2013-06-11

The Effects of Mismatches and Probe Tethering Configurations on the Stability of DNA Duplexes on Surfaces

Kyle Evan Pratt
Brigham Young University - Provo

The Effects of Mismatches and Probe Tethering Configurations on the Stability of DNA Duplexes on Surfaces

Kyle Evan Pratt

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Thomas A. Knotts IV, Chair
Vincent Wilding
William Hecker

Department of Chemical Engineering

Brigham Young University

June 2013

ABSTRACT

The Effects of Mismatches and Probe Tethering Configurations,
on the Stability of DNA Duplexes on Surfaces

Kyle Pratt
Department of Chemical Engineering BYU
Master of Science

DNA microarrays are chip-based, analysis tools which can perform hundreds of thousands of parallel assays to determine the identity of genes or gene expression levels present in a sample. They have been identified as a key technology in genomic sciences and emergent medical techniques; however, despite their abundant use in research laboratories, microarrays have not been used in the clinical setting to the fullest potential due to the difficulty of obtaining reproducible results. Microarrays work on the principle of DNA hybridization, and can only be as accurate as this process is robust. Fundamental, molecular-level understanding of hybridization on surfaces is needed to further refine these devices.

This work shows how orientation of DNA probes with respect to the surface affects the thermodynamics and stability of hybridization. Ideal surface hybridization (a DNA duplex bound to the surface on one end) is compared to more realistic conditions such as interaction between DNA and the surface in multiple locations. This research also describes the effect of mismatch location and number of mismatches on a single target strand. The results clarify key details of the biophysics involved in microarray performance and this knowledge can be used to improve next-generation devices. The disparity between surface and bulk hybridization behavior is examined here in molecular level detail that is not currently possible with experimental techniques.

Keywords: DNA, molecular modeling, microarray, simulation, hybridization

ACKNOWLEDGMENTS

Thank you to my supportive wife Sara and to my advisor Dr. Knotts for all of their help through my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 DNA Microarrays

DNA microarrays are chip-based, analysis tools that can screen biological samples for hundreds of thousands of genes (sections of DNA containing genetic code on how to make proteins) simultaneously. This ability means that microarrays have potential use in a variety of fields including criminal investigations and medicine. The latter is of particular interest in designing precise, patient-tailored treatment strategies based upon individual genetics. However, despite the promise of the technology, challenges remain to ensure the devices produce reproducible and reliable results.

Microarrays are essentially dots of specific gene sequences arranged on a surface at addressable locations (See Figure 1). Each dot is composed of many molecules (called *probes*) of single-stranded DNA (ssDNA) of the same sequence corresponding to specific genes of importance. A typical experiment consists of fluorescently labeling a test sample of ssDNA (called targets) whose gene composition is unknown and incubating the sample over the microarray. During the incubation, targets in the sample will hybridize with complementary surface probes. The array is then analyzed, and the genes present in the sample can be identified by the address of the fluorescence on the chip.

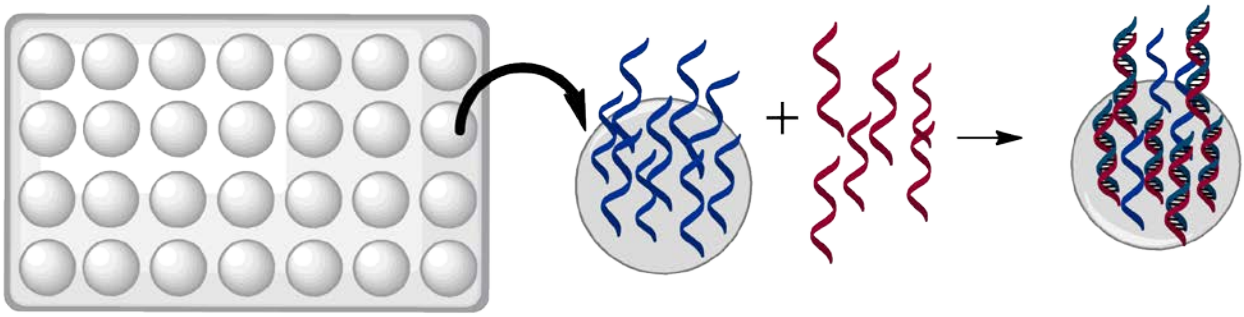

Figure 1 - A microarray is spotted with thousands of dots at addressable locations. Each dot is covered with probe DNA strands, and each dot only contains strands with the exact same sequence. Target strands of DNA from a sample are added and complementary strands hybridize. Locations with matches are determined via florescent tags on target strands.

Microarrays can be used in ways other than simply identifying the genes present in a sample. Specifically, the fluorescent intensity of a specific dot on the chip varies proportionally to the amount of the corresponding gene present in the sample. This means microarrays can also be used to identify gene expression levels (the amount of genes) present and for comparative genomics. The latter, described in Figure 2, is of particular importance in the medical field where it can be used to determine the genetic origins of disease.

Control and test cells are harvested for sample DNA or RNA. Each sample is labeled with fluorescent molecules. The fluorophore of one sample is designed to emit at a different wavelength than the other so that the effects of both populations can be tracked. The labeled samples are then incubated over the microarray. Imaged microarrays reveal expression level differences between control and test cells through variations in fluorescent colors. For the system depicted in Figure 2, those dots which fluoresce green mean that the cells in the control population express a particular gene that the test cells do not. Those dots that fluoresce red mean

that the test cells express a gene that the control cells do not. Yellow and orange mean that those genes are expressed to some degree by both sets of cells.

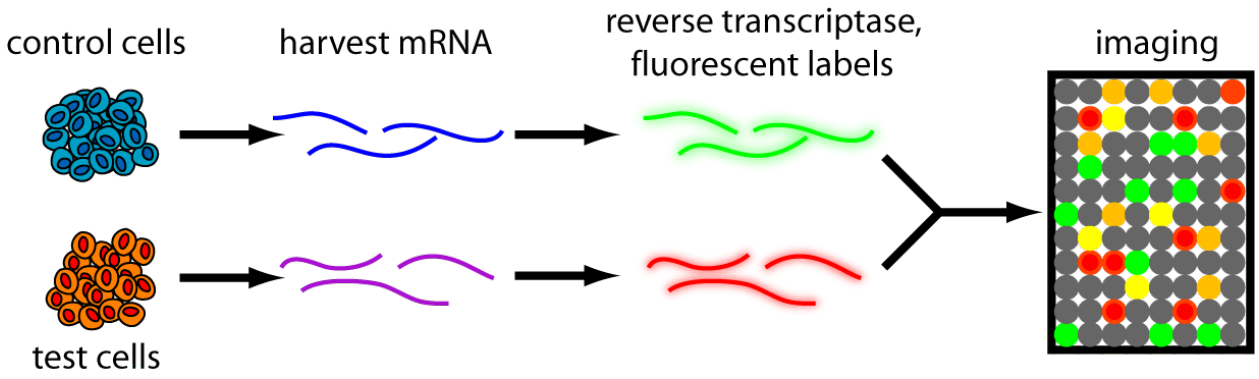

Figure 2 - Comparison of sample DNA or RNA from control and test cells using a Microarray platform. Used with permission.

Microarrays are used extensively in research settings and have the potential to revolutionize the health care industry by facilitating patient-tailored diagnoses and therapies. Despite their potential they have not been adopted into mainstream use by the medical field. The main reason for this is that microarrays provide very inconsistent results[1].

Much effort has gone into optimizing the technology, but efforts are hindered because the behavior of DNA hybridization on surfaces is not fully understood. A significant limitation is that experiments using DNA microarrays do not provide detail at the molecular level. This makes it difficult to identify the mechanisms by which hybridization occurs, the hindrances to the process, and possible approaches to improve the situation. As will be described later, the purpose of this research is to develop an improved, molecular-level view of DNA hybridization as it occurs on microarrays. Specifically, this research illustrates the effects that the microarray

environment has on the hybridization process compared to the typical circumstances that occur in a bulk phase.

1.2 Limitations of Microarrays

Some research has been performed to test the reproducibility of DNA microarrays. Ivanova et al.[2] and Matsuzaki et al.[3] performed complementary studies using microarrays. These two groups both used DNA microarrays to examine mouse DNA samples for genes with stem cell characteristics. The two groups used the same DNA sample preparation and the same microarrays, but only 4 genes were identified in common. Fortunel et al.[1] repeated these experiments in order to validate the conclusions of the studies. They found 359 sequence matches when comparing multiple samples, the other two studies found 275 and 201 matches, but only 1 sequence was common among the three experiments. These discrepancies indicate that reliability must be improved for microarrays to be useful in research and as clinical diagnostics.

The MicroArray Quality Control Project[4] was an intensive study of the reliability of DNA Microarrays. 137 researchers from 51 organizations participated in the study. The researchers examined 6 microarray platforms and attempted to use highly-standardized laboratory practices and reporting methods to improve reproducibility. They did find some success in improving reliability by comparing the results of multiple tests from the same platform. However even these improvements still resulted in significant sample variance, sometimes in the 5% to 20% range. Such error is still too high for clinical use, and improvements in the base technology were called for by the researchers. Other research has been done in increasing reliability of the results by signal amplification,[5] but such did not solve the problem.

Despite all these experimental efforts microarrays still require improvement before widespread adoption of the devices is possible.

1.3 Previous Simulation Research into Hybridization

As briefly mentioned, one of the difficulties encountered when trying to improve DNA microarrays is that very little is understood about DNA hybridization on the molecular level, especially when surfaces are involved. This is due to the fact that no experimental techniques exist to directly measure or visualize the hybridization process as it occurs. Because of this, simulation has emerged as the primary tool to study hybridization[6,7,8].

Wong et al.[9] performed an atomistic study of a 12 base pair strand of hybridized DNA. This simulation illustrated how the presence of a surface affects stability simply by orientation of the duplex. They later extended this simulation over a longer time scale[10] but the atomistic nature of the simulation restricted the number of base pairs that could be represented. Neither did the probe and target strands separate to a significant distance during the simulation. To obtain reliable thermodynamic data on hybridization, separation distances must be large and simulation times must be long enough to allow hybridization to occur many times.

Hagan et al.[11] performed a computationally intensive atomistic study of 3 base pairs, but only simulated the unstacking of one base. Coarse grain models can simulate the behavior of much longer strands of DNA, even thousands of base pairs,[12] but to describe behavior relevant to the scale of DNA microarrays the appropriate level of detail must be designed in. Hybridization of a long strand was simulated by Jayaraman et al.[13] through use of a coarse grain model. This was a Monte Carlo simulation and looked at how various factors affect the extent of hybridization, including probe length and temperature.

Schmitt et al.[14], using an advanced model for DNA, compared hybridization in the bulk to hybridization on a surface and found that the presence of a surface increased the stability of hybridization, even when the strands were not perfectly matched. This is significant because conventional wisdom at the time suggested that hybridization was hindered on the surface. In fact, some microarrays had been designed to counteract the suspected hindrance to hybridization by making the surface more attractive to the DNA. The results of Schmitt et al. showed that such a design plan would not aid the accuracy or effectiveness of microarrays. They also demonstrate the importance of additional studies into the subject. Specifically, microarray fabrication and procedure has been based upon the assumption that surfaces inhibit hybridization. Such an approach is likely to contribute to the poor reproducibility of microarrays, but more work is needed before the community sees the need to change standard practice.

The model used by Schmitt et al. was re-parameterized by Sambriski et al.[15-16] who were able to uncover more details[17] of the actual mechanism of hybridization. Specifically, their findings challenged the intuitive idea of hybridization occurring through a "zipping" motion. Their simulations showed a preference to align non-complementary ends and then shift into correct alignment, but this work was only performed in the bulk.

Schmitt et al. also examined the effects of a single mismatch in the center of a strand on hybridization. Their study showed a very small decrease in stability with the addition of a single mismatch. They also found that the mismatched strands on the surface were more stable than perfectly matched strands in the bulk and, based on this finding, proposed ways to improve microarray specificity by slightly decreasing overall sensitivity, but these have not been tested.

