Real-time Three-Dimensional DNA Microarrays:

How Well Can We Distinguish Between Related Target Sequences?

Maris Lemba

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Paul Yager (Chair)

Albert Benight

Herbert Sauro

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Real-time Three-dimensional Microarrays:

How Well Can We Distinguish Between Related Target Sequences?

Maris Lemba

Chair of the Supervisory Committee:

Prof. Paul Yager

Dept. of Bioengineering

DNA microarrays, due to their highly parallel nature, are in principle well suited for rapid identification of known or related microbial species, but our ability to extract meaningful information from microarray images is still at a rudimentary level. The use of DNA microarrays is currently hampered by a few key analytical and theoretical challenges [1-3]. In particular, the nucleic acid sequence space to be explored can be very large [4], the genetic sequences of many species are very similar, and the concentrations at which the different species are present is typically not known at the time of the sample collection [5], which can result in complex overlapping hybridization patterns. There is much disagreement in the literature regarding the merits of different microarray data analysis approaches, as they have been tested on different

experimental platforms with samples of varying complexity. Experimental validation of analysis methods is limited, and not feasible as a general strategy [4]. Advances in microarray data analysis would accelerate the employment of the powerful DNA microarray technology, already integrated into "lab-on-a-chip" instruments [2, 4, 6-8], in routine clinical practice. This dissertation proposes to improve the diagnostic accuracy of microarrays and characterize their detection limits with respect to distinguishing between closely related target sequences by utilizing computational microarray modeling as a tool for design and validation of microarray data analysis methods and experimental approaches. Mass transport and binding kinetics of oligonucleotide targets in gel matrices was modeled based on current understanding of the thermodynamics of DNA stability, and the performance of gel matrix microarrays as an emerging platform was characterized. Bi-phasic behavior was demonstrated in thermal dissociation curves in a multi-component system, showing promise for distinguishing between related sequences. Finally, the utility of including mismatch probes on the array for distinguishing between related sequences was explored computationally.

Table of Contents

List of Figures

List of Tables

Acknowledgements

All of the research reported here has been performed under the supervision of Dr. Paul Yager at the University of Washington. Dr. Dave Stahl has co-advised this project and all the microarray experimental work has been conducted in his laboratory. Committee members Dr. Albert Benight, Dr. Herbert Sauro, Dr. Paolo Vicini and Dr. Larry Ruzzo all provided helpful discussions, guidance and support while completing this work. Dr. Elain Fu's invaluable advice has had a big influence on shaping this project. Peter Kauffman provided technical assistance through both action and discussion. Dr. Michelle E. Starke advised some of the gel array microarray experimental work. Joshua Thompson helped characterize the microfluidic platform. Mike Purfield assisted in collecting some of the solution-based experimental data for Chapter 2. Dr. Bruce Finlayson has given helpful advice regarding finite element modeling with COMSOL. Mandy Williams also gave helpful regarding computational modeling. Other Yager lab members, past and present, including Dr. Jennifer Foley, Dr. Melissa Hasenbank, Dr. Kenneth Hawkins, Dr. Kristen Helton, Dr. Paolo Spicar-Mihalic, and Dr. Rahber Thariani, are gratefully acknowledged for their advice and support.

Chapter 1: Background and Significance

1.1 DNA microarray technology

DNA microarray technology is widely used in a broad range of fields, from applied genomics and drug discovery to disease diagnostics to agricultural and food industries. Practical applications of microarray technology range from measuring gene expression to detecting single nucleotide polymorphisms. This subsection will review the basics of DNA microarray technology as it relates to this work.

1.1.1 Basic principle of DNA microarrays

Microarray platforms vary in terms of chemistries and detection methods employed, but there is one common theme: immobilization of nucleic acid probes. A microarray platform consists of an array of short nucleic acid capture molecules (probes) immobilized either onto a surface (forming a planar array) or into a three-dimensional gel element (forming a three-dimensional gel array). The array is brought into contact with nucleic acids (targets) in the sample solution, and complementary sequences are hybridized to form probe:target duplexes. If the targets are fluorescently labeled, the duplexes can be detected via fluorescence microscopy on the array after a wash step, and the resulting image can be compared with a spatial map of probe chemistries on the array. This process is illustrated in Figure 1.

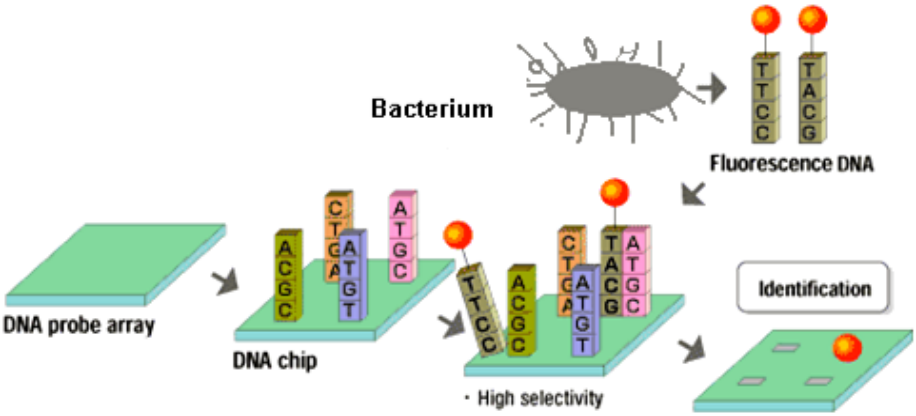

Figure 1 – Basic steps in DNA microarray hybridization experiment. Target DNA is fluorescently labeled, and then applied to a DNA chip where target sequences will find their complementary matches, if present. After washing away unbound target, fluorescent signal is detected and linked back to the original design of the chip to determine potential target sequences. Image modified from http://www.canon.com/technology/future/images/07_1.gif

The substrate onto which the probes are immobilized is dictated partly by the method of signal transduction employed [9]: for example, piezoelectric and electrochemical detection devices would have probes immobilized on the surface of gold electrodes, while devices relying on fluorescence detection would use silica or glass surfaces [9].

1.1.2 Microarrays as an emerging technology for organism identification

Microarrays are particularly well suited for characterizing microbial communities, as hundreds to thousands of nucleic acid signatures can be probed for in a single hybridization assay to determine the identity and approximate abundance of different microbial species. Microarrays have been successfully used to identify microbes in human clinical samples [10-12], and they hold great promise for characterizing human-associated microbial communities in clinical diagnostics as well as for a range of other applications such as monitoring food and water quality [6] and biothreat assessment [13]. One of the most useful markers for organism identification is 16S ribosomal RNA (rRNA). First, it is naturally abundant as a target, and second, as variation in rRNA sequence is related to phylogenetic relationships, probes can be designed to identify organisms at varying taxonomic (phylogenetic) resolution, e.g., from family to species [14, 15].

Several recent advances in microarray technology have brought microarrays close to being employed in rapid direct detection – in particular, the development of gel-based microarrays and integration of microarrays into automated microfluidic devices.

1.1.3 Microarray chemistries

While most commercially available microarrays are currently planar, gel-based microarrays may offer distinct advantages for organism identification [16, 17]. First, the probe concentration can be up to several orders of magnitude higher in gel-based arrays compared to planar microarrays leading to increased signal [18]. Moreover, steric hindrance to duplex formation may be reduced in gel-based arrays compared to planar arrays with surface-immobilized probes, which could further increase the signal [19, 20]. For example, Shchepinov *et al.* showed that adding spacers of at least 40 atoms in length could bring about as high as 150 fold increase in the yield of hybridization [20] The overall greater sensitivity facilitates the detection of low-abundance species without the need for enzymatic amplification, which may introduce bias [21-23]. Second, the gel matrix consists mostly of water, and the hybridization and dissociation in gel elements follow predictable solution kinetics [16], which is important for probe:target duplex stability analysis as discussed below. Lastly, gel-based microarrays are re-usable, which reduces cost and variability between arrays [16, 24]. For example, different types of gel microarrays have been reported to sustain 15-50 rounds of hybridization without a decrease in signal intensity [18, 25]. These are key features for identifying and enumerating microbial populations with high diagnostic accuracy in either clinical or environmental samples. Sorokin *et al.* compared hybridization on gel-based and planar arrays, both made in-house with similar protocols [19]. They found that the fluorescence signal and discrimination between perfect-match and mismatch appeared to be higher in gel-based arrays, and this advantage kept growing as the time of

hybridization progressed [19]. This could be the result of either different kinetic constants for gel pad array, or better probe immobilization in gel arrays [19].

1.2 Microarrays in microfluidic devices

Conventional microarray platforms often suffer from long hybridization times to reach thermodynamic equilibrium. This can be due to either limiting kinetic processes or due to diffusion limitations, and this project investigated both. Diffusion limitation has recently been mitigated by combining microarrays with microfluidic devices.

1.2.1 Significance and potential of microfluidics

Microfluidics is a technology for manipulating and exploiting tiny quantities of fluids, which holds the promise to accomplish biological assays on "lab-on-a-chip" microscale devices [26, 27]. In these devices, small quantities of sample and reagents are guided through microscale channels, mixed, and can be analyzed with a variety of techniques, including traditional laboratory methods such as fluorescence detection [26, 27]. The employment of microfluidic technologies in biosensor applications is a very active research field, as microfluidics has the potential to offer several unique advantages: in particular, the reduction of assay time and smaller quantities of required reagent/sample, and the opportunity for assay automation and integration [26]. In microfluidics, the dominating forces are not necessarily the ones important in macroscale fluidic systems, and microfluidic devices cannot be viewed simply as shrunken counterparts of larger devices [27]. Important effects to consider in microscale are laminar flow, diffusion, fluidic resistance, surface area to volume ratio, and surface tension [27].

Conventional microarray technology features prolonged static incubation of the target solution on the array under a coverslip. This process has several drawbacks that could be alleviated with the implementation of microarray technology on microfluidic platforms [28]. Most importantly, long incubation times are required to achieve optimal signal in conventional assays, as the target nucleic acid must diffuse across the relatively large array (diffusion distances on the order of a centimeter) to interrogate a relatively large high-density array to find its potential match. Slow rate of hybridization reflects the depletion of target from the local environment surrounding the probe [28]. Not only does this prolong the duration of the detection process, but subjects probe linkage chemistry to potential solution-dependent cleavage, reducing the sensitivity and reproducibility of the hybridization [28]. In addition to mitigating the mass transport limitation of the hybridization reaction and lowering total reaction times for achieving optimal signal, microfluidic technologies can potentially offer reduction in sample and reagent volumes, automation of process flow, reduction of labor costs, and integration with upstream and downstream process that reduce the likelihood of sample contamination.

1.2.2 Integration of microarrays into microfluidic devices

Recent integration of microarrays into microfluidic devices has reduced the hybridization time – a significant step towards employing microarrays as a rapid detection technology. Traditional microarray hybridizations typically take 14-16 hours to complete: some researchers argue that for some target compositions, equilibrium during hybridization is not reached until 24 or more hours, depending on the sample composition and target length [29, 30]. Long hybridization times are the result of slow diffusion of nucleic acids in microarrays, limiting the formation of probe:target duplexes. Among the approaches proposed to overcome slow diffusion are mixing [31-33], surface acoustic wave microagitation [34], target recirculation [32, 35, 36], and the

application of electric fields [37] – so-called dynamic hybridizations. The collaborating Stahl and Yager laboratories have demonstrated an improvement in the rate of increase of signal intensity during hybridization upon recirculating the target solution with a peristaltic pump [38]. Furthermore, shorter hybridization times make it more practical to monitor the increase in signal intensity during hybridization and explore the utilization of hybridization kinetics for differentiating specific from non-specific binding [30, 39, 40]. These recent advances in microarray technology, coupled with integration into temperature-controlled imaging platforms, allow rapid collection of real-time hybridization data, previously not available. However, the potential of these advances to improve the diagnostic accuracy of identification microarrays has not yet been explored in depth.

1.3 Microarray data analysis

1.3.1 Challenges in identification microarray data analysis

Despite the advances in microarray technology, one of the major obstacles for widespread use of diagnostic microarrays is data interpretation (see for example [3], for overview and references). A widely acknowledged issue in data interpretation is that the measured hybridization signal does not necessarily correlate with target concentrations [41, 42]. The signal intensity, measured at the probe can be broken into four different terms: specific binding (binding of targets with complementary sequence), non-specific binding (binding of targets with sequences that differ from the complementary one to various extents), background, and system noise [2]. While the probability of probe:target duplex formation decreases with decreasing sequence complementarity, the likelihood of targets with sequences containing one or more non-complementary bases hybridizing on-to a probe is still relatively high. This can seriously

confound the analysis of hybridization data, as will be discussed below. Also, the concentration at which different targets are present (if at all) in identification microarrays is typically not known at the time of the experiment [5]. The added complication for diagnostic arrays applied to polymicrobial specimens stems from the less constrained sequence space and the fact that genetic material of many species may differ by only a few nucleotides [4]. This means that the requirements for distinguishing between different closely related hybridizations are more stringent than for other types of microarrays. The persistent issues with data interpretation have led some authors to recently label the microarray technology as being at a crossroads: "*The strategic question is whether there are any realistic ways to significantly improve the quality of microarray data or it will remain a preliminary screening tool soon to be replaced by next-generation sequencing*" [1]. The inadequacy of conventional end-point hybridization signal measurement has been shown by several different groups, as will be discussed below.

1.3.2 Current analysis algorithms and experimental designs to increase the specificity of identification microarrays

Several approaches have been proposed in the literature to compensate for the confounding effect of non-specific hybridizations: a few selected ones relevant to this work are summarized in Table 1. The approaches can be roughly divided into two categories: those focusing on including probes that vary in sequence at selected positions, and those focusing on altering the experimental conditions during a microarray experiment to measure probe:target duplex stability. The first set of approaches were pioneered by Affymetrix, who proposed the inclusion of so-called mismatch (MM) probes along with perfect-match (PM) probes in the array – the MM probe has a non-complementary base in the sequence [43]. The MM probe is expected to gauge the amount of non-specific binding to the PM probe, and the simplest

proposed method to make use of this information is to subtract the MM signal intensity from that of the PM [43]. However, several studies have shown that under some circumstances, the intensity on the MM probe can exceed that of the PM probe [5, 44], and it has been concluded that using the information obtained through MM probes in a meaningful way is not trivial [45-49].

Table 1 - Published data analysis methods to improve discrimination between closely related hybridizations.

Methods for analyzing signal intensities at the end of hybridization	
Name and reference	**Description**
Comparison of end-point signal intensities on PM and MM probes [43]	Comparison of the signal intensities from PM and MM probes at the end of hybridization, by looking at the 1) difference in intensity values; 2) ratio of signal intensities.
Optimization algorithm for noise reduction [2]	An analytical predictor of nonspecific probe:target interactions combined with an optimization algorithm that iteratively deconvolutes true probe:target signal from raw signal affected by spurious contributions (cross-hybridization, noise, background, and unequal specific hybridization response)
Methods for thermal analysis	
Name and reference	**Description**
T_d comparison [50-52]	Comparison of the temperature at which the normalized signal intensity has dropped to 50% of its original value for PM and MM probes during a thermal dissociation.
Functional ANOVA (Bugli *et al.*, unpublished, [53]	Comparison of the maximum differences and the temperatures at which they occur between thermal dissociation curves from PM and MM, by using a functional ANOVA calculator, adaptation of linear modeling to analyze curve data.
Melt curve first derivative analysis [54-56]	Analysis the first derivatives of the thermal dissociation curves by either comparing data from PM and MM probes, or taking into account PM data only, in terms of 1) the temperatures at which the derivative reaches its maximum value (peak); 2) the width of the peak.

The second approach to estimating specificity of hybridization is to assess the relative stability of duplexes. While in conventional hybridization assays, one would adjust the temperature, hybridization and washing conditions to ensure optimal conditions for the probe:target duplex formation, this is not feasible in DNA microarrays, as the optimal conditions differ across different probes [52, 57, 58]. The stability of the duplexes should therefore be studied under a

range of different experimental conditions. An example of this is conducting a "melt" at the end of the hybridization by changing the temperature on the array [50-52, 54-56, 59]. As the temperature in the hybridization chamber increases, the duplexes start to dissociate based on their relative stability, with the non-specific and MM pairs dissociating before the PM duplexes, as the latter are the most stable [50-52, 58]. The subsequent diffusion of the target away from the probe brings about a decrease in signal intensity as measured on the probe spot (Figure 1). Several approaches have been proposed for comparing the respective thermal dissociation curves, obtained from a graph of signal intensity versus temperature [17, 50-56, 58]: for example, T_d comparison, functional ANOVA, melt curve first derivative comparison (see Table 1 for details).

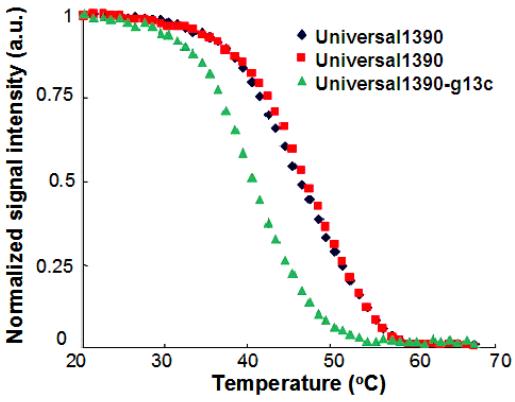

Figure 1 – Normalized signal intensity as a function of temperature during a thermal dissociation, from Smoot *et al.* [60]. Universal1390-g13c is a MM probe to Universal1390, present in two replicas on the array (diamond and square markers on the graph).

Given the time- and labor-intensiveness and relatively high cost of microarray experiments, different analysis methods are usually validated only under a limited set of experimental conditions and for a particular sample of interest, and the performance of an algorithm cannot be subsequently easily extrapolated to a different experimental system. This work aims to demonstrate the power of a computational model to guide the development and evaluation of microarray data analysis methods, especially important in the light of the emerging kinetic microarray data.

1.4 Thermodynamics of DNA hybridization: development of the Nearest-Neighbor model

One of the aims of this work is to develop a physicochemical model of a microarray platform, rooted in an understanding of the thermodynamics of oligonucleotide binding. The latter is a very active research field: active experimental investigation in the thermodynamics of DNA binding and developing of the nearest-neighbor model of DNA stability has been partly motivated by the emergence of applications such as PCR [61], sequencing by hybridization [62], and Southern blotting. However, the majority of the studies have been conducted in solution. In the following subsections, we will review the experimental and theoretical work on DNA hybridization and dissociation relevant to this work.

1.4.1 Unified view of DNA nearest-neighbor thermodynamics

There is a common theme in the numerous publications discussing physical properties of DNA or RNA: an attempt to relate these physical properties to the structure and sequence of polynucleotide and later oligonucleotide subunits with the ultimate goal of predicting physical properties from them. While some physical properties depend only on the gross polynucleotide structure and molecular weight, like light-scattering properties, numerous physical properties depend on the base composition and/or the different interactions by the neighboring bases [63].

In a general sense, these dependencies can be divided into "zeroth-neighbor", "first-neighbor", or "Nth-neighbor" dependencies [63]. In the case of "zeroth"-neighbor" dependency, the property is only a linear function of base composition, *i.e.* the relative frequency of each of the bases. A "first-neighbor" property depends linearly on the dinucleotide frequency: a base and its neighbor in case of a single-stranded DNA, and a basepair and its neighboring pair in case of double-

stranded DNA. "N-th neighbor" property depends respectively on sequences containing N+1 bases (or basepairs) [63]. Gray and Tinoco proposed an approach to studying sequence-dependent properties of polynucleotides based on this idea: "*/A/a sequence-dependent property of any long polynucleotide can be related to the properties of a limited number of other polynucleotides*" [63]. They argued that only the properties of a set of linearly independent sequence combinations need be known in order to predict the properties of polynucleotides. Gray and Tinoco further calculated the number of linearly independent sequences for single- and double-stranded sequences for which the properties would need to be determined [63].

The development of theories around DNA stability, rooted in nearest-neighbor dynamics, dates back to the beginning of 1960s. Authors of one of the first papers in this area, DeVoe and Tinoco [64] used a theoretical approach to calculate the relative contributions of the bases of DNA to the free energies of two possible configurations, the two-stranded DNA helix, and single-stranded random coil. They took into account the contributions of non-paired bases adjacent to base-pairs, and calculated the interaction energies for 10 possible combinations of adjacent base-pairs [64]. They also showed that the inclusion of more than just nearest-neighbor interactions did not significantly influence their calculations [64].

A seminal paper was published by Tinoco and co-workers, studying the thermodynamic parameters, extracted from melting curves, of RNA 6-11 base oligonucleotide sequences [65]. Their data confirmed the very strong sequence-dependence of the thermodynamic parameters. Borer *et al.* proposed that the simplest interpretation of the results is to assume that the contribution of each base pair to the stability of the helix only depends on its nearest neighbors:

i.e., the stability of, for example, a G·C pair depends on whether it is next to another G·C pair or A·U pair [65]. The authors treated dimer thermodynamic parameters as unknown in the 19 linear equations corresponding to 19 different oligonucleotide thermodynamic parameters they had measured, and obtained best least-squares values for them [65]. Although Borer *et al.* had reduced the number of nearest-neighbor parameters from 10 to 6, they did not consider end effects, and used a limited experimental set to derive the parameters, the obtained parameter set gave realistic predictions for melting temperatures of duplexes not used in the original set [65].

Following the groundwork laid by Borer *et al.* [65], a number of groups worked on validating the nearest-neighbor approach to understanding DNA stability, and on refining the parameter sets of dimer stabilities, which has led to the widespread acceptance of the so-called nearest-neighbor model in describing DNA and RNA duplex stability. As already outlined in the previous paragraph, the nearest-neighbor model for nucleic acid assumes that the stability of a given basepair is dependent on the neighboring base pairs. Calculating the stability of an oligonucleotide duplex then means summing over the standard free energy changes of basepairs, as well as an initiation parameter of duplex formation (the parameter has two different values, depending on whether the initiation occurs with terminal G·C or A·T basepair) and an additional entropic penalty if the duplex is self-complementary [66]. The total free energy of a duplex can then be written down in the following manner:

$$\Delta G^0(total) = \sum_i n_i \Delta G^0(i) + \Delta G^0(init\ with\ term\ G \cdot C) + \Delta G^0(init\ with\ term\ A \cdot T) + \Delta G^0(sym),$$

where ΔG(i) are the standard free-energy changes for the 10 possible Watson-Crick nearest-neighbor parameters, n_i is the number of occurrences for each of the nearest-neighbor pairs, and ΔG^0(sym) equals +0.43 kcal/mol if the duplex is self-complementary, and 0 if it is not [66]. The calculation of standard free energy change of an oligonucleotide duplex based on nearest-neighbor parameters is illustrated in Figure 2, taken from SantaLucia's paper [66].

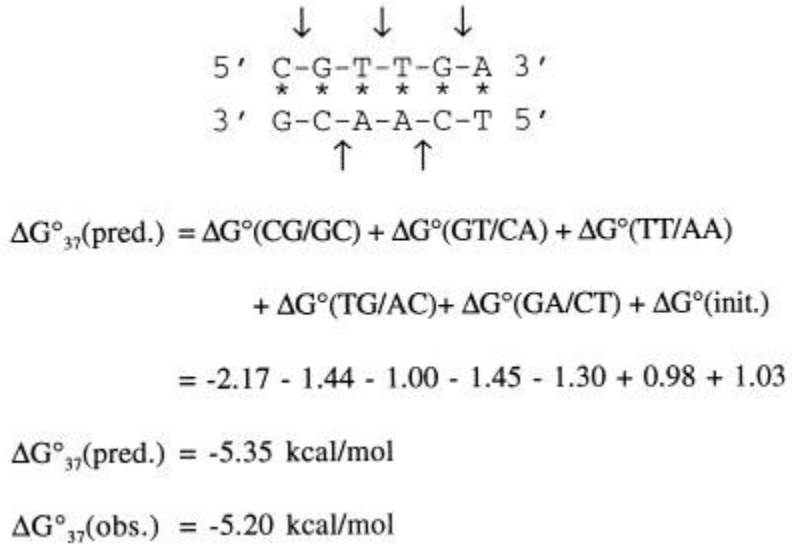

Figure 2 – Calculating standard free energy change of an oligonucleotide duplex CGTTGA·TCAACG, using nearest-neighbor parameters. The arrows point to the middle of each of the nearest-neighbor dimers. As the duplex is not self-complementary, ΔG^0(sym) is zero. Figure taken from SantaLucia [66].

Parameters ΔH^0 and ΔS^0, the change in enthalpy and entropy of the reaction, respectively, can be calculated similarly from nearest-neighbor parameters, or by making use of the following relationship:

$$\Delta G_T^0 = \Delta H^0 - T\Delta S^0.$$

The melting temperature, T_m, at which half of the strands are in the duplex and half are in the random coil state, can also be calculated using the predictions for ΔH^0 and ΔS^0 and oligonucleotide strand concentrations. The equation for T_m for non-self complementary oligonucleotides with equal concentration of both strands is as follows:

$$T_m = \frac{\Delta H^0}{\Delta S^0 + RlnC_T/4},$$

where R is the universal gas constant and C_T the total oligonucleotide strand concentration [66]. If the strands are not present in equal concentration, $C_T/4$ must be replaced by $(C_A-C_B)/2$, where C_A and C_B are the concentrations of the more concentrated and less concentrated oligonucleotide, respectively [66]. If the oligonucleotides are self-complementary, $C_T/4$ should be replaced by C_T [66]. The reader can consult Appendix section A.1 for the derivation of the T_m equation. Appendix section A.2 outlines the methods to derive thermodynamic parameters from melt data.

A seminal paper regarding DNA nearest-neighbor thermodynamics was published in 1998 by SantaLucia [66], setting out to unify thermodynamic parameters that had been published by different laboratories on data acquired using different salt concentrations, different data analysis methods, and different formats to present the data. He compared the nearest-neighbor parameters from seven different studies that had presented data from natural polymers, synthetic polymers, oligonucleotide dumbbells, and oligonucleotide duplexes, with the "unified" oligonucleotide parameters his group had published earlier [67]. SantaLucia found that the parameter sets across the different studies were actually in good agreement with one another, and presented a single set of nearest-neighbor thermodynamic parameters, applicable for DNA hybridization at 1M NaCl concentration, describing both polymer and oligomer thermodynamics [66]. This parameter set is the current gold standard for predicting oligonucleotide stabilities (as of this writing, SantaLucia's paper has been cited over 1,000 times). The parameters are summarized in Appendix section A.3.

While the unified parameter set is valid for Watson-Crick complementary pairs, there are eight possible mismatch pairs in duplexes. Figure 3 shows the 3D structure of DNA, and the Watson-Crick and mismatch pairings between bases. Allawi and SantaLucia hypothesized that the nearest-neighbor model can also be applied to duplexes with internal mismatches, and published nearest-neighbor parameters for internal G·T mismatches along with showing the validity of the approach [67]. SantaLucia and coworkers later published nearest-neighbor parameters for all other seven possible internal mismatches, forming a complete set, along with the unified parameters for Watson-Crick complementary pairs, to describe DNA thermodynamics with or without internal mismatches [68-71]. As the previous datasets had been obtained from perfectly overlapping duplexes, the work was followed up by characterization of the contributions of all possible 32 single-nucleotide dangling ends to duplex stability [72].

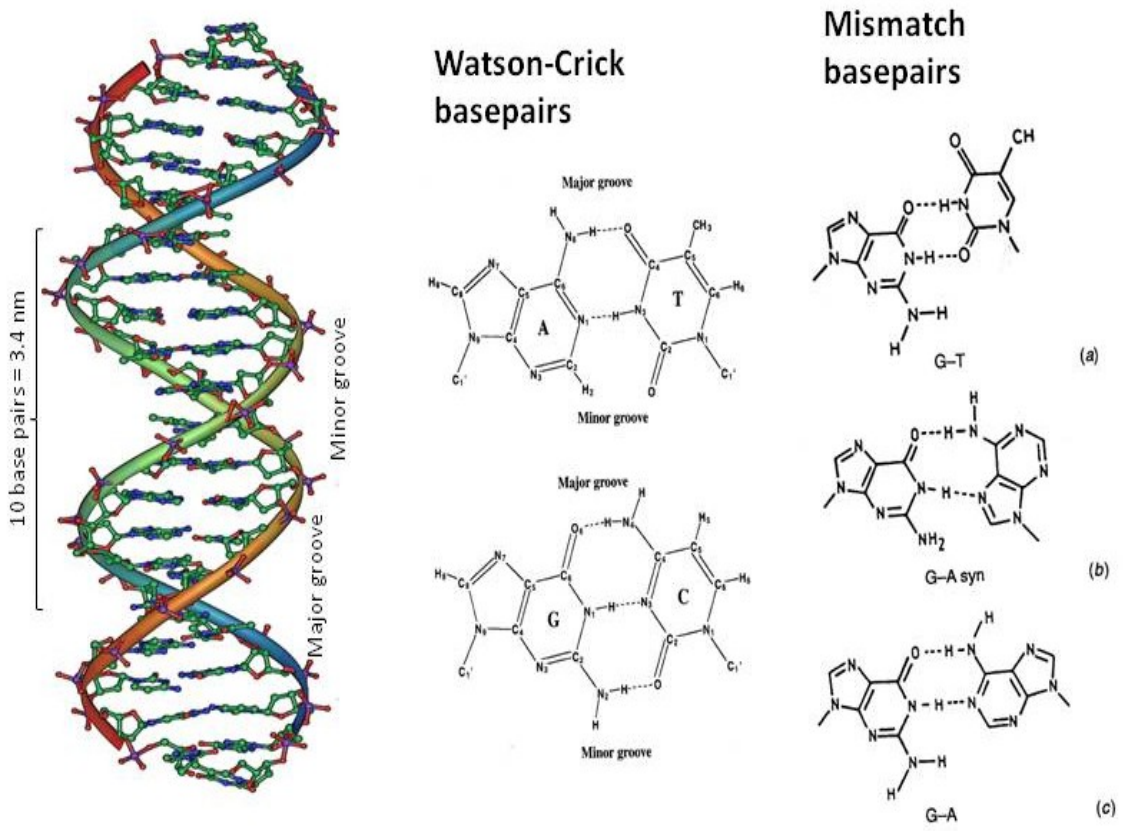

Figure 3 – DNA 3D structure (modified from http://www.molecularstation.com/molecular-biology-images), and pairings between Watson-Crick and mismatch bases (taken from [73]).

1.4.2 Applicability of nearest-neighbor parameters to microarray hybridization

The nearest-neighbor model parameters have been measured in solution, as opposed to in a microarray platform where one of the two hybridizing strands is immobilized. The applicability of solution-based models and their parameters to microarrays has been a matter of debate, especially for mismatch duplexes, with only a few published studies carrying out a direct comparison. For example, Fotin *et al.* compared the thermodynamic parameters (hybridization free energy change) of very short oligonucleotides either forming a perfect-match or containing a terminal mismatch in solution and immobilized in three-dimensional gel pads and found that a

good correlation between ΔG_{chip} and $\Delta G_{solution}$: $\Delta G_{chip}=a\Delta G_{solution}+b$, where a and b are constants [74]. The findings of Fotin *et al.* are shown in Figure 4.

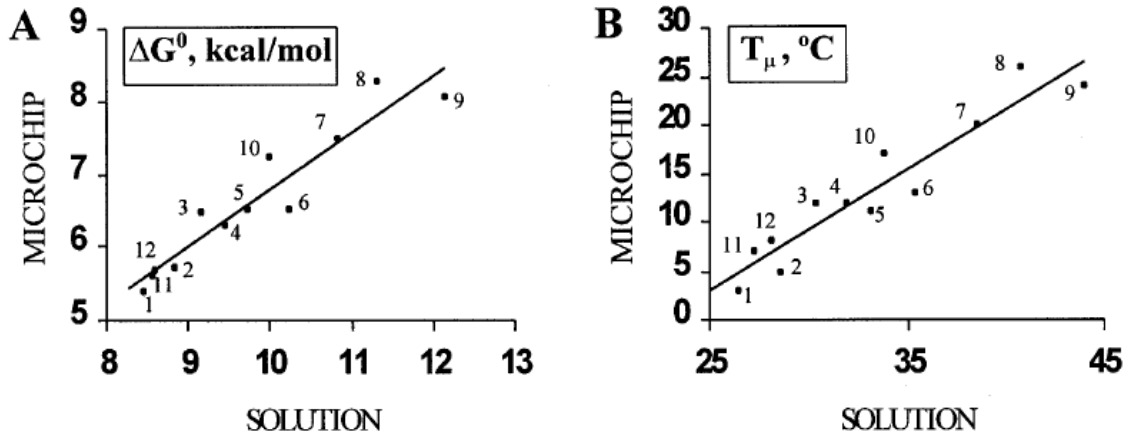

Figure 4 – Correlation between binding free energy change of short duplex formation (left) and melting temperatures (right) in solution and in gel-based microarray. The solution experiments were conducted with 8-base duplexes, while microchip experiments involved the hybridization of 13-base targets onto 8-base immobilized probes. Different numbered points correspond to different perfect-match duplexes (1-10) and one-base terminal mismatch duplexes (11-12). Figure taken from Fotin *et al.* [74], data from solution-based measurements in this figure was taken from Doktycz *et al.* [75] by Fotin *et al.* and combined with microchip data.

Pozhitkov *et al.* on the other hand found poor statistical correlation between measured signal intensities and nearest-neighbor based binding free energies [76]. Weckx *et al.* compared the end-of hybridization signal intensities on CodeLink three-dimensional arrays to intensities computed using a Langmuir model (see section 1.9.2 for a mathematical description of Langmuir models) with ΔG values calculated from solution-based nearest-neighbor parameters [77]. They tested oligonucleotides presenting either a perfect-match to probe sequences or 1-3 mismatches and found good agreement between the computed values and experimental measurements [77]. Fish *et al.* investigated the effect of duplex length, number of mismatches and mismatch position on stability of the duplexes in solution and on the measured signal intensities in planar microarrays [78]. Unlike other studies, the authors measured the thermodynamic parameters in the solution themselves, instead of relying on published nearest-neighbor parameters, making this the first published direct comparison [78]. Fish *et al.* obtained good agreement between

solution free energy and measured signal intensities for perfect-match and most mismatch duplexes, and found that nearest-neighbor model parameters tend to underestimate the stability of mismatch duplexes [78]. On the other hand, Hooyberghs *et al.* found a good correlation between nearest-neighbor based solution binding free energy and microarray free energy, computed from a Langmuir model. However, that finding was limited to lower hybridization intensities only. At higher intensities, the data deviated from Langmuir model for all of the three target concentrations (50-5000 pM) explored [79].

1.5 Effect of buffer composition on DNA thermodynamics

The nearest-neighbor parameters discussed in the previous subsection are typically given for 1M NaCl. As DNA is a charged polymer, altering the salt concentration or adding denaturing agents in the buffer such as formamide can significantly change the relative stability of the duplexes. As the present work is aiming to build a predictive tool to study duplex hybridization and dissociation in microarrays, understanding the influence of buffer composition is important. The effects of salt and formamide concentration on duplex stabilities and how these effects are currently incorporated into theoretical predictions of duplex stabilities in the literature are discussed below.

1.5.1 Effects of salt concentration on DNA stability

Salt is known to affect the structure of both single-stranded DNA and DNA double helix. It also increases the melting point of DNA duplex as it stabilizes the DNA helix by suppressing the electrostatic repulsion between the negatively charged phosphate groups in the complementary strands [80]. Analyzing the thermodynamic data published in literature for oligonucleotide

duplexes in various salt concentrations and comparing it to nearest-neighbor based prediction for 1M NaCl, SantaLucia proposed the following correction for the oligonucleotide standard free energy change:

$$\Delta G^0([Na^+]) = \Delta G^0(1M\ NaCl) - 0.114 \times N \times \ln[Na^+],$$

where $\Delta G^0([Na^+])$ is the free energy change for an oligonucleotide duplex dissolved in a given sodium concentration, ΔG^0(1M NaCl) is the free energy change predicted from the unified nearest-neighbor parameters for 1M NaCl, and N is the total number of phosphates in the duplex divided by 2, *e.g.*, for a 10 basepair duplex without terminal phosphates, N=9 [66]. Assuming that ΔH^0 is independent of salt-concentration – as shown by previous experimental studies by other groups [81-83], SantaLucia proposed the salt correction for entropy as follows:

$$\Delta S^0([Na^+]) = \Delta S^0(1M\ NaCl) + 0.368 \times N \times \ln[Na^+],$$

where again $\Delta S^0([Na^+])$ is the entropy change for an oligonucleotide duplex dissolved in a given sodium concentration, ΔS^0(1M NaCl) is the entropy change predicted from the unified nearest-neighbor parameters for 1M NaCl, and N is the total number of phosphates in the duplex divided by 2 [66].

The effect of salt concentration on DNA hybridization in microarrays has only been considered in a small handful of studies, although it is not clear whether the effect of salt concentration on hybridization would be similar to that observed in solution. The electrostatic penalty in microarray hybridization could depend on the fraction of hybridized probes since the surface charge is increasing as soon as target molecules start hybridizing [84, 85]. Peterlink *et al.* measured the hybridization onto surface-tethered DNA in various NaCl concentrations, and compared the measured T_m with T_m values predicted for solution: the measured T_m, although

lower across all NaCl concentrations, followed a similar logarithmic dependence on NaCl concentration [86]. Fuchs *et al.* also measured equilibrium melting curves from a planar microarray as a function of salt concentration, and proposed a modified Langmuir isotherm that includes electrostatic interactions to explain the data [84].

1.5.2 Effects of formamide on DNA stability

Effect of formamide on hybridization and T_m

Early studies of renaturation of DNA and hybridization of RNA with DNA were conducted at elevated temperatures, as it had been shown that the maximum rate of reaction occurs roughly around 25° C below the T_m [87]. However, as the degradation due to prolonged exposure of nucleic acids to high temperatures is not desirable, adding high concentrations of certain salts or various organic solvents was explored to reduce the stability of double-stranded DNA. Formamide was found to be a particularly useful organic solvent towards this end in several studies [88, 89]. The chemical structure of formamide, an amide derived from formic acid, is shown in Figure 5. Formamide is able to form four hydrogen bonds like water, and is a strong donor and stronger acceptor than water (as reviewed in [90]. It is believed that at least some of the destabilizing effect of formamide on DNA is due to formamide forming hydrogen bonds with DNA or DNA hydrate [90, 91], especially as formamide-water bonds have been calculated to be around 20% stronger than water-water hydrogen bonds (reviewed in [90]).

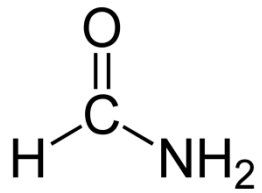

Figure 5 - Chemical structure of formamide, CH_3NO.

However, the utility of adding formamide was soon shown to be not only in lowering the melting temperature and reducing the exposure time of nucleic acids to elevated temperatures [92, 93], but in increasing specificity by reducing non-specific hybridization [94, 95]. One of the first papers to explore the relationship between the rate and specificity of reaction, temperature, and salt and formamide concentration was that of [96]. They first demonstrated that for their polynucleotides, T_m decreased on the average by 0.7 ^oC per percent formamide [96]. When comparing the extent of hybridization between *B. subtilis* DNA and either filter-bound homologous *B. subtilis* or filter-bound heterologous rabbit DNA, they found a similar percent reaction at room temperature and 0% formamide, but adding formamide increased both the extent as well as specificity of binding as demonstrated by the binding onto homologous and heterologous DNA filters [96]. Increasing salt concentration in formamide solutions increased the rate of initial reaction as well as extent of reaction completed in long incubations [96]. Increasing the temperature in a 30% formamide solution revealed that the largest extent of reaction completed was achieved at about 20-25 ^oC below the T_m [96]. Experiments comparing the thermal stabilities of DNA-DNA duplexes formed when mouse DNA was incubated with filters containing either mouse or hamster DNA revealed higher thermal stability (implying higher specificity) in hybridization solutions containing higher formamide concentrations (formamide was washed away from the solution before conducting thermal melts to assess stability) [96]. The authors also outlined guidelines for optimizing incubation mixture conditions. For example, for an incubation temperature of ~25 ^oC, optimal reaction rates will be obtained for T_ms around 45-50 ^oC [96]. Knowing the T_m for a given salt concentration with no formamide in the incubation mixture, one can then calculate the required formamide concentration needed to lower T_m, using the established linear relationship between T_m and

formamide concentration. Several later studies confirmed that T_m decreases approximately by 0.60-0.65 ^oC per % formamide (volume fraction × 100) [97, 98]. Hutton used single-stranded DNA ranging from 1,600 nucleotides to 20,000 nucleotides to study the effect of formamide on DNA stability in salt solutions of different concentration and found the dependency of T_m on formamide concentration to be slightly stronger in high salt solutions [98]. Specifically, while T_m decreased by 0.6 ^oC per percent formamide for NaCl concentrations ranging from 0.035 to 0.88 M, it decreased by 0.65 ^oC per percent formamide at 1.12 M NaCl [98]. Furthermore, he showed that the salt concentration at which the salt dependence of T_m levels off decreases with increasing concentrations of formamide. The NaCl concentration for a particular formamide concentration, above which the T_m no longer increased with the addition of NaCl, could be calculated with the following experimentally based formula:

$$\text{NaCl molarity} = -0.0096 \times \text{percent formamide} + 1.06 \text{ [98].}$$

Casey and Davidson [97] showed using synthetic polynucleotides that the base composition of polynucleotide duplex influences the extent to which adding formamide can decrease duplex stability: for example, the T_m of dA:dT duplex decreased by 0.75 ^oC per percent formamide, while the T_m of dG:dC duplex decreased only by 0.5 ^oC per percent formamide [97]. The T_m of most synthetic RNA:DNA duplexes decreased linearly with increasing formamide concentration, except the T_m of rG:dC duplex, which decreased nonlinearly, similarly to RNA:DNA hybrids. RNA:DNA hybrids were also shown to be more stable at higher formamide concentrations than DNA:DNA duplexes [97].

Blake and Delcourt [90] carried out a systematic characterization of the effects of different concentrations of formamide on DNAs of different (G+C) content. They found that the effect of formamide on the melting temperature depends on the (G+C) content of sequences, helix conformation, and the state of hydration. Their results showed that formamide did not influence the inherent cooperativity of melting, or the strong dependence of the conformational state of the base pairs on those of their neighbors due to hydrogen bonding and stacking forces, and that formamide both destabilizes the helical state and stabilizes the coiled state of DNA [90]. Quantitatively, Blake and Delcourt showed that the dT_m/dC_F for plasmid DNA domains is a function of (G+C) content in the following manner:

$$\frac{dT_m}{dC_F} = 0.453 \cdot (G + C) - 2.88.$$

The authors attributed deviations from this relationship to *"sequence-dependent variations in the energy of hydration at selected sites"* and showed that the deviations are greater for synthetic polynucleotide duplexes [90]. Calculated transition enthalpy drops steeply as a function of formamide concentration between 0-1.0 M formamide, roughly -0.8 kcal/mol-bp per 1M formamide/l, arguably corresponding to "*a small enthalpic loss associated with the exchange of formamide for bound water and possibly counterion to the helical state*" [90]. The authors argued that the subsequent gradual drop (0.05 kcal/mol-bp) in transition enthalpy above 1M formamide is probably due to a small reduction in residual single-strand stacking enthalpy with increasing T_m.

