„Microarray based real-time analysis of nucleic acid hybridization kinetics and thermodynamics“

Dissertation
Zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie
Der Johannes Gutenberg-Universität Mainz

DI Siegfried Krainer
geb. am 21.01.1968 in Friesach/Österreich

Mainz, 2011

Contents

List of Figures

List of Tables

Acknowledgements

I want to thank the small atoms for their ability and willingness of forming hydrogen bonds. The smallest of the interatomic interactions is the cause of the unusually low melting temperature of water and thus the basis for the existence of the oceans, the clouds, the snow and the stable helix of the biggest molecule in the known universe. It would be a boring planet without.

Zusammenfassung

Ziel dieser Dissertation war die experimentelle Charakterisierung und quantitative Beschreibung der Hybridisierung von komplementären Nukleinsäuresträngen mit oberflächengebundenen Fängermolekülen für die Entwicklung von integrierten Biosensoren. Im Gegensatz zu lösungsbasierten Verfahren ist dabei die Untersuchung vieler Nukleinsäurekombinationen parallel möglich. Zusätzlich erlaubt die online Messung eine Reduktion der Prozesschritte und bietet sich somit für den Einsatz in integrierten Lab on Chip Systemen an. Ausgangspunkt der experimentellen Charakterisierung war die Entwicklung eines, auf evanszenter optischer Anregung beruhenden, Microarrayreaders für handelsübliche Substrate. Basierend auf den Ergebnissen wurde der gesamte Signalpfad von Nukleinsärekonzentration bis zum digitalen Wert modelliert. Als molekularbiologisches Testsystem wurden das in Eukaryoten universell exprimierte *Actin*-Gen gewählt. Aufbauend auf diesem Gen wurde ein umfassendes Microarray System bestehend aus kurzen und langen Oligonukleotiden (mit eingebauten LNA-Molekülen), cDNA sowie Targets in DNA und RNA realisiert. Die aus der Entwicklung und den Experimenten gewonnen Erkenntnisse über die Kinetik und Thermodynamik von Hybridisierung sind in drei Publikationen zusammengefasst die das Rückgrat dieser Dissertation bilden. Die erste Publikation beschreibt die Verbesserung der Reproduzierbarkeit und Spezifizität durch online Messung von Kinetik und Thermodynamik gegenüber endpunktbasierten Messungen mit Standard Microarrays. Für die Auswertung der riesigen Datenmengen wurden zwei Algorithmen entwickelt, eine reaktionskinetische Modellierung der Isothermen und ein auf der Fermi-Dirac Statistik beruhende Beschreibung des Schmelzüberganges. Diese Algorithmen werden in der zweiten Publikation beschrieben. Durch die Realisierung von gleichen Sequenzen in unterschiedlichen Nukleinsäuren (DNA, RNA und LNA) ist es möglich definierte Unterschiede in der Konformation des Riboserings und der C5-Methylgruppe der Pyrimidine zu generieren. Die kompetitive Wechselwirkung dieser unerschiedlichen Nukleinsäuren gleicher Sequenz und die Auswirkungen auf Kinetik und Thermodynamik ist das Thema der dritten Publikation. Neben der technologischen Entwicklung im Bereich der Sensorik von Hybridisierungsreaktionen von oberflächengebundene Nukleinsäuremmolekülen, der automatisierten Auswertung und Modellierung der anfallenden Datenmengen und dem damit verbundenen besseren quantitiativen Beschreibung von Kinetik und Thermodynamik dieser Reaktionen tragen die Ergebnisse zum besseren Verständnis der physikalisch-chemischen Struktur des elementarsten biologischen Moleküls und seiner nach wir vor nicht vollständig verstandenen Spezifizität bei.

