 isopropanol:ethyl acetate:ammonium hydroxide:water, detected with H_2SO_4 in ethanol, R_f = 0.47), with the presence of the cyclodextrin product marked by treatment with 5% sulfuric acid in ethanol with heating. After completion of the reaction, the product mixture was rotary-evaporated down to ~ 2 ml of DMF. Acetone was added to the solution; the resulting precipitate was gravity filtered and washed with acetone. The precipitate was redissolved in the minimum amount of DMF, acetone was again added, and the precipitate collected. This process was repeated one more time. The final precipitate was dried *in vacuo* at 40ºC overnight. The product was furnished as yellowish powder (73%). MALDI-TOF Mass (m/z): 1348

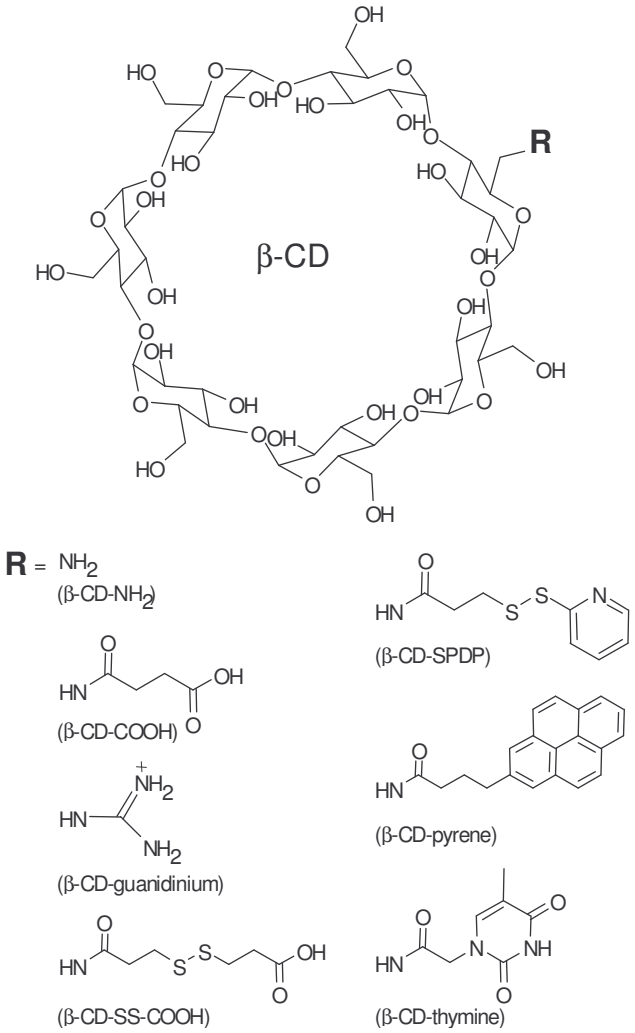


Figure 3.4. Structures of the modifications made to β-cyclodextrin.[77]

for $[M + Na]^+$ (calculated mass for $C_{48}H_{79}O_{37}NS_2$: 1325) Further purification was carried out using preparative TLC (5:3:3:3 isopropanol:ethyl acetate:ammonium hydroxide:water, detected with H_2SO_4 in ethanol) with *N*-hydroxysuccinimide byproduct used as a UV detectable marker and methanol used to extract the cyclodextrin (yield 49%). ^{1}H NMR (D_2O) δ 5.24 (s, 1 H), 5.07 (s, 6 H), 3.94-3.61 (m, 42 H), 2.94 (t, 2 H), 2.60 (t, 2 H).

Mono-6-deoxy-6-amino-β-cyclodextrin (β-CD-NH_2 in Fig. 3.4) was prepared following the literature procedure.[79] The reaction was monitored by TLC (5:3:3:3, isopropanol:ethyl acetate:ammonium hydroxide:water, primary amine detected with ninhydrin solution and heating, cyclodextrin with 5% sulfuric acid in ethanol with heating, R_f = 0.28) (yield 48%). ^{1}H NMR (D_2O) δ 5.06 (singlet with shoulder, 7 H), 3.9-3.7 (m, 20 H), 3.70-3.58 (m, 8 H), 3.58-3.42 (m, 14H). MALDI MS m/z: 1156 for $M+Na^+$ (calcd for $C_{42}H_{71}O_{34}N$: 1134).

4-(Mono-6'-deoxy-6'-amino-β-cyclodextrin)-4-oxobutanoic acid (β-CD-COOH in Fig. 3.4) was prepared using a solution of β-CD-NH_2 (0.200 g, 0.176 mmol) in anhydrous pyridine (10 ml) in which was added an excess of succinic anhydride (0.034 g, 0.352 mmol). The reaction was heated to 50ºC and stirred under N_2 while being monitored by TLC (5:4:3, butanol:ethanol:water, detected with H_2SO_4 in ethanol, R_f = 0.16). Products were separated using preparative TLC (5:4:3, butanol:ethanol:water) with *N*-hydroxysuccinimide added as a UV detectable marker and methanol used to extract the desired product (yield 27%). ^{1}H NMR (D_2O) δ5.09 (singlet with shoulder 1 H) 4.93 (s, 6

H), 3.87-3.67 (m, 26 H), 3.61-3.24 (m, 16 H). MALDI-TOF m/z: 1255 for $[M + Na]^+$ (calculated mass for $C_{46}H_{75}O_{37}N$: 1233).

Mono-6-deoxy-6-guanidinium-β̃-cyclodextrin (β-CD-guanidinium in Fig. 3.4) was prepared using a solution of β-CD-NH_2 (0.050 g, 0.004 mmol) in anhydrous DMF (0.5 ml) in which were added 1*H*-Pyrazole-1-carboxamidine hydrochloride (0.027 g, 0.018 mmol) and *N*,*N*-Diisopropylethylamine (0.024 g, 0.018 mmol) under N_2.[80] The reaction proceeded for 22 hours, monitored by TLC (7:7:5:4, EtOAc:isopropanol:NH_4OH :water, detected with H_2SO_4 in ethanol, $R_f = 0.06$). Once the reaction was finished the solution was poured into a 50 ml round bottom flask and 20 ml of ether was added. After 2hrs of stirring the product, a whitish precipitate, was collected by suction filtration and dried *in vacuo* at 40 ºC overnight (yield 30%). ^{1}H NMR (D_2O) δ 4.94 (singlet with shoulder, 7 H), 3.84-3.68 (m, 26 H), 3.62-3.39 (m, 16 H); and ^{13}C NMR (D_2O, ref 125 MHz) δ160.4, 107.6, 104.5, 85.1, 83.9, 75.7, 74.6, 63.1, 62.8 56.9, 45.2. MALDI-TOF m/z: 1199 for $[M + Na]^+$ (calculated mass for $C_{43}H_{75}O_{34}N_3$: 1177).

N-(mono-6'-deoxy-β-cyclodextrin)-4-(pyren-2-yl)butanamide (β-CD-pyrene in Fig. 3.4) was synthesized using a solution of β-CD-NH_2 (0.100 g, 0.088 mmol) in anhydrous DMF (12 ml) to which was added 1-pyrenebutyric acid *N*-hydroxysuccinimide ester (0.040 g, 0.1 mmol) was added under N_2. The reaction was stirred at room temperature overnight. The next day more 1-pyrenebutyric acid *N*-hydroxysuccinimide ester (20 mg, 0.05 mmol) was added because TLC (5:3:3:3 isopropanol:ethyl acetate:ammonium hydroxide:water, detected with H_2SO_4 in ethanol and UV, $R_f = 0.6$) showed β-CD-NH_2 remained. Once the starting material was consumed the solvent was removed by rotary

evaporation to give a yellow oil. The product was purified using preparative TLC (isopropanol:ethyl acetate:ammonium hydroxide:water, 5:3:3:3, detected with H_2SO_4 in ethanol and UV), extracted (yield 64%). ^{1}H NMR (DMSO with D_2O) δ 8.32 (d, 1H), 8.2 4(t, 2H) 8.19 (m, 2H), 8.09 (m, 2H), 8.03 (t, 1H), 7.91 (d, 1H), 4.81 (singlet with shoulder, 7 H), 3.66-3.53 (m, 27H), 3.37-3.22 (m, 15H), 3.25 (m, 2H), 2.86 (t, 2H), 2.18 (t, 2H). MALDI-TOF Mass (m/z): 1428 for $[M + Na]^+$ (calculated mass for $C_{62}H_{87}O_{35}N$: 1405).

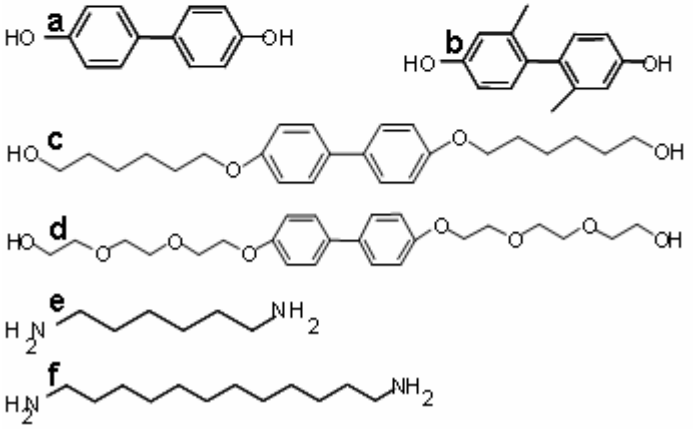
N-(mono-6'-deoxy-β-cyclodextrin)-2(thymin-1-yl)acetamide (β-CD-thymine in Fig. 3.4) was made using thymine-1-acetic acid (0.024 g, 0.13 mmol) and *N*,*N*′-Dicyclohexylcarbodiimide (0.063 g, 0.31 mmmol) which were mixed together in DMF (20 ml) under nitrogen (24hrs). β-CD-NH_2 (0.100 g, 0.088 mmol) was then added. The reaction was stirred overnight. The reaction was monitored with TLC (5:4:3 butanol:ethanol:water, detected with H_2SO_4 in ethanol and UV, $R_f = 0.3$) until no β-CD-NH_2 remained. The reaction was dried down to give a white crystalline solid. Product was separated using preparative TLC (5:4:3 butanol:ethanol:water, detected with H_2SO_4 in ethanol and UV) (yield 40%). 1H NMR (DMSO with D_2O) δ 7.24 (s, 1H), 4.08 (singlet with shoulder, 7H), 3.92 (s, 2H), 3.60 (m, 28H), 3.31 (m, 14H), 1.71 (s, 3H). MALDI-TOF Mass (m/z): 1322 for $[M + Na]^+$ (calculated mass for $C_{49}H_{77}O_{37}N_3$: 1299).

3.3. Threading molecules

We have observed that single stranded DNA molecules cannot self-assemble into rotaxanes with cyclodextrin, although aminodeoxy-β-CDs form complexes with adenosine phosphates.[81] The addition of a threading molecule at the end of the DNA is

Table 3.1. Structure of threading molecules and results from solubility tests(S- soluble, NS- not soluble) with 8mM β-CD.

Guest Molecule	Solubility in β-CD Solution
a. Biphenyl-4,4'-diol	S
b. 2,2'-Dimethylbiphenyl-4,4'-diol	NS
c. 4,4'-bis(6-hydroxyhexyloxy)biphenyl	NS
d. 4,4'-bis(triethylene glycol)biphenyl	NS
e. 1,6-diaminohexane	NS
f. 1,12-diaminododecane	S



therefore necessary for loading the cyclodextrin. A number of molecules have been screened for their threading capability (Table 3.1). The procedure for the synthesis of 4,4'-bis(6-hydroxyhexyloxy)biphenyl (c in Table 3.1) was taken from Bagheri *et al*.[82] Synthesis of 4,4'-bis(triethylene glycol)biphenyl (d in Table 3.1) was taken from Cordova *et al*.[83] A 1 to 1.5 ratio of biphenyl-4,4'diol (1mg, 5.4µmol) was mixed with β-CD (9.1mg, 8.1µmol) in 100mM, pH 7.5 phosphate buffer (1ml). The mixture was stirred overnight with initial heating and cooling. Solubility was determined visually. A similar procedure was used for all the other threading molecules. Based on the solubility study, 1,12-diaminododecane (DOD) was further investigated as a threading molecule.

NMR was used in order to prove that DOD does, in fact, spontaneously thread the cyclodextrin in aqueous solution. Before threading, accurate chemical shifts for the hydrogens in β-CD had to be established. A COSY (Correlated SpectroscopY) 2D NMR

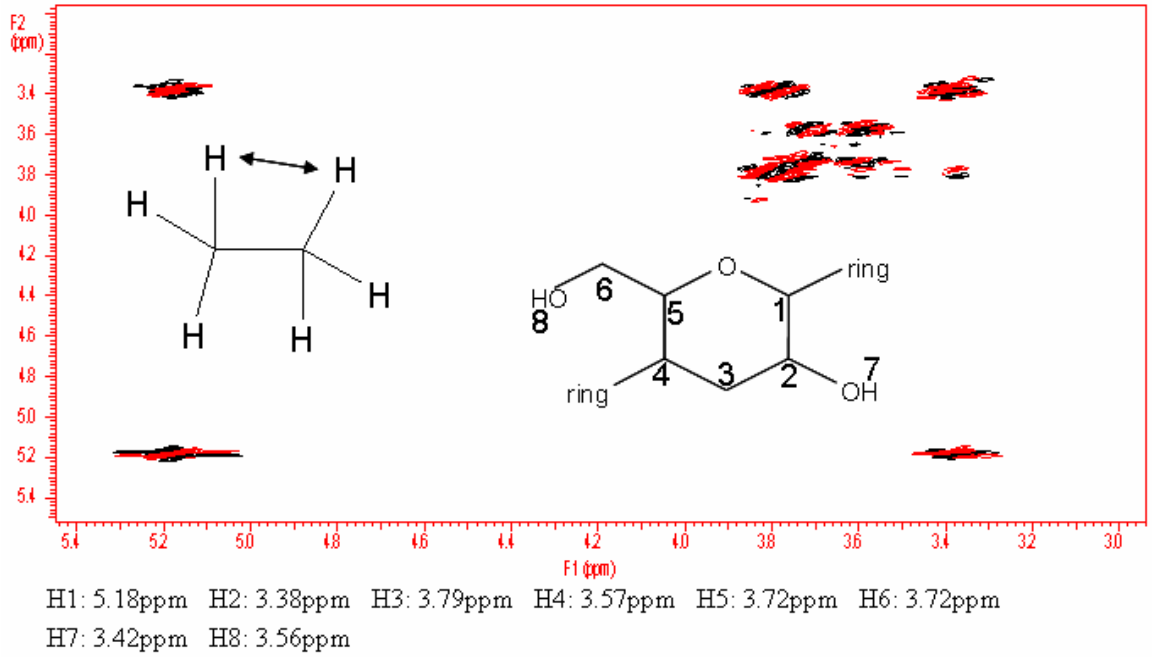


Figure 3.5. COSY of β-CD with chemical shift assignments of hydrogens in one glucose subunit. The insert on the left shows what is meant by hydrogens on neighboring carbons and the one on the left show the location of the hydrogens.

experiment was used to assign the hydrogens of the cyclodextrin. COSY experiments are able to give information about which hydrogens are on neighboring carbons in the same molecule. By starting with the knowledge that H1 of a single glucose unit of symmetric β-CD has a chemical shift at 5.18 ppm and knowing that its neighboring hydrogen is on C2, by looking at cross peaks one can determine that H2's chemical shift is 3.38 ppm (Figure 3.5). The shift for all the other hydrogens follow in a sequential manner; H2 neighbors H3 and, based on the cross peaks the shift for H3 is 3.79 ppm, etc. It has been proven through crystal structures of β-CD that H3 and H5 of β-CD are located in the interior of the cyclodextrin.[84] Therefore, if DOD threads β-CD then H3 and H5 will be the closest to the hydrogens of DOD.

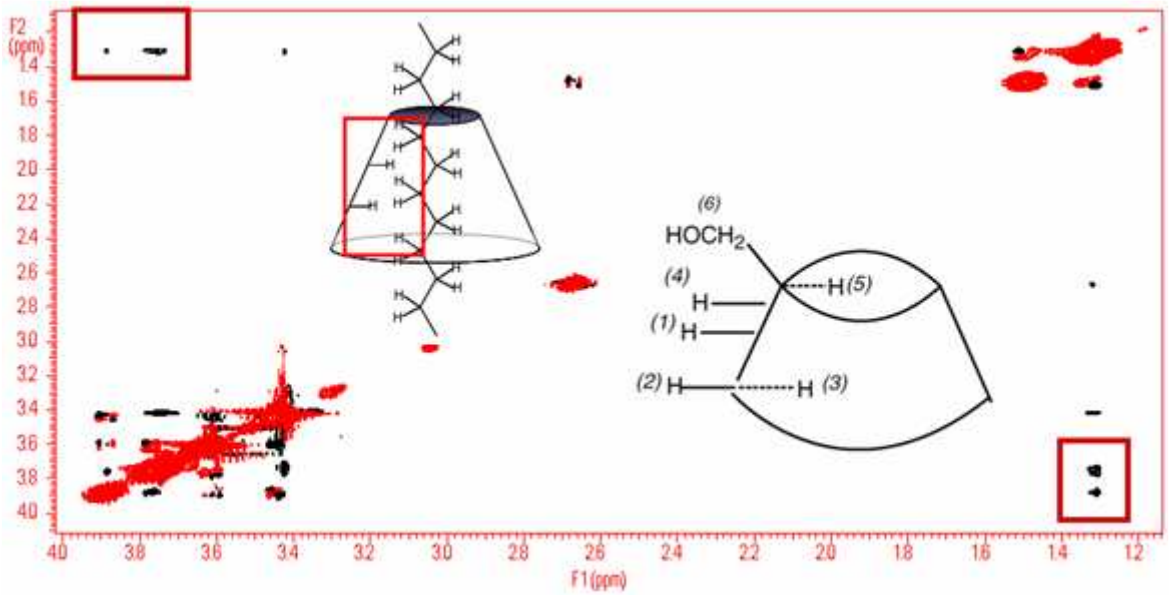


Figure 3.6. ROESY of β-CD mixed with DOD. The inset on the left shows what is meant by close in space and on the right shows the position of hydrogens in the ring, H3 and H5 are the ones of interest. ROESY also contains COSY information, that is why there are so many additional cross peaks.[77]

Rotational Overhauser Effect SpectroscopY (ROESY) can determine if hydrogens are close in space, if they are not in the same molecule. During a ROESY experiment the relaxation rate for nuclei that have been pulsed by the fluctuating magnetic fields caused as their neighbors in space relax is proportional to r^{-6}, where *r* is the distance between the nuclei, so in order to detect cross peaks the neighboring nuclei have to be close.[23] Figure 3.6 is a ROESY spectrum collected of a 1:1 mixture of β-CD (2 mg, 1.4 µmol) and 1,12-diaminododecane (0.3 mg, 1.4 µmol) in D_2O, stirred overnight and filtered. It confirms that a pseudorotaxane (no bulky groups are present to keep the ring in place) of β-CD with DOD indeed exists in aqueous solution. The red rectangles in Figure 3.10 highlight the cross peaks between central hydrogens of DOD (1.28 ppm) and those in the interior of the β-CD (H3, 3.89 ppm, and H5, 3.76 ppm, some shift is to be expected because of the presence of DOD). The dodecane chain can readily be incorporated into

DNA using an automated DNA synthesizer since the aminododecane phosphoramidite is commercially available; therefore, DNA with a dodecane chain with an amine group on the end is also commercially available and DOD is an obvious choice as a threading molecule.

3.4. Rotaxanes

Once a threading molecule was determined and tested to confirm that it can form pseudo-rotaxanes (a structure where the macrocycle threads the axis but no stoppers are present to hold it in place) with unmodified β-CD, rotaxanes using the modified β-CD needed to be synthesized and verified. Two different types of rotaxanes were formed for this purpose. First, MALDI-TOF was used to prove that solution phase rotaxanes were constructed. Then rotaxanes were built on a surface similar to the schematic in Figure 1.8 except that, in order to simplify the results, the DNA was not added at the end. ATR-FTIR was used for verification.

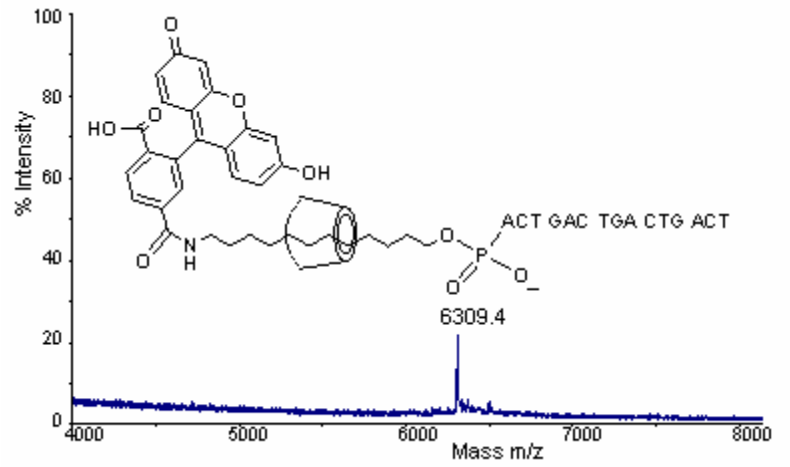


Figure 3.7. MALDI-TOF of rotaxane. Mass (m/z): 4816 for DNA, 5175 for DNA plus fluorescein stopper, and 6309 for the rotaxane, all had sodium adducts.[77]

3.4.1. Solution phase DNA-rotaxane conjugates

The first method consisted of forming rotaxanes in solution and using MALDI-TOF to verify the process.[77] This was done for all the cyclodextrins that were mentioned earlier. The synthesis of the DNA-rotaxane

conjugate with β-CD used a 5'-aminododecane-ACT GAC TGA CTG ATC oligonucleotide (11.65 nmol) which was mixed with β-CD (0.03 mg, 23.3 nmol) in pH 7.5, 100 mM phosphate buffer (50 µl). The solution was stirred overnight. An excess of 6-carboxyfluorescein *N*-succinimidyl ester in DMSO (0.06 mg, 126 nmol) was added and stirred for 2 hours to complete the rotaxane. The fluorescein stopper was picked because we found it gave the most reproducible results with rotaxane formation and its spectroscopic properties helped with separation. Experiments show that the DNA can function as the other stopper as well, so these are rotaxanes.[85] The rotaxane was purified using HPLC. A gradient of 0-30% acetonitrile with 100mM fresh triethylammonium acetate buffer was used for separation with the detector set at 260nm and 490nm. The yield was determined by integration of the HPLC spectrograph (excluding peaks without any DNA) (see Figure 3.7 for MALDI of pure rotaxane and Table 3.2 for yields). A similar procedure was used for all the other modified cyclodextrins. Control experiments were done using the same sequence of DNA without the dodecane amine at the 5' end, and with the dodecaneamine-DNA conjugate without adding the fluorescein. Based on MALDI data, no rotaxanes were formed.

Comparing the yields of the rotaxanes (Table 3.2), we see that all the modifications of β-CD lower the yield of the diaminododecane rotaxanes. In contrast, the presence of DNA greatly increases rotaxane formation with all but the three modified cyclodextrins containing aromatic moieties. The lack of solubility due to the hydrophobic modifications of β-CD-pyrene, β-CD-thymine, and β-CD-SPDP may contribute to the lack of rotaxane formation. It is also possible that these moieties block formation of the

Table 3.2. Table of yields for solution phase rotaxanes formed with different modified cyclodextrins. For these rotaxanes a 15mer poly T sequence was used.[77]

Cyclodextrin	DNA-Rotaxane Conjugate		Diaminododecane Rotaxane	
	Yield	Exp m/z w/ Na^+ adducts [calc]	Yield	Exp m/z w/ Na^+ adducts [calc]
β-CD	14.0%	6281.65 [6257.31]	6.5%	2078.77 [2052.06]
β-CD-NH_2	18.2%	6279.51 [6256.32]	3.1%	2211.48 [2051.01]
β-CD-COOH	28.8%	6464.03 [6356.40]	1.0%	2173.19 [2151.09]
β-CD-guanidinium	21.3%	6304.61 [6298.36]	3.4%	2115.67 [2093.05]
β-CD-SS-COOH	13.4%	6471.12 [6448.58]	1.1%	2265.82 [2243.27]
β-CD-SPDP	0.0%	[6453.6]	0.4%	2266.71 [2248.29]
β-CD-pyrene	0.0%	[6528.67]	1.8%	2344.37 [2323.36]
β-CD-thymine	0.0%	[6422.46]	3.3%	2243.93 [2217.51]

rotaxane by staying in the cyclodextrin cavity. On the other hand, the presence of a small amount of the diaminododecane rotaxane may indicate that those hydrophobic interactions may be so strong with the DNA that the cyclodextrin is "too busy" to form a rotaxane. The length of the side chain on the cyclodextrin also seems to affect formation. β-CD-COOH and β-CD-SS-COOH have the same end group, carboxylic acid, interacting with the amino group at the 5'-end of DNA. The difference in DNA rotaxane formation could be due to the increased flexibility of carboxylic acid in β-CD-SS-COOH, resulting in a reduced interaction with the amino group. Both the negatively charged, β-CD-COOH and β-CD-SS-COOH, and positively charged, β-CD-NH_2 and β-CD-guanidinium, cyclodextrins were able to form reasonable amounts of rotaxanes. One of the modifications that can base stack or hydrogen bond with a DNA base may be best for the purpose of slowing down DNA as it translocates through a nanopore.[77] They prevent the formation of rotaxanes when DNA is present but, as described in the next section, surface bound rotaxanes are first formed with DOD before DNA is present.

3.4.2. Surface bound rotaxanes

Once we knew that it was possible for modified β-CDs to slide onto the dodecaneamine-DNA, we assembled rotaxanes on a silicon surface and used FTIR to verify their formation.[85] β-CD-SS-COOH was the only modified cyclodextrin used for these experiments. The results proved that β-CD-SS-COOH does thread DOD and is effectively stopped by PEG, trapping it and forming a surface-bound rotaxane. DNA was not used in these experiments in order to simplify the FTIR spectra and because a PEG tether is used in the final construct (discussed in Chapter 4). In addition to DOD, when constructing rotaxanes on the surface, 1,3-adamantane diacetic acid was used to hold cyclodextrins in place. The concern was that, even if the cyclodextrin threads DOD on the surface, until the addition of a stopper cyclodextrin would be able to slide back off; the stability constant for a 1 to 1 ratio of DOD to β-CD is 4,760 M^{-1}.[86] By integrating adamantane at the base of the carbon chain the cyclodextrin can be held more tightly onto the surface pseudorotaxane until PEG is used to stopper it because of the high affinity β-CD has for adamantane, the 1 to 1 stability constant is 20,000 M^{-1}.[86] The diaminododecane is still a necessary piece because, once the cyclodextrin threads the surface complex, the PEG needs to find the pseudorotaxane to bind to it which is an easier task with the long carbon chain extending up off the surface rather than the cubic-shaped adamantane with the cyclodextrin around it (see Figure 3.9).