Effect of formamide on non-equilibrium dissociation in microarrays

While the majority of work detailing the effect of formamide on DNA hybridization and melting has been conducted with long polynucleotide sequences, equally important from the viewpoint of

practical applications of hybridization chemistry is to understand the influence of formamide the context of a microarray platform. Urawaka and colleagues were first to publish a study assessing the effect of formamide on DNA duplex formation in the context of a microarray platform [17]. In particular, they used a polyacrylamide gel-based three-dimensional microarray platform, using 39-base long DNA oligonucleotide targets hybridizing onto 19-base oligonucleotide probes, immobilized in gel elements [17]. The hybridization buffer the authors used contained 40% formamide, and the melt buffer contained 0-40% formamide in 10% increments. The hybridization was carried out at room temperature for 12 h. They characterized the effect on melt curves in terms of the effect on T_d: a non-equilibrium parameter referring to the temperature at which the initial signal had decreased by 50% (or the temperature calculated as the mean of two T_ds: one based on maximum signal intensity, and one on the initial intensity) [17]. The non-equilibrium melts were conducted over the range of 52 ^oC at a rate 1 ^oC/min. The study compared the change in T_d as a function of formamide concentration for melt curves from two different probes and their one-base mismatches in five different positions. The average decrease in T_d was -0.56 ^oC/%FA when averaging over all probe-target duplexes, including mismatches [17]. Analysis of sensitivity of T_d to individual inputs revealed that over 75% of the variation in T_d was due to changes in formamide concentration, around 19% due to different mismatch positions, and less than 6% due to the type of mismatch [17]. While the authors provided a table, listing the T_d values as a function of mismatch position and type and formamide concentration, they did not present an analysis of the sequence-dependent effects of formamide beyond commenting that formamide appeared to decrease the T_d of G or C mismatches slightly stronger than A or T mismatches [17].

Only in 2010 was the first work published that investigated the influence of formamide on DNA hybridization in planar microarrays [99], where the hybridization is affected by proximity of the surface, and interactions between different probes, targets, and the surface. Furthermore, they were interested in elucidating whether the effect of formamide is different on perfect-match versus mismatch duplexes in a microarray, and whether the effect depends on the surface chemistry of probe immobilization [99]. For the study, Fuchs *et al.* used probes with a hybridizing sequence of 15-16 bases, and targets of 15 or 20 bases, with mismatches in the middle of the sequences. The hybridization was carried out at 25 ^{o}C for 10 min., while hybridization buffer was constantly flowing through the device. They conducted flow-through non-equilibrium melting with a temperature scan rate of 2^{o}C/min on an SPR system, from 25 to 70 ^{o}C, and only the hybridization buffer contained formamide. While comparing the two grafting methods used – polypyrrole electrospotting and spotting of thiol-modified DNA probes – rendered higher T_d values for thiol chips, the reduction in T_d as a function of formamide concentration was comparable for both chips for perfect-match duplexes [99]. Experiments comparing the effect of formamide concentration on T_d values of perfect-match and one and two-base mismatches showed no significant sequence-dependent reduction in T_d, although two-base mismatches showed a slightly weaker dependency of T_d on formamide concentration, calculated as the slope ΔT_d per percentage of formamide [99]. Average slopes were -0.57 ^{o}C/% FA for polypyrrole surfaces and -0.59 ^{o}C/% FA for thiol surfaces.

One note of caution when comparing the effects of formamide on T_d values in all these studies, as there are vast variations in the protocols of these studies: a) the studies are generally conducted with sequences of different composition and lengths; b) the melt buffers differed in

composition in terms of salt concentration; c) except for studies by [17] and [99], the melting experiments have been conducted under equilibrium conditions; d) the studies vary from static (*e.g.* [17]) to flow-through melts [99]; e) the pre-melt hybridizations have taken place over vastly different condition, ranging from no formamide in hybridization buffer [96] to 40% formamide in the hybridization buffer [17], not to mention hybridizing under different salt concentration and hybridization times.

Thermodynamic models of formamide effect

While the effect of salt concentration on nearest-neighbor thermodynamic parameters is described fairly well and validated by several groups [66], there seems to be a paucity of similar work relating formamide concentration to nearest-neighbor parameters. The only work to the best of our knowledge aiming to simulate the effect of formamide on probe dissociation is that by Yilmaz and Noguera [100]. They were motivated by the need to develop better theoretical tools to optimize protocols for FISH (Fluorescence In Situ Hybridization). FISH is conducted by hybridizing fluorescently labeled DNA oligonucleotides – probes – with rRNA molecules – targets - present in fixed whole cells. The specificity of FISH is achieved by designing the probe to be a match to a sequence on the rRNA of the organism of interest with minimal potential for cross-hybridization to sites on other organisms, and by minimizing the binding of the probe to non-target organisms by adjusting the stringency conditions of the hybridization and wash through the use of formamide [100]. The generally practiced approach to optimizing FISH protocols involves performing FISH with pure cultures, but the feasibility of this approach is however limited by the availability of target or non-target organisms in pure culture, and by the sheer number of non-targets of interest [100]. Contrary to the dissociation studies cited above

that looked at the effect of a constant concentration of formamide on the thermally induced dissociation profiles, in this study, Yilmaz and Noguera obtained formamide-induced equilibrium dissociation profiles at a constant temperature, increasing formamide concentration from 0 to 80% in 10% increments. They formulated the effect of formamide on binding free energy in the following manner:

$$\Delta G^0_{i,[FA]} = \Delta G^0_{i,0\%} + m_i \cdot [FA],$$

adopting a notion from the literature on thermodynamics of interactions of biopolymers with co-solvents and other solutes that the free energy change of denaturation is roughly proportional to the concentration of the denaturant [100]. The index "i" assumes values between one and 3, referring to different reactions that their equilibrium model covers: interaction between rRNA and the DNA probe (i=1); intramolecular DNA/DNA interactions due to the self-complementarity of the probe (i=2); and intramolecular RNA/RNA interactions forming the secondary structure of rRNA (i=3) [100]. The proportionality constant in the equation above is called the "m-value". Using 16-25 base long DNA probes and 16S rRNA of *E. coli*, the authors used curve-fitting of the model to experimental data, and obtained best estimates for m_1, m_2, and m_3 either as global formulations applicable to every probe, or functions of probe length, G/C content and duplex stability (m_2), or target site location, characteristics of the target structure, and ΔG_3 (m_3) [100]. The authors omitted reaction 2 from further consideration as it was shown to be insignificant in the overall hybridization process. Their analysis with perfect-match probe-target duplexes concluded that several model formulations presented an improvement over the predictive power of the simplest model that assumed constant global values for m_1 and m_3 [100]. Interestingly, the authors did not find the G/C content to influence the effectiveness of formamide-induced denaturation of RNA/DNA duplexes, contrary to DNA/DNA data reported

by Blake and Delcourt [90, 100]. The model chosen for in-depth analysis had the following form: $m_1=\theta_1+\theta_2 L$, $m_3=\theta_3\Delta G^0_{3,0\%}$, where L is the length of the probe, and θ_1, θ_2, and θ_3 are constants [100]. While the model predicted the dissociation profiles of perfect-match probe-target pairs not used during the fitting phase quite well, the situation was different for single-mismatch duplexes. The only parameter that should have changed, if formamide denaturation mechanisms in mismatched duplexes are the same as in perfect-match duplexes, is ΔG^0_1, which was not explicitly included in the model formulation. However, to obtain agreements comparable to those with perfect-match duplexes, Yilmaz and Noguera had to re-optimize parameters θ_1 and θ_2 through curve-fitting using the experimental dissociation profiles on all available mismatch duplexes [100]. It is important to keep in mind two things when reflecting on the significance of that finding. First, the accuracy of the model proposed by Yilmaz and Noguera is significantly influenced by the input ΔG values, as the authors themselves acknowledge [100]. The authors do not discuss whether $\Delta G^0_{1,0\%}$ predictions were accurate for the mismatch duplexes used in this study. A follow-up study by Yilmaz *et al*. discusses the incompleteness of the data set of thermodynamic stability parameters for DNA/RNA mismatched duplexes and suggests using an average of DNA/DNA and RNA/RNA parameters [101]. Second, the data set used for model fitting did not include any data from mismatch duplexes: one cannot help but wonder if using a data set comprised of both mismatch and perfect-match dissociation profiles for model fitting would have rendered model parameters capable of predicting dissociation profiles not used in the original set, regardless of the degree of match.

1.6 Hybridization kinetics of short DNA oligonucleotides

While the bulk of the work on using microarrays to distinguish between perfect-match and mismatch binding has been on thermodynamic distinction, marked by differences in binding free energy, several recent publications have focused on kinetic effects for discrimination. These are especially important for applying microarray technology as a biosensor platform. One of the first experimental studies to investigate the hybridization kinetics of short DNA oligonucleotides in microarrays was done by Herning *et al.* [39]. They monitored the hybridization of short (16-25 bases) DNA probes onto long DNA targets using fluorescence polarization measurements, and investigated the influence of three-base mismatches in the probe and probe length on hybridization kinetics. Herning *et al.* reported that the presence of mismatches had a significant effect on hybridization kinetics, and while they did not find any difference in the hybridization of 16-base probe versus 25-base probe, the 16-base probe was more sensitive to the presence of mismatches than the 25-base probe [39].

The differences in stabilities between perfect-match and mismatch duplexes are mainly due to differences in dissociation constants. Gotoh *et al.* used a 20-mer oligonucleotide immobilized on a planar surface and 11- to 20-mer target oligonucleotides to study the effect of mismatches on association and dissociation constants via surface plasmon resonance measurements [102]. They found the effect of mismatches to be more pronounced on the dissociation constant than the association constant. In addition, in agreement with findings from Herning *et al.* [39], Gotoh *et al.* found that the effect of mismatches was more pronounced on the stability (equilibrium constant) of the duplex as the target length decreased from 20 to 15, 13 and 11 bases [102]. In addition, the overall equilibrium constant of the perfect-match duplex decreased as well, mainly

through changes in the dissociation constant, while the association constant was not significantly altered. For example, the equilibrium constant is roughly 3.5 times larger for the duplex with a 20-mer target compared to 13-mer target, while the ratio of perfect-match versus one-base mismatch equilibrium constants was 2.5 for the 20-mer and 7 for the 15-mer [102].

Sorokin *et al.* demonstrated experimentally with gel-based microarrays that higher probe concentration correlates with higher duplex concentrations, but also with longer times to equilibrium [103]. They showed that as mismatched duplexes hybridize faster, the perfect-to-mismatch signal ratio is lower in the transient hybridization regime before equilibrium is reached. This is however augmented if the diffusion coefficient in the probe is increased or if the initial concentration of the probe is decreased [103]. This presents a trade-off between perfect-match and mismatch discrimination and fluorescence signal intensity. Khomyakova *et al.* demonstrated that 50-mer oligonucleotide probes corresponding to the same genes showing very different hybridization signals, and compared the differences in melting temperatures between these probes [104]. The majority of the probes studied had fairly similar melting temperatures, but anecdotal evidence collected from just one probe complex showed differing hybridization kinetics [104].

1.7 Influence of target length on hybridization

Although this work is not immediately concerned with investigating the effect of target length on hybridization and melt in microarrays, a comment on the influence of target length is warranted, as it is hard to imagine an application of microarray technology where the sample would consist of targets of uniform lengths. One of the classic papers (by number of citations) on DNA

renaturation was written by Wetmur and Davidson in 1968 [105]. Wetmur and Davidson confirmed that DNA renaturation can be described as a second order reaction, and proposed an equation relating the observed association rate constant k_a to the (length-independent) rate constant of nucleation k_N', DNA length L and sequence complexity N:

$$k_a = k_N' \frac{\sqrt{L}}{N}.$$

Complexity is defined as the number of DNA base-pairs in non-repeating sequences [105]. The step with the lowest rate constant in DNA renaturation is the initial nucleation step, the formation of one or a few basepairs in some complementary point along the two DNA strands, while the formation of remaining basepairs ("zippering up") occurs very fast [105]. Wetmur and Davidson also remark that "*The reaction rate increases as the temperature decreases below T_m, reaching a broad flat maximum from about 15 to 30 °C below T_m and then it decreases with a further decrease in temperature*" [105]. The dependence of k_N' on sequence composition (*i.e.*, GC percentage) and solution properties was later reviewed by Wetmur [106].

It is important to note that most biophysical studies have focused on short oligonucleotide microarrays. Long oligonucleotide arrays have not received the same amount of attention, nor has the binding of long oligonucleotides in solution been characterized as well. For long oligonucleotide microarrays, multi-step models should be employed that would include target and probe folding, probe-probe and target-target interactions [107, 108].

Suzuki *et al.* experimentally investigated the effect of probe length on the sensitivity and specificity of hybridization signal [109]. They used probes of 14 to 25 bases tethered to the surface of an array, and used synthetic 25-mer DNA oligos to characterize the dependence on the

probe length of the hybridization end-point signal intensity (at 16 h) and the ratio of signals from the probes presenting a perfect-match and mismatch sequence to the target [109]. While the signal intensity as a function of probe length exhibited a sigmoidal shape, the signal intensities at longer probe lengths (23- to 25-mer) were saturated at both lower and higher target concentrations, corresponding to finite availability of target and probe molecules, respectively [109]. Comparing the ratios of signal intensity of PM to those of cognate MM probes as a function of probe length and target concentration, Suzuki *et al.* found that the PM/MM signal intensity ratio decreased for all target concentrations as the probe length increased beyond 21 bases, indicating decreased specificity of longer probes [109]. This agrees qualitatively well with earlier published studies [110, 111].

1.8 Secondary structure effects on hybridization

In this work, we are not concerned with considering secondary structure of targets and/or probes, as these effects can be mitigated by elevating hybridization temperature or appropriately designing shortened probe sequences and employing shorter target sequences. There are only a few studies that look quantitatively at the effect of oligonucleotide secondary structure on hybridization. For example, Gao *et al.* use 25-mer DNA oligonucleotide targets of varying degrees of secondary structure, and compared the effect of secondary structure on the rate of hybridization in solution and in a planar microarray platform, using either UV absorbance or surface plasmon resonance spectroscopy [112]. While the hybridization rate was suppressed on the surface 20 to 40-fold compared to that in the solution, the inclusion of three intramolecular basepairs in the probe and target decreased the apparent association constant in the two-state

binding model roughly two-fold [112]. The data from probe and target sequences with four or more intramolecular basepairs could not be fitted to a two-state binding model anymore [112].

1.9 Mathematical modeling of microarray processes

1.9.1 Significance and potential of microarray modeling

For the purpose of fulfilling the promise of microarray technology for rapid organism identification, it is essential to characterize the limits of detection and recognize instances where the identification of a particular species is not possible due to confounding target cross-hybridization [51]. Mathematical modeling as a scientific method has been successfully applied in biological research and technology development. Several recent papers have pointed out the lack of an analytical/theoretical framework to guide the development of data analysis methods for diagnostic microarrays [4, 51, 53], and the potential usefulness of computational models to better understand PM/MM hybridizations [29, 100, 113]. Simulations can offer the opportunity to characterize the response of a system to a wide-range variation of several parameters in a manner that is significantly faster, cheaper and more general than actual experiments. Several important lessons about the interpretation of microarray data analysis have already been learned from a virtual experiment, as will be outlined below. However, hybridization and dissociation within a three-dimensional gel element of samples of complex target mixtures with different target concentrations and sequence heterogeneities, subject to diffusion, convective transport, and temperature effects, have not been explored in detail in a single mathematical model. We anticipate that a comprehensive computational model of microarray hybridization will provide an important tool for systematic evaluation of analysis algorithms and/or experimental designs for a

large variety of sample compositions, and guide the development of methods to further discriminate between hybridizations by similar targets [114].

1.9.2 Langmuir models

Many important biological processes involve the binding of a solute molecule onto a surface-immobilized molecule, for example cell signaling after receptor-ligand interactions. Models of receptor-ligand binding are summarized by Lauffenburger and Linderman [115]. As a first approximation, DNA target binding onto immobilized probes can be treated as receptor-ligand binding [116]. In this work, we are concerned with an approach to modeling DNA hybridization and dissociation that is firmly rooted in current understanding of the thermodynamics of DNA binding. The simplest way to incorporate thermodynamics into the modeling of surface-bound DNA duplexes in a microarray is using the Langmuir model. The so-called Langmuir equation (or isotherm) relates the coverage, or adsorption of molecules onto a surface, to the concentration of the adsorbing species above the surface at a fixed temperature. If the rate of change in the fraction of surface-bound species θ in a complex is described as follows:

$$\frac{d\theta}{dt} = ck_1(1-\theta) - k_{-1}\theta,$$

where c is concentration of the adsorbing species, and k_1 and k_{-1} are the forward and reverse reaction kinetics constants, respectively, then in equilibrium conditions we obtain

$$ck_1(1-\theta) = k_{-1}\theta.$$

Taking into account that $k_1/k_{-1}=\exp(-\Delta G/RT)$ and rearranging, we obtain

$$\theta = \frac{ce^{-\Delta G/RT}}{1+ce^{-\Delta G/RT}}.$$

This is called the Langmuir equation (or isotherm).

The simplest equilibrium thermodynamic model of DNA hybridization in microarrays is the Langmuir model, which represents signal intensity as a function of target concentration in solution, temperature, and binding free energy:

$$I = I_o + \frac{Ace^{-\Delta G/RT}}{1 + ce^{-\Delta G/RT}},$$

where I_o is the non-specific background signal, A is the proportionality constant, c is target concentration in solution, T is temperature, ΔG is binding free energy, and R is the universal gas constant. An in-depth discussion of the use of the Langmuir model has been presented by Halperin *et al.* [117]. As outlined in section 1.4.2 above, the Langmuir model has been employed by a number of studies to fit data from microarray experiments measuring end-point equilibrium values for signal intensities.

1.9.3 Mechanistic models of microarray hybridization

In this section, we will discuss published mechanistic models of DNA binding that are not limited to equilibrium and are more detailed than the Langmuir model presented in 2.9.2. Several groups have called for a deeper theoretical understanding of physicochemical aspects of microarray hybridizations [17, 40, 113, 116, 118, 119]. Let us survey the development literature of DNA microarray models.

Several mechanistic models of DNA hybridization combined with mass transport to a surface have been proposed in the literature. One of the first and most cited studies was published by Chan *et al*. [116], modeling DNA diffusion and hybridization. They considered DNA hybridization by two mechanisms: either direct or initial non-specific adsorption following by surface diffusion to the probe. Chan *et al.* assumed irreversible binding, the existence of a constant concentration of soluble targets at a certain distance away from the binding surface that

itself was considered as an infinite flat plate, and that the number of available probes is constant and not dependent on reaction rates [116]. These assumptions would be most appropriate for the initial time points in the hybridization. They investigated the effect of target DNA molecular weight, size of immobilized DNA, and surface probe density on hybridization reaction [116].

Livshits and Mirzabekov [40] published a theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligonucleotides. They developed analytical solutions to two differential equations describing the concentration of free target in the gel and bound duplex, determined by the diffusion of targets in the gel and association/dissociation of duplexes. They made two significant assumptions to arrive at analytical solutions: the concentration of target outside the gel was set to be constant at all times, and the ratio of bound target to total probe concentration was taken to be significantly less than one (low saturation conditions). They argue that "*for optimal discrimination of perfect duplexes from duplexes with mismatches the hybridization process should be brought to equilibrium under low-temperature nonsaturation conditions for all cells*" [40]

Wang and Drlica [120] developed a model to describe the hybridization of short complementary regions within large nucleic acids, relating the steady-state rate constant to experimentally measured free energy parameters (in their case, pertaining to the formation of locally single strands and the free energy gained as a result of hybridization).

Erickson *et al.* presented a general theory of solid-phase hybridization using a two-mechanism approach in which the target can hybridize with a surface-bound probe directly from the

solution, or non-specifically adsorb onto the surface and then hybridize with the probe through a surface diffusion step [121]. They used a 3D finite element model to simulate the hybridization of targets onto a 2-dimensional probe surface in a 3D channel, and investigated the effect of flow rate, target concentration, sequence (comparing perfect-match and mismatch target/probe hybridization) and channel height on the hybridization time and pattern. However, they only considered a flow-through system (their model included a fixed concentration as a boundary condition for target concentration at the inlet), and they compared the hybridization of perfect-match and mismatch targets as separate hybridization experiments [121].

Gadgil *et al.* [118] published a finite element reaction-diffusion model of DNA hybridization using a realistic assay geometry with a finite volume and reversible reaction kinetics. They considered the effect of target concentration and the effect of mismatch (modeled with an increase of dissociation constant). Gadgil *et al.* [118] investigated the development of a depletion zone in the solution surrounding the probe surface and duplex formation over time as a function of target concentration.

The Stahl and Yager groups have co-published a preliminary computational fluid dynamics model of the microarray hybridization assay, investigating the effect of convective flow on hybridization signal in a single-analyte system. Using a computational software package, a model of the hybridization was composed, that incorporates reversible hybridization, incompressible fluid dynamics, and mass transport of the labeled target oligonucleotide, taking also into account the three-dimensional geometry of the DNA target binding spots [38]. The model was used to calculate three-dimensional profiles of fluid velocity in the hybridization

device, concentration of the target in the bulk, and the concentration of probe-target duplexes [38], see Figure 6. The results demonstrate the local depletion of target DNA in the solution behind the gel elements that are specific to the target, compared to other non-specific gel elements [38]. This is due to the difference in binding kinetics, with the dissociation constant for duplexes of target DNA and non-specific probes being significantly larger than that for target duplexes with specific probes. The results presented in [38] demonstrate the increase in signal intensity as a function of time under dynamic hybridization compared to the static hybridization. Comparison of the simulation results with experimental data showed good qualitative agreement [38].

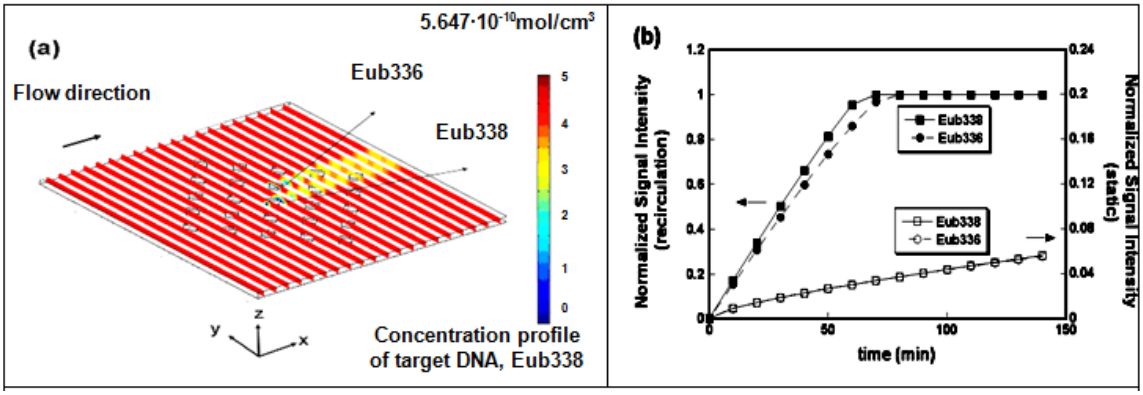

Figure 6 - Data from Dr. Yager's laboratory on the computational modeling of microarray hybridization assay with three-dimensional gel probe structures. Panel (a) displays the DNA concentration profile in the bulk. Panel (b) displays the normalized signal intensity on selected probes. Eub338 refers to a probe with perfect complementary match sequence to target sequence and Eub336 refers to two-base mis-match probe to target sequence. Figure taken from Lee *et al.*, 2006 [38].

All the studies summarized above have one important shortcoming when it comes to understanding DNA hybridization in potentially complex mixtures. The studies of the kinetics of oligo- or polynucleotide hybridization had assumed that the interaction between each probe/target pair can be treated as a separate event [113]. Most real-life applications of microarray hybridization kinetics however involve the simultaneous interactions of targets of

varying affinities with probe molecules. The works by Zhang *et al.* (Zhang, Hammer et al. 2005), Bishop *et al.* [29], and Horne *et al.* [122] are some of the first to consider a realistic binding scenario: that of a heterogeneous mix of multiple targets competing for probe sites. The main lessons from both the computational and experimental investigations into competitive binding have very aptly been summarized by Bishop *et al.* [123]:

"*1. The specificity of recognition is controlled by the competitive displacement of the lower-affinity species by the higher-affinity one;*
2. The signature of this mechanism is a nonmonotonic growth curve of the lower-affinity species;
3. Depending on the relative concentration and rate constants of the primary and secondary species, the latter may become a major contributor to the observed signal, especially in the transient regime;
4. The presence of the secondary species extends the time to reach equilibrium;
5. The maximum specificity is obtained at equilibrium; and
6. In the absence of the primary target, the secondary species may produce a signal comparable to the primary target alone."

Let us review the studies investigating competitive hybridization in detail.

Zhang *et al.* considered a hybridization system consisting of either a single soluble species binding on two immobilized spots, or two soluble species binding on two spots. The kinetic model used to mathematically describe the system consisted of differential equations describing the change of bound concentrations for the duplexes, and the effects of diffusion were entirely ignored as if in a situation of a well mixed solution [113]. The authors found that the ratio of duplexes formed on the immobilized spots can change quite dramatically over time, and that the results can be influenced heavily by the ratio of soluble (target) to immobilized (probe) molecules. Zhang *et al.* also simulated the wash after hybridization with the same model used for hybridization by simply setting the initial conditions of the duplexes equal to the end concentrations in the hybridization simulations, and the concentration of free solute equal to zero, and introducing a dilution factor for the free solute concentration to account for the wash

solution volume being typically significantly larger than the volume of the hybridization buffer [113]. They found that in their simulations, the total wash time and wash volume can significantly influence the ratio of high affinity to low affinity bound duplexes, and that theoretically, there exists an optimal wash time [113]. While it is not clear what conclusion one can really draw for experimental implementation of washing from a kinetic model with a well mixed solution, the authors' emphasis is that the washing step has not quite received the attention it deserves.

Bishop *et al*. [29] built a finite element model, describing oligonucleotide target diffusion in a hybridization chamber and binding onto a two-dimensional sensing surface. They compare the hybridization profiles in a single-component system, consisting of one type of a target, with those in a two-component system, consisting of two types of targets: a perfect-match and either one-, two- or three-base mismatch targets to the probe [29]. The theoretical model assumed that the association constant of binding is constant over a range of temperatures and sequences, while the dissociation constant is dependent on both duplex sequence and temperature. Each binding scenario considered only one type of probe. Bishop *et al.* characterized the effects of target and mismatch species concentrations, temperature and total hybridization time on the specificity of the hybridization in the two-component system [29]. Important findings of the work are the following. In the two-component model, containing a perfect-match and lower affinity mismatch target competing for binding on the same sensing zone, the surface binding reactions proceed in two phases [29]. In the early phase, where probe molecules are in excess, binding of perfect-match and mismatch targets happens independently. In the second phase where most if not all probe molecules have formed duplexes, the perfect-match target species gradually replaces

mismatch duplexes due to their lower dissociation rate [29]. This has the following implications for interpreting microarray hybridization data. Bishop *et al.* showed that at low initial target concentrations, it can take up to several days to reach 90% of thermodynamic equilibrium, and these times are prolonged even more by the presence of mismatch species [29]. Also, variations in hybridization time can significantly impact the fraction of signal on a probe originating in bound mismatch versus bound perfect-match. The simulation results indicate that in scenarios where mismatch species are present at equal or higher concentrations than the perfect-match, employing mismatch probes to assess nonspecific background is likely to produce errors [29]. Bishop *et al.* explore the theoretical limit of discrimination in a two-component system by using the thermodynamic equilibrium equations for the match target and corresponding mismatch equation in a two-component system and show the dynamic range as a function of concentrations of match and mismatch and dissociation constant ratios. The theoretical limit for the dynamic range is shown to decrease if the mismatch concentration is larger than the matched target concentration [29].

Bishop *et al.* subsequently verified the predictions of their theoretical model, presenting experimental results that aligned well with simulated data [124]. Namely, the authors presented the first direct evidence of kinetic competition and displacement of low-affinity targets in the literature by monitoring in real time the hybridization of a high (perfect-match) and lower affinity (mismatch) target onto a two-dimensional sensing zone [124]. This was achieved by labeling the two species in the two-component system with different fluorescent labels [124]. In a follow-up paper, the group proposed an interesting experimental approach to monitor the hybridization of two species in a two-component system. The so-called Competitive

Displacement Detection Method uses fluorescently labeled "competitor" target species that are mixed in with the sample species of interest [125]. As the competitor has a lower affinity than the target the binding kinetics of the competitor gives information about target binding via competitive displacement. The requirement for this method is real-time data collection. The authors demonstrated sensitivity of one tenth of competitor concentration with a dynamic range of two orders of magnitude [125].

Bishop [126] explored the theoretical limits of the effects of convective flux on the hybridization signal in a multi-component system as a function of target concentrations and sample volume, using a comparison between a single analyte system and a two-component system: a perfect-match and lower affinity mismatch target competing for hybridizing onto the same sensing surface (probe). They compared the effect of flow-through with that of recirculation on the time to 90% equilibrium in a single-component and two-component competitive system using a computational model similar to the one presented earlier [29]. Their results show that the enhancement of convective mass transport (defined as the reduction in time to 90% equilibrium) for a single analyte system is the upper limit for the achievable enhancement in a multi-component system [126]. This is not surprising, since convective flow would enhance reactions that are limited by mass transport instead of reaction kinetics, and competitive binding in a multi-component system where probe sites are not in an overwhelming excess, contributes to the overall rate of target capture [126]. However, no experimental validation of the predictions was presented. Also, it is curious that the authors chose to focus on time to equilibrium as a metric, given that their later work is focused on early hybridization kinetic effects and argues that early phase kinetics can provide information about the composition of the hybridizing sample.

In a paper from 2008, Bishop *et al.* [123] proposed a framework to consider multiplex hybridization in a general sense. They considered a general model of hybridization of N different species onto a single probe with corresponding solution concentrations and kinetic constants, and proposed a simplified three-component model of multiplex hybridizations, arguing that it is sufficient to describe the kinetics of hybridization with only three components in mind: the perfectly matched target, a competitor of high affinity (typically one-base mismatch), and a low affinity "non-target" representing background [123]. The ratio of target and competitor dissociation constants was ~100. They moved away from the finite element modeling that the group employed in their first paper [29], and proposed a two-compartment model to include mass transport instead, with a mass transfer coefficient representing the diffusion of targets across the compartment interface [123]. First, they simulated the hybridization of a perfect-match target, competitor and up to five background species with decreasing affinities onto the same probe surface, computing the composite curve (sum of all species bound) and individual hybridization curves. As already suggested by the group's previous publications, the highest affinity species of the system (in this case the target) is the only species to have a monotonously increasing binding curve, while the hybridization curves of the competitor and background species all exhibit a bi-phasic hybridization profile: after initial increase in the amount bound, the lower affinity species are displaced by higher affinity species in the second phase of the hybridization [123]. They then conducted experiments using a target species, fluorescently labeled competitor, and two background species with different affinities, varying the concentrations between experiments. This data was fitted to a three-component model, using known competitor concentration and known kinetics constants for target, but unknown target and background concentration and background kinetics [123]. The fitting

routine obtained target concentrations within 25% of experimental concentration. This is an exciting demonstration that the concentration of the target can be evaluated by monitoring the hybridization curve of a competitor. As the authors point out, this approach can eliminate false positives, while not requiring labeling of the target, and reduces assay time as it does not rely on reaching thermodynamic equilibrium [123]. It and the work of Dai *et al.* [30] are among the few studies to demonstrate improved microarray signal specificity using real-time binding kinetics. However, there are several important issues to keep in mind while interpreting the significance of the results of the study. First, as the authors themselves admit, this approach may not work when the concentration of the background is considerably higher than that of the competitor as they might outcompete the competitor. Second, the approach would also heavily rely on proper and optimal design of competitors: this method is relying on the competitors to be the highest affinity target to the probe after the perfect-match target of interest. Third, while this study is constrained to looking at the binding of multiple targets onto one probe, it is important not to lose sight of the notion that part of the promise and beauty of microarray technology is the parallel probing of the sample for hundreds or thousands of different targets. It does not sound very feasible to include a fluorescently labeled competitor species for each and every one of the probes. And lastly, the authors do not discuss the sensitivity of the method to different relative ratios of probe/target/competitor concentrations. Biphasic behavior of the competitor species in the presence of the target will only occur if probe sites are limited. Brief discussion of the effect of relative target/probe concentrations is offered by this group in [125], using experiments with two different competitor concentrations to show that the lower limit of detection for the target depends on the competitor concentration.

The competitive models have been proposed by a few more groups. Horne, Fish and Benight published a comprehensive analytical framework describing equilibrium and reaction kinetics of DNA multiplex hybridizations. Their work extends that of Bishop *et al.* [29] by moving from focusing on the binding of two targets on one probe to considering a system of any number of probes and targets. The number of probes and targets is limited only by computational power to solve the set of equations. However, their system does not consider transport effects. Horne *et al.* present numerical solutions of their model, which highlight the importance of relative concentrations of probe and target, and the relative stabilities of perfect-match and mismatch duplexes [122].

Chechetkin employed a two-compartment model for a two-component hybridization problem and presented approximate analytical solutions to the two differential equations, showing that the non-monotonous growth of the fraction of non-specific species bound on the probe is a feature under both reaction-limited and diffusion-limited scenarios [127]. In yet another paper, Chechetkin investigated the effect of geometry and shape of binding elements on the kinetics of the binding between the target and probe molecules, again simplifying a set of differential equations and obtaining an analytical solution [128]. He considered probes immobilized on the surface of hemispherical and flat circular cells, as well as hemispherical gel pads and gel slabs. Comparing the binding kinetics in the latter two, as they are the most relevant to this work, Chechetkin found that the binding kinetics for hemispherical pads should provide for faster binding kinetics. The overall conclusion from his work was that the shape and geometry of the microarray binding elements can have a significant influence on the binding kinetics and therefore influence the measurement from a microarray in the transient regime [128].

Li *et al.* used an equilibrium model of competitive hybridization, considering specific and lumped non-specific hybridization (marked by one unified dissociation constant and species concentration) [119]. They employed a nearest-neighbor based probe-specific dissociation constant and four global parameters that were fitted to the Affymetrix Latin Square spike-in dataset: initial probe concentration, a lumped parameter of non-specific binding, association constant, and detection coefficient of fluorescence [119].

1.9.4 Understanding surface effects in microarray hybridization: steric hindrances and electrostatics

Experiments comparing the stabilities of DNA duplexes in solution and in planar microarrays have revealed the decreased stability of DNA duplexes on the surface. Several reasons for this have been proposed in the literature. Immobilization of one ligand restricts the ways in which the two ligands can interact with each other, compared to hybridization in solution, which can affect the reaction rate [20]. Furthermore, steric hindrances can cause a problem if immobilized oligonucleotides are too close to each other. And second, electrostatic repulsion due to the negatively charged backbone of DNA molecules can affect target hybridization. One of the first systematic studies of the effect of spacers on hybridization efficiency was conducted by Shchepinov *et al.*, showing that tethering immobilized DNA to a spacer of optimal length can dramatically increase hybridization yield [20]. Watterson *et al.* compared the dissociation profiles and T_m values of short perfect-match and one-base mismatch duplexes formed in planar microarrays to those in bulk solution at different ionic strengths for three different surface densities [129]. They found that while the general trends were the same with respect to ionic strength of the buffers and destabilization of a single mismatch, the T_m values in the solution were higher than in microarray experiments, and the presence of one-base mismatch was more

destabilizing for duplexes in which one strand was immobilized than in solution [129]. Watterson *et al.* also observed a broadening of the melt curve in microarray experiments compared to solution-based experiments [129]. The electrostatic effects are not thoroughly understood and the theoretical analysis of such phenomena has only been considered in a handful of papers. Vainrub and Pettitt were the first to propose a theory describing the repulsion between the immobilized probe layer and target DNA in a planar microarray [85], as an extension of their earlier work, characterizing the effect of electrostatic interaction between nucleic acids and a surface on the thermodynamics of hybridization [130, 131]. Vainrub and Pettitt introduced a modification into the Langmuir isotherm that accounts for the repulsion of the target DNA from the probe layer [85]. They found good agreement with experimental data from the literature, published by Peterson *et al.* [132] relating increased hybridization efficiency to decreasing probe density. Vainrub and Pettitt were able to also reproduce the broadening of the duplex melt curve and decrease in the T_m value [85].

Gong and Levicky [133] characterized experimentally the surface coverage as a function of probe density in a planar microarray at different ionic strengths, and observed three different hybridization regimes for the probe densities they tested. Non-hybridizing regime occurred at high probe densities and low ionic strengths, where the hybridization signal was below the detection limit (all points in 0.012 M buffer, and some measured at 0.037 M and 0.11 M). Suppressed hybridization regime corresponded to a region from lower probe densities at lower ionic strengths to higher probe densities at very high ionic strengths, and was characterized by the non-Langmuir dependence of percentage of probe coverage on probe density. And finally, a pseudo-Langmuir regime at low probe densities ($2 \cdot 10^{12}$ probes/cm^2) up to medium probe

densities (~$6 \cdot 10^{12}$ - $8 \cdot 10^{12}$ probes/cm^2) at high ionic strength (0.33-1.0 M) was characterized by a near independence of percentage of probe coverage on probe density, a property of Langmuir equation [133].

Wong and Melosh [134] proposed an electrostatic model of DNA surface hybridization, reproducing the three regimes proposed by Gong and Levicky [133]. The numeric model of Wong and Melosh accounted for surface density of probe coverage, salt concentration and applied voltage using a fully nonlinear modified Poisson-Boltzmann scheme, assuming that all probe molecules can be hybridized and any suppression of hybridization is purely due to an electrostatic barrier [134]. Wong and Melosh treated the negative electric fields resulting from charged DNA as a kinetic activation barrier that limits hybridization of DNA from solution onto surface-immobilized probes, and solved numerically a modified Langmuir equation using a hybridization-dependent association constant k_a [134]. Their simulations showed that optimal conditions for maximizing the hybridized target number density occurs at high probe concentrations, high salt concentrations, and high voltages, as opposed to the maximum occurring at some intermediate probe density in case of zero applied voltage [134]. An interesting side note from their model fitting was that mobile counterions appear to be excluded from the immobilized DNA layer at high probe densities (the simulations assumed that the charge on DNA was screened to 55% by Manning-condensed counterions) [134]. Lastly, focusing on dissociation instead of surface coverage during hybridization, Fuchs *et al.* conducted thermal dissociation experiments with planar microarrays at various salt concentrations, and obtained good agreement between the data and a model consisting of a modified Langmuir curve similar to that proposed by Vainrub and Pettitt [84].

1.9.5 Concluding remarks

In conclusion, all these theoretical and computational investigations into DNA binding have highlighted the power of such methods towards optimizing DNA microarray performance. Of particular importance is the finding, both computational and experimental, of the existence of bi-phasic binding curves in competitive binding situations. However, temperature-dependent hybridization or non-equilibrium thermal dissociation have not been modeled in detail and compared with hybridization data in terms of their ability to discriminate between targets of varying affinities, using experimental and computational methods. This would add to the discussion on the future of microarray technology. Furthermore, in this work, we are investigating the possibility of composing a physico-chemical model of a three-dimensional gel-based microarray platform that agrees with experimental data using nearest-neighbor thermodynamics-based understanding of probe/target binding kinetics, and integrating that with diffusion of targets in solution and in three-dimensional binding elements.

Chapter 2: Mathematical model for DNA duplex melting in solution

2.1 Objectives

The ability to accurately describe DNA duplex formation kinetics in mathematical terms is essential to achieve the project goals. The microarray model will be a tool we will be using to simulate hybridization and melt patterns of different oligonucleotide sequences in potentially different buffer compositions, and the dependency of hybridization and melt curves on duplex sequences and buffer composition will be introduced through the mathematical description of binding kinetics. The objectives of this Chapter are to

a) develop a mathematical model of binding kinetics;
b) develop the formalism for introducing sequence-, temperature-, and buffer composition dependency into the model;
c) conduct solution-based DNA melting experiments, monitoring DNA hyperchromic effect with a UV spectrophotometer, to gather experimental data to validate model simulations with different DNA oligonucleotides, including perfect-match, one- and two-base mismatch duplexes in buffers of different salt and formamide concentrations;
d) validate the model using the experimental data, making adjustments in the model as necessary.

The model will be used as the basis for hybridization and dissociation modeling in the microarray 3D model in Chapter 3.

2.2 Theory & Design Considerations

The DNA hyperchromic effect has been the basis of interrogating DNA stability as a function of temperature throughout the study of DNA thermodynamics. Hyperchromicity refers to an increase in the absorbance of a material. DNA hyperchromicity refers to the strong increase in UV absorbance of DNA as the double-helix separates into single strands, for example as a response to increasing temperature or adding a denaturant, and the bases become unstacked. The aromatic bases in DNA interact via their electron π clouds when stacked together in the double helix, and this stacking in turn limits the absorbance as the latter is due to π electron transitions by limiting the potential for these transitions [80]. The peak of DNA absorbance falls around 260nm, for both double-helix and random coil conformations.

We chose to use DNA oligonucleotide melting in solution as an experimental system to support the description of binding kinetics mainly for two reasons. First, analyzing experimental melt data from a microarray platform is part of the objective of this work. Second, it is widely accepted as an assumption in the field of computational modeling of DNA oligonucleotide binding and microarray systems that the association constant in duplex formation is largely independent of oligonucleotide sequence and temperature within the range studied in this work [29]. For example, the changes in standard free energy of binding as a result of considering a one-base mismatch duplex instead of a perfect-match duplex, or the change in the binding free energy due to increasing the temperature of the system, are assumed to be mainly manifested through corresponding changes in the dissociation constant, while the association constant remains unaltered [29]**.** The thermodynamic parameters characterizing the stability of different

DNA duplexes will be calculated using the unified Nearest-Neighbor parameters proposed for perfect-match duplexes by SantaLucia [66], by Peyret *et al.* for internal mismatches [71].
It is important to note the difference between the quantity modeled and experimental measurements in this Chapter. The mathematical model describes the decrease in the concentration of the duplex as a function of increasing temperature. The experimentally measured increase in absorbance as a function of temperature represents both unstacking of bases and breaking of bonds between bases, meaning that even a solution of single-stranded non-selfcomplementary DNA oligonucleotides would exhibit an increase in absorbance (see for example [135]). Hence there is no direct one-to-one correspondence between the experimental and numerical melting curve. Instead, agreement between the model and experiments is defined as the agreement of measured and computed melting temperature T_m as a function of duplex sequence composition and buffer composition.

2.3 Materials & Methods

Experimental

DNA oligonucleotides and buffer compositions

Experiments were performed with synthetic custom-ordered oligonucleotides (Eurofins MWG Operon, Huntsville, AL), representing probes and their corresponding matches or mismatches on the microarray platforms that are used to collect experimental data for subsequent chapters. Table 2 displays the four sets (Set 1, Set 2, Set 3 and Set 4) of sequences of oligonucleotides used in this study. In subsequent discussion about duplex melting, the 5'→3' and 3'→5' strands in the duplex will be separated by a slash in the notation: for example, 1537/1271 will refer to 5'→3' strand of 1537, and 3'→5' strand of 1271.

Table 2 – DNA oligonucleotide sequences used in this Chapter. Mismatch bases are highlighted in red and underlined.