Abstract

This thesis comprises the development of a real-time measurement system for the detection of hybridization and temperature induced denaturation of surface immobilized nucleic acid including the analysis of the kinetic and thermodynamic data obtained. Based on the results from the development of the evanescence based real-time microarray reader prototype a modeling of the whole signal transfer function from the probe and target concentration to the digital output of the electronic signal path was done. Compared to solution based systems the real-time analysis of microarray formats gives the opportunity of parallel investigation of a large number of probe-target combination in parallel. For the evaluation of the prototype the *Actin* X-chip was designed. The *Actin* gene is ubiquitous expressed in the cells of all eukaryote species. Based on this gene a set of probes of different length was designed, including probes with LNA molecules incorporated. Additionally a cDNA array was prepared. Targets were realized in DNA and RNA with different labeling for optimization of signal to background ratio. Based on the results achieved, three publication have been submitted: The first describes the improvement of specificity and reproducibility of real-time measurement compared to endpoint based systems. The algorithms for the automated extraction of kinetic and thermodynamic parameters from the huge dataset are comprised in the second publication. For the modeling of the melting curves a Fermi-Dirac statistics based approach was used. To differentiate sequence effects from structural ones different types of nucleic acid (DNA, RNA, LNA) were preparated and hybridized simultaneously. This gives to opportunity for modification of the ribose and the C5-methylgroup in nucleic acid strands of the same sequence and for investigations on the impact of the chemical groups on hybridization and thermodynamics. The competitive interaction of this isosequential nucleic acid strands is described in the third publication. Beside the technological development in the field of biosensors and the algorithms for the automated analysis of the kinetics and thermodynamics this work is trying to improve the understanding of the physico-chemical behavior of the most elementary molecule of life.

Chapter 1

Introduction

1.1 Motivation

Biology is specificity. This simple statement summarizes the most fascinating aspect of biological systems, the hybridization of complementary biomolecules with almost perfect specificity. Microarrays and especially real-time experiments with microarrays allow an investigation of binding kinetics and thermodynamics for various biomolecules. While endpoint based microarray technology is already widely used to generate data with biological relevance, the physicochemical basics of hybridization kinetics and thermodynamics are far from being understood. This was a motivation to start research on hybridization to solid-phase bound oligonucleotides with real-time monitoring. This thesis summarizes our interdisciplinary research in different fields of physics, chemistry and genetics: extraction, amplification, cloning and labeling of the nucleic acids, design and preparation of the microarray slides, optical and electronic hardware development, the data analysis including the programming of analysis tools based on physical modeling and the extraction of kinetic and thermodynamic parameters from the results. But the most important motivation was the fascination for the four (five) elementary particles of life and their complex cooperative interaction when bound to a nucleic acid strand.

1.2 The topic

The development of advanced molecular biological reactions with solid-phase bound biomolecules offers a wide field of opportunities for future lab-on-chip systems. Most of the molecularbiological processes like amplification, transcription, translation can be realized 'on-chip'. While the simulation of all kind of 'technological' building blocks, e.g. electronic circuits, optical systems, FEM-based simulation of thermal flow or fluid dynamics, in this interdisciplinary system is already a mature technology, the quantitative description of biological systems is still a matter of basic research. So there is need for research towards a better understanding of hybridization to solid-phase bound oligonucleotides for the improvement of microarrays and for the development of advanced lab-on-chip systems. For our analysis we developed a molecular biological evaluation system based on the housekeeping gene actin gene of different plant species. Long and short oligo probesets and a cDNA based array were designed. Probes with chemically modified nucleic acids (LNA) were included. DNA and RNA targets were realized with different labeling techniques. Reference measurement have been done on established nucleic acid and antibody based hybridization systems. It could be shown

that reproducibility and quality of microarray based real-time analysis generates valuable data for biological analysis and for a better understanding of the hybridization of nucleic acid.

1.3 Organization of the content

Three thematic topics and the corresponding publications are the building blocks of this thesis, one dealing with the evaluation of the improvement of specificity and reproducibility due to real-time measurement, one explaining the physical model based data analysis and one providing insight into the competitive thermodynamics and kinetics of DNA•DNA and DNA•RNA duplexes including quantitative modeling. The thesis consists of an introduction of the physicochemical basics of nucleic acid hybridization and real-time experiments (Chapter 2). Chapter 3 deals with nucleic acid thermodynamics which is summarized in the publication on specificity improvement of melting analysis and Chapter 4 describes the important part of surface based hybridization kinetics and the influence on microarray hybridization. In Chapter 5 the algorithms for the modeling of nucleic acid hybridization are introduced including the publication about the modeling of kinetic and thermodynamic data based on physical models. The paper on competitive hybridization of RNA and DNA is central part of Chapter 6. Chapter 7 gives the final discussion of the thesis and the scientific outlook. The most important data on the *Actin* genes and the preparation of the Actin X-chip are shown in Appendix A. Reprints of the patents are given in Appendix B-E, Appendix F includes a CV of myself.