The procedure for forming the entire surface bound rotaxane construct begins with cleaning the silicon surface. Silicon wafers were cut to 1 cm square pieces. They were then placed in the ozone cleaner for 10 minutes, removed, and immediately placed in piranha (3:1 sulfuric acid and hydrogen peroxide) for 3 minutes. The wafers were

removed and immediately rinsed in 18 MΩ water and placed in the silane solution. The silane was prepared with 200 µl *n*-propyl silane and 5 µl APDM in 1 ml 95% ethanol. The wafers were shaken for 3 minutes, rinsed lightly with water, put into a clean, argon-filled desiccator, and placed under vacuum for 1 hour to cure the silanes. Then 1,3-adamantane diacetic acid (25mg, 0.1mmol), N, N'-dicyclohexylcarbodiimide (80 mg, 0.4 mmol), and N-hydroxysuccinimide (11 mg, 0.1 mmol) were added to DMF (1 ml) and allowed to react for 15 minutes at 50˚ C. The wafers were added to this solution for 1 hour at 50˚ C and then rinsed with water. Then they were placed in a solution of N, N'-dicyclohexylcarbodiimide (80 mg, 0.4 mmol) and N-hydroxysuccinimide (11 mg, 0.1 mmol) for 10 minutes, and then rinsed with DMF and transferred to 1,12-diaminododecane (20 mg, 0.1 mmol) in DMF (1ml). The wafers were allowed to sit for 10 minutes at 50˚C and then rinsed well with water. The wafers were then placed in a solution of β-CD-SS-COOH (50 mg, 0.03 mmol). After 20 minutes, the VS-PEG-NHS (30 mg, 0.09 mmol) was added to the solution and allowed to react for 30 minutes. VS-PEG-NHS was added to the controls without the cyclodextrin. After 30 minutes the surfaces were rinsed well with water.

3.4.3. ATR-FTIR measurement

Attenuated total reflection (ATR) FTIR was used in order to prove that surface bound rotaxanes were indeed formed.[85] At each step of the surface bound rotaxane preparation FTIR data was collected in order to verify every level of the chemistry.

Aminopropylsilylate silicon surfaces (Figure 3.8): Thickness: less than 1 nm. Contact angle: 42°. FTIR (cm^{-1}): 3403 (N-H stretching), 2926 and 2858 (CH_2 stretching), 1641 (NH_2 scissoring), 1446 and 1379 (stretching of bicarbonate salt).[87].

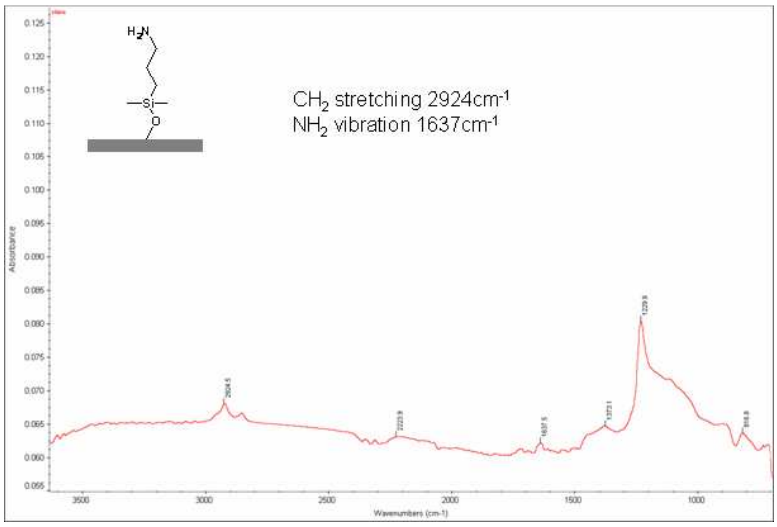


Figure 3.8. FTIR of aminopropylsilylate (inset), the first attachment step in building a surface bound rotaxane.[85]

Adamantyl surface (Figure 3.9): Thickness: 7.8Å. Contact angle: 62°. FTIR (cm^{-1}): 3398 (OH stretching), 2925 and 2855 (CH_2 stretching), 1716 (C=O stretching), 1636 (amide II bend).

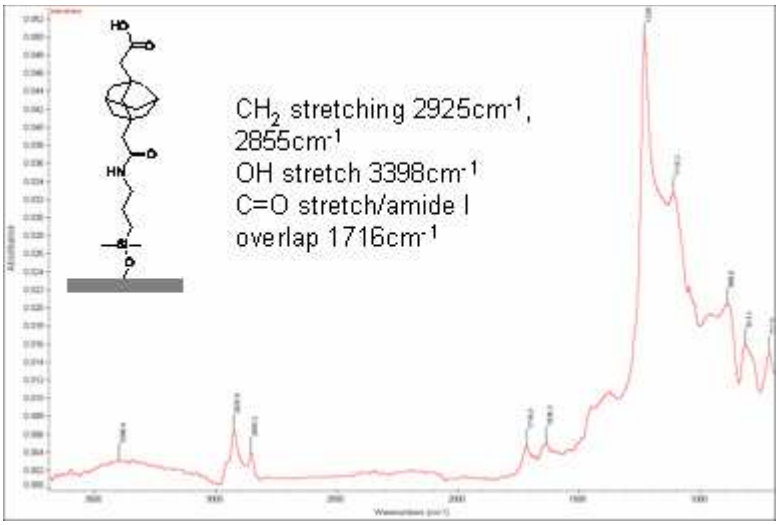


Figure 3.9. FTIR of adamantyl surface (inset), the second step in building a surface bound rotaxane.[85]

Dodecylamine surface (Figure 3.10)**:** Thickness: 10.9Å. Contact angle: 72°. FTIR (cm^{-1}): 3325 and 3277 (primary amine stretching), 2926 and 2857 (CH_2 stretching), 1717 (overlap with C=O), 1609, and 1376 (amide I, amide II, and amide III, respectively).

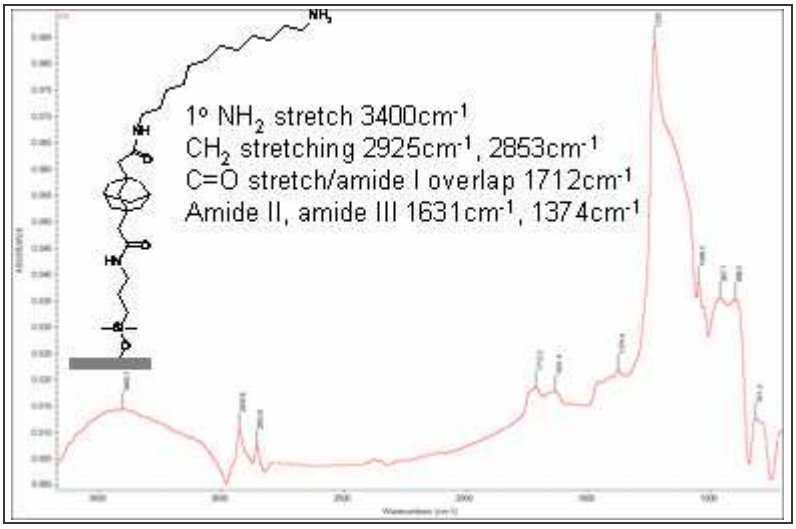


Figure 3.10. FTIR of dodecylamine surface (inset), the third step in building a surface bound rotaxane.[85]

CD-rotaxane surface: A substrate with the rinsed dodecylamine surface (Figure 3.14, inset) was placed in a solution of β-CD-SS-COOH (50 mg, 0.03 mmol) in pH 7, 50 mM phosphate buffer for molecular threading, forming β-CD-SS-COOH pseudorotaxane. After 20 minutes, vinylsulfone-PEG-NHS (30 mg, 0.09 mmol) was added to the solution and allowed to react for 30 minutes. The resultant CD-rotaxane surface (Figure 3.11) was rinsed with water, blown with nitrogen, and dried in vacuum at 100°C overnight for surface characterization. XPS showed a single S 2p peak at ~ 167 eV resulting from the disulfide groups (-S-S-) on the CD rotaxane.

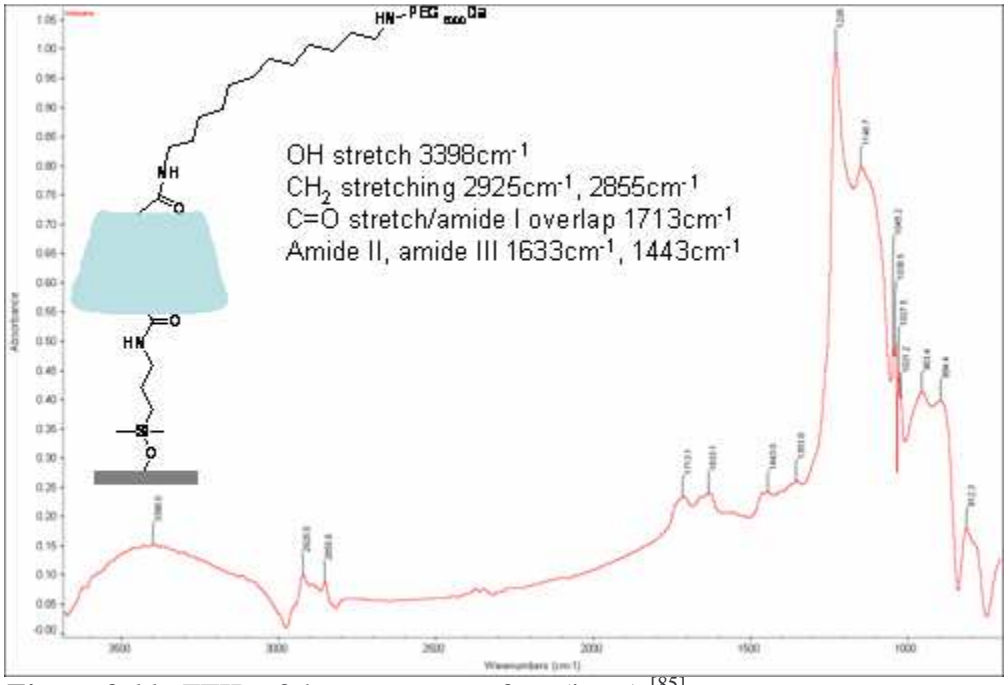


Figure 3.11. FTIR of the rotaxane surface (inset).[85]

An ATR-FTIR spectrum of the β-CD-SS-COOH thin layer, generated by depositing its aqueous solution on the silicon surface and drying in vacuum, was collected. The spectrum shows a characteristic OH band between 3650 cm^{-1} and 3000 cm^{-1} in addition to weak CH_2 stretching in the region of 2950 to 2750 cm^{-1}, strong C=O stretching/amide I overlap around 1646 cm^{-1}, and amide stretch-bend and open (1572 cm^{-1} and 1407 cm^{-1}, respectively). The CD-rotaxane surface shows the strong OH band solely associated with the cyclodextrin molecule in the same region as the β-CD-SS-COOH thin layer. It also shows CH_2 stretching at 2925 cm^{-1} and 2856 cm^{-1}, amide peaks at 1737 cm^{-1}, 1633 cm^{-1}, and 1443 cm^{-1}. Additionally, a surface was prepared using maleimide-PEG-NHS (MW 5000) skipping the step of adding the cyclodextrin in order to make comparisons. Thickness: 47.9 Å. Contact angle: 52°. FTIR (cm^{-1}): 2925 and 2855 (CH_2 stretching),

1737 (overlap with C=O), 1631, and 1372 (amide I, amide II, and amide III, respectively).

An overlay of the CD-rotaxane surface, β-CD-SS-COOH, and PEG spectra is shown in Figure 3.12. It was expected that the concentration of modified cyclodextrin on the rotaxane surface would be less than that of the cyclodextrin dried on silicon, which is consistent with the absorbance intensities of the OH stretching band on each spectrum.

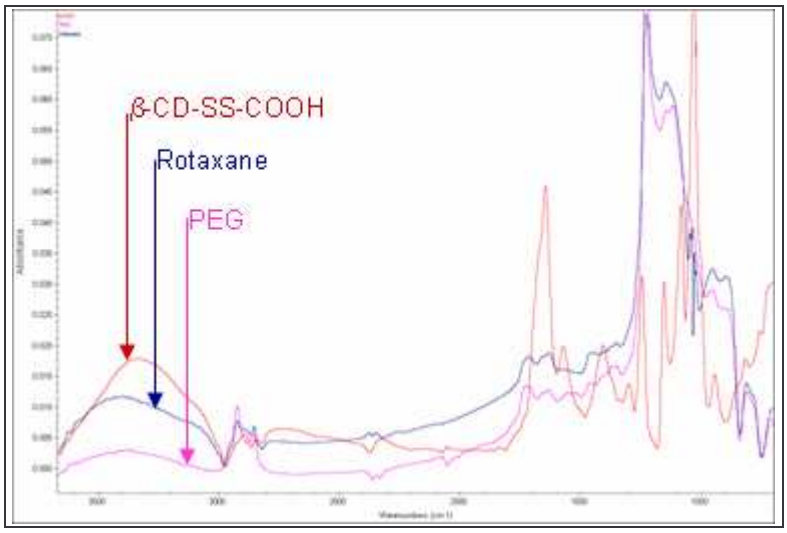


Figure 3.12. Overlay of FTIR of dried β-CD-SS-COOH (red), surface bound rotaxane (blue), and PEG-ylated surface (pink).[85]

Additionally, another control experiment was carried out to prove that cyclodextrin on the surface is not nonspecifically adsorbed to the construct. After the surface was PEG-ylated by reacting maleimide-PEG-NHS with the dodecylamine surface, the modified cyclodextrin β-CD-SS-COOH was deposited on the surface, allowed to dry, and FTIR data were collected. Then the same surface was washed with water to remove the cyclodextrin and another spectrum was taken. An overlay of the PEG surface before and after the cyclodextrin was washed away, a PEG surface that was never exposed to cyclodextrin, β-CD-SS-COOH alone, and the rotaxane surface is shown in Figure 3.13. The most telling part of the overlay is the OH stretching region (3000 to 3650 cm^{-1})

which shows strong absorbance for the cyclodextrin, rotaxane, and PEG with cyclodextrin dried on. The rinsed PEG surface and original PEG surface both lack the apparent OH absorbance.

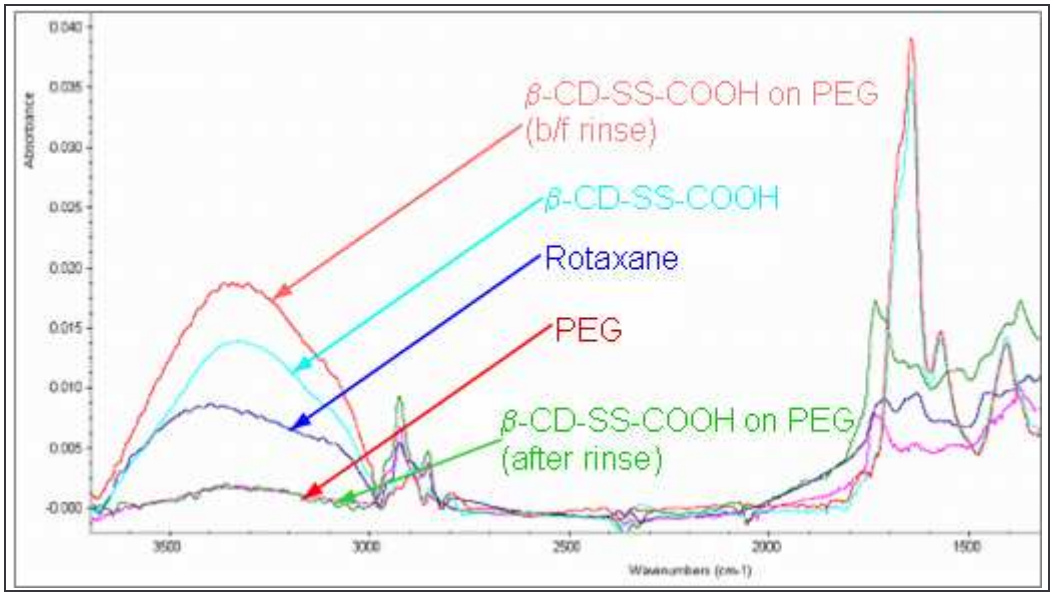


Figure 3.13. Overlay of FTIR of dried β-CD-SS-COOH on PEG-ylated surface (pink), β-CD-SS-COOH (light blue), surface bound rotaxane (blue), PEG-ylated surface (red), and rinsed dried β-CD-SS-COOH on PEG-ylated surface (green).[85]

3.5. Non-adhesive surface

An important aspect of this DNA sequencing system is that the forces measured should be the result of the cyclodextrin sliding over DNA, not from the cyclodextrin pulling DNA off of a surface it has adhered to.[85] This required the development of a surface that allowed attachment of the rotaxane while at the same time eliminating non-specific DNA adhesion. The solution is a mixed monolayer on silicon with a very low percentage of amine functionalized silanes for attachment and then a silane with a group that is the most non-stick for DNA. Silanes are compounds that have four substituent

groups attached to a silicon atom. These groups can have a combination of reactivity. There can be a hydroxyl group or chlorine for binding the silane to a silicon surface, an amine or thiol group for attaching something to the surface, non-reactive groups, like methyls, or other hydroxyl groups to cause polymerization of the silanes. This is not easy because DNA can adhere to surfaces via Coulomb, hydrophobic, salt-bridge and specific chemical interactions.[88, 89]

In order to determine the best non-stick surface Dr. Brian Ashcroft and I tested multiple candidates. They included 2-(methoxy(polyethyleneoxy) propyl)trimethoxysilane (PEG), (3,3,3-trifluoropropyl) dimethylchlorosilane (TFC), *n*-propyldimethylchlorosilane (*n*-propylsilane), 3-aminopropyldimethylethoxy-silane (APDM), and carboxyethylsilanetriol sodium salt (COOH).[90-92] A monolayer of each silane was prepared and DNA was covalently bound to an AFM tip. After multiple approach/retraction experiments the best surface was determined by the number of adhesion events recorded. It was decided that an *n*-propylsilane surface would most greatly decrease non-specific adhesion.

Chips of silicon ~1 cm^2 were cleaned as described above and treated with 200 µL *n*-propylsilane in 1mL 95% ethanol, 5% water. The chips were shaken for 3 minutes, rinsed lightly with water, put into a clean desiccator, flushed with argon and placed under vacuum for 1 hour to cure the silanes. The TFC and PEG surfaces used a 0.1% solution in 95% ethanol solution taken to pH 5 with acetic acid. The APDM surface was made in a similar way as the TFC surface except on acetic acid was used. The COOH surface was made using a 2% solution of the silane in 100mM PBS, pH7. Ellipsometry, FTIR and

AFM were used to verify the surface modification and determine the uniformity of the modified surfaces.

Ultrasharp CSC11/AlBS probes were placed in the ozone cleaner for 10 minutes, immediately dipped into fresh piranha solution for no more than 30 seconds (to prevent damage to the metallization), rinsed with water and put into a solution of APDM (200 µl, 1.4 µmol) in 95% ethanol (1 mL). After 5 minutes, the tips were rinsed with water, placed in a clean desiccator flushed with argon, and placed under vacuum (10 Torr) for 1 hour. After curing, the tips were placed in a solution of N-β(maleimidopropyloxy)-succinimide ester (BMPS – from Pierce Chemical) (10 mg) in 1 mL DMF for 15 minutes to convert the amine to the maleimide. The cantilevers were then placed in a solution of deprotected, thiol - 5'- $(TTT)_8$ CCC CTT TTG GGG $(TTT)_6$ TGC GCG CGT TGT TCG CGC GCT TTT TTT -3' DNA in TCEP buffer for 1 hour. The tips were rinsed with water before the experiment. Adhesion measurements were carried out in a 50 mM ph 7 phosphate buffer.

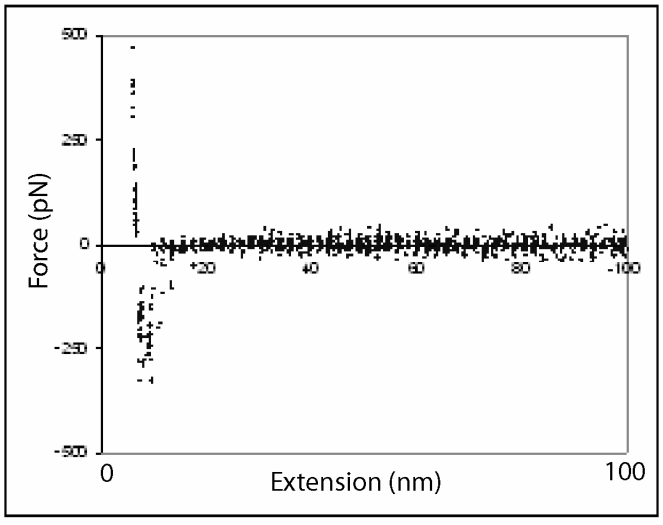


Figure 3.14. 30 superimposed withdrawal curves of DNA adhering to *n*-propylsilane.[85]

DNA functionalized probes approached the surface until a 100pN trigger force was reached, held for 1-8 s, and then withdrawn while the deflection signal was recorded. The *n*-propylsilane surface showed significant adhesion out to only about 10 nm, implying

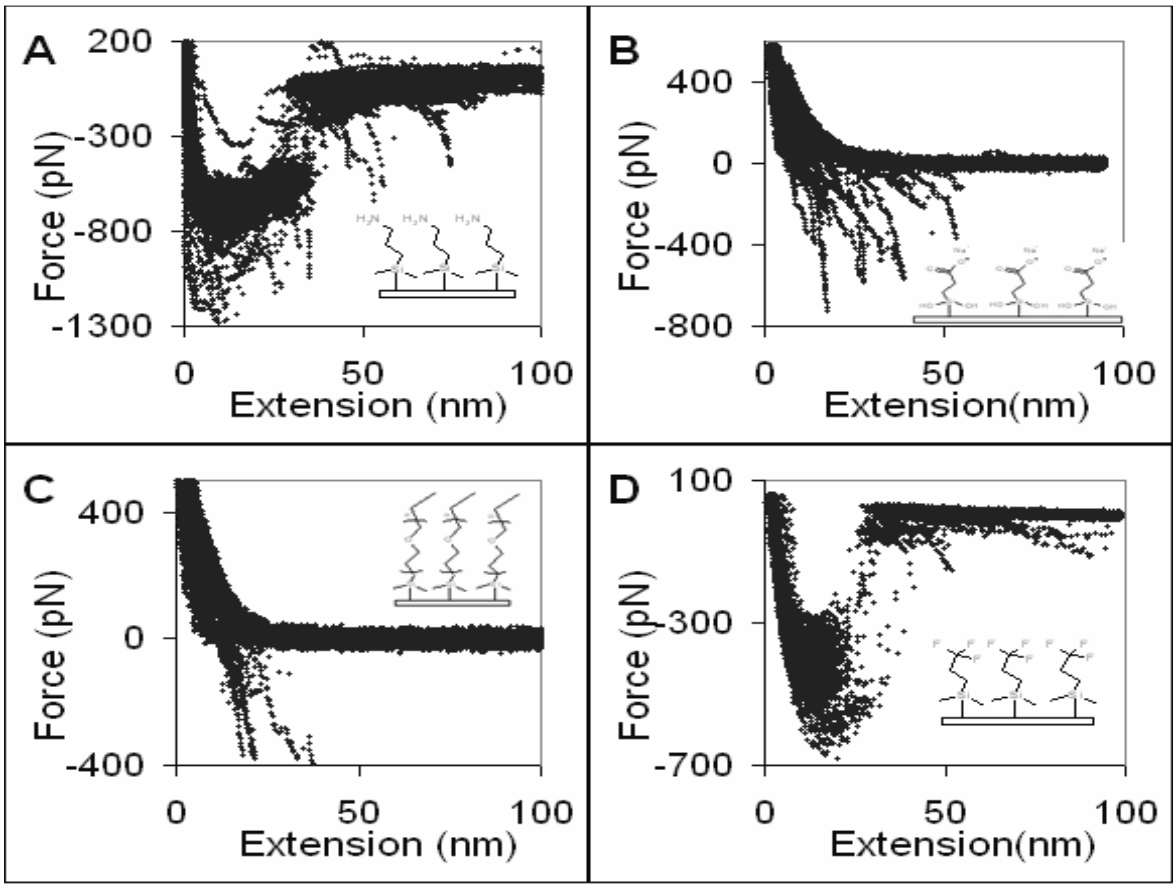


Figure 3.15. APDM (A) shows long range Coulomb interactions, COOH (B) and PEG (C) generally repel the DNA but also show a significant number of long-range adhesive interactions, and TFC (D) shows long range hydrophobic interactions.

that there were no measurable interactions with the DNA which is approximately 49nm long. (Figure 3.14).

In contrast the APDM, COOH, PEG and TFC surface all showed a number of long range interactions (Figure 3.15, overlay of 300 curves). The highly charged APDM and COOH surfaces are easily contaminated. It is the DNA's adhesion to the contamination that is measured on the COOH surface and a combination of contamination and the oppositely charged surface and DNA backbone attraction with the APDM. One would expect that a TFC surface (basically a Teflon surface) would be very non-stick.

Unfortunately, the highly hydrophobic nature of the surface causes small air bubbles to be trapped on the surface once water is added that interact with the DNA.[93] A PEG surface resulted in very few long range adhesion events but the *n*-propyl surface never showed any long range events.

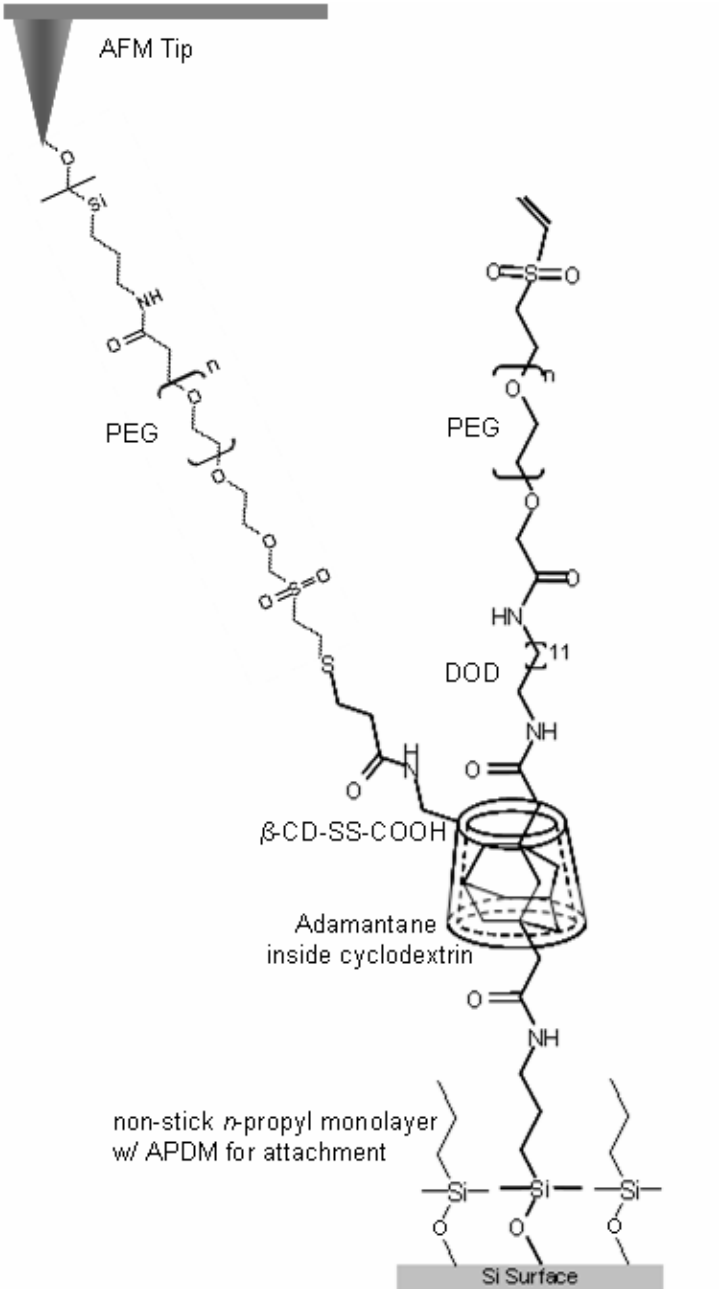


Figure 3.16. Complete structure of the surface bound rotaxane, the AFM tip is bound to deprotected β-CD-SS-COOH.