One factor that has not been thoroughly examined at the molecular level is the number and location of mismatches in the duplex. As already mentioned, Schmitt et al. studied duplexes

with only one mismatch in one location. Others have also taken this approach. For example, Held et al.[18] experimentally compared perfectly matched strands and strands with a single mismatch in the center. However, Held et al. only reported the variation in melting temperature, which was approximately 4 K and the information has not really proven useful in the design realm. Kim et al.[19] measured the thermodynamics of a single mismatch at the end of a strand, two mismatches at the same end, a single mismatch at the center, and two mismatches at the center mismatches as a function of temperature. The study showed that thermodynamics could be predicted by the nearest neighbor base-pair model, but these data were done only for bulk hybridization so their applicability to the microarray system is questionable.

Peterson et al.[20] studied hybridization of mismatches on surfaces using experimental techniques. The study showed dependence on target strand sequence as well as probe density. Strands with up to two mismatches were examined; however, this study focused on the kinetics of hybridization and did not address the stability of the process. Though other research has been done comparing surface and bulk hybridization, more complete and detailed work remains. The effects of mismatches need to be more completely and systematically researched. Both the thermodynamic differences in hybridization strength and the effects on mechanism of various numbers of mismatches need deeper exploration.

1.4 Non-Ideal Binding Conformations

The study performed by Schmitt et al. also highlights a common limitation of simulation work done to date, namely, the system modeled was idealized in that only a single probe molecule interacted with a single target molecule. In such a situation, the probe strand was perfectly accessible to the target strand and ready to hybridize. It was also only attached to the

surface at one location on the end of the molecule. Real microarrays involve phenomena that cannot be captured with this simple, two-molecule system. For example, many probe strands are present in each "dot" on a microarray surface and nothing prevents tangling from occurring (shown in Figure 3).

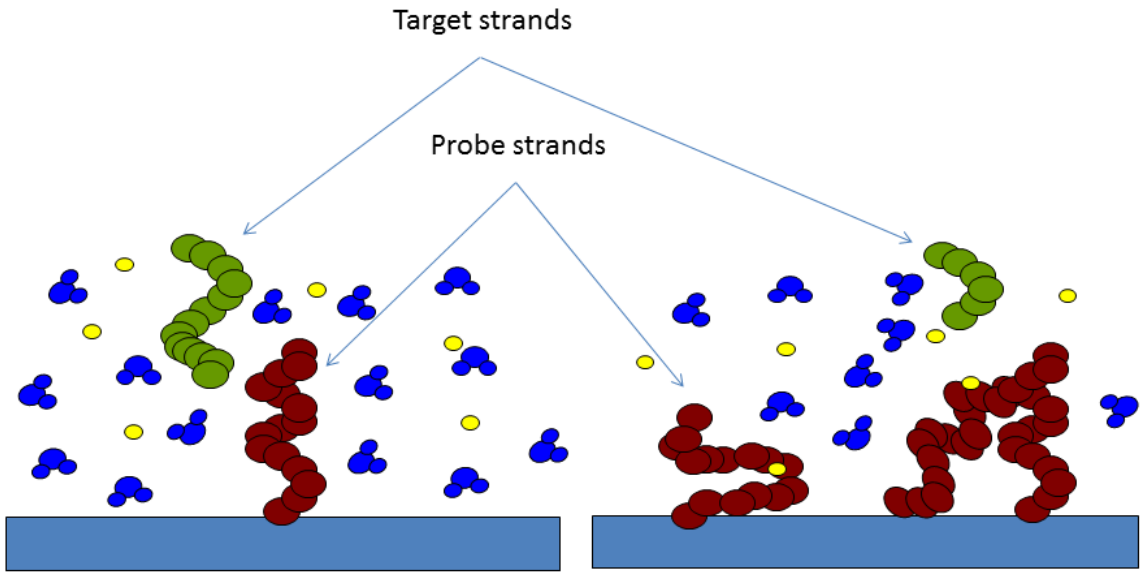

Figure 3 - An ideal DNA microarray would have probe strands perpendicular to the surface, easily accessible to target strands in the bulk. This situation is pictured on the left. On the right, a more realistic situation is depicted. Probe strands are likely tangled and in positions more parallel to the surface. They may also be attached to the surface at multiple locations.

Also not prohibited is that probes can be adsorbed to the surface or even be attached to the surface in multiple locations (shown in Figure 4). In one microarray manufacturing technique, the spotted array method, DNA probes are literally placed on the surface in a dot of solution. This method can easily result in high levels of surface interaction. Strands in such a non-ideal setting will likely be less accessible to bonding with target strands, and may show

differences in overall bond strength. One microarray manufacturing technique is the spotted array method. In this procedure, DNA probes are literally placed on the surface in a dot of solution. This method can easily result in high levels of surface interaction.

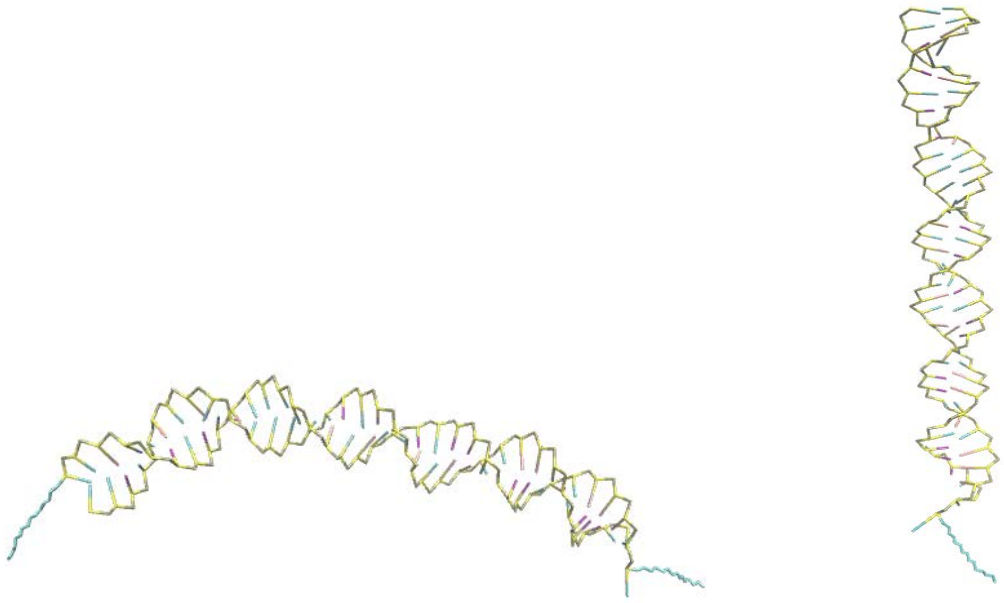

Figure 4 - On the left is a simulation of a hybridized probe strand that is held parallel to the surface by 2-tethers. On the right for comparison is a hybridized probe strand held to the surface by only one tether and is perpendicular to the surface.

Such tangling or other surface interaction is not accounted for in current preparation schemes or models, but is likely to be significant. For example Schreiner et al.[21] experimentally examined DNA chemically adsorbed to a surface. The adsorbed strands were determined to be less stable than hybridized pairs in the bulk or tethered to the surface. The strands they tested were comprised of only one base pair type, potentially neglecting many effects of strands comprised of multiple base types. A strand held near the surface by tangling with other nearby strands would likely have more mobility than an adsorbed strand and therefore different

behavior. Much work needs to be done in this area to understand the implications of tangling on a molecular level.

Related to binding at both ends are the effects of multiple strands on a surface. Previous simulation studies have only compared the effects of DNA hybridization in bulk solution and of a single probe strand on a surface, but as mentioned above, multiple strands are present on the surface and are likely tangled. This effect is not taken into account in current microarray design procedures. Schreiner et al. examined the effects of multiple strands on a gold surface experimentally and found minimal effect through competition on stability or signal intensity. But the study was limited to a single type of probe strand and did not quantify the stability of the process.

Ambia-Garrido et al.[22] used simulation to examine multiple probe effects with a rudimentary meso-scale model of 3 spheres for every 3 base pairs. Simulations were performed on gold and SiO_2 surfaces and electrical and probe density effects were examined. They found that probe molecules increased length in the presence of close neighbors and were oriented perpendicular to the surface rather than aggregating together. The effects were attributed to repulsion forces, the degree of which was enough to separate the strands but not cause them to orient parallel to the surface. However, the model used was not parameterized against experimental data and hybridization was not considered, both of which call into question the applicability of the results to the microarray situation.

Other studies have also been done to study the effects of multiple probes on the surface. For example, Irving et al.[23] modeled the effects of high surface packing (DNA volume fraction >25%) and compared predictions to experimental data. Probe-probe interactions, including binding of complementary segments were accounted for in the model which helped expand the

results beyond the repulsive effects reported by Ambia-Garrido et al., but the study still did not examine hybridization.

Cherepinsky et al.[24] studied the competitive effects of multiple probe strands with relatively few target strands, a more realistic microarray system. In these simulations two different probe strands compete for a single target which is complementary to one of the probes. While the results help give one interpretation of the variation between signal intensity of targets with pure matching probes and targets with multiple probe types, the simulation was done with a rudimentary mathematical model based on affinity constants. As such, it neglects conformational factors and does not provide the detail needed to understand the molecular-level interactions in a way that can help future design.

Finally, Peterson et al.[20] experimentally studied the effects of probe density finding that increased probe density led to inefficiency in target hybridization. Crowding effects were suspected by the researchers, which is the opposite seen in the study by Ambia-Garrido et al. In short, a review of the relevant literatures shows the opportunity for further research into the molecular-level behavior of microarrays. In general, previous simulations have focused on ideal hybridization cases. Moreover, in the few cases where multiple probes were considered, hybridization was not simulated and the results are inconsistent with each other. Some predict crowding of probes on a surface while others predict little probe/probe interactions. As such, molecular simulation studies using an experimentally-parameterized model, which can capture the relevant phenomena such as hybridization and surface-induced conformational changes, would be a welcomed addition to the field of microarrays. The research here presented will be useful in providing quantitative data about the thermodynamic effects of non-ideal DNA microarray conditions that can be used to design next-generation technologies. The insights

gained from this research can also be applied to the design of other self-assembling DNA technologies, such as DNA origami.[25]

1.5 Goal and Outline

The objective of this research was to determine *how probe design affects the stability of DNA duplexes on surfaces on the molecular level*. Two factors will be investigated:

1. Probe tether configurations
2. Mismatched sequences

These are selected as they move the understanding of microarray function away from ideal cases to more real-life situations. As was discussed in the previous section, understanding of non-ideal effects is lacking.

A major issue with microarrays is that the probes are designed based on melting and stability data for duplexed oligonucleotides in the bulk. As has been discussed, the presence of a surface can drastically change the thermodynamics of the hybridization process, but the nature of these changes is not currently known. The purpose of the proposed research is to further the understanding of how DNA hybridization on a surface differs from that in the bulk so that future microarrays will be designed using more accurate stability data.

With the tether studies, it was expected that increasing numbers of tethers would make duplexes progressively less stable. It was also expected that longer tethers bring melting temperatures closer to bulk values. In mismatch studies it was hypothesized that mismatch location would have a small but noticeable effect on melting temperature. The exact hypotheses examined in each case are discussed in more detail in later chapters.

2 METHODOLOGY

2.1 Model

To achieve my research objectives I used advanced molecular simulation to probe various factors involved in DNA hybridization on surfaces in a manner not done previously. My methods included using a novel, coarse grain model for DNA which facilitates the calculation of the free energy of hybridization on surfaces. The general approach is to first compute the free energy of hybridization of various DNA sequences in the bulk (to serve as the controls), and then compare these to the free energies of hybridization that were computed for other situations such as multiple tether locations, mismatches, etc. (the treatments).

Models that account for *each* atom in a system with DNA are accurate, but the computational demands are high. State-of-the-art microarrays currently use probes on the order of 20-30 bases and measure from 68 to 102 Angstroms in length. To properly simulate hybridization the two strands need to be separated to a distance where each does not feel the effect of the other. Such occurs at distances of approximately 250 Angstroms. Add in enough water molecules to properly solvate both strands, and the size of the simulation box and the number of molecules needed to fill the volume far exceeds resources available from even the largest supercomputers. Moreover, thousands of such simulations are needed to calculate thermodynamics properties such as free energies which means all-atom models cannot be used.