Chapter 2

Physical chemistry of nucleic acid hybridization

2.1 Introduction

The discovery of the double helix by Watson and Crick in 1953 immediately provided fundamental new insights into the nature of genetic events and the mechanism of duplication. Today there is a lot of knowledge of both the detail and the variety of DNA and RNA structures themselves, together with the manner in which they are recognized by regulatory, repair, and other proteins, as well as by small molecules. These advances in nucleic acid structural studies have been largely due to the increased power of the experimental approaches of X-ray crystallography and NMR spectroscopy, which have provided most of the highly detailed structural information to date. The dominance of this experimental approaches still continues but molecular modeling and simulation, biochemical probe techniques and solution and surface based hybridization analysis also play important roles in providing information on structure, dynamics, and flexibility, mostly on a macroscopic level. Underlying all of this progress have been the significant technical advances and the rising interest in small, cheap and reliable sensor systems for biological reactions. In this chapter the physico-chemical fundamentals of nucleic acid structure and of microarray based melting analysis will be presented. The second part gives a short overview about the experimental basics of real time microarray analysis. Focus is a quantitative understanding of the signalpath from molecular concentrations of probe and target molecules to digital numbers. Results were achieved with an AIT-developed evanescent wave prototype, a modified Tecan LS scanner [191] and a Genewave Hyblive real-time hybridization station [66, 125]. Focus of this chapter is on the physicochemical and biophysical issues which are relevant for the understanding hybridization kinetics and thermodynamics. So the differences between DNA and RNA are described in detail while there is only a glance on the basics of nucleic acid structure.

2.2 Building blocks of DNA and RNA

In this section the complex field of nucleic acid structure will be touched. Figure 2.1 shows the typical helical structure of an DNA•RNA hybrid. The building blocks of nucleic acids (DNA and RNA) are the nucleotides, the monomers. Each nucleotide includes three components: a phosphate, a sugar, and a nitrogenous base. The phosphate is bonded to the sugar through phophodiester bonds and makes up the backbone of the molecule. The nitrogenous bases form the "rungs" of the ladder and

are connected through hydrogen bonds. The phosphate is the same in DNA and RNA, but the sugar is a ribose for RNA and a deoxyribose for DNA. Very good books about DNA and RNA structure are [209, 90, 5, 170].

2.2.1 D-Ribose: Determining structure and chirality

Pentoses are five-carbon monosaccharides. The most important of these are ribose and deoxyribose, which are found in nucleic acids. Due to the aldehyde (R-CHO) group, ribose and deoxyribose are aldopentoses. Ribose forms part of ribonucleic acid (RNA), and deoxyribose forms part of deoxyribonucleic acid (DNA). The prefix *de-* means without, so *deoxy-* means without oxygen. The stereoisomers of monosacharides can be categorized due to the orientation of the most distant chiral carbon atom of the carbonyl-group. The structures in Figure 2.2 are called open-chain structures. From theory there are $2^n = 8$ stereoisomeres with $n = 3$ for Ribose. D-ribose represents one such permutation and the enantiomer of D-ribose is L-ribose Figure 2.2. While D-ribose determines the right-handed DNA conformation (except Z-DNA) it is possible to synthesize L-handed mirror images of DNA with L-ribose. Thus L-ribose accounts for one stereoisomer of D-ribose. There are six others,