3.6. Modified AFM tips

In order for this method to work an AFM tip must be able to locate and bind to a modified cyclodextrin on a surface. An effective way to do this is add a "fishing line" to the tip in the form of a long (~28nm) PEG molecule. This adds flexibility and broadens the surface area the tip can search when trying to find and bind to the cyclodextrin. The reactive group on the end of the PEG is vinylsulfone. It was chosen because it binds to thiols, is relatively stable, unreacted groups produced less non-specific

adhesion events on AFM curves, TCEP will not cleave the thiol-vinyl sulfone reaction, and because PEG with a vinylsufone group at one end and N-hydroxysuccinimide ester (to react with the amine on the tip) is commercially available.[85]

The first step in preparing AFM tips for sequencing is the same as the one described above for the surface adhesion study. After the tips were removed from the vacuum, they were placed in a solution of vinylsulfone -PEG-NHS (MW: 3200 Dalton) (45 mg, 0.01 mmol) in PBS buffer (1 mL) pH 7 for 15 minutes. This produced tethered PEG molecules terminated by a vinylsulfone (Figure 3.16)**.** The probes were then rinsed in water and used immediately. The maximum number of successful pulls with any one probe indicated that there were typically 15 to 45 functional groups near enough to the end of a probe to be used in the experiment.

3.7. Conclusion

In conclusion, all the necessary pieces for sequencing by AFM were identified, synthesized, and combined (Figure 3.16). β-cyclodextrin was identified as an acceptable nanopore and it was successfully modified with a protected thiol group so it could bind to a functionalized AFM tip. The formation of rotaxanes consisting of the modified cyclodextrin and 1,12-diaminododecane was proven both in solution using MALDI and NMR, and on the surface using ATR-FTIR. The problem of DNA sticking down to the surface (which can significantly complicate force data) was resolved and *n*-propyl-dimethylchlorosilane was chosen as part of the mixed monolayer. In the next chapter I will review some of the mechanics of the AFM, theory of dynamic force spectroscopy

and polymer pulling, and the results of experiments in which the entire construct, including DNA, was successfully built.

4. Attempting Nanopore Sequencing with Force Spectroscopy

4.1. Atomic Force Microscopy (AFM)

The AFM was invented in 1986 by Gerd Binnig, Calvin Quate, and Christoph Gerber.[94] It was originally invented to image insulating surfaces, something the scanning tunneling microscope could not do. Today, in addition to imaging a surface with atomic resolution, the AFM is also used for dynamic force spectroscopy. In our experiments, we primarily used the AFM for force spectroscopy. This technique allows one to measure the force (down to pNs) required to rupture molecular interactions (e.g. covalent bonding, electrostatic attraction, hydrogen bonding) between a (usually modified) AFM tip and surface.

Figure 4.1 shows a schematic of an AFM. Deflections of an AFM cantilever are monitored using the reflection of a laser off the back of the cantilever onto a four diode detector in what is known as the optical lever method. Deflections of the cantilever which cause the reflected laser to move around the face of the detector are translated into a voltage signal. The relationship between the change in voltage and deflection, the sensitivity of the tip, is linear. Once the sensitivity is determined (a necessary step for every AFM cantilever) by pushing the tip into a flat, hard surface, the voltage can be converted to distance and the force can be calculated using Hooke's law,

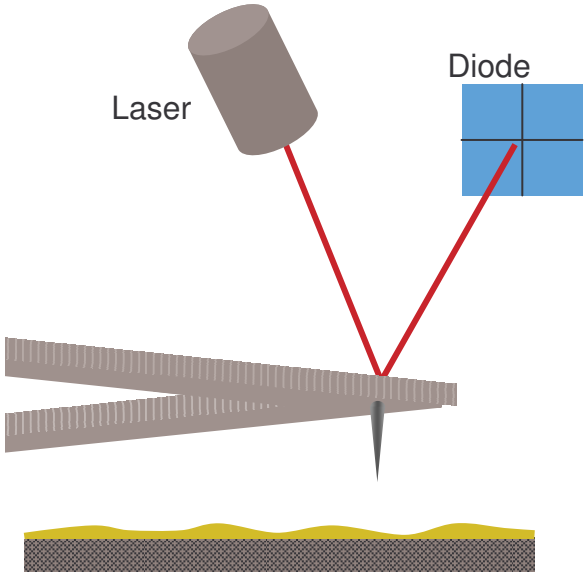


Figure 4.1. Schematic of AFM. A laser reflects off the back of an AFM tip onto a diode.

$F=kx$, where k is the force constant of the AFM tip. The force constant of cantilevers is provided by the manufacturer, but as mentioned later in this chapter determining the force constant of modified tips may be necessary. The movement of the cantilever is controlled by a piezo tube, a tube made of a ceramic that responds to applied voltages (accomplished by an electrode coating around it) by either contracting or dilating the material. It is through this applied voltage across the tube, that the tip is moved up and down, as well as laterally, with atomic resolution.

The speed with which the nanopore can be pulled over the DNA and the length of DNA that can be used are both limited by the capabilities of the AFM. Currently, the range for z-direction movement for an AFM can be as far as 8μm.[95] If the tip is moved at a speed of 100nm/sec then in 80sec the tip can travel 8μm and the nanopore can travel over an approximately 12,000 base long strand of DNA (based on a base to base separation of 0.6nm). This means that 2 million bases could be sequenced in 3.7hrs, as opposed to the currently available automated sequencing instruments that take a full day to sequence the same amount.[95]

The tip of the cantilever can be considered a spring with force constant, k. Depending on the purpose of the experiment, a tip is chosen based on its force constant and resonant frequency. For imaging, softer tips (k~0.01N/m) can be used to avoid damaging samples while, with force spectroscopy, stiffer tips (k~1N/m, up to 100N/m) can be used to help reduce noise. Figure 4.2 shows the progression of an AFM tip during a force spectroscopy experiment as it approaches a surface, binds to a molecule on the surface, retracts, and the bond is ruptured.

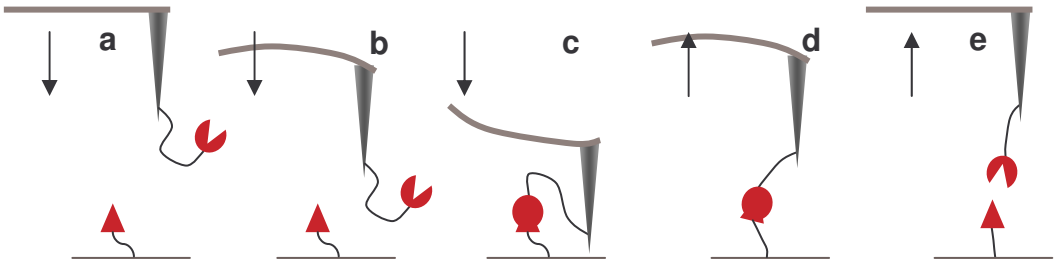


Figure 4.2. A representation of the response of an AFM tip during a force spectroscopy experiment. Panel a) shows a functionalized tip approaching a functionalized surface, b) the tip deflects towards the surface due to the attraction of the tip to the surface, c) the tip is pressed into the surface until a preset trigger value of deflection occurs, during which time the groups on the tip and surface bind, d) the tip is withdrawn from the surface and the tethers connecting the two groups are stretched under the force, e) the rupture force is reached and the bonds between the groups are broken and the tip returns to its original position with no deflection.

The ability to measure the forces on the pico-Newton scale is limited by noise. Some sources of noise can be minimized by performing experiments in an environment that blocks stray light from hitting the diode detector and protects the AFM from vibrations due to nearby traffic, people, etc. The majority of the experiments that follow were done at night to limit outside vibrations. Electrical noise can also be decreased with properly grounding the instrument; however there will be some unavoidable noise due to pick-up of stray electronic noise. Even after optimizing a system to decrease the sources of noise just mentioned, there will be noise from the thermal fluctuations of the cantilever with length, *L*. This is usually derived from the equipartion theorem, $\frac{1}{2}k(\Delta z)^2 = \frac{1}{2}k_B T$. This makes the minimum noise $\Delta z = \sqrt{\frac{k_B T}{k}}$. However, this approximation does not take into account all the vibrational modes of the cantilever or that it is the inclination ($\frac{dz(L)}{dx}$) not

the deflection ($z(L)$) that is actually measured with the optical lever method. With these considerations, the minimum noise becomes,

$$\Delta z = \sqrt{\frac{4k_B T}{3k}}\ .$$ [96]

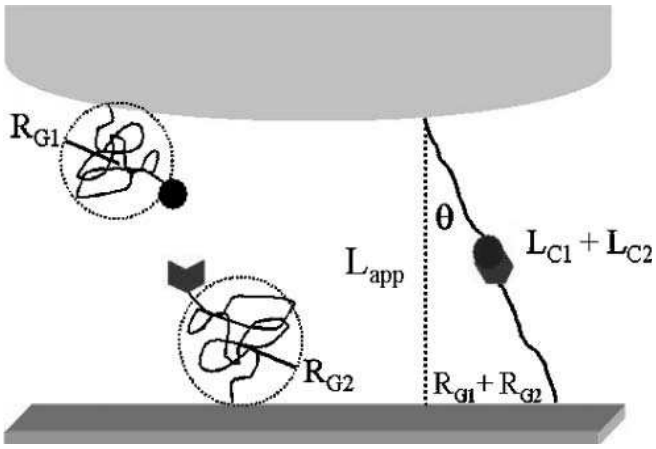


Figure 4.3. Relationship between the measured distance (L_{app}) between tip and surface and the angle (θ) formed between bonded tethers (L_{C1} and L_{C2}) with radii of gyration, R_{G1} and R_{G2}.[97]

Tethers are often employed to link reactive groups to surfaces and tips – our experiments used poly(ethylene glycol), abbreviated PEG. This allows for the interesting part of the experiment (the change in forces or bond rupture) to occur far from the surface, and away from the non-specific adhesion that usually occurs closer to it. It also decreases the occurrence of rebinding during the experiment. This, however, results in a distribution of distances at which bond rupture occurs. The major reason is that not all PEG polymers used in a set of pulls will have the same length. Another cause is that binding of the tethered groups may occur at some angle from the measured vertical distance from tip to surface, which may vary from one pull to another (Figure 4.3).[97]

At the timescale of AFM dynamic force experiments the breaking of bonds is still a thermally activated process.[98] When force is applied, the effect is to change the probability that the bound state will reach the top of the activation energy barrier via thermal fluctuations.[99] Without the application of force there are multiple paths that can be taken, with the force (f) a path is selected because the energy barrier is decreased by

that force times the projection of the force onto the reaction coordinate, $x_\beta = \langle x_{ts} \cos\theta \rangle$ where x_{ts} is the distance to the transition state. The new energy barrier becomes $E_b(f) = E_b(0) - fx_\beta$ (Figure 4.4).[99]

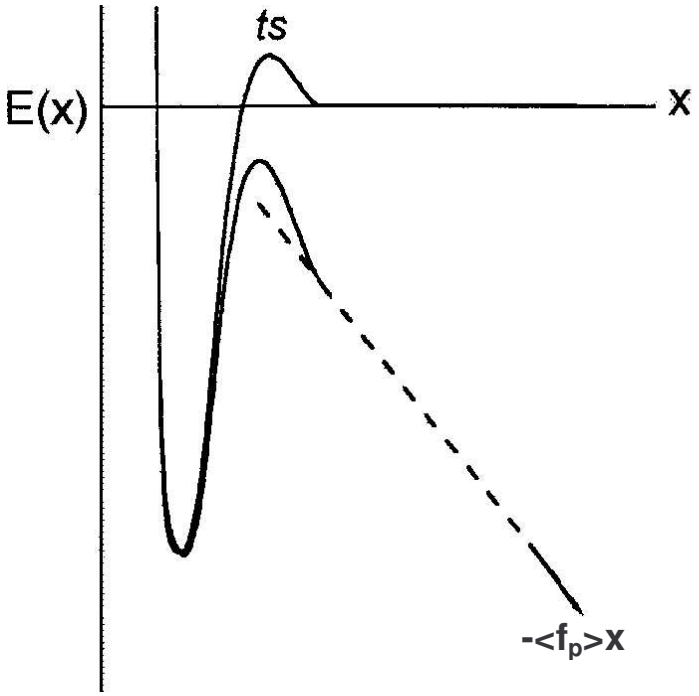


Figure 4.4. Decrease in the energy barrier by the thermally averaged projected force $\langle f_p \rangle$ multiplied by the molecular coordinate, x.[99]

The use of tethers decreases the chance of rebinding during these force spectroscopy experiments, this makes the chance of a bond surviving over time,

$$\frac{dS_1}{dt} = -k_{\rightarrow}(t)S_1(t) \qquad 4.1$$

where $S_1(t)$ the likelihood of being in the bound state and $k_{\rightarrow} = \frac{1}{t_{off}} e^{(\frac{f}{f_\beta})}$, the rate of unbinding or escape where t_{off} is the time it takes for spontaneous disassociation and $f_\beta = \frac{k_B T}{x_\beta}$.[99] By using $\frac{df}{dt} = r_f$, the loading rate, which is linear because constant velocity pulling leads to a linear increase in force with time, bond survival can be described in terms of force by

$$\frac{dS_1}{df} = -\frac{1}{r_f} k_{\rightarrow}(f) S_1(f). \qquad 4.2$$

Integration of Equation 4.2 gives the likelihood of being bound as

$$S_1(f) = \exp\left[-\int_0^f \frac{k_{\rightarrow}(y)}{r_f} dy \right] \qquad 4.3$$

Multiplying Equation 4.3 by the unbinding rate gives the distribution of rupture between f and $f + \Delta f$.

$$p(f) = k_{\rightarrow} \exp\left[-\int_0^f \frac{k_{\rightarrow}(y)}{r_f} dy \right] \qquad 4.4$$

Substituting in $k_{\rightarrow}$ and integrating gives

$$p(f) = \frac{1}{t_{off}} e^{\frac{f}{f_\beta}} \left(\exp\left(\frac{f_\beta}{r_f t_{off}} \left[1 - e^{\frac{f}{f_\beta}} \right] \right) \right) \qquad 4.5$$

The peak of this distribution will be the most probable force at unbinding, f^*. Before taking the derivative of Equation 4.5 and setting it to zero in order to find f^*, taking the natural log simplifies the calculus without changing the location of the peak,

$$\frac{d}{df}\left(\ln\left(p(f)\right)\right) = \frac{1}{f_\beta} - \frac{1}{r_f t_{off}} e^{\frac{f^*}{f_\beta}} = 0 \qquad 4.6$$

Remembering that $f_\beta = \frac{k_B T}{x_\beta}$, the most probable unbinding force is

$$f^* = \frac{k_B T}{x_\beta} \ln \frac{r_f x_\beta t_{off}}{k_B T} \qquad 4.7$$

For our system, it is not the breaking of covalent bonds that we are interested in but, as discussed later, Equation 4.7 can also be used to find the most probable force for each base to translocate through the cyclodextrin and for the opening of hairpins.

Work is required to extend DNA because of the reduction in conformational entropy. DNA is a polymer and the energy associated with extending a polymer has components associated with bending, twisting, stretching, and twist-stretch coupling.[100] The twist components can be ignored because the monomers of most polymers are free to rotate around and in most experiments extending polymers (including our own) the polymer is free to swivel around a single attachment point.[100] Additionally, the stretching term can be eliminated if we pull at a force less than that required to stretch the polymer (DNA), i.e. if we treat it to have a fixed full length, the contour length, L_c.[101] As a result, DNA can be modeled based on the linear elasticity of a thin, uniform rod.[101] This is done using the worm like chain model (WLC). The force to extend DNA along the x axis from the WLC model is,

$$f(x) = \frac{k_B T}{L_p}\left(\frac{1}{4(1-\frac{x}{L_c})^2} - \frac{1}{4} + \frac{x}{L_c} \right) \qquad 4.8$$

where L_p is the persistence length. L_p = 0.6nm for ssDNA, as mentioned in Chapter 1. In our configuration, however, there is DNA and a PEG linker being extended and its contour length and persistence length (0.3nm) must be considered also.[102, 103] As a result the total contour length is the contour length of the two sections added together; $L_{ctot} = L_{cd} + L_{cp}$, where L_{cd} and L_{cp} are the contour lengths of the DNA and PEG,

respectively. The persistence length of the total construct becomes $L_{ptot} = \frac{L_{pd} L_{pp} \left(L_{cd} + L_{cp}\right)^2}{\left(\sqrt{L_{pp}} L_{cd} + \sqrt{L_{pd}} L_{cp}\right)^2}$, where L_{pd} and L_{pp} are the persistence lengths of the DNA and PEG, respectively.[85]

4.2. Sequencing Attempts Using Surface Bound Rotaxanes and AFM

4.2.1. Materials and Methods

Modified DNA was acquired from Integrated DNA Technologies (Skokie, IL) and purified by HPLC. All water was 18 mΩ from Nanopure Diamond of Barnstead. Scanning probe microscopy measurements were carried out on a Molecular Imaging PicoPlus system working in contact mode for force spectroscopy and magnetic tapping mode for imaging. Mikromasch CSC-11 (0.35 nN/nm) cantilevers were used for the force spectroscopy. A Picoplus AFM (Molecular Imaging/ Agilent) was used in acquiring force data.

The attachment of DNA to the PEG already on the surface (a full description of preceding steps is in Chapter 3) was done by first deprotecting the protected thiol on the 5' end with tris (2-carboxyethyl)phosphine hydrochloride (TCEP), 0.02% in 50mM phosphate buffer solution (PBS), pH 7 for 15 minutes and running it through a Sephadex G-25 size exclusion column before use, leaving a thiol linked to the DNA via a hexane linker. The thiolated DNA (0.75 nmol in 17 μL of PBS solution) was deposited on the CD-rotaxane surface and allowed to react for 2 hours with the vinylsulfone group on the PEG. The final step in preparing the system was deprotecting the thiol on β-CD-SS-COOH. This was accomplished by adding TCEP (0.02% in 50mM PBS, pH 7) for 20

min to the surfaces. The surfaces were then rinsed with 50 mM PBS, pH = 7, and force curves were obtained in PBS using the liquid cell of a Picoplus AFM (Molecular Imaging/ Agilent).

Dr. Brian Ashcroft interfaced the AFM to a LabView control system via a custom modification of the AFM head electronics. The AFM approach was controlled with custom software using the Measurement Studio (National Instruments) in Visual Basic (Microsoft)**.** Once all reagents were activated, the sample was placed immediately into the liquid cell of the microscope and covered in PBS buffer. A freshly-functionalized probe was inserted into the scanner and calibrated after each run using the thermal spectrum method.[104] The probe was lowered onto the surface using Measurement Studio until a deflection increase of ~100 pN was detected, held for 6-8 s to allow the linkers to react, and then retracted while force-distance curves were recorded. Pulls showing significant deflection at a distance greater than 50 nm from the surface were flagged for further analysis. Typically, 80 retractions from the surface were required to locate one successful pull because the substrate surface had been sparsely functionalized in order to present only an individual linker for reaction with the tip for any given pull. Each probe lasted ~800 to 1000 pulls before all the active tethers were depleted (through the collection of ~15 to 45 curves that were indicative of successful attachment to the β-CD-SS-COOH). Each sample survived through 3 probe replacements to give a total of 3000 curves per run. The end of the experiment was signaled by the appearance of spurious features at distances greater than the sum of the linker and DNA lengths

(indicative of contamination). 2000 data points per second were collected for each force-distance curve at pulling rates that varied from 30nm/s to 600nm/s. Displacement data were corrected for the tip deflection, and all plots shown here reflect the true probe-sample distance.

4.2.2. Control Experiments

Once all the different pieces for this technique were developed and combined on the surface (see Chapter 3) the next step was to try to sequence a strand of DNA. Control experiments were first performed in order to confirm that the functionalized AFM tip was able to bind to the modified β-CD and pull it up over surface bound DNA.

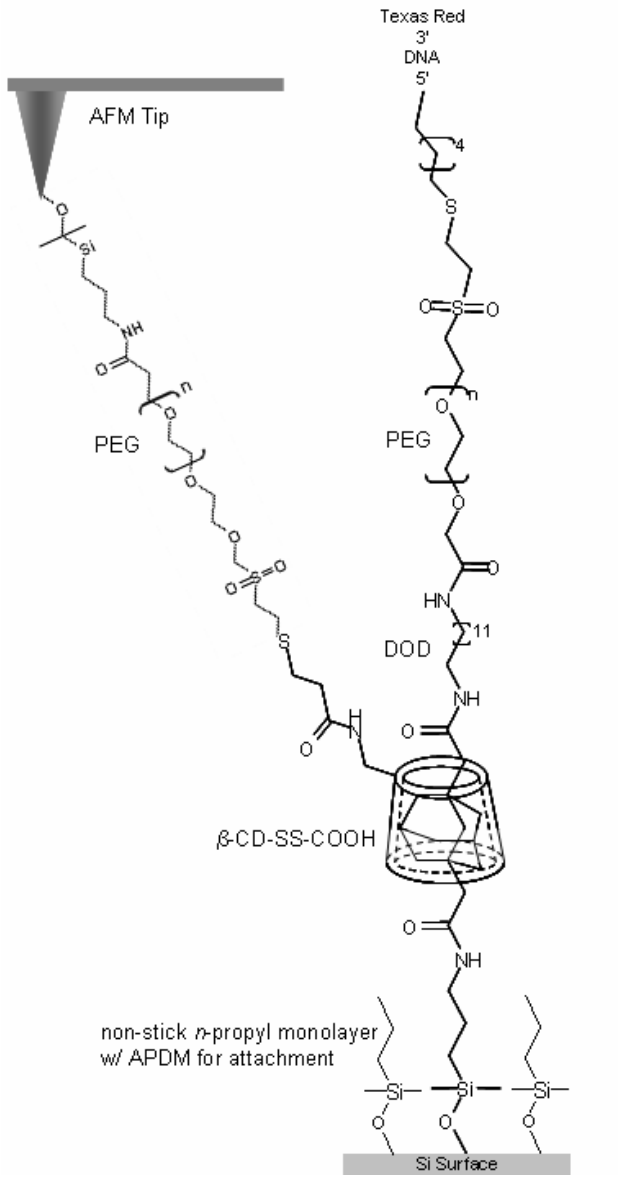


Figure 4.5. Full construct for sequencing using the AFM. The AFM tip is bound to deprotected β-CD-SS-COOH.

These experiments used DNA that had a large group bound at its free end (Texas Red dye, see Figure 3.21 A). As the cyclodextrin slide along the DNA it would be "snagged" by the large stoppering group and only once the force needed to break covalent bonds (>1nN) was reached would the AFM tip snap back to zero deflection.

4.2.3. Results

Figure 4.5 shows the finished experimental construct for sequencing by AFM. Fully extended, the 5'-thiolated oligo-T_{35} stoppered at its 3' end with a Texas Red dye molecule (Figure 4.6 A) has a length of 21.4 nm (using a stretched base-to-base distance of 0.6nm).[80] Each PEG linker (one on the surface and one on the probe – see Figure 4.5) has a Gaussian distribution of fully-stretched lengths with a mean value of 28 nm and a half-width at half-height (HWHH) of 2.8 nm.[97] The other minor linking components (APDM, DOD, adamantane) add a further 4.4 nm at the surface and 1.2 nm at the probe.[85] As a result, the cyclodextrin-DNA interaction cannot occur until the probe and surface are separated by 62±5.6 nm.

Examples of force curves showing features beyond 62nm are given in Figure 4.6 B. These curves reasonably fit the worm like chain model of DNA stretching (Eq. 4.8) with peak lengths (L_C) near the 21 + 62 = 83 nm expected.[103, 105] Control experiments using DNA that lacked a large stopper did not show these large force peaks. The distribution of measured peak forces is shown in Figure 4.6 D. The largest peak forces in control experiments (white bars) lacking a β-CD or using an unfunctionalized β-CD, were significantly smaller than the largest peak forces measured for the full construct (black bars). These large forces (≥1nN) indicate that covalent bonds are being broken, the expected result once the cyclodextrin reaches the end of the DNA and is snagged by the Texas Red while the AFM continues to pull it up.[106]

The distribution of fitted contour lengths is shown in Figure 4.6 C (black bars). There are many events below 62 nm, but these are also observed in control experiments (white

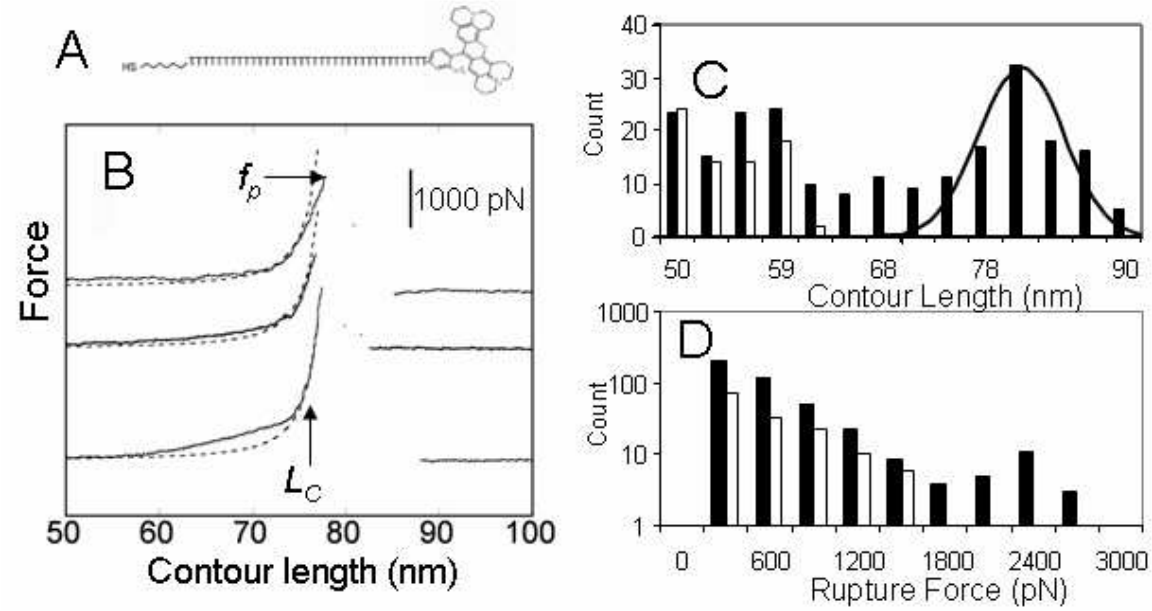


Figure 4.6. Passage of DNA tethered to the surface via PEG linker with Texas Red dye bound to the end of the DNA in order to snag the cyclodextrin. A) The T_{35} oligomer is thiolated at its 5' and Texas Red at the 3' end. B) Examples of force-distance curves (with arbitrary vertical displacements) that show features beyond the stretched tether length of 62 nm (all distances are corrected for tip displacement). The polymer contour length, L_C is derived from fits of the worm-like chain model (dashed lines). C) Distribution of measured values of L_C; black bars are data for the full construct using all curves that yielded data beyond 62 nm, white bars show data for controls lacking β-CD-SS-COOH or taken with a unmodified β-CD. The solid line is a Gaussian centered at 80 nm with a HWHH of 5.6 nm. D) Distribution of peak rupture forces; black bars are for the full construct, white bars are the controls.[85]

bars) in which the β-CD was not functionalized or was omitted entirely. Thus, the features below 62 nm correspond to various types of non-specific interactions between the components of the system. The mean length of all pulls with features above 62 nm lies near the predicted mean of 82 nm and the distribution follows a Gaussian with a HWHH close to the expected 5.6 nm.