Coarse grain models are simplified representations where multiple atomistic sites are groups together into one interaction site. Knotts et al.[26] developed a coarse grain DNA model that represents each nucleotide as 3 sites: a phosphate group, a sugar, and a base (Figure 5, below). This model has been shown to reproduce both the thermal and mechanical properties of DNA, such as salt-dependent melting and the characteristically-long persistence length of the molecule. As it is coarse grain, it does not include explicit water molecules. Instead it uses an implicit solvent to capture the shielding properties of water with a much lower computational expense. A representation of the coarse grain DNA is shown in Figure 6.

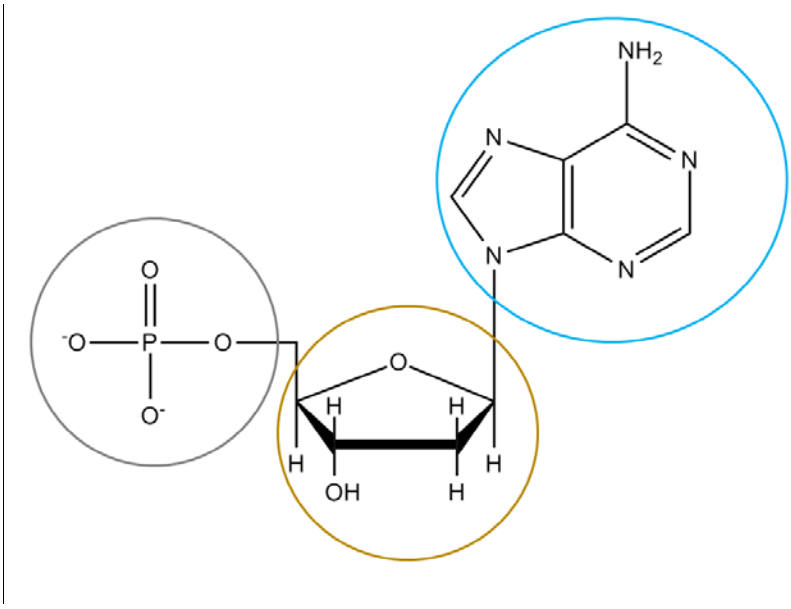

Figure 5 - An illustration of the Knotts model. Each nucleotide is represented by 3 sites. The left-most in this image is the phosphate site. The center is the sugar site. The right-most is the base site, in this case Adenine.

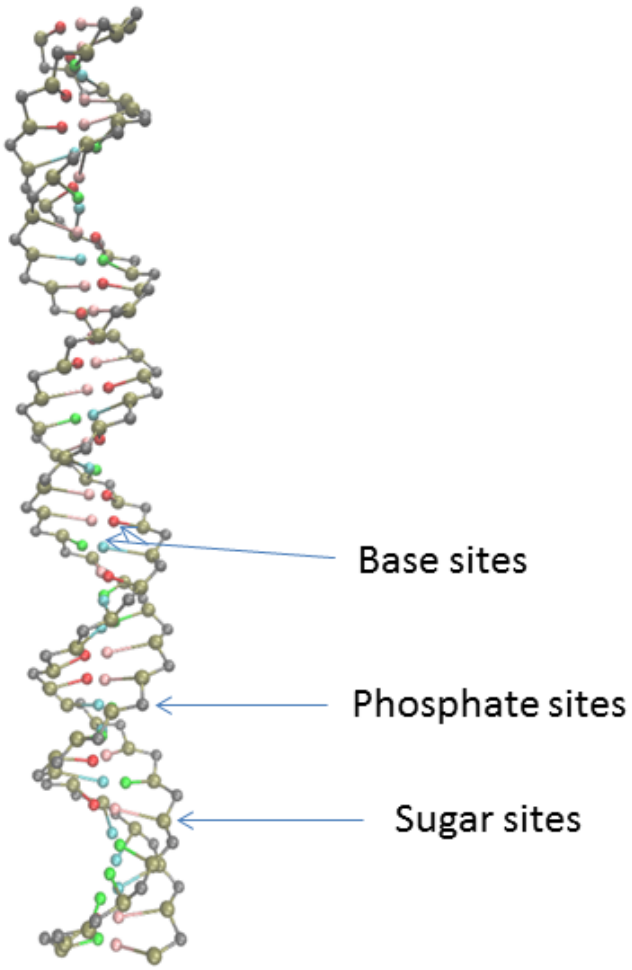

Figure 6 - An image of the coarse grain representation of a hybridized pairof DNA molecules.

The surface used in these simulations is designed to mimic the behavior of a glass slide on which DNA microarrays are typically built. Surfaces are capable of being modified to exhibit varying degrees of repulsion or attraction to DNA and this model is able to duplicate each of these cases. The simulations performed in this study used a short-ranged, purely-repulsive surface which is the type needed for DNA microarrays. Attractive surfaces aren't used because non-specific adsorption of the targets to the surface would occur obscuring the signals from real,

complementary hybridization. Repulsive surfaces aren't used as the targets would not be able to transport to the surface and hybridize. The exact surface used was modeled as a purely repulsive surface using a Weeks-Chandler-Anderson potential[27]. Tethers were based on the atomistic Wong and Pettit model, modeled as a bead-spring, coarse-grain linker[28].

2.2 Simulation Techniques

A significant advantage of simulation over experimentation is the ability to isolate and test very specific effects. Not only do molecular dynamics simulations allow visualization of DNA hybridization, they allow direct calculation of thermodynamic properties. Simulation accurately returns thermodynamic data by sampling properties values as the system reaches different configurations. In molecular dynamics simulations, the type used in this research, this is achieved by calculating the forces in the system and integrating Newton's equations of motion through time.

These types of simulations do have the limitation of potentially getting "stuck" in local energy minima rather than finding the globally, thermodynamically stable state of the system. Advanced simulation techniques are available that overcome these sampling issues and ensure that the relevant molecular configurations visited during the simulation which increases the likelihood of finding the most stable state. Replica exchange is one such technique. This method still uses molecular dynamics principles, but it applies temperatures to bias the simulation. The bias is then removed during analysis according to standard statistical mechanical principles.

Replica exchange is a method used to ensure that simulations run at low temperatures are less likely to become stuck in local energy minima. A replica exchange system has a certain number of simulations running at once. Each replica exchange simulation or "box" is identical in

composition but vary in the temperature of the system. At specified intervals, swaps between systems with the closest temperatures are attempted. These swaps are accepted or rejected based on the metropolis algorithm. Essentially, the metropolis algorithm is that if the proposed state is more likely than the previous state, it is accepted. If the proposed state is less likely than the previous state, it is accepted or rejected based on a given probability. In these swaps the molecules effectively trade places. This enables the lower temperature boxes to sample phase space more completely and escape from local energy minima. As these simulations only look at a single pair of DNA rather than sampling a large number of cases, complete sampling of phase space is critical to determining accurate average behavior. Replica exchange enables the DNA system to more consistently arrive at the most stable physical conformation. The resulting data more accurately reflect experimental systems than a simulation using only simple molecular dynamics simulation.

The main quantity of interest calculated from the replica exchange is the melting temperature of the DNA duplex. In simulation, the melting temperature corresponds to a state where half of the bases are no longer hybridized and is a measure of the stability of a duplex. Higher melting temperatures correspond to higher stabilities. Melting temperature is also the ubiquitous metric used when designing microarray probes. Microarrays are designed based on the expected stability of the potential hybridized probe/target complexes. If designed and manufactured correctly, each dot and associated probe/target complex on the array should exhibit nearly the same stability. In other words, the melting temperature of each possible probe/target duplex should be the same across the whole chip. This ensures that washing only removes truly mismatched strands and leaves perfectly-matched strands intact to eliminate both false positives and negatives.

Melting temperatures are calculated from the heat capacity curves produced by replica exchange simulations. When a duplex melts, there is a large spike in heat capacity. The temperature at which this spike occurs is the melting temperature. The general method for testing the hypotheses in this research is to compare the melting temperature of the control case (perfectly complementary duplexes in the bulk) with the melting temperature for different treatments. As discussed in the Introduction, these treatments include systems with mismatches and tethering to the surface at multiple locations. If the melting temperature of the treatment is higher than the control, the treatment conditions stabilized the duplex. This comparison allows the effects of the non-ideal hybridization conditions to be determined.

The simulations began with an "equilibrium" phase, which allows the sites to interact for a sufficient amount of time to remove any biases imposed by the initial positions and velocities. After the equilibrium phase of the simulation, the "production" phase began. Data for analysis are collected during this stage.

The primary goal of the analysis of the simulation results is the calculation of the partition function. Once the partition function is known, any thermodynamic property of interest can be calculated. WHAM is the technique used to calculate the partition function from replica exchange and umbrella sampling simulations. WHAM refers to the Weighted Histogram Analysis Method[29] and in addition to producing the partition function it is also used to find the free of the system. The simulation program stores data as histograms of the potential energy or some other property of the system. WHAM takes these data and calculates smooth property curves as a function of temperature.

The density of states of the system, $\Omega(U)$ is the number of ways to arrange the system and still maintain the same total energy. As already mentioned, it is calculated directly via

WHAM. Once the partition function is known, Equation 1 can be used to calculate the desired properties of the system that will aid analysis, in this case, determining the temperature dependence of property. X requires the canonical partition function (Q), the density of states for energy state I, $\Omega(U_i)$, and $\beta = 1/(k_B T)$. The $< >_T$'s denote the ensemble average at temperature T, and the summation over i includes all populated energy states.

$$X(T) = \langle X \rangle_T = \frac{1}{Q} \sum_i X(U_i) \Omega(U_i) e^{-\beta U_i} \quad (1)$$

One of the primary properties calculated using Equation 1 is the heat capacity. A plot of the heat capacity for a hybridized pair over a range of temperatures is relatively flat at low and high temperatures but peak in the middle. The peak in the heat capacity occurs when the hybridized DNA begins to come apart, a process known as melting. The temperature at which the heat capacity is a maximum is defined as the melting temperature. Figure 7 below gives a characteristic example of this behavior. As explained earlier, the melting temperature is important because it is an indicator of the stability of a duplex. A high melting temperature means that the DNA is more stable.