Sequence code	5'→3' sequence
Set 1	
EUB338	GCTGCCTCCCGTAGGAGT
EUB338MM	GCTGCCTCCCCTAGGAGT
Set 2	
1537	CTCACACACGTTCTTGACT
1271	CTCACACACGTTCTTCACT
1538	CTCACACACCTTCTTGACT
Set 3	
62	GACGGGCGGTGTGTACAA
399	GACGGGCGGTGTCTACAA
Set 4	
323	CCACAGCCTTTTACTTCAG
1282	CCACACCTTTTACTTCAG

Hybridization and melt were carried out in a buffer containing 10 mM Tris, and varying amounts of EDTA, NaCl, and formamide (as specified for each data figure or table below). Although some of the microarray buffers used in subsequent Chapters also contain EDTA, in view of minimizing the components of the buffer the model would explicitly have to account for in this Aim, we collected experimental melt curves using buffer containing only Tris and NaCl.

Hybridization and melt protocols

Total concentration of oligonucleotides in the melting solution was 4 μM: either 2 μM each of the two strands expected to form a perfect-match or a mismatch duplex, or 4 μM of single-stranded DNA, used as a control for certain experiments. The concentration was chosen to be in the linear range of the spectrophotometer. Hybridization prior to starting the melt experiments was carried out by two methods: a) adding the oligonucleotides to the buffer, vortexing, and waiting for 30 minutes, making sure that the absorbance reading on the spectrophotometer was stabilized before starting the melt – all at room temperature (RT); 2) adding the oligonucleotides

to the buffer at RT, heating the mixture rapidly up to 90 ^{o}C on a custom-made heating block designed to fit 1.5 ml Eppendorf tubes (courtesy of Peter Kauffman), letting the mixture stay at 90 ^{o}C for 5 min., and then letting it slowly cool back to RT in the span of ~70 min., and starting the temperature ramp roughly 20 min. after the end of the cooling, again making sure that absorbance reading had stabilized.

Melt experiments were carried out with HP 8452A diode array spectrophotometer with an HP 8909A Peltier temperature control accessory, using a quartz cuvette with 10 mm pathlength. Absorbance at either 260 nm or 270 nm (buffer containing formamide) is reported, using buffer absorbance as reference. The temperature ramp was chosen to be the same as the ramp rate used in microarray experiments: 0.83 ^{o}C/min., scanning from 20 ^{o}C to 70 ^{o}C in the span of 60 minutes. Temperature setting of the Peltier temperature control accessory of the HP was manually changed either after every two minutes or a minute through-out the temperature scan. The temperature of the solution in the cuvette was independently measured for calibration: the temperature in the cuvette was shown to lag behind the input temperature by ~1.5 ^{o}C. This correction is introduced in the reported experimental T_m values, *i.e.*, the reported T_m values are the T_m values of raw data minus 1.5 ^{o}C.

Data analysis

Data were analyzed in MATLAB, and first order derivative was calculated using the finite difference approximation, employing forward differences. T_m was calculated as the location (temperature value) of the peak of the first order derivative. Since numerical derivatives are known to be very sensitive to noise, cubic spline smoothing of the original raw data was used in

MATLAB for some of the experiments before computing the derivative. For each smoothed curve, the adequacy of the fit was determined by minimizing simultaneously the mean difference and absolute difference.

Computational

Mathematical description of duplex formation and dissociation

In this Chapter, we are considering the dissociation of oligonucleotide duplexes as a function of duplex sequences and buffer composition. We assumed a two-state model of DNA binding and dissociation, and no formation of secondary structure [29, 113, 116, 118, 120, 121]. The hybridization between two oligonucleotides is described by the following formalism:

$$C + P \underset{k_d}{\overset{k_a}{\rightleftharpoons}} B$$

where C and P are the concentrations of the two oligonucleotide strands, B is the oligonucleotide duplex, k_a association rate constant, and k_d dissociation rate constant for the reaction. The change in duplex concentration and single strand concentration over time can then be described by the following system of differential equations, assuming a well mixed solution:

$$\begin{cases} \frac{dB}{dt} = k_a \cdot C \cdot (P_0 - B) - k_d \cdot B \\ \frac{dC}{dt} = -\frac{dB}{dt} \end{cases} \quad (1)$$

where P_0 is the initial single strand concentration.

Temperature-dependence of the rate of the binding reaction for thermal dissociation modeling was introduced to the model through the dissociation constant, since empirical studies have found the association constant to be fairly insensitive to temperature in the temperature range of interest:

$$k_d = k_a \cdot e^{\frac{-\Delta G}{RT}} = k_a \cdot e^{\frac{-\Delta H + T\Delta S}{RT}},$$

where G is binding free energy, H is enthalpy, S entropy, R universal gas constant, and T temperature [29]. ΔH and ΔS for a given DNA oligonucleotide sequence can be estimated thermodynamic nearest-neighbor parameters [66], see also Appendix section A.3 for a table of nearest-neighbor parameters for perfect-match duplexes. Nearest-neighbor parameters for internal mismatches were obtained from [71]. Since the two-base mismatch duplex considered has the two internal mismatches five bases apart, the effect of these mismatches was considered to be independent of each other. Dependence on salt concentration for both perfect-match and mismatch duplexes was introduced through modifying ΔS while keeping ΔH independent of salt concentration:

$$\Delta S^0([Na^+]) = \Delta S^0(1M\ NaCl) + 0.368 \times N \times \ln[Na^+],$$

where $\Delta S^0([Na^+])$ is the entropy change for an oligonucleotide duplex dissolved in a given sodium concentration, ΔS^0(1 M NaCl) is the entropy change predicted from the unified nearest-neighbor parameters for 1 M NaCl, and N is the total number of phosphates in the duplex divided by 2 [66]. The values for duplex stability, ΔS and ΔH, are given in Appendix, section B for all the duplexes investigated in this Chapter.

No explicit way to directly account for the effect of formamide on DNA thermodynamic parameters ΔH and ΔS has been reported. Dependence of melt temperature on formamide was introduced into the model simply via a temperature offset of 0.6 ^{o}C per percent formamide: this is a typical offset reported for T_m in experiments where formamide is present [17, 90, 96, 99]. Also, no sequence-dependent effect on the formamide was introduced in the model as there are contradicting reports on that in the literature [17, 90, 99, 100].

Numerical model

The system of differential equations (1) was implemented in MATLAB. The concentration of single-stranded DNA *C* was assumed to be 0 at time t=0, and the concentration of bound duplex *B* was assumed to be 2 μM. The implication of the assumption that all oligonucleotides were bound in a duplex at time t=0 was tested by running the simulations with initial conditions of a) 50% of DNA bound in a duplex at time t=0; and b) 10% of oligonucleotides bound in a duplex at time t=0. This was done for 1537/1537 and 1538/1537 duplexes. For both cases, the largest difference in the calculated T_m value from simulations with these varying initial conditions was ~0.0001% of the T_m value calculated using the original assumption for initial conditions. The set of equations was solved using MATLAB's *ode23* solver, which is an explicit Runge-Kutta method implementation based on Bogacki and Shampine's method [136]. The solver was run for time points between 0 and 3600 s. Temperature was calculated at each time point as a function of time, assuming the same ramp rate as in the experimental system: $T=273+t\cdot 5/360$, where T is temperature in Kelvin and t time in seconds. The value of the dissociation rate constant k_d was calculated as a function of temperature at each time point.

2.4 Results & Discussion

Comparison of pre-melt hybridization protocols

As a first part of the study, we compared the effect of pre-melt hybridization protocol on the melt data as outlined above: hybridizing at RT after vortexing the solution of oligonucleotides in the buffer (subsequently referred to as "RT hybridization" protocol), or bringing the mixture up to 90 ^{o}C for 5 min., and slowly cooling down to RT before starting the melt (subsequently referred to as "slow cooling" protocol). The comparison of solutions containing only one single-stranded

DNA versus a mix of complementary DNA was conducted to ensure that any temperature-dependent absorbance observed is not simply due to the base unstacking that is characteristic of single-stranded DNA as well. Figure 7 displays the comparison of melt curves after conducting the pre-melt hybridization either at room temperature, or through slow cooling from 90 ^{o}C. First, note that both single-stranded and double-stranded DNA exhibit an increase in absorbance as a function of increasing temperature, but the curve for double-stranded DNA includes a section with a sharper rise than that of the single-stranded DNA. Second, note that the melting curves are affected by the pre-melt hybridization protocol, however, the effect is most evident when looking at the single-stranded DNA melt curve.

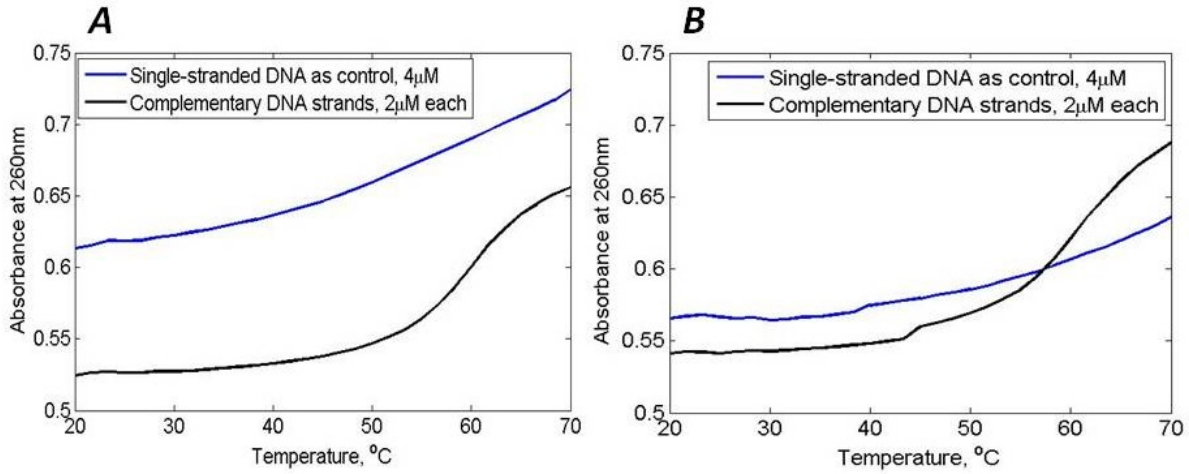

Figure 7 – Effect of pre-melt hybridization protocol on the melt curves of single-stranded and double-stranded DNA oligonucleotides: targets mixed at room temperature, vortexed, and melt started after 30 min. (panel A), and targets heated up to 90 ^{o}C, and slowly cooled to room temperature (panel B). Buffer composition: 10 mM Tris, 50 mM NaCl, 1 mM EDTA. Absorbance plotted with buffer as reference. Target sequence used: EUB338 and its perfect complement for complementary DNA, and EUB338(5'→3') for single-stranded DNA. Temperature ramp rate: 0.83 ^{o}C/min.

Figure 8 displays the effect of pre-melt hybridization treatment on normalized melt curves of single-stranded and double-stranded DNA. In addition to the EUB338 single-strand and complementary duplex melt displayed in Figure 7, Figure 8 shows the melts of single-stranded EUB338MM and EUB338/EUB338MM duplex. Note that here and below in referring to

duplexes, we mark the 5'→3' strand on the left side of the forward slash '/', and the 3'→5' sequence on the right.

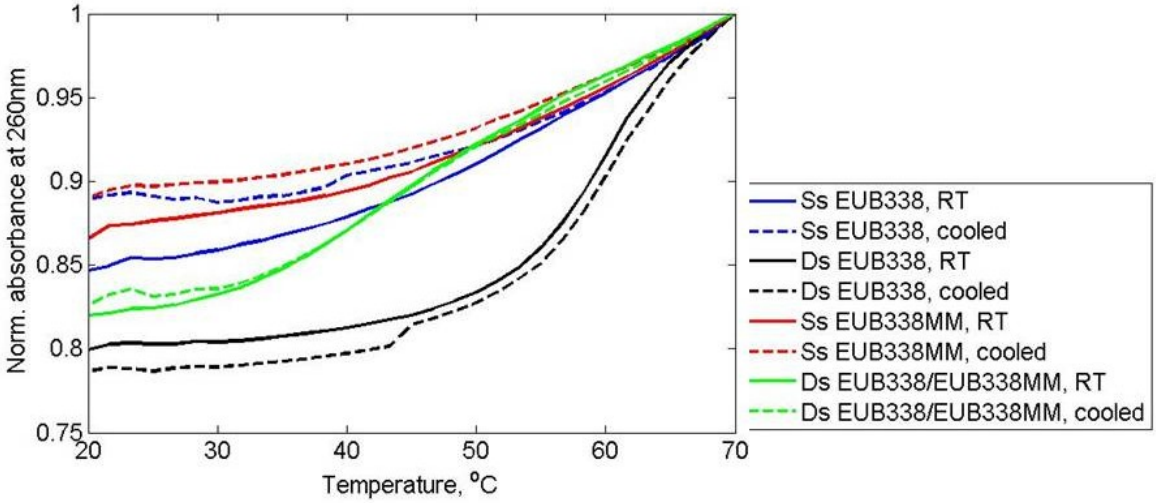

Figure 8 – Effect of pre-melt hybridization protocol on the normalized melt curves of single-stranded and double-stranded perfect-match and mismatch DNA oligonucleotides: targets mixed at room temperature ('RT' in legend), and slowly cooled ('cooled' in legend). Ss stands for single-stranded, and Ds for double-stranded. Buffer composition: 10 mM Tris, 50 mM NaCl, 1 mM EDTA; total oligonucleotide concentration 4 μM (2 μM of each complementary strand, or 4 μM of single-stranded DNA). Absorbance plotted with buffer as reference. Temperature ramp rate: 0.83 °C/min.

While the curves of single-stranded DNA after RT and slow cooling hybridization are not quite overlapping, the agreement between double-stranded DNA melt curves is very good. This allowed us to consider in the subsequent work only RT hybridization protocol.

Sequence effects on experimental and computational melt curves

We collected experimental melt data, using targets from Sets 1, 3, and 4, which rendered perfect-match and one-base mismatch duplexes, and targets 1537, 1271 and 1538 from Set 2, which rendered perfect-match duplexes, two different one-base mismatch pairs (1537 and 1271, 1537 and 1538), and one two-base mismatch pair (1271 and 1538). The particular mismatches displayed were: one internal G·G mismatches for EUB338/EUB338MM, 62/399, 323/1282 and 1537/1271, one internal C·C mismatch for P1538/P1537, and two internal mismatches for

1538/1271, G·G and C·C. Figure 9 shows the melt curves of 1537, 1271 and 1538 duplexes, and the first order derivatives of these melt curves. Note the expected overlap of 1537 and 1271 perfect-match duplexes, the two mismatches (1537/1271 and 1537/1538) being of different stability, manifested by shifting of the melt curves along the temperature axis, and the two-base mismatch being of the lowest stability. The melt curves of Set 1 are discussed later as part of investigating the effect of different salt concentrations.

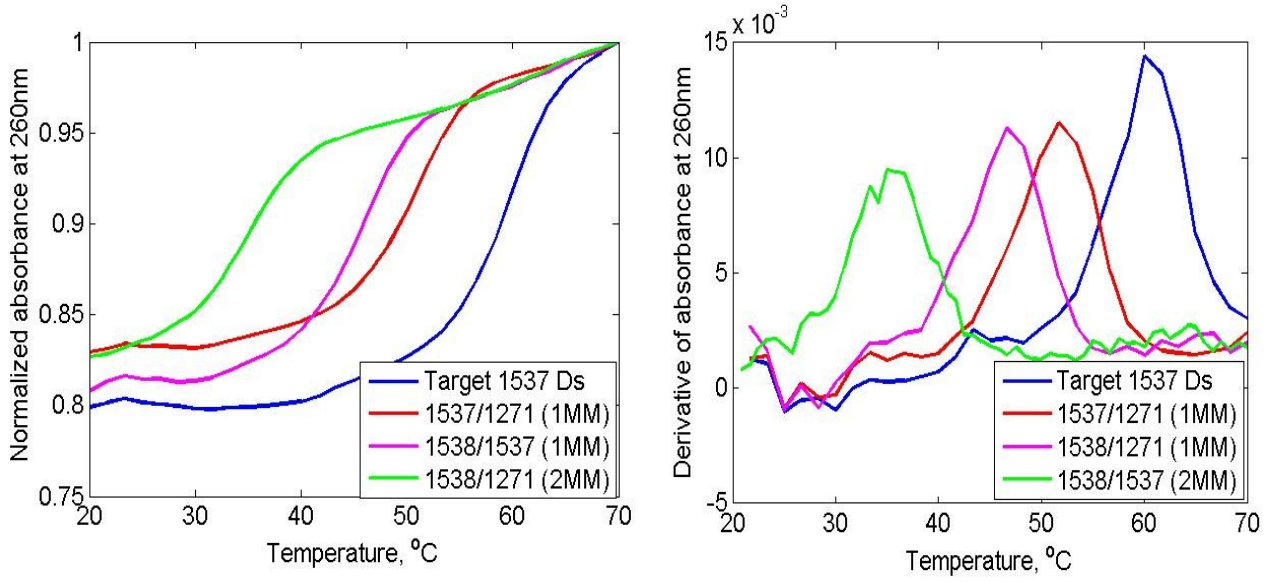

Figure 9 – Normalized melt curves (left) and first order derivative of melt curves (right) for targets 1537, 1271, and 1538. Buffer composition: 10 mM Tris, 50 mM NaCl; total oligonucleotide concentration 4 µM (2 µM of each complementary strand). Absorbance plotted with buffer as reference. Temperature ramp rate: 0.83 °C/min.

As outlined above, sequence- and temperature-dependence of the affinity of the DNA duplex was introduced into the model through sequence-dependent thermodynamic parameters used to calculate the dissociation constant. Figure 10 displays the dissociation constants as a function of temperature for the duplexes used for experiments. Note the larger difference between kinetic constants at lower temperatures compared to elevated temperatures.

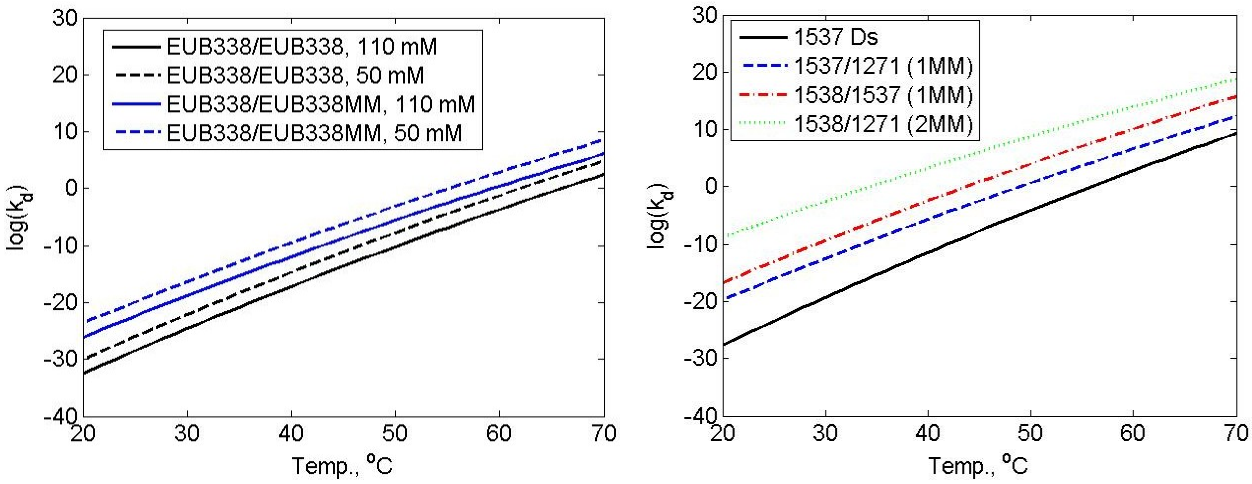

Figure 10 – Theoretical dependence of the dissociation constant on temperature, sequence, and NaCl concentration, for Set 1 (left panel) and Set 2 duplexes (right panel). NaCl concentration for the left panel was 50 mM.

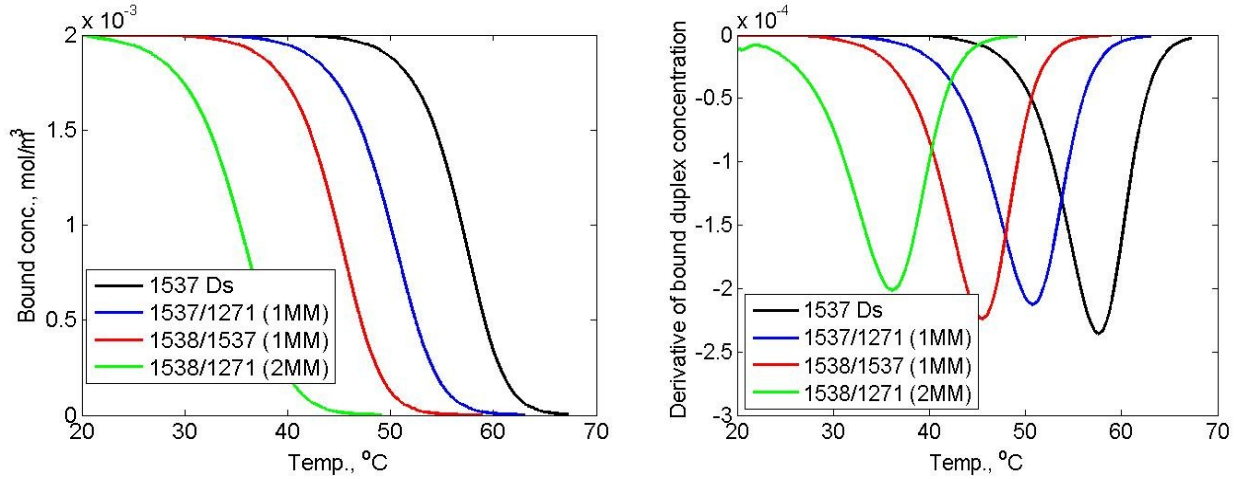

Figure 11 – Simulated melt curves for targets 1537, 1271 and 1538, Ds stands for double stranded. Temperature ramp rate: 0.83 °C/min. Assumed salt concentration: 50 mM.

Figure 11 shows the simulated melt curves for Set 2 (1537, 1271 and 1538), Set 1 melt curves will be discussed below. Note that unlike absorbance-based melt data, the variable plotted from the simulations is the bound concentration of duplex, which decreases with increasing temperature. Table 3 presents the computed and experimental T_m values for Set 2. Note the close correspondence between the values.

Table 3 – Comparison of computed and experimental T_m values for Set 2 targets.

T_m, °C	1537/1537	1537/1271	1538/1537	1538/1271
Experimental	58.5 ± 0.4	50.2 ± 0.7	45.2 ± 0.4	33.5 ± 0.8
Computational	57.6	50.7	45.5	36.1

Salt effects on the experimental and computational melt curves

Next, we explored the ability of the model to account accurately for different salt concentrations by conducting melting studies at two different salt concentrations and comparing the T_m values with computed values. 110 mM NaCl concentration was chosen as the highest salt concentration to still render a melting temperature within the range of the temperature controlled for the spectrophotometer (70 °C). Figure 12 displays the effect of increased salt concentration on Set 1 duplexes. Note the dissociation occurring at a higher temperature for a higher salt concentration.

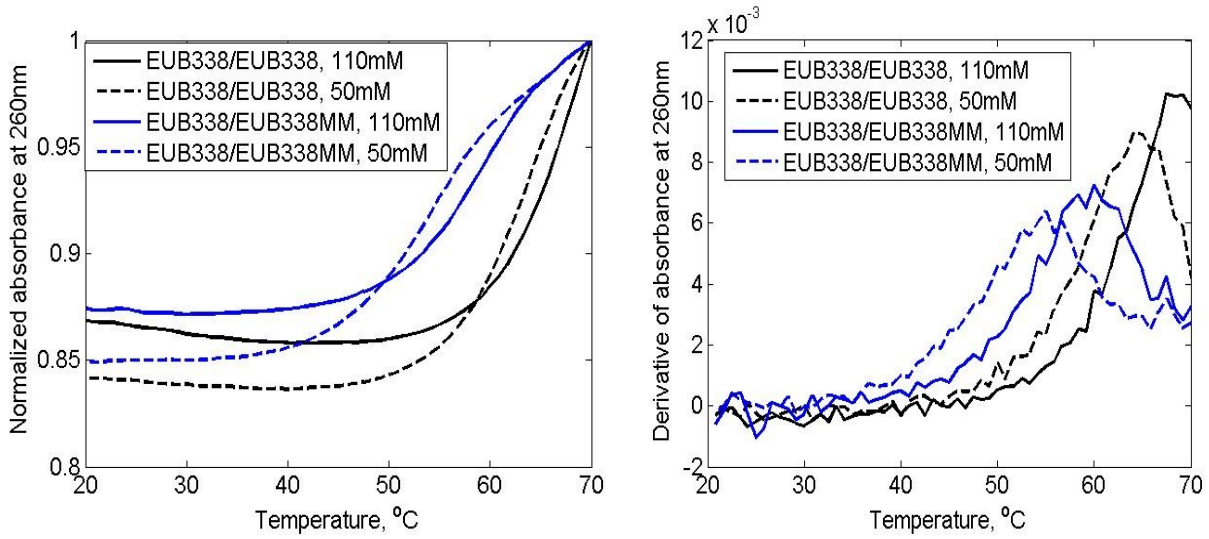

Figure 12 – The measured melt curves and first order derivative of EUB338/EUB338 and EUB338/EUB338MM duplexes at two different salt concentrations. Total oligonucleotide concentration 4 μM (2 μM of each complementary strand). Absorbance plotted with buffer as reference. Temperature ramp rate: 0.83 °C/min. Buffer also contained 10 mM Tris.

Figure 13 displays the simulated melt curves and first order derivatives for Set 1 duplexes.

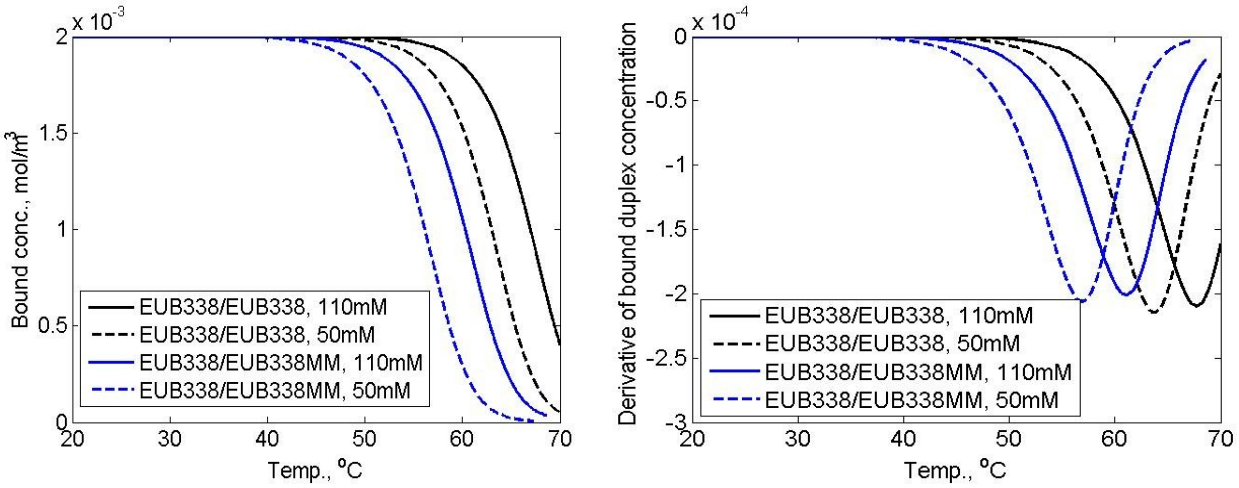

Figure 13 – Simulated melt curves for targets EUB338 and EUB338MM at two different salt concentrations. Temperature ramp rate: 0.83 °C/min.

Table 4 – Comparison of computed and experimental T_m values for Set 1 at two different salt concentrations (n=3).

	EUB338/ EUB338		**EUB338/ EUB338MM**	
T_m, °C	**50 mM**	**110 mM**	**50 mM**	**110 mM**
Experimental	62.7 ±0.7	67.4 ±0.5	53 ±1.9	57 ±1.8
Computational	63.7	67.7	56.9	61.1

Table 4 summarizes the T_m values of the two sets of oligonucleotides at two different salt concentrations: 50 mM and 110 mM. Note again the close agreement between experimental and computed T_m values for EUB338/EUB338 duplexes. The model is slightly overestimating the T_m of EUB338/EUB338MM duplex, but also the experimental data is less consistent and has a larger standard deviation. Note that while the first order derivative peaks are very narrow for 1537 and its respective mismatch duplexes (Figure 9), the peaks are broader for EUB338 duplexes. This might contribute to the larger standard deviations in the experimental data, and perhaps between the experimental data and model simulations.

Formamide effects on the experimental and computational melt curves

As a first step, the absorbance spectra of buffers containing 0-40% formamide were measured. Formamide has a very high absorbance around 260 nm, resulting in absorbance flattening, so even buffers containing very low percentage of formamide were not suitable to track DNA duplex related absorbance changes at 260 nm. Furthermore, as the concentration of formamide increased in the buffer, the first wavelength at which formamide's effects were negligible, kept increasing. Therefore, absorbance values at 270 nm were used in the subsequent DNA melting experiments to study the effect of formamide, and only 10% and 20% formamide concentrations were considered. Figure 14 displays the melting curves for 0, 10 and 20% formamide at 50 mM NaCl.

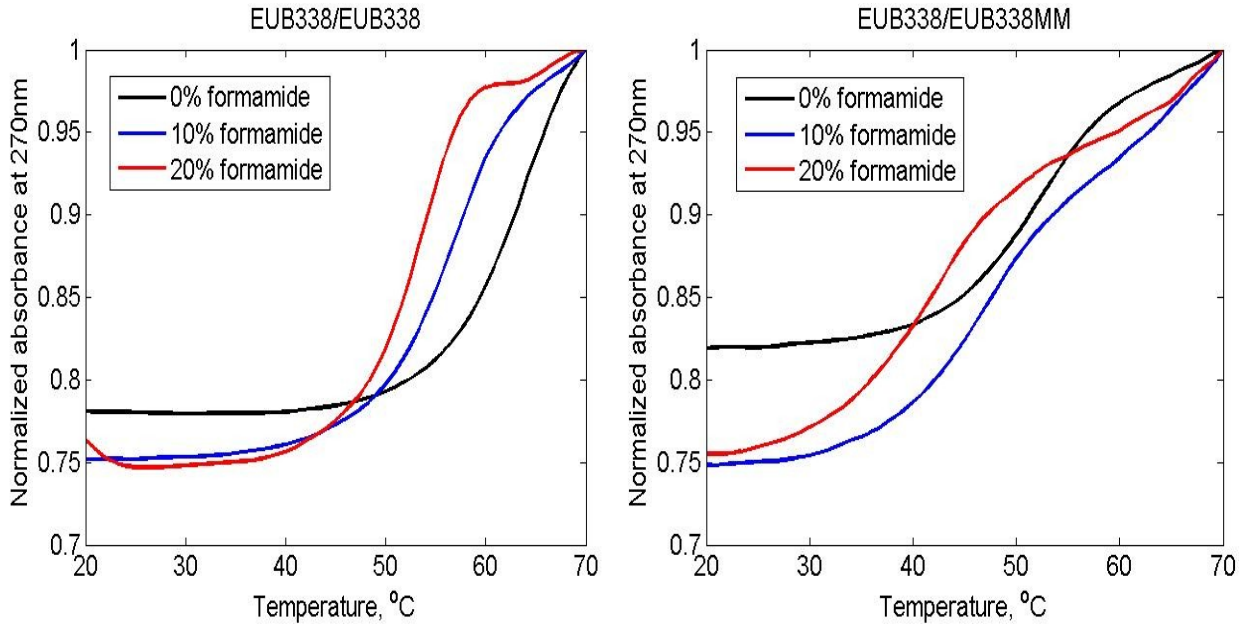

Figure 14 – Effect of formamide concentration of melt curves for EUB338/EUB338 and EUB338/EUB338MM duplexes. Temperature ramp rate: 0.83 °C/min. Total oligonucleotide concentration 4 μM (2 μM of each complementary strand). Absorbance plotted with buffer as reference. Buffer composition beyond formamide: 10 mM Tris, 50 mM NaCl.

Table 5 displays the experimental and computed T_m values for Set 1 duplexes at different concentrations of formamide. Just like already commented above, the model is slightly underestimating the T_m of EUB338/EUB338MM duplex.

Table 5 – Comparison of experimental and computational T_m values for Set 1 duplexes at different concentrations of formamide (n=3). Buffer also contained 50 mM NaCl and 10 mM Tris.

	EUB338/ EUB338			**EUB338/ EUB338MM**		
T_m, °C	0%	10%	20%	0%	10%	20%
Experimental	62.7 ±0.7	57.1 ±0.5	53.8 ±1.3	53 ±1.9	47.2 ±2.1	41.0 ±1.6
Computational	63.7	57.7	51.7	56.9	50.9	44.9

2.5 Conclusions

In this Chapter, we composed a simple kinetic model of oligonucleotide thermal dissociation and compared the results of the simulations with collected experimental data from solution-based melts. The simulations and experiments considered two different oligonucleotides sets, each containing a perfect-match and one-base mismatch duplex, and one of the two oligonucleotide sets also contained a two-base mismatch duplex. In addition to exploring different sequences, we varied salt and formamide concentration in the hybridization and melt buffer. Our results demonstrated that a kinetic model of oligonucleotide dissociation, based on thermodynamic nearest-neighbor parameters, is able to simulate sequence effects and effects of salt and formamide concentration. These results will form the basis of modeling sequence-dependent and buffer composition effects in subsequent Chapters.

However, two points are worth keeping in mind, as we transition from modeling solution-based dissociation to modeling microarray hybridization and binding. First, the thermodynamic

nearest-neighbor parameters used in this model are collected from solution-based experiments, and the stabilities of duplexes in a microarray platform could differ from those in the solution. However, there are published studies that have found generally good correlation between binding free energies in the solution and in microarrays [77, 78]. Second, as the experimental data focused on thermally introduced dissociation, we did not explicitly verify the assumptions in the hybridization model, most notably that the association constant is sequence-independent.

Chapter 3: Finite element computational model of single-target DNA hybridization and melt in a microarray platform with three-dimensional binding elements.

3.1 Objectives

This model will incorporate the kinetic description developed in Chapter 2, and expand from solution-based dissociation to hybridization and dissociation in a microarray platform. Desired features of this model will include the description of the three-dimensional geometry of DNA microarray platform, and the following physico-chemical processes: diffusion of DNA oligonucleotide target in bulk solution, diffusion of target into and in the gel elements, and the binding reaction as a function of temperature. In both experiments and simulations, we will focus on binding and hybridization of one single target analyte only.

The objectives of this Chapter are to

a) collect experimental data of single target hybridization and melt in gel-based microarrays, using different buffers;
b) use a one-dimensional model of hybridization and dissociation, developed in Chapter 2, to compare hybridization and dissociation predictions with experimental data;
c) develop a 3D finite element model of single target hybridization and melt in a three-dimensional gel microarray;
d) calculate scaling factors for nearest-neighbor based ΔH^o and ΔS^o for use in the microarray model;
e) perform a constrained optimization to find the best match of model to experiment, and a sensitivity analysis to characterize the model sensitivity to experimental inputs that have

not been experimentally determined (probe density, diffusion coefficient in the gel elements);

f) conduct an *in silico* implementation and validation of image analysis strategies;
g) characterize inter- and intra-array reproducibility of hybridization and thermal dissociation metrics in experiments with three-dimensional gel-based microarrays
h) validate the model against single target hybridization experimental data.

This model will be the basis for developing a competitive binding model in Chapter 4.

3.2 Theory & Design Considerations

In this Chapter, we will compose and validate a model of a three-dimensional gel-based microarray. The model will simulate target hybridization and dissociation in three-dimensional gel probes, and incorporate the effects of diffusion in the gel elements and in the solution, and temperature. Figure 15 presents a simplified schematic of the geometries and physico-chemical processes to be modeled. This has not been done before according to the literature for three-dimensional arrays. Another novel aspect of this work is the investigation of the link between solution-based and microarray (planar or 3D) thermodynamic constants.

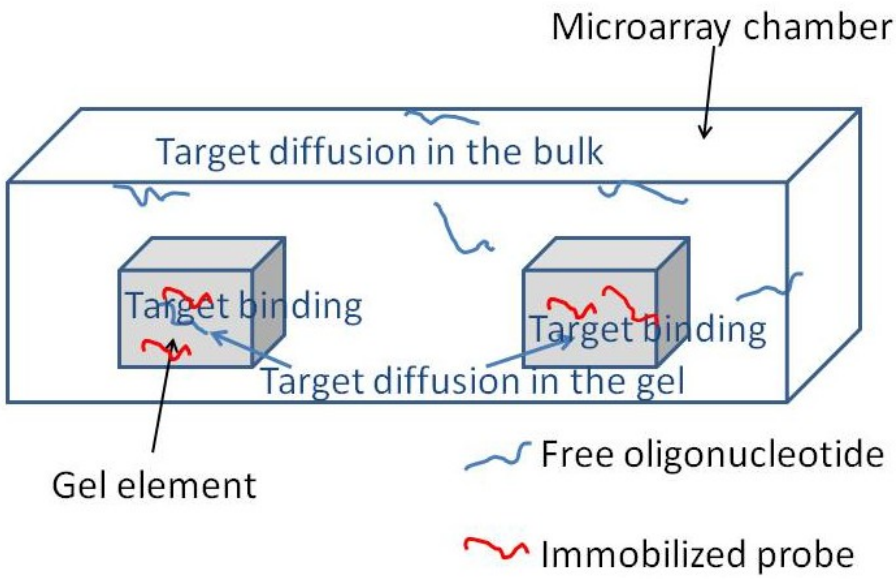

Figure 15- A simplified schematic of the geometries and processes included in the microarray model. Three-dimensional gel-based binding elements are embedded in a microarray chamber. Target oligonucleotides diffuse into the gel elements where they can bind to form complementary duplexes with probe oligonucleotides, immobilized in the gel matrix. Schematic not drawn to scale.

In describing the hybridization and dissociation kinetics, we will make use of the formalism developed and validated in Chapter 2 to describe the hybridization and temperature-dependent dissociation of DNA oligonucleotide complexes. Strategies for extending the solution-based model of Chapter 2 to a microarray context will be discussed below.

Thermodynamic parameters of duplex stability: solution versus microarray

One of the most important overall goals of this Chapter is to determine how the stability of duplexes, measured and simulated with a kinetic model in solution in Chapter 2, correlates with stabilities in the microarray. In Chapter 2, we utilized SantaLucia's unified nearest-neighbor parameters [66] to account for sequence-dependent duplex stability, along with a logarithmic correction to the entropy term based on NaCl concentration [66], and a constant offset in the temperature variable of the model to account for formamide-lowered stability of the duplex.

Several studies have found a general correlation between the binding free energies in the solution and in microarrays [74, 77, 78]. For example, Fotin *et al.* [74] compared the binding free energies of very short oligonucleotides in solution and in three-dimensional gel elements, and found the following relationship: $\Delta G_{solution}$: $\Delta G_{chip}=a\Delta G_{solution}+b$. However, it is not clear how the nearest-neighbor parameters ΔH and ΔS will need to be modified to account for binding in a microarray instead of solution, and whether they can be all scaled by the same factor irrespective of sequences.

Equal consideration should be given to extending the modeling of formamide and salt effects from solution to microarrays. The effect of formamide on decreasing the melting temperature of the duplex in microarrays has been investigated in two separate studies. Urakawa *et al.* found that the average decrease in T_d was 0.56 ^oC/%FA in a three-dimensional gel-based microarray, averaged over all target:probe duplexes, including mismatches [17]. They used 39-base targets hybridizing onto 19-base probes. Fuchs *et al.* used 15-16 base probes hybridizing with 15-20 base targets in planar microarrays, and found the average decrease in T_d to be 0.57-0.59 ^oC/%FA, depending on immobilization chemistry. Based on these two studies, we propose a simple offset in the temperature variable as a function of formamide concentration to account for formamide effect (*i.e.,* the temperature variable in the model will be lowered by a constant offset of 5.8 ^oC in a buffer of 10% formamide). However, the effect of salt concentration on duplex stabilities in a microarray can be more complex than that characterized in the solution. In the solution, the influence of salt concentration on binding free energy is described through a change in standard entropy: standard entropy increases proportionally with the natural logarithm of salt concentration. Only a handful of studies have looked at the effect of salt concentration in

microarray binding and dissociation, and the conclusions are not in agreement with each other. For example, Peterlinz *et al.* found that T_d followed a logarithmic dependence on NaCl similar to that in solution [86], while Fuchs *et al.* used a modified Langmuir isotherm including a term for electrostatic interactions to explain their data of dissociation curves in different salt concentrations [84]. Both studies were concerned with planar arrays only. Electrostatic effects have not been studied previously in a three-dimensional gel-based microarray, and investigating the effect of salt concentration on hybridization intensities and melting temperatures will be a contribution to the field.

Gel-based three-dimensional microarray platform

In this work, we are utilizing three-dimensional gel-based microarrays, which have not been studied as extensively as more widespread planar microarrays. It has been shown, however, to offer distinct advantages over more traditional planar microarrays, most importantly for the purposes of this Chapter, solution-like kinetics [8]. Figure 16 displays two images of two different formats of gel arrays used in this study, the gel pad and gel drop arrays. The gel pad array consists of an array of 100 μm × 100 μm × 20 μm polyacrylamide pads, polymerized in place before loaded with oligonucleotide probes. Polyacrylamide gel is activated by hydrazine-hydrate treatment, substituting some amide groups with hydrazide groups [25]. Oligonucleotides were immobilized through coupling with the hydrazide groups of the gel via dialdehyde groups, which were produced by activating the 3'-terminal 3-methyluridine on oligonucleotides by oxidizing with $NaIO_4$ [25]. A different manufacturing technique, more suitable for large-scale manufacturing, is used to fabricate the gel drop microarray where gel polymerization and probe immobilization steps are combined and take place simultaneously [18]. Polymerization solution

containing modified DNA is spotted onto glass slides for photoinduced polymerization [18]. This can render a more uniform distribution of probe molecules compared to gel pad technology [18]. In this Chapter, we have utilized both platforms for collecting experimental data. The effect of the geometry of a three-dimensional binding element was studied by Chechetkin [128], who developed an analytical solution to compare the binding kinetics of targets hybridizing with probes immobilized in a hemispherical gel drops and gel slabs (referred to as pads in the discussion above). He found the hemispherical geometry to provide much faster rate of hybridization, which is consistent with findings from a computational project by Mike Purfield in the Yager lab. Experimental comparison of these two gel microarray platforms is not within the scope of this work.

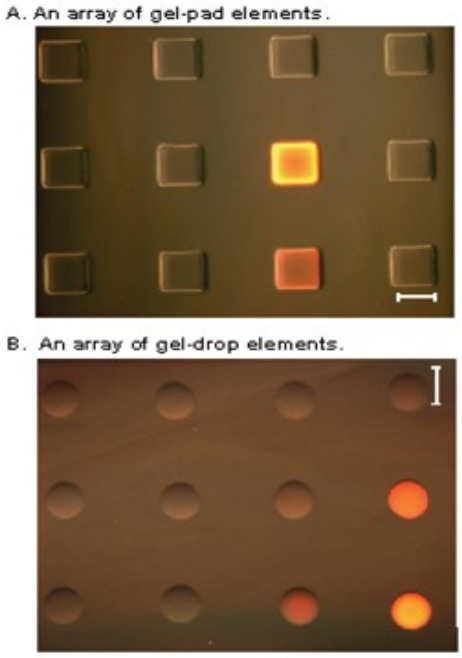

Figure 16 – Fluorescent and white-light image of Alexa 594-labeled oligonucleotides hybridized on binding elements in a gel-pad array (top panel, A), or gel-drop array (lower panel, B). Scale bar in both panels is 10 µm. Image from Starke *et al.* [7].