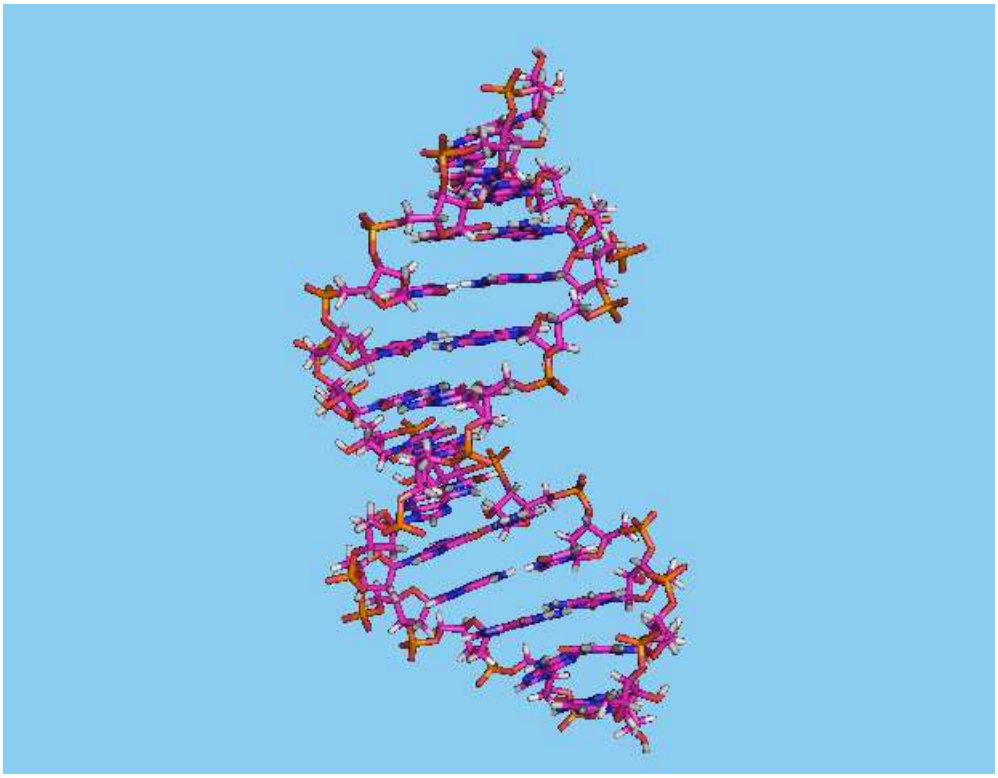


Figure 2.1: Structure of an RNA•DNA hybrid. From [141], Accession code: 1EFS

.

all of them diastereomers of D-ribose, i.e. one or more, but not all, chiral centers have been inverted. In nucleic acids the pentoses are forming ring structures (Figure 2.3). The aldehyde can react with the hydroxyl oxygen at carbon number 4 to form two different hemiacetals. This stereoisomers are labeled with $\alpha-$ and $\beta-$ ribose of nucleic acids. In solutions only a small fraction of the ribose forms a ring structure. In natural nucleic acids the glycosidic bond is always β, the base is above the plane of the sugar and therefore on the same face of the plane as the 5′ hydroxyl substituent (Figure 2.3, Figure 2.9). To form a nucleoside the ribose sugar is attached via glycosidic bond to the base.

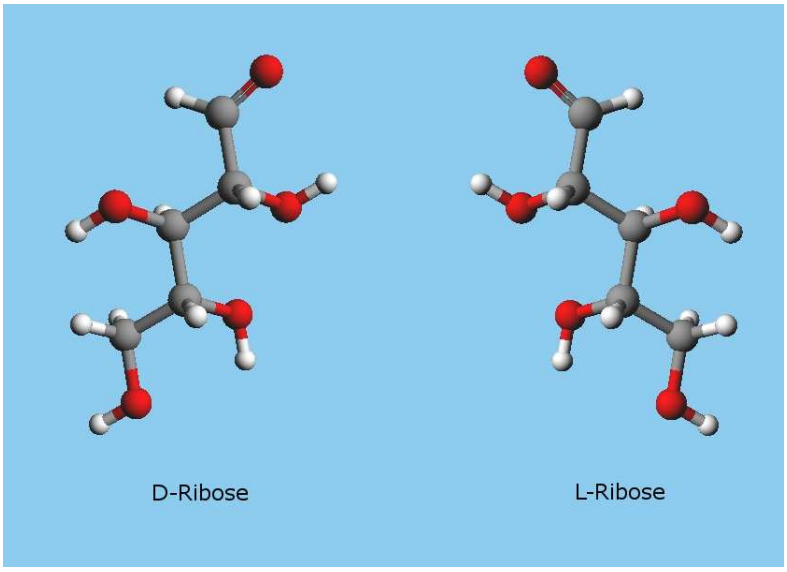


Figure 2.2: D-ribose and the enantiomer L-ribose.

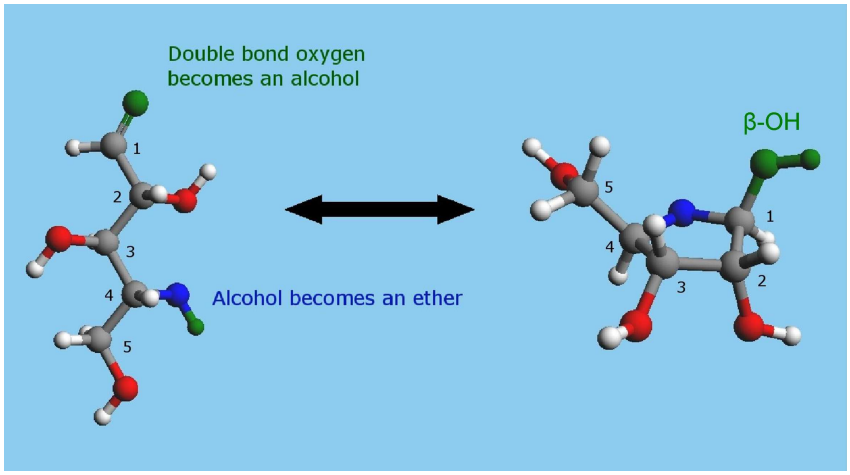


Figure 2.3: Formation of the β-ribose ring from the linear molecule. In the α conformation the green OH-group would point in 3′ direction downward.