4.2.4. Sequencing Attempts

Experiments performed using DNA with a large dye molecule on the end proved that the construct was formed as expected on the surface and that it was possible for a

functionalized AFM tip to find, bind to, and pull the nanopore over the DNA. DNA without any end modifications was then used to collect data in order to find sequence-specific force changes. A strand of alternating purine (A) and pyrimidine (C) that will not form hairpins through Watson-Crick base pairing was used, again with a protected thiol group on the 5' (5'-$(AAA\ CCC)_7$AAA-3') end. The fully extend length of the DNA plus PEGs is 88±5.6nm. Pulling was performed in the same manner as the control experiments with a 100nm/s pulling speed (loading rate = 35nN/s). The results of these experiments are in Figure 4.7 and show they lacked the large force at the end that was present with the stoppered DNA. They show that there is occasionally a small force at the extended length of the DNA (Figure 4.7 A). It is too small for covalent bond breaking and is most likely due to the CD forming a complex with the terminal base, causing it to adhere to the end of the DNA.[107] There is also no distribution of pulls at the extension points where the strand changes from purine to pyrimidine (Figure 4.7 B). Due to the lack of force curves reported these experiments were repeated three times without any change to the results.

4.2.5. Results

Based on these experimental results, it is not possible to determine the sequence of a single strand of DNA with this method. The minor difference in size between the four bases and their nearness on the strand make it difficult to detect any changes in force. When DNA is driven through a nanopore the bases tilt towards the 5' end.[108] This difference causes a measurable difference in blockage current in experiments where is it the ionic current that is measured when DNA translocates (Chapter 1) depending on

whether DNA passes through in the 3'-to-5' or the 5'-to-3' direction.[108] A similar situation arises when pushing a Christmas tree though a door – the branches tilt towards the tip so it is easier to push the tree through trunk first. In the same way, ssDNA passes through a pore more easily in the 3'-to-5' direction. Our sequences have the thiol on the 5' end and so the resistance of the bases to pass through the cyclodextrin (and therefore the force needed to pull the cyclodextrin) should already be maximized due to the directionality of the DNA.

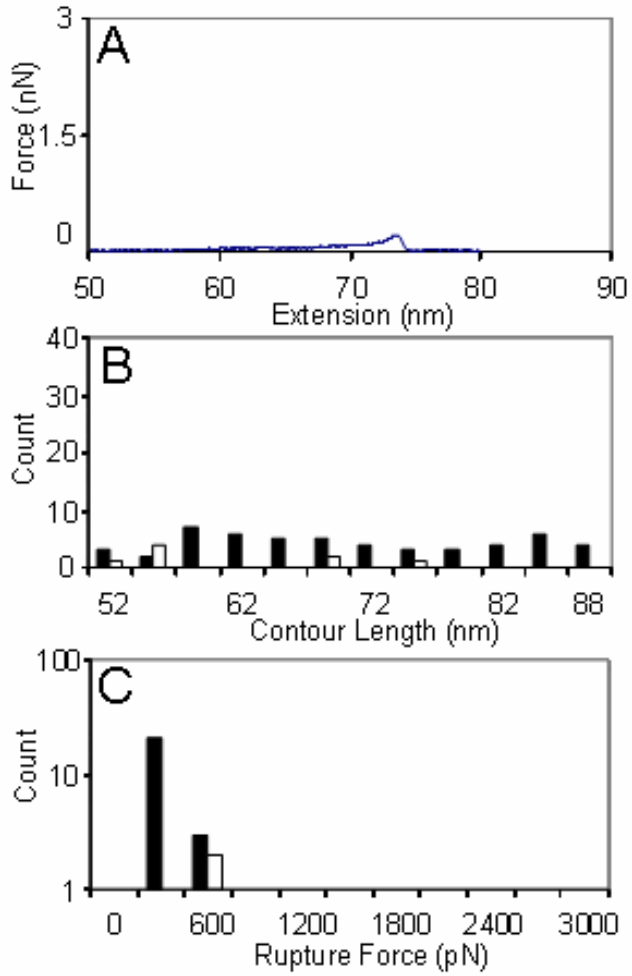


Figure 4.7. Results of sequencing experiments using unstoppered DNA. Panel A shows an example of a force vs. extension curve that shows the small interaction between the CD and end of the DNA. Panel B shows there is no Gaussian distribution of extension lengths at the boundaries between adenine and cytosine. Panel C is the distribution of measured forces. All fall below that for breaking covalent bonds; the cyclodextrin is able to clear the entire length of DNA. Black bars are with cyclodextrin and white are for control experiments performed without cyclodextrin.[109]

The result of a lack of sequence information or even a distinction between purines and pyrimidines is not surprising based on modeling work performed by Shahid Qamar after our experimental results showed no sequence specific responses. He used a combination of steered molecular dynamic (SMD) simulations and

milestoning to determine at what speeds the difference in forces are large enough to detect and what the difference is at realistic AFM pulling speeds (100nm/s).[110] The SMD results were determined using speeds on the order of meters per second to pull β-CD over first a poly(ethylene glycol) thread and then a nucleoside (a sugar ring and a base). The results showed that at high enough speeds there is a force difference that would, in principle, be measurable with an AFM (Table 4.1). These speeds are not experimentally feasible, however, so Equation 4.7 is used to get more realistic results. This method requires the calculation of the activation energy barrier, ΔE, and diffusion rate constant, k_o, because $t_{off} = \frac{1}{k_o \exp\left(\frac{\Delta E}{k_B T}\right)}$. This was accomplished using milestoneing; splitting the space between the initial and final result (in this case the cyclodextrin starting on one side

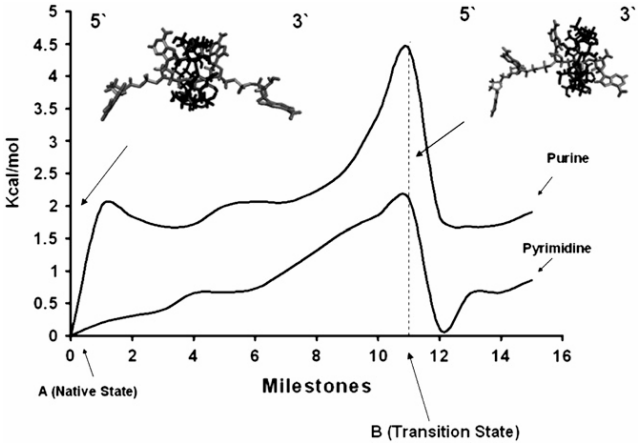


Figure 4.8. Energy profile, from milestoning data, for the passage of β-cyclodextrin over a purine and a pyrimidine. The peak is the transition state. The insets show β-CD initially on the left of the base and then on the right. A and B indicate the points used for computing transition rates over the barrier.[110]

Table 4.1. The forces required to pull β-CD over the different bases at different speeds.[110]

Speed (m/s)	A (pN)	G (pN)	C (pN)	T (pN)
0.2	230	300	100	120
0.3	300	350	120	150
0.4	350	400	150	180
0.5	425	500	210	260
0.6	500	580	315	380
1	700	600	400	450

and ending on the other side of a base) into equal parts and, using simulations, determining the rate of movement of the cyclodextrin randomly walking from one milestone to the next. These rates (k_f is forward and k_r is reverse) give the activation energy at each milestone through, $\Delta E = -k_b T \ln\left(\frac{k_f}{k_r}\right)$. A plot of these energies vs. milestone determines the maximum free energy, the activation energy barrier (Figure 4.8). For purines it is 4.5kcal/mol, more than twice that for pyrimidines (2kcal/mol). The diffusion rate constant, k_o, is determined by taking the slope of the line when $k_o t = -\ln(1 - P_B)$ vs. time is plotted, where P_B is the probability of the cyclodextrin being in state B (defined as being just to the right of the transition state, the base it passes over). Plugging these values into Equation 4.7, and using a distance to transition state, X_{ts}, of 0.2nm, yields forces dependent on loading rate, as shown in Figure 4.9. Again, at realistic loading rates (with a AFM pulling speed of 100nm/s and our tips the loading rate is 35nN/s) the difference in force is too small (<20pN) to be detected with an AFM.

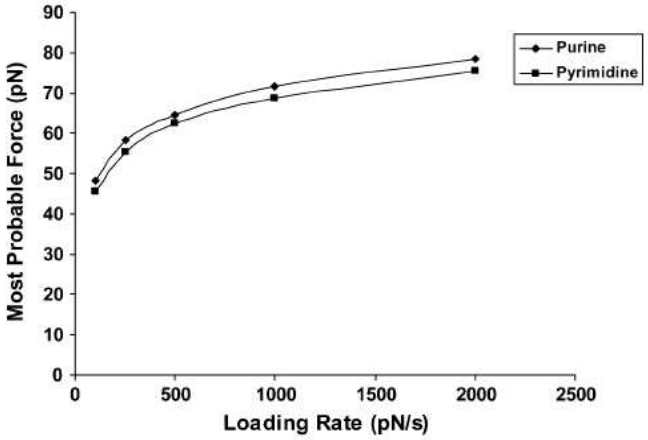


Figure 4.9. Plot of the most probably force for either a purine or pyrimidine vs. loading rate based on values calculated using Equation 4.7.[110]

While this method of sequencing proved to be unfeasible there may be ways to improve the system. The simulations above implied that if we could increase the speed of translocation (ironically the very opposite problem of other nanopore sequencing methods) the difference in force between the bases may rise above the noise. This is

limited, however, by the actual construct; pulling the rotaxane free of the surface before the cyclodextrin progresses down the DNA because of a high loading rate defeats the purpose in the first place. Another possibility may be to add a second group onto the cyclodextrin. The work in Chapter 3 on solution phase rotaxanes showed that some modifications may more strongly interact with DNA. This molecular friction, whether due to base stacking (β-CD-pyrene), base pairing (β-CD-thymine), or electrostatic repulsion/attraction (β-CD-COOH, β-CD-guanidinium) might be able to increase the signal from each base. Another possibility would be to use a smaller cyclodextrin ring (α-cyclodetrin has 6 glucose rings, for example) or make base-specific modifications to the DNA and run the experiment for each base, overlaying the results.

4.3. Secondary Structure Study

While this method proved to be ineffective in determining sequences, the situation of a pore passing over DNA mimics the way RNA polymerase slides over DNA during transcription. Usually, when studying the formation of secondary structures of DNA or RNA (such as hairpins) using force spectroscopy, the two ends that go on to base pair and form the stem of the hairpin are pulled apart. [47] This is not how polymerase encounters a hairpin while transcribing DNA. The sliding clamp must break open the hairpin through a shearing force applied at the bases that begin the stem of the hairpin, not by pulling apart the ends.[103] This situation is similar to our setup, where only one end of the DNA is tethered and the ring is pulled over the DNA. Additionally, the AFM allows us to know the force on and the position of the pore at the same time. This is different from experiments where a strand of DNA with a hairpin is driven through a nanopore

electrophoretically – with such methods it is impossible to directly measure the force to break open the hairpin, and the location along the strand where the hairpin occurs cannot be determined. [108, 111, 112]

4.3.1. Hairpins

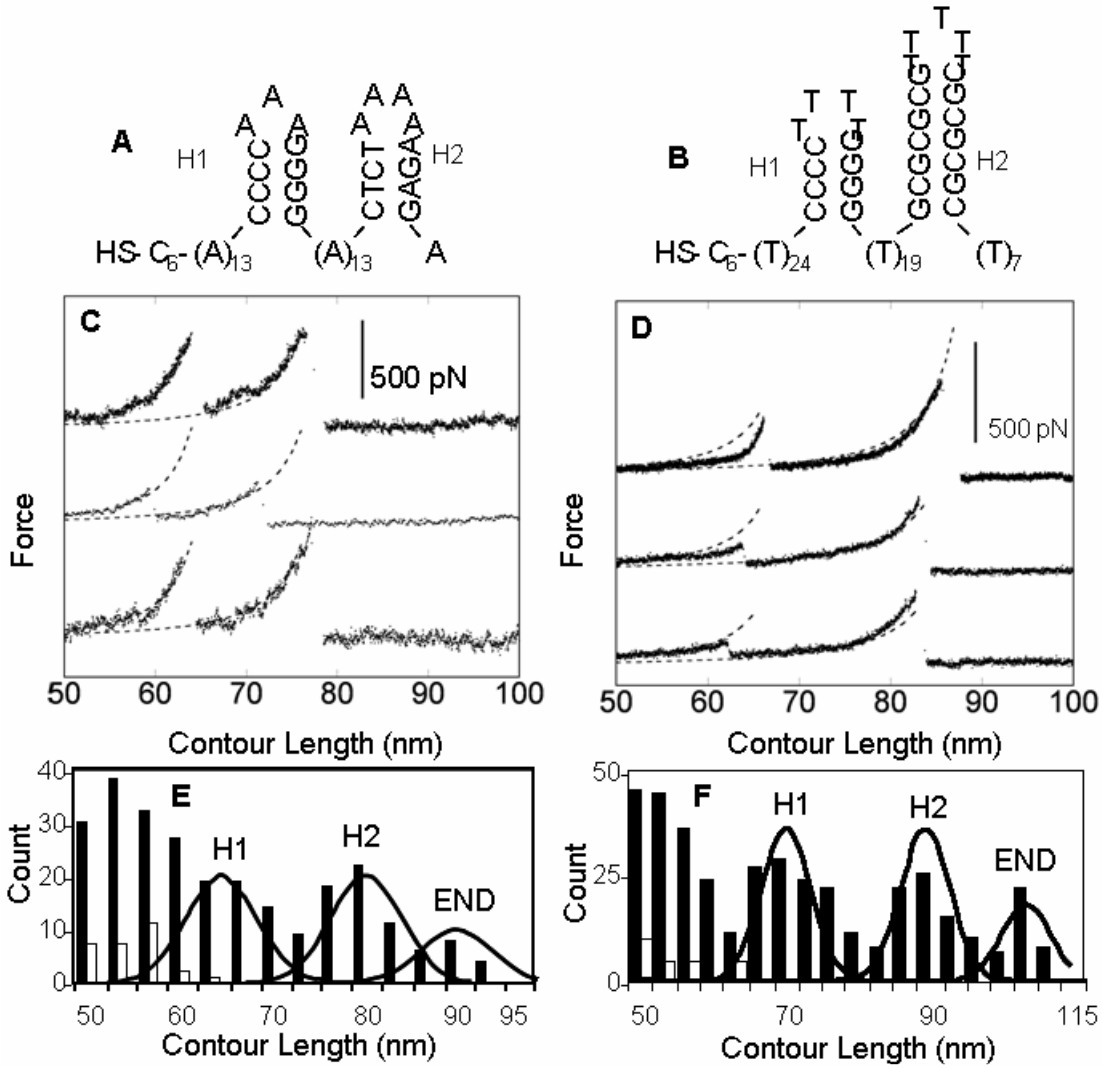


Figure 4.10. Results from pulling β-CD over DNA with hairpins. Panels A and B show the structure of the two DNA strands studied based on mFold calculations. Panels C and D show force curves with pulls at both hairpins, the dotted lines are fits to Equation 4.8. Panels E and F show the measured contour lengths of the two DNA strands. Control experiments either without β-CD or unfunctionalized cyclodextrin show no events beyond 62 nm (white bars).[85]

For the experiments to study the force required to open hairpins two different DNA strands were used.[85] They were attached to the surface and data was collected in the manner described above. Each strand has two hairpins, this way if two peaks are present in the force curve then the first force peak is due to the opening of the hairpin and not the breaking of a covalent bond. The lowest energy structures for these two strands (one with 50 bases, one with 81) were determined using mFold (Figure 4.10 A and B).[95]

Force curves collected using the DNA with double hairpins show events that correspond to the extended length of the DNA before each hairpin (Figure 4.10 C and D) In order to predict the expected extensions (Table 4.2), only two hairpin states were assumed: open and closed.[113-115] If a hairpin was closed at the time of the first encounter, the β-CD should stick at the first paired base in the stem. The measured contour length distributions are in good agreement with the predicted lengths, and the widths of the distributions are close to the expected 5.6 nm (Table 4.2, Figure 4.10 E and F). Similar to the sequencing experiments, there was also a pull-off force as the cyclodextrin cleared the DNA (Table 4.2, end).

4.3.2. Results

Although the free energy for forming these hairpins is favorable, it is also small (Table 4.3). The ratio of the number of curves with a single first hairpin event and the ones that show both indicate that the first hairpin is spontaneously open (no applied force) 30% of the time for both the 50 and 81 base strands. The free energy based on this ratio can be calculated using,

Table 4.2. Predicted (*d)* and measured (L_C) distances to peaks in force-distance curves. Measured values are systematically a few nm lower than the predicted values, this is probably due to tip geometry (radius ca. 5 nm, so that most active tethers are suspended some nm above the point of contact). L_C was obtained from fits of the WLC model. Corresponding values of the effective persistence length L_p are listed as the mean values and a range (because the s.d. usually exceeds the mean). The number of curves analyzed is *n* (with the subset that yielded reliable values for L_p in parenthesis).[85]

Feature	d (nm)	L_C (nm)	*Lp* (nm)	*N*
1st hairpin, 49 base oligomer	70±5.6	66±5	0.4 (0.05 to 1.0)	58(36)
2nd hairpin, 49 base oligomer	85±5.6	81±5.6	0.3 (0.05 to 0.8)	66(27)
End, 49 base oligomer	91±5.6	89±5	0.3 (0.05 to 0.9)	15(7)
1st hairpin, 81 base oligomer	78±5.6	73±6	0.4 (0.05 to 0.8)	125(20)
2nd hairpin, 81 base oligomer	96±5.6	94±6	0.3 (0.05 to 1.0)	81(20)
End, 81 base oligomer	110±5.6	103±8	0.3 (0.05 to 1.0)	37(4)

$$\Delta G = -RT \ln\left(\frac{open}{closed}\right) \qquad 4.9^{[115]}$$

This gives a free energy for the first hairpin as ΔG = 0.2kcal/mol, well below that predicted by mFold (Table 4.3). This lower value is explained by the fact that if the peak force needed to open the first hairpin falls below the noise (~20pN) then it would not be detected or counted in the closed value.

The force needed to break apart hairpins by methods that pull on the ends of the DNA is in the range from 10 to 15pN.[115, 116] The forces required to separate hairpins in our method are much larger, up to almost 400pN (Table 4.3). The change in the geometry of

Table 4.3. Average opening forces for the hairpins (data for the hairpins with identical stems have been combined). x_{ts} and t_{off} were derived from fits to Equation 4.7. Stabilization free energies (ΔG) were predicted by mFold.[85]

Feature	$\langle f_p \rangle$ (pN)	x_{ts} (nm)	τ_{open} (ms)	ΔG_{calc} (kcal/mole)
1st hairpin, both oligomers	376	0.05±0.01	53±56	-5.95
2nd hairpin, 49 base oligomer	248	0.06±0.01	100±37	-2.94
2nd hairpin, 81 base oligomer	371	0.03±0.01	43±26	-12.19

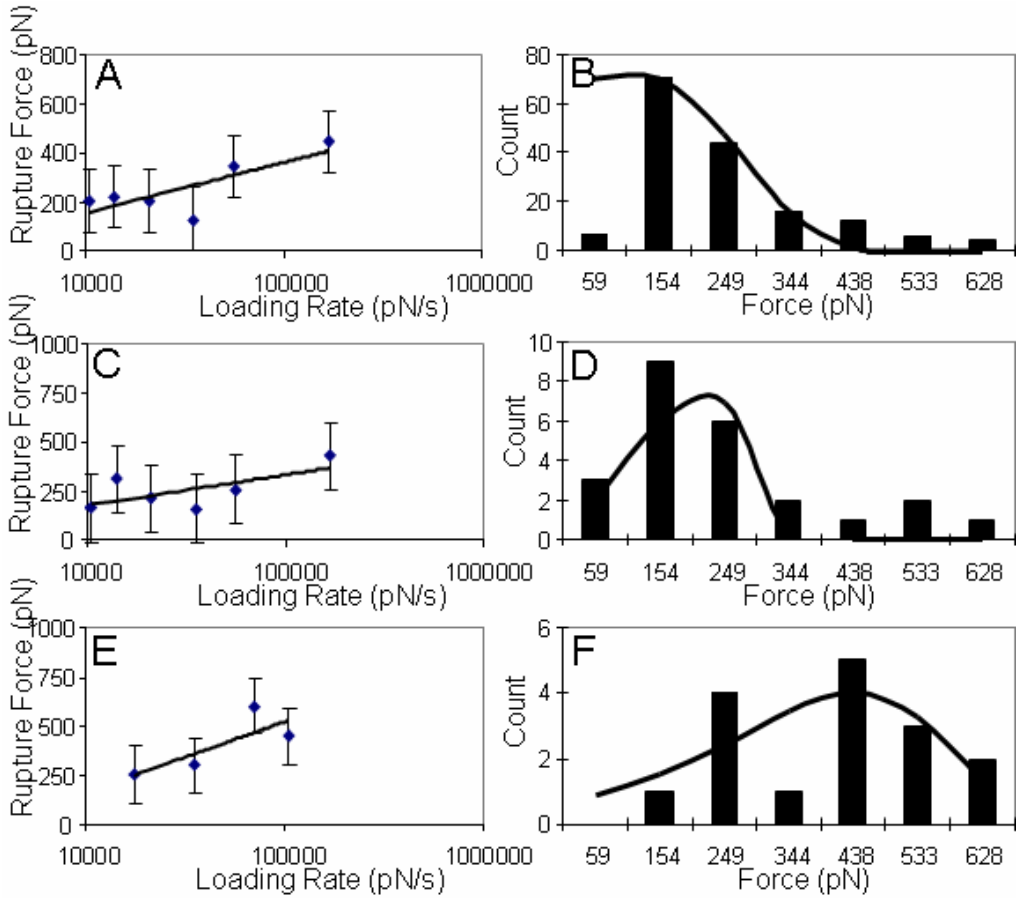


Figure 4.11. Kinetic analysis of hairpin opening: Panels A, C, and E plot the modal force (error bars are ±1 sd) as a function of the logarithm of the loading rate for the first hairpin in both strands (A) (data were identical and are aggregated here), the second hairpin in the 50 base strand (C) and the second hairpin in the 81 base strand (E). The corresponding distribution of measured forces is plotted on the right at loading rates of 10.5 nN/s (B), 21 nN/s (D), and 105 nN/s (F). The solid lines are calculated with Equation 4.5 using the parameters shown in Table 4.3.[85]

unfolding greatly affects the force required. As mentioned earlier, the application of force will tilt the energy landscape towards the unfolded state. Equation 4.5 gives the distribution of unfolding forces and Equation 4.7 gives the most probable unfolding force. With the experimental determination of f^*, Equation 4.7 allows one to solve for the lifetime of a hairpin when no force is applied, t_{off}, and the distance to the transition state, x_{ts} (Figure 4.11 A, C, and E, Table 4.3). In order to check these values they were substituted into Equation 4.5 and are plotted with the experimental distribution of forces (Figure 4.11 B, D, and F). The calculated thermal opening time, t_{off}, is similar to those reported by using other methods.[108, 111, 116, 117] It is the distance to the transition state, x_{ts}, that differs and causes the large difference in force to open hairpins; 0.05 nm with this method as compared to 5 to 26 nm when the force is applied to either end.[115, 116] This means that opening a hairpin by forcing it through a pore requires less strain on the system to destabilize it but more force because it acts over a shorter distance. The path followed for opening a hairpin using a nanopore is "brittle" compared to the pathway chosen in experiments that pull on opposite ends of the hairpin stem.

4.4. Conclusion

Sequencing single strands of DNA by measuring the force required to pull β-cyclodextrin over each base using an AFM is not possible with our current setup. However, a nanopore encircling DNA mimics polymerases and this setup was successfully used to study the shearing forces required to open three different hairpins. This geometry results in the opening of hairpins with a stronger force and less strain as compared to methods that pull apart the ends of the DNA. The diameter of RNA

polymerase II from yeast is 25Å, compared to 15Å at the widest end of β-CD.[118] The slightly tighter fit may increase these values from those actually experienced in nature.

Additionally, studying the behavior of DNA sliding through a nanopore and the breaking of hydrogen bonds between bases has informed the conception of another method of sequencing that uses the scanning tunneling microscope, sequencing by recognition. This method, which will be discussed in greater detail in the next chapter, relies on the interaction of single stranded DNA with a nanopore via hydrogen bonds. Unlike the AFM method, the translocation of the DNA will once again be controlled by an electric field, and the molecular friction due to the hydrogen bonds will be relied upon to slow the DNA's progress.

5. Sequencing by Recognition

5.1. Scanning Tunneling Microscopy

The scanning tunneling microscope (STM) allows a conductive tip to be brought very close (less than 10Å) to a conducting or semiconducting surface (see Figure 5.1). A voltage is applied across the gap between the tip and the surface and electrons tunnel through the space between them, creating the tunneling current. The distance between tip and surface can be understood as the width of a potential barrier (V_0) through which the electrons can tunnel. The wavefunction of the electron in this region consists only of the decay term, $e^{-2\kappa x}$, with $\kappa = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}}$.[120] In experiments where the distance between tip and surface is not a vacuum but is occupied by molecules, that space is no longer a barrier but a set of molecular orbitals. Now the barrier height (ϕ_o in Figure 5.2) becomes the difference between the energy of the metal contacts (E) and the energy of the lowest unoccupied molecular orbital, LUMO (or highest occupied molecular orbital, HOMO, whichever is closer to E, generally E falls in between the two) of the molecule.[121, 122] The decay parameter, $\beta = 2\kappa$, becomes $\beta = 2\sqrt{\frac{2m^*(E_{LUMO} - E)}{\hbar^2}}$, where the mass is replaced by the effective mass (the effective mass is used in order to

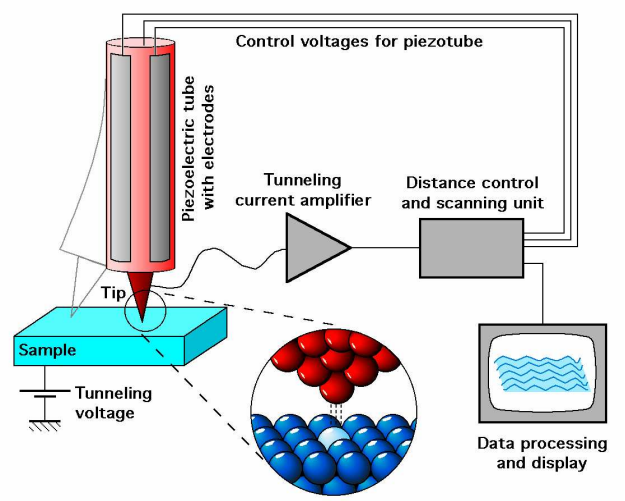


Figure 5.1. Schematic by Michael Schmid and Tu Wien of a scanning tunneling microscope.[119]

take into account the interactions electrons will have as they travel through a material). If a voltage, *V*, is applied to the contact on the right in Figure 5.2 its Fermi energy (E_f) level (at zero bias) will decrease by $eV/2$. Treating both contacts as symmetrical, it is approximated that the left contact's Fermi energy level (at zero bias) increases by eV/2.[122] It is the electrons between the left and right contact's adjusted Fermi levels (after applied voltage) that will dominate tunneling.[122] Tunnel current, *I* is related to the distance of the tip from the surface (x) through an exponential decay, $e^{-\beta x}$.[123] One type of STM imaging is accomplished by maintaining tunneling current through adjustments to that distance. If a region has a small β (tunneling is easier) then the distance must increase in order to keep the current constant. Conductance (G) of a molecule is related to the current through, I=GV. It is determined by setting the applied voltage, measuring current vs. voltage curves, and then taking the slope if the relationship is linear, taking the derivative of I with respect to V if it is not.