Approaches for computational modeling

We will be using a computational model similar to the one employed in Chapter 2, describing a well mixed solution, to initially compare the solution-based stability predictions with actual data from three-dimensional microarrays. The desired agreement between the model and experiment is defined as agreement between 1) absolute values of melting temperatures in non-equilibrium thermal dissociation (T_d, temperature at which the signal is 50% of its original value); 2) signal ratio from hybridizations with a single target on perfect-match and mismatch probes; 3) time constant of the hybridization reactions. We will then compose a three-dimensional finite element model of the microarray hybridization and thermal dissociation processes. We will also compose a compartmental model and validate it with the finite element model to facilitate rapid parameter optimization. While the three-dimensional finite element COMSOL model comprises physically realistic processes, it is computationally complex and not really suited for fast parameter optimization. A compartmental model has been used to study kinetics in SPR surface binding assays [137] and also in modeling hybridization on planar microarrays [123, 127].

In modeling the hybridization and dissociation in the microarray platform, we are only considering two probe elements at a time due to computational constraints, assuming the binding of one target analyte onto different probes in the array to be quasi-independent.

3.3 Materials and Methods

Experimental

Materials

Experiments were performed with synthetic custom-ordered labeled oligonucleotides (Eurofins MWG Operon, Huntsville, AL) that had perfect matches and mismatches among the probes displayed on the microarrays. The sequences of probes used on the three different microarrays are given in Table 6. The single target used presented a perfect-match to either probe EUB338, probe 1537, probe 62, or probe 323. The targets were labeled with either Texas Red or OregonGreen488 fluorophore. The original microarray hybridization buffer contained 900 mM NaCl, 20 mM Tris-HCl, and 40% formamide [52], while the original melt buffer contained 4 mM NaCl, 20 mM Tris-HCl, and 5 mM EDTA, but other buffer formulations were used in some of the experiments as specified and explained below. The three gel microarray platforms used in this Chapter, gel pad and two different generations of gel drop arrays (gel binding elements printed on a microscope slide), were obtained from Akonni Biosystems (Frederick, Maryland).

The gel drops of spherical cap shape in the gel drop array are 20-40 μm high, and have a diameter of 170 μm. The distance between the centers of gel drops is 300 μm. The standard oligonucleotide concentration for probe immobilization is 0.125 mM, out of which roughly 60 μM is expected to be functional upon immobilization. The volume V of the spherical cap can be calculated as follows:

$$V = \frac{\pi h}{6}(3a^2 + h^2),$$

where h is the height of the spherical cap and a the radius. Therefore, the volume of a spherical cap on a gel microarray with the radius of 85 μm and height of 30 μm is $3.55 \cdot 10^{-13}$ m^3.

Assuming that the height of the drops varies between 20 and 40 µm, the volume will vary between $2.31 \cdot 10^{-13}$ m^3 and $4.87 \cdot 10^{-13}$ m^3.

Two different generations (the word 'generation' here refers to arrays manufactured at different times, not to actual differences in the general manufacturing process) of gel drop arrays, both manufactured by Akonni Biosystems (Frederick, Maryland), were used in this Chapter. While the first generation of gel drop arrays was fully populated by gel drops with immobilized probes in them and had been designed by the Stahl lab, the second generation of arrays was designed specifically for this project and contained a mix of empty gel drops without probe and probe-immobilized gel drops to cut down on the cost of the array. One microarray slide consisted of a 9 mm × 9 mm printable area, divided into four quadrants (see Figure 17 for details). Each quadrant contained two beacon probes, one immobilized with Texas Red and the other with Oregon Green 488 fluorophore.

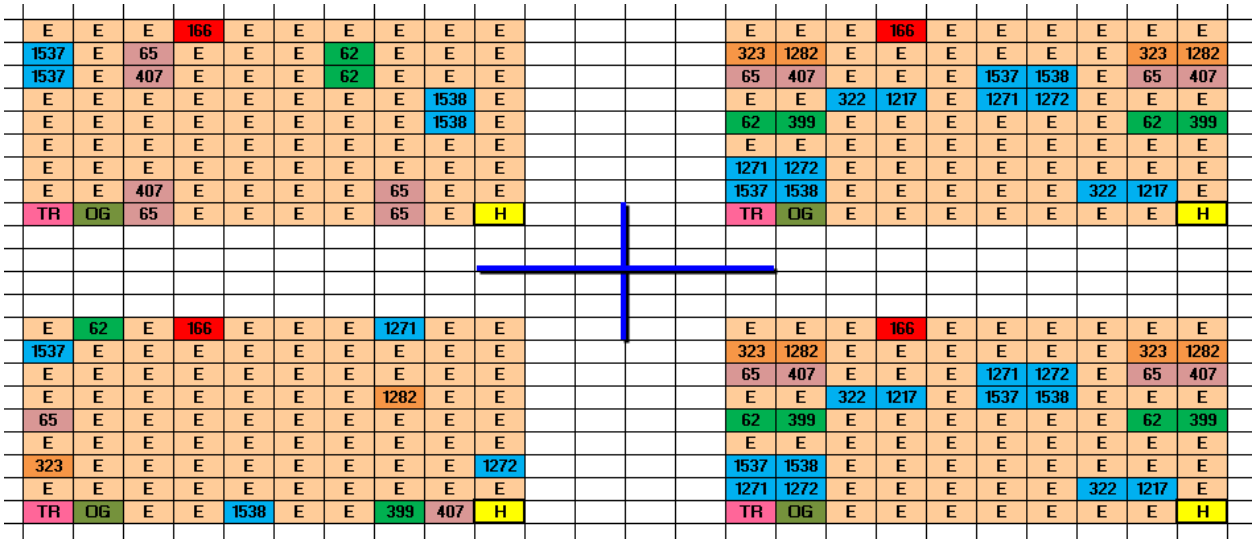

Figure 17 – A map of the second generation gel drop array used in Chapters 3 and 4 of this work. The letter 'E' refers to empty gel drop (*i.e.* no probe immobilized in it), and numbers refer to different probes as listed in Table 6. 'H' and '166' refer to manufacturer placed hybridization control probes. Each color refers to 'related' probes, *i.e.* a perfect-match probe and its 1- or 2-base mismatch companions. Total printed area of the array is 9 mm × 9 mm.

Table 6 - DNA oligonucleotide probe sequences used in Chapter 2. Mismatch bases are highlighted in red and underlined.

Sequence code	5'→3' sequence
Gel pad array	
EUB338	GCTGCCTCCCGTAGGAGT
EUB338MM	GCTGCCTCCCCTAGGAGT
EUB338III (2MM)	GCTGCCACCCGTAGGTGT
EUB336	TGCCTCCCGTAGGAGTCT
EUB336MM	TGCCTCCCCTAGGAGTCT
Spir	CTTARCTGCTGCCTCCCG
SpirMM	CTTARCTCCTGCCTCCCG
Gel drop array	
1537	CTCACACACGTTCTTGACT
1271	CTCACACACGTTCTTCACT
1538	CTCACACACCTTCTTGACT
322	CTCACACTCGTTCTTGACT
1272	CTCACACACCTTCTTCACT
1217	CTCACACTCCTTCTTGACT
62	GACGGGCGGTGTGTACAA
399	GACGGGCGGTGTCTACAA
323	CCACAGCCTTTTACTTCAG
1282	CCACACCCTTTTACTTCAG
65	GCTGCCTCCCGTAGGAGT
407	GCTGCCTCCCCTAGGAGT

Microarray reader, image capture and data analysis

Imaging was performed with a custom-built fluorescent microscope (Argonne National Laboratory), illuminated by a short-arc Mercury lamp, with a cooled CCD camera and a thermal table that is connected to a thermoelectric temperature controller (LFI-3751; Wavelength Electronics, Inc., Bozeman, MT) and a water bath (VWR, West Chester, PA). Images were recorded, using a custom-written MCI program (Argonne National Lab). The sample was only exposed to light when taking an image; otherwise, the sample was shielded from the light by a closed mechanical shutter in the light path. The intensities were extracted from the images with iStackX, in-house written software obtained from the Stahl lab. MATLAB-based image

conversion scripts were developed and validated as part of this work to facilitate data flow between those two programs, as MCI records images in a customized IMG format, and iStackX accepts images in TIF format.

iStackX places a grid of rectangular boxes on an image, and in case of a stack of images such as a sequence of hybridization or melt images, subsequently applies the same grid automatically to all the images in the stack. Currently, iStackX does not have image registration capabilities, and all the images in one stack were checked for sliding relative to the grid from the first image. Also, the grid size was altered in the analysis to confirm that it did not have an effect on the data. The data was then further processed through background adjustment and normalization. The need for such data transformations in processing microarray data has been established [138, 139]. Specifically, in our case, each box in the iStackX grid contains an inner border and outer border, with the intensity in the area within the inner border representing the estimate of foreground (spot) intensity, and the intensity in the area in between the outer and inner border representing the estimate of (local) background intensity (see Figure 18). Both foreground and background intensities were averaged over the area. Currently, the intensities are background adjusted for analysis, subtracting the background from the foreground intensity, and then divided by the background value, which is a standard approach in microarray data analysis [139]. The division by background was done only for hybridization experiments but not for thermal dissociation experiments, as explained in the sections below. This data transformation and all downstream processing were conducted in MATLAB. For determining the T_d values of melt curves, the melt curves were 1) normalized between 0 and 1; 2) using the two intensity values closest to 0.5, the temperature corresponding to 0.5 was found via interpolation.

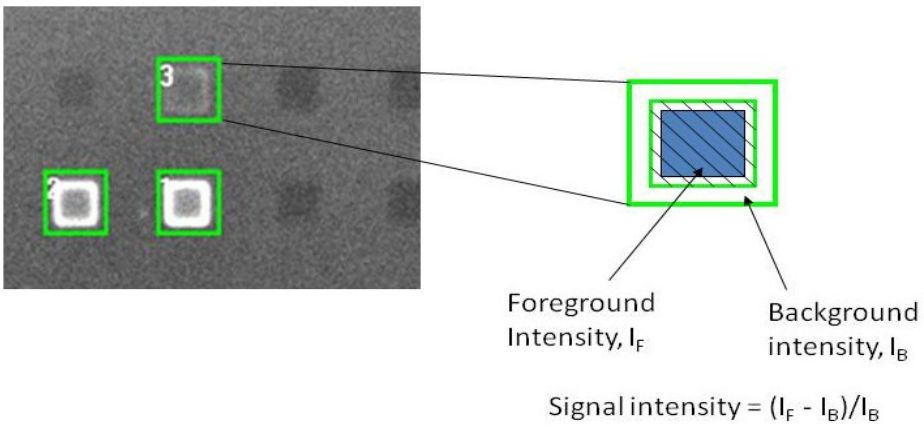

Figure 18 – Transformations of raw data before downstream processing. Left side of the Figure shows a part of a gel pad array image with a grid, and right side displays the location of 'foreground' and 'background' regions mentioned in the text. Local background is estimated in the immediate vicinity of the pad (band between the outer and inner box), and subtracted from the foreground signal (value in the inner box). The division by background was carried out for data from hybridization experiments, but not for thermal dissociation experiments.

Experimental protocols

Experiments with gel drop arrays were conducted in a hybridization chamber (Grace Biolabs, Bend, Oregon) affixed to the surface of the microarray glass slide (see Figure 41 in Chapter 4), containing 40 µl of hybridization buffer, while gel pad array experiments were conducted using a microfluidic platform later employed and further discussed in Chapter 4 (see Figure 41 in Chapter 4). All hybridizations and melts occurred in static conditions. The total volume of the hybridization chamber in the microfluidic platform was roughly 40 µl. In both platforms, a wash was conducted at the end of the hybridization and before starting the melt, washing away all unbound target. For the wash step, the hybridization chamber was removed from the glass slide in the gel drop experiments, and an imaging chamber (Grace Biolabs, Bend, Oregon) was affixed on the glass slide. Hybridization experiments were conducted at 20 ^{o}C, while in the melt experiments, the temperature was ramped from 20 ^{o}C to 70 ^{o}C in the span of an hour, rendering a ramp rate of 0.83 ^{o}C/min. The gel drop arrays were hybridized in the dark, with total time of

hybridization 17-18h (unless specified otherwise), since only end-of-hybridization signal was collected. During thermal dissociation experiments, images were taken after every minute. Gel pad arrays were continuously imaged in the microarray reader during the hybridization (image taken every 5 minutes, 1500 ms exposure time) as well as the melt (image taken every minute, exposure time 1500 ms).

Computational

Mathematical description of duplex formation and dissociation

In this Chapter, we are considering the simultaneous binding and dissociation of one oligonucleotide target onto two pads with different probe sequences. We assumed a two-state model of DNA binding and dissociation, and no formation of secondary structure [29, 113, 116, 118, 120, 121]. The hybridization between a target and probe is described by the following formalism:

$$C + P \underset{k_d}{\overset{k_a}{\rightleftharpoons}} B$$

where C is target, P probe, B probe-target duplex, k_a association constant, and k_d dissociation constant for the reaction. The change in duplex concentration over time then can be described by the following differential equation:

$$\frac{dB}{dt} = k_a \cdot C \cdot (P_0 - B) - k_d \cdot B,$$

where P_0 is the initial probe concentration.

Temperature-dependence of the binding reaction for thermal dissociation modeling was introduced to the model through the dissociation constant, since empirical studies have found the

association constant to be fairly insensitive to temperature in the temperature range of our interest:

$$k_d = k_a \cdot e^{\frac{-\Delta H + T\Delta S}{RT}},$$

where *H* is enthalpy, *S* entropy, *R* universal gas constant, and *T* temperature [29]. ΔH and ΔS for a given DNA oligonucleotide sequence can be estimated using the thermodynamic nearest-neighbor parameters [66].

Thermodynamic parameters of duplex stability

Temperature-dependence of the binding reaction for thermal dissociation modeling was introduced to the model through the dissociation constant, similarly to Chapter 2. In brief,

$$k_d = k_a \cdot e^{\frac{-\Delta G}{RT}} = k_a \cdot e^{\frac{-\Delta H + T\Delta S}{RT}},$$

where Δ*G* is the change in binding free energy, *H* is enthalpy, *S* entropy, *R* universal gas constant, and *T* temperature [29]. For solution binding, considered in Chapter 2, ΔH and ΔS for a given DNA oligonucleotide sequence can be estimated using thermodynamic nearest-neighbor parameters [66], see also Appendix section A.3 for a table of nearest-neighbor parameters for perfect-match duplexes and Peyret *et al.* for nearest-neighbor parameters for internal mismatches [71]. Dependence on salt concentration for both perfect-match and mismatch duplexes in the solution was introduced through modifying ΔS while keeping ΔH independent of salt concentration :

$$\Delta S^0([Na^+]) = \Delta S^0(1M\ NaCl) + 0.368 \times N \times \ln[Na^+],$$

where $\Delta S^0([Na^+])$ is the entropy change for an oligonucleotide duplex dissolved in a given sodium concentration, ΔS^0(1M NaCl) is the entropy change predicted from the unified nearest-neighbor parameters for 1M NaCl, and N is the total number of phosphates in the duplex divided

by 2 [66]. The values for duplex stability, ΔS and ΔH, are given in Appendix, section C. Microarray hybridization and dissociation buffers have varying amounts of salt and formamide, and our model accounts for different stabilities during the hybridization and dissociation phase of microarray experiments. The formamide effect, similarly to Chapter 2, was introduced into the model simply via offset of the temperature variable by $0.6^{o}C$ per percent formamide.

One-compartment model

Initial investigation of parameter sensitivity was performed using the implementation of the binding reaction for single target in MATLAB, assuming a well mixed solution, to gain a preliminary understanding of the relationship between target/probe concentration and binding kinetics without carrying out computationally expensive finite element simulations. This was similar to the MATLAB model described in Chapter 2, except we also considered binding of a single target onto two different probes, and initial conditions were appropriate for simulating hybridization (duplex concentrations were set to 0 at time t=0). The local concentration of the probe molecules was not considered; instead, the total amount of probe molecules was computed using the geometry of the gel pad and probe concentration, and divided by the total volume of the reaction chamber. The same one-compartment model was also used for the initial comparison between a compartmental model and three-dimensional finite element model, except dissociation from only one probe was considered, and initial conditions were those reflecting of melt, similarly to the model described in Chapter 2.

In Chapter 2, the T_m value for both experimental and computational melt curves was taken to be the location of the maximum of the first order derivative of the melt curves. In this Chapter, the

parameter calculated from the experimental data is T_d, the temperature at which the signal has decreased to 50% of its original value during the melt, and the same parameter was computed from the simulations for comparison: the temperature at which the bound concentration was decreased to 50% of its original maximum. A comparison of the two methods for calculating the melt temperature (T_m and T_d) from simulation results showed that the difference was less than 0.8 ^oC.

Two-compartment model

A schematic description of the two models is given in Figure 19.

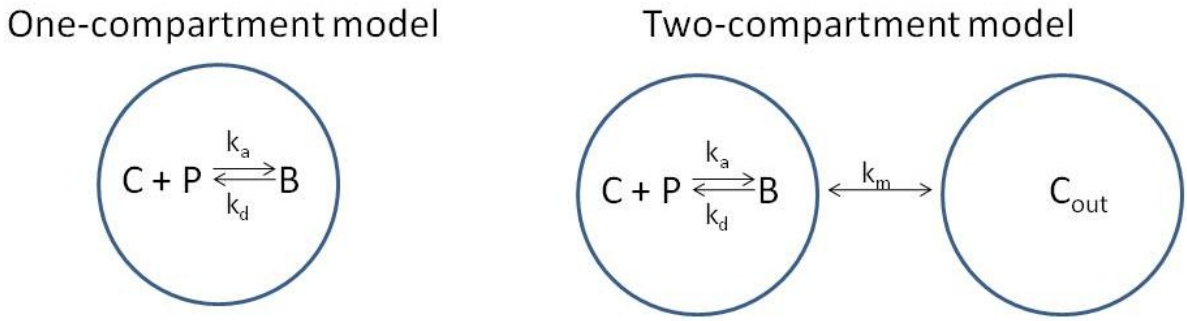

Figure 19 - Schematic description of one- and two-compartment model. C refers to target available for binding, P available probe, and B bound duplex.

For the 1-compartment model, the set of differential equations to be solved was as follows:

$$\begin{cases} \dfrac{dB}{dt} = k_a \cdot C \cdot (P_o - B) - k_d \cdot B \\ \dfrac{dC}{dt} = -\dfrac{dB}{dt}\dfrac{V_{gel}}{V_{solution}} \end{cases}$$

where B and C are the concentrations of bound duplex and free target, respectively; k_a and k_d are kinetic constants; P_o is the initial probe concentration; and $V_{gel}/V_{solution}$ is the ratio of the volume of one gel pad versus the volume of the hybridization chamber, equal to $5 \cdot 10^4$ (dimensions taken

from actual experimental system). At time 0, beginning of melt, free target concentration was assumed to be 0. The two-compartment model consisted of the following differential equations:

$$\begin{cases} \frac{dB}{dt} = k_a \cdot C \cdot (P_o - B) - k_d \cdot B \\ \frac{dC}{dt} = -\frac{dB}{dt} + k_m \cdot (C_{out} - C) \end{cases}$$

where k_m is mass transfer coefficient between the two compartments. C_{out} was assumed to be zero at all times, so the outer compartment functions as a permanent sink.

Three-dimensional finite element model

Numerical simulations of hybridization and melt in the microarray platform were performed with finite element modeling package COMSOL Multiphysics™ (Burlington, MA, versions 3.5a to 4.2a). The binding elements were assumed to be embedded in a hybridization chamber with a square footprint, with a surface area of $2.29 \cdot 10^{-4}$ m^2, height 100 μm. The dimensions of the three-dimensional cube-like pads were assumed to be 100 μm × 100 μm × 20 μm; the dimensions of the drop shaped gel elements was assumed to be 170 μm, height 30 μm, as specified by the manufacturer.

The binding in three-dimensional gel elements was modeled using the kinetic equation above, defining it for both the two pads considered, rendering the following set of binding equations:

$$\frac{dB_{PM}}{dt} = k_a \cdot C \cdot (P_{0PM} - B_{PM}) - k_{dPM} \cdot B_{PM}$$

$$\frac{dB_{MM}}{dt} = k_a \cdot C \cdot (P_{0MM} - B_{MM}) - k_{dMM} \cdot B_{MM},$$

where B_{PM} and B_{MM} refer to duplex concentrations on perfect-match and mismatch probes. P_{0PM} and P_{0MM} refer to initial concentrations of these probes (assumed to be equal), and k_{dPM} and k_{dMM} are the dissociation constants for PM and MM duplexes, respectively.

The diffusion of the free targets in the bulk and in the gel element was modeled using Fick's law:

$$\frac{\partial C}{\partial t} = \mathrm{D}\nabla^2 C,$$

where C is the concentration of the target, and D diffusion coefficient.

Modeling of free target diffusion in bulk and in gel elements was implemented in COMSOL by applying the diffusion mode (essentially defining the diffusion equation), and defining two separate diffusion constants, one for the bulk solution and one for the gel environment. The mobility of DNA fragments in the gel is expected to be lower than in the free solution and is determined by the size of the diffusing DNA fragments and properties of the gel. The binding of free target was reflected in the target concentration through a reaction term linked to the kinetic equation shown above. Modeling of the binding of target DNA onto immobilized probe molecules was implemented by applying two different approaches in the subdomain of the gel elements: 1) using COMSOL's subdomain PDE weak mode for the binding reaction; 2) defining a new diffusion mode to account for the bound probe species, with D=0 diffusion coefficient and a reaction term linked to the kinetic equation. The results from both were compared for a range of model input parameters, and the results were identical. Implementing a separate diffusion mode for the immobilized species was chosen and has been used throughout this work.

For thermal dissociation modeling, temperature as a linear function of temperature was defined under global expressions in COMSOL modeling interface, and k_d was calculated at every solver

step as a function of temperature. The vertical walls of the hybridization chamber, the top and bottom of the chamber were set to an insulation boundary. The boundaries along three-dimensional binding elements were set to continuity for target concentration. The initial conditions for the time-dependent solver when modeling hybridization were set to 0 for duplex concentration, 0 for free target concentration inside three-dimensional binding elements, and a pre-determined value in the solution. For dissociation, all non-bound target concentration (inside and outside binding elements) was assumed to be zero, while the concentration of bound duplex was set to a pre-determined value.

The shape of the spherical cap shaped gel drops in the model was created in two steps. First, a sphere was created with a 135 μm radius, lowering the center of the sphere 105 μm below the bottom of the hybridization chamber, and a square block was created with its height equal to 30 μm and length of the side more than the desired diameter of the spherical cap. Second, an intersection of those two objects was formed which rendered a spherical cap with the radius of 170 μm and height 30 μm.

Main issues with building the model had to do with negative concentrations in cases where reaction flux was large compared to the diffusional flux. Concentration of free target would consistently dip into negative, with the ratio of integrated absolute concentration to integrated concentration in the binding pad subdomains being close to 2. Three different approaches were employed simultaneously to avoid this problem: 1) forcing the solver to take smaller time steps; 2) optimizing mesh element size around the pads; 3) for some modeling cases, avoiding spatially discontinuous initial conditions by using a smoothed Heaviside function (flc2hs in COMSOL).

The list of COMSOL model constants and their values is given in Table 7.

Table 7 – List of COMSOL model constants with sources (if applicable) and flexibility in the value for purposes of model fitting.

Parameter	(Starting) Value	Source	Flexibility in the value
Association constant k_a	10^6 1/(M·s)	[29, 118, 126]	Fairly little flexibility; literature claims k_a to be fairly independent of target sequence and temperature within our range of interest
Dissociation constant for PM k_{dPM}	Starting value computed as a function of ΔH_{PM}, $\Delta S_{PM,}$ and T	[29]	We hypothesize that ΔH_{PM}, ΔS_{PM} for microarray binding are linear functions of solution thermodynamic parameters: flexibility in the constants of the linear functions
Initial target concentration	Nano- to micromolar range	Experimental value	No flexibility
Initial probe concentration Θ_0	60 μM – 1 mM	Depends on the microarray type (gel pad versus gel drop) [7, 8]	Possible to explore lower probe concentrations in the model to account for hindered target penetration and a population of inaccessible probe molecules
Diffusion constant in solution D_s	~$5\cdot10^{-7}$ -10^{-6} 1/cm^2·s	[29, 118, 126]	No flexibility
Diffusion constant in gel matrix D_g	<D_s		Leaving it as a rather open parameter: upper limit would be the value of solution diffusion constant

Parameter optimization

Optimization of thermodynamic parameters was conducted with the two-compartment model, utilizing MATLAB's Optimization toolbox. Specifically, a nonlinear least-squares based Levenberg-Marquardt optimization algorithm was used from MATLAB optimization toolbox [140, 141]. The thermodynamic parameters in the two-compartment model were optimized

either based on the comparison with thermal dissociation modeling results from a 3D model, or based on the experimental dissociation curves.

3.4 Results & Discussion

Experimental: photostability of the fluorescent labels over several hours of light exposure

As the fluorescently labeled targets are imaged over several hours during hybridization, the photostability of two different fluorescent labels was characterized. Texas Red (TR) or Oregon Green (OgG) labeled target solution was subjected to same conditions as in the hybridization experiments, only on a blank microscope array with no binding elements present. Images were collected every 5 minutes, exposure time 1500 ms, at 20 ^{o}C and dark-frame subtracted. Note the very good stability of the TR and OgG fluorophore in the span of several hours of imaging (Figure 20 and Figure 21). ROI1-ROI5 refer to SI taken from five separate spots on the array: four corners, and the center of the image. Control refers to a spot on the image outside the hybridization chamber. Relative standard deviations for the corresponding regions of interest (1-5) and control region in Figure 20 were 1.2%, 1.2%, 1.0%, 2.0%, 3.0%, 0.8% (average values 684.6, 644.6, 901.5, 796.3, 705.3, 306.6, respectively).

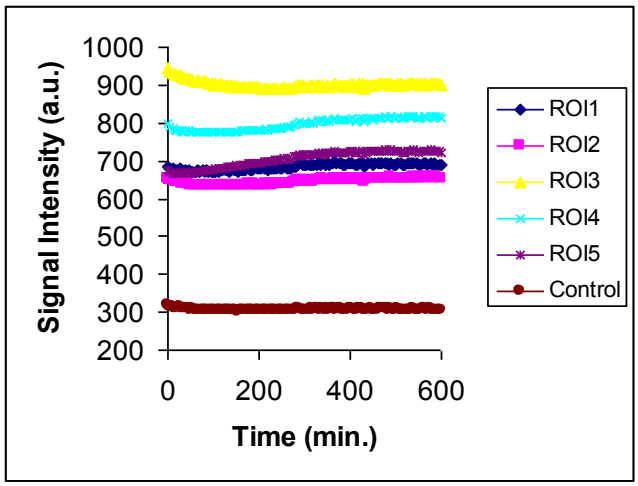

Figure 20 – Signal intensity as a function of time in a hybridization chamber for Texas Red labeled target solution in five different regions of interest in the chamber and a control spot. See text for detailed description of the ROI locations.

Relative standard deviations for the corresponding regions of interest (1-5) in Figure 21 were 1.0%, 1.9%, 1.0%, 0.7%, and 1.9% (average values 38.6, 36.4, 64, 56.7, 52.7, respectively).

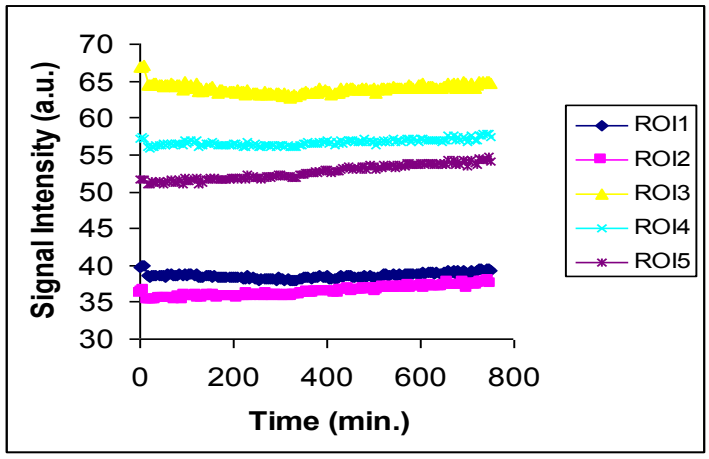

Figure 21 – Signal intensity as a function of time in a hybridization chamber for Oregon Green labeled target solution in five different regions of interest in the chamber and a control spot. See text for detailed description of the ROI locations.

In conclusion, both labels exhibited good photostability over the course of the experiments, and are suitable to use for monitoring hybridization reaction over several hours with frequent light exposure.

Experimental: Microarray hybridization and thermal dissociation for model validation

Sample image from hybridization inside the microfluidic platform is shown in Figure 22. The locations of probes that are used later in the Chapter for analysis are labeled. Figure 23 displays the hybridization signal as a function of time on multiple probes from experiments with a single hybridizing target. Note that the signal from probe EUB338 has not reached a plateau yet, illustrating the potentially long time to equilibrium in static experiments. Also, note the striking difference between the signal intensities from EUB338 and EUB338MM probes.

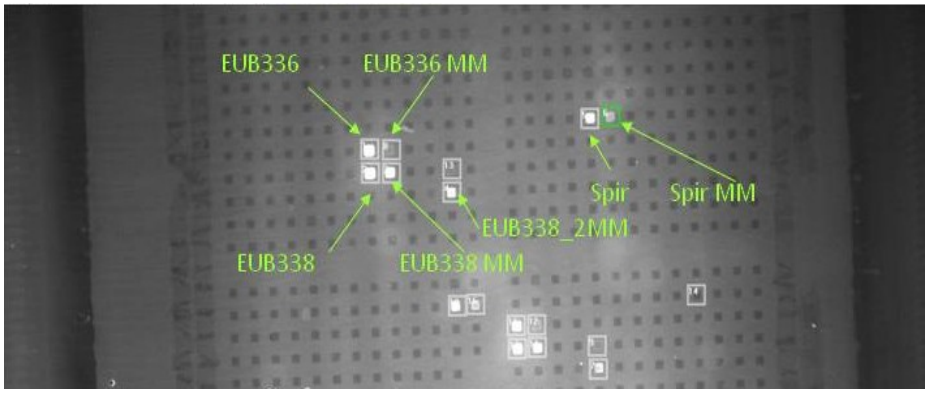

Figure 22 - Sample image from a hybridization experiment with a gel pad microarray. Dark squares are gel elements, bright squares probes with fluorescently labeled target hybridized. Light background is due to fluorescently labeled unbound target. The hybridization solution contains target that is perfect match to EUB338 probe.

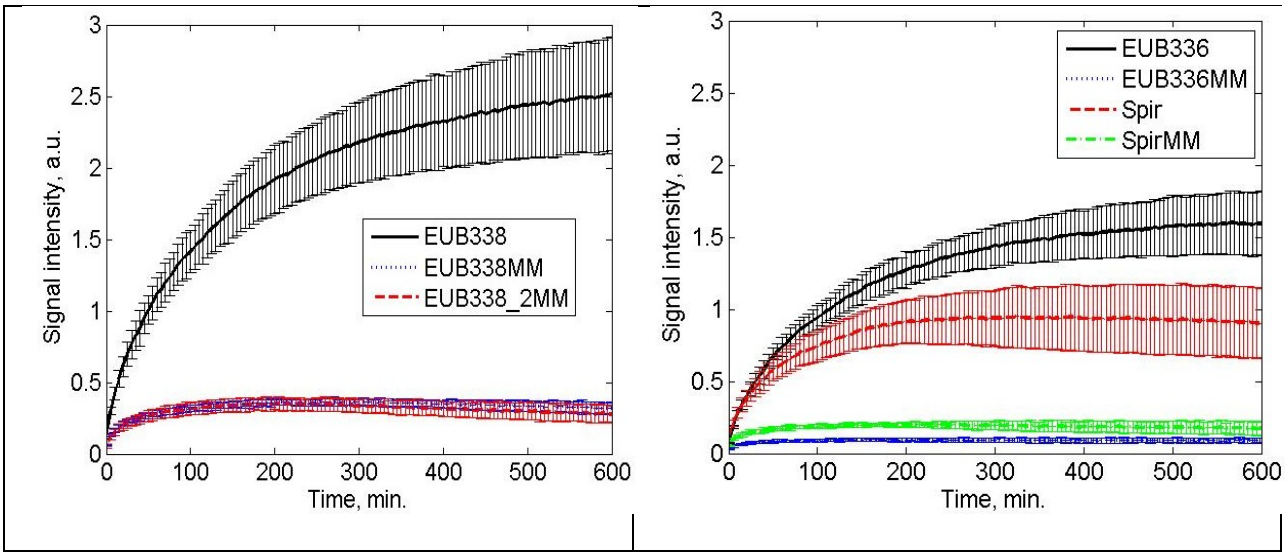

Figure 23 – Signal intensity with standard deviation from different probes on the gel pad microarray during the course of hybridization of a single target, perfect-match to probe EUB338 (n=3). Hybridization temperature 20°C, target concentration 500 nM, buffer composition 900 mM NaCl, 40% formamide, 20 mM Tris-HCl.

Figure 24 shows normalized melting curves from non-equilibrium thermal dissociation experiments following the initial hybridization, and Table 8 lists the T_d values (temperature at which the signal is at 50% from its maximum value) for the probes considered in Figure 24. Note the marked difference in EUB338 and EUB338MM T_d values, but curious overlap of T_d values between probes EUB338 and EUB336.

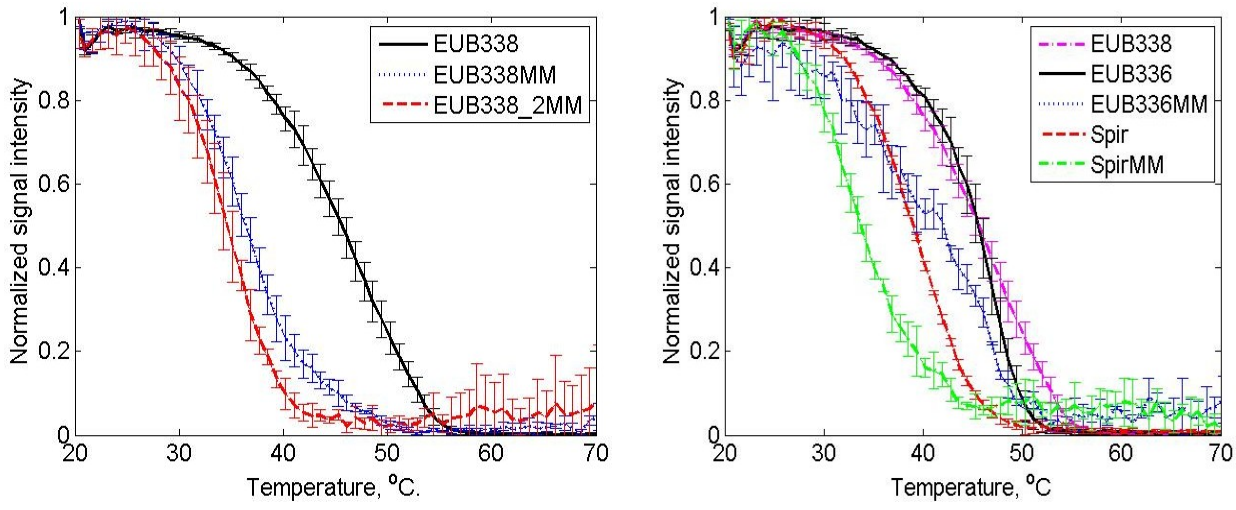

Figure 24 – Normalized and averaged (n=3) melt curves with standard deviations for different probes in the gel pad microarray, after hybridization of single perfect-match target to probe EUB338 at 500 nM. Temperature ramp rate:0.83 °C/min. Buffer composition: 4 mM NaCl, 20 mM Tris-HCl, 5 mM EDTA.

Table 8 – Average T_d (temperature at which normalized signal intensity is at 50%) values and standard deviations for melt curves in Figure 24 (n=3). The target was a perfect match to EUB338 probe. Hybridizing target: 5′-ACTCCTACGGGAGGCAGC-TexasRed.

Probe ID	EUB338	EUB338MM	EUB336	EUB336MM	Spir	SpirMM	EUB338_2MM
T_d, °C	45.6 ±0.77	36.4 ±0.71	45.5 ±0.73	41.6 ±1.0	39.0 ±0.1	33.8 ±0.5	34.5 ±1.0

Computational: will the solution-based thermodynamic constants render stabilities similar to those seen experimentally?

We are aiming to describe accurately both the microarray hybridization as well as thermal dissociation experiments. In particular, the agreement between model predictions and experimental data would mean the following: 1) alignment of the ratio of hybridized amount on perfect-match and mismatch probes; 2) time to equilibrium in the hybridization reaction; 3) T_d values of thermal dissociation. In Chapter 2, we developed a solution-based kinetic model of thermal dissociation, and supported it with solution-based experimental data, using DNA oligonucleotides of different sequences in different buffers. As a first step in examining the applicability of our kinetic model to microarray experimental data, we neglect the effects of mass transport and therefore are not concerned with the time course of hybridization reaction, and investigate a one-dimensional model of ordinary differential equations of binding with a well mixed solution, similar to the one described in Chapter 2, before delving into the model of hybridization chamber with spatially discrete binding elements.

PM/MM ratio: experimental versus one-compartment kinetic model

Figure 25 displays the experimental PM/MM signal ratio from EUB338 and EUB338MM probe in the presence of one target, a perfect-match to EUB338 probe. At the end of 10 hours of hybridization, the signal from perfect-match probe on the average exceeds that from a mismatch

probe roughly 8 times. Also note that the PM/MM ratio has not reached a plateau yet, meaning the true equilibrium value could even be higher.

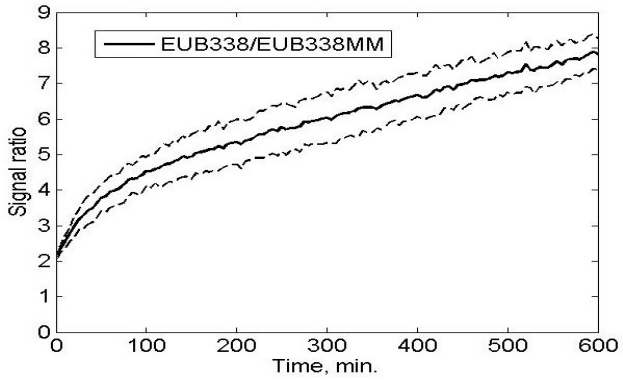

Figure 25 – Ratio of signal from probes presenting a perfect match (EUB338) to the hybridizing target and one-base mismatch (EUB338MM), corresponding to Figure 23. Dashed lines represent mean +/- 1 SD.

In order to characterize the specificity of binding reactions in a mathematical model of hybridization with a well mixed solution (solution-based model), taking into account the thermodynamic nearest-neighbor based parameters for the appropriate salt concentration (Appendix, section C) and formamide concentration, we calculated the equilibrium fraction of probe occupied as a function of probe concentration, target concentration, and dissociation constant, assuming the binding of one target onto one probe. The results of these simulations are displayed on Figure 26. While the experimental PM/MM ratio is obtained from an experimental set-up where one target is simultaneously hybridizing onto two separate probes, Figure 26 different panels show the fraction of probe bound for separate hybridizations of one target onto one probe as a function of dissociation constant. A hypothetical PM/MM ratio can be inferred from the figure assuming an average of three-fold difference in dissociation constants between a perfect-match and mismatch duplex. For example, one can see from the uppermost left panel, that regardless of the target concentration, the fraction of probe bound is insensitive to the

dissociation constant at values 10^{-4} 1/s and lower, and the hypothetical PM/MM ratio for $k_d < 10^{-4}$ 1/s is 1.

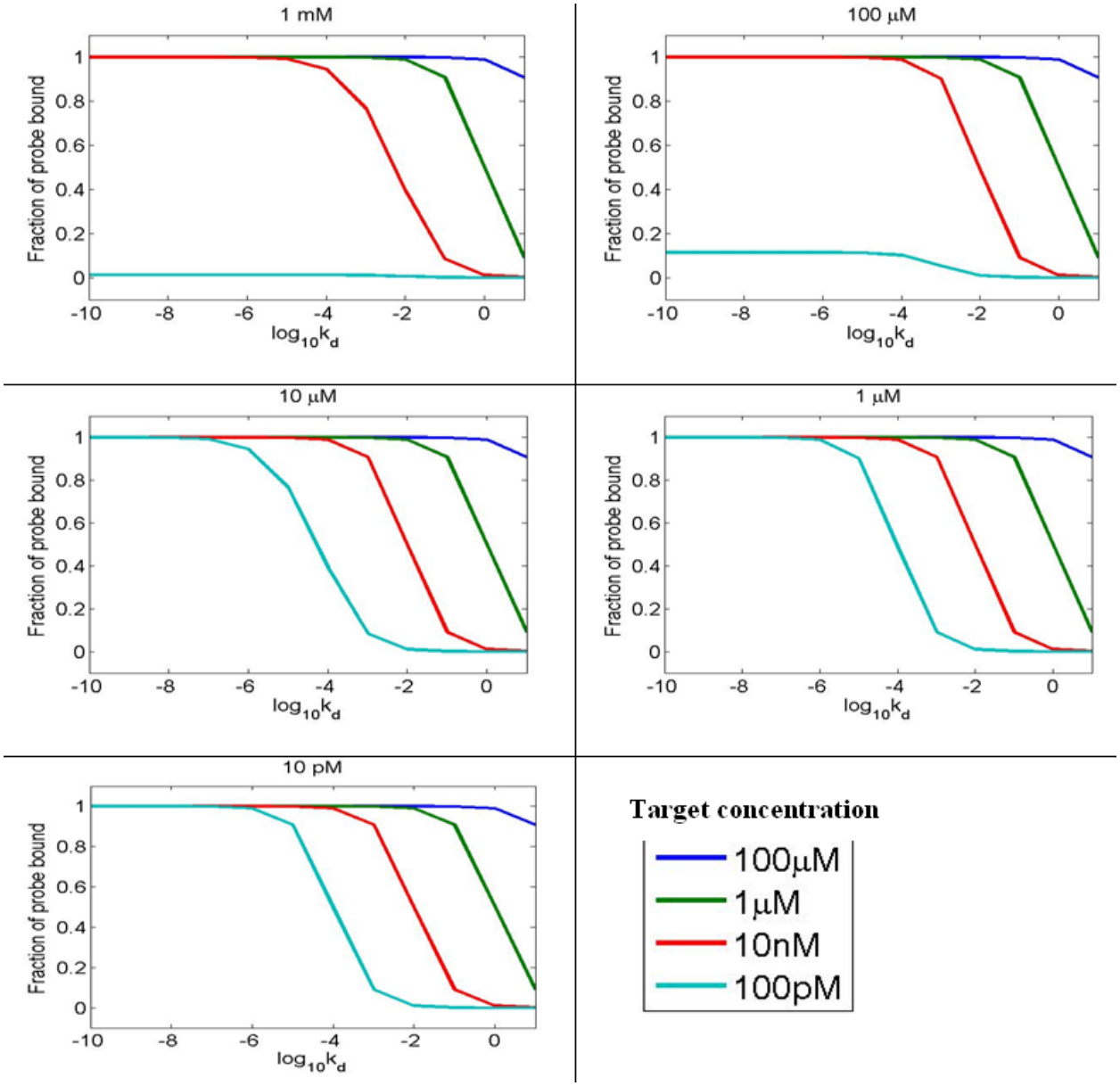

Figure 26 - Fraction of probe bound as a function of dissociation constant for various probe concentrations (different subfigures) and target concentrations (different colored lines on each subfigure). The probe and target concentrations correspond to actual concentrations in microarray experiments as the probe concentration was "diluted" in the solution-based model by multiplying it by the ratio of probe gel element.