The nucleoside units are jointed together through phosphate groups attached to the 3′ and 5′ via phosphodiesterbonds, see Figure 2.4. The full repeating unit in a nucleic acid is a 3′,5′-nucleotide.

2.2.2 Sugar pucker and nucleic acid structure

The existence of 2′-hydroxyl in ribose has important consequences on the structure of nucleic acids [57]. A planar pentose ring is sterically and energeticall very unfavorable. By pulling one atom out of the plain, the strain is released and the energy is lowered. This results in a stable conformation, called sugar pucker. The conformation with the carbon $C3'$ out of the plan and on the same side as the base is called $C3'$-*endo*. Analogous is the conformation with the $C2'$ carbon out of the plane and on the same side as the base called $C2'$-*endo* (Figure 2.5). If the carbon atoms would be on the opposite side of the plane the conformation would be $C3'$-*exo* and $C2'$-*exo*. The

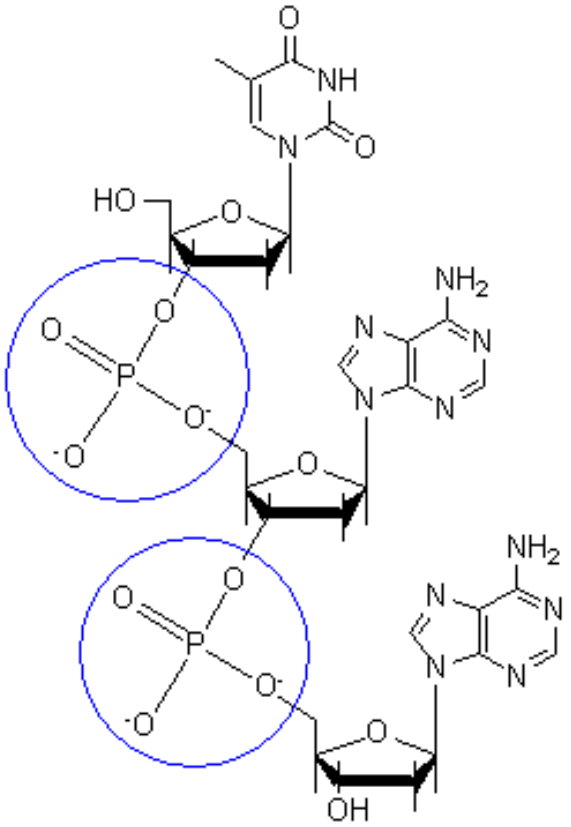


Figure 2.4: Schematic representation of DNA phosphodiesterbonds.

displacement from the plane is 0.5Å or less. The other carbon atom is either in the plane (a pure *endo* conformation) or on the other side of the plane (a mixed *endo*, *exo* conformation). These two conformations approximate those found in DNA and RNA [90]: in double-stranded B-form the deoxyribose sugars are approximately $C2'$-*endo*, while in double-stranded A-form RNA (or A-form DNA) the ribose sugars are in a conformation close to $C3'$-*endo*. In general sugars in DNA•RNA hybrids adopt an intermediate conformation between the A-Form of RNA and the B-form of DNA (see Figure 2.1). A rule of thumb is, the higher the fraction of RNA-purines, the closer to A-structure [75]. A model for the RNA preference for the $C3'$-*endo* is shown in Figure 2.6 [209]. Intrastrand $O2'H - O4'$ hydrogen bonds between adjacent sugar rings and water mediated interaction between the $O2'$ and the phosphate oxygen are stabilizing RNA. It is interesting that structure-stability studies on chemically modified DNA•RNA duplexes showed that the most stable duplexes contained a heteroatom at the 2′-position of the sugar, *i.e.* any atom that is not carbon or hydrogen [60]. Other sugar modifications usually led to diminished hybrid stability. Most backbone modifications that led to improved hybridization restricted backbone mobility and resulted in an A-type sugar pucker [60].

2.2.3 Locked nucleic acid (LNA)

Locked nucleic acid has physical properties similar to RNA but is not degraded by RNAses. LNA nucleosides are a class of nucleic acid analogues in which the ribose ring is locked by a methylene bridge connecting the 2′-O atom and the 4′-C atom 2.8. Thus the ribose is forced into an C3′-*endo* configuration. The structure of a DNA•LNA hybrid is shown in Figure 2.7.