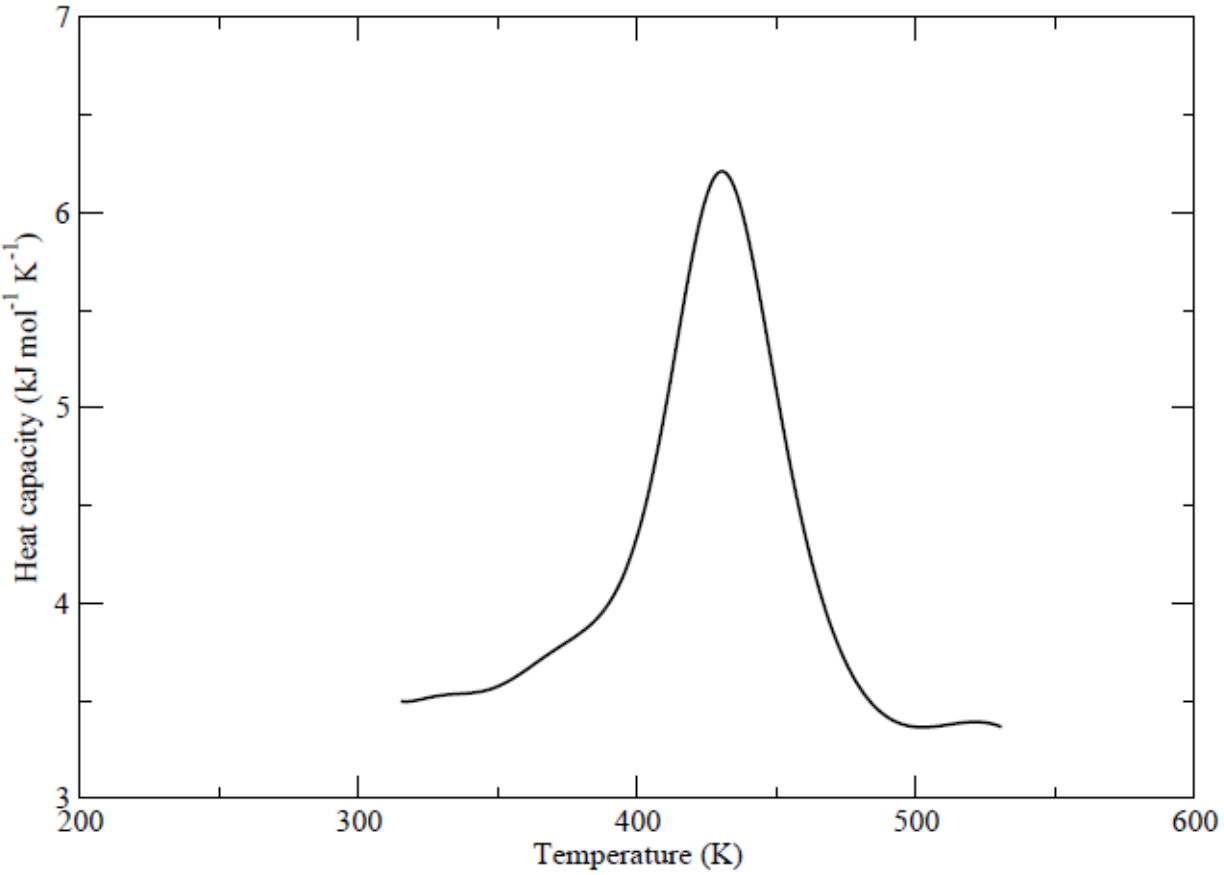

Figure 7 - Heat capacity of a DNA duplex vs. temperature. Relatively stable values are seen to the left and the right of the sharp peak. The peak occurs at the melting temperature, which is the point at which half of the bases are no longer hybridized.

To test the validity of the model, NVT MD simulations were performed to measure bubble formation. DNA bubbles are an important part of the melting mechanism of the duplex, and the results presented in Figure 8 show that the model accurately captures these phenomena. The melting mechanism often begins with the formation of a "bubble," a section within the DNA double helix where the bases have increased their separation distance and are effectively no longer bonded. At temperatures near, but still below, the melting temperature bubbles form and then collapse frequently. When a hybridized pair is heated above the melting temperature,

simulations show bubbles forming more frequently until one or both of the ends of DNA break the bond and begin to move freely. The degree of separation between the strands increases until they are fully separated. After the bonds between base pairs are broken, the strands often stay in close proximity for a brief period before rapidly separating from each other.

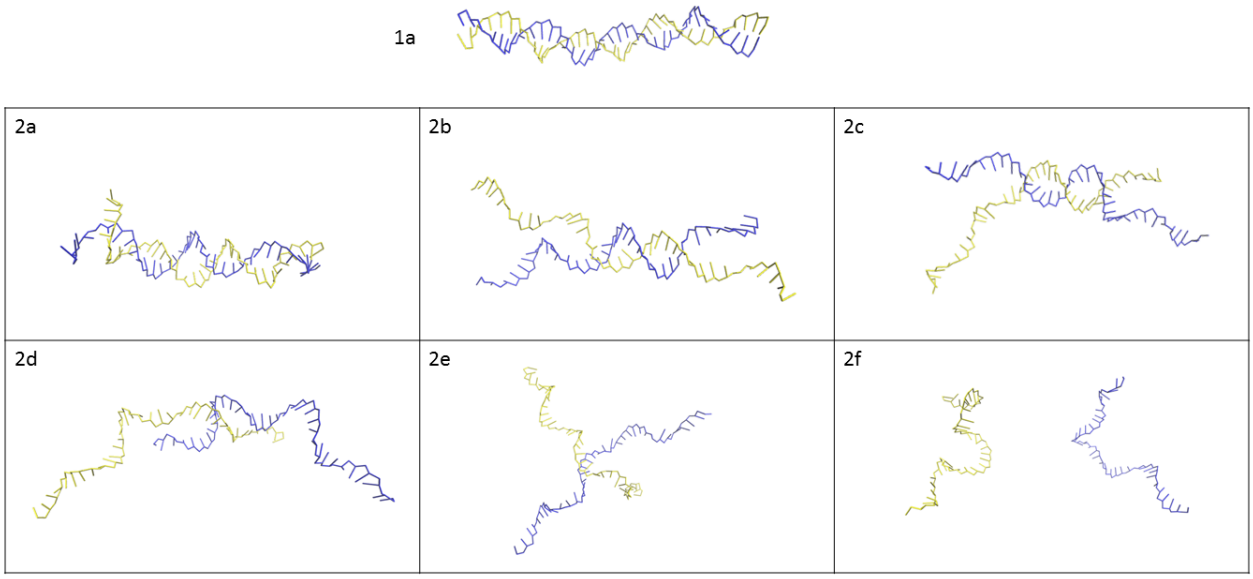

Figure 8 - Steps in melting process of DNA duplex. 1a – a fully hybridized duplex. 2a – a bubble in the center and melted end. 2b – both ends melted. 2c – increased melting and shifting. 2d – strands "crawled" almost completed free of each other. 2e – contact at only one point. 2f – complete separation.

The general experimental approach can be summarized using the following steps.

1) Simulate, using replica exchange method, DNA in the bulk (the control) and on the surface in various configurations (the treatments).
2) Calculate the heat capacity of the system in each case using WHAM and then determine the melting temperatures.

3) Compare the melting temperatures from each treatment to the control to analyze the effect of the surface configurations on stability.

In these simulations, 36 base pair and 51 base pair segments of the Insulin DNA sequences are used. Below are listed the exact sequences used.

36 base pair

Probe CCAGGTGTGAGCCGCACAGGTGTTGGTTCACAAAGG
Target GGTCCACACTCGGCGTGTCCACAACCAAGTGTTTCC

51 base pair

Probe CCAGGTGTGAGCCGCACAGGTGTTGGTTCACAAAGGCTGCGGCTGGGTCAG
Target GGTCCACACTCGGCGTGTCCACAACCAAGTGTTTCCGACGCCGACCCAGTC

In studies of mismatches the 36 base pair sequence was used as a basis.

A key finding of previous research[14] into surface hybridization that will be referred to throughout this study is the idea of surface stabilization. That is, a hybridized duplex attached to a surface is more stable than the same hybridized duplex in a bulk solution. This finding was contrary to previous assumptions, that a surface would destabilize hybridization. This research was designed in part to further develop that research and determine more details. Mismatches between hybridized DNA are expected to destabilize hybridization, and the exact effects of more complex interaction with a surface were unknown at the onset of this research. This study compares all results to surface stabilization effects, and will elaborate on changes to stability from the bulk and simple surface case.

3 TETHER STUDY

3.1 Introduction

The ability to bind ssDNA probes to specific locations on a surface is critical to the functionality of many DNA based technologies. However, as discussed previously, the presence of the surface also influences the behavior and properties of DNA. The binding can be accomplished through multiple means: adsorption to the surface, physical tethers, and others. Regardless of the method, the configurations of the probes on the surface are the dominant factor affecting the ability of the target to recognize the probes.

As mentioned previously, simulation research to date has focused on the DNA being tethered to the surface at only one end. This is an idealized case, and the research described in this section seeks to expand current understanding to the more realistic case that occurs when the probes are bound to the surface at multiple locations. These non-ideal effects restrict mobility, increase proximity to the surface, impose strain, and limit orientation to the surface.

Efforts to understand multiple binding to the surface focused on tethers. Multiple tethers were used to bind the probe to the surface. As mentioned above, tethering is not the only means by which non-idealities are manifest, but they serve as a good model to approximate many cases. For example, a probe strand of DNA on the surface of a microarray could be tangled with another strand and thereby held parallel to the surface rather than perpendicular. These tether

studies isolate the specific effects of being held near the surface from the many complex effects of strands tangling.

The studies included looking at longer strands of DNA (See Figure 9) as well as longer tethers. 3-tether configurations are examined: 1 tether at the end of a probe strand (the ideal case), 2-tethers (one at each end of the probe strand), and 3-tethers (one at each end and one in the middle of the probe strand). These cases are illustrated in Figure 10.

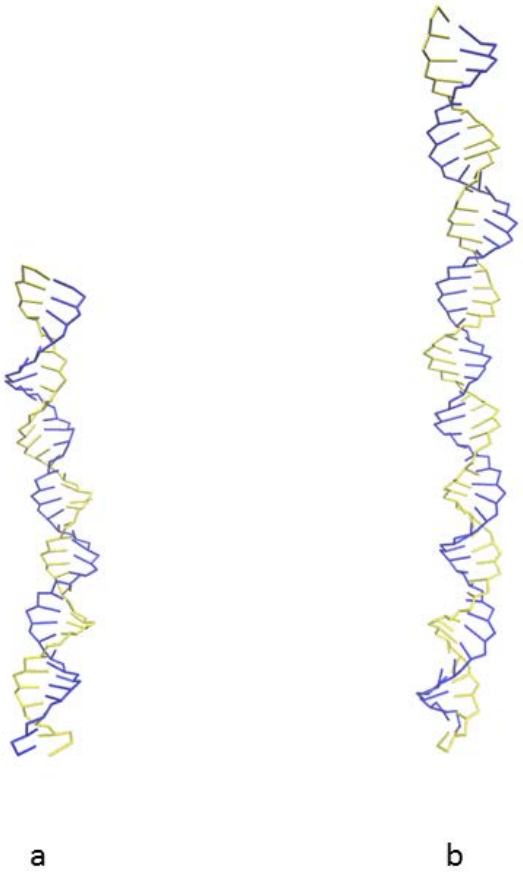

Figure 9 - 36 base pair (a) and 51 base pair (b) duplexes.

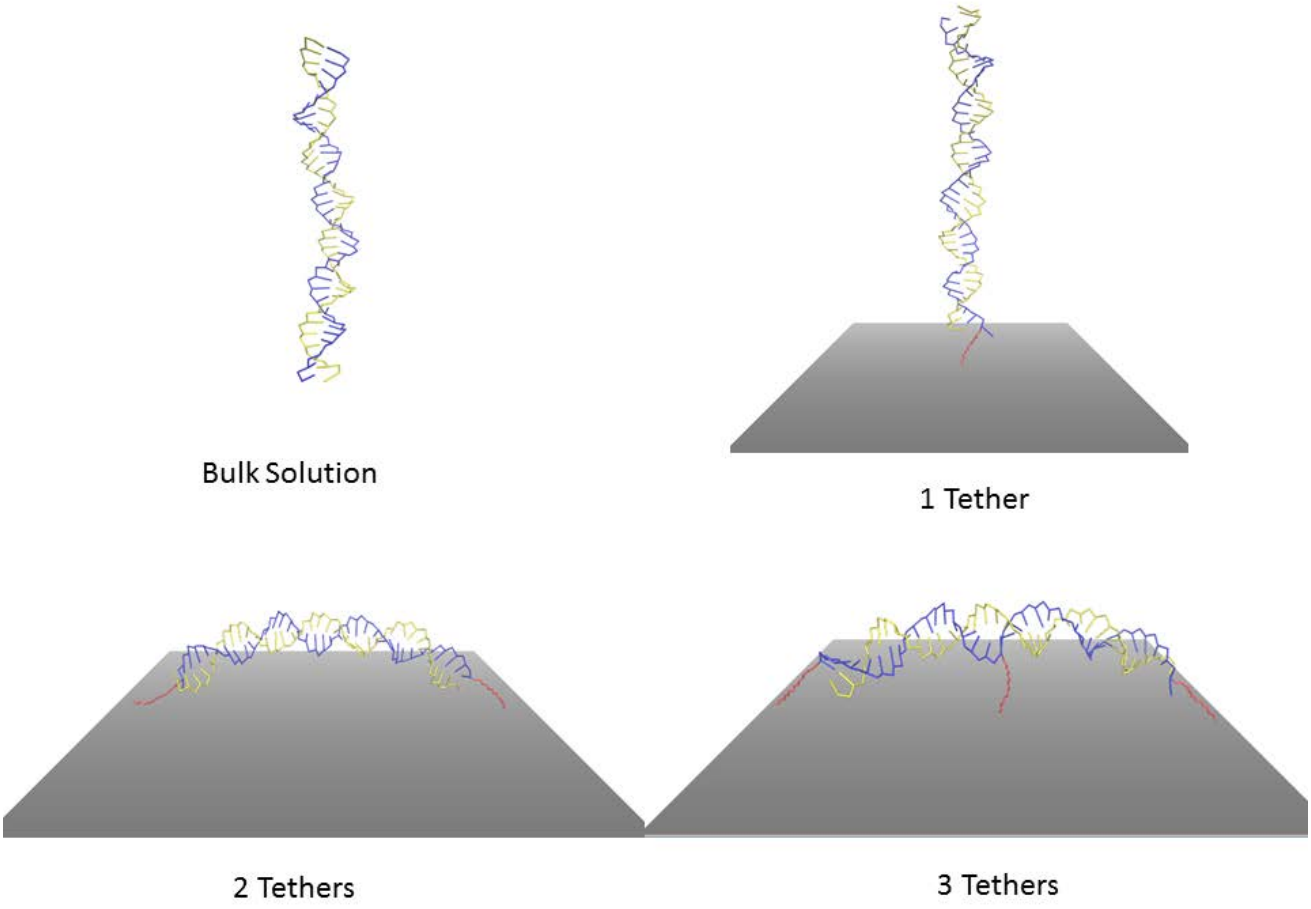

Figure 10 - Duplex in bulk solution compared to 3-tether configurations.