If we consider the following conditions for the hybridization reaction: T=293 K plus an offset to account for 40% formamide, 0.9M NaCl buffer, we obtain the following dissociation constant values for EUB338/EUB338 (PM) and EUB338/EUB338MM (MM) duplexes: $k_{dPM}=7.7\ 10^{-10}$ 1/s; $k_{dMM}=1.1\ 10^{-7}$ 1/s. Consulting Figure 26, one can see that regardless of probe or target concentrations, it is not possible to replicate a PM/MM ratio above 1 with these dissociation constants.

T_d values: experimental versus one-compartment kinetic model

It is important to realize that T_d, the temperature at which the original end-of-hybridization signal has decreased by 50%, depends on the concentration of the bound duplex. This is similar to the concentration-dependence of the equilibrium parameter T_m (the derivation of this parameter and a figure illustrating its dependence on DNA concentration is given in Appendix, Section A.1). Figure 24 displays the dissociation curves for EUB338 PM and MM duplexes in the experimental section; we are going to compare it with Figure 27 that displays the computed dissociation curves for four different initial probe concentrations, assuming that at the end of the hybridization, all probe molecules had formed a duplex with target molecules. This assumption is not critical: we ran simulations where the initial probe concentration was the same as in the four cases displayed in Figure 27 but only 50% of the total probe was bound, and the maximum difference between the T_d values of the two scenarios was only 0.15%. The T_d values of the EUB338 PM duplex dissociation curves shown in Figure 27 and of EUB338 MM duplex (dissociation curves not shown), and the difference between these T_d values are given in Table 9. Note that the computed dissociation curves have a higher T_d value for all but very low initial probe concentrations compared to experimental data in Table 8. However, the difference

between the EUB338PM and EUB338MM T_d values is comparable between the experimental data and one-compartment model (minimum of 7.0 ^oC at the highest probe concentration considered in the simulation versus 9.2 ^oC in the experiments).

Results of repeating the same comparison for the probes on gel drop array are given in Table 10, listing the T_d values for perfect-match and one-base mismatch duplexes. The same qualitative trends as are displayed as in the corresponding experimental data (Table 9): the T_d of probe 62 with its perfect-match target is higher than that of probe 1537 with its respective perfect-match target, and the mismatch probe 1271 has a higher T_d than probe 1538.

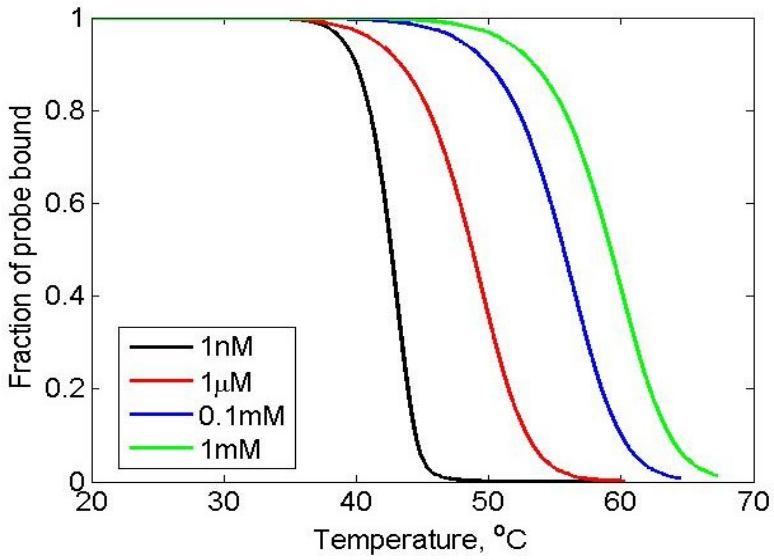

Figure 27 – Computed dissociation curves of EUB338 PM duplex for four different initial probe concentrations, assuming all of the probe was involved in a duplex at the beginning of the dissociation. Buffer conditions: 4mM NaCl. Thermodynamic parameters used in the model are given in Appendix, Section C.

Table 9 – Computed T_d values of EUB338 PM and EUB338 MM duplexes for different probe concentrations. ΔT_d values are the differences between EUB338PM and EUB338MM T_d values. Buffer conditions: 4 mM NaCl. Thermodynamic parameters used in the model given in Appendix Section C.

Initial probe conc.	T_d, °C		ΔT_d, °C
	Hybridizing target: 5′-ACTCCTACGGGAGGCAGC-TexasRed		
	EUB338 PM	EUB338 MM	
1 nM	43.0	29.0	14
1 μM	49.4	35.5	13.9
0.1mM	56.4	42.7	13.7
1mM	60.0	46.5	13.5

Similarly to the k_d values calculated above for hybridization, if we consider that at the beginning of the melt, T=293 K and the buffer contains 4 mM NaCl and no formamide, we obtain the following dissociation constant values for the EUB338 PM and MM duplex: $k_{dPM}=2.3\ 10^{-10}$ 1/s; $k_{dMM}=1.4\ 10^{-7}$ 1/s.

The conclusion from the comparison of hybridization simulations with experimental data was that the high PM/MM signal ratios could be replicated only if one considers the k_d values to be significantly higher than computed based on the nearest-neighbor model, while keeping the same relative ratio of the dissociation constants of perfect-match and mismatch duplexes. Similarly, in dissociation modeling, the observed lower experimental T_d values can be replicated by starting the thermal dissociation at a higher starting dissociation constant value in the model, which would be equivalent of introducing a temperature offset in the model (*e.g.,* experimental temperature T=293 K would correspond to model temperature T=293K+ *offset.* That will lower the calculated T_d values by the magnitude of the offset: for example, offset of 14K will give us starting values of $k_{dPM}=1\ 10^{-5}$ 1/s; $k_{dMM}=3.5\ 10^{-3}$ 1/s and lower calculated T_d values by 14 °C.

Table 10 - Computed T_d values of duplexes formed on probes 1537, 1271, 1538, 62 and 399 at different probe concentrations for two different hybridizing targets. Buffer conditions: 4 mM NaCl. Thermodynamic parameters used in the model given in Appendix Section C.

Initial probe conc.	T_d, °C				
	Set a Hybridizing target: 5'-TTGTACACACCGCCCGTC-OgG		Set b Hybridizing target: 5'-AGTCAAGAACGTGTGTGAG-OgG		
	62	**399**	**1537**	**1271**	**1538**
1 nM	42.4	29.7	37.5	25.6	24.7
1 µM	48.3	36.4	43.0	32.1	30.4
0.1mM	55.0	43.8	49.3	39.4	37.0
1mM	58.5	47.6	52.5	43.2	40.4

Experimental: effect of initial pre-hybridization target concentration on thermal dissociation

While the amount of target hybridized on a probe is generally proportional to the initial target concentration below saturation of probe, as well as a function of affinity, the effect of initial target concentration on thermal dissociation curves has not been studied to the best of our knowledge. Generally, the T_d value is considered to be an indicator of duplex stability – lower T_d values corresponding to less stable duplexes, the effect of initial target concentration on the T_d values has not systematically been studied in a simple system. Table 11 displays T_d values from non-equilibrium thermal dissociation experiments with gel drop microarrays, varying single target concentration over two orders of magnitude. Note the decrease in T_d as a function of increasing target concentration for all probes except 1538 – the T_d at higher target concentration (mM) for a perfect-match is almost comparable to the T_d value of mismatch at a lower temperature. The decrease in T_d is especially pronounced for perfect-match duplexes formed on probes 62 and 1537, and the difference between perfect-match and corresponding one-base mismatch decreases with increasing target concentration. For example, ΔT_d for probes 62 and 399 decreases from 8.1 °C to 6.8 °C (by 1.3 °C), while the ΔT_d for probes 1537 and 1271 decreases from 7.4 °C to 5.6 °C (1.8 °C). The possible reasons behind these trends will be discussed below.

Table 11 – Average T_d (temperature at which normalized signal intensity is at 50%) values and standard deviations (n=4) from gel drop arrays, as a function of probe sequence and added target concentrations. Set a corresponds to experiments with an added target that is perfect match to probe 62 (399 is a one-base mismatch to 62), and Set b corresponds to experiments with an added target that is perfect match to probe 1537 (probes 1271 and 1538 are different one-base mismatch probes).

Target concentration	**Probe ID**				
	Set a **Hybridizing target:** 5'-TTGTACACACCGCCCGTC-OgG		**Set b** **Hybridizing target:** 5'-AGTCAAGAACGTGTGTGAG-OgG		
	62	**399**	**1537**	**1271**	**1538**
10 nM	54.5 ±1.0	46.4 ±10.6	50.2 ±0.5	42.8 ±0.4	37.62 ±4.5
100 nM	51.8 ±0.8	44.2 ±0.4	46.3 ±0.7	40.0 ±0.2	38.4 ±0.2
200 nM	48.7 ±0.5	41.3 ±0.2	43.6 ±0.2	37.3 ±0.1	36.1 ±0.2
1 μM	49.5 ±1.4	42.7 ±0.3	44.0 ±0.7	38.4 ±0.1	37.3 ±0.1

Three-dimensional finite element model: effect of diffusion coefficient, free target inside the gel element, and thermal dissociation initial conditions

Before continuing with the discussion of experimental results, we will introduce the three-dimensional finite element model of microarray hybridization and dissociation. Figure 28 shows the geometry of the COMSOL three-dimensional model, containing two gel drops, corresponding to perfect-match and mismatch probe.

The only open parameter in the model is intra-gel diffusion coefficient (see Table 7). While the intra-gel diffusion coefficient will not affect the expected equilibrium values of bound duplex concentration, it can alter the time course of the hybridization. We conducted a sensitivity analysis to determine the effect of diffusion coefficient on hybridization time course and thermal dissociation curve. The second question of interest was the contribution that the free targets inside the three-dimensional binding element will make to the total amount of targets (bound and

unbound) in the binding element, specifically during non-equilibrium thermal dissociation. In a hybridization experiment, free fluorescently labeled targets will diffuse into the gel element and, depending on the presence of complementary base pairs, form duplexes. The experimentally observed signal is the light emitted by fluorophores attached to targets in a specific area – the binding element, regardless of whether the target was bound to a probe molecule or not. Similarly, in a thermal dissociation experiment, the observed signal is again the light emitted by fluorophores attached to targets in the binding element, regardless of whether they are still bound to a probe or already freely diffusing inside the gel. The third question of interest was what effect the fraction of probe bound at the beginning of a thermal dissociation has on the shape of the dissociation curve.

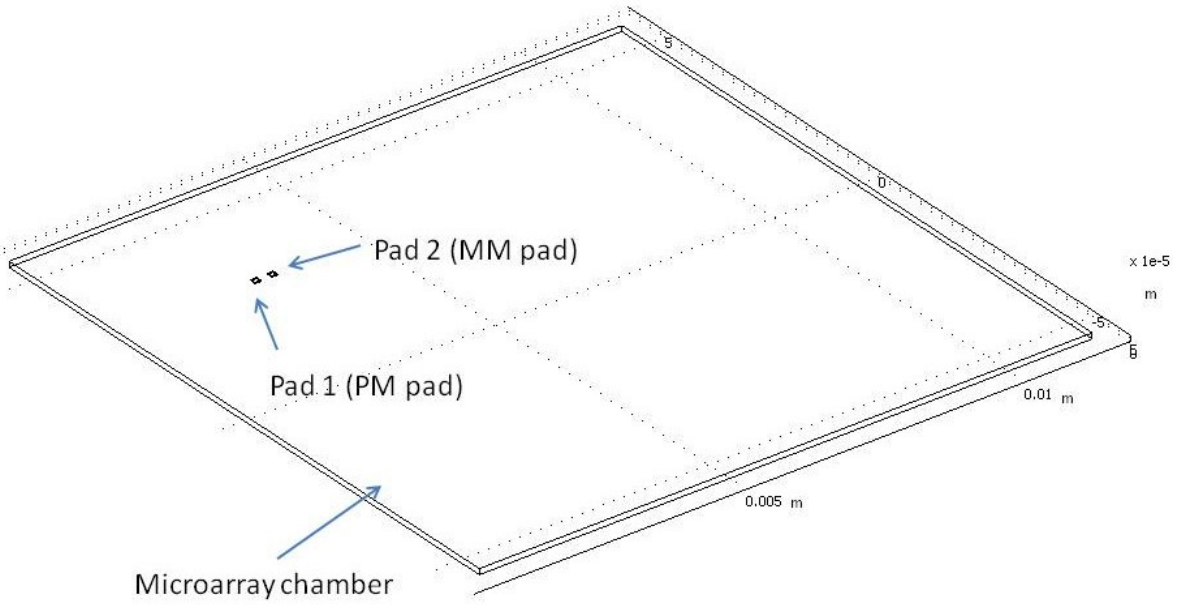

Figure 28 – Geometry of the COMSOL three-dimensional model: microarray chamber (bulk solution) and two three-dimensional binding elements.

To explore the effect of intra-diffusion coefficient on the hybridization simulations, three different cases were considered: 1) intra-gel diffusion coefficient equal to that in the solution; 2) intra-gel diffusion coefficient 10 times smaller than the solution coefficient; 3) intra-gel diffusion

coefficient 100 times smaller than the solution coefficient. Two different initial probe concentrations ($6 \cdot 10^{-2}$ mol/m^3 and $6 \cdot 10^{-4}$ mol/m^3, the first concentration being in the range of what we expect our experimental arrays to contain) and two initial target concentrations (10^{-4} mol/m^3 and 10^{-8} mol/m^3) were considered. Figure 29 displays the effect (or lack of thereof) of altering the diffusion coefficient on the time course of hybridization for the probe concentration we expect to have in our binding elements.

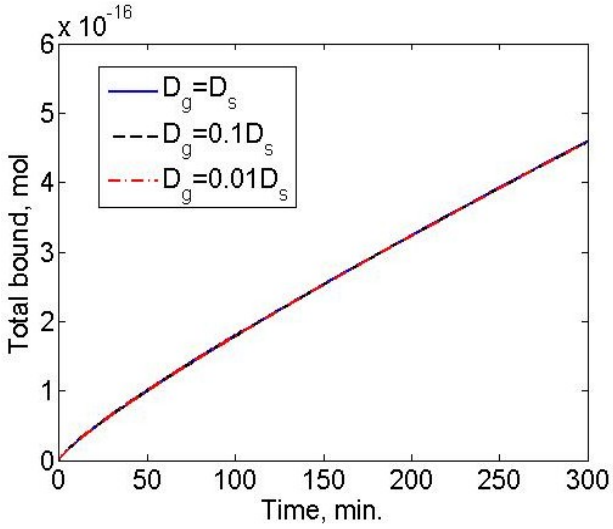

Figure 29 – Simulated hybridization curves for different intra-gel diffusion coefficient values. Solution-based diffusion coefficient was 10^{-10} m^2/s. Initial probe concentration $6 \cdot 10^{-2}$ mol/m^3, dissociation constant $k_d=10^{-7}$ 1/s, association constant $k_a=10^6$ 1/(M·s). Initial target concentration $c_0=10^{-6}$ mol/m^3.

For comparison, Figure 30 displays the dramatic effect of reducing the solution diffusion coefficient. We also computed the amount of free target in the binding element to estimate its effect on the total amount of target in the binding element, but found the ratio of free target to bound target to be less than 0.5 %. We therefore conclude that the three-dimensional model of the hybridization system is rather insensitive to the value of intra-gel diffusion coefficient, given our expected probe density, and that there is no significant accumulation of free target inside the gel element.

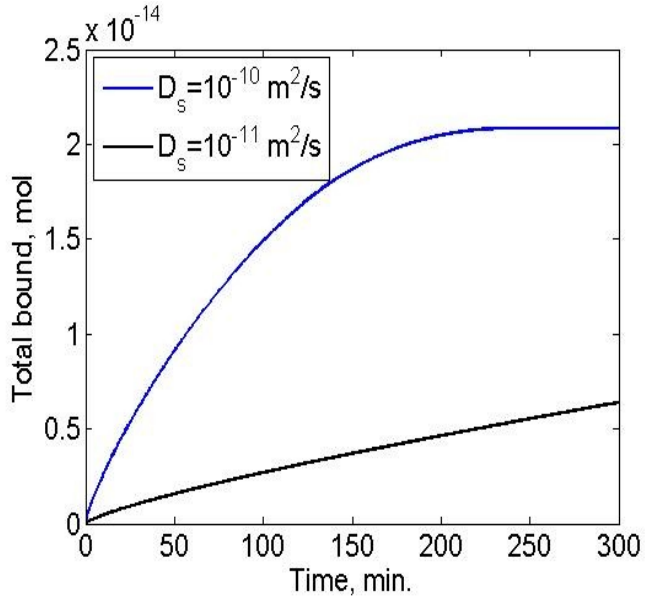

Figure 30 – Simulated hybridization curves for two different values of the diffusion coefficient in the solution. Intra-gel diffusion coefficient was 10^{-11} m^2/s. Initial probe concentration $6 \cdot 10^{-2}$ mol/m^3, dissociation constant $k_d=10^{-7}$ 1/s, association constant $k_a=10^6$ 1/(M·s).

Conducting a similar analysis for the model of thermal dissociation, Figure 31 displays computed thermal dissociation curves, showing the effect of reducing the value of the intra-gel diffusion coefficient below the solution value for two different ways of computing the 'total amount' of targets: either just the amount of targets that are part of a duplex, or the sum of targets in the duplex and free targets in the binding element. It is only at an intra-gel diffusion coefficient that is a hundred times lower than the solution coefficient that one begins to see a small contribution by the free targets in the gel element to the total amount of targets (difference between solid and dashed lines in the bottom panel of Figure 31). However, it is only appreciable in a scenario where targets cannot reassociate after dissociation (compare blue and black lines in bottom panel of Figure 31). Figure 31 also demonstrates the profound effect of target rebinding on the shape of the dissociation curves: the blue lines on all three panels of Figure 31 are shifted to the left compared to black lines, rendering a reduction in apparent T_d values by several degrees. We will return to discussing this observation later in this Chapter.

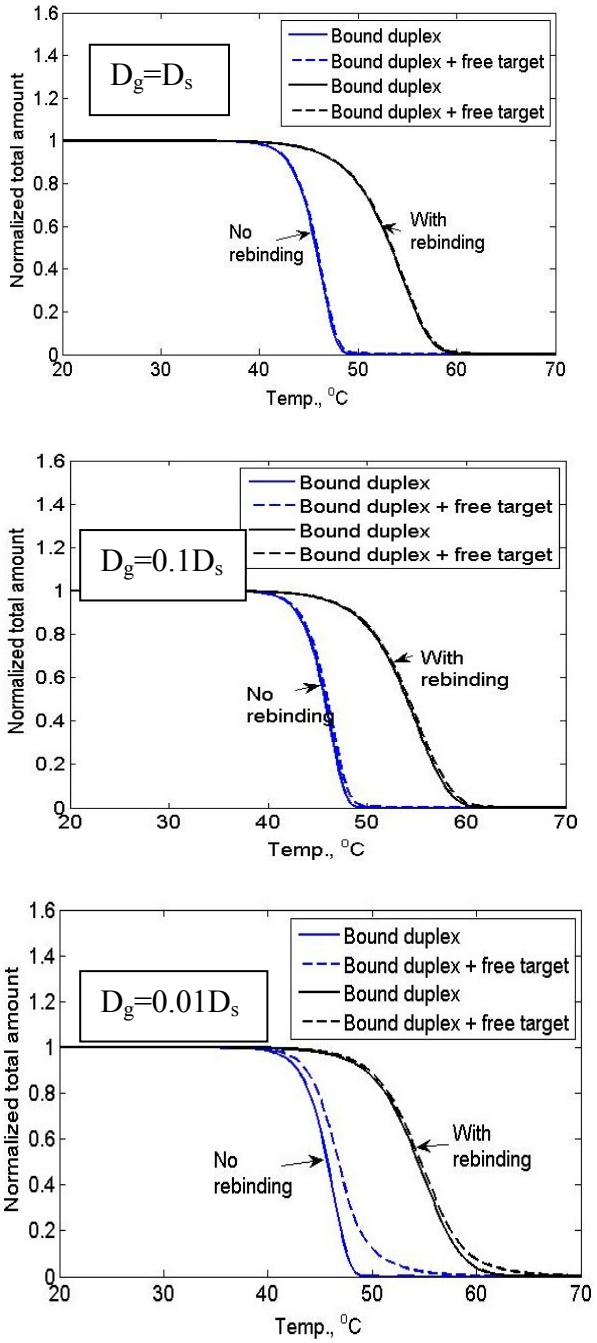

Figure 31 – Computed effect of free target on the total amount of target in the binding element (total amount) during thermal dissociation, for three different values of intra-gel diffusion coefficient D_g as a fraction of solution diffusion coefficient $D_s=10^{-10}$ m^2/s. 'No rebinding' refers to target dissociation as an irreversible process, while 'With rebinding' refers to target having the option of rebinding after initial dissociation. Solid lines refer to the total target amount only being composed of targets that are in the duplex, while dashed lines refer to the total target amount being composed of the sum of free targets in the binding element and targets in the duplexes. Model parameters: initial fraction of probe bound = 100%, initial probe concentration $6\cdot10^{-2}$ mol/m^3, ΔH=-148 kcal/mol, ΔS=-403.4 cal/(mol·K), $D_s=10^{-10}$ mol/m^2, association constant $k_a=10^6$ 1/(M·s). $D_g=10^{-11}$ mol/m^2. Assumed temperature ramp rate 0.83 °C/min.

Lastly, we explored the effect that the fraction of probe bound at the beginning of thermal dissociation will have on the simulated thermal dissociation curves. Specifically, we considered two extreme cases: 100% of total probe concentration bound in a duplex, and 5% of total probe concentration bound in a duplex at the beginning of the dissociation. Figure 32 displays the shift in the dissociation curve to the left that accompanies the increase in the fraction of probe involved in a duplex at the beginning of the melt. We will compare this simulation result with experimental observations in the following section. Also, note that similarly to modeling hybridization, accounting for free target in the gel element does not have an appreciable effect on the total amount of target in the gel element.

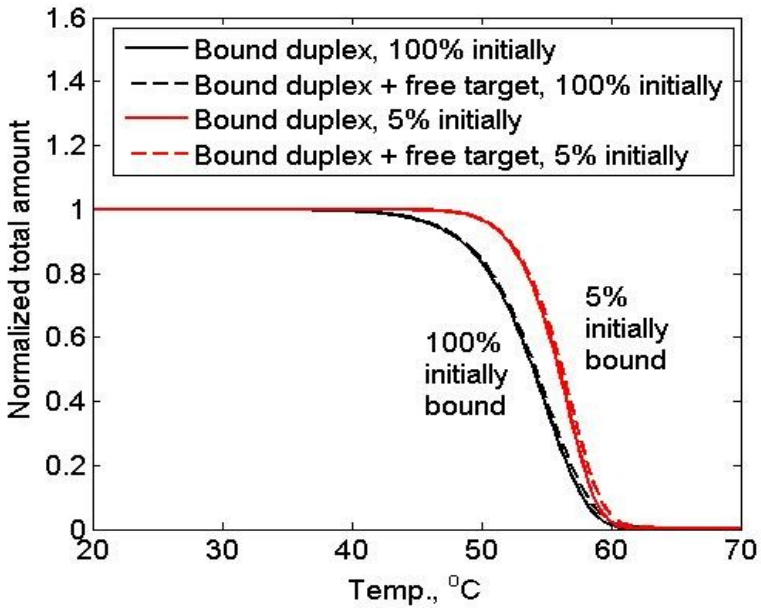

Figure 32 – Normalized simulated thermal dissociation curves, varying in the fraction of probe that is initially bound in a duplex (100% *vs* 5%). Also, the sum of bound duplex and free target in the gel element was included for comparison. Initial conditions: initial probe concentration $6 \cdot 10^{-2}$ mol/m^3, ΔH=-148 kcal/mol, ΔS=-403.4 cal/(mol·K), $D_s=10^{-10}$ mol/m^2, $D_g=10^{-11}$ mol/m^2. Association constant $k_a=10^6$ 1/(M·s). Assumed temperature ramp rate 0.83 °C/min.

Alternative explanation for the concentration dependence of T_d: experiments versus simulations

As shown earlier in this Chapter (Table 11), the T_d value of thermal dissociation curves depends on the initial concentration of target at the beginning of hybridization – specifically, the value of T_d was shown to decrease with increasing target concentration at the beginning of the hybridization. Our first hypothesis was that this phenomenon is due to electrostatic effects, as the initial experimental protocol included a buffer switch at the end of the hybridization: while the hybridization itself was conducted at a very high salt concentration (900 mM NaCl), the wash buffer and melt buffer contained only 4 mM NaCl. To probe this hypothesis, we conducted thermal dissociation experiments in three different buffers, using the newest generation of gel drop arrays: 1) hybridization and melt with the original protocol that prescribes a buffer swich, to observe the phenomenon also in the new gel arrays; 2) hybridization and melt both in the original hybridization buffer (40% formamide, 900 mM NaCl, 20 mM Tris-HCl) without switching buffers but still washing away unbound targets at the end of the hybridization; 2) hybridization in the old buffer and melt in a high salt melt buffer (900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl); 4) both hybridization and melt in the high salt buffer (900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl).

Repeating the experiments with the old protocol of buffer switch rendered qualitatively similar results to those obtained earlier. Note that one should not directly compare the absolute T_d values of these experiments with the old T_d values from the previous generation gel drop arrays due to potential changes in the platforms. However, the trend of decreasing T_d value with increasing initial target concentration was present also in the other three protocols that prescribed high salt concentration for both hybridization and dissociation, with or without a buffer switch, although absolute T_d values varied. Note that even in the 'no buffer switch' experiments a wash

with the hybridization/melt buffer was carried out to wash away unbound targets at the end of the hybridization. Table 13 reports the T_d values for experiments conducted with high salt hybridization and melt buffers (the same buffer used for both hybridization and melt).

Table 12 – Average T_d values, in ^{o}C (temperature at which the signal has decreased by 50%), and standard deviations for three different initial target concentrations. The hybridizing target was perfect-match to probe 1537 and one-base mismatch to probes 1271 and 1538. Hybridization buffer contained 40% formamide, 900 mM NaCl, 20 mM Tris-HCl, and melt buffer contained 900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl. Note the decrease in T_d values as the initial target concentration increases, similarly to earlier observations.

	Initial target concentration		
Probe ID	**10 nM**	**100 nM**	**1 µM**
1537	40.2 ± 0.3	37.8 ± 0.3	35.9 ± 0.2
1271	35.7 ± 0.1	34.7 ± 0.2	32.7 ± 0.2
1538	33.8 ± 0.2	33.0 ± 0.2	31.1 ± 0.2

Table 13 - Average T_d values, in ^{o}C (temperature at which the signal has decreased by 50%), and standard deviations for two different initial target concentrations. The hybridizing target was perfect-match to probe 1537 and one-base mismatch to probes 1271 and 1538. Hybridization and melt buffer contained 900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl. Note the decrease in T_d values as the initial target concentration increases, similarly to Table 12.

	Initial target concentration	
Probe ID	**10 nM**	**100 nM**
1537	54.0 ± 0.5	49.8 ± 0.5
1271	49.7 ± 0.6	45.3 ± 0.5
1538	45.9 ± 1.2	41.3 ± 0.4

This data leads me to believe that while electrostatic effects could still be playing a role in the experiments that include a switch from high salt hybridization buffer to low salt melt, there is an additional reason for the decrease in T_d value with increasing initial target concentration, given that the decrease in T_d values is also seen with very high salt concentration in the melt buffer. Going back to the simulation data presented above, exploring the effect of the initial fraction of probe bound in a duplex on the shape of the dissociation curves, one can see a qualitative similarity between experimentally observed dissociation curves for two different dissociation buffers (see Figure 33). We can also estimate the difference in the fraction of probe bound by

normalizing the intensity of the probe at the beginning of the melt to the Texas Red beacon on the array, assuming identical behavior of beacons on different arrays: we find that the initial signal intensity is five times higher between the initial 1 μM and 10 nM target concentrations in case of the experiments with high initial salt concentration during hybridization and low salt concentration during a melt; and 13 times higher in case of the high salt hybridization and melt experiments. We propose that at least part of the observed decrease in T_d value as a function of initial target concentration and indirectly initial bound duplex concentration is due to the vacant probe sites slowing down the departure of dissociated target molecules during a non-equilibrium melt.

Hybridization buffer contained 40% formamide, 900 mM NaCl, 20 mM Tris-HCl, and melt buffer contained 900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl

Hybridization and melt buffer: 900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl

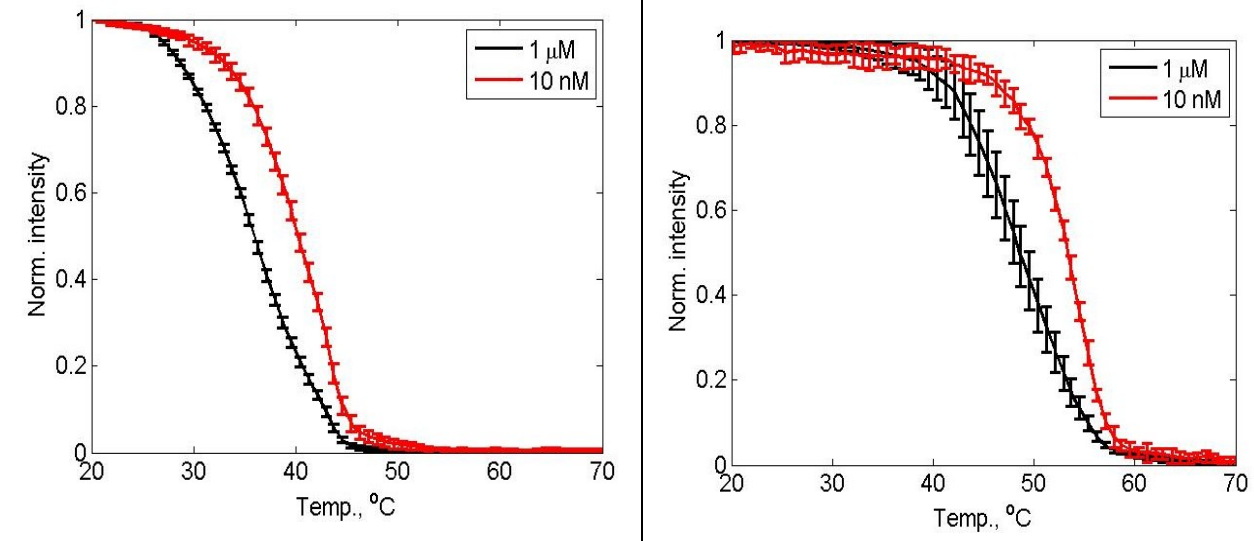

Figure 33 – Normalized thermal dissociation curves from probe 1537, using two different buffer compositions, initial target concentration either 1 μm or 10 nM. Hybridizing target was a perfect complement to probe 1537. Note the shift to the left in the dissociation curves with higher initial target concentration.

Simulating thermal dissociation: compartmental model versus 3D COMSOL model

The remainder of this Chapter will discuss optimization of thermodynamic parameters in the model to achieve a quantitative fit between the model and the experiment, specifically for thermal dissociations. To determine whether for modeling non-equilibrium thermal dissociation a simple compartmental model could be used in lieu of the significantly more complicated finite element three-dimensional model, we compared simulated thermal dissociation curves between a compartmental model (1- and 2-compartments) and 3D COMSOL finite element model of three-dimensional gel microarray.

In the one-compartment model, the target has a very slim chance of rehybridizing after it has unbound from the probe: this is due to accounting for different volumes in the equation corresponding for free target. Therefore, the results from the one-compartment model were compared with those from a 3D COMSOL model that did not include target rebinding after dissociation – *i.e.* the association constant k_a was equal to 0 in the 3D model. All other input parameters were the same between the one-compartment model and the 3D model, no parameters were optimized. Temperature-dependence was introduced through the dissociation constant k_d as explained above. Excellent agreement between the two models is displayed in Figure 34. The same was true for a significantly lower probe concentration ($6 \cdot 10^{-6}$ mol/m^3), and for the initial condition of 5% of total probe bound.

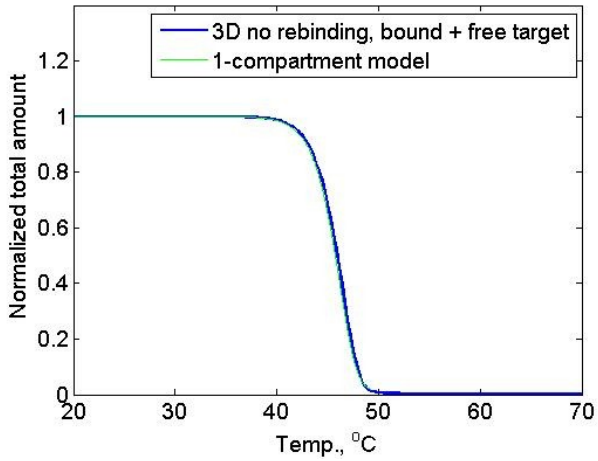

Figure 34 - Normalized simulated thermal dissociation curves using one-compartment model and 3D COMSOL model with no target rebinding after dissociation. Temperature ramp rate in the model was 0.83 °C/min. Note how the two curves overlap completely. Initial conditions for both models: initial probe concentration $60 \cdot 10^{-3}$ mol/m^3, 100 % probe bound at time t=0, ΔH=-148 kcal/mol, ΔS=-403.4 cal/(mol·K), $D_s=10^{-10}$ mol/m^2, $D_g=10^{-11}$ mol/m^2. For compartmental model, $k_a=10^6$ 1/(M·s), for 3D model k_a=0. Assumed temperature ramp rate: 0.83 °C/min.

For solving the two-compartmental model, the mass transfer coefficient k_m was considered an open parameter – one whose value can be optimized. The question of interest was whether the mass transfer coefficient is highly dependent on the initial fraction of probe bound and/or kinetic constants when simulating thermal dissociation – if yes, then the mass-transfer coefficient would need to be re-optimized for every case of modeling thermal dissociation. Figure 35 depicts the results of simulations where the mass transfer coefficient k_m was first optimized to fit the 3D simulations of thermal dissociation from a 100 % saturated probe with perfect-match kinetics, and the *same* k_m value was used to then simulate 1) dissociation with mismatch kinetics; 2) dissociation starting at 5 % initial saturation level. Optimization of k_m for perfect-match kinetics and 100 % of probe initial bound rendered k_m value of 0.0671 1/s (solid green line in the left panel of Figure 35). Simulating the dissociation from a probe that was occupied at a lower level by target at time 0 (5 % bound) rendered a shift in dissociation curves, but using the k_m value of 0.0671 1/s still resulted in good agreement between the compartmental and three-dimensional

model (dashed lines in the left panel of Figure 35). The good agreement was present also when repeating the same simulations or mismatch kinetics, while still assuming k_m=0.0671 1/s (right panel of Figure 35)

In conclusion, the results show close agreement between compartmental model and 3D COMSOL model for all four cases, although the mass transfer coefficient was only optimized for one of the four scenarios. This suggests that the two-compartment model is a good approximation for the 3D COMSOL model for thermal dissociation modeling in the three-dimensional binding elements, and the mass transfer coefficient does not need to optimized for every new set of initial conditions. Therefore, the two-compartment model was used later in this work for optimizing thermodynamic parameters in lieu of the finite element model.

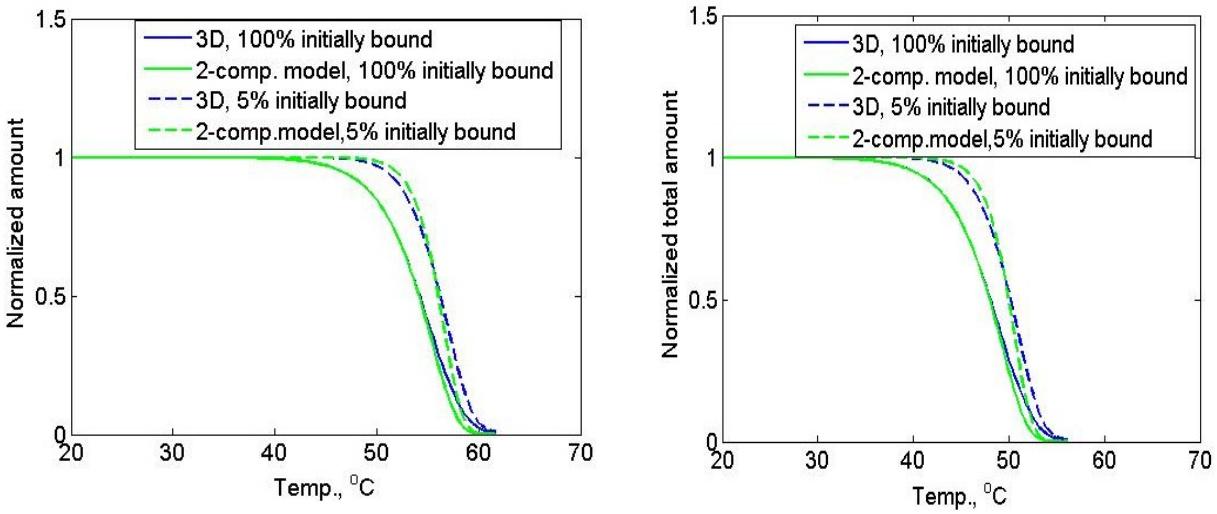

Figure 35 - Normalized thermal dissociation curves, comparing the results from a 2-compartment model and 3D COMSOL model. Left panel – assuming perfect-match kinetics, comparison of simulations with 100% and 5% initially bound probe. Mass transfer k_m was optimized for perfect-match kinetics only, assuming that 100% of probe was initially bound in a duplex. Right panel – assuming mismatch kinetics, comparison of simulations with 100% and 5% initially bound probe with no optimization of k_m. Note the close agreement between compartmental model and 3D COMSOL model for all four cases, although the mass transfer coefficient was only optimized for one of the four scenarios. Initial conditions for both models: initial probe concentration $60 \cdot 10^{-3}$ mol/m^3, ΔH=-148 kcal/mol, ΔS=-403.4 cal/(mol·K), $D_s=10^{-10}$ mol/m^2, $D_g=10^{-11}$ mol/m^2. Association constant $k_a=10^6$ 1/(M·s). Assumed temperature ramp rate 0.83 °C/min.

Optimizing thermodynamic parameters: is there a link between the solution and microarray parameters?

As we have shown in Chapter 2 (see *e.g.* Figure 26), nearest-neighbor based thermodynamic parameters describe the solution-based melt curves with good accuracy for 18-19 basepairs long perfect-match duplexes and one-base mismatches. However, as we saw earlier in this Chapter, using the same thermodynamic parameters to describe binding and dissociation in gel-based microarrays leads to overestimation of duplex stability in the predictions. In this section, we are testing the hypothesis that the thermodynamic parameters of DNA binding in microarrays, ΔH^o_A and ΔS^o_A, can be represented by solution-based parameters in the following way:

$$\Delta H_A^0 = c\Delta H_S^0 + d$$

$$\Delta S_A^0 = a\Delta S_S^0 + b,$$

where ΔH^o_S and ΔS^o_S are the thermodynamic parameters in the solution, and a, b, c, and d are scaling constants that are obtained via fitting the data to the equations above. Table 14 lists the solution-based thermodynamic parameters for four perfect-match probes used in the experiment. Specifically, we are going to assume that these are global constants, applicable to every probe on one platform. Built into the hypothesis are two assumptions: 1) the dependence of ΔS^o_S and ΔS^o_A on salt concentration is the same; 2) electrostatic effects do not play a role.

Table 14 – Nearest-neighbor based calculated thermodynamic parameters for probes 1537, 62, 65 and 323, assuming the hybridization of a perfect-match target. Buffer contains 50 mM NaCl.

Thermodynamic parameter	Probe 1537	Probe 62	Probe 65	Probe 323
ΔH^o, kcal/mol	-148	-143.4	-139.9	-146.1
ΔS^o, cal/(K·mol)	-422.5	-401.7	-390.2	-418.8

As we have shown above, the fraction of probe bound at the beginning of the thermal dissociation can affect the shape of the dissociation curve. To minimize this difference between

the probes when calculating global scaling constants, we used a very high target concentration to saturate all probe sites in the experiments (10 μM) and long hybridization time before starting the dissociation, 24h. We also used a salt concentration similar to the one used in experiments for Chapter 2 for best comparison.

As a first pass, we fitted the thermodynamic parameters to the thermal dissociation model independently for all four perfect-match probes. The experimental curves and model best fits are given in Figure 36. Figure 37 plots the values of optimized thermodynamic parameters against the nearest-neighbor parameters. Note that for either thermodynamic parameter, we obtain a reasonably good linear fit, meaning one can use the solution-based parameters to predict microarray thermodynamic parameters for perfect-match duplexes, using the linear relationship obtained from select probes.

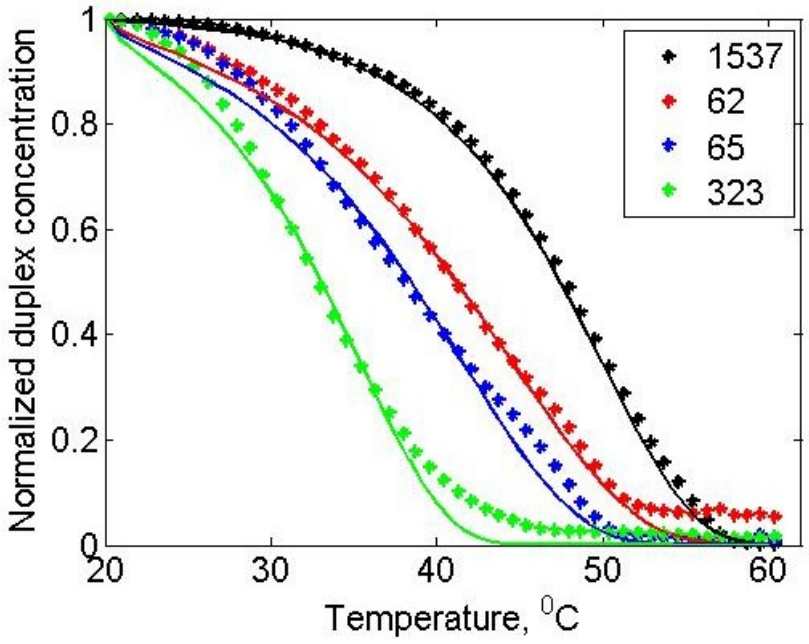

Figure 36 – Normalized thermal dissociation curves from four different probes, corresponding to four different perfect-match duplexes. Solid lines refer to model simulations, symbols to mean experimental values. Experimental hybridization and melt buffer consisted of 50 mM NaCl, 20 mM Tris-HCl. Initial target concentration was 10 μm. Note the close correspondence between the model simulations and experimental curves.