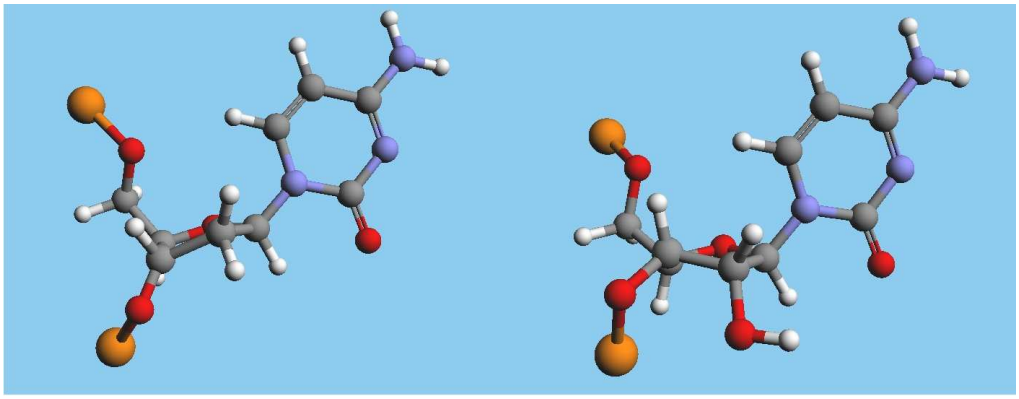


Figure 2.5: Comparison C2′-endo conformation of DNA (left) with C3′-endo of RNA (right).

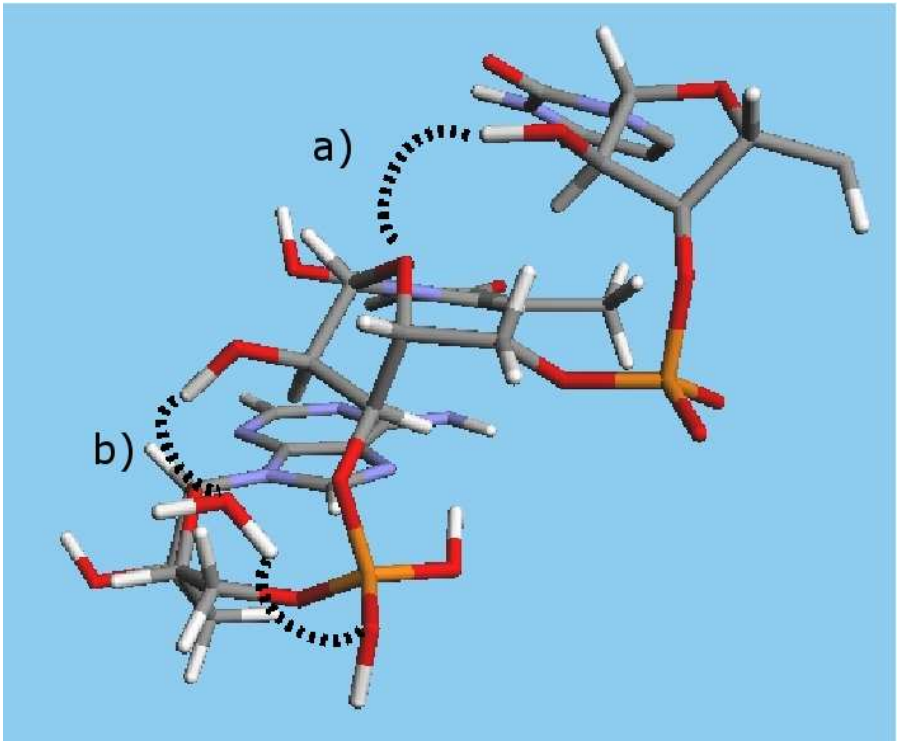


Figure 2.6: Stabilization of C3′-endo puckering in RNA. Intrastrand O2′H-O4′ hydrogen bond (a) and water-mediated O2′-phosphate interaction (b).