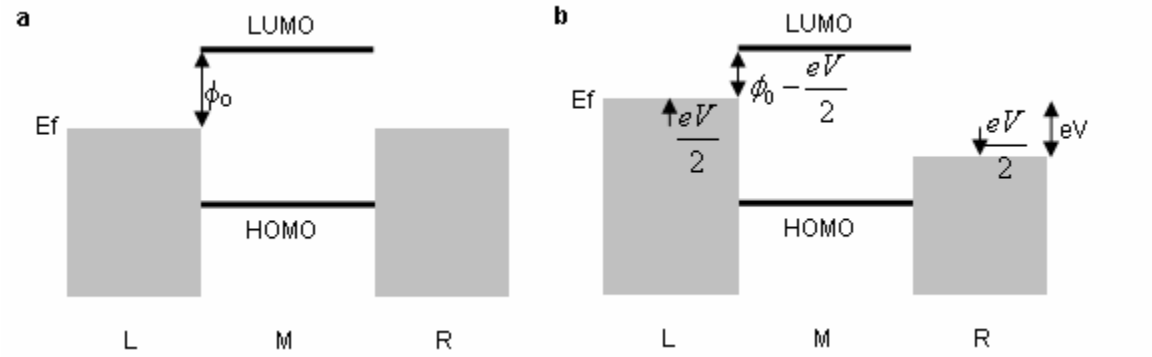


Figure 5.2. On the left (a) is the energy differences of the HOMO and LUMO of a molecule (M) with the Fermi energy levels of the metal contacts (L and R) when no bias is applied with a barrier height, ϕ_0. The right (b) shows the shift in Fermi energy levels after an applied voltage and the new barrier height, $\phi_0 - \frac{eV}{2}$.[124]

5.2. Motivation

In 2006 Ohshiro and Umezawa published on "intermolecular tunneling microscopy."[123] Using STM they studied the electron charge transfer between complementary and non-complementary DNA bases. Using thiol-modified bases (see Figure 5.3) bound to a gold surface and a gold STM tip, they determined that when complementary bases were brought together and formed Watson-Crick base pairing hydrogen bonds there was an increase in tunneling current. When non-complementary bases were used there was a noticeably smaller increase in current. Figure 5.4 shows an STM image of a surface with a mixture of adenine and guanine bases. A cytosine functionalized tip was used to scan the image and the bright spots correspond to its complementary base, guanine (a decrease in β, so an increase in height) and the dark spots to the non-complementary adenine. They expanded their experiments by depositing single stranded peptide nucleic acid (PNA) (similar to DNA except the phosphates and sugars are replaced by peptide linkages) onto a gold surface and probing it using an STM tip with complementary and non-complementary bases. They were able to see points of increased tunnel current along the strands, indicating Watson-Crick base pairing.

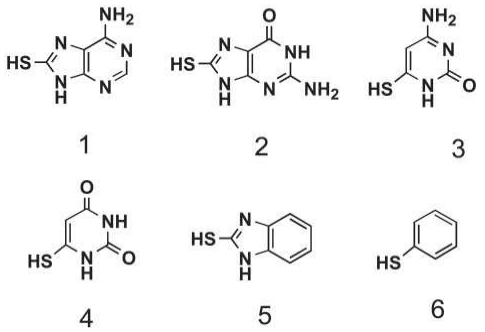


Figure 5.3. Structure of modified bases and controls (5 and 6). 1) adenine, 2) guanine, 3) cytosine, 4) uracil, 5) 2-mercaptobenzimidazole, 6) thiophenol.[123]

It has been shown experimentally that the level of conjugation and twist within a molecule will affect its conductance (an increase in conjugation causes an increase, an increase in twist causes a decrease)(see Figure 5.5).[125] In Ohshiro and Umezawa's

experiments, not only is the current traveling through the bases bound to the gold contacts but must also travel through the hydrogen bonds between them. It is possible for non-Watson-Crick hydrogen bonding to occur between bases (I will discuss these alternatives later in this chapter) but based on their results, complementary bases result in the strongest current. They believe this is because the Watson-Crick base pairing requires the bases to be coplanar (effectively reducing twist).[123]

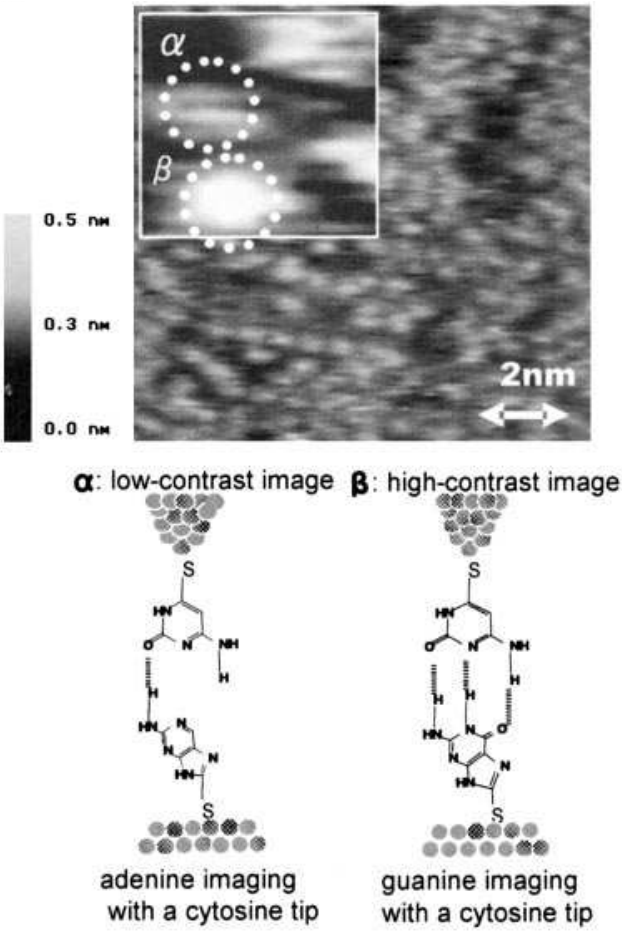


Figure 5.4. STM images of a mixed monolayer of guanine and adenine (10:1 ratio), indicating an increase in tunnel current with complementary bases (β in the blow up of the image) and a decrese with non-complemtary (α).[123]

Based on Ohshiro and Umezawa's work our lab began work repeating their measurements. It is difficult to interpret STM images so a different method was used to determine tunneling through complementary bases.[126] Individual current vs. retraction distance curves (collected as an STM tip withdraws from the surface after forming hydrogen bonds, 0nA indicates that the tip has traveled far enough away that those bonds have been broken) were collected by Dr. Jin He and Lisha Lin which showed a marked difference between complementary,

non-complementary, and control tips (see Figure 5.6).[126] When complementary bases are brought close enough together to form hydrogen bonds, the current vs. distance curves (Figure 5.6 d) are obviously different from those collected when non-complementary bases (Figure 5.6 b and c) or control experiments (only a bare gold substrate, Figure 5.6 a) were run. This work is the foundation for a new sequencing technique being called "Sequencing by Recognition."

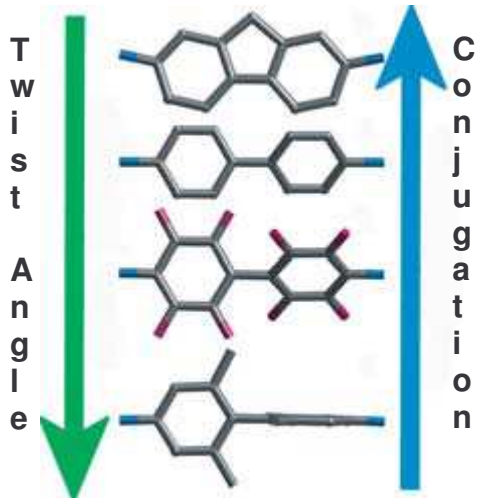


Figure 5.5. Examples of twist angle and conjugation and their relationship to conductance.[125]

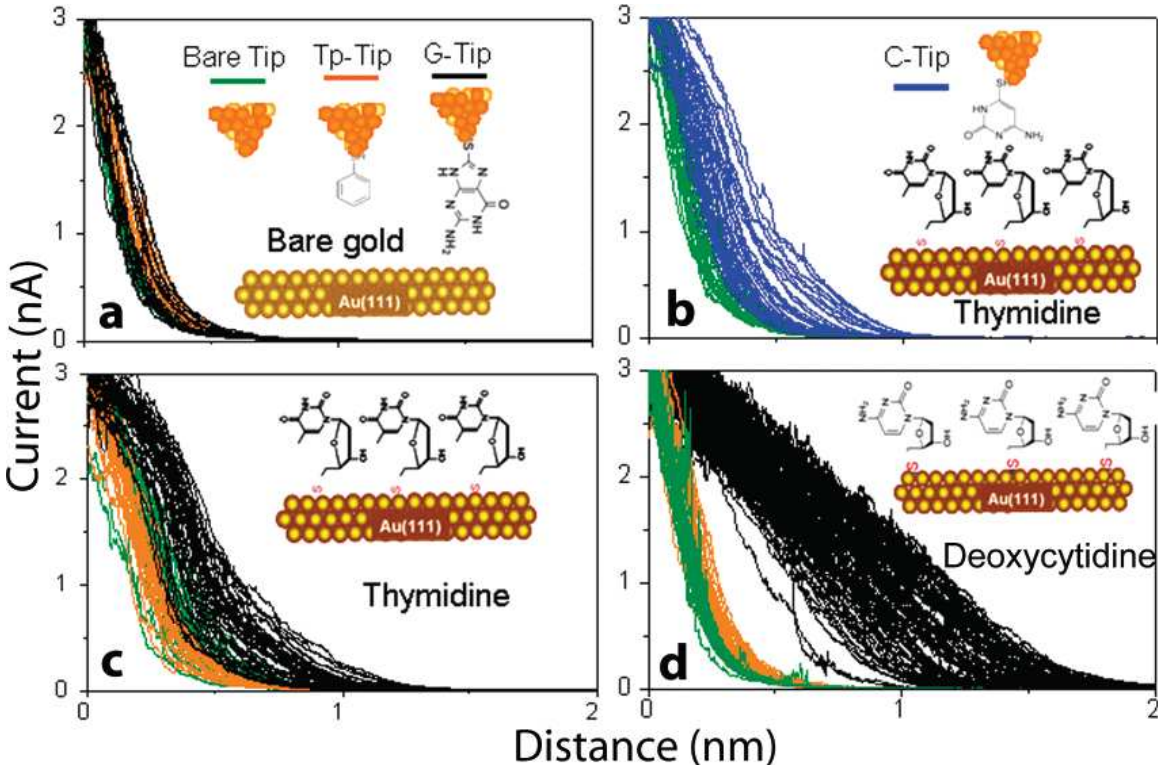


Figure 5.6. Current vs. distance curves of a) control tips (clean, benzenethiol, and guanine with a bare surface, green, orange, and black curves, respectively), b) and c) non-complementary (cytosine with thymidine, blue curves, and guanine with thymidine, black), and d) complimentary (guanine with deoxycytidine, black).[126]

5.3. Sequencing by Recognition

Sequencing by recognition relies on the increased current that flows through complementary base pairs (Figure 5.7). In order to read off a sequence with this method a single base on a single strand of DNA must be probed at a time. A nanopore and nanoelectrodes are used for this purpose. The DNA is threaded through the pore to eliminate any secondary structure and to ensure only one strand is being read by the electrodes, and fed to the nanoelectrodes. The nanoelectrodes are functionalized with guanidinium (a molecule capable of hydrogen bonding to the phosphate backbone, it can "grab" the DNA) on one side and one of the 4 bases on the other (capable of "reading" the bases). As the DNA passes through via an applied electric field, hydrogen bonding takes place with both electrodes, completing the

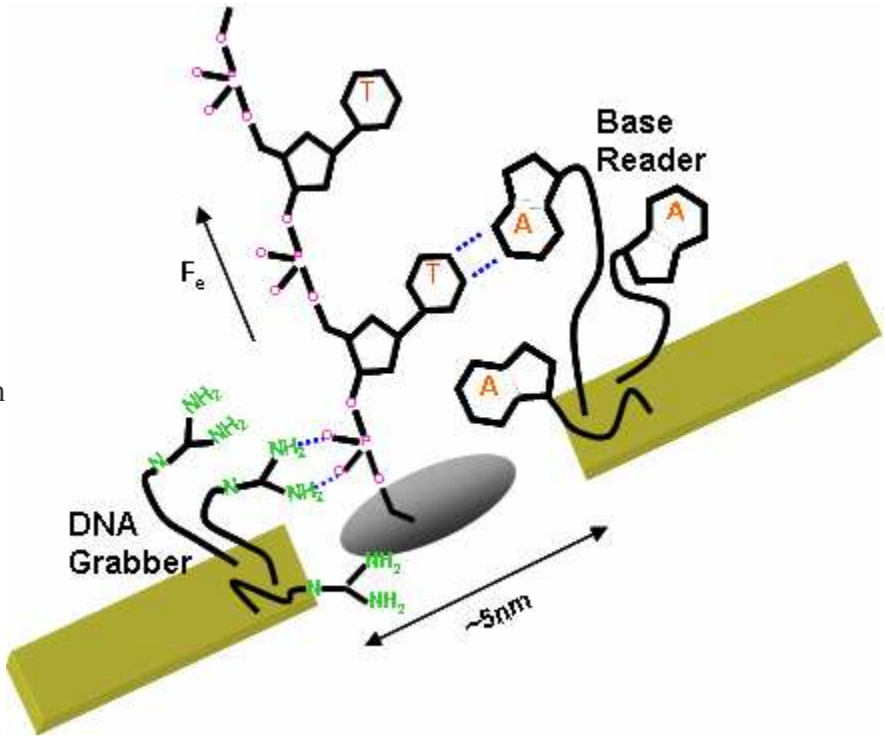


Figure 5.7. Sequencing by recognition set-up. Single stranded DNA translocates through a nanopore (gray) under an electrophoretic force (F_e) and, when it passes through the gap between nanoelectrodes (gold), guanidinium (on the left, green) and a base reader (on the right, orange, in this case A) bound to the electrodes form hydrogen bonds (the blue dashed lines) with the DNA. A current passes through the base reader-DNA-guanidinium signaling Watson-Crick base pairing is occurring, identifying the base on the strand as T.

circuit, and depending on current flow and knowledge of which base is on the electrode, its complimentary bases on the strand are determined.

5.3.1. DNA readers

For this method to be successful, single stranded DNA must be able to chemically interact with the nanoelectrodes. Unlike the STM experiments described above, the DNA bases being probed can not be covalently bound to or electrostatically stuck to one of the electrodes.

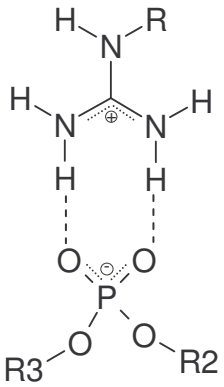


Figure 5.8. Parallel H-bonding and electrostatic interactions between guanidinium and phosphate.

The alternative is hydrogen bonding. The bases on the DNA can hydrogen bond to their complements bound to an electrode. In order to complete the circuit, the phosphate backbone must also hydrogen bond to something covalently linked to the other electrode. Guanidinium is a molecule that can form parallel hydrogen bonds with the oxygens on the DNA backbone (Figure 5.8).

Watson-Crick base pairing (see Figure 1.1) is not the only hydrogen bonding that can occur between bases which is why there is a measurable current between non-complementary bases. An alternative is called Hoogsteen base pairing (Figure 5.9). In order to decrease these non-Watson-Crick base pairing interactions modified bases will be synthesized that will form desired, stable hydrogen bonds, bind to electrodes, and not be oxygen, light, water, or electrochemically sensitive so as to be easily integrated into the sequencing system.

Experiments were carried out by Dr. Jin He and Lisha Lin in order to determine how well the guanidinium group can interact with the DNA backbone on the non-reading electrode (Fig. 5-7, *left*). Lisha Lin synthesized guanidinium with a thiol in order to bind the group to gold; a description of the synthesis and deposition on gold is in Chapter 6. Imaging of DNA monolayers (12 to 79 base ssDNA, 2.8Kbp dsDNA, 6.4Kb ssDNA) using STM (scanning the same spot before and after DNA deposition) was able to prove that the DNA will adhere to the guanidinium.[127] Shorter pieces of DNA and DNA minicircles could not form ordered monolayers, most likely because they are unable to form strong DNA-DNA interactions. Interestingly, when phosphate buffer was used, instead of TrisHCl, the phosphate outcompeted the DNA for the guanidinium and no monolayers were observed. Therefore, it was the phosphate backbone of the DNA that was sticking to the guanidinium. Work done using Surface Plasmon Resonance

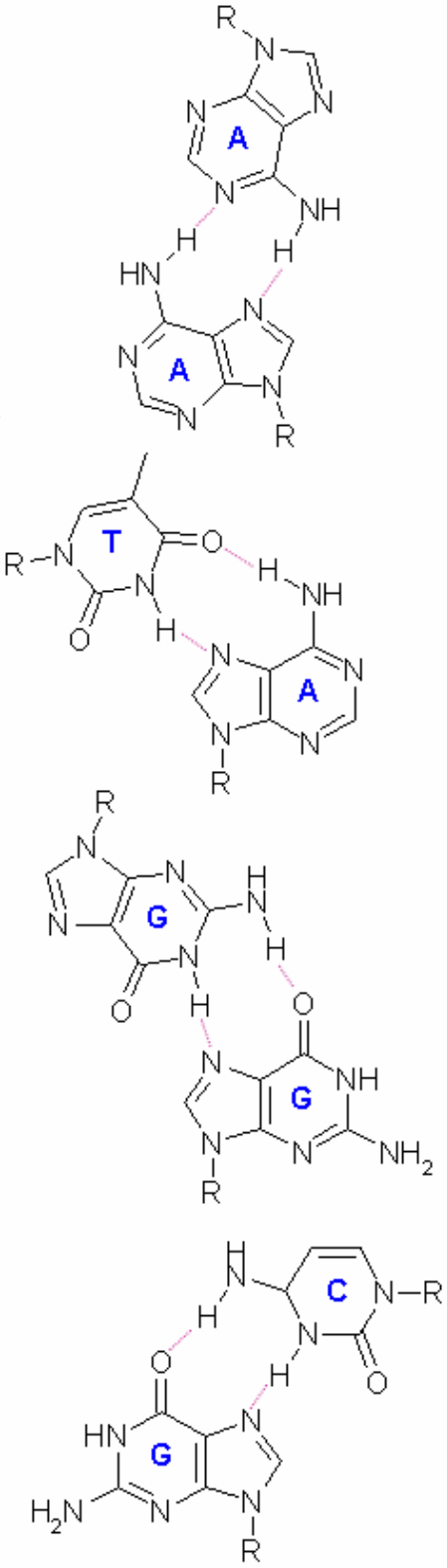


Figure 5.9. Hoogsteen base pairing, hydrogen bonds in pink.

(SPR) verified these results. Figure 5.10 shows that when ssDNA is deposited on a gold surface, already functionalized with guanidinium, a monolayer forms, shifting the resonance angle (the angle of incidence that corresponds to the minimum in reflected light intensity which indicates resonance between the plasmons and the incident light). The DNA so strongly adheres to the guanidinium that even when washed with 1M HCl, DNA would not unbind from the surface (Figure 5.10).[127] The irreversibility of the DNA-guanidinium interaction was troubling, Chapter 6 will describe the experiments that were performed in order to prove that – in contrast to the cooperative binding of a DNA monolayer – single molecules of DNA stick to guanidinium via hydrogen bonds at a strength that indicates it is easily reversible.

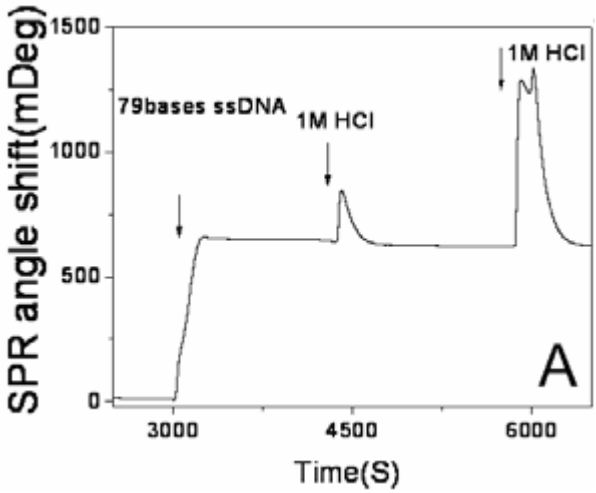


Figure 5.10. SPR of 79 base ssDNA adsorbed onto a guanidinium surface. Two injections of 1M HCl were unable to unbind the DNA.[127]

STM experiments using a guanidinium monolayer, ssDNA deposited on the monolayer, and STM tips functionalized with thiol modified bases were performed to determine if Watson-Crick base-pairing specific increases in tunneling current are still observed (Figure 5.6) once guanidinium was added to the gap. Figure 5.11 shows that when a base complementary to bases in the ssDNA on the guanidinium are on the STM tip (Figure 5.11 D and E) there is a response distinguishable from experiments when non-complementary bases (Figure 5.11 C and F) are on the tip or no DNA is on the surface

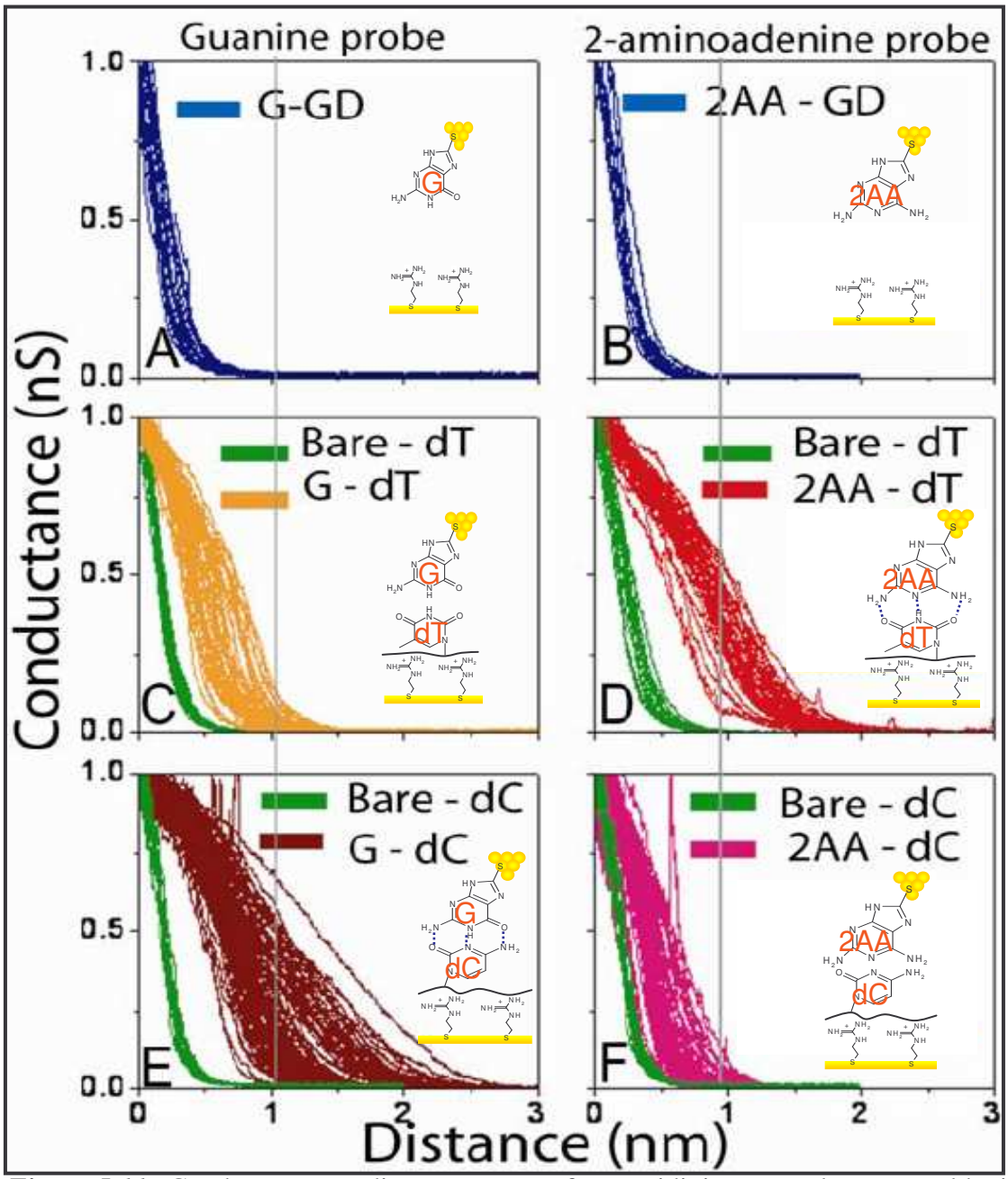


Figure 5.11. Conductance vs. distance curves of a guanidinium monolayer on gold with guanine or 2-aminoadenine on an STM tip. In panels A and B no DNA is on the surface, in panels C and F non-complementary base pairs between 45mer poly T DNA and guanine (C) and 45mer poly C DNA and 2-aminoadenine (F), and panels D and E show complementary base pairing between poly T DNA (D) and 2-aminoadenine and poly C DNA and guanine (E). There is clearly a difference in the curves based on complementarity.[127]

(Figure 5.11 A and B).[127] The addition of guanidinium does not eliminate the use of tunneling current to distinguish base complementarity.