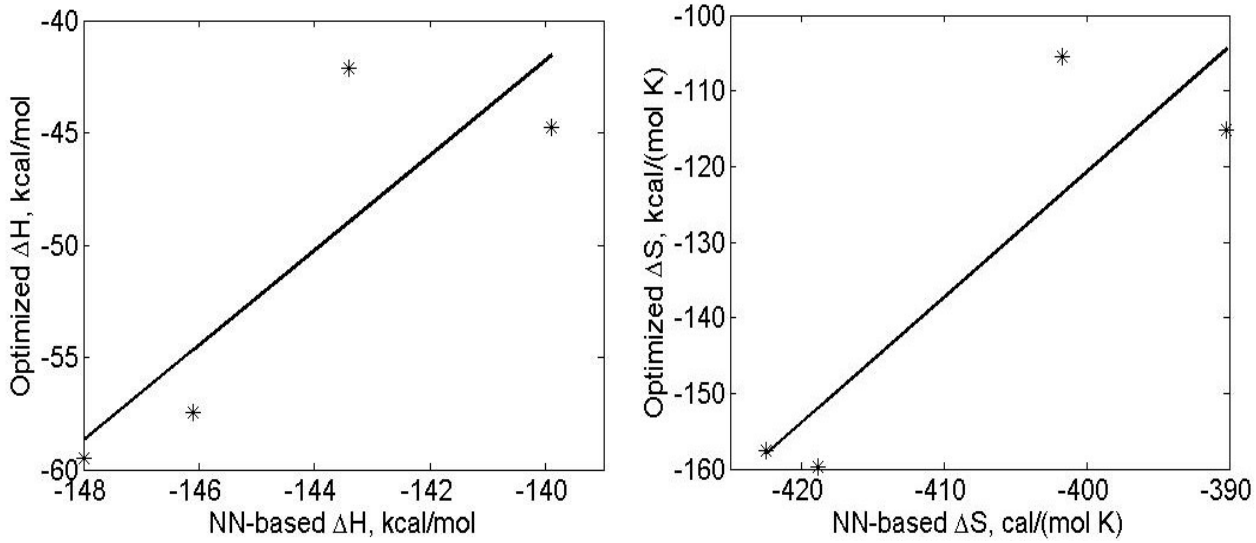

Figure 37 – Optimized and nearest-neighbor based solution thermodynamic parameters - ΔH° in the left, and ΔS° in the right. Linear regression for ΔH° yielded y=2.35x+282.39, R^2=0.72; for ΔS° y=1.75x+572.90, R^2=0.79.

Using the same linear relationships between nearest-neighbor and microarray thermodynamic parameters for mismatch probes gives so large dissociation constant values that the duplex is predicted to dissociate in an instant, rendering a duplex concentration of 0 for any time point after time t=0. The corresponding nearest-neighbor based thermodynamic parameters for the mismatch duplexes are given in Table 15. This leads us to conclude that in the case of the probes utilized in this work, the linear relationship that holds between the nearest-neighbor parameters and microarray thermodynamic parameters for perfect-match duplexes does not predict the parameters for mismatch duplexes: the latter appear to be more stable in the microarray platform than predicted.

Table 15 - Nearest-neighbor based calculated thermodynamic parameters for probes 1538, 399, 407 and 1282, assuming the hybridization of a target complementary to probes 1537, 62, 65 and 323, respectively. Buffer contains 50 mM NaCl.

Thermodynamic parameter	Probe 1538 (MM to 1537)	Probe 399 (MM to 62)	Probe 407 (MM to 65)	Probe 1282 (MM to 323)
ΔH°, kcal/mol	-130.5	-120.4	-128.9	-124.9
ΔS°, cal/(K·mol)	-384.5	-344.6	-365.4	-368.9

A note about reusability and reproducibility of gel-based microarrays

While the scope of a study of reusability and reproducibility of the three-dimensional gel-based microarrays is not in the scope of this work, a short discussion of the issues as they relate to the development of applications for the microarray platform is presented below. Note that these observations should not be generalized to all gel-based platforms, but simply serve to answer questions about this specific commercial platform as the manufacturer in our experience has been rather opaque about performance expectations. Specifically, we looked at the reproducibility of two metrics on the newest generation of gel-based microarrays: 1) absolute T_d values; 2) the difference in T_d values between a probe that is a perfect match to the target and a probe that is a one-base mismatch. These metrics were compared in experiments with new arrays stored at different amounts of time and reused arrays. Overall, the absolute values of T_d between fresh and used arrays could differ by as much as 5 ^oC; the absolute T_d values could also differ by 2-3 degrees between new arrays that were stored at different amounts of time unused (although in the same conditions – at 4 ^oC in the dark). The used arrays were stored in water at 4 ^oC in the dark. This cautions to develop applications focusing just on the absolute T_d values as indicators of duplex stability – the differences between the absolute values were often the same magnitude as those seen between perfect-match and mismatch probes. However, comparing the differences between T_d values, collected from perfect-match and mismatch probes, rendered very consistent results irrespective of the length of time the arrays had been stored or whether they were used as new, after one use or two uses. Figure 38 shows the ΔT_d values between a perfect-match probe and three different one-base mismatch probes, combining those measures from two new slides, used in experiments four months apart, and 2 used slides.

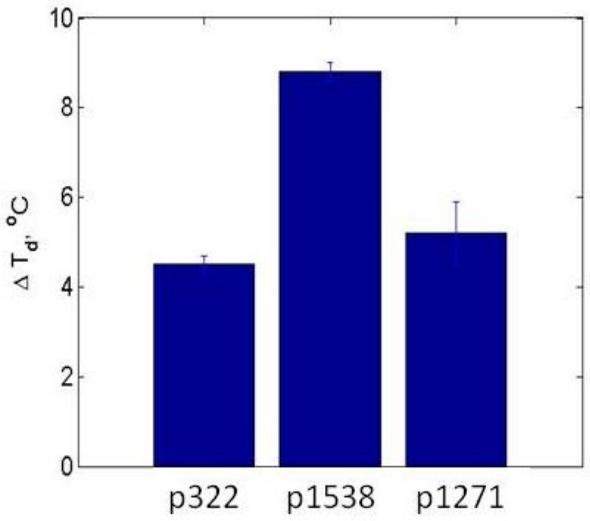

Figure 38 – Differences in T_d values between a perfect-match probe 1537, and three different mismatch probes 322, 1538 and 1271 following a single-component hybridization of a target complementary to probe 1537. Hybridization and melt buffer contained 900 mM NaCl, 5 mM EDTA, and 20 mM Tris-HCl. Initial target concentration 100 nM. Thermal dissociation ramp rate was 0.83 °C/min.

3.5 Conclusions

In this Chapter, we have collected experimental hybridization and melt data from gel-based three-dimensional microarray platforms for a number of different probe sequences and at different target concentration for single target analytes. We also analyzed the predicted perfect-match/mismatch signal ratios and T_d values using a one-compartment mathematical model, employing thermodynamic parameters calculated from solution-based nearest-neighbor parameters. Building on the solution-based model, we developed a three-dimensional finite element model in COMSOL for modeling the diffusion of analyte in the bulk solution, within the gel elements, and binding of analyte onto immobilized probe molecules. We used the model to compute the effect of intra-gel diffusion coefficient on the simulations, the contribution by the free target inside the gel element to the overall amount of target in the binding elements, and the effect of the fraction of probe bound in a duplex at the beginning of the hybridization on the

thermal dissociation curves. For optimizing thermodynamic parameters for thermal dissociation modeling, we developed and validated a two-compartment mathematical model. We were able to obtain good agreement between model simulations of thermal dissociations and experimental data, and saw a linear relationship between the nearest-neighbor based thermodynamic parameters and those optimized for the microarray model.

The findings can be summarized as follows:

1) The melting temperature T_d is lower in the microarray than predicted by the model or observed in the solution;
2) The perfect-match/mismatch signal intensity ratio corresponds to dissociation constants significantly larger than those predicted by the model;
3) The differences between T_d values from different probes are predicted accurately by the model using solution-based parameters;
4) T_d values decrease as the target concentration increases, which is more pronounced the higher the original signal intensity – even in experiments with high salt concentration;
5) Higher fraction of probe bound at the beginning of the thermal dissociation lowers the T_d value in the model, similarly to experimental results;
6) In the three-dimensional finite element model, lowering intra-gel diffusion coefficient by an order of magnitude will not substantially affect the hybridization nor thermal dissociation simulations;
7) In the three-dimensional finite element model, free target in the gel element contributes very little to the overall amount of target in the probe;
8) A two-compartment model can substitute for three-dimensional finite element model for modeling non-equilibrium thermal dissociation;

9) Thermodynamic parameters for perfect-match duplexes can be predicted using a linear relationship between microarray parameters and nearest-neighbor based parameters.

Chapter 4: Competitive binding of targets in the microarray

4.1 Objectives

The purpose of this Chapter is to extend the previous work in Chapters 2-3 in two specific ways. First, to introduce targets that are competing for binding sites on the same probe into the experimental system and into the model, mimicking hybridization and melt in conditions similar to real-world applications of microarrays, involving a potentially complex system of targets of similar sequences. We will explore the utility of thermal dissociation curves in a three-dimensional microarray after a multicomponent binding for distinguishing between related targets. Second, we introduce convective transport into the experimental system to mitigate mass transport limited hybridization in static systems.

The objectives of this Chapter are the following:

a) collect experimental melt data of multiple target melt (two targets of closely related sequences, 1537 and 1271 in gel drop array), varying the relative concentration of targets;
b) collect hybridization data of multiple targets with two different labels to track each individually;
c) extend the microarray hybridization model from previous Chapter to include a second target and validate the model;
d) characterize the performance of the microfluidic recirculation platform and collect hybridization and melt data with the recirculating microfluidic device;
e) characterize the utility of thermal dissociation after a multicomponent hybridization to resolve between related sequences.

4.2 Theory and Design Considerations

In the previous Chapter, a three-dimensional model of a gel-based microarray platform was developed, and experimental data on single target hybridization in static conditions was collected to support the model. While this is an important step in learning about the physico-chemical processes in the model, the ultimate objective of this work is to inform the design of DNA microarray-based biosensors. The applications for these kinds of biosensors are likely to include targets of mixed composition and also put a premium on reducing the time it takes to complete a measurement with desired sensitivity and specificity. In this Chapter, we extend the single analyte model to a multi-analyte model, investigate multicomponent melts in a controlled system, and incorporate recirculation of target sample. An updated schematic from Chapter 3 of the processes considered is shown in Figure 39.

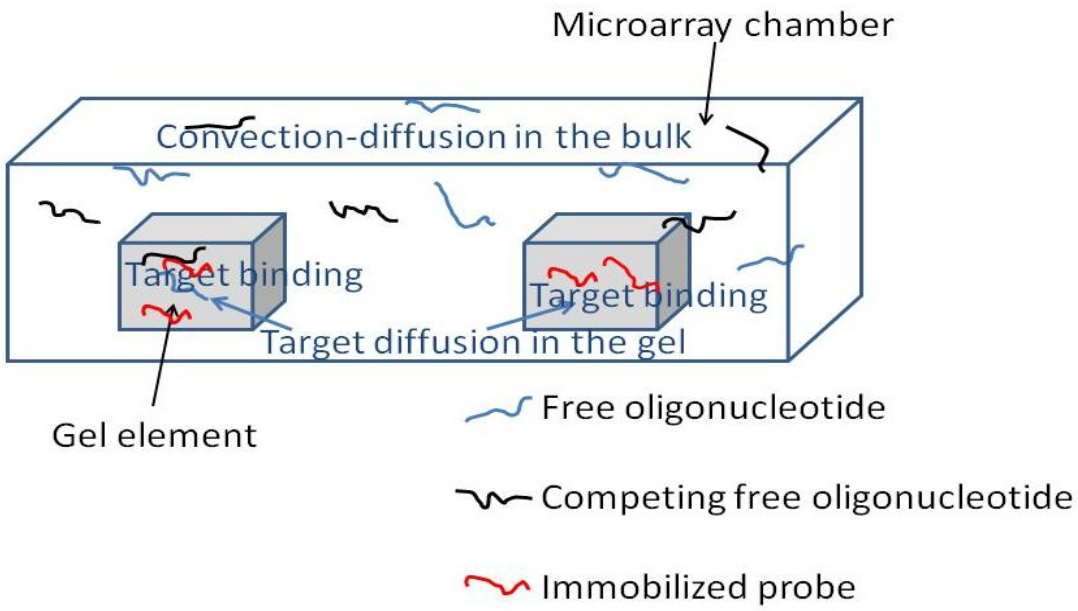

Figure 39 - A simplified schematic of the geometries and processes included in the microarray model. Three-dimensional gel-based binding elements are embedded in a microarray chamber. Target oligonucleotides are transported via convection and diffusion into the gel elements where they can bind to form complementary duplexes with probe oligonucleotides, immobilized in the gel matrix. Schematic not drawn to scale.

Up until only a few years ago, studies aiming to model the transport and binding of DNA oligonucleotides onto DNA microarrays only considered the hybridization of one target at a time. While this certainly can give much insight into the hybridization process, and the hybridization of multiple oligonucleotides with dissimilar sequences could be viewed as independent events, the simultaneous hybridization of multiple related targets, however, *cannot* be approximated as a superposition of single-target hybridizations. This has been shown in several recent publications for planar microarrays, through both computational and experimental work [29, 113, 122, 123, 125, 142]. Melting in a multi-component system on microarrays has not emerged as a topic of thorough investigation until recently. Multiplex DNA melting in planar arrays under equilibrium conditions was recently explored by Williams *et al*. [143], concluding that equilibrium dissociation curves do not help in resolving heterogeneous duplexes in the probe limited regime. While non-equilibrium thermal dissociation as a potential tool to distinguish between related sequences even in complex systems has been suggested in the literature [51, 52, 58], we are not aware of studies examining the non-equilibrium melt with a controlled set of closely related targets in a gel-based microarray, attempting to describe the system with a computational physics-based model, as opposed to statistical methods.

In this Chapter, we are characterizing the effect of competitive binding on hybridization and melt signals for gel-based microarrays. For collecting experimental data, we are again using artificial samples of synthetic oligonucleotides with known sequences and concentrations, as the collected data will be compared with the results from the extended mathematical model, describing the simultaneous binding of multiple analytes. Slow DNA diffusion is one of the reasons for why it can take a long time to reach equilibrium in the microarray hybridization, and a number of

approaches to overcome mass transport limitations have been proposed in the literature (see Background section for a summary). For sample recirculation, we are using a microfluidic recirculation platform developed in collaboration with Micronics, Inc. (Redmond, WA). The main emphasis of this Chapter is to consider the implications of multicomponent binding and subsequent thermal dissociation on resolving between closely related DNA sequences.

4.3 Materials and Methods

Experimental platform

The microfluidic microarray hybridization chip was fabricated in partnership with Micronics (Redmond, WA), and is displayed in Figure 40. The card features a pneumatically-activated pump for solution recirculation within the card itself under the control of a microfluidic controller (Micronics' MicroFlow System), which is, in turn, controlled by a computer (panel B in Figure 41). Each of the valves on the card can be controlled individually. The hybridization/wash solution is loaded into the card manually with a pipette, isolating the pump line and hybridization chamber via closing off the respective valves for separate loading. The gel element array is situated in an imaging chamber for continuous monitoring of fluorescence intensity of the gel pads during hybridization and melting (Figure 40). The total volume of the recirculating loop within the card is around 40 μL. During recirculation, the valves connecting the inlet and outlet channels with the main fluidic loop are closed off, allowing for recirculation of the target solution upon executing the pump line. The microfluidic card is placed into a manifold, interfaced with the MicroFlow System (panel B in Figure 41), which in turn is placed onto a thermal table that is connected to a thermoelectric temperature controller and a water bath. Imaging is permitted with the custom-built fluorescence microscope (Argonne National

Laboratory), allowing on-line imaging without removing the card from the MicroFlow System's manifold.

Materials

Experiments were performed with synthetic custom-ordered labeled oligonucleotides (Eurofins MWG Operon, Huntsville, AL) that had perfect matches and mismatches among the probes displayed on the microarrays. The sequences of probes used on the two different microarrays are the same as in Chapter 3, but are reproduced here for reference (see Table 16). The targets used in gel drop arrays presented a perfect-match to one of the probes EUB338, probe 1537, probe 1271, or probe 62. In gel pad experiments, the target was either an 18-base perfect match to probe EUB338, or a 38-base sequence that in its middle contained an 18-base sequence perfectly complementary to probe EUB338 (5'-AACCACACCAACTCCTACGGGAGGCAGCACCACACCAA-3'). Table 17 shows the relative concentrations of perfect-match and mismatch targets used in competitive binding experiments on gel drops. The targets were labeled with either Texas Red or OregonGreen488 fluorophore at the 3′ terminus. Microarray hybridization buffer contained 900 mM NaCl, 20 mM Tris-HCl, and 40% formamide [52], while the melt buffer contained 4 mM NaCl, 20 mM Tris-HCl, and 5 mM EDTA. The two gel microarray platforms used in this Chapter, gel pad and gel drop array (gel binding elements printed on a microscope slide), were obtained from Akonni Biosystems (Frederick, Maryland).

Table 16 - DNA oligonucleotide probe sequences used in Chapter 4. Mismatch bases are highlighted in red and underlined.

Sequence code	5'→3' sequence
Gel pad array	
EUB338	GCTGCCTCCCGTAGGAGT
EUB338MM	GCTGCCTCCCCTAGGAGT
Gel drop array	
1537	CTCACACACGTTCTTGACT
1271	CTCACACACGTTCTTCACT
1538	CTCACACACCTTCTTGACT
62	GACGGGCGGTGTGTACAA
399	GACGGGCGGTGTCTACAA

Table 17- Concentrations of targets to probes 62, 1537 and 1271 used in competitive binding experiments (A, B, C, D). Each row in the table represents a separate experiment.

Exp. ID	Target to probe 62	Target to probe 1537	Target to probe 1271
A	20 nM	10 nM	10 nM
B	210 nM	10 nM	200 nM
C	400 nM	200 nM	200 nM
D	210 nM	200 nM	10 nM

Experimental protocols

Experiments with gel drop arrays were conducted covering the hybridization or melt solution with a hybridization chamber as discussed in the Methods section of Chapter 3 (panel A in Figure 41), while gel pad array experiments were conducted using the microfluidic platform with and without sample recirculation. In both platforms, a wash was conducted at the end of the hybridization and before starting the melt, washing away unbound target. Hybridization experiments were conducted at 20 ^{o}C, while in the melt experiments, the temperature was ramped from 20 to 70 ^{o}C in the span of an hour, rendering a ramp rate of 0.83 ^{o}C/min. The first generation of gel drop arrays were hybridized in the dark, with total time of hybridization 17-18 h, since only end-of-hybridization signal and subsequent melt data (imaging every minute, 1500 ms exposure time) was collected. The second generation of gel drop arrays was also used for continuous imaging of hybridization. Gel pad arrays were continuously imaged in the

microarray reader during the hybridization as well as the melt (image was taken every 10 minutes, 1500 ms exposure time).

Microarray reader, image capture and data analysis

For details on single-color imaging, the reader can see Methods section in Chapter 3. For multicolored imaging, the oligonucleotides were labeled with Oregon Green and Texas Red labels that were characterized in Chapter 3, corresponding to Oregon Green and Texas Red beacons on the array. Instead of tracking the emission from one fluorophore during the experiment, dual color imaging was used. Peak excitation and emission wavelengths for Texas Red and Oregon Green (488) fluorophores are 596 nm and 615 nm for Texas Red, and for Oregon Green 496 nm and 524 nm, respectively, giving rise to little overlap in the excitation and emission spectra. Multicolor imaging was accomplished by switching the excitation and emission filters in the microscope after each image. As described in the previous Chapter, the sample was only exposed to light when an image was being taken; otherwise, the sample was shielded from light by a closed mechanical shutter in the light path. The filter change was automated as follows: a LabView program was designed to interface with the current reader system and to automatically change the filter as specified at the beginning of the experiment immediately after each image capture (implemented by Mike Purfield in the Yager lab, with help from Peter Kauffman). The time between taking each image, irrespective of the filter set, was constant, meaning the data for each fluorophore (filter set) was not captured at the same exact time, which was taken into account when analyzing images. Both real-time hybridization and subsequent melt data were recorded.

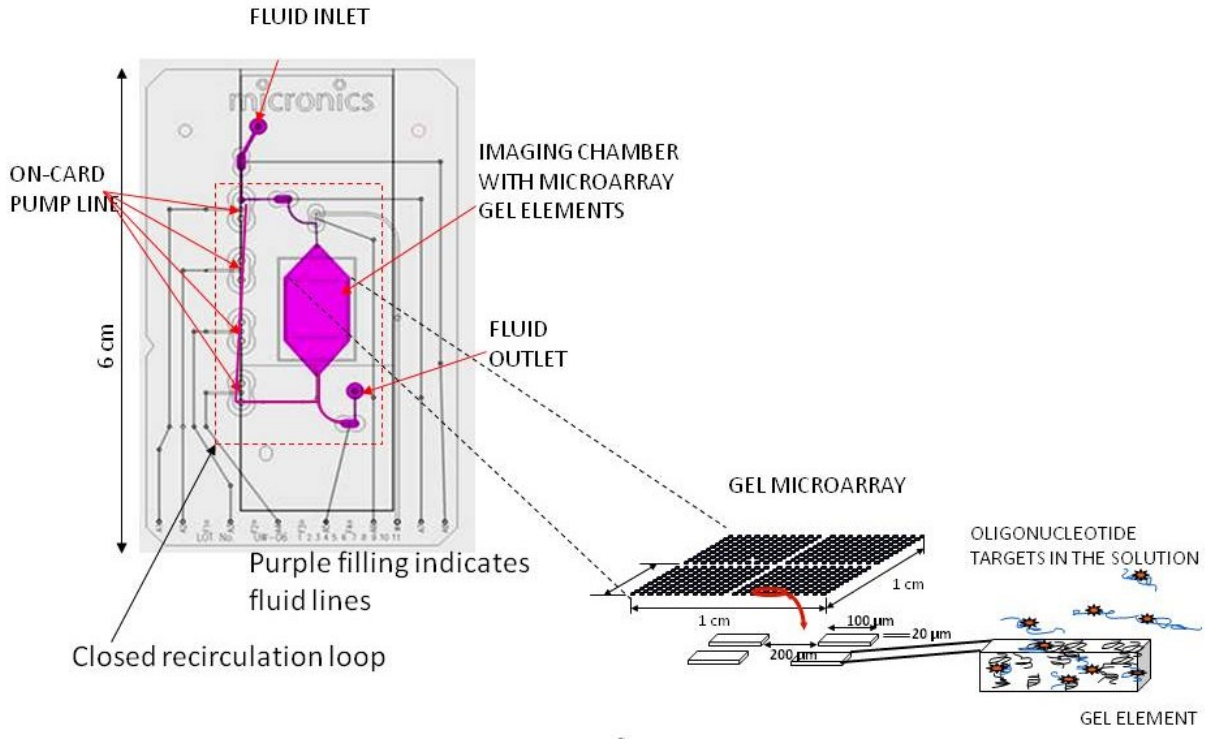

Figure 40 – Microfluidic microarray platform used in Chapters 3 and 4 of this dissertation. Blow-up of the microarray imaging chamber shows the gel pad microarray.

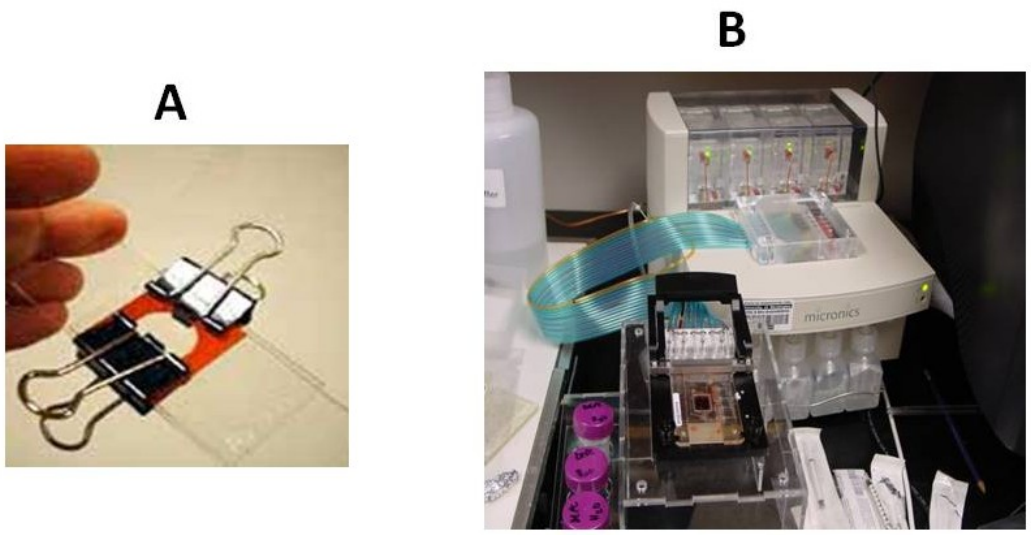

Figure 41 – Two experimental platforms used in Chapters 3 and 4 of this dissertation: a hybridization chamber affixed to a microarray glass slide (panel A); microarray glass slide embedded in a microfluidic chip and interfaced with a microfluidic controller (Micronics, Redmond, WA).

Computational

Competitive binding

The model developed in Chapter 3 served as a basis of model development for this Aim. The only modification with respect to considering competitive binding was adding binding of a second analyte to the mathematical description to account for multi-species competitive binding. Instead of one binding reaction, keeping track of the binding of one target onto one probe, we considered the binding of two targets onto two probes:

$$\frac{dB_{m,i}}{dt} = k_a \cdot C_i \cdot \left(P_{m0} - B_{m,i} - B_{m,j}\right) - k_{di,m} \cdot B_{m,i}$$

$$\frac{dB_{m,j}}{dt} = k_a \cdot C_i \cdot \left(P_{m0} - B_{m,i} - B_{m,j}\right) - k_{dj,m} \cdot B_{m,j}$$

$$\frac{dB_{n,i}}{dt} = k_a \cdot C_j \cdot \left(P_{n0} - B_{n,i} - B_{n,j}\right) - k_{di,n} \cdot B_{n,i}$$

$$\frac{dB_{n,j}}{dt} = k_a \cdot C_j \cdot \left(P_{n0} - B_{n,i} - B_{n,j}\right) - k_{dj,n} \cdot B_{n,j,}$$

where the subscripts *m* and *n* refer to two different probes, subscripts *i* and *j* to two different targets. $B_{m,i}$ marks the bound concentration of target i on probe m, $B_{m,j}$ marks the bound concentration of target j on probe m, and so on. P_{m0} and P_{n0} stand for the initial concentrations of probes m and n. C_i and C_j stand for the free concentration of targets i and j in the solution, and $k_{di,m}$, $k_{dj,m}$ *etc.* denote the dissociation constants for target *i* in a duplex with probe *m*, and target *j* in a duplex with probe *m*, respectively. At time 0 at the beginning of hybridization, all duplex concentrations were assumed to be zero. The diffusion of both analytes in the solution was described using Fick's Law.

For modeling thermal dissociation, the initial free target concentration was assumed to be 0, and the fraction of the total probe bound by each possible target was used as an input parameter for the simulations.

4.4 Results and Discussion

Assessing fluid recirculation in the microfluidic platform

The performance of the fabricated card, connected to the Micronics Microflow System™, was tested first to verify the recirculation of the fluid within the microfluidic card, and second, to estimate the range of circulation flow rate achievable by executing the chip. Recirculation of liquid sample was visually confirmed using two differently colored samples, loading one into the imaging chamber, and the other into the pump line (Figure 39). The flow rate within the chip is controlled by the timing of pump execution, which in turn is controlled by the computer-operated microfluidic controller (MicroFlow System™, Micronics, WA). The timing of pump execution is controlled by the delay time specified in the executable script for the MicroFlow System. A range of valve timings was tested by weighing the amount of water pumped out of the card when the recirculation circuit is broken by an appropriate selection of valve states. Figure 43 shows the flow rate as a function of valve timings, measured by weighing the amount of water pumped out.

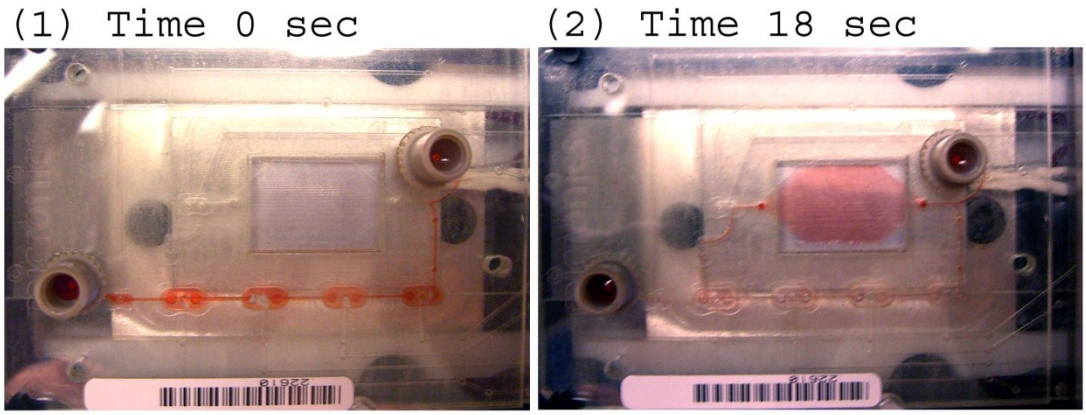

Figure 42 – Recirculation of sample in the microfluidic platform. At time 0, red liquid has been loaded into the pump line, and clear liquid into the imaging chamber (left). Upon starting the execution of the on-card pumps, the sample is recirculated (right). Photo courtesy of Joshua Thompson.

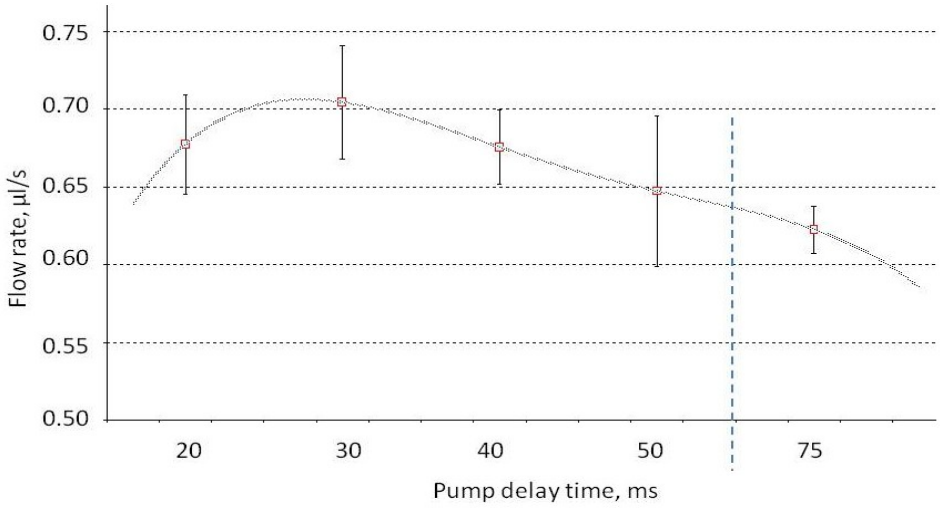

Figure 43 – Mean flow rate in the microfluidic hybridization platform as a function of delay time between pump executions. The experiments were carried out at 30 °C, using water. Data courtesy of Joshua Thompson.

Target recirculation in the microfluidic platform

Initial experiments, employing this microfluidic hybridization platform, characterized the rate of hybridization and temperature-dependent dissociation of targets from the DNA probes under static and dynamic (recirculation) conditions as a function of target length and concentration.

Figure 44a and Figure 44c display the results of hybridizing a synthetic target to perfect-match (PM) and mismatch (MM) probes, demonstrating that recirculation accelerates the hybridization rate (more pronounced at lower target concentrations). Also, the magnitude of signal intensity is dependent on the length on the target. Figure 44b displays the T_d values (temperature at which the normalized hybridization signal has dropped to 50% of its original value) collected after hybridization during dissociation for 18-and 38-base targets under static and dynamic conditions. In addition to observing the expected reduced stability (lower T_d value) for the mismatched hybrid, relative to the perfect match, the results shown in Figure 44 also demonstrate a decrease in apparent T_d for the 18-nucleotide target with recirculation compared to static dissociation.

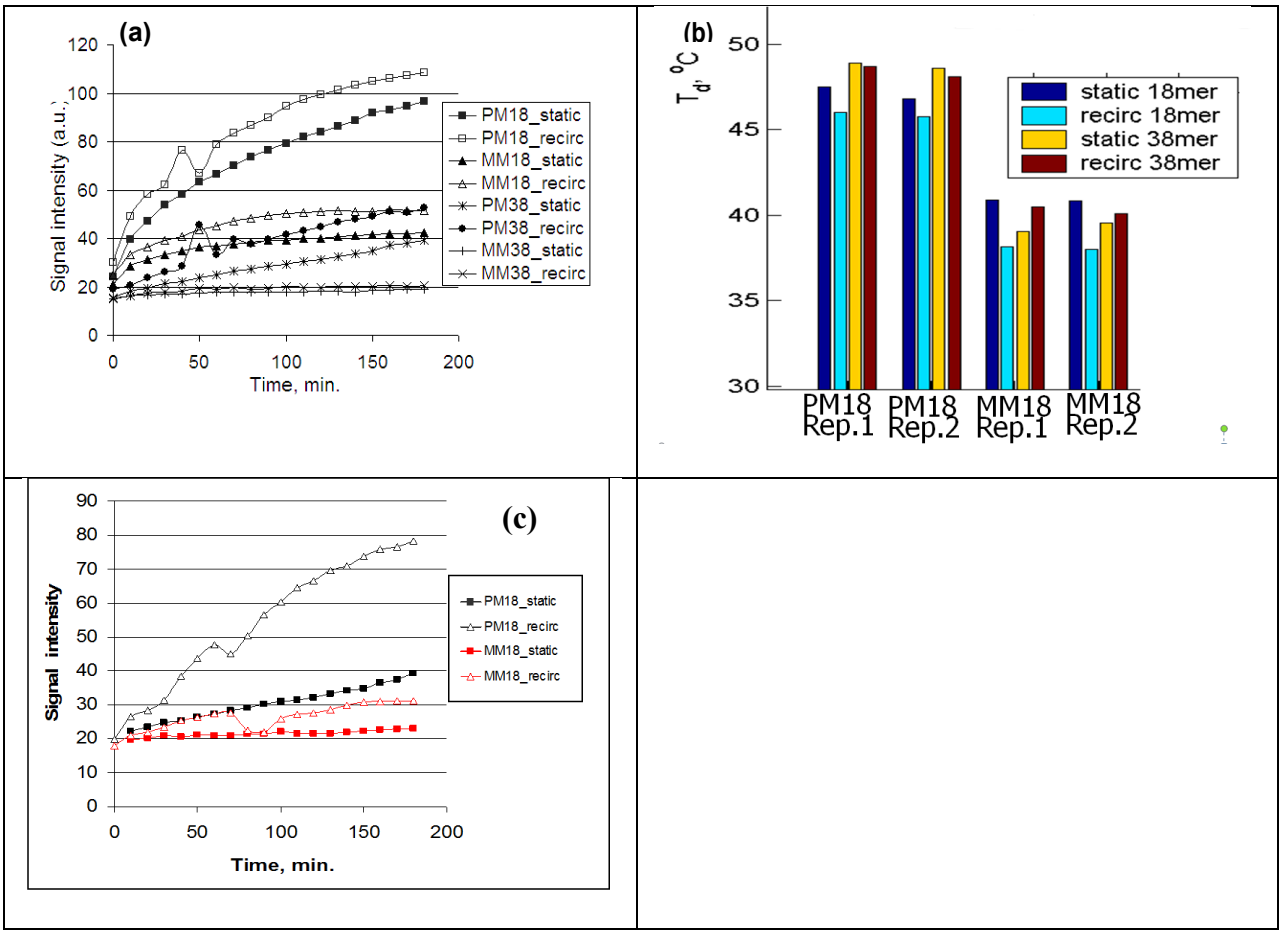

Figure 44 - Isothermal hybridization and thermal dissociation of target complementary to EUB338 probe. Panel (a) displays the hybridization signal as a function of time under static and dynamic conditions; panel (b) displays T_d, temperature at which the normalized signal during a thermal dissociation is 50% of its original value; panel (c) displays the hybridization signal as a function of time under static and dynamic conditions for a 10 times less concentrated target solution than panel (a). PM and MM refer to signal from EUB338 and EUB338MM probes, 18 and 38 refer to the length of the target in number of bases, under static and recirculation (recirc) conditions, respectively. Rep.1 and Rep.2 on panel (b) refer to two replicas in the same experiment.

Competitive hybridization and melt in the three-dimensional microarray platform

At first, two targets, one complementary to probe 62 and the other to probe 1537, were hybridized simultaneously, and the results compared with single-target hybridizations. As expected, since these dissimilar targets do not compete for hybridizations onto the same probe, the hybridization patterns of their co-hybridization were superpositions of their individual hybridization experiments.

This was not the case when using related target sequences. Figure 45 displays the end-of hybridization signal intensity ratios from probes 1537 and 1271 for competitive binding experiments with four different relative target concentrations. As expected, the PM/MM signal ratio changes significantly as the relative concentrations of targets vary. While the PM/MM >1 values for experiments does seem to correctly suggest, irrespective of the presence of mismatch target at low levels, that a high affinity target to probe 1537 is present (experiments A and D), the interpretation of the signal intensity ratios from experiments B and C is more complex. The ratios from these experiments could also stem from scenarios that do not include the presence of EUB338 target: for example, simply high level of mismatch target (B), or even higher level of mismatch probe that would overwhelm both probes, irrespective of degrees of affinity.

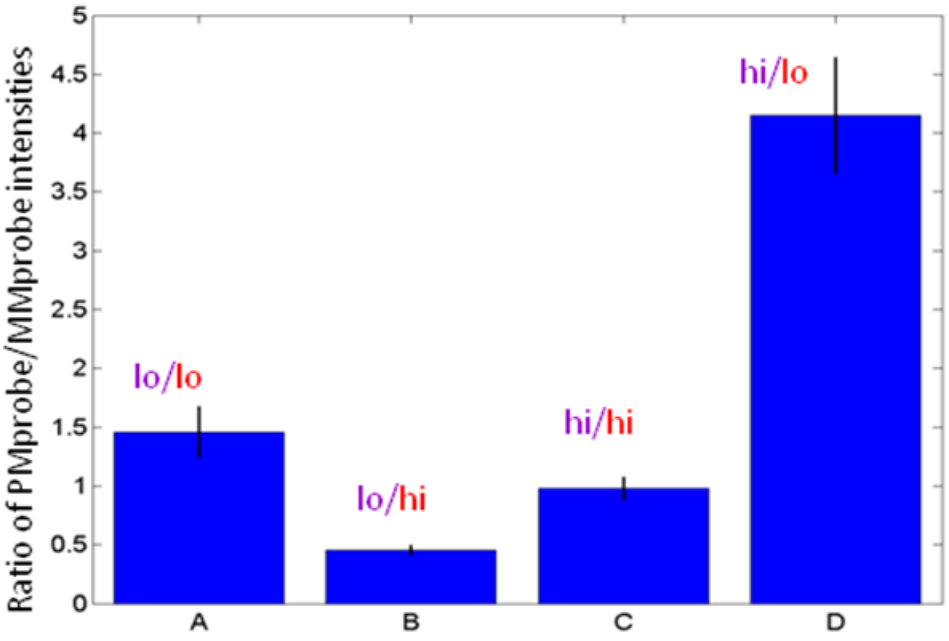

Figure 45 - Ratio of signal intensities from probes 1537 and 1271 in competitive binding experiments A-D on the gel drop microarray (Table 17). Text above the bars refers to the relative concentrations of perfect-match and mismatch targets: purple text to perfect-match, and red to mismatch. n=4.

Table 18 lists the T_d values from experiments A-D for different competitive binding scenarios. The typical approach to using melt curves for discrimination between related targets relies on the assumption that the duplex formed on a mismatch probe will have a lower T_d than the perfect-match duplex. This indeed holds true in Chapter 3 with single target hybridizations. However, in case of competitive binding, the T_d values from experiments A and C are virtually undistinguishable, and the only "clear" interpretation scenario corresponds to experiment D.

Table 18 - Average T_d, °C (temperature at which normalized signal intensity is at 50%) values and standard deviations (n=4) from gel drop arrays, as a function of probe sequence and added target concentrations. Set a corresponds to experiments with an added target that is perfect match to probe 62 (399 is a one-base mismatch to 62), and Set b corresponds to experiments with an added target that is perfect match to probe 1537 (probes 1271 and 1538 are different one-base mismatch probes).

Exp. ID	Probe ID				
	Set a		Set b		
	62	399	1537	1271	1538
A	49.5 ±0.1	36.0 ±16.4	45.3 ±0.2	45.0 ±0.3	41.7 ±3.7
B	54.0 ±1.4	46.2 ±0.3	45.3 ±0.7	47.6 ±0.6	41.5 ±6.5
C	52.1 ±1.7	45.2 ±0.1	46.4 ±0.6	46.0 ±0.5	39.3 ±0.1
D	51.5 ±2.1	44.5 ±0.1	46.7 ±0.2	41.8 ±0.2	37.8 ±0.2

To further explore the utility of melt curves in distinguishing between related sequences, let us explore the actual shapes of the melt curves. Specifically, we will look at the melt curves from the probe 1271, comparing experiments B, C, and D, where the relative concentrations of perfect-match target (1271) to mismatch target (1537) was 20:1, 20:20, and 1:20, respectively

(see Table 17). Figure 46 displays the thermal dissociation curves from these experiments. Note the difference in shapes: the sharpest transition corresponds to the experiments where the ratio of perfect-match to mismatch target was 20:1 (blue line in Figure 46): one would not expect to see any notable contribution by mismatch targets to the overall signal following an overnight hybridization in a stringent buffer. Slightly earlier thermal dissociation is observed for the 20:20 ratio (red line in Figure 46), most likely due to a small contribution by the mismatch target to the overall signal. And lastly, in the case of mismatch target in excess at the beginning of hybridization (1:20 ratio), one sees distinct two-phase behavior of the dissociation process. This finding is particularly important when one re-examines the T_d values of the probe for all three scenarios, given in Table 18. While the T_d values for experiments B and C – either perfect-match and mismatch target present at equal concentrations or perfect-match in excess – are fairly similar, the T_d value for experiment D (mismatch target in excess) is lower. This argues that T_d value by itself cannot necessarily point to the presence of high affinity target, if lower affinity target is present in excess. However, the shape of the dissociation curves show promise for displaying multi-phasic behavior, indicative of a mixed population of dissociating duplexes.