Conformation of LNA hybrids

Due to the potential use of LNA as an antisense drug there was a lot of research work done on chemically modified nucleosides [165, 51, 204]. An interesting diastereisomeric form of LNA is the right-handed helix made of LNA nucleotides in α-L-*ribo*-configuration [51, 61, 204]. Depending on the location of the $2'O-4'C$-oxymethylene bridge it is possible to lock the ribose either to $2'-$*endo* with $\beta-D-$LNA (S-type) or to $3'-$*endo* conformation ($\alpha-L-$LNA) (N-tpe). The two conformations are shown in Figure 2.8. Nevertheless, as α-L-LNA is an L-*ribo* configured nucleotide, comparison with D-*ribo* configured nucleotides at the monomer level is difficult and gives no indication of features when built into oligonucleotides [51]. In general $\beta-D-$LNA is locked

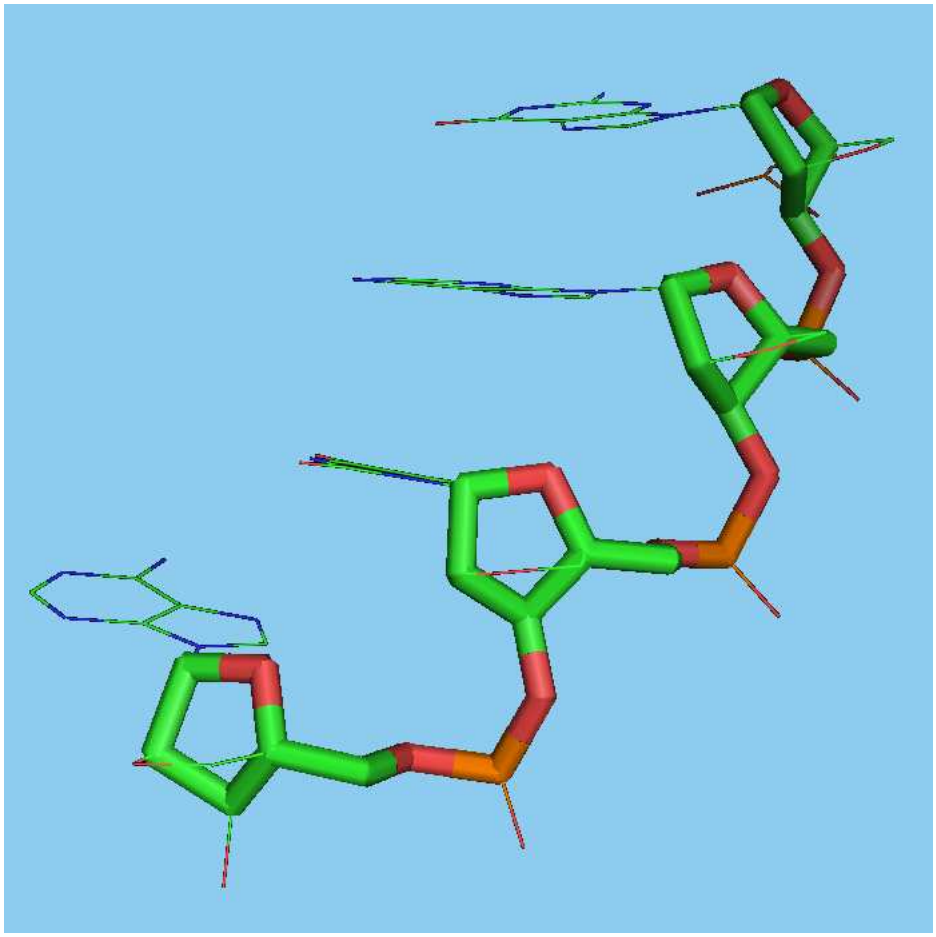


Figure 2.7: Structure of LNA in a DNA•LNA duplex. The phosphor backbone and the ribose is marked with thick sticks, the C3′*endo* configuration is clearly visible. The $2'O-4'C$-oxymethylene bridge of the LNA and the nucleic acid bases are shown in thin lines for visibility.

in a N-type conformation and thus is yielding an A-form duplex with complementary DNA and almost canonical A-form with complementary RNA (the natural form of double stranded RNA) [61]. Duplexes between $\alpha-L-$LNA and DNA adopt a B-form (the natural form of double stranded DNA), whereas duplexes $\alpha-L-$LNA•RNA generate an intermediate structure between A and B form [147].

2.2.4 Bases: The letters of the genetic code

The bases are planar aromatic heterocyclic molecules and are divided into two groups - the pyrimidine (Cytosine (C) and Thymine (T))and the purine (Adenine (A) and Guanine (G)) bases. Their major tautomeric forms are shown in Figure 2.9. The free bases bear a hydrogen atom in positions 9 (purines) and 1 (pyrimidines) which in the nucleotide is replaced by the sugar moiety. Beside the basic nucleotides there exist numerous naturally occurring and chemically synthesized, modified nucleotides, some with antibiotic activity. In general the pyrimidine and purine heterocycles of the bases are planar. The purine bases and cytosine have an amino group, which is important for the immobilization on aldehyde surfaces via Schiff bases. The amino groups are integrated into the aromatic $\pi-$resonance system of the corresponding bases, as the exocyclic C-NH_2 bond distance is considerable shorter than aliphatic C-N single bond (1.34 Å instead of 1.47 Å) [209].