5.3.2. Nanopores and nanoelectrodes

Once appropriate DNA grabbers and base readers are determined, the next step will be to integrate them into a nanopore/nanoelectrodes system. The goal, again, is to develop a system that produces results that distinguish between complementary and non-complementary base pairing just as the STM experiments have (Figure 5.6). This requires a gap between electrodes of 2nm to 5nm; the base readers will have flexible linkers so the requirement that the atomically small ends of the electrodes be in perfect alignment for molecular binding is relaxed. The presence of the linkers may result in a drop in conductance across the electrodes but, with the modified bases, the presence of a signal, not its strength, will be what is important. Building nanoelectrodes is a challenge in and of itself, but aligning them with a nanopore, a key to this technique, is an additional challenge. However, this necessary alignment may be the key to a method of controllable nanoelectrodes construction. The idea, as illustrated in Figure 5.12, is to create electrodes with a 20nm gap by drilling through deposited

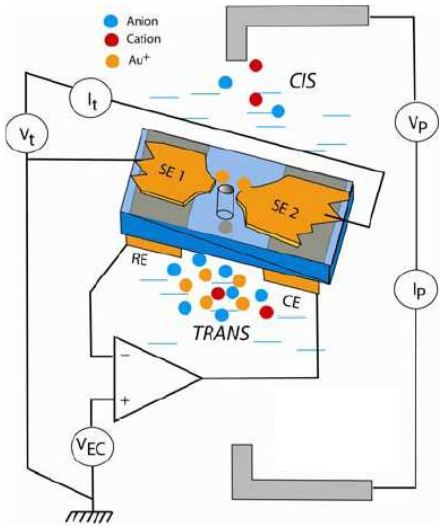


Figure 5.12. Schematic of nanoelectrodes (SE1 and SE2, with a potential difference Vt) narrowing over a nanopore via electroplating of gold ions that are directed to the ends of the electrodes by first traveling through the pore from the trans side. RE and CE are reference and counter electrodes. Through monitoring the pore current (Ip) and tunneling current (It) the final pore and tunneling gap size can be adjusted.[128]

gold, forming a nanopore by drilling through the Si_3N_4 beneath the gold from the other side, and then electroplating the ends of the electrodes, shrinking their distance, by feeding gold ions through the nanopore.[128] Before building nanopores with nanoelectrodes, tests must be preformed on nanoelectrodes to determine how well the base-DNA-guanidinium bridge will conduct current.

There are two ways the electrodes can be modified. A mixture of thiolated base with thiolated guanidinium can be deposited, resulting in both on either electrode. This process may result in an increase in either nucleoside monophosphates (for control experiments) or ssDNA only hydrogen bonding to a single electrode. Without completing the bridge between electrodes, tunneling current specific to base pairing could not be detected and recognition could not be achieved. Specific deposition to each electrode can be achieved, therefore ensuring the gap is closed, by maintaining a potential difference between the electrodes and relying on the cleavage of the gold thiol bond at -1V. One electrode can be modified, then rinsing the electrodes and depositing the second reagent on the clean electrode will result in specific attachment.

5.3.3. Translocation

As mentioned in the introductory chapter, the slowing of the translocation speed of DNA through a nanopore is the largest barrier to single base reading with more traditional nanopore sequencing. This technique has many possible ways to solve this problem. The hydrogen bonding may create a friction, slowing the DNA's progress. Additionally, magnetic tweezers can be used to counter the electrophoretic force. Magnetic tweezers are able to control multiple beads in parallel so their use would allow

an array of nanopores to be used to sequence multiple DNA strands. Figure 5.13 shows how the electrophoretic force (F_{elec}) will be opposed by the friction due to hydrogen bonding of the DNA at the nanoelectrodes, secondary structures of the DNA, and entropy fluctuations of the DNA. Additionally, magnetic tweezers could be used to further slow down the translocation of the DNA (F_{bead}), stretch the DNA to remove secondary structures (attaching a second bead and forming a rotaxane with the pore), or increasing the force on the DNA in the direction of F_{elec} if the hydrogen bond friction, entropy fluctuations, and secondary structures prove to be too strong for the DNA to pass through the pore without back slippage with just the application of an electric field. An electrophoretic force would still be needed to thread the DNA (with a magnetic bead attached) through the nanopore and then the application of a magnetic field could control its movement.

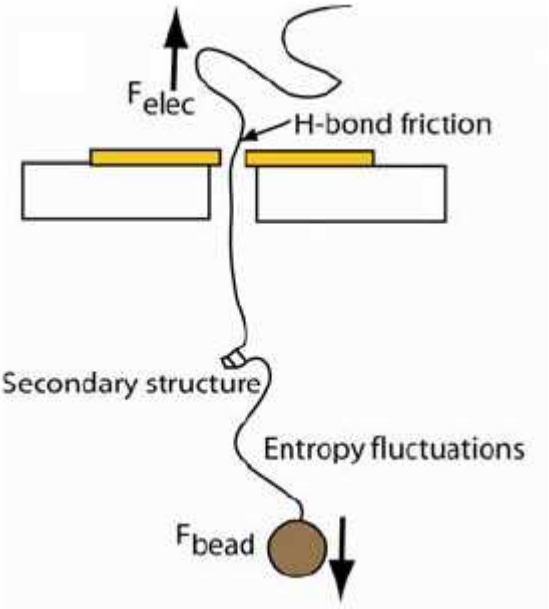


Figure 5.13. Diagram of the forces at play when DNA translocates a modified nanopore.[128]

5.3.4. Advantages

This method has similar advantages to other proposed DNA sequencing methods involving nanopores. It will allow for long contiguous reads (> tens of thousands of bases), minimal sample prep, and can read unmodified DNA. Its accuracy can meet the

standards of the National Human Genome Research Institutes (1 error in 10,000 bases) with just 13 duplicate reading heads, based on STM results that show 50% of reads give unambiguous positive reads ($0.5^{13} << 10^{-4}$).[128] It has the additional advantage of multiple methods to control translocation speed. It will also not rely on a change in tunneling current (like other nanopore methods that look at changes in ionic current) but will rely on a binary signal of binding/not binding thanks to the design of specific base readers.

Many of the advantages of this new technique were believed to be true for the previously mentioned AFM based sequencing method. Beyond its accuracy and binary nature there is one more very important difference. As opposed to the AFM based technique which had relied on a previously untested method of determining each nucleotide (the steric hindrance as the bases passed through the cyclodextrin), this method of recognizing each base has already been tested and proven to be highly accurate (Figure 5.11). In order to translate this advantage into a sequencing method there are many steps that still must be completed. One of these is to determine how individual DNA molecules will interact with guanidinium. As described earlier, solutions of ssDNA and dsDNA will form monolayers on a guanidinium surface that can not be washed away. In order to determine if this is a cooperative effect among strands and reversible with single molecules, AFM experiments were carried out to study single DNA molecule interactions with a guanidinium surface. The strength of that interaction is also important. The sequencing by recognition technique requires that the oxygens on the DNA backbone and guanidinium on the nano-electrode form hydrogen bonds and then

those bonds must be easily broken for the next nucleotide to move through the nanopore and be recognized. The next chapter will discuss the experiments performed that verified that guanidinium is acceptable for sequencing by recognition and better determined how DNA interacts with it.

6. Adhesion of Single Molecules of DNA

6.1. Introduction

Experiments discussed in Chapter 5 proved that multiple DNA strands do not interact in a reversible manner with guanidinium. In order for the sequencing by recognition method to work single strands of DNA must hydrogen bond with guanidinium and then those bonds must be broken without damaging the rest of the construct; the force to break apart the phosphate backbone and guanidinium must be less than that needed to break covalent bonds. The interaction of single strands of DNA with guanidinium was studied using the AFM. DNA was covalently bound to an AFM tip, brought into contact with the surface, and then the tip with DNA was retracted. In this manner the force required to break the hydrogen bonds between the DNA and guanidinium was measured and the strength of the DNA's interaction with the electrodes estimated.

During these experiments DNA is collapsing out of solution and adhering to the surface. When DNA collapses into a compact structure (e.g. chromosomes in eukaryotes or packed into a virus capsid) it has condensed. Condensation is distinguished from aggregation or precipitation in that the result of condensation is of finite size and an orderly morphology.[129] The amount by which DNA is able to condense is very impressive. The DNA of T4 phage undergoes a 540 fold compression when its 160,000 base pairs condense and fill a 100nm diameter capsid.[129]

DNA can also collapse out of solution and form well ordered, flat-lying monolayers on a surface that consist of DNA-DNA and DNA-surface interactions.[130] Ordinarily, when DNA spontaneously condenses onto a surface or packages itself into a capsid, positively charged ions or polyamines (like spermine, a polyamine that has a 4^+ charge

and spermidine with 3^+ at physiological pH) must be present in order to shield its negatively charged backbone from itself or to form bridges to the surface.[130, 131] Other factors that affect condensation are the type and concentration of bridging ion, the pH of the solution, the temperature, and the structure of the DNA (e.g. double stranded or circular).[130, 132] Different bridging ions will affect how compact a condensed DNA monolayer on a surface will be.[130] The pH will determine if surface groups are protonated (pH < pKa of molecules), affecting the electrostatic attraction and hydrogen bonding. On hydrophobic surfaces a low pH (pH ~ 3) causes double stranded DNA to melt (the hydrogen bonds between the strands break and the dsDNA unwinds separating into single stranded pieces) exposing the hydrophobic bases, increasing adhesion.[132] Experiments have also shown that above 40^{o}C DNA monolayers disappear.[130] It has been confirmed experimentally that cations are not evenly distributed along a strand of DNA, they gather in the middle and the first and last 20 phosphate groups will not be shielded to the same degree.[42] This results in the ends of the DNA being more negative than the midsection, as such it is the ends that adhere to a protonated surface more strongly.[132] As a result, strands of DNA are ten times more likely to stick to a surface than a plasmid DNA lacking ends.[132]

In the case of DNA adhering to a guanidinium monolayer (or any surface), the DNA is originally a flexible polymer in solution and then collapses down and is constrained to the two dimensional surface. The experiments that follow use single molecules of DNA attached to an AFM tip as opposed to most studies of condensation and monolayer formation which use much longer strands of DNA (kilo-base pairs) and higher

concentrations (e.g. 25ng/ml) which are free in solution.[129, 130] The tip is brought into contact with the guanidinium surface, allowed to sit for three seconds, and then withdrawn, pulling the DNA off the surface (Figure 6.1).

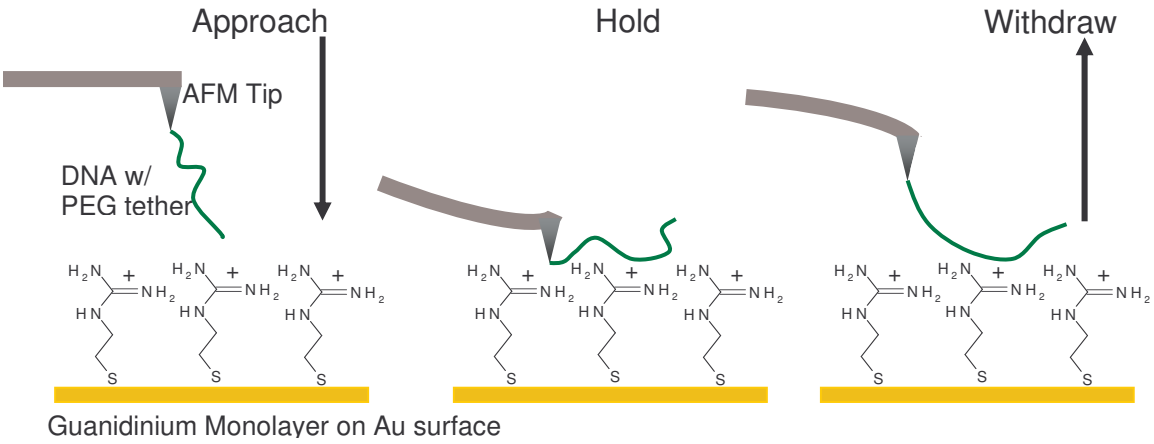


Figure 6.1. Schematic of the experimental set-up. DNA is covalently bound to an AFM tip, the tip approaches a guanidinium monolayer on a gold surface, the tip is held on the surface for 3 seconds, and then the tip is withdrawn, pulling the DNA off the surface.

As a result, condensation may not be the best word to accurately describe what is happening since the DNA can not form an orderly morphology or aggregates (with itself or other strands). However, some of the forces that either stabilize DNA condensation or oppose its collapse are similar to the ones that act on the DNA when it lays down on the guanidinium surface. Therefore, it is helpful to discuss some of the theory that has been developed around DNA condensation.

The primary forces involved in DNA condensation arise from bending, mixing entropy, and Coulombic attraction/repulsion.[129] The free energy for bending DNA takes into account the total length of the DNA (L), its persistence length (a), and the radius of curvature (R_c) of the bend,

$$\Delta G_{\text{bend}} = \frac{RTaL}{2R_c^2}, \tag{6.1}$$

R is the gas constant and T temperature.[129] The mixing entropy results from the mixing of DNA with the solvent; DNA and solvent have more microscopic states accessible to them when mixed than when segregated. The negative of that value, multiplied by temperature, gives the free energy of segregating the DNA when it condenses. The flexibility of the DNA is again a factor in determining the free energy,

$$\Delta G_{\text{cond}} = -\Delta G_{\text{mix}} = RT\frac{L}{a}.^{[129, 131]} \tag{6.2}$$

The Coulombic force consists of two parts. When DNA condenses, its negatively charged backbone is forced to come into closer contact with itself and that of other DNA. The presence of counterions neutralizes some of the charge but it is estimated that 10% remains.[129] The electrostatic repulsion from this remaining charge contributes a free energy of

$$\Delta G_{\text{elec}} = \frac{n_{\text{tot}} k_B T}{2\xi z_2^2} \ln \frac{V_{\text{uncond}}}{V_{\text{cond}}}, \tag{6.3}$$

n_{tot} is the total number of DNA phosphate charges before neutralization, z_2 is the valence of the condensing ion, ζ is the counter ion condensing parameter which relates the distance between charges to the Bjerrum length, the length where two charges' electrostatic interaction is similar in magnitude to k_BT, V_{uncond} is the volume of the uncondensed DNA (taken as a sphere with a radius equal to the radius of gyration), and V_{cond} is the volume of the condensed DNA, k_B is the Boltzmann constant, and T is the temperature.[129] Van der Waals forces due to ion fluctuations around the condensed

DNA creating induced dipoles which momentarily attract each other are responsible for stabilizing the condensed DNA

$$\Delta G_{\text{fluct}} = -\frac{3Lk_{\text{B}}T}{X_{\text{cond}}}\frac{(\theta_2 z_2^2 \xi)^2}{(1+\theta_2 z_2^2 \xi)^2}, \qquad (6.4)$$

where θ_2 is the fraction of phosphates being neutralized by ions, and X_{cond} is the center-to-center interhelix distance of the condensed DNA.[129] These four types of free energies combine to equal the total free energy for DNA condensation:

$$\Delta G_{tot} = \Delta G_{bend} + \Delta G_{mix} + \Delta G_{elec} + \Delta G_{fluct}.$$

For double stranded, kilo-base pair lengths of DNA, at room temperature, in solution with ions with a charge of +3 (like spermidine), ΔG_{tot} can range from -0.044 to -0.097$k_{\text{B}}T$/base pair, indicating condensation is a favorable reaction.[129]

6.2. Single Molecules of DNA Adhering to a Surface

As mentioned earlier, the experiments performed do not cause DNA condensation as described above. The DNA is not forced to bend and form a toroid shape in order to fit in a capsid; therefore the electrostatic repulsion between the close packed phosphate backbones is not an issue. The guanidinium monolayer is not free to surround the DNA like the ions in more traditional condensation reactions so the contribution from induced dipoles is also not as relevant. The DNA must still collapse onto the surface so the mixing entropy continues to be an important repulsive contribution to the free energy. There is a strong

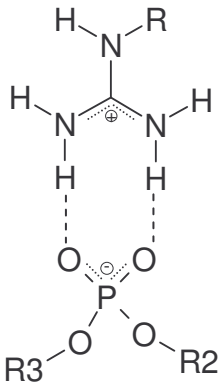


Figure 6.2. Parallel H-bonding and electrostatic interactions between guanidinium and phosphate.

electrostatic attraction between the negative DNA and positive surface and the guanidinium and phosphates will form parallel hydrogen bonds adding stability to the DNA collapse (Figure 6.2).[133] The contribution to the free energy of DNA electrostatically binding to the guanidinium is

$$\Delta G = -zRT \ln\left(\frac{2\pi\sigma^2}{\varepsilon k_B T N_1}\right) \tag{6.5}$$

where σ is the surface charge density of the DNA, ε is the dielectric constant of water, N_1 is the concentration of the monovalent ions in solution (e.g. Na^+), and z is the valence of the condensing ion (guanidinium).[134] The non-specific electrostatic enthalpy for DNA condensation is then

$$\Delta H_{\text{electro}} = -T^2 \frac{\partial(\Delta G/T)}{\partial T} = zRT(\nu - 1) \tag{6.6}$$

The value of ν was experimentally determined to be 1.4 and is related to ΔH through the dielectric constant of water, $\varepsilon = \varepsilon^*(T^*/T)^\nu$, where T^* is 298K and ε^* is the dielectric constant at T^*.[134] What is important is that ν is slightly larger than one which results in small, positive values for ΔH.[134] These factors, the unfavorable mixing entropy and enthalpy electrostatic interactions and the favorable enthalpy of hydrogen bonding, will determine the free energy for the DNA to lay down on the surface.

The strength of adhesion of the DNA to the guanidinium is measured by determining the force required to pull it off of the surface. According to Di Marzio and Guttman, the force to pull a polymer off a surface by breaking hydrogen bonds is:

$$f = \left(\frac{k_B T}{a}\right)\sqrt{6\{\frac{\Delta\varepsilon}{k_B T} + \ln[1 + e^{\frac{-\Delta\varepsilon_2}{k_B T}}] - \ln z\}}\,, \qquad (6.7)$$

where the length of a monomer of a polymer is the persistence length (a) for DNA, the energy of attachment for the polymer is $\Delta\varepsilon$, $\Delta\varepsilon_2$ is the energy of breaking a hydrogen bond ($\Delta\varepsilon$ and $\Delta\varepsilon_2$ may not be the same), and z is the number of configurations of the length of DNA not stuck to the surface (Figure 6.3).[135] What is important from this is that (somewhat surprisingly) the force does not depend on the distance (R) the end being pulled is from the surface.[135]

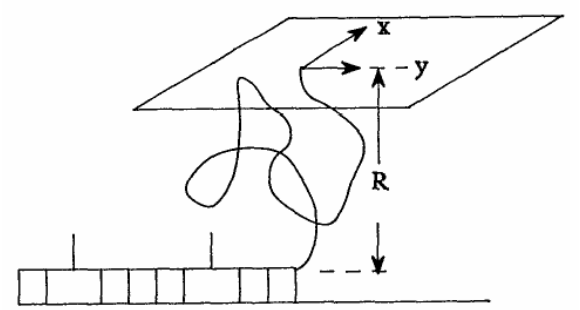


Figure 6.3. Unzippering model when the polymer is held to the surface with hydrogen bonds that are allowed to break.

As a result, the withdraw curves from AFM experiments where DNA is being peeled off of a guanidinium surface should have plateau regions (see Figure 6.10) where the force is not changing as the tip is retracting from the surface.

Understanding the interaction between the phosphate backbone of single molecules of DNA and the protonated guanidinium monolayer is important to the sequencing by recognition technique. The protonated guanidinium functions as the bridging ion for DNA adhesion. A description of the experiments used to examine these interactions and a discussion of results follow.

6.3. Experimental Set-up

AFM deflection curves were collected using a Molecular Imaging/Agilent Picoplus AFM. The ozone cleaner is an UV Clean #135500 (Boekel Inc.). The thickness

measurement was performed on LSE Stokes Ellipsometer (Gaertner Scientific Corp.). NMR spectra were recorded on the Varian Inova 500 MHz instrument in the Magnetic Resonance Research Center at ASU. Matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra were recorded on a VG TofSpec spectrometer, Proteomics and Protein Chemistry Lab at ASU. 3-aminopropyldimethylethoxysilane (APDM) was purchased from Gelest, Inc. (Morrisville, PA). 3,3'-dithiobis(succinimidyl propionate) was from Pierce Biotechnology, Inc. (Rockford, IL). Maleimide-PEG-NHS (MW=5000Da) (PEG is poly(ethylene glycol)) was from Nektar Therapeutics (San Carlos, CA). BamHI, EcoRI, CIP, and T4 DNA Ligase were all purchased from New England BioLabs (Ipswich, MA). Polybead Carboxylate 10μm Microspheres was purchased from Polysciences Inc. (Warrinton PA). All other chemicals and anhydrous solvents are from Aldrich (Milwaukee, WI). DNA was purchased from Integrated DNA Technologies (Coralville, IA). Three different strands were used: single stranded poly T 15mer with dodecaneamine on the 5' end (poly T DNA), hairpin DNA used as an example of double stranded DNA (TTT CAT TGG TAA CCT GAG GTT ACC AAT G) with hexaneamine on the 5' end (hpDNA), and a 60mer (AAC CAG AGA CCC TCA GAG CTA CGG ACA GTC CGT GTT AGG TTA GTT AAC CGC CAG GGG TCA) with hexaneamine on the 5' end (60 DNA), which does have some secondary structure but with long runs of single stranded regions it is used as an example of single stranded DNA (Figure 6.4).

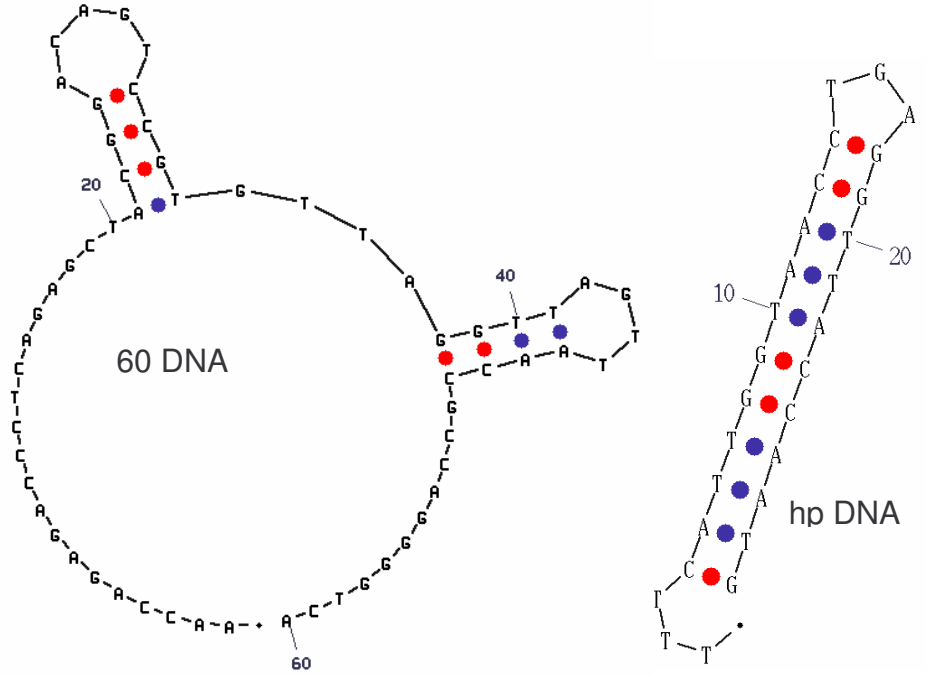


Figure 6.4. The structures that 60 DNA and hp DNA form in 100mM salt concentration, based on mFold calculations.[95]

6.3.1. 4.6Kbp DNA Preparation

TOPO 10N plasmid was provided by Dr. Jim Wilson, assistant research professor in the Center for Infectious Diseases and Vaccinology at the Biodesign Institute, Arizona State University. It was digested with BamHI and EcoRI for 2hrs at 37°C. The enzymes were removed using a phenol/chloroform extraction and then Alkaline Phosphatase, Calf Intestinal (CIP) was added (two additions for 0.5hrs at 37°C each) to remove phosphates from the ends of the two strands of DNA (4.6Kbp and 3.9Kbp). Another phenol/chloroform extraction was used to remove CIP, followed by an ethanol precipitation of the DNA. Ligation between the 4.6Kbp strand and a 24 base "handle" with a 5' phosphate group, complimentary sequence to BamHI overhang and a 3' aminopropane group (IDT DNA) was preformed with T4 DNA Ligase at room

temperature overnight. Gel electrophoresis was used to separate the 4.6Kbp strand with the amine "handle," the 3.9Kbp unmodified strand, and any unused handle. The final concentration of 4.6Kbp amine modified DNA (4.6Kbp DNA), after gel extraction, was 45µg/ml.

6.3.2. Guanidinium Surface Preparation

β-mercaptoethylguanidine was synthesized, characterized, and provided by Lisha Lin (see Figure 6.1). Cystamine dihydrochloride (0.25g, 1.1mmol) was dissolved in anhydrous DMF (1.5ml). N,N-bis(tert-butoxycarbonyl)thiourea (729mg, 2.2mmol) was added to the above solution followed by addition of anhydrous triethylamine. (0.373ml, 2.64mmol). A suspension of Mukaiyanma's reagent (675mg, 3 mmol) in anhydrous DMF (1.5ml) was added dropwise to the reaction mixture which was stirred at room temperature overnight. The crude product was diluted with water (2.2ml) and extracted with diethyl ether (3x2.2ml). The organic layer was dried with sodium sulfate and evaporated. The crude product was purified by column chromatography (ethyl acetate: hexane=1:4). The yield was about 14%. The tert-butoxycarbonyl group was deprotected by 50% TFA in DCM for two hours. The excess TFA was evaporated under vacuum. The product was then dissolved in 2ml water and washed with DCM twice to remove organic impurities. Pure bis(2-guanidinoethyl) disulfide was obtained as a trifluoroacetato salt. 1HNMR (D2O) δ 3.338(t, 2H), 2.727 (t, 2H); MALDI MS m/z: 237.098 (calcd for 237.095) β-mercaptoethylguanidine was obtained by treatment of the bis(2-guanidinoethyl) with immobilized TCEP (Tris[2-carboxyethyl] phosphine hydrochloride) disulfide reducing gel (Pierce) immediately prior to use.[127]

6.3.3. Monolayer preparation and characterization

A freshly hydrogen-flame annealed gold substrate (Agilent, Chandler, AZ) was immersed in a 0.5 mM aqueous solution of β-mercaptoethylguanidine for 15-30 minutes and then rinsed with 1M NaCl solution and doubly distilled deionized water. The thickness of the guanidinium SAM was determined by averaging ellipsometry readings at eight different locations on the sample, yielding a thickness of 0.99±0.03 nm, close to the expected height for an upright molecule of 1nm (estimated by Chemdraw 3D).[127]

6.3.4. AFM measurements of single DNA molecule adhesion to guanidinium

Ultrasharp CSC11/AlBS cantilevers (k = 0.35N/m) were used. In order to functionalize the probes they were first placed in the ozone cleaner for 10 minutes, immediately dipped into fresh piranha solution for no more than 30 seconds (to prevent damage to the metallization), rinsed with water and put into a solution of APDM (200 µl, 1.4 µmol) in 95% ethanol (1 ml). After 5 minutes, the tips were rinsed with water, placed in a clean dessicator flushed with argon, and placed under vacuum (10 torr) for 1 hour. The tips were removed from the vacuum and placed in solution of 3,3'-dithiobis(succinimidyl propionate) (0.025g, 0.062mmol) in DMF (1ml) for half an hour. The tips were then rinsed in and then allowed to sit in 10mM tris(2-carboxyethyl)phosphine hydrochloride in 5mM PBS (pH 7) for 15 minutes. While waiting for the tips, maleimide-PEG-NHS (0.1mg, 22nmol) was added to DNA (10.5nmol) in nanopure water (25µl). Before adding the tips to the DNA, the tips were rinsed in 100mM phosphate buffer (pH 6.5) with 5mM EDTA and 1M NaCl. The phosphate buffer with EDTA (50µl) was also added to the DNA-PEG solution. The tips

were allowed to react overnight in the DNA-PEG solution. The probes were then rinsed in water and used immediately (Figure 6.5).