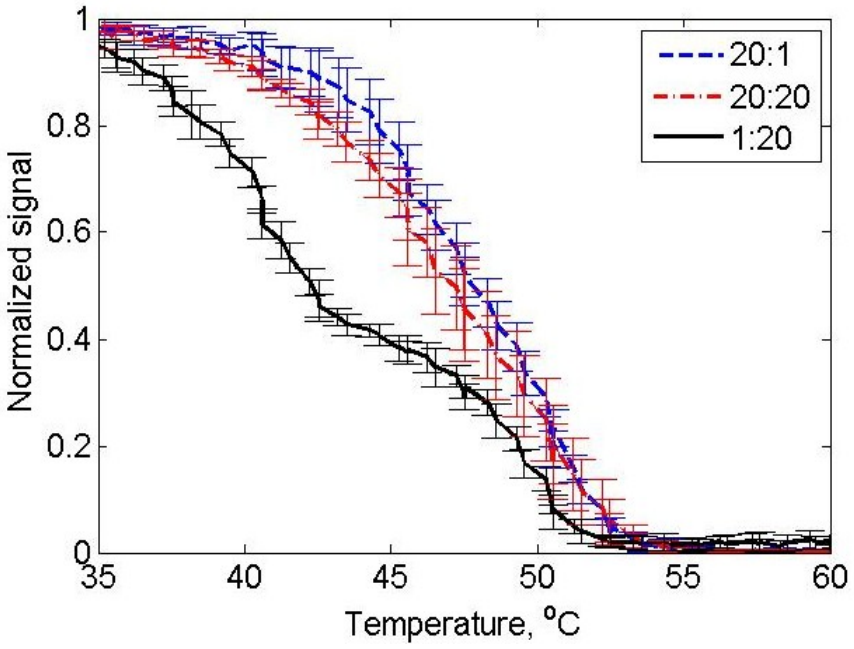

Figure 46 – Normalized thermal dissociation curves from probe 1271. Hybridization preceding the thermal dissociations were conducted with different relative ratios of target 1271 and one-base mismatch target 1537, as specified in the legend. Ratio 20:1 corresponds to experiment B, ratio 20:20 to experiment C, and ratio 1:20 to experiment D in Table 17. Temperature ramp rate: 0.83 ^{0}C/min. Hybridization buffer contained 900 mM NaCl, 20 mM Tris-HCl, and 40% formamide; melt buffer contained 4 mM NaCl, 20 mM Tris-HCl, 5 mM EDTA.

Multi-color imaging of competitive hybridization and dissociation

To confirm that we see biphasic hybridization of lower affinity target during the hybridization of a mixture of two targets, we used a Texas Red labeled target 1537 and Oregon Green labeled target 1538. Lower affinity target t1538 was present at 10 times higher concentration than higher affinity target: 1 μM versus 100 nM. Figure 47 displays the hybridization curves from probe 1537 (perfect-match probe to target 1538). Note the bi-phasic hybridization behavior of the lower affinity target, reflecting the competitive displacement of the mismatch target by perfect-match higher affinity target after the initial phase of hybridization. This is also indirect confirmation that at this total target concentration, we have saturated the probe within the first

~30 minutes, as the decay in the hybridization curve of the mismatch target will only occur if there is an excess of target relative to the total available number of probe sites.

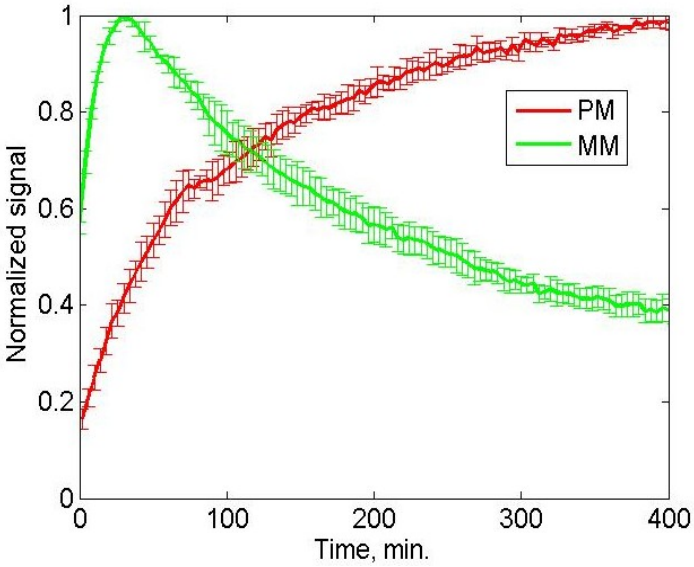

Figure 47 – Normalized hybridization curves, corresponding to a perfect-match and lower affinity mismatch target competing for hybridization onto the same probe. Initial concentrations were 100 nM for perfect-match target, and 1 μM for mismatch target. Perfect-match target was labeled with Texas Red, and lower affinity target with Oregon Green fluorophore. Note the bi-phasic hybridization curve of the lower affinity target. Hybridization buffer contained 900 mM NaCl, 20 mM Tris-HCl, and 40% formamide.

Figure 48 displays the normalized thermal dissociation curves from the same probe 1537, tracking separately the Texas Red labeled target 1537 and Oregon Green labeled target 1538. The dissociation curves and their T_d values are very similar to those obtained from single-component melt (*e.g.* T_d for the melt of target 1537 in the multicomponent dissociation was 38.5 ± 0.5 ^oC, while the T_d value for the melt of 100 nM target 1537 in a single-component experiment in the same buffer was 37.8 ± 0.6 ^oC).

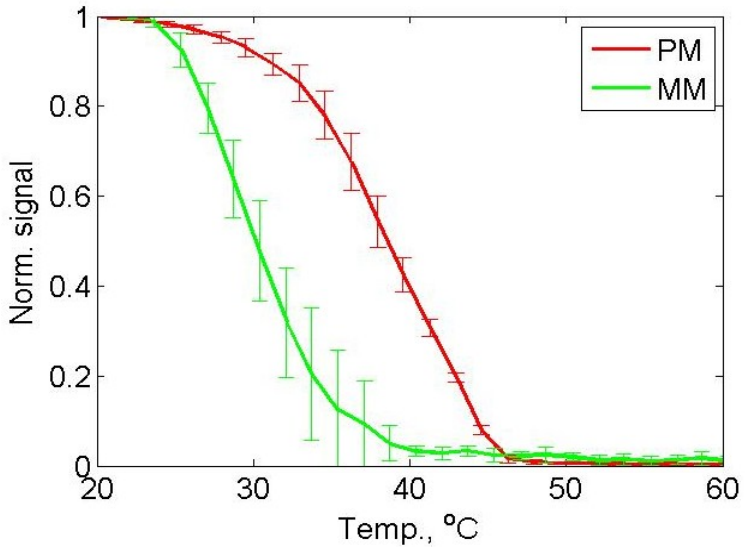

Figure 48 – Normalized thermal dissociation curves for perfect-match and mismatch target dissociation from the same probe. Temperature ramp rate: 0.83 °C/min. Melt buffer contained 4 mM NaCl, 20 mM Tris-HCl, 5 mM EDTA.

Computational

Modeling multi-component hybridization

To validate the model, we simulated experimental conditions described above and reported in Figure 47. Specifically, we used as model inputs experimental target concentrations, and dissociation constant values typically reported for perfect-match and mismatch duplexes for 18-20 bases long sequences [29, 118, 142]. Note the good agreement in Figure 49 between the experimental data and model simulations – no model parameters were optimized for the fit.

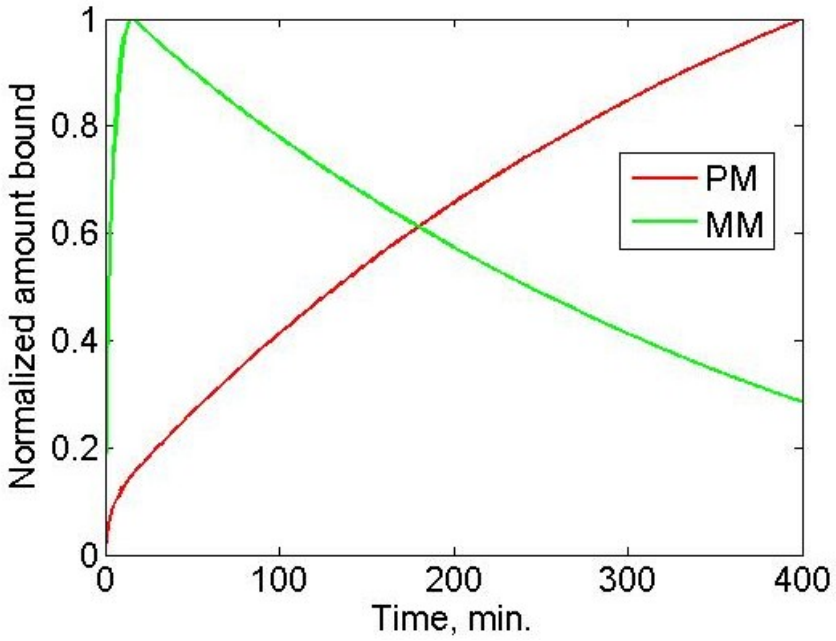

Figure 49 – Simulated normalized competitive hybridization curves from one probe for a high affinity (PM, perfect-match) and lower affinity (MM, mismatch) targets. Intra-gel diffusion coefficient $D_g=10^{-11}$ m^2/s, initial probe concentration $4 \cdot 10^{-2}$ mol/m^3, initial target concentrations 100 nM for perfect-match, and 1 μM for mismatch target.

Modeling thermal dissociation following a multi-component hybridization

To computationally explore the thermal dissociation of two targets hybridized onto the same probe, we altered the fraction of probe bound by high affinity target versus lower affinity target at the beginning of the thermal dissociation. Note that this does not directly correspond to the initial target concentrations at the beginning of hybridization, as the total time of hybridization, and relative concentrations of probe and targets will impact the fraction of total available probe bound by one target versus the other. Figure 50 displays the simulated thermal dissociation curves corresponding to different initial bound fractions.

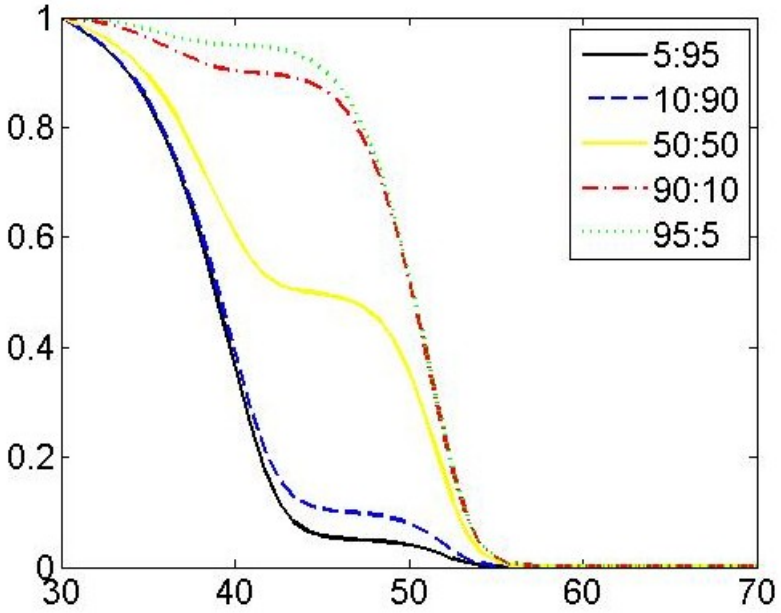

Figure 50 – Simulated normalized thermal dissociation curves, corresponding to different fractions of probe bound to high affinity versus low affinity target. The numbers in the legend for each line refer to the fraction of probe bound by affinity target (first number) and lower affinity target (second number). Assumed temperature ramp rate was 0.83 ^{o}C/min.

4.5 Conclusions

In this Chapter, we investigated multi-component hybridization and subsequent thermal dissociation of a heterogeneous duplex population in three-dimensional gel microarrays. We also demonstrated increased apparent rate of hybridization due to sample recirculation in a microfluidic platform. The increase in the rate was larger for longer nucleic acid fragments, and for lower concentration of targets, as those hybridization scenarios would be the most affected by mass transport limitations.

We demonstrated bi-phasic behavior of the hybridization curves belonging to lower affinity target in three-dimensional gel-based arrays, using multi-color imaging. Also, we showed that our three-dimensional model of the hybridization and dissociation process gives rise to results

similar to those observed in the experiments. Specifically, we replicated the bi-phasic competitive binding observed by multi-color imaging, obtaining good agreement between experimental results and model simulations, and also demonstrated how different relative fractions of probe bound to targets with different affinities result in different shapes of thermal dissociation curves. We also demonstrated bi-phasic behavior of thermal dissociation curves in case of low concentration of high affinity target and high concentration of mismatch target, showing promise for using three-dimensional gel arrays for distinguishing between related sequences in a multi-component system.

Chapter 5: Computational characterization of the potential of microarray real-time hybridization to distinguish between related sequences

5.1 Objectives

The final Chapter of this work utilizes the computational models developed in previous Chapters, taking advantage of the ability of the model to rapidly interrogate a large set of experimental parameters with respect to their influence on the sensitivity of the signal to variations in target affinities. This Chapter also serves as an example of the kinds of questions one could explore with the computational models characterized in previous Chapters. Specifically, we characterize the potential of microarray real-time hybridization data to distinguish between related target sequences.

The objectives of the Chapter are the following:

a) investigate the potential of 'cooling' hybridizations – hybridizations started at an elevated temperature, cooling down to room temperature – in preferentially binding perfect-match targets;

b) analyze the contribution of the data from an added mismatch probe on the array to distinguish between the binding of targets of different affinities.

5.2 Theory and Design Considerations

There are two main types of data that are currently collected from microarray experiments to elucidate the target sequences present in the sample: 1) data from the hybridization phase, either time-dependent signal relating to bound concentration, or end-of-hybridization time point that

may or may not reflect equilibrium values (the latter significantly more widespread than the former); 2) data from experiments where the stringency of hybridization is altered after an initial hybridization phase, resulting in differential dissociation of duplexes of different affinities (*e.g.* non-equilibrium thermal dissociation experiments). Recent theoretical and experimental work towards understanding the effects of competitive binding on hybridization profiles demonstrate that the relative contribution of lower affinity targets to the signal compared to the contribution of the perfect match target can change significantly over the course of the experiment – as the hybridization experiment progresses, higher affinity targets will gradually replace lower affinity ones [29, 113, 122, 123, 127, 142]. While the equilibrium signal in the hybridization is the most specific, time to equilibrium can often be very long (up to days!), depending on the concentrations of the targets in the sample, which is not known *a priori*. Therefore, attaining equilibrium can be unrealistic for practical applications, meaning that aiming to determine the specificity of the bound target from a single time-point collected at the end of the hybridization is rather questionable if not downright impossible. In light of this, analysis of real-time hybridization data or melting analysis has been suggested as a potential promising alternative to standard hybridization timepoint analysis. However, current theoretical and experimental studies of hybridization kinetics have been limited to considering the hybridization of multiple related targets onto one probe, real-time data hybridization data from arrays that include a purposefully designed mismatch probe with each perfect-match probe to a target of interest has not been considered before. This current Chapter explores the limitations of promise of using real-time microarray hybridization data to distinguish between closely related targets and answer the question whether a target of interest is present in the sample based on the signal collected from its perfect-match and/or mismatch probe.

The three-dimensional model of the two-dimensional binding elements was not separately validated with experimental data, as a similar two-dimensional finite element binding model with one binding zone has already been published [29, 118, 142] and we validated our model by reproducing the published results. For simulating the presence of different targets in the hybridizing solution in this Chapter, we do not strictly need to tie the thermodynamic parameters, such as binding free energy change, to specific sequences. What we are interested in in this Chapter, is the theoretical sensitivity of data typically measured in experiments to target affinities under a variety of experimental conditions. This Chapter also serves as an example of the kinds of question one can explore in assay development with the use of computational methods.

5.3 Materials and Methods

Computational model

This Chapter is largely utilizing the mathematical models described in previous Chapters. For modeling cooling hybridization, the diffusion coefficient for 18- and 19-base targets was assumed to be 10^{-10} m^2/s at 20 ^oC [144, 145] and a temperature-correction was used for the diffusion coefficient based on the Stokes-Einstein equation:

$$\frac{D_{T_1}}{D_{T_2}} = \frac{T_1}{T_2}\frac{\mu_{T_2}}{\mu_{T_1}}$$

where T_1 and T_2 are two different temperatures, D_{T1} and D_{T2} are diffusion coefficients corresponding to these two temperatures, and μ_{T1} and μ_{T2} dynamic viscosities of the liquid. Taking D_{T1} equal to 10^{-10} m^2/s at 20 ^oC and using the dynamic viscosities of water (1.002 mPa s), we obtain the relationship displayed on Figure 51.

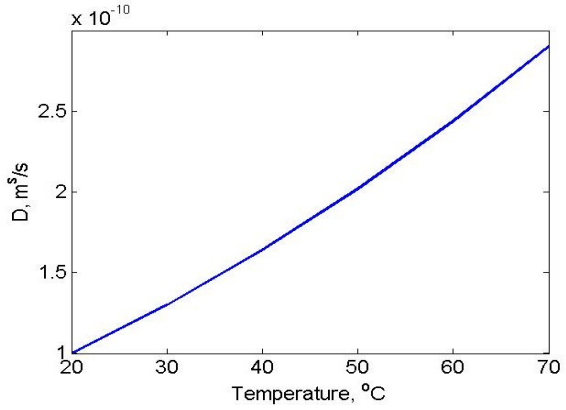

Figure 51 - Relationship between oligonucleotide diffusion coefficient and temperature.

The 2D binding model was composed similarly to the development of the 3D binding model in Chapters 3 and 4 with respect to the diffusion of the free analytes in the bulk solution. The main difference between the 3D binding and 2D binding model is in the treatment of the planar binding regions. The binding reaction was described similarly to Chapters 3 and 4, surface diffusion in the binding region was not considered, and boundary PDE weak mode was utilized. The two-dimensional binding region was coupled to the bulk compartment via a flux boundary condition. Processing and analysis of simulation data was conducted with MATLAB.

Dimensions of the hybridization chamber in the 3D model were 1cm × 1 cm × 100 μm, binding element diameter was 100 μm. Diffusion coefficient in the solution was 10^{-10} m^2/s, intra-gel diffusion coefficient was 10^{-11} m^2/s. Association constant k_a was 10^6 $1/(M \cdot s)$, dissociation constant k_d value will be specified for each simulation. At the beginning of the hybridization, duplex concentration and free target concentration inside the gel were assumed to be 0. Initial free probe and target concentration will be specified for each simulation.

5.4 Results and Discussion

"Cooling" hybridization – can it replace the melt in resolving related sequences?

Cooling hybridization, unlike an isothermal hybridization, would start at an elevated temperature (*e.g.* 70 ^{o}C), and continue at a decreasing temperature down to RT (or 20 ^{o}C). Unlike during isothermal hybridization, where lower affinity targets get slowly replaced by higher affinity targets and the time to equilibrium can be unreasonably long for practical applications (up to days), during cooling hybridization, higher affinity targets would preferentially bind at higher temperatures.

The first question we explored was the effect of the ramp rate on the hybridization curves. For a system where a target and a competitor are competing for binding onto one probe, we see a slight effect of the ramp rate on the maximum slope in the total hybridization curve (Figure 52) and a more profound effect on the ratio of target bound versus competitor bound curves (Figure 53).

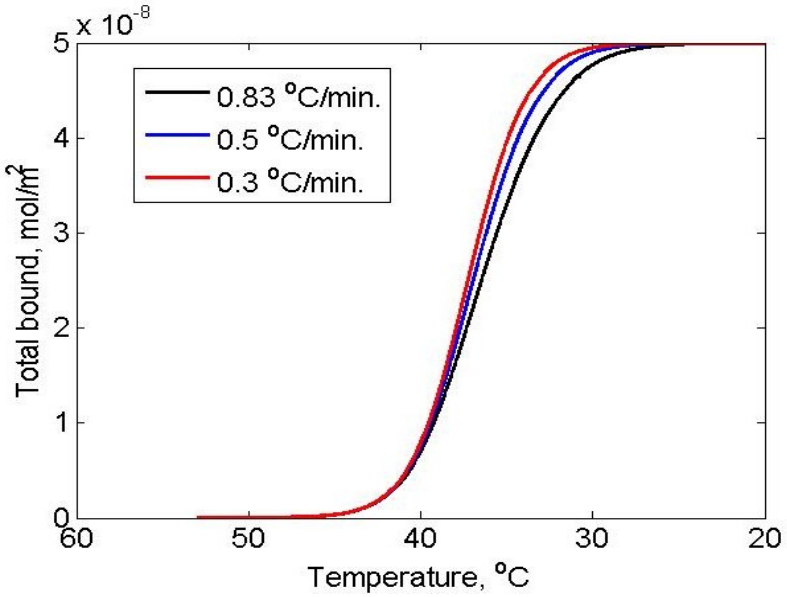

Figure 52 – Total amount bound (target of interest plus competitor) as a function of temperature for three different ramp rates. Initial probe concentration: $5 \cdot 10^{-8}$ mol/m^2, initial target of interest and competitor concentration $5 \cdot 10^{-6}$ mol/m^3.

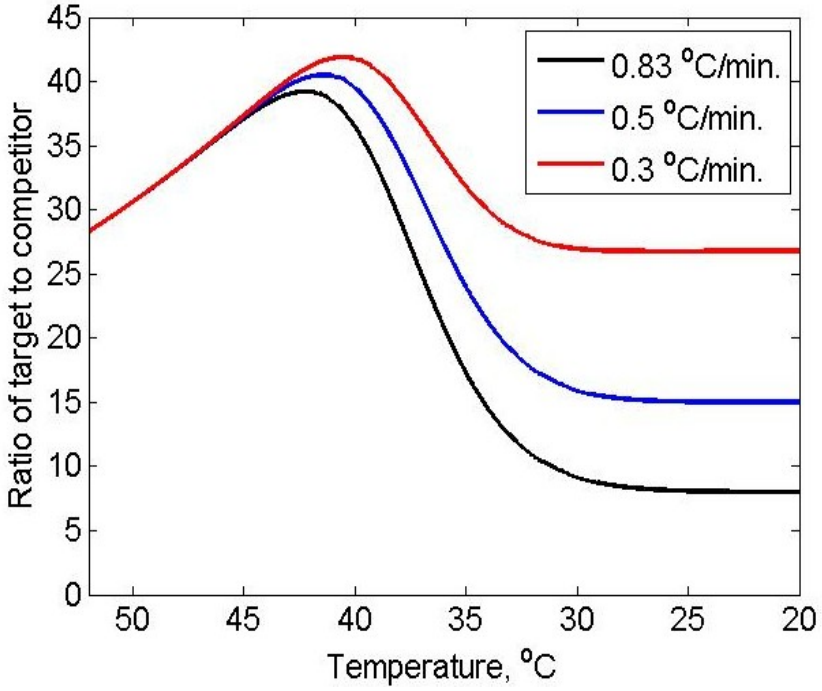

Figure 53 – Ratio of target versus competitor bound on a single probe as a function of temperature, for three different ramp rates. Initial probe concentration: $5 \cdot 10^{-8}$ mol/m^2, initial target of interest and competitor concentration $5 \cdot 10^{-6}$ mol/m^3.

We can compare this with isothermal hybridization at either 20 °C or 30 °C, with the target of interest and competitor present at equal concentrations ($5 \cdot 10^{-6}$ mol/m^3) with $5 \cdot 10^{-8}$ mol/m^3 probe (Figure 54). What is interesting about Figure 54, is that although the relative contributions from target of interest and competitor are very different between hybridizations at 20 °C and 30 °C, the total amount bound does not vary significantly (compare red and black solid lines) – and the latter would be observed in the applications where all sample DNA is labeled.

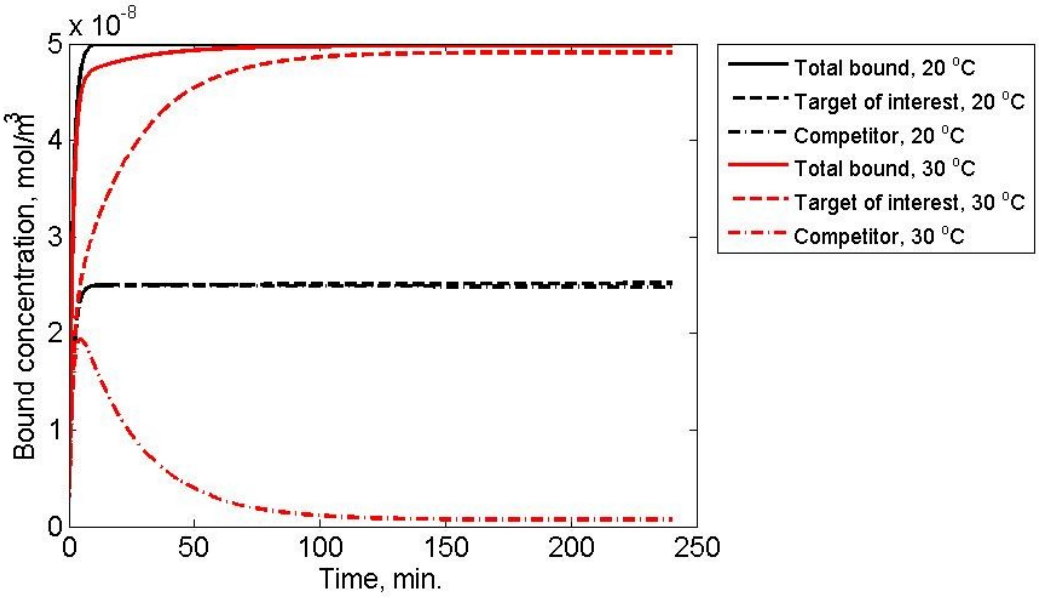

Figure 54 – Isothermal hybridization at either 20 °C or 30 °C as a function of time, for total bound and target and competitor bound.

Looking at cooling hybridization, we then simulated a case where only one of the two is present: either target of interest, or competitor. In order to distinguish between these two cases, the hybridization curves should fall between different temperature ranges. In the parallel case of melting, we would expect the melting point for a target-of-interest-only case to be higher than for competitor-only. The cooling hybridization curves however do depend on the concentration of the hybridizing target. Figure 55 and Figure 56 display the hybridization curves for target of interest and competitor in a one-component system. Ten-fold difference in concentration shifts the curves significantly (compare solid and dashed lines), and the steepest part of the solid curves in Figure 56 (higher concentration of competitor) is similar to the location of the steepest part of dashed lines in Figure 55. This means that cooling hybridization is influenced by the concentration of the hybridizing target – lower affinity target, if present at higher concentration, can produce hybridization curves similar to a higher affinity target at lower concentration. This means that the cooling hybridization does not appear to hold any more promise for distinguishing between related species than isothermal hybridization.

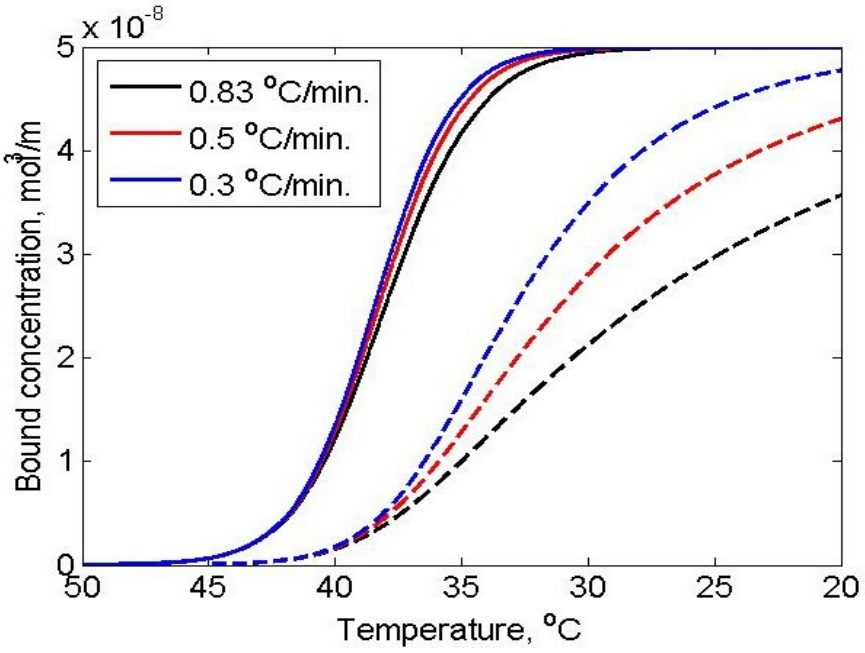

Figure 55 – Bound concentration of target of interest in a 1-component system for three different temperature ramp rates. Dashed lines refer to 10^{-6} mol/m^2, while the solid lines to 10^{-5} mol/m^2 initial target concentration.

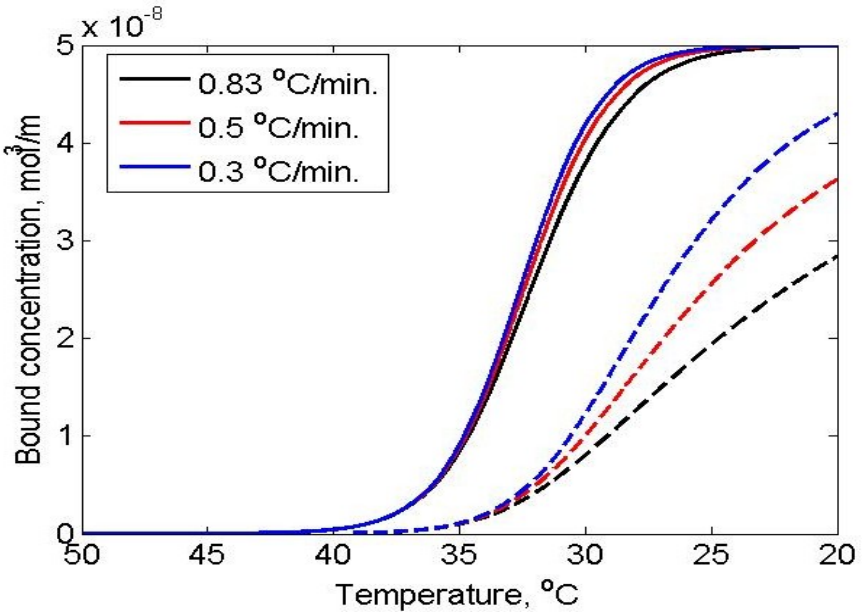

Figure 56 - Bound concentration of competitor in a 1-component system for three different temperature ramp rates. Dashed lines refer to 10^{-6} mol/m^2, while the solid lines to 10^{-5} mol/m^2 initial target concentration.

Does adding a mismatch probe make a difference for resolving between closely related sequences?

We implemented a second binding element in the three-dimensional microarray hybridization model to investigate the potential utility of collecting data from a mismatch probe in parallel with a probe that is a perfect complement for the hybridizing target. Specifically, we focused on the initial phase of the hybridization, without necessarily waiting for equilibrium. The different affinity of the target towards the two different probes was modeled through altering the dissociation constant: k_d equal to 10^{-4} 1/s represents values typically reported for a single-mismatch probe, and k_d equal to 10^{-2} 1/s corresponds to two-base mismatch probe. Figure 57 displays the normalized amount bound on both the perfect-match and mismatch probes for different probe and target concentrations in a single-component system. The figure illustrates the wide range of possible duplex concentration ratios, depending on the relative probe and target concentration and also on the kinetic constants of the binding reaction. Seeing the ratio of amount bound to perfect-match probe over the mismatch probe considerably over 1 would lead to unambiguous evidence about the difference in affinity of the target towards the two probes. As Figure 57 demonstrates, the ratio of the amount bound to the two probes can be equal to 1 either in cases of high target concentration, leading to rapid equilibrium and full saturation of both probes (for example blue lines in the upper right panel), or in case of low target concentration relative to probe concentration, leading to mass transport limitations for the hybridization reaction (red and black lines in the upper right panel). The unambiguous case of the ratio of the duplex concentration over 1 in the Figure below corresponds to either large dissociation constant value and moderate to low target concentrations, or low probe concentration and moderate to low initial target concentrations.

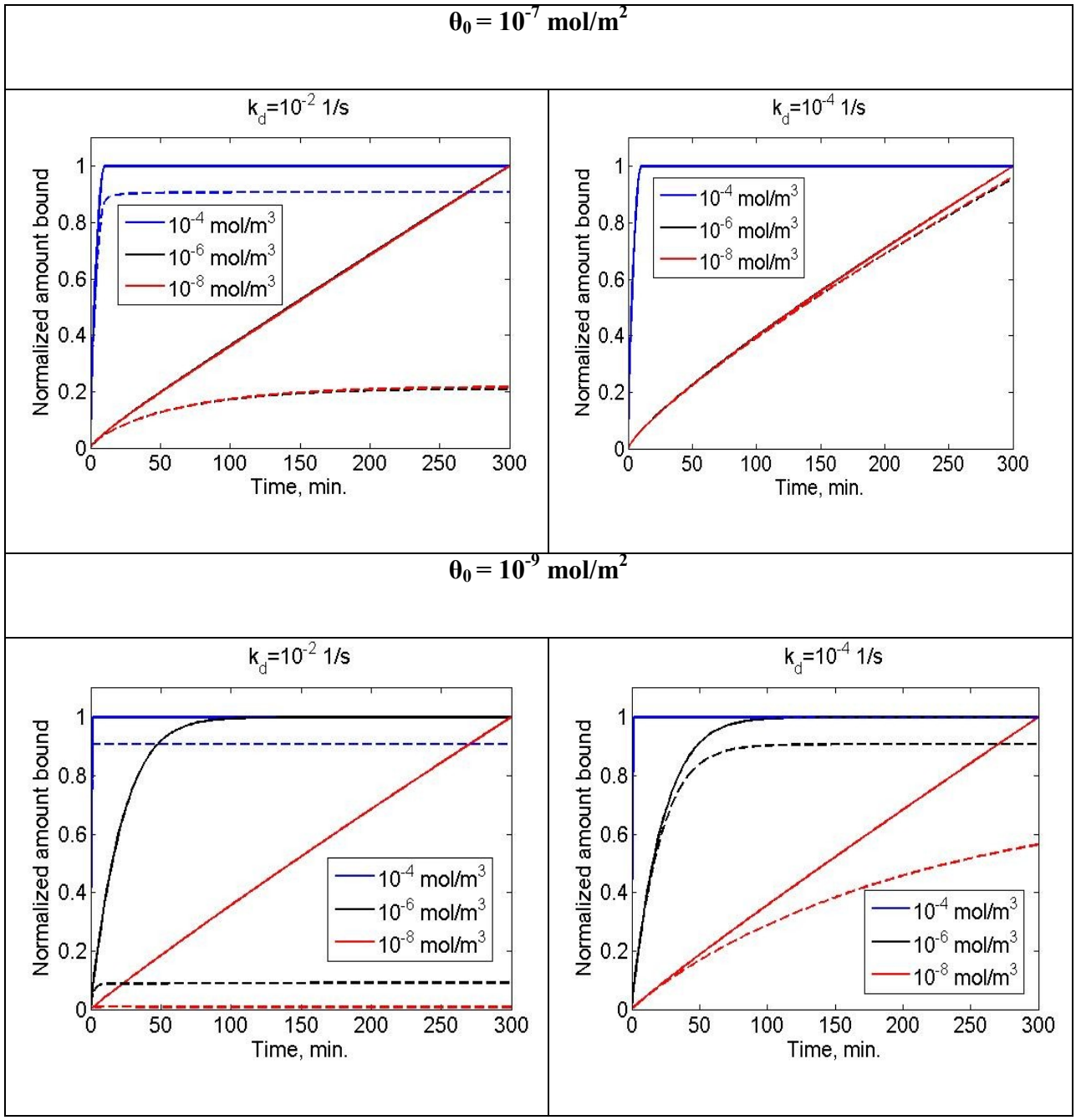

Figure 57 – Normalized hybridization curves from a perfect-match and mismatch probe upon a single-component hybridization. Normalization was done with respect to the maximum value of duplex concentration on the perfect-match probe for each target concentration. Different affinities to the two different probes were modeled by altering the dissociation constant k_d (see above each subplot). Upper figures refer to initial probe concentration of 10^{-7} mol/m^3, while the bottom row corresponds to initial probe concentration of 10^{-9} mol/m^3. The solid lines refer to perfect-match probe, and dashed lines to mismatch probe. Different colors represent different initial target concentration.

Chapter 6: Overall Conclusions

The goal for undertaking this work was two-fold. One was to advance the development of physico-chemical computational models to aid in the fundamental understanding of the microarray system, and in the development of applications either based on microarray technology, or on binding and thermal dissociation of short oligonucleotides in a three-dimensional matrix or on a planar platform. The second goal was to further explore the limits of the data currently obtained from microarray experiments in distinguishing between closely related target sequences, using both experimental and computational methods.

Several models of planar microarray hybridization processes have been proposed, exploring either a single-component system with only one target hybridizing onto a probe [118, 121], or a multicomponent system with two or more targets competing to hybridize onto the same probe [29, 78, 123, 124, 127, 142]. The current work is the first to propose a three-dimensional model of the hybridization and thermal dissociation processes on a three-dimensional microarray platform, and obtain a good fit between simulations and experiments with single-component thermal dissociation and multi-component competitive binding and thermal dissociation. The three-dimensional gel-based microarray technology is a rather new type of a microarray platform that has been less extensively characterized than planar microarray technology. Furthermore, beyond presenting a finite element model for a new type of microarray platform, we explored the link between nearest-neighbor based thermodynamic parameters characterizing duplex stability, and microarray thermodynamic parameters. Current planar array models that are concerned with a comparison with experimental data either use optimized kinetic constants to describe the

binding reaction, or introduce a temperature offset into the model to account for experimentally observed reduced stability of the duplexes compared to binding reactions in the solution [29]. The gel-based microarray platform has been proposed in the literature to potentially present solution-like binding kinetics. While our kinetic model for solution-based thermal dissociation, utilizing nearest-neighbor based thermodynamic parameters, agreed very well with experimental results, the parameters significantly overestimated the stability of the duplexes in the microarray platform. Optimizing thermodynamic parameters to obtain a good fit between the experimental microarray thermal dissociation curves and model simulations rendered a linear relationship between the nearest-neighbor based parameters and microarray parameters for perfect-match duplexes. This demonstrates the possibility of predicting the thermodynamic parameters for a number of different perfect-match duplexes on the array based on a small set of duplexes used to establish the linear relationship. However, we did not find the same linear relationship to hold between nearest-neighbor parameters and microarray parameters for mismatch probes, and the small selection of mismatch probes presented on the array did not allow us to study this effect further based on the position and nature of mismatch.

The remainder of this work utilized the composed computational models, in concert with experimental data, to explore the limits of the microarray hybridization and thermal dissociation data in distinguishing between related sequences. In addition to thermal dissociation data, collecting real-time hybridization data in microarray experiments has been suggested to aid in distinguishing between related sequences, although latest research papers have expressed skepticism towards that idea [1]. This work has been limited to collecting data on one probe of interest. In this work, we explored the potential advantage of adding a mismatch probe to the

hybridizing target in determining the affinity of the target towards the perfect-match probe, and found that even in the simplest system of one target hybridizing onto two probes, the relative ratio of duplex concentrations on the two probes will only lead to unambiguous determination of the relative affinity of the target towards the two probes in a limited range of probe and target concentrations. Lastly, we explored thermal dissociation in a multicomponent system, utilizing experimental data from gel-based microarrays and computational results. While the T_d values of the thermal dissociation curves (temperature at which the signal from a probe has decreased by 50 %) were misleading in cases with high mismatch target concentration and low perfect-match target concentration, we observed bi-phasic behavior of the dissociation curves, indicative of a heterogeneous population and showing potential to distinguish between targets of different relative affinities.

Chapter 7: Current Industry Trends and Future Research Directions.

Although commercialized DNA microarray technology dates back to the mid-1990s (the first on the market, Affymetrix introduced its first array product in 1994), published studies investigating the reproducibility and reliability of microarray hybridization experiments are less than ten years old [146]. As FDA declared pharmacogenomics and toxicogenomics key areas of interest, it initiated a wide-scale project to address concerns about reliability of data from microarray experiments, involving scientists from over 50 organizations [146]. Gene expression data from seven different microarray platforms, generated at multiple test sites, were compared for intra- and inter-array variability. Six of the tested microarray platforms were commercially available (Applied Biosystems, Affymetrix, Agilent Technologies, GE Healthcare, Illumina, Eppendorf), and one was a spotted microarray, prepared in-house by a team at the National Cancer Institute [146]. The study demonstrated good intra-array reproducibility and also a "*high level of interplatform concordance in terms of genes identified as differentially expressed"* [146]. The conclusions from that study have been cited as one of the possible reasons for consolidation of the microarray market in the mid-2000s – contrary to marketing campaigns by array manufacturers, the performance of different platforms was shown to be quite similar, and the main differences were suspected to be rooted in sample preparation and processing [147]. As of the beginning of year 2012, the four largest manufacturers of DNA microarrays were Affymetrix, Agilent, Illumina and Roche [147], but Roche recently announced its decision to phase out its DNA microarray business by the end of 2012 [148]. As of Summer 2012, Affymetrix, Agilent and Illumina have stated their firm commitment to producing DNA microarrays [149]. The other competitors, GE Healthcare, Nanogen and Applied Biosystems

(now Life Technologies) had discontinued their microarray production a few years ago [147]. This has been attributed generically to the maturing of the market, and also to a variety of reasons from selling too closed systems that could not be integrated with other array instrumentation (Applied Biosystems) to tight control of intellectual property by early entry companies [147].

In the past few years, next-generation sequencing (NGS), also referred to as 'massively parallel sequencing', has emerged as a competitor to DNA microarray technology, as the price of sequencing is getting closer to that of a microarray experiment [150, 151]. The advantages of NGS over DNA microarray technology are several: NGS has a higher dynamic range than the DNA microarray technology; NGS is quantitative while microarray technology is qualitative; NGS is more sensitive than microarray technology, offering a lower background than microarray technology; and finally, NGS, unlike microarray technology, is not hampered by cross-hybridization, which means accurate single-nucleotide resolution [151-155]. Microarray technology also requires generally *a priori* knowledge of potential hybridizing sequences to guide probe design. Despite the advantages, reviews in the literature that compare the performance of NGS and DNA microarray technology conclude that DNA microarrays will remain relevant as a research tool in the foreseeable future [151, 153-155]. As of right now, DNA microarray technology is considerably more familiar and in general more accessible to the research community as a whole. Also, NGS requires access to a sequencing facility and massive infrastructure for information technology for data analysis, and the bioinformatics support for NGS is considerably less mature than for microarray technology [156]. Even as the cost of sequencing is rapidly going down, storage and analysis of massive volumes of data are predicted

to form a bottleneck for the user [152-155]. For example, Illumina's Solexa Genome Analyzer II will produce over 115,000 files during one run, amounting to 1 terabyte of data [156]. While a growing number of researchers is likely to begin to use NGS instead of DNA microarray technology for genomic studies, the degree of adoption of NGS will probably vary across applications: for example, NGS is predicted to be a strong competitor in the area of global expression profiling, while DNA microarrays will retain their stronghold for custom genotyping and copy-number variation analysis [151]. Employing DNA microarrays will also remain considerably cheaper and less time-consuming when analyzing the DNA or RNA in a large number of clinical samples [154]. Another key area where DNA microarrays have been predicted to occupy a middle ground between PCR and NGS, is the detection and analysis of microorganisms [155]. Several different DNA microarray platforms for that purpose have been proposed in the literature, differing in the fabrication, the organisms detected, and the sensitivity/specificity [155]. The ability to design probes against conserved regions allows the use the microarrays both for detecting existing and discovering novel microorganisms [155]. Furthermore, the so-called universal detection microarrays that have been proposed in the literature do away with the limitation of needing to know sequenced microbial genomes beforehand for probe design; instead, the probe sequences are generated randomly, but hybridization of the genomic DNA from different organisms on the array still gives rise to unique hybridization patterns [155]. Overall, DNA microarray technology continues to show great potential for fast and affordable microbial detection in point-of-care type devices, provided further advances are made in automated sample preparation, DNA/RNA extraction, faster time to read-out, and efficient analysis algorithms that do not require more computing power than is typically found in a smart-phone [155]. However, achieving fast, reliable and reproducible

application of DNA microarray technology, either in medical setting, food safety, or environmental monitoring, requires a deeper understanding of the complex biophysical phenomena that occur on the array [155]. The aim of this dissertation was to contribute towards that goal.