The single tether case shows the specific effects on hybridization caused by the presence of a surface. Parallel alignment to the surface would likely be caused by tangling with another probe or non-specific binding to the surface, simulating this with an additional tether allows the effects of the surface to be viewed separately from the effects of the second probe. Similarly, 3-tethers on the probe strand further restrict the mobility of the probe.

Two lengths of tethers are also examined. Previous studies have shown that tethering to a surface increases hybridization stability[14]. The hypothesis tested in this regard is that a longer

tether allows a tethered probe to approach bulk solution hybridization situation. The effects could be complicated. For example, as a single tether probe moves away from the surface, the stability of the duplex is expected to go down. However, for a probe tethered in multiple locations, where the inhibition of configurations is expected to destabilize the duplex, moving towards the bulk may give the system more flexibility and improve stability. The actual results are discussed below.

As described in Table 1 below, simulations were performed with 36 base pair strands with 1, 2, and 3-tethers. The 1 tether case was simulated with tethers of 15 links and 30 links, all other cases with 36 base pairs were performed with 15 links. Simulations with 51 base pair strands were performed with 1 tether with 15 and 30 links.

Table 1 – Design variables in Tether Study.

Relation to Surface	Length of Tether (# of sites)	Length of DNA (# of Base Pairs)
Bulk	N/A	36 Base Pairs 51 Base Pairs
1-tether	15	36 Base Pairs 51 Base Pairs
	30	36 Base Pairs 51 Base Pairs
2-tether	15	36 Base Pairs
3-tethers	15	36 Base Pairs

3.2 Methods

3.2.1 Experimental Design

The effects of the surface on hybridization stability are quantified through melting temperature. That is, the temperature at which two hybridized strands separate. A higher melting

temperature indicates higher stability. As explained previously, melting temperature is in this research is determined through replica exchange simulations.

In the tether studies, replica exchange was performed over a range of 250 K at 10 K increments. The exact range sampled in each case was shifted to center approximately on the melting temperature. This shifting ensured accurate sampling and calculation of heat capacity. A range of 250 K was found to enable the heat capacity calculation to achieve high accuracy. Narrower temperature ranges were attempted in early simulations, but resulted in broad peaks and therefore less confidence in the calculated values of heat capacity and melting temperature.

3.2.2 Simulation Protocols

Previous work featured simulations where the probe strand is attached to the surface at only one end. Tethering a probe strand to the surface at multiple locations simulated the effects of a strand being held near the surface by tangling with other probes or adsorption onto the surface. The difference in stability between a strand in bulk, a strand tethered at one end, the strand tethered at two ends, and the strand tethered in the middle and at the two ends was compared. These were compared to a perfectly matched pair.

The first step of the simulation process is generating input files. These files provide the program with vital data about the system to be simulated. This includes DNA length and sequence, initial position of all sites, and notation of interacting sites. The initial position of all sites in the system is also used as a reference when determining the energy of the system and the value of forces acting between sites. For tethered cases, data on the tethered sites was also provided.

While input file generation may seem straight forward, careful calculation of tether positions relative to the DNA strands was required to ensure accurate calculation of system energy. With a hybridized pair as the starting point, tethers of the appropriate length and in the desired numbers were added to the probe strand. Once tethers were in place, forces were applied in simulation to "pull" the tether into position to attach to the surface. The simulations were then allowed to come to equilibrium to remove any biases in the data that may have been introduced by pulling the probe to the surface.

A time step of 1 fs was used in these simulations. This was chosen in order to ensure that energy is conserved during integration. Previous analysis[26] showed that time steps as large as 10 fs could be used, but 1 fs was chosen for this study to ensure accuracy even over long simulation times. Simulations were run for approximately 5 to 10 million time steps. Temperature values were used between 270K to a peak of 540K, though each replica exchange set used only a range of 240K within the stated values. With these conditions, a single replica exchange set ran for 1 to 4 hours.

As mention in Chapter 2, the data produced by the simulations were analyzed using the WHAM method to obtain heat capacity curves and melting temperatures. Five replicates were performed for each case and an average melting temperature was determined.

Simulations were also performed with two different lengths of tethers. Comparing the results of these two situations is extremely valuable in further determining the effects of close proximity to the surface. Especially significant will be the ability to better separate the effects of stability of close proximity and orientation from the effects of reduced mobility and increased strain. Longer tethers were expected to decrease any direct effects of the surface.

In single tether cases, the hybridized duplex stays near perpendicular to the surface, so no large strain is imposed on the tether. In the 2-tether cases when both tethers are attached to the surface some strain is introduced into the tethers (See Figure 11). The hybridized pair is also held in a mildly strained position, which is a significant and accurate reproduction of conditions that would be seen *in situ*. As discussed above, the 2-tether case is approximating the DNA probe being held in close proximity to the surface, either through tangling with another probe or through some other interaction with the surface such as direct binding. Observing additional strain on the DNA in those cases is not only reasonable, it would be expected.

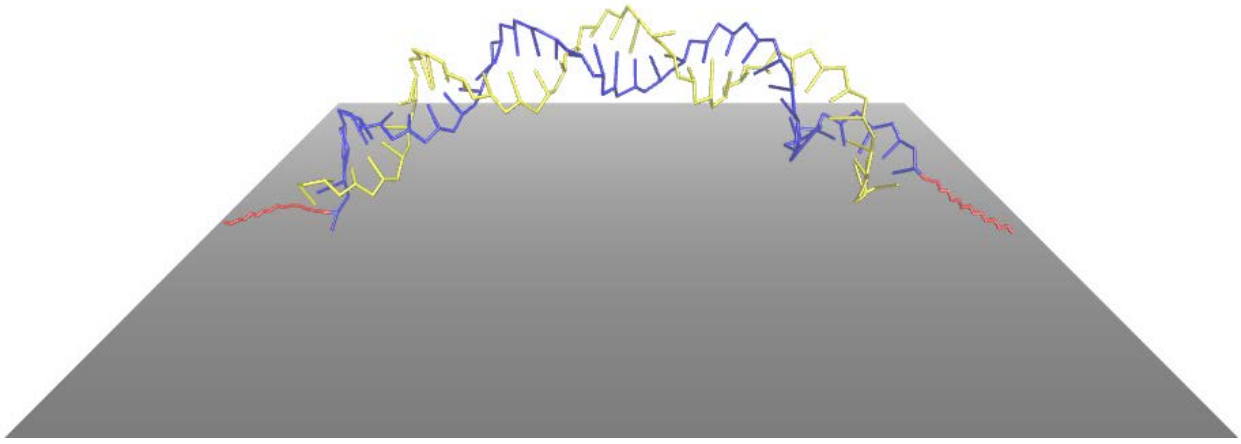

Figure 11 - Illustration of strain imposed on DNA when tethered to a surface.

In the 3-tether case, the two end tethers were formed in precisely the same manner as the 2-tether case. The third tether was placed on a central site, attached to a sugar. The 3-tether cases also introduced additional strain into the system. This is also reasonable as a representation of actual behavior, as in the 2-tether cases. *It was expected that this strain, as well as the limited motion and close proximity to the surface would destabilized hybridization.* As mentioned

previously, the length of the tethers was also investigated. *It was hypothesized that longer tethers would increase hybridization stability, bringing the melting temperature closer to bulk values.*

In summary, three major hypotheses were tested:

1) Tethering the probe to the surface at multiple locations will decrease stability.
2) Increasing tether length will decrease stability.
3) Longer segments of DNA will experience smaller changes in stability due to their increased flexibility.

These hypotheses were chosen as understanding the stabilizing effects of the surface is critical to design of functional and consistent DNA microarrays. If increasing tether length decreases stability, then that would be a viable method for researchers to implement to bring microarray behavior closer in line with known bulk behavior. If tethering at multiple locations is found to decrease stability, this would have significant implications for manufacturing methods that are likely to result in high degrees of surface interaction, such as the spotted method. This result would confirm that these methods do not contribute to microarray inconsistency. If longer strands of DNA help obtain more consistent results, this design feature would be easy to implement.

3.3 Results and Discussion

3.3.1 1-tether

Table 2 contains the melting temperatures of the duplexes of 36 and 51 base pairs in the bulk and tethered to the surface at one location. The melting temperature was lowest for the 36 base pair duplex in the bulk and highest for the 51 base pair duplex on the surface. For the 36

base pair cases, the tethered case has a melting temperature that is ~36K higher than the bulk case. For the 51 base pair cases, the tethered case is only 7K higher than the bulk case.

A single tether was found to increase stability for the tested cases compared to hybridization in the bulk solution. This is in agreement with previous studies performed on 21 base pair DNA strands. Comparing the 36 and 51 base pair strands shows an interesting trend. The absolute difference in stability between the bulk cases and the corresponding single tether cases is much smaller for the 51 base pair case than the 36 base pair case. In both cases, tethering the DNA to the surface increases stability, but the effect is greater for the 36 base pair case. The reason for this difference is likely the increased flexibility of the longer DNA strand: the portion furthest from the surface has much the same ability to move as a strand in the bulk. The section of the strand that is not connected to the surface "feels" less of the effect of the surface. Any stabilizing effect the surface has on the strand would likely be diminished for sections of the strand that are further from it.

Table 2 – Comparison of Bulk and 1 Tether cases. Melting temperatures and errors. The tethered cases use 15 links in the tethers. Higher melting temperatures indicate higher stability and stronger bonding.

	T_{melt} (K)	Error
Bulk Complementary, 36bp	427.9	3.9
1 Tether Complementary, 36bp	463.6	3.5
Bulk Complementary, 51bp	460.5	2.9
1 Tether Complementary, 51bp	467.5	5.2

3.3.2 2-tethers

Table 3 compares the results of studies with 2-tethers to bulk and single tether cases. The 2-tether case has a melting temperature that is ~34K higher than the bulk case and 1.6K lower than the 1 tether case. The difference between the melting temperatures of the 1 and 2-tether cases is within statistical error suggesting that increasing the number of tethers to two does not drastically affect stability of the duplex.

Simulations with a tether at both ends of the probe strand yielded unexpected results. It was theorized before the simulations were performed that the close proximity to the surface and additional restriction to mobility would destabilize hybridization. Instead, the results of 2-tether studies show statistically equal binding strength with single tether studies. The additional restriction in mobility and more direct interaction with the surface did not appreciably change the overall stability. However, as described above, it was not known what features of the tethered case resulted in these effects. Therefore, additional conditions were tested in order to isolate the effects of features such as proximity to the surface and restriction of mobility.