For experiments with the 4.6Kbp long DNA, 10µm, carboxylate functionalized polystyrene spheres were first glued to AFM tips using Varian TorrSeal. Once the epoxy set (24hrs), spring constants of the now modified AFM tips were determined using the thermal calibration method (custom software for the spring constant provided by Ashley Kibel).[104] Next, the tips were immersed in an aqueous solution (1ml) of *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide (0.010g, 0.065mmol) and *N*-hydroxysuccinimide (0.008g, 0.070mmol) for 30min. They were then added to the DNA and allowed to sit for 24hrs.

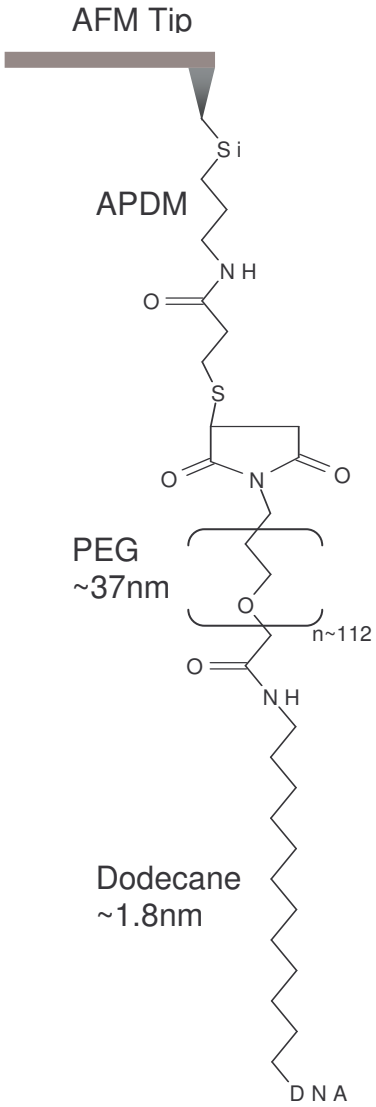


Figure 6.5. Chemical structure of tip chemistry, with some lengths.

Dr. Brian Ashcroft interfaced the AFM to a LabView control system via a custom modification of the head electronics. The AFM approach was controlled with custom software (written by Dr. Ashcroft) using the Measurement Studio (National Instruments) in Visual Basic (Microsoft).[85] The guanidinium surface was placed into the liquid cell of the microscope and covered in appropriate solvent, either nanopore water (slightly acidic

pH), 100mM phosphate (pH 7.5), or 100mM tris (hydroxymethyl) aminomethane hydrochloride (TrisHCl) (pH 7.5) buffer. A freshly-functionalized probe was inserted into the scanner and lowered onto a guanidinium functionalized gold surface using Measurement Studio until a deflection increase of ~100pN (a larger trigger force was used for the 4.6Kbp DNA, 700pN) was detected, held for three seconds to allow the DNA to make contacts with the guanidinium then retracted while force-distance curves were recorded. A withdraw speed of 100nm/s was used because it allowed for the most stable sample collection and is the most similar to the estimated expected speed that DNA will pass through the functionalized nanogaps.[33, 85, 109, 128]

6.4. Results

The full length for each of the constructs (DNA plus PEG linker) is 63.4 ± 2.8nm for 60 DNA (excluding the bases that form the hairpins), 48.8±2.8nm for poly T DNA, and 45.3±2.8nm for hp DNA, and 1,564nm for 4.6Kbp DNA (there was no PEG used on the 4.6Kbp functionalized AFM tips) based on a base to base distance of 0.34nm for double stranded and 0.6nm for single stranded DNA. The major contribution to the distribution in length is due to the polydispersity of the PEG.[97] It is the interaction of the DNA with the surface that is of interest -- the events that occur within the first 40nm from the surface (this corresponds to the PEG molecule and other linkers) are therefore ignored when analyzing the results for DNA-surface interactions. All adhesion forces, even those occurring non-specifically close to the surface, are included when the entire construct is specified. Roughly 800 approach and withdraw curves are collected for each tip prepared; histograms show the data collected over 800 pulls for each condition listed. As

compared to the literature cited earlier which used DNA on the order of kilo base pairs long to study condensation and adhesion,[129-132] very short strands of DNA (tens of base pairs, 5 to 25nm long) were used in most of these experiments, except for experiments with 4.6Kbp DNA. The pH of the solvents used (doubly distilled water with a slightly acidic pH, phosphate pH 7.5, TrisHCl pH 7.5) all result in the protonation of the guanidinium surface (pKa of guanidinium, 13.6[136]). The charged nature of the surface makes it very susceptible to contamination. This can result in withdraw curves with events out past the distance expected for the length of the construct being used. Those events are not selected out when the entire construct is specified but are when just the DNA is being presented in the results that follow.

Control experiments were carried out in TrisHCl buffer at a withdraw speed of 100nm/s on a clean gold surface with 60 DNA on the AFM tip, a guanidinium surface with a bare tip, and a cystamine surface with 60DNA on the AFM tip (Figure 6.6). Cystamine was used in order to compare a guanidinium surface that can form hydrogen bonds to an aminated surface which will only interact electrostatically with DNA. The cystamine surface was made by immersing a freshly annealed gold surface in 1mM aqueous solution of cystamine for 30 minutes. DNA interacting with a gold or cystamine surface resulted in force vs. distance curves that occurred within the length range of the DNA and had a saw tooth shape. The forces (measured from the minimum of peaks in the withdraw curves to zero deflection) of interaction with the gold surface, 126±58pN, fell below those expected for hydrogen bonding. The average force for the DNA interacting with cystamine was 128±40pN and for the DNA interacting with a

guanidinium surface it was 168.5±48pN. The difference in force between these two monolayers is expected because guanidinium/phosphate interactions are stronger than those between the cystamine and phosphate.[80] The bare tips resulted in the majority (93%) of force vs. distance curves showing adhesion of the tip to the surface within the first 15nm from the surface which was not shaped like either plateaus or saw tooth. The highly charged nature of the guanidinium monolayer causes it to be easily contaminated; that can explain the occasional pull out past the expected length due to the interaction of the tip with the contamination on the surface.

When the DNA comes into contact with the guanidinium surface hydrogen bonds should form. As the AFM tip is pulled away those bonds are broken and the rupture forces measured should be in the range expected for hydrogen bonding. Figure 6.7a is a histogram of measured forces for poly T, 60, and hp DNA adhering to guanidinium in TrisHCl buffer. The pink line shows the excepted rupture force of single hydrogen bonds, 181±35pN (AFM experiments were performed where silicon nitride tips with silanols were brought into contact with a mica surface in water, hydrogen bonding occurs and then the tip withdrawn, a plot of force variance vs. the average force of each experiment gave a slope of 181pN).[137] The distribution shows that the DNA is hydrogen bonding with the guanidinium on the surface and the bonds are individually broken as the DNA is pulled away from the surface. The average forces are 145.3±55 for poly T, 168.5±48 for 60, and 277.2±52 for hp DNA. Later in this chapter the much stronger force for hp DNA adhering to guanidinium in TrisHCl will be explained. Experimentally, electrostatic interactions have been found to rupture at 70±15pN (AFM

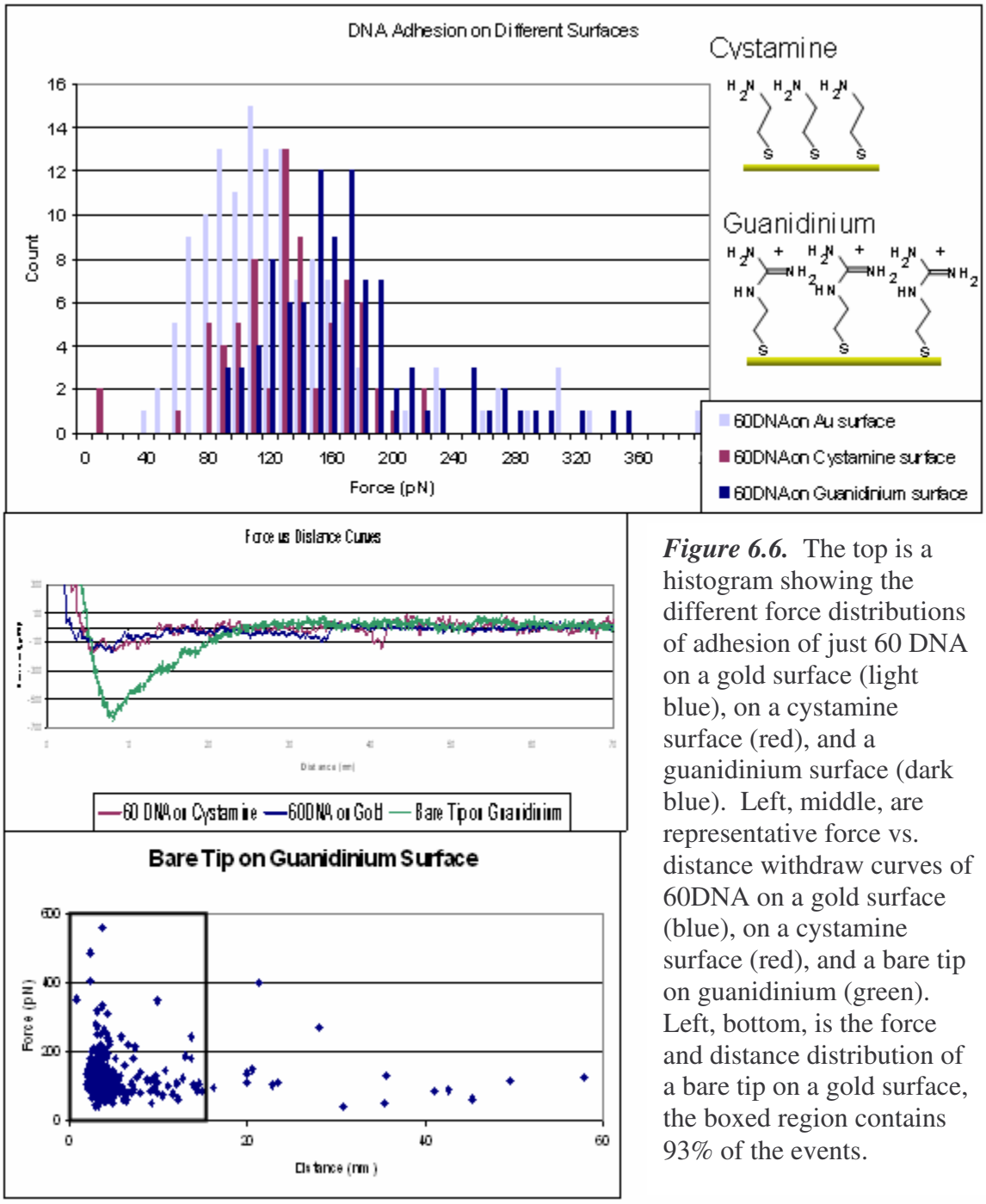


Figure 6.6. The top is a histogram showing the different force distributions of adhesion of just 60 DNA on a gold surface (light blue), on a cystamine surface (red), and a guanidinium surface (dark blue). Left, middle, are representative force vs. distance withdraw curves of 60DNA on a gold surface (blue), on a cystamine surface (red), and a bare tip on guanidinium (green). Left, bottom, is the force and distance distribution of a bare tip on a gold surface, the boxed region contains 93% of the events.

experiments were performed where positively charged *N*,*N*,*N*′,*N*′-tetramethyl-*p*-phenylenediamine and negatively charged 7,7,8,8-tetracyanoquinodimethane were brought together and then pulled apart and histogram of the pull-off forces showed peaks

with a 70pN periodicitiy),[138] therefore the DNA backbone is forming a strong connection with the guanidinium than can not be explained by only considering their different charges. Figure 6.7b is an overlay of force vs. distance curves of 60 DNA

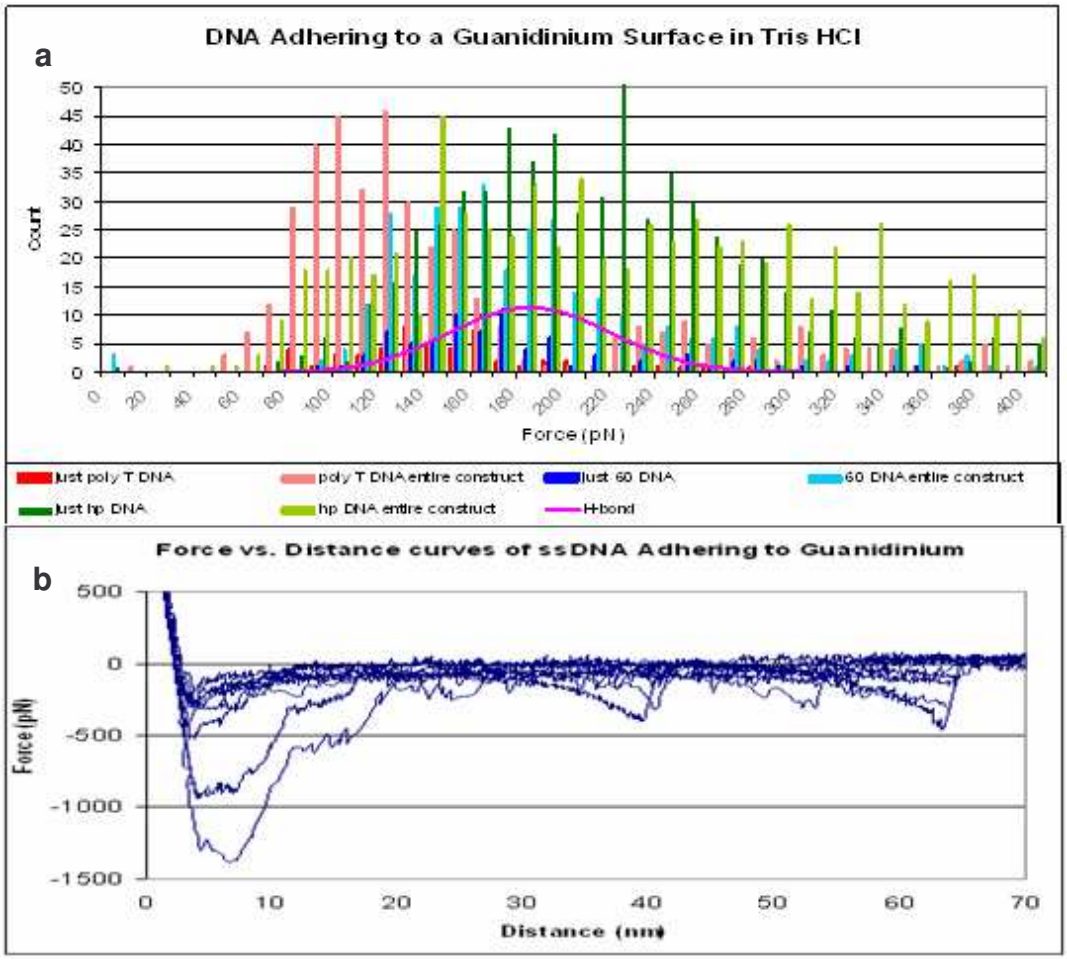


Figure 6.7. Top is the force distribution of poly T, 60, and hp DNA adhering to guanidinium, bottom overlay of 8 force vs. distance curves of 60 DNA in TrisHCl.

interacting with the guanidinium surface.

The type of DNA (both single stranded or double stranded and the length) and the type of solution the experiments were carried out in affect the adhesion of the DNA.

Work done studying the hydrogen bonding between guanidinium and phosphate groups has shown that the alignment of the two groups is more important to the formation of

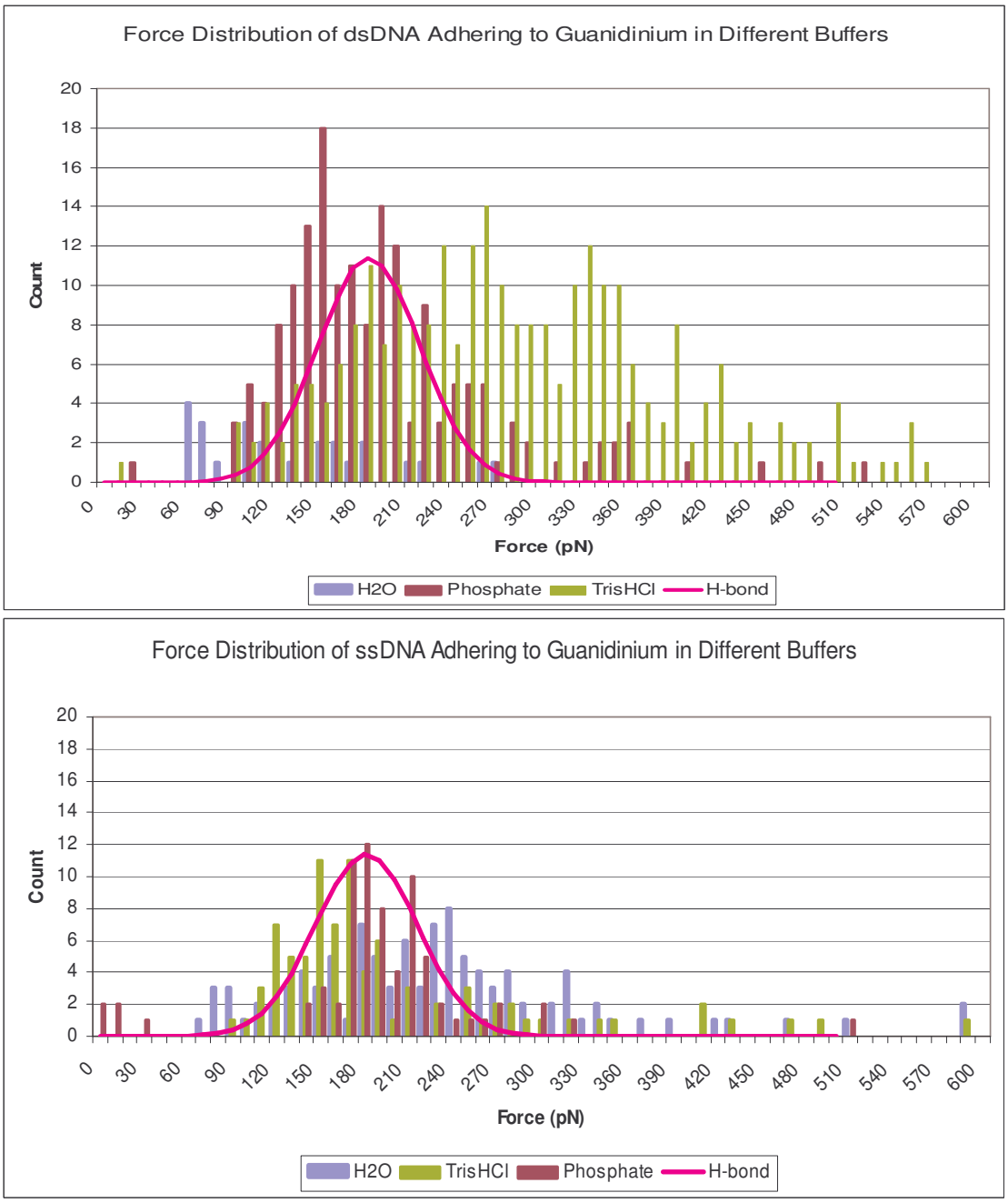


Figure 6.8. Comparison of buffer effects on dsDNA (hp DNA) and ssDNA (60 DNA), just the interaction with the DNA is being shown. The pink line shows the distribution of forces for hydrogen bonds, 181±35pN.[137]

hydrogen bonds than with ammonium groups.[80] This would indicate that ssDNA could form more hydrogen bond connections with the surface than the double stranded because of its ability to obey the strict restraints in bonding due to its greater flexibility. However, there is a greater entropy cost for ssDNA to lay flat on the surface because of its shorter persistence length and greater flexibility, than for dsDNA.

It was expected that using phosphate buffer, with its negatively charged phosphate groups, would result in fewer pulls because of the competition between the DNA and the buffer as opposed to TrisHCl or water. This is true for the case for dsDNA (hp DNA); there are far more events and stronger rupture forces measured in TrisHCl than in phosphate (Figure 6.8). These buffer effects are less pronounced in the case of ssDNA (60 DNA) where the forces for all three types fall around that for single hydrogen bond rupture and the number of events are about the same (Table 6.1). The length of dsDNA used falls below the persistence length of the DNA so it is like a single semi-rigid rod. In the presence of phosphate buffer it lays on the surface but is only able to form single hydrogen bonds because of the competing buffer. When TrisHCl is used multiple bonds are able to form, increasing the number of events and the strength of the measured rupture forces. In the case of ssDNA, its flexibility and resulting additional entropy cost of laying down results in only single hydrogen bonds forming in any solution.

Table 6.1. Average forces for hp and 60 DNA adhering to guanidinium when the experiments are carried out in different solutions.

DNA	Solution	Avg. Force (pN)
hp DNA	Phosphate	180±39
	TrisHCl	277±52
	Water	140±35
60 DNA	Phosphate	192±35
	TrisHCl	168±48
	Water	199±75

As mentioned in the introduction, the free energy of the DNA adhering to the surface will equal the negative of the mixing entropy and the electrostatic and hydrogen bonding free energies, $\Delta G_{tot} = \Delta H_{H-bond} + \Delta H_{electr} + RT(\frac{L}{a})$. The persistence length of dsDNA is more than 50 times greater than for ssDNA (50nm vs. 0.6nm). By substituting those values into Equation 6.2 one can see a large difference in their condensing free energies; 0.08kcal/mol (3.5meV) for hp DNA and 12.8kcal/mol (0.56eV) for 60 DNA. This is expected since it should be more entropically unfavorable to pull a flexible polymer out of solution than a semi-rigid rod. The enthalpy from the non-specific electrostatic interaction between the DNA and guanidinium is 0.24kcal/mol (10meV) (Equation 6.5). This is a small and positive value due to the fact that the exponent *v* in the temperature dependence of the dielectric constant of water is 1.4, as explained earlier. Based on experiments to determine the enthalpy of hydrogen bonding between guanidinium and phosphate group $\Delta H_{\mathrm{H-bond}} = -3.42\,\mathrm{kcal/mol}$ (-0.15eV).[80] Using these values, the free energy for hp DNA to adhere to the guanidinium monolayer is -3.1kcal/mol (-0.13eV) and for 60 DNA it is 9.62kcal/mol (0.42eV). This assumes that only a single hydrogen bond is formed between the DNA and the surface. This appears to be a reasonable assumption because less than 0.3% of withdraw curves show multiple pulls in the DNA region and none show more than two. Multiple hydrogen bonds may be breaking at one time, which would increase the measured force. This appears to only be true in the case of hp DNA in TrisHCl (Figure 6.8). From that data it appears that a maximum of two hydrogen bonds will form between the hp DNA. If we double the enthalpy from

hydrogen bonding ($\Delta H_{H-bond} = -6.84$ kcal/mol (-0.30eV)) then the free energy for hp DNA to adhere to guanidinium in the presence of TrisHCl is now -6.52kcal/mol (-0.29eV) (Table 6.1). What is interesting about the calculated free energies is that it is energetically favorable for hp DNA to hydrogen bond to the surface and not for 60 DNA due to the difference in the entropy of condensing.

Table 6.2. Free energy of condensation and total free energy of adhesion of different DNAs to a guanidinium monolayer.

DNA	ΔG cond (kcal/mol)	ΔG adhering (kcal/mol)
hp DNA	0.08	-3.10
poly T	5.12	1.94
60 DNA	12.80	9.62
4.6Kbp DNA	18.2	15.02
hpDNA (2 H-bonds)	0.08	-6.52

The shape of the withdraw curves also gives information on the type of interaction between the DNA and guanidinium. As stated in the introduction, when a polymer is peeled off of a surface it has formed hydrogen bonds with, the force is not dependent on the distance from the surface of the AFM tip (see Equation 6.6). As a result, if multiple hydrogen bonds are formed between the DNA and guanidinium, then as they are broken and the DNA is pulled up, the withdraw curves should have plateau regions (see

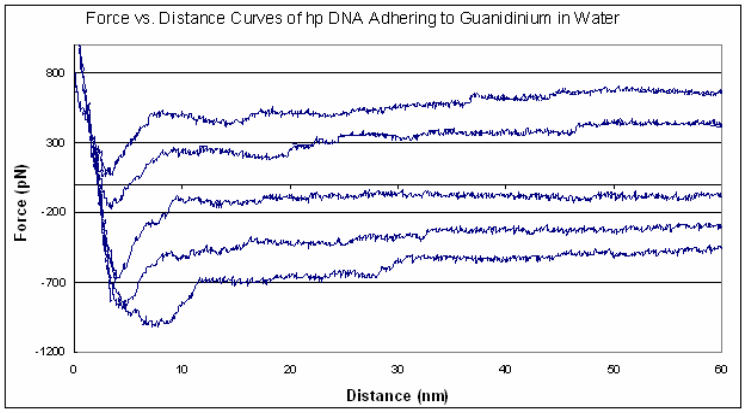


Figure 6.9. Force vs. distance curves of hp DNA adhering to guanidinium showing plateau regions. The curves are off-set from each other in order to better see their shape.

Figure 6.9). If the DNA is only able to form single contacts then one would expect the curves to have the more familiar saw-tooth shape seen in force spectroscopy as that single bond is pulled apart (see Figure 6.7b). While plateaus are observed, they make up a very small percentage of curves for both 60 DNA (~6%) and hp DNA (~14%) and only occur when experiments were run in water. Not only is the secondary structure of DNA much less stable without the presence of ions, any charge neutralization of the DNA will be eliminated. While only multi-valent ions can cause condensation, it has been found that the concentration of monovalent salts (Na^+) does make the enthalpy of binding more favorable, but still endothermic.[134] Therefore, the lack of monovalent ions leads to a slightly more favorable free energy of adhesion. Since the free energy for hp DNA is already more favorable than for 60 DNA it is not surprising that plateau shaped curves occur twice as often. However, the lack of ions in solution will also destabilize secondary structures; hp DNA will denature (opening the hairpin). This may explain why there are not more plateau shaped curves as it becomes more like single stranded DNA and adhesion becomes more entropically unfavorable.