Future directions for research, arising from the lessons learned in this work, can be broadly divided into three categories: research towards a deeper understanding of DNA binding in the solution; investigations of the stability of duplexes in the microarray; and further study of multiplex non-equilibrium melt in gel-based microarrays. While the nearest-neighbor model performed well, predicting the melting temperatures for perfect-match and even one-base mismatch sequences tested in this work, there were notable discrepancies between the melting temperatures predicted by the nearest-neighbor based model and experimental solution-based thermal dissociation data for two-base mismatch duplexes. Specifically, the model estimated the stability of the duplexes to be higher than what was seen in the solution-based experiments. In order to move closer to a predictive quantitative mathematical model of the microarray processes to guide the development of practical applications (for example universal detection arrays as outlined above), a solid understanding of the affinities of both perfect-match and mismatch duplexes in the solution is needed as a foundation. It can be hypothesized that longer than nearest-neighbor influences by the two (or more) mismatches weaken the duplex more than is predicted simply by considering the mismatches independent of each other. Regarding the binding process in microarrays, further work is needed to fully understand the link between solution-based duplex stability and stability in the microarray. While the selected perfect-match duplexes considered in this study rendered a linear relationship between the nearest-neighbor

based thermodynamic parameters and microarray thermodynamic parameters, a larger selection of duplexes would be needed to fully validate the findings. Also, further work would need to be done to explore the relationship between nearest-neighbor based and microarray thermodynamic parameters for one-base mismatch duplexes. Investigation of the relationship for more than one-base mismatch duplexes is awaiting improvements in the description of the stability of duplexes in the solution as pointed out above. While the current work only focused on duplexes where the target and probe molecule were of equal length, future work on investigating DNA duplex stability in microarrays should also take into account anticipated variability in the target lengths in practical applications of microarray technology. The length of the DNA target will have an impact on the rate of mass transport of the target, and will influence the predicted stability of the duplex (see work on dangling end duplexes in [72, 157-160]). Lastly, this dissertation demonstrates the promise of non-equilibrium thermal dissociation to distinguish between closely related target sequences, as the observed biphasic shape of the melt curve allowed to detect low levels of perfect-match target in the presence of high levels of mismatch target. Further experiments would need to be done, using more complex samples, to explore the feasibility of this detection technique.

Bibliography

1. Chagovetz, A. and S. Blair, *Real-time DNA microarrays: reality check.* Biochem Soc Trans, 2009. **37**(Pt 2): p. 471-5.
2. Marcelino, L.A., et al., *Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data.* Proc Natl Acad Sci U S A, 2006. **103**(37): p. 13629-34.
3. Mehta, T., M. Tanik, and D.B. Allison, *Towards sound epistemological foundations of statistical methods for high-dimensional biology.* Nat Genet, 2004. **36**(9): p. 943-7.
4. Urisman, A., et al., *E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns.* Genome Biol, 2005. **6**(9): p. R78.
5. Binder, H. and S. Preibisch, *Specific and nonspecific hybridization of oligonucleotide probes on microarrays.* Biophys J, 2005. **89**(1): p. 337-52.
6. Gilbride, K.A., D.Y. Lee, and L.A. Beaudette, *Molecular techniques in wastewater: Understanding microbial communities, detecting pathogens, and real-time process control.* J Microbiol Methods, 2006. **66**(1): p. 1-20.
7. Starke, E.M., et al., *Technology development to explore the relationship between oral health and the oral microbial community.* BMC Oral Health, 2006. **6 Suppl 1**: p. S10.
8. Starke, E.M., et al., *Saliva-based diagnostics using 16S rRNA microarrays and microfluidics.* Ann N Y Acad Sci, 2007. **1098**: p. 345-61.
9. Watterson, J.H., P.A.E. Piunno, and U.J. Krull, *Towards the optimization of an optical DNA sensor: control of selectivity coefficients and relative surface affinities.* Analytica Chimica Acta, 2002. **457**: p. 29-38.
10. Bekal, S., et al., *Rapid identification of Escherichia coli pathotypes by virulence gene detection with DNA microarrays.* J Clin Microbiol, 2003. **41**(5): p. 2113-25.
11. Bruant, G., et al., *Development and validation of an oligonucleotide microarray for detection of multiple virulence and antimicrobial resistance genes in Escherichia coli.* Appl Environ Microbiol, 2006. **72**(5): p. 3780-4.
12. Troesch, A., et al., *Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays.* J Clin Microbiol, 1999. **37**(1): p. 49-55.
13. Pannucci, J., et al., *Virulence signatures: microarray-based approaches to discovery and analysis.* Biosens Bioelectron, 2004. **20**(4): p. 706-18.
14. Woese, C.R., *Bacterial evolution.* Microbiol Rev, 1987. **51**(2): p. 221-71.
15. Stahl, D.A., et al., *Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology.* Appl Environ Microbiol, 1988. **54**(5): p. 1079-84.
16. Starke, E.M., et al., *Saliva-based diagnostics using 16S rRNA microarrays and microfluidics.* Annals of N.Y. Academy of Sciences, 2007.
17. Urakawa, H., et al., *Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses.* Appl Environ Microbiol, 2002. **68**(1): p. 235-44.
18. Rubina, A.Y., et al., *Hydrogel drop microchips with immobilized DNA: properties and methods for large-scale production.* Anal Biochem, 2004. **325**(1): p. 92-106.
19. Sorokin, N.V., et al., *Kinetics of hybridization on surface oligonucleotide microchips: theory, experiment, and comparison with hybridization on gel-based microchips.* J Biomol Struct Dyn, 2006. **24**(1): p. 57-66.

20. Shchepinov, M.S., S.C. Case-Green, and E.M. Southern, *Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays.* Nucleic Acids Res, 1997. **25**(6): p. 1155-61.
21. Suzuki, M.T. and S.J. Giovannoni, *Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR.* Appl Environ Microbiol, 1996. **62**(2): p. 625-30.
22. Speksnijder, A.G., et al., *Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences.* Appl Environ Microbiol, 2001. **67**(1): p. 469-72.
23. Acinas, S.G., et al., *PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample.* Appl Environ Microbiol, 2005. **71**(12): p. 8966-9.
24. Guschin, D.Y., et al., *Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology.* Appl Environ Microbiol, 1997. **63**(6): p. 2397-402.
25. Yershov, G., et al., *DNA analysis and diagnostics on oligonucleotide microchips.* Proc Natl Acad Sci U S A, 1996. **93**(10): p. 4913-8.
26. Mitchell, P., *Microfluidics--downsizing large-scale biology.* Nat Biotechnol, 2001. **19**(8): p. 717-21.
27. Beebe, D.J., G.A. Mensing, and G.M. Walker, *Physics and applications of microfluidics in biology.* Annu Rev Biomed Eng, 2002. **4**: p. 261-86.
28. Situma, C., M. Hashimoto, and S.A. Soper, *Merging microfluidics with microarray-based bioassays.* Biomol Eng, 2006. **23**(5): p. 213-31.
29. Bishop, J., S. Blair, and A.M. Chagovetz, *A competitive kinetic model of nucleic acid surface hybridization in the presence of point mutants.* Biophys J, 2006. **90**(3): p. 831-40.
30. Dai, H., et al., *Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays.* Nucleic Acids Res, 2002. **30**(16): p. e86.
31. Liu, R.H., et al., *Hybridization enhancement using cavitation microstreaming.* Anal Chem, 2003. **75**(8): p. 1911-7.
32. Yuen, P.K., et al., *Microfluidic devices for fluidic circulation and mixing improve hybridization signal intensity on DNA arrays.* Lab Chip, 2003. **3**(1): p. 46-50.
33. Vanderhoeven, J., et al., *DNA microarray enhancement using a continuously and discontinuously rotating microchamber.* Anal Chem, 2005. **77**(14): p. 4474-80.
34. Toegl, A., et al., *Enhancing results of microarray hybridizations through microagitation.* J Biomol Tech, 2003. **14**(3): p. 197-204.
35. Noerholm, M., et al., *Polymer microfluidic chip for online monitoring of microarray hybridizations.* Lab Chip, 2004. **4**(1): p. 28-37.
36. Vanderhoeven, J., et al., *Comparison of a pump-around, a diffusion-driven, and a shear-driven system for the hybridization of mouse lung and testis total RNA on microarrays.* Electrophoresis, 2005. **26**(19): p. 3773-9.
37. Gurtner, C., et al., *Microelectronic array devices and techniques for electric field enhanced DNA hybridization in low-conductance buffers.* Electrophoresis, 2002. **23**(10): p. 1543-50.
38. Lee, H.H., et al., *Recirculating flow accelerates DNA microarray hybridization in a microfluidic device.* Lab Chip, 2006. **6**(9): p. 1163-70.

39. Herning, T., E. Tamiya, and I. Karube, *Specific liquid DNA hybridization kinetics measured by fluorescence polarization.* Analytica Chimica Acta, 1991. **244**: p. 207-213.
40. Livshits, M.A. and A.D. Mirzabekov, *Theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligonucleotides.* Biophys J, 1996. **71**(5): p. 2795-801.
41. Hekstra, D., et al., *Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays.* Nucleic Acids Res, 2003. **31**(7): p. 1962-8.
42. Held, G.A., G. Grinstein, and Y. Tu, *Relationship between gene expression and observed intensities in DNA microarrays--a modeling study.* Nucleic Acids Res, 2006. **34**(9): p. e70.
43. Affymetrix, *Affymetrix Microarray Suite 5.0*, in *User Guide*. 2001, Affymetrix, Inc., Santa Clara, CA.
44. Naef, F., et al., *DNA hybridization to mismatched templates: a chip study.* Phys Rev E Stat Nonlin Soft Matter Phys, 2002. **65**(4 Pt 1): p. 040902.
45. Halperin, A., A. Buhot, and E.B. Zhulina, *Sensitivity, specificity, and the hybridization isotherms of DNA chips.* Biophys J, 2004. **86**(2): p. 718-30.
46. Bhanot, G., et al., *The importance of thermodynamic equilibrium for high throughput gene expression arrays.* Biophys J, 2003. **84**(1): p. 124-35.
47. Naef, F. and M.O. Magnasco, *Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays.* Phys Rev E Stat Nonlin Soft Matter Phys, 2003. **68**(1 Pt 1): p. 011906.
48. Zhou, Y. and R. Abagyan, *Match-only integral distribution (MOID) algorithm for high-density oligonucleotide array analysis.* BMC Bioinformatics, 2002. **3**: p. 3.
49. Irizarry, R.A., et al., *Summaries of Affymetrix GeneChip probe level data.* Nucleic Acids Res, 2003. **31**(4): p. e15.
50. Liu, W.T., A.D. Mirzabekov, and D.A. Stahl, *Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach.* Environ Microbiol, 2001. **3**(10): p. 619-29.
51. El Fantroussi, S., et al., *Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays.* Appl Environ Microbiol, 2003. **69**(4): p. 2377-82.
52. Urakawa, H., et al., *Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays.* Appl Environ Microbiol, 2003. **69**(5): p. 2848-56.
53. Eyers, L., et al., *Discrimination of shifts in a soil microbial community associated with TNT-contamination using a functional ANOVA of 16S rRNA hybridized to oligonucleotide microarrays.* Environ Sci Technol, 2006. **40**(19): p. 5867-73.
54. Russom, A., et al., *Genotyping by dynamic heating of monolayered beads on a microheated surface.* Electrophoresis, 2004. **25**(21-22): p. 3712-9.
55. Russom, A., et al., *Rapid melting curve analysis on monolayered beads for high-throughput genotyping of single-nucleotide polymorphisms.* Anal Chem, 2006. **78**(7): p. 2220-5.
56. Howell, W.M., et al., *Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms.* Nat Biotechnol, 1999. **17**(1): p. 87-8.
57. Kajiyama, T., et al., *Genotyping on a thermal gradient DNA chip.* Genome Res, 2003. **13**(3): p. 467-75.

58. Wick, L.M., et al., *On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes.* Nucleic Acids Res, 2006. **34**(3): p. e26.
59. Stromqvist Meuzelaar, L., et al., *DNA diagnostics by surface-bound melt-curve reactions.* J Mol Diagn, 2007. **9**(1): p. 30-41.
60. Smoot, L.M., et al., *DNA microarrays as salivary diagnostic tools for characterizing the oral cavity's microbial community.* Adv Dent Res, 2005. **18**(1): p. 6-11.
61. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science, 1988. **239**(4839): p. 487-91.
62. Fodor, S.P., et al., *Multiplexed biochemical assays with biological chips.* Nature, 1993. **364**(6437): p. 555-6.
63. Gray, D.M. and I. Tinoco, Jr., *A new approach to the study of sequence-dependent properties of polynucleotides.* Biopolymers, 1970. **9**: p. 223-244.
64. Devoe, H. and I. Tinoco, Jr., *The stability of helical polynucleotides: base contributions.* J Mol Biol, 1962. **4**: p. 500-17.
65. Borer, P.N., et al., *Stability of ribonucleic acid double-stranded helices.* J Mol Biol, 1974. **86**(4): p. 843-53.
66. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1460-5.
67. Allawi, H.T. and J. SantaLucia, Jr., *Thermodynamics and NMR of internal G.T mismatches in DNA.* Biochemistry, 1997. **36**(34): p. 10581-94.
68. Allawi, H.T. and J. SantaLucia, Jr., *Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects.* Biochemistry, 1998. **37**(26): p. 9435-44.
69. Allawi, H.T. and J. SantaLucia, Jr., *Thermodynamics of internal C.T mismatches in DNA.* Nucleic Acids Res, 1998. **26**(11): p. 2694-701.
70. Allawi, H.T. and J. SantaLucia, Jr., *Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA.* Biochemistry, 1998. **37**(8): p. 2170-9.
71. Peyret, N., et al., *Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches.* Biochemistry, 1999. **38**(12): p. 3468-77.
72. Bommarito, S., N. Peyret, and J. SantaLucia, Jr., *Thermodynamic parameters for DNA sequences with dangling ends.* Nucleic Acids Res, 2000. **28**(9): p. 1929-34.
73. Bloomfield, V.A., D.M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structures, Properties, and Functions.* 2000, Sausalito, California: University Science Books.
74. Fotin, A.V., et al., *Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips.* Nucleic Acids Res, 1998. **26**(6): p. 1515-21.
75. Doktycz, M.J., et al., *Optical melting of 128 octamer DNA duplexes. Effects of base pair location and nearest neighbors on thermal stability.* J Biol Chem, 1995. **270**(15): p. 8439-45.
76. Pozhitkov, A., et al., *Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.* Nucleic Acids Res, 2006. **34**(9): p. e66.
77. Weckx, S., et al., *Thermodynamic behavior of short oligonucleotides in microarray hybridizations can be described using Gibbs free energy in a nearest-neighbor model.* J Phys Chem B, 2007. **111**(48): p. 13583-90.

78. Fish, D.J., et al., *DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison.* Nucleic Acids Res, 2007. **35**(21): p. 7197-208.
79. Hooyberghs, J., P. Van Hummelen, and E. Carlon, *The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters.* Nucleic Acids Res, 2009. **37**(7): p. e53.
80. Garrett, H.G. and C.M. Grisham, *Biochemistry*. Second Edition ed. 1999: Saunders College Publishing.
81. Breslauer, K.J., J.M. Sturtevant, and I. Tinoco, Jr., *Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: rA7U7.* J Mol Biol, 1975. **99**(4): p. 549-65.
82. Erie, D., et al., *A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation.* Biochemistry, 1987. **26**(22): p. 7150-9.
83. Record, M.T., Jr. and M.L. Lohman, *A semiempirical extension of polyelectrolyte theory to the treatment of oligoelectrolytes: application to oligonucleotide helix-coil transitions.* Biopolymers, 1978. **17**(1): p. 159-166.
84. Fuchs, J., et al., *Salt concentration effects on equilibrium melting curves from DNA microarrays.* Biophys J, 2010. **99**(6): p. 1886-95.
85. Vainrub, A. and B.M. Pettitt, *Coulomb blockage of hybridization in two-dimensional DNA arrays.* Phys Rev E Stat Nonlin Soft Matter Phys, 2002. **66**(4 Pt 1): p. 041905.
86. Peterlinz, K.A., et al., *Observation of hybridization and dehybridization of thiol-tethered DNA using two-color surface plasmon resonance spectroscopy.* Journal of American Chemical Society, 1997. **119**: p. 3401-3402.
87. Marmur, J. and P. Doty, *Thermal renaturation of deoxyribonucleic acids.* J Mol Biol, 1961. **3**: p. 585-94.
88. Marmur, J. and P.O. Ts'O, *Denaturation of deoxyribonucleic acid by formamide.* Biochim Biophys Acta, 1961. **51**: p. 32-6.
89. Subirana, J.A. and P. Doty, *Kinetics of Renaturation of Denatured DNA. I. Spectrophotometric results.* Biopolymers, 1966. **4**(2): p. 171-87.
90. Blake, R.D. and S.G. Delcourt, *Thermodynamic effects of formamide on DNA stability.* Nucleic Acids Res, 1996. **24**(11): p. 2095-103.
91. Sen, A. and P.E. Nielsen, *On the stability of peptide nucleic acid duplexes in the presence of organic solvents.* Nucleic Acids Res, 2007. **35**(10): p. 3367-74.
92. Bonner, J., G. Kung, and I. Bekhor, *A method for the hybridization of nucleic acid molecules at low temperature.* Biochemistry, 1967. **6**(12): p. 3650-3.
93. Weiss, S.B., et al., *Transfer RNA coded by the T4 bacteriophage genome.* Proc Natl Acad Sci U S A, 1968. **61**(1): p. 114-21.
94. Church, R.B. and B.J. McCarthy, *Related base sequences in the DNA of simple and complex organisms. II. The interpretation of DNA-RNA hybridization studies with mammalian nucleic acids*. Biochem Genet, 1968. **2**(1): p. 55-73.
95. McCarthy, B.J. and B.L. McConaughy, *Related base sequences in the DNA of simple and complex organisms. I. DNA-DNA duplex formation and the incidence of partially related base sequences in DNA.* Biochem Genet, 1968. **2**(1): p. 37-53.
96. McConaughy, B.L., C.D. Laird, and B.J. McCarthy, *Nucleic acid reassociation in formamide.* Biochemistry, 1969. **8**(8): p. 3289-95.

97. Casey, J. and N. Davidson, *Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide.* Nucleic Acids Res, 1977. **4**(5): p. 1539-52.
98. Hutton, J.R., *Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea.* Nucleic Acids Res, 1977. **4**(10): p. 3537-55.
99. Fuchs, J., et al., *Effects of formamide on the thermal stability of DNA duplexes on biochips.* Anal Biochem, 2010. **397**(1): p. 132-4.
100. Yilmaz, L.S. and D.R. Noguera, *Development of thermodynamic models for simulating probe dissociation profiles in fluorescence in situ hybridization.* Biotechnol Bioeng, 2007. **96**(2): p. 349-63.
101. Yilmaz, L.S., L.I. Bergsven, and D.R. Noguera, *Systematic evaluation of single mismatch stability predictors for fluorescence in situ hybridization.* Environ Microbiol, 2008. **10**(10): p. 2872-85.
102. Gotoh, M., et al., *A new approach to determine the effect of mismatches on kinetic parameters in DNA hybridization using an optical biosensor.* DNA Res, 1995. **2**(6): p. 285-93.
103. Sorokin, N.V., et al., *Discrimination between perfect and mismatched duplexes with oligonucleotide gel microchips: role of thermodynamic and kinetic effects during hybridization.* J Biomol Struct Dyn, 2005. **22**(6): p. 725-34.
104. Khomyakova, E., et al., *On-chip hybridization kinetics for optimization of gene expression experiments.* Biotechniques, 2008. **44**(1): p. 109-17.
105. Wetmur, J.G. and N. Davidson, *Kinetics of renaturation of DNA.* J Mol Biol, 1968. **31**(3): p. 349-70.
106. Wetmur, J.G., *Hybridization and renaturation kinetics of nucleic acids.* Annu Rev Biophys Bioeng, 1976. **5**: p. 337-61.
107. Gharaibeh, R.Z., et al., *Application of equilibrium models of solution hybridization to microarray design and analysis.* PLoS One, 2010. **5**(6): p. e11048.
108. SantaLucia, J., Jr., *Physical principles and visual-OMP software for optimal PCR design.* Methods Mol Biol, 2007. **402**: p. 3-34.
109. Suzuki, S., et al., *Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays.* BMC Genomics, 2007. **8**: p. 373.
110. Relogio, A., et al., *Optimization of oligonucleotide-based DNA microarrays.* Nucleic Acids Res, 2002. **30**(11): p. e51.
111. Chou, C.C., et al., *Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression.* Nucleic Acids Res, 2004. **32**(12): p. e99.
112. Gao, Y., L.K. Wolf, and R.M. Georgiadis, *Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison.* Nucleic Acids Res, 2006. **34**(11): p. 3370-7.
113. Zhang, Y., D.A. Hammer, and D.J. Graves, *Competitive hybridization kinetics reveals unexpected behavior patterns.* Biophys J, 2005. **89**(5): p. 2950-9.
114. Pozhitkov, A., et al., *Evaluation of gel-pad oligonucleotide microarray technology by using artificial neural networks.* Appl Environ Microbiol, 2005. **71**(12): p. 8663-76.
115. Lauffenburger, D.A. and J.J. Linderman, *Receptors: models for binding, trafficking, and signaling*. 1993, Oxford: Oxford University Press.
116. Chan, V., D.J. Graves, and S.E. McKenzie, *The biophysics of DNA hybridization with immobilized oligonucleotide probes.* Biophys J, 1995. **69**(6): p. 2243-55.

117. Halperin, A., A. Buhot, and E.B. Zhulina, *On the hybridization isotherms of DNA microarrays: the Langmuir model and its extensions.* Journal of Physics: Condensed Matter, 2006. **18**(18): p. S463-S490.
118. Gadgil, C., et al., *A diffusion-reaction model for DNA microarray assays.* J Biotechnol, 2004. **114**(1-2): p. 31-45.
119. Li, S., A. Pozhitkov, and M. Brouwer, *A competitive hybridization model predicts probe signal intensity on high density DNA microarrays.* Nucleic Acids Res, 2008. **36**(20): p. 6585-91.
120. Wang, J.Y. and K. Drlica, *Modeling hybridization kinetics.* Math Biosci, 2003. **183**(1): p. 37-47.
121. Erickson, D., D. Li, and U.J. Krull, *Modeling of DNA hybridization kinetics for spatially resolved biochips.* Anal Biochem, 2003. **317**(2): p. 186-200.
122. Horne, M.T., D.J. Fish, and A.S. Benight, *Statistical thermodynamics and kinetics of DNA multiplex hybridization reactions.* Biophys J, 2006. **91**(11): p. 4133-53.
123. Bishop, J., A.M. Chagovetz, and S. Blair, *Kinetics of multiplex hybridization: mechanisms and implications.* Biophys J, 2008. **94**(5): p. 1726-34.
124. Bishop, J., et al., *Competitive displacement of DNA during surface hybridization.* Biophys J, 2007. **92**(1): p. L10-2.
125. Bishop, J., A.M. Chagovetz, and S. Blair, *Competitive displacement: a sensitive and selective method for the detection of unlabeled molecules.* Opt Express, 2007. **15**(8): p. 4390-7.
126. Bishop, J., S. Blair, and A. Chagovetz, *Convective flow effects on DNA biosensors.* Biosens Bioelectron, 2007. **22**(9-10): p. 2192-8.
127. Chechetkin, V.R., *Two-compartment model for competitive hybridization on molecular biochips.* Physics Letters A, 2007(360): p. 491-494.
128. Chechetkin, V.R., *Kinetics of binding and geometry of cells on molecular biochips.* Physics Letters A, 2007(366): p. 460-465.
129. Watterson, J.H., et al., *Effects of oligonucleotide immobilization density on selectivity of quantitative transduction of hybridization of immobilized DNA.* Langmuir, 2000. **16**: p. 4984-4992.
130. Vainrub, A. and B.M. Pettitt, *Thermodynamics of association to a molecule immobilized in an electric double layer.* Chemical Physics Letters, 2000. **323**: p. 160-166.
131. Vainrub, A. and B.M. Pettitt, *Surface electrostatic effects in oligonucleotide microarrays: control and optimization of binding thermodynamics.* Biopolymers, 2003. **68**: p. 265-270.
132. Peterson, A.W., R.J. Heaton, and R.M. Georgiadis, *The effect of surface probe density on DNA hybridization.* Nucleic Acids Res, 2001. **29**(24): p. 5163-8.
133. Gong, P. and R. Levicky, *DNA surface hybridization regimes.* Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5301-6.
134. Wong, I.Y. and N.A. Melosh, *An electrostatic model for DNA surface hybridization.* Biophys J, 2010. **98**(12): p. 2954-63.
135. Vesnaver, G. and K.J. Breslauer, *The contribution of DNA single-stranded order to the thermodynamics of duplex formation.* Proc Natl Acad Sci U S A, 1991. **88**(9): p. 3569-73.
136. Bogacki, P. and L.F. Shampine, *A 3(2) pair of Runge–Kutta formulas.* Applied Mathematics Letters, 1989. **2**(4): p. 321–325.
137. Myszka, D.G., et al., *Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data.* Biophys J, 1998. **75**(2): p. 583-94.

138. Gracey, A.Y. and A.R. Cossins, *Application of microarray technology in environmental and comparative physiology.* Annu Rev Physiol, 2003. **65**: p. 231-59.
139. Qin, L.X. and K.F. Kerr, *Empirical evaluation of data transformations and ranking statistics for microarray analysis.* Nucleic Acids Res, 2004. **32**(18): p. 5471-9.
140. Levenberg, K., *A Method for the Solution of Certain Non-Linear Problems in Least Squares.* Quarterly of Applied Mathematics, 1944. **2**: p. 164-168.
141. Marquardt, D., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters.* SIAM Journal on Applied Mathematics, 1963. **11**(2): p. 431-441.
142. Fish, D.J., et al., *Multiplex SNP discrimination.* Biophys J, 2007. **92**(10): p. L89-91.
143. Williams, L., et al., *The paradox of multiplex DNA melting on a surface.* Anal Biochem, 2011. **409**(1): p. 150-2.
144. Eimer, W. and R. Pecora, *Rotational and translational diffusion of short rodlike molecules in solution: Oligonucleotides.* Journal of Chemical Physics, 1991. **94**(3): p. 2324-2329.
145. Garcia de la Torre, J., S. Navarro, and M.C. Lopez Martinez, *Hydrodynamic properties of a double-helical model for DNA.* Biophys J, 1994. **66**(5): p. 1573-9.
146. Shi, L., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-61.
147. Petrone, J., *Exodus of GE, ABI, and Others from Array Space a Sign of Market Maturity, Users Say.* GenomeWeb BioArray News, 2008.
148. Petrone, J., *Roche to Phase out Chip-Based Sequence Capture; Peptide Arrays 'Close' to Commercialization.* GenomeWeb Bioarray News, 2012.
149. Petrone, J., *Affy, Agilent, Illumina Affirm Commitment to Array Market in Light of Roche's Planned Exit.* GenomeWeb Bioarray News, 2012.
150. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
151. Shendure, J., *The beginning of the end for microarrays?* Nat Methods, 2008. **5**(7): p. 585-7.
152. Asmann, Y.W., M.B. Wallace, and E.A. Thompson, *Transcriptome profiling using next-generation sequencing.* Gastroenterology, 2008. **135**(5): p. 1466-8.
153. Hey, Y. and S.D. Pepper, *Interesting times for microarray expression profiling.* Brief Funct Genomic Proteomic, 2009. **8**(3): p. 170-3.
154. Hurd, P.J. and C.J. Nelson, *Advantages of next-generation sequencing versus the microarray in epigenetic research.* Brief Funct Genomic Proteomic, 2009. **8**(3): p. 174-83.
155. McLoughlin, K.S., *Microarrays for pathogen detection and analysis.* Brief Funct Genomics, 2011. **10**(6): p. 342-53.
156. Richter, B.G. and D.P. Sexton, *Managing and analyzing next-generation sequence data.* PLoS Comput Biol, 2009. **5**(6): p. e1000369.
157. Doktycz, M.J., et al., *Thermodynamic stability of the 5' dangling-ended DNA hairpins formed from sequences 5'-(XY)2GGATAC(T)4GTATCC-3', where X, Y = A, T, G, C.* Biopolymers, 1990. **30**(7-8): p. 829-45.
158. Riccelli, P.V., K.E. Mandell, and A.S. Benight, *Melting studies of dangling-ended DNA hairpins: effects of end length, loop sequence and biotinylation of loop bases.* Nucleic Acids Res, 2002. **30**(18): p. 4088-93.

159. SantaLucia, J., Jr. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 415-40.
160. Senior, M., R.A. Jones, and K.J. Breslauer, *Influence of dangling thymidine residues on the stability and structure of two DNA duplexes.* Biochemistry, 1988. **27**(10): p. 3879-85.

Appendix A: DNA Thermodynamics

A.1 Derivation of T_M equation

The melting temperature T_M is defined as the temperature at which half of the DNA strands are in the double-helical state, and half are in the random coil state. Marking K as the equilibrium constant of the reaction and C_T, S_{1eq}, S_{2eq}, D_{eq} as the total strand concentration and concentrations of the two strands and duplex at equilibrium, respectively, we can write

$$K = \frac{D_{eq}}{S_{1eq} \cdot S_{2eq}} = \frac{C_T/4}{(C_T/4) \cdot (C_T/4)} = \frac{4}{C_T},$$

where we assumed that the two strands S_1 and S_2 are present at equal concentrations. From the other hand, the standard free energy change of the reaction, ΔG^0 can be written out as follows:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RTlnK,$$

where R is the universal gas constant and T is temperature. Combining the two relationships, we obtain

$$(\Delta S^0 - RlnK) \cdot T_M = \left(\Delta S^0 + Rln(\frac{1}{K})\right) \cdot T_M = \Delta H^0$$

$$T_M = \frac{\Delta H^0}{\Delta S^0 + Rln(C_T/4)}$$

Figure 58 depicts T_m values as a function of total target concentration for duplex 1537/1537. Increase in total oligonucleotide concentration by two orders of magnitude translates roughly into an increase in T_m by 7 ^oC.

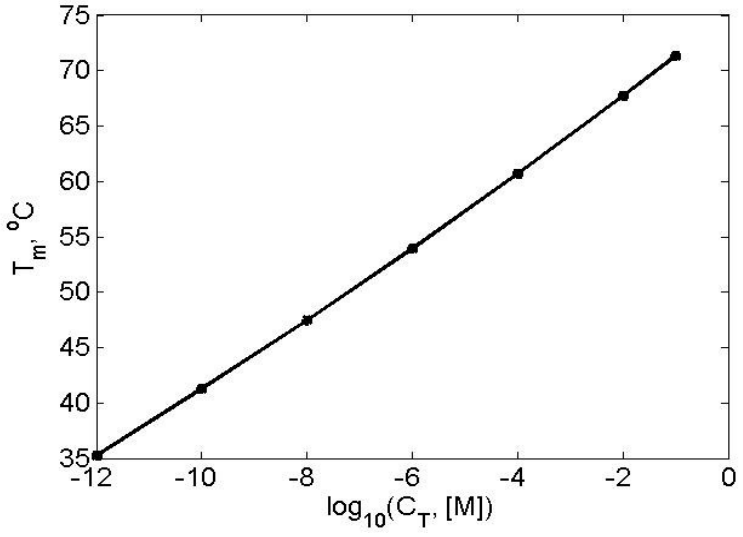

Figure 58 – T_m as a function of total oligonucleotide concentration. ΔH°=-148 kcal/mol, ΔS°=-422.5 cal/(K·mol), values corresponding to duplex 1537/1537 at 50 mM NaCl.

A.2 Derivation of thermodynamic parameters from melt curves

The hyperchromicity method

The thermodynamic parameters are mainly derived using either the van't Hoff method or the concentration method. Van't Hoff equation relates the change in temperature T to the change in the equilibrium constant K given the change in standard enthalpy H^0 for the process:

$$\frac{d\ lnK}{dT} = \frac{\Delta H^0}{RT^2},$$

R is the universal gas constant. Integrating this between temperatures T_1 and T_2 gives the following relationship:

$$ln\left(\frac{K_2}{K_1}\right) = \frac{\Delta H^0}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right),$$

where K_1 and K_2 are equilibrium constants at absolute temperatures T_1 and T_2, respectively. Using the following relationships:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0$$

$$\Delta G^0 = -RTlnK,$$

we obtain

$$lnK = -\frac{\Delta H^0}{RT} + \frac{\Delta S^0}{R}.$$

Plotting *lnK* as a function of inverse temperature, one can obtain ΔH^0 from the slope and ΔS^0 from the Y-intercept of the so-called van't Hoff plot. K as a function of temperature can be determined from thermal melting experiments using the hyperchromicity method with the following rationale. The fraction α_T of DNA strands that remains hybridized in a duplex at a particular temperature T can be expressed as

$$\alpha_T = \frac{A_S - A}{A_S - A_d},$$

where A_s is the absorbance of the single strands in fully dehybridized condition, A_d the absorbance of duplex in fully hybridized condition, and A the absorbance at a particular point on the equilibrium thermal melting curve at temperature T. Assuming a two-state model of DNA binding, and denoting the total concentration of strands c_{TS}, K_T can be calculated using α_T in the following manner:

$$K_T = \frac{\alpha_T \cdot c_{TS}/2}{(1-\alpha_T)^2 \cdot \left(\frac{c_{TS}}{2}\right)^2} = \frac{2\alpha_T}{(1-\alpha_T)^2 \cdot c_{TS}}.$$

Concentration method

This method obtains the thermodynamic parameters using the relationship between T_m and c_{TS}. At T_m, by definition α_T=0.5. Substituting this into K_T equation above, and then substituting that into lnK_T expression from van't Hoff equation, we obtain the following:

$$K_T = \frac{4}{c_{TS}}$$

$$\frac{1}{T_m} = \frac{R}{\Delta H^0} lnc_{TS} + \frac{\Delta S^0 R ln4}{\Delta H^0}.$$

This equation is used to obtain thermodynamic parameters from a linear fit to a plot of $1/T_m$ versus ln c_{TS}.

A.3 Unified oligonucleotide nearest-neighbor parameters ΔH^0 and ΔS^0

The unified nearest-neighbor parameters ΔH^0 and ΔS^0 used throughout this work are summarized in Table 19.

Table 19 – Unified oligonucleotide nearest-neighbor thermodynamic parameters in 1 M NaCl [66, 67]. The dimer duplexes as marked with a slash separating strands in antiparallel notation. For example, notation CA/GT means sequence 5'-CA-3' paired with 3'-GT-5' sequence.

Sequence	ΔH^o kcal/mol	ΔS^o cal/K·mol
AA/TT	-7.9	-22.2
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
CA/GT	-8.5	-22.7
GT/CA	-8.4	-22.4
CT/GA	-7.8	-21.0
GA/CT	-8.2	-22.2
CG/GC	-10.6	-27.2
GC/CG	-9.8	-24.4
GG/CC	-8.0	-19.9
Init. w/term G·C	0.1	-2.8
Init. w/term A·T	2.3	4.1
Symmetry correction	0	-1.4

Appendix B: Calculated Thermodynamic Parameters for Duplexes Used in Chapter 2

Table 20 – Thermodynamic parameters for duplex combinations used in Chapter 2, calculated from nearest-neighbor parameters for perfect matches (Table 19) and internal mismatches [71], for three difference NaCl concentrations. Salt concentration for ΔS^0 was calculated based on the empirical relationship outlined in section 2.5.1. ΔH^0 is considered independent of salt concentration.

	ΔH^o kcal/mol	ΔS^o (1M NaCl) cal/K·mol	ΔS^o(110 mM NaCl) cal/K·mol	ΔS^o(50 mM NaCl) cal/K·mol
EUB338/EUB338	-139.9	-371.5	-385.3	-390.2
EUB338/EUB338MM	-128.9	-346.7	-360.5	-365.4
1537/1537	-148.0	-402.7	-417.3	-422.5
1537/1271	-128.1	-350.6	-365.2	-370.4
1538/1537	-130.5	-364.7	-379.3	-384.5
1538/1271	-110.6	-312.6	-327.2	-332.4
62/62	-143.4	-383.0	-396.8	-401.7
399/62	-120.4	-325.9	-339.7	-344.6
323/323	-146.1	-399.0	-413.6	-418.8
1282/323	-124.9	-349.1	-363.72	-368.9

Appendix C: Calculated Thermodynamic Parameters for Duplexes Used in Chapter 3

Table 21 – Thermodynamic parameters for duplex combinations used in Chapter 3, calculated from nearest-neighbor parameters for perfect matches (Table 19) and internal mismatches [71], for three (or two in case of 1537 and 62 and their mismatches) different NaCl concentrations. Salt concentration for ΔS^0 was calculated based on the empirical relationship outlined in section 2.5.1. ΔH^0 is considered independent of salt concentration.

	ΔH^o kcal/mol	ΔS^o (1 M NaCl) cal/K·mol	ΔS^o (900 mM NaCl) cal/K·mol	ΔS^o (4 mM NaCl) cal/K·mol
EUB338/EUB338	-139.9	-371.5	-372.2	-406.0
EUB338MM/EUB338	-122.4	-333.5	-334.2	-368.0
1537/1537	-148	-402.7		-439.3
1271/1537	-119.1	-325.0		-361.6
1538/1537	-130.5	-364.7		-401.3
62/62	-143.4	-383.0		-417.5
399/62	-120.4	-325.9		-360.4

Appendix D: MATLAB Code for One-compartment and Two-compartment Models

The following code for the one-compartment model was used for thermal dissociation simulations in Chapter 2 and Chapter 3. The code for the two-compartment model was used for simulations in Chapter 3. The parameter values – initial probe and target concentration, thermodynamic parameters, and the initial fraction of probe bound at the beginning of the melt – are given below only as examples.

The solver for differential equations was called out from program main.m. For the one-compartment model, main.m was the following:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
options=odeset('AbsTol',1e-20,'RelTol',1e-17);

%%Until what time to run the simulations
tmax=3600;

%%Specify deltaH and deltas; the values below belong to duplex 1537
deltaH=-148000; deltaS=-417.3;

%%Third value in par is initial probe concentration
par=[deltaH,deltaS,2e-3];

%initial duplex concentration
initD=2e-3;

[t1_1537,y1_1537]=ode23(@diffmelt_1duplex_solution,[t0 tmax],[initD 0]', options, par);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%One-compartment model for solution-based melting

function concdiff = diffmelt_1duplex_solution(t,conc,par)

% define the derivatives of these variables from equations

deltaH=par(1,1);
deltaS=par(1,2);
```

```
b0=par(1,3);
ka=1e6;

%Temperature ramp: 0.83^oC/min, starting temperature 20^oC
T=293+5*t/360;

kdPM=ka*exp(0.5035246727089627*(deltaH-deltaS*T)/T);

bpm=b0-conc(1);

cdotPM=(ka/1000)*conc(2)*bpm-kdPM*conc(1);
adot=-cdotPM;
%
%return the derivatives in dxy in the right order
%
concdiff = [cdotPM adot]';
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The above code assumes that the probe molecules and free target molecules occupy the same volume in the reaction chamber, as is the case during solution-based melting. When approximating the gel-based microarray system with the one-compartment model as was done in Chapter 3, one must also take into account the volume differences: in that case, the function diffmelt_1duplex_solution was modified to include the ratio of probe versus hybridization chamber volume in the adot term.

For two-compartment model for simulating thermal dissociation, the program main.C was modified in the following way:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
options=odeset('AbsTol',1e-20,'RelTol',1e-17);

b0=60e-3;

tmax=3600;
tspan=0:60:tmax;

%Thermodynamic parameters
deltaH=-148000; deltaS=-403.4;

%fraction of probe bound at time t=0
F=1;
```

```
%initial bound duplex concentration
initb=F*b0;

km=0.0671;
par=[deltaH,deltaS,b0,km];

[t2comp,y2comp]=ode23(@diffmelt_1duplex_2compartments,[0 tmax],[initb 0]', options, par);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
```

The function main.C calls out is as follows.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%
%Solving the differential equation of melting

function concdiff = diffmelt_1duplex_2compartments(t, conc,par)

%b0 is initial probe concentration
b0=par(1,3);

deltaH=par(1,1);
deltaS=par(1,2);
km=par(1,4);
ka=1e6;

T=293+5*t/360;

kdPM=ka*exp(0.5035246727089627*(deltaH-deltaS*T)/T);

bpm=b0-conc(1);

cdotPM=(ka/1000)*conc(2)*bpm-kdPM*conc(1);

adot=-(ka/1000)*conc(2)*bpm+kdPM*conc(1)-km*conc(2);

%return the derivatives in dxy in the right order
%
concdiff = [cdotPM adot]';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
```

Appendix E: MATLAB Code for Optimization of Thermodynamic Parameters

For optimizing thermodynamic parameters, using the two-compartment model, the following experimental data was used: a vector of average normalized signal intensity, each component corresponding to a different time point during the melt. The code below requires MATLAB Optimization Toolbox in order to run.

The optimization was started by calling out a program optimizer.m:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
OPTIONS = optimset('TolFun',1e-12, 'Algorithm','levenberg-marquardt');

%Set initial conditions for deltaH and deltaS: these vary depending on the %duplex
X0=[-66089, -168];

[X,RESNORM] = lsqnonlin(@ObjFun1537,X0, [],[],OPTIONS);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
```

The function ObjFun1537 is a duplex-specific function – ObjFun1537 refers to duplex 1537, and is given here as an example. The function diffmelt_1duplex_2compartments that the code below calls out is similar to the function described in Appendix D.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%
function [ yn ] = ObjFun1537( X0)

%Computes the objective function for optimization
%   Calls out the solver for differential equations, and computes the %difference with respect to
experimental curves - total n differences, where %n is the number of experimental measurements
%  The parameters in X are deltaH, deltaS.
```

```
%load the temperature points at which experimental data was collected
T=load('temperature.txt');

%Now load and manipulate the rest of the experimental data

%load the ‘in’ signal
intensities=load('30_35sumintensities1537.txt');

%load the ‘background’ signal
intout=load('30_35sumintensitiesout1537.txt');
IN=30;
OUT=35;
avgint=intensities(:,2:8)/(IN*IN);
avgintout=intout(:,2:8)/(OUT*OUT-IN*IN);

corrint=(avgint-avgintout);

NT=length(T);
corrint0=zeros(NT,7);

for i=1:7
   corrint0(:,i)=corrint(1:NT,i)-min(corrint(1:NT,i));
end

corrint01=zeros(NT,7);
for i=1:7
   corrint01(:,i)=corrint0(:,i)/max(corrint0(:,i));
end

expdata=mean(corrint01');
expdata=expdata(1:49);

options=odeset('AbsTol',1e-20,'RelTol',1e-17);

b0=60e-3;

tspan=0:60:3600;

%Fraction of probe in a duplex at time t=0
F=1;

deltaH=X0(1); deltaS=X0(2); initb=F*b0; km=0.0671;
par=[deltaH,deltaS,b0, km];

[t1it,y1it]=ode23(@diffmelt_1duplex_2compartments,[tspan],[initb 0]', options, par);

y1itnorm=Normalize1(y1it(:,1));

yn=1000*(y1itnorm'-expdata);

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%
```

VITA

Maris Lemba was born in Tallinn, Estonia. She earned a BSc and MSc in Engineering Physics from Tallinn University of Technology before starting graduate school in Seattle. She obtained her PhD in Bioengineering at the University of Washington in Seattle in Fall 2012.