3.3.3 3-tethers

The melting temperature of the three-tether configuration, as shown in Table 3 is by far the highest. It is ~112K higher than the bulk value, ~76K higher than one tether, and ~78K higher than 2-tether. The error for the three-tether case is much higher than the other treatment, but the difference is still statistically significant.

Table 3 – Comparison of 3-tether cases: Melting temperatures and errors. The tethered cases use 15 links in the tethers. Higher melting temperatures indicate higher stability and stronger bonding.

	T_{melt} (K)	Error
Bulk Complementary, 36bp	427.9	3.9
1 Tether Complementary, 36bp	463.6	3.5
2-tether Complementary, 36bp	462.0	5.7
3-tether Complementary, 36bp*	539.7	15.6

*This value was averaged from only 4 replicates

Three-tether configurations of DNA were also expected initially to weaken hybridization, especially due to the much more severe restriction in mobility caused by the presence of the tether attached to the center of the probe strand. The lengths of the tethers are the same as in the two-tether case, so changes any changes in stability are not related to the distance of the probe from the surface but to the restricted configuration of the surface.

To help understand this unexpected result of tethers increasing stability, a visual inspection of the images produced by the model was performed. This inspection showed that the additional restriction in motion of the hybridized pair made it common for the target strand of DNA to become trapped between the probe strand and the surface even when melting was almost complete. The restricted physical orientation of the probe with the surface prevents melting from occurring as easily as in the bulk case. In essence, once the target strand is close to the probe strand, it is difficult for the target to leave again. These results run counter to the expected outcome: restriction of mobility and introduction of additional strain would have destabilizing effect on hybridization.

It appears that the driving force of the stabilization is the ability of the surface to restrict the movement of the target away from the probe once it is there. This effect must dominate the decrease in the ability of probe to orient. In other words, the dominant force is the keeping of the target near the probe once it is there. It is also important to consider that having a target hybridize with this highly restricted case would be much more difficult.

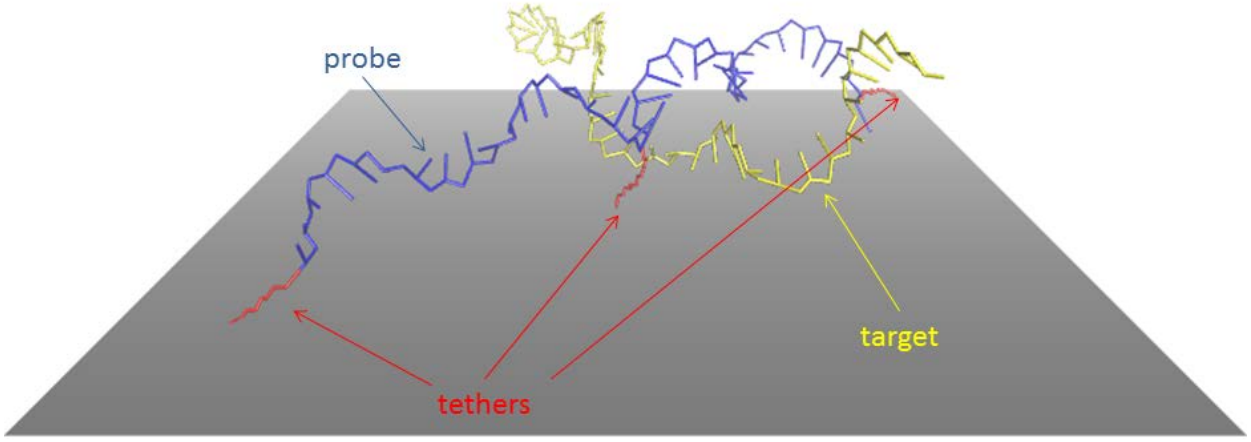

Figure 12 - Image of target strand trapped by probe with 3-tethers.

3.3.4 Tether Length Effects

Tether lengths used in many of the simulation discussed in Chapter 2 were 15 segments long. This same length was used to produce the results explained just above. In order to determine how tether length affected the stability, studies were also performed with tethers of 30 segments (see Figure 13). As stated earlier, the longer tethers were expected to reduce hybridization stability compared to shorter tethers, bringing the strength closer to bulk values. It

was also hoped that the results would help clarify the effects of close proximity to the surface and restriction in motion.

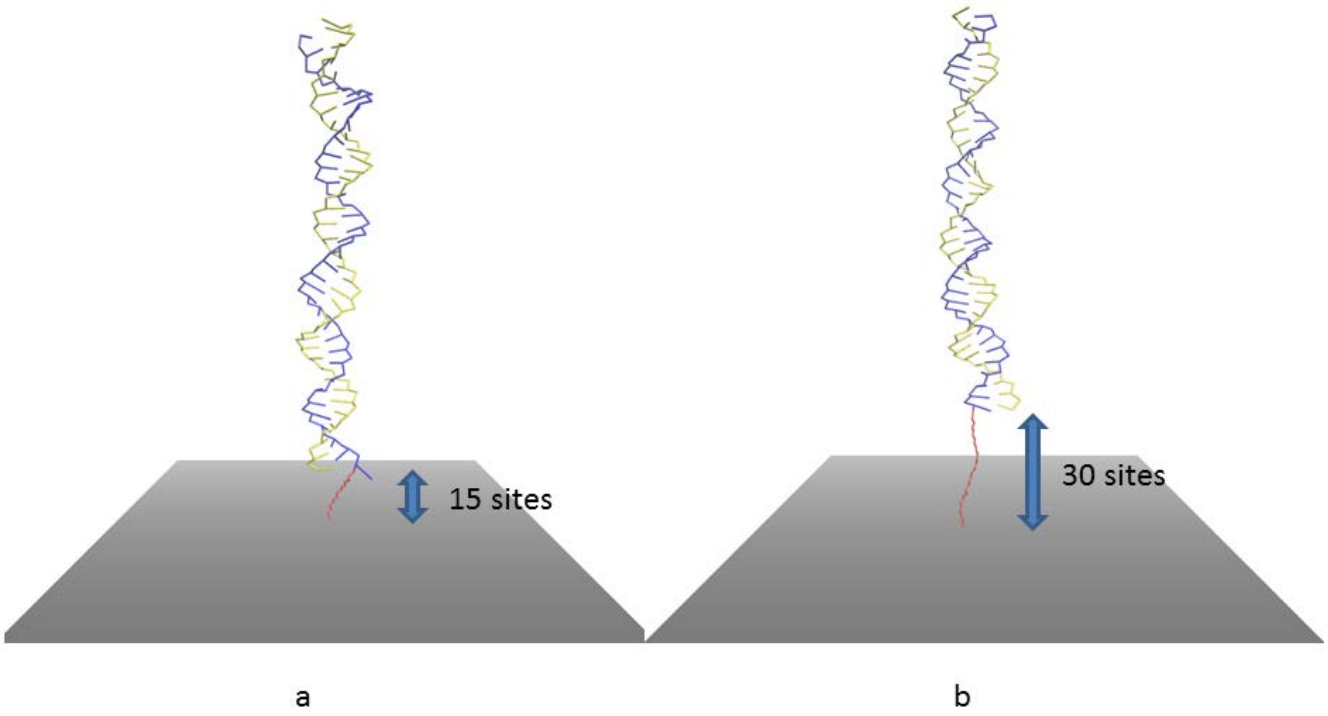

Figure 13 - 2-tether lengths. 15 sites (a) and 30 sites (b).

Table 4 shows the comparison of the melting temperatures for the 36 and 51 base pair cases with long and short tethers. The melting temperatures for the longer tethers are statistically identical to that of the short tether. The long tether in the 36 base pair case is only 1.3K higher in melting temperature than the short tether case. In the 51 base pair case, the melting temperature of the longer tether case is 3.2K lower than the short tether case.

Table 4 – Comparison of 2-tether lengths. Short tethers were 15 segments in length, long tethers were 30 segments.

	T_{melt} (K)	Error
Bulk Complementary, 36bp	427.9	3.9
1 Tether Complementary, 36bp, Short Tether	463.6	3.5
1 Tether Complementary, 36bp, Long Tether	464.9	2.8
Bulk Complementary, 51bp	460.5	2.9
1 Tether Complementary, 51bp, Short Tether	467.5	5.2
1 Tether Complementary, 51bp, Long Tether	464.3	1.7

These results are contrary to expectations, as longer tethers were predicted to yield melting temperatures more similar to bulk values. Instead, within the experimental errors, the differences between long and short tether cases are insignificant. This result suggests that the proximity to the surface isn't as important as the mere presence of the surface and its ability to keep the target and probe closer together than they would be in the bulk.

3.3.5 Conclusions

The purpose of this chapter was to understand how attaching the probe to the surface at multiple locations would affect the stability of duplex. Such a situation more accurately represents the experimental system compared to the 1 tether, ideal case. Simulations were done with the probe strand tethered at one, two, and three locations. It was hypothesized that two or 3-tethers would destabilize the duplexes, and that longer tethers would bring melting temperature

values closer to bulk values. The results ran contrary to expectations in many cases. For example the inclusion of a second tether did not affect stability, and the inclusion of the third tether made melting more difficult, stabilizing the duplex.

These results are significant to researchers using both simulation and experimental methods. Interaction with the surface was uniformly found to increase stability. Simple attachment to the surface, regardless of the exact way it was done, was found to increase melting temperature. In the case with the most severely restricted mobility the surface was found to strongly inhibit complete melting. By analogy, it is likely that such restrictive interaction with the surface would also make initial hybridization difficult.

The results of the longer tether studies confirm the stabilizing effects of the surface compared to the bulk. That is, even with longer tether simulations, attachment to the surface still made the duplex much more stable compared to the bulk. Since the binding strength does not approach the behavior of bulk DNA compared to shorter tethers, experimental researchers would not be able to simply use longer tethers to make microarrays perform more as expected based on bulk experiments.

Simulations with one additional tether holding the DNA probe strand in close proximity to the surface had a very small effect on hybridization stability compared to the addition of a single tether. This result is encouraging for future simulation and experimental work. This suggests that the critical behavior of a microarray system can be accurately accounted for with the simple and more idealized system with a probe held to the surface by a single tether. It also suggests that the experimental systems, where the probes are likely bound to the surface at multiple locations, are not adversely affected by the increased interaction with the surface.

Finally, the results seen from the 3-tether study show that small differences in interaction with the surface can cause large changes in behavior. The orientation to the surface and proximity did not change between the 2- and 3-tether cases, but the addition of another interaction site did make melting much more difficult.

4 MISMATCHES

4.1 Introduction

A key aspect of tools and devices that rely on DNA hybridization is the sensitivity to mismatches. The complementary base pairs are A-T and C-G; any other attempted alignment of bases constitutes a mismatch. These complementary base pairs, the so-called Watson-Crick base pairs result in the DNA double helix. Specifically, two strands of DNA form a hybridized pair if each of their sequences of base pairs are complementary. It is the binding ability of complementary bases that lead to hybridization.

In microarrays, the sample of DNA to be tested will have many target strands with many different sequences. Accurate analysis of a sample requires that only perfectly complementary target and probe strands hybridize. A difficulty in using microarrays is that the presence of a surface increases binding stability of both matched and mismatched sequences. Previous studies have shown that the stability of the DNA duplex is weakened by even a single mismatch, but that this effect is less significant than the stabilizing effect of the surface as a whole. A complementary pair of DNA sequences is compared to a pair with a mismatched section in Figure 14 below.

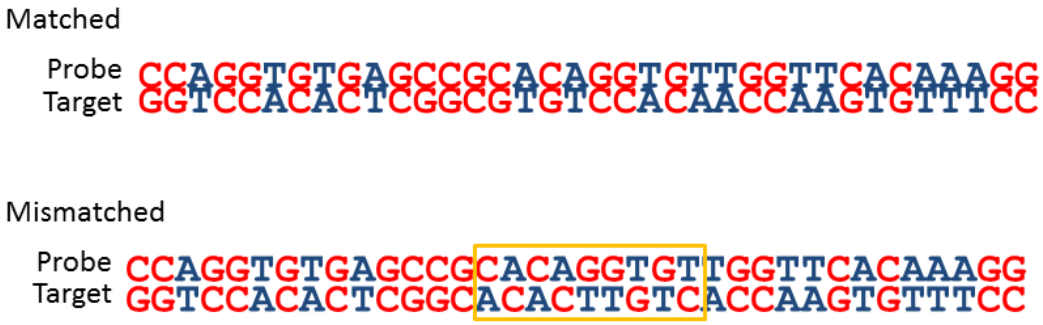

Figure 14 - Illustration of mismatches between two strands.