As mentioned earlier, all the previous experiments were carried out using short DNA. In order to determine if the results would differ for very long DNA, 4.6Kbp (1564nm) dsDNA was attached to AFM tips. Control experiments were carried out with a beaded AFM tip and guanidinium surface in water. The results show there is non-specific adhesion near the surface but almost no adhesion pulls past 20nm (Figure 6.10). Despite the increased length and after eliminating non-specific adhesion pulls close to the surface

(the first 20nm are excluded) from the addition of the functionalized polystyrene beads, which were added to give more surface area to the tips in order to increase the chances of the 4.6Kbp DNA to "find" and bind to the tip, most of the adhesion events measured occurred with in the first 50nm of the DNA (Figure 6.10). The dominance of the condensing free energy which is directly proportional to the total length of the DNA and the persistence length of dsDNA being 50nm, it is not surprising that just the DNA brought closest to the surface by the AFM and within a persistence length, sticks to it, while the rest stays in solution. As a result, one would expect that this data would be similar to that for hp DNA. When experiments were carried out in TrisHCl there were more than twice as many events than in phosphate and 1.5 times more in water (Figure 6.10). Comparing forces is a bit more difficult because the longer DNA seems to have been able to make more hydrogen bonds with the guanidinium surface, even in the presence of phosphate (Figure 6.11). There is the expected increase in the distribution of

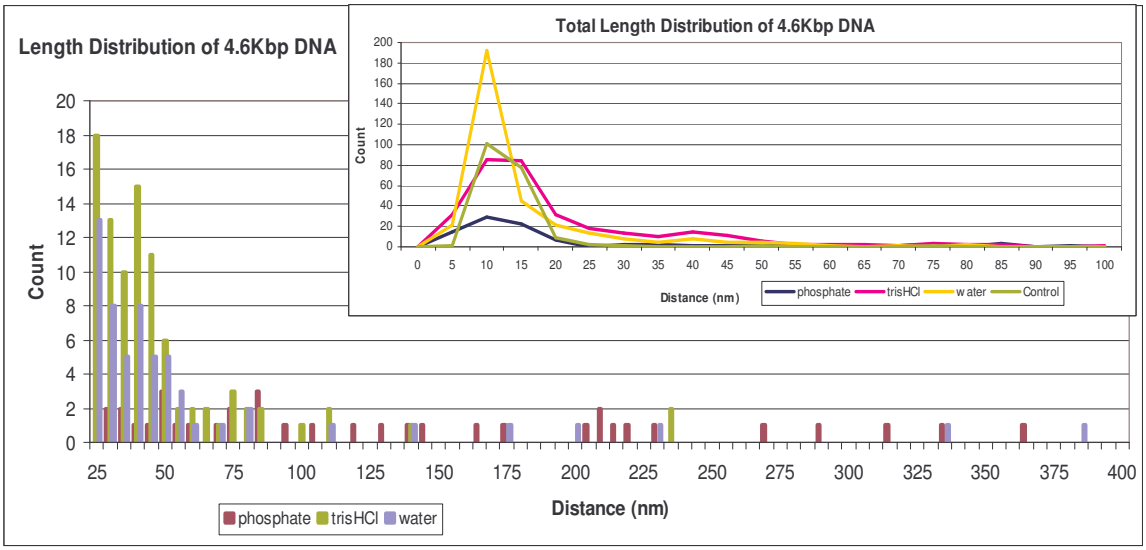


Figure 6.10. Length distribution of 4.6Kbp DNA adhering to guanidinium in different solutions. The insert shows the all the data, the larger plot show the distribution excluding events that occurred within 20nm of the surface.

force when the non-competitive TrisHCl was used but in the presence of phosphate forces did not cluster around the rupture of a hydrogen bond. Despite this increase in forces implying that more hydrogen bonds were formed, none of the withdraw curves had the expected plateau shape (Figure 6.12). It appears that rather than peeling off, the DNA would stretch and then multiple hydrogen bonds would break at a time. This is reasonable since the DNA does have a 50nm persistence length allowing for 150 phosphate groups (only spaced 0.34nm apart) to be brought into close contact with the surface. Only when water was used did the forces cluster around 181pN. The lack of ions result in denaturing the dsDNA,

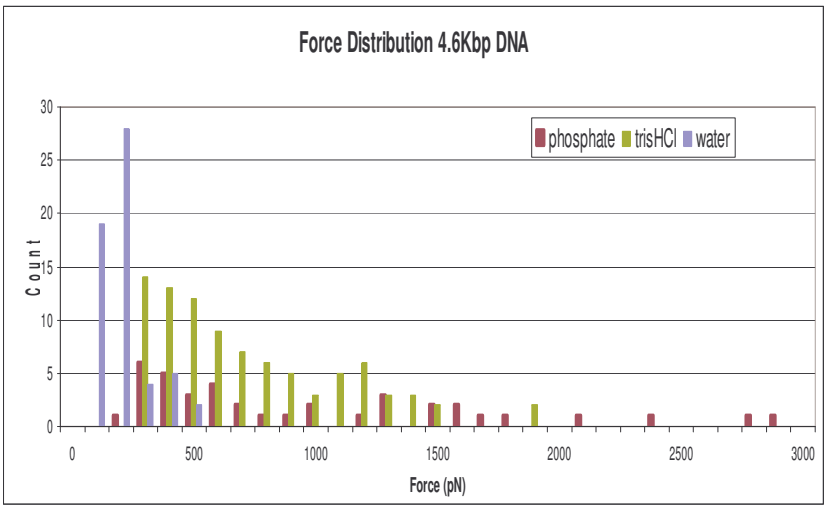


Figure 6.11. Force distribution of just 4.6Kbp DNA adhering to guanidinium in different solutions.

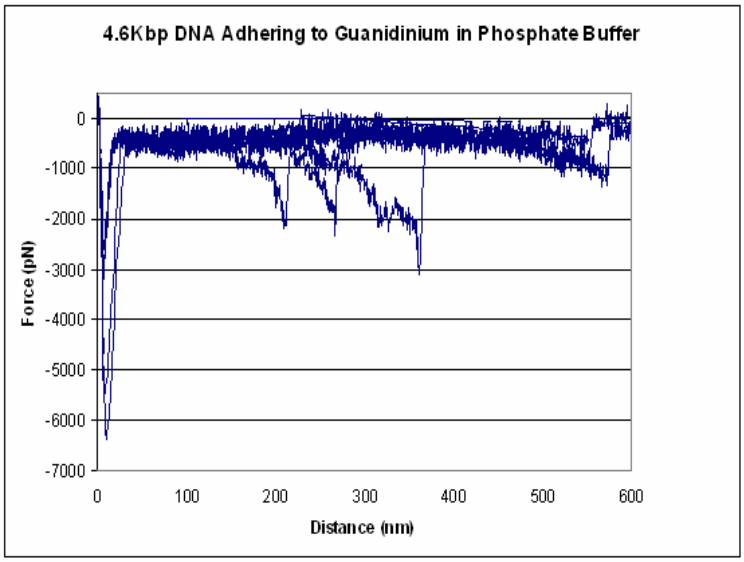


Figure 6.12. Overlaid force vs. distance curves of 4.6Kbp DNA adhering to guanidinium.

decreasing the persistence length to approximately 1nm, therefore it is not surprising that under that condition only single hydrogen bonds were able to form and break at a time.

6.5. Conclusion

The loss in entropy associated with segregating DNA from solution dominates its interactions with a guanidinium monolayer. Only for short dsDNA is it energetically favorable to condense out of solution and hydrogen bond to a surface, even in the presence of a competing buffer. The longer dsDNA only shows adhesion to the surface within the first persistence length from the AFM tip; it behaves similarly to the short dsDNA on that length scale but, due to the mixing entropy, the rest of the long DNA rarely condenses out of solution and hydrogen bonds to the guanidinium surface. Short ssDNA and dsDNA in the presence of a competitive buffer adhere to a guanidinium surface with the strength of a single hydrogen bond. The few instances of plateau shaped withdraw curves show that in the time allowed (3sec) the majority of DNA tested was not able to lay down on the surface and form enough hydrogen bonds to behave in the manner expected for peeling a polymer adhering via hydrogen bonds off the surface. The presence (or absence) of monovalent ions in solution does cause some differences in the way the DNA adheres to the surface.

This work was performed in order to determine if guanidinium is an acceptable DNA "grabber" for the sequencing by recognition method. These results are greatly affected by the DNA's ability to condense out of solution and form contacts with the guanidinium monolayer. The condensing free energy will no longer be an issue when sequencing; in the sequencing set-up, guanidinium is attached, via a flexible linker, to one of the

electrodes and DNA is feed through a nanopore to the electrodes one base/phosphate at a time. Based on the distributions of force (Figure 6.8), the extremely small percentage of withdraw pulls that showed multiple events with a single strand of DNA (<0.3%), and mixing entropy that oppose the DNA from condensing down on the surface, it is reasonable to consider the results from the experiments using 60 DNA, even though it is much shorter than the length of DNA hoped to be used for sequencing (tens of thousands of bases) to give the most accurate forces required to break the hydrogen bonds between individual phosphate and guanidinium groups. Table 6.1 shows that the forces required are not so strong that covalent bonds in the DNA molecule are broken (a force greater than 1nN would be required[106]) while being pulled away, therefore guanidinium is an acceptable base grabber when used with single molecules of DNA.

Translocation through the nanopore will be controlled electrophoretically, which can apply 0.24±0.002pN/mV[139] of force to the DNA (100pN with 100mV), and with magnetic tweezers, which can apply up to 150pN.[128] A slightly larger force than either of these two methods alone is required to break the hydrogen bonds between guanidinium and phosphate (Table 6.1), and this does not even take into account the hydrogen bonding between the bases on the DNA and the base reader on the other electrode. As a result, even though the interaction between guanidinium and the phosphate backbone of DNA is reversible when single molecules are used, the use of an electric field and magnetic tweezers in parallel may be necessary to overcome the molecular friction between the DNA and the guanidinium and base reader on the electrodes.

REFERENCES

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, 4th ed., Garland Science, New York, NY, **2002**.

[2] *How Different is One Human Genome From Another?* **2003**, *J. Craig Venter Intitute Genome News Network*, http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp4_1.shtml.

[3] J. B. Heng, A. Aksimentiev, C. Ho, P. Marks, Y. V. Grinkova, S. Sligar, K. Schulten, G. Timp, *Biophys. J.* **2006**, *90*, 1098-1106.

[4] N. M. Toan, D. Marenduzzo, C. Micheletti, *Biophys. J.* **2005**, *89*, 80-86.

[5] J. Yan, J. F. Marko, *Phys. Rev. Lett.* **2004**, *93*, 108108-108112.

[6] S. V. Kuznetsov, Y. Shen, A. S. Benight, A. Ansari, *Biophys. J.* **2001**, *81*, 2864-2875.

[7] S. B. Smith, Y. Cui, C. Bustamante, *Science* **1996**, *271*, 795-799.

[8] M. P. Ball, Wikipedia, **2007**.

[9] *Human Genome Project Information* **2006**, www.ornl.gov/hgmis.

[10] *National Institutes of Health* **2008**, *Revolutionary Genome Sequencing Technologies – The $1000 Genome*, http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml.

[11] J. Shendure, R. D. Mitra, C. Varma, G. M. Church, *Nature Reviews* **2004**, *5*, 335-344.

[12] L. Franca, E. Carrilho, T. Kist, *Quarterly Reviews of Biophysics* **2002**, *35*, 169-200.

[13] F. Sanger, A. R. Coulson, *Journal of Molecular Biology* **1975**, *94*, 441-448.

[14] F. Sanger, S. Nicklen, A. R. Coulson, *Proceedings of the National Academy of Science* **1977**, *74*, 5463-5467.

[15] A. H. Ahsan, *Sanger Sequencing* **2002**, *McGill University*.

[16] M. L. Metzker, *Genome Research* **2005**, *15*, 1767-1776.

[17] A. Chien, D. B. Edgar, J. M. Trela, *Journal of Bacteriology* **1976**, *127*, 1550-1557.

[18] James W. Jorgenson, K. D. Lukacs, *Science* **1983**, *222*, 266-272.

[19] A. S. Cohen, D. R. Najarian, B. L. Karger, *Journal of Chromatography* **1990**, *516*, 49-60.

[20] Wolfgang Goetzinger, Lev Kotler, Emanuel Carrilho, Marie C. Ruiz-Martinez, Oscar Salas-Solano, B. L. Karger, *Electrophoresis* **1998**, *19*, 242-248.

[21] Haihong Zhou, Arthur W. Miller, Zoran Sosic, Brett Buchholz, Annelise E. Barron, Lev Kotler, B. L. Karger, *Analytical Chemistry* **2000**, *72*, 1045-1052.

[22] Xiaohua C. Huang, Mark A. Quesada, R. A. Mathies, *Analytical Chemistry* **1992**, *64*, 967-972.

[23] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, L. E. Hood, *Nature* **1986**, *321*, 674-679.

[24] J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jensen, K. Baumeister, *Science* **1987**, *238*, 336-341.

[25] Jingyue Ju, Chihchuan Ruan, Carl W. Fuller, R. A. M. Alexander N. Glazer, *Proceedings of the National Academy of Science* **1995**, *92*, 4347-4351.

[26] Su-Chun Hung, Jingyue Ju, Richard A. Mathies, A. N. Glazer, *Analytical Biochemistry* **1996**, *243*, 15-27.

[27] Andre Marziali, M. Akeson, *Annual Review of Biomedical Engineering* **2001**, *3*, 195-223.

[28] Minsoung Rhee, M. A. Burns, *Trends in Biotechnology* **2006**, *24*, 580-586.

[29] Jens Stephan, Klaus Dorre, Susanne Brakmann, Thorsten Winkler, Timm Wetzel, Markus Lapczyna, Michael Stuke, Bernhard Angerer, Waltraut Ankenbauer, Zeno Foldes-Papp, Rudolf Rigler, M. Eigen, *Journal of Biotechnology* **2001**, *86*, 255-267.

[30] Jianxun Mou, Daniel M. Czajkowsky, Yiyi Zhang, Z. Shao, *FEBS Letters* **1995**, *371*, 279-282.

[31] Stanislaus S. Wong, Adam T. Woolley, Teri Wang Odom, Jin-Lin Huang, Philip Kim, Dimitri V. Vezenov, C. M. Lieber, *Applied Physics Letters* **1998**, *73*, 3465-3467.

[32] Adam T. Woolley, Chantal Guillemette, Chin Li Cheung, David E. Housman, C. M. Lieber, *Nature Biotechnology* **2000**, *18*, 760-763.

[33] Jordanka Zlatanovaa, Stuart M. Lindsay, S. H. Leuba, *Progress in Biophysics & Molecular Biology* **2000**, *74*, 37-61.

[34] Matthias Rief, Hauke Clausen-Schaumann, H. E. Gaub, *Nature Structural Biology* **1999**, *6*, 346-349.

[35] S.B. Hladky, D. A. Haydon, *Nature* **1970**, *5231*, 451-453.

[36] David W. Deamer, M. Akeson, *Trends in Biotechnology* **2000**, *18*, 147-151.

[37] J. J. Kasianowicz, ERIC Brandin, DANIEL Branton, D. W. Deamer, *Proceedings of the National Academy of Science* **1996**, *93*, 13770-13773.

[38] Langzhou Song, Michael R. Hobaugh, Christopher Shustak, Stephen Cheley, Hagan Bayley, J. E. Gouaux, *Science* **1996**, *274*, 1859-1866.

[39] Mark Akeson, Daniel Branton, John J. Kasianowicz, Eric Brandin, D. W. Deamer, *Biophysical Journal* **1999**, *77*, 3227-3233.

[40] C. Dekker, *Nanture Nanotechnology* **2007**, *2*, 209-215.

[41] H. Chang, B. M. Venkatesan, S. M. Iqbal, G. Andreadakis, F. Kosari, G. Vasmatzis, D. Peroulis, R. Bashir, *Biomedical Microdevices* **2006**, *8*, 263-269.

[42] Jiali Li, Derek Stein, Ciaran McMullan, Daniel Branton, M. J. Aziz, J. A. Golovchenko, *Nature* **2001**, *412*, 166-169.

[43] J. B. Heng, A. Aksimentiev, C. Ho, P. Marks, Y. V. Grinkova, S. Sligar, K. Schulten, G. Timp, *Biophysical Journal* **2006**, *90*, 1098-1106.

[44] A. J. Storm, J. H. Chen, H. W. Zandbergen, C. Dekker, *Physical Review E* **2005**, *71*, 051903-051910.

[45] Arnold J. Storm, Cornelis Storm, Jianghua Chen, Henny Zandbergen, Jean-Francüois Joanny, C. Dekker, *Nano Letters* **2005**, *5*, 1193-1197.

[46] H. Chang, F. Kosari, G. Andreadakis, G. V. M. A. Alam, R. Bashir, *Nano Letters* **2004**, *4*, 1551-1556.

[47] Ulrich F. Keyser, Bernard N. Koeleman, Stijn van Dorp, Diego Krapf, Ralph M. M. Smeets, Serge G. Lemay, N. H. Dekker, C. Dekker, *Nature Physics* **2006**, *2*, 473-477.

[48] H.-H. Lei, *Eye on DNA* **2008**, *Whole Genome Sequencing Costs Continue to Drop*.

[49] J. A. Robertson, *AJOB* **2003**, *3*, W35-W42.

[50] A. Filipski, S. Kumar, *Comparative genomics in eukaryotes* Elsevier, San Diego, **2005**.

[51] *Human Genome Program U.S. Dept. of Energy Office of Science* **2006**, *H. G. P., Functional and Comparative Genomics Fact Sheet*, http://www.ornl.gov/sci/techresources/Human_Genome/faq/compgen.shtml.

[52] *Dept. Energy Microbial Genome Program* **2006**, http://microbialgenomics.energy.gov/organisms.shtml.

[53] *Genetic Genealogy* **2006**, *http://www.dnaancestryproject.com*.

[54] S. Parthasarathy, *SSS* **2005**, *35*, 5-40.

[55] J. L. Hustead, A. Cunningham, J. Goldman, California HealthCare Foundation, **2002**.

[56] B. McKibben, *Enough*, Henry Holt and Co., LLC, New York, NY, **2003**.

[57] J. Mortimer, *Medicine at Michigan* **2004**, *6*, 2.

[58] A. Lipmann, *Am. J.L. & Med.* **1991**, *17*, 15-39.

[59] *World Health Organization* **2006**, *Working Together for Health: The World Health Report*, http://www.who.int/nha/country/en/index.html.

[60] *447 U.S. 303* **1980**, *Diamond v. Chakrabarty*.

[61] *American Medical Association's Council on Ethical and Judicial Affairs* **2008**, *Gene patent guidelines*, http://www.ama-assn.org/ama/pub/category/3607.html.

[62] *Human Genome Project Information, Genetics and Patenting* **2006**, *2006*, www.ornl.gov/hgmis.

[63] S. Mayer, *Bio-IT World* **2002**, *http://www.bio-itworld.com/archive/111202/insights_public.html*.

[64] M. Tan, in *The Berkeley Electronic Press*, **2007**.

[65] *Human Genome Project, Legal, Ethical, and Social Issues* **2006**, *www.ornl.gov/hgmis*.

[66] M. K. Schwartz, *Clin. Chem.* **1999**, *45*, 5739-5745.

[67] *Human Genome Project, Genetics Privacy and Legislation* **2006**, *www.ornl.gov/hgmis*.

[68] *Genetic Information Nondiscrimination Act* **2005**, *109th Congress*, S 306.

[69] N. L. Jones, in *Congressional Research Service*, **2007**, p. 37.

[70] K. A. Lee, in *Congress Daily*, Washington, D.C., **2007**.

[71] F. S. Collins, J. D.Watson, *Science* **2003**, *302*, 745.

[72] V. G. Gregoriou, M. S. Braiman, CRC Press, Boca Raton FL, **2006**, p. 430.

[73] D. B. Mawhinney, J. John A. Glass, J. John T. Yates, *J. Phys. Chem. B* **1997**, *101*, 1202-1206.

[74] W. Saenger, *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 344-362.

[75] M. R. deJong, J. Huskens, D. N. Reinhoudt, *Chem. Eur. J.* **2001**, *7*, 4164-4170.

[76] M. T. Rojas, R. Kiiniger, J. F. Stoddart, A. E. Kaifer, *J. Am. Chem. Soc.* **1995**, *117*, 336-343.

[77] Q. Spadola, S. Lindsay, R. Bension, P. Zhang, *Org. Lett.* **2008**, *submitted*.

[78] C. Henke, C. Steinem, A. Janshoff, G. Steffan, H. Luftmann, M. Sieber, H.-J. Galla, *Anal. Chem.* **1996**, *68*, 3158-3165.

[79] R. C. Petter, J. S. Salek, C. T. Sikorski, G. Kumaravel, F.-T. Lin, *J. Am. Chem. Soc.* **1990**, *112*, 3860-3868.

[80] Steven L. Hauser, Eric W. Johanson, Heather P. Green, P. J. Smith, *Organic Letters* **2000**, *2*, 3575-3578.

[81] A. V. Eliseev, H.-J. Schneider, *JACS* **1994**, *116*, 6081-6088.

[82] K. D. Massoumeh Bagheri, Zolfaghar Rezvani, Ali Akbar Entezami, *European Polymer Journal* **2004**, *40*, 865-871.

[83] R. A. B. Emilio Cordova, Neil Spencer, Peter R. Ashton, J. Fraser Stoddart, Angel E. Kaifer, *Journal of Organic Chemistry* **1993**, *58*, 6550-6552.

[84] K. Eliadou, K. Yannakopoulou, A. Rontoyianni, I. M. Mavridis, *J. Org. Chem.* **1999**, *64*, 6217-6226.

[85] B. Ashcroft, Q. Spadola, S. Qamar, P. Zhang, G. Kada, R. Bension, S. M. Lindsay, *Small* **2008**, *accepted*.

[86] K. A. Connors, *J Pharm Sci* **1995**, *84*, 843-848.

[87] O. Wolter, T. Bayer, J. Greschner, *J. Vac. Sci. Tech.* **1991**, *B 9*, 1353-1357.

[88] N. H. Thomson, S. Kasas, B. Smith, P. K. Hansma, H. G. Hansma, *Langmuir* **1996**, *12*, 5905-5908.

[89] D. Pastre, O. Pietrement, S. Fusil, F. Landousy, J. Jeusset, M.-O. David, L. c. Hamon, E. L. Cam, A. Zozime, *Biophys. J.* **2003**, *85*, 2507-2518.

[90] A. Kichler, *J Gene Med* **2004**, *6*, S3-S10.

[91] K. Feldman, G. Hahner, N. D. Spencer, P. Harder, M. Grunze, *J. Am. Chem. Soc.* **1999**, *121*, 10134-10141.

[92] K. Umemura, M. Ishikawa, R. Kuroda, *Anal. Biochem.* **2001**, *290*, 232-237.

[93] T. Ederth, K. Tamada, P. M. Claesson, R. Valiokas, J. R. Colorado, M. Graupe, O. E. Shmakova, T. R. Lee, *J. Colloid Interface Sci.* **2001**, *235*, 391-397.

[94] G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **1986**, *56*, 930-933.

[95] *Integrated DNA Technologies* **2008**, *SciTools OligoAnalyzer 3.1*.

[96] H.-J. Butt, M. Jaschke, *Nanotechnology* **1995**, *6*, 1-7.

[97] T. V. Ratto, K. C. Langry, R. E. Rudd, R. L. Balhorn, M. J. Allen, M. W. McElfresh, *Biophys. J.* **2004**, *86*, 2430-2437.

[98] F. Hanke, H. J. Kreuzer, *Phys. Rev. E* **2006**, *74*, 0319091 - 0319095.
[99] E. Evans, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 105-128.

[100] P. Nelson, *Biological Physics: Energy, Information, Life*, W.H. Freeman and Company, New York, NY, **2004**.

[101] J. F. Marko, E. D. Siggia, *Macromolecules* **1995**, *28*, 8759-8770.

[102] A. Storm, J. Chen, X. Ling, H. Zandbergen, C. Dekker, *Nature Mat., 2003* **2003**, *2*, 537-540.

[103] C. Bustamante, *Science* **1994**, *265*, 1599-1600.

[104] J. L. Hutter, J. Bechhoefer, *Rev. Sci. Instruments* **1993**, *64*, 1868-1873.

[105] F. Oesterhelt, M. Rief, H. E. Gaub, *N. J. Ph.* **1999**, *1*, 6.1-6.11.

[106] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, H. E. Gaub, *Science* **1999**, *283*, 1727-1729.

[107] M. A. Spies, R. L. Schowen, *J. Am. Chem. Soc.* **2002**, *124*, 14049-14053.

[108] J. Mathe, A. Aksimentiev, D. R. Nelson, K. Schulten, A. Meller, *PNAS* **2005**, *102*, 12377-12382.

[109] B. A. Ashcroft, Dissertation thesis, Arizona State University (Tempe AZ), **2007**.

[110] S. Qamar, P. M. Williams, S. M. Lindsay, *Biophys. J.* **2008**, *94*, 1233-1240.

[111] W. A. Vercoutere, S. Winters-Hilt, V. S. DeGuzman, D. Deamer, S. E. Ridino, J. T. Rodgers, H. E. Olsen, A. Marziali, M. Akeson, *Nucleic Acids Research,* **2003**, *31*, 1311-1318.

[112] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, D. Branton, *Phys. Rev. Let.* **2003**, *93*, 238101-238105.

[113] S. Cuesta-Lopez, M. Peyrard, D. J. Graham, *The European Physical Journal E* **2005**, *16*, 235-246.

[114] J. Muth, P. M. Williams, S. J. Williams, M. D. Brown, D. C. Wallace, B. L. Karger, *Electrophoresis* **1996**, *17*, 1875-1883.

[115] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr., C. Bustamante, *Science,* **2001**, *292*, 733-737.

[116] W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, *Phys. Rev. Lett* **2005**, *95*.

[117] U. Dornberger, M. Leijon, H. Fritzsche, *J. Biol. Chem.* **1999**, *274*, 6957-6962.

[118] S. A. Darst, A. M. Edwards, E. W. Kubalek, R. D. Kornberg, *Cell* **1991**, *66*, 121-128.

[119] M. Schmid, T. Wien, Wikipedia, **2005**.

[120] G. Baym, *Lectures on Quantum Mechanics*, Westview Press, New York, **1990**.

[121] S. M. Lindsay, *Journal of Chemical Education* **2005**, *82*, 727-733.

[122] X. Cui, Arizona State University (Tempe), **2001**.

[123] T. Ohshiro, Y. Umezawa, *Proceedings of the National Academy of Sciences* **2006**, *103*, 10-14.

[124] M. Lee, *personal communication* **2008**.

[125] L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, M. L. Steigerwald, *Nature* **2006**, *442*, 904-907.

[126] Jin He, Lisha Lin, Peiming Zhang, S. Lindsay, *Nano Letters* **2007**, *7*, 3854-3858.

[127] J. He, L. Lin, P. Zhang, Q. Spadola, Z. Xi, Q. Fu, M. Lee, O. Sankey, S. Lindsay, *submitted* **2008**.

[128] S. Lindsay, *National Human Genome Research Institute* **2007**, *Grant Proposal*.

[129] V. A. Bloomfield, *Biopolymers* **1997**, *44*, 269-282.

[130] Yonghai Song, Cunlan Guo, Lanlan Sun, Gang Wei, Chongyang Peng, Li Wang, Yujing Sun, Z. Li, *Journal of Physcial Chemistry B* **2007**, *111*, 461-468.

[131] S. C. Riemer, V. A. Bloomfield, *Biopolymers* **1978**, *17*, 785-794.

[132] J.F. Allemand, D. Bensimon, L. Jullien, A. Bensimon, V. Croquette, *Biophysical Journal* **1997**, *73*, 2064-2070.

[133] Lukasz Berlicki, Ewa Rudzinska, P. Mlynarz, P. Kafarski, *Current Organic Chemistry* **2006**, *10*, 2285-2306.

[134] Daumantas Matulis, Ioulia Rouzina, V. A. Bloomfield, *Journal of Molecular Biology* **2000**, *296*, 1053-1063.

[135] E. A. DiMarzio, C. M. Guttman, *Journal of Chemical Physics* **1991**, *95*, 1189-1197.

[136] H. Collaer, M. C. Haulait-Pirson, P. Huyskens, *Journal of Solution Chemistry* **1989**, *18*, 1037-1053.

[137] J. M. Williams, T. Han, T. P. Beebe, *Langmuir* **1996**, *12*, 1291-1295.

[138] H. Skulason, C. D. Frisbie, *Journal of the American Chemical Society* **2002**, *124*, 15125-15133.

[139] U. F. Keyser, B. N. Koeleman, S. VanDorp, D. Krapf, R. M. M. Smeets, S. G. Lemay, N. H. Dekker, C. Dekker, *Nature Physics* **2006**, *2*, 473-477.