This research expands on past work to look at multiple mismatches in multiple locations. The past research only examined a single mismatch. The goal of this study is to determine how the location and number of mismatches affect the stability of a duplex and the fidelity with which correctly matched sequence can displace mismatched molecules. This understanding is critical to effective design of microarrays, because the accurate analysis of a DNA sample for specific genes hinges on the sensitivity to mismatches.

It is expected that greater numbers of mismatches will result in lower melting temperature. What is unknown is the exact effect of the surface on hybridization with mismatches present. It is expected, based on previous research, that small numbers of mismatches will not significantly reduce the melting temperature of a duplex. That is, a hybridized pair on the surface with a small number of mismatches will still have a higher melting temperature than a perfectly matched pair in the bulk.

It was theorized that for a given number of mismatches, the absolute difference in stability between the complementary case and the mismatched case would be the same for

surface cases and bulk cases. In other words, though surface cases have higher melting temperatures than bulk cases, introducing X number of mismatches would lower the melting temperature of the surface and bulk case by the same amount. If this were found to be the case, then designing microarrays based on surface melting temperatures would be relatively simple if bulk melting temperatures are known. However, it was unknown if this would be the case, or if the surface would cause some other unexpected behavior.

4.2 Methods

A 36bp segment of Insulin DNA was used as the basis for the mismatch studies. Separate simulations were performed for numbers of mismatches varying from one to twelve. The locations of mismatches were also varied. As in the tether studies, mismatch effects were studied using replica exchange to calculate melting temperature and overall stability.

The effects of mismatches on bulk hybridization were well understood prior to this study. However, it was unknown how many mismatches within a hybridized pair would be necessary to have a measurable effect on melting temperature for surface bonded DNA. Therefore, preliminary studies were performed with 1, 2, 3, 6, 9, and 12 mismatches. The preliminary results showed how many mismatches were required to get statistically significant changes in melting temperature. Once optimal numbers of mismatches was known, more studies were performed varying the location of the mismatches.

As an example of a few of the tested mismatch cases, below the sequences are listed. The sequences of the perfectly matched 36 base pair strands are:

Matched

With 9 mismatches in the center, the sequence becomes:

Mismatched

With 9 mismatches on one end, the sequence becomes:

Mismatched

Table 5 below describes all the conditions simulated as a part of the mismatch study.

Table 5 - Design variables and conditions of Mismatch study.

Relation to Surface	Deviation from Match	Position of Mismatch
Bulk	Perfect match	N/A
	1 mismatch	Bulk End
	2 mismatches	Ends
	3 mismatches	End, Center
	6 mismatches	Center
	9 mismatches	Bulk End, Center, Wall End
1 Tether	Perfect match	N/A
	1 mismatch	End
	2 mismatches	Ends
	6 mismatches	Center
	9 mismatches	Bulk End, Center
	12 mismatches	Center

The type of mismatch was controlled. Type of mismatch, i.e. aligning of different types of non-complimentary base pairs, could significantly affect free energy of hybridization. The nucleotides Adenine and Guanine are larger than Thymine and Cytosine, so a mismatch of Adenine to Guanine for example would likely have greater destabilizing effect than a mismatch of Thymine to Cytosine. For consistency, only mismatches of Thymine to Guanine and Adenine to Cytosine will be implemented. This study will focus on the effects of multiple mismatches on a single strand and how their locations on the strand affect stability and melting temperature.

4.3 Results and Discussion

4.3.1 Number of Mismatches

The same simulation protocols were used in these replica exchange simulations as in the simulations for the tether studies. The effect on melting temperature of the studies with 1, 2, and 3 mismatches are shown in Table 6. Column 1 describes the conditions that were simulated. Column 2 gives the melting temperature and column 3 gives the standard error of the melting temperature. For the bulk simulations, the 1 mismatch and 2 mismatch cases were both about 6K higher than the complementary case. Similarly, the cases with 3 mismatches in the bulk were ~8K. However, these differences are within the statistical noise.

In all cases, the 1 tether cases were more stable than any of the bulk cases. In the 1 tether simulations the 1 mismatch melting temperature was 2.7K lower than the complementary case. Once again this difference is not statistically significant compared to the error. The 2 mismatch case is 12.7K lower than the complementary case, and this difference appears to be significant. However, the small differences in melting temperature suggest that the presence and location of a small number of mismatches only minorly affect stability.

Table 6 – Comparison of small numbers of mismatches. Melting temperatures and errors. The tethered cases use 15 links in the tethers.

	T_{melt} (K)	Error
Bulk Complementary	427.9	3.9
Bulk, 1 Mismatch, End	434.0	1.5
Bulk 2 Mismatches, Ends	434.2	5.2
Bulk 3 Mismatches, End	436.3	4.6
Bulk 3 Mismatches, Center	435.9	3.4
1 Tether Complementary	463.6	3.5
1 Tether, 1 Mismatch	460.9	2.6
1 Tether, 2 Mismatches	450.9	4.0

Table 7 shows the melting temperature data for mismatches of 6, 9, and 12 base pairs found in the center of a strand. Increasing mismatches decreased the melting temperature significantly. The most significant decrease in melting temperature was found with 9 mismatches for both the bulk and single tether case. In the bulk, the 6 mismatch melting temperature was 6.6K lower than the complementary case. The 9 mismatch case was 23.5K lower than the complementary, a difference that is statistically significant.

In the single tether cases, each melting temperature was greater than the bulk complementary case, even when increased to 12 mismatches. Compared to the single tether complementary case, the 6 mismatch case was 12.2K lower, the 9 mismatch case was 29.3K lower, and the 12 mismatch case was 31.9K lower.

Table 7 - Comparison of larger numbers of mismatches. Melting temperatures and errors. The tethered cases use 15 links in the tethers.

	T_{melt} (K)	Error
Bulk Complementary	427.9	3.9
Bulk, 6 Mismatch, Center	421.3	5.0
Bulk, 9 Mismatches, Center	404.4	5.7
1 Tether Complementary	463.6	3.5
1 Tether, 6 Mismatches, Center	451.4	7.4
1 Tether, 9 Mismatches, Center	434.3	4.7
1 Tether, 12 Mismatches, Center	431.7	5.0

4.3.2 Location of Mismatch

The location of the mismatch was found to have a stronger effect on hybridization than expected. Nine mismatches were further studied by simulating the cluster of mismatches at either end of the strands of DNA. Comparing the three locations simulated with 9 mismatches (Table 8), the center was the least stable being 16.2K lower than the lowest end case for the bulk and 22.6K lower than the end case for the single tether. In both of the bulk cases a cluster of 9 mismatches on the end did not have a statistically significant effect. When located in the center, the melting temperature was decreased by ~20K.

In the surface case, a cluster of 9 mismatches on the end also had a small effect, but a cluster in the center decreased the melting temperature by ~30K. Comparing the bulk cases to the single tether case for 9 mismatches at the mismatch positions studied suggests that mismatches may have a slightly larger effect on surface tethered cases than in the bulk. The decrease in

melting temperature for the bulk end cluster and the center cluster is larger for the tethered strands than the bulk.

Table 8 - Comparison of locations of mismatches. Melting temperatures and errors. The tethered cases use 15 links in the tethers.

	T_{melt} (K)	Error
Bulk Complementary	427.9	3.9
Bulk, 9 Mismatch, Bulk End	426.3	3.5
Bulk, 9 Mismatches, Center	404.4	5.7
Bulk, 9 Mismatches, Wall End	420.6	2.7
1 Tether Complementary	463.6	3.5
1 Tether, 9 Mismatches, Bulk End	456.9	3.2
1 Tether, 9 Mismatches, Center	434.3	4.7

In summary, the significant results are:

- The center is the most destabilizing location for mismatches to occur.
- For a 36 base pair strand, a cluster of 9 mismatches caused the most significant destabilization.
- Tethered cases seem to be slightly more sensitive to mismatches than bulk cases.

4.3.3 Conclusions

In agreement with previous studies, hybridization on the surface is more stable than in the bulk for both perfectly complementary strands and mismatched strands. Also as expected,

mismatches of base pairs between strands of DNA decrease stability compared to perfectly complementary strands. This comparison of the stabilizing effects of the surface to the destabilizing effect of mismatches is extremely valuable in determining the sensitivity and accurate responsiveness of microarrays. For the 36 base pair simulations, it required 9 mismatches (in the center of a hybridized pair) to effectively counter the stabilizing effects of the surface, bringing it to approximately the same melting temperature as the perfectly matched case in the bulk. Microarrays are designed based on the melting temperatures of DNA as measured in the bulk with the assumption that bulk and surface melting temperatures are the same. If the effects of the surface are not accounted for, a highly mismatched sequence of DNA might register as matching because the surface stabilizes both mismatched and complementary sequences. Thus, this research helps illustrate the potential magnitude of the error possible in microarray analysis.

The presence of mismatches serves to destabilize hybridization through two effects. First, there is the simple decrease in the number of complementary bases. Fewer bonds results in lower binding energy. Second, an important step in the separation of two hybridized strands is the formation of a bubble, as discussed previously. Mismatches, particularly when in large clusters, act as catalysts for bubble formation even at low temperatures. This leads to greater mobility and flexibility in the unmatched sections of the strands which consequently leads to more tugging and pulling on the bonded pairs near the mismatched sections.

In both bulk and tethered cases, partial melting is observed at the ends of the DNA strands even far below the melting temperature. In the perfectly matched cases simulated for tether studies and discussed earlier, “bubbles” in the hybridized pair of significant size only appeared near the melting temperature. In the mismatch cases, especially the simulations of

larger clusters of mismatches, “bubbles” or sections of un-hybridized DNA were essentially always present. The mismatched sections of the DNA strands did not experience the binding of complementary bases. This lack of bonding was apparent on inspection of the generated images, as illustrated in Figure 15, and served to verify the correct implementation of the desired design.

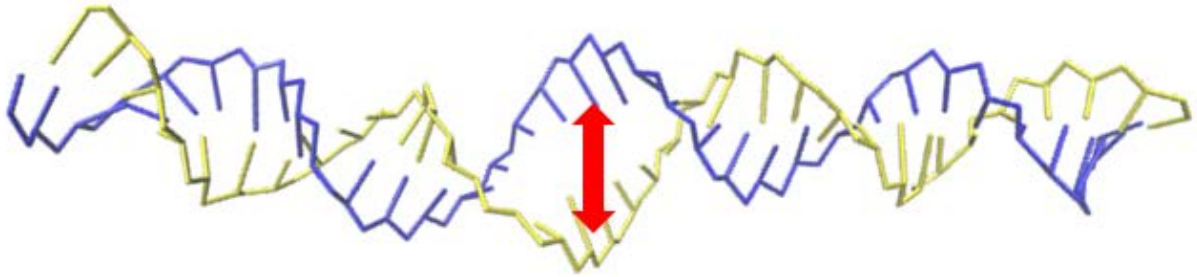

Figure 15 - Image of bubble formation in case with 9 mismatches in the center of a duplex at 300K.

In general the ends of the strands appear to be the weakest section of the molecule, which helps explain why mismatches at the ends of strands have a relatively small effect. The overall strength of hybridization is not as reliant on the ends as it is on the center. When a significant cluster of mismatches is located in the center of strands of DNA, there is effectively a permanent bubble in what should be the mostly strongly bound section.

5 CONCLUSIONS

5.1 Summary of Results

This study was intended to provide a deeper and more complete understanding of the effects a surface has on DNA hybridization. Two features of non-ideal behavior were examined: tether configuration and degree and location of mismatch. The specific hypotheses tested were:

1) Tethering the probe to the surface at multiple locations will decrease stability.
2) Increasing tether length will decrease stability.
3) Longer segments of DNA will experience smaller changes in stability due to their increased flexibility.
4) For a given number of mismatches, the absolute difference in stability between the complementary case and the mismatched case would be the same for surface cases and bulk cases.
5) Location of mismatch would have only a small effect on overall stability.

The results, as now described, ran contrary to expectation.

1) Increasing the number of tethers from 1 or 2 to 3 increased stability substantially.
2) Increasing tether length had no statistically significant effect.
3) Tethered cases appeared slightly more sensitive to mismatches than bulk cases.

4) A duplex with mismatches in the center was significantly less stable than duplexes with mismatches on the ends.

5.1.1 Tethers

This study found that tethering DNA to a surface greatly increases the melting temperature and therefore the stability of a duplex. The increase in melting temperature was approximately the same in cases where only one end was tethered to the surface or if both ends were tethered. Also, a shorter and longer tether resulted in approximately the same melting temperature. When a third tether attached the center of the strand to the surface, melting temperature increased drastically. These results suggest that simulations of microarrays accurately predict behavior with only a single tether if the degree of interaction with the surface is known to be low. However, in cases of very high degrees of interaction with the surface, the highly constrained mobility with respect to the surface is expected to alter behavior and thermodynamics considerably.

Most significantly, it was found with the 3-tether case that complete melting is inhibited by the target becoming trapped between the highly constrained probe and the surface. This effectively stabilized the hybridization to a much greater degree than simple attachment to a surface by one or 2-tethers. The variability in the measured melting temperature for the 3-tether case was very high, suggesting that this high degree of interaction with the surface should be avoided to gather accurate data from DNA microarrays.

5.1.2 Mismatches

This research also showed that for the 36 base pair strands studied, a cluster of mismatches comprising 25% of the total sequence was required in order to bring the melting temperature of tethered DNA down to the approximately the same value as a perfectly matched pair in the bulk. Even this high percentage of mismatch was only sufficiently destabilizing if the cluster is located in the center of the strand. The simulations performed show that the location of mismatches is extremely significant to stability, and this is likely due to the destabilizing effect of a "bubble" of non-hybridized DNA in the center of the strand. This bubble is present at temperatures far below the melting temperature and allows the target to move much more easily.

This additional mobility, especially when located in the center of the strand, makes melting much easier. This center bubble counters the most significant stabilizing effect of the surface: the ability to force the target and probe to stay close together. The observation of a slightly greater destabilizing effect from mismatches in tethered cases than bulk cases supports this idea that mismatches directly counter the major cause of surface stabilization.

5.2 Implications of Results

This study has numerous implications in the design of Microarrays and other DNA based technology. The binding to the microarray surface, as shown previously and confirmed in this study, results in much higher melting temperatures than would be predicted from bulk DNA. Longer tethers are not recommended as a means to bring melting temperatures of a microarray closer to bulk values. Tethered DNA appears to be slightly more sensitive to mismatches than DNA located in the bulk. However, the melting temperature of tethered DNA is still much greater than bulk DNA even with larger numbers of mismatches.

The direct comparison completed in this research of melting temperature effects from tethers and mismatches gives researchers using DNA based technology a more concrete understanding of the effect of a surface on hybridization stability. It further emphasizes the need to design microarrays based on surface hybridization rather than behavior in the bulk solution. As discussed earlier, microarrays are designed with the intention that each dot on the chip should have the same melting temperature. If this is successfully accomplished, then washing steps only remove non-complementary sequence and keep perfectly-matched duplexes intact. If each dot has a different stability then it becomes likely that false positives and negatives will result as the analysis method will not be able to consistently remove mismatched strands and leave matched pairs intact.

The current procedures to design microarray probes use correlations[30,31,32,33,34,35] to predict probe/target melting temperatures, but the correlations were developed from experimental data done in the bulk. This research shows that if the complications of a surface are not accounted for in design then achieving the same melting temperature across the whole chip would be almost impossible. Developing new correlations that take into account the non-idealities found in the microarray environment could improve microarray performance. This research also suggests that more predictable behavior could be achieved if the interaction between the surface and DNA probes is minimized. Some potential methods to minimize detrimental interactions include the following: large separation distance between probe strands to avoid tangling, using production methods that are more controlled than drop methods, using curved surfaces to avoid tangling or limit surface interaction, or using microfluidic methods to control flow and system behavior.

These or other design strategies are likely to improve reproducibility of results if they limit interaction with the surface. This research suggests that relatively simple binding of one or even both ends and the surface gives a relatively consistent shift in melting temperature. Higher degrees of interaction, as demonstrated in the 3-tether case, resulted in much less consistent data and a much greater overall shift in melting temperature. This could help explain the highly inconsistent results found in current DNA microarray studies. The results of the 3-tether study, as stated earlier, featured a high degree of error and more complex melting behavior.

5.3 Future Work

One of the major limitations of this research is that the kinetics of the process were not considered. The changes in the kinetics of the hybridization process when moving from the bulk to the microarray environment are likely significant. For example, the kinetics of the three-tether system, discussed in Section 3.3.3, are likely significantly reduced compared to the bulk situation. The thermodynamic analysis revealed that the stability in this case was greatly increased once the two strands are close to each other due to the trapping of the target between the probe and the surface, but it is likely that the kinetics of the process are much slower to arrive at such a trapped state from the fully-separated state. Future work should address the kinetics of the process.

Understanding of the fundamental behavior of DNA based technologies would be further improved by simulation of more complex and realistic cases, such as simulation of multiple probe strands and target strands. Further increasing the length of target strands would provide potentially valuable insight into a design technique if there were a limiting distance after which hybridization may actually approach bulk values.

REFERENCES

1. Fortunel, N. O.; Otu, H. H.; Ng, H. H.; Chen, J. H.; Mu, X. Q.; Chevassut, T.; Li, X. Y.; Joseph, M.; Bailey, C.; Hatzfeld, J. A.; Hatzfeld, A.; Usta, F.; Vega, V. B.; Long, P. M.; Libermann, T. A.; Lim, B., (2003), Comment on " 'Stemness': Transcriptional profiling of embryonic and adult stem cells" and "A stem cell molecular signature" (I). *Science 302* (5644).

2. Ivanova, N. B.; Dimos, J. T.; Hackney, J. A.; Moore, K. A.; Lemischka, I. R., (2002), A stem cell molecular signature. *Blood 100* (11), 60A-61A.

3. Ramalho-Santos, M.; Yoon, S.; Matsuzaki, Y.; Mulligan, R. C.; Melton, D. A., (2002), "Stemness": Transcriptional profiling of embryonic and adult stem cells. *Science 298* (5593), 597-600.

4. Shi, L. M.; al., e., (2006), The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol. 24* (9), 1151-1161.

5. Vora, G. J.; Meador, C. E.; Anderson, G. P.; Taitt, C. R., (2008), Comparison of detection and signal amplification methods for DNA microarrays. *Mol. Cell. Probes 22* (5-6), 294-300.

6. Prytkova, T. R.; Eryazici, I.; Stepp, B.; Nguyen, S. B.; Schatz, G. C., (2010), DNA Melting in Small-Molecule-DNA-Hybrid Dimer Structures: Experimental Characterization and Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B 114* (8), 2627-2634.

7. Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J., (2011), A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J. Chem. Phys. 135* (16).

8. Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K., (2011), Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys. 134* (8).

9. Wong, K.-Y.; Montgomery Pettitt, B., (2001), A study of DNA tethered to a surface by an all-atom molecular dynamics simulation. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) 106* (3), 233-235.

10. Wong, K.-Y.; Pettitt, B. M., (2004), Orientation of DNA on a surface from simulation. *Biopolymers 73* (5), 570-578.

11. Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K., (2003), Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proceedings of the National Academy of Sciences 100* (24), 13922-13927.

12. Qamhieh, K.; Nylander, T.; Ainalem, M.-L., (2009), Analytical Model Study of Dendrimer/DNA Complexes. *Biomacromolecules 10* (7), 1720-1726.

13. Jayaraman, A.; Hall, C. K.; Genzer, J., (2006), Computer Simulation Study of Molecular Recognition in Model DNA Microarrays. *Biophysical Journal 91* (6), 2227-2236.

14. Schmitt, T. J.; Knotts, T. A., (2011), Thermodynamics of DNA hybridization on surfaces. *J. Chem. Phys. 134* (20).

15. Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J., (2009), Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *Proceedings of the National Academy of Sciences 106* (43), 18125-18130.

16. Sambriski, E. J.; Ortiz, V.; de Pablo, J. J., (2009), Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways. *J. Phys.-Condes. Matter 21* (3).

17. Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J., (2009), A Mesoscale Model of DNA and Its Renaturation. *Biophysical Journal 96* (5), 1675-1690.

18. Held, G. A.; Grinstein, G.; Tu, Y., (2003), Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl. Acad. Sci. U. S. A. 100* (13), 7575-7580.

19. Kim, D. K.; Kwon, Y. S.; Takamura, Y.; Tamiya, E., (2006), Detection of DNA hybridization properties using thermodynamic method. *Jpn. J. Appl. Phys. Part 1 - Regul. Pap. Brief Commun. Rev. Pap. 45* (1B), 509-512.

20. Peterson, A. W.; Wolf, L. K.; Georgiadis, R. M., (2002), Hybridization of Mismatched or Partially Matched DNA at Surfaces. *Journal of the American Chemical Society 124* (49), 14601-14607.

21. Schreiner, S. M.; Hatch, A. L.; Shudy, D. F.; Howard, D. R.; Howell, C.; Zhao, J. L.; Koelsch, P.; Zharnikov, M.; Petrovykh, D. Y.; Opdahl, A., (2011), Impact of DNA-Surface Interactions on the Stability of DNA Hybrids. *Anal. Chem. 83* (11), 4288-4295.

22. Ambia-Garrido, J.; Vainrub, A.; Pettitt, B. M., (2010), A model for structure and thermodynamics of ssDNA and dsDNA near a surface: A coarse grained approach. *Comput. Phys. Commun. 181* (12), 2001-2007.

23. Irving, D.; Gong, P.; Levicky, R., (2010), DNA Surface Hybridization: Comparison of Theory and Experiment. *J. Phys. Chem. B 114* (22), 7631-7640.

24. Cherepinsky, V.; Hashmi, G.; Mishra, B., (2010), Competitive hybridization models. *Physical Review E 82* (5), 051914.

25. Rothemund, P. W. K., (2006), Folding DNA to create nanoscale shapes and patterns. *Nature 440* (7082), 297-302.

26. Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J., (2007), A coarse grain model for DNA. *J. Chem. Phys. 126* (8), 084901.

27. Weeks, J. D.; Chandler, D.; Andersen, H. C., (1971), Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *The Journal of Chemical Physics 54* (12), 5237-5247.

28. Wong, K. Y.; Pettitt, B. M., (2001), A study of DNA tethered to surface by an all-atom molecular dynamics simulation. *Theoretical Chemistry Accounts 106* (3), 233-235.

29. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., (1992), THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD. *Journal of Computational Chemistry 13* (8), 1011-1021.

30. Owczarzy, R., (2005), Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophysical Chemistry 117* (3), 207-215.

31. Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L. Y.; Behlke, M. A.; Walder, J. A., (2004), Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry 43* (12), 3537-3554.

32. Benight, A. S.; Pancoska, P.; Owczarzy, R.; Vallone, P. M.; Nesetril, J.; Riccelli, P. V., (2001), Calculating sequence-dependent melting stability of duplex DNA oligomers and multiplex sequence analysis by graphs. *Drug-Nucleic Acid Interactions 340*, 165-192.

33. Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A., (1986), PREDICTING DNA DUPLEX STABILITY FROM THE BASE SEQUENCE. *Proc. Natl. Acad. Sci. U. S. A. 83* (11), 3746-3750.

34. Sugimoto, N.; Nakano, S.; Yoneyama, M.; Honda, K., (1996), Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic acids research 24* (22), 4501-5.

35. Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H., (1998), Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry 37* (42), 14719-35.