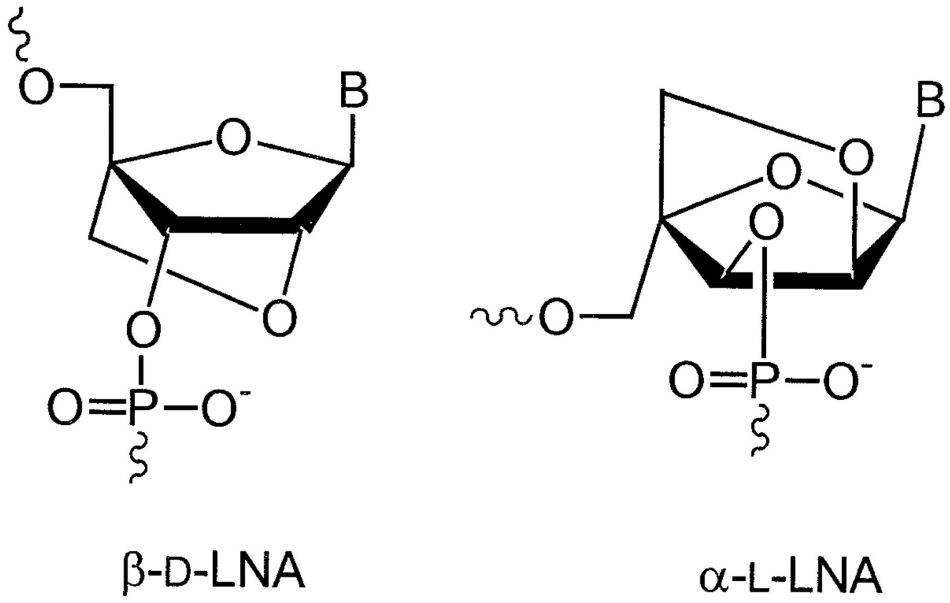


Figure 2.8: Structure of $\beta - L-$LNA and $\alpha - D-$LNA [61]. Depending on the location of the $2'O - 4'C$-oxymethylene bridge it is possible to lock the ribose either to $2'-endo$ or to $3'-endo$ conformation.

2.2.5 Interaction between bases

There are two kinds of interactions: those in the plane of the bases due to hydrogen bonding and the base stacking perpendicular to the base plane stabilized mainly by London dispersion forces and hydrophobic effects (Figure 2.10). Hydrogen bonding is most pronounced in nonpolar solvents where base stacking is negligible. Higher ionic content is also shielding the electrostatic hydrogen bonds. Base stacking dominates in water where base-base hydrogen bonding is greatly suppressed due to competition of binding sites by water molecules. From enthalpic point of view it is not possible to explain the Watson-Crick hydrogen bonds because of the competition to water molecules. The difference between the hydrogen bonding of a water molecule and an opposite base should be negligible from enthalpic point of view. The only possiblity for an explanation are entropic forces. Nevertheless, there is a big influence of stacking on nucleic acid structure.

Hydrogen bonding

Hydrogen bonds are mainly electrostatic in character and interact between bases of the type N-H$\cdots$ N and N-H$\cdots$ O with the donor N-H group of either the amino or imino type [209]. Compared to covalent bonds of well-defined length, strength, and orientation, hydrogen bonds are about 20 to 30 times weaker (3...6 kcal/mol). Under the influence of a hydrogen bond, the charges on the atoms involved are modified due to polarization, H becoming more electropositive and the two hydrogen bond acceptors becoming more negative. This effect leads to increased affinity for accepting further hydrogen bonds [209]. Under the influence of the cooperative effect, hydrogens in base-pairs can jump in concerted mechanism from the donor in one base to the acceptor on the partner base (Figure 2.11). Under the assumption that at least two N-H$\cdots$ O or N-H$\cdots$ N hydrogen bonds must form in order to produce a stable base-pair the four bases can be arranged in 28 different configurations, for an overview see [209, 90]. Fluctuations in local helical conformation of DNA, the phenomenon known as DNA breathing, lead to infrequent events of base pair opening thus making normally buried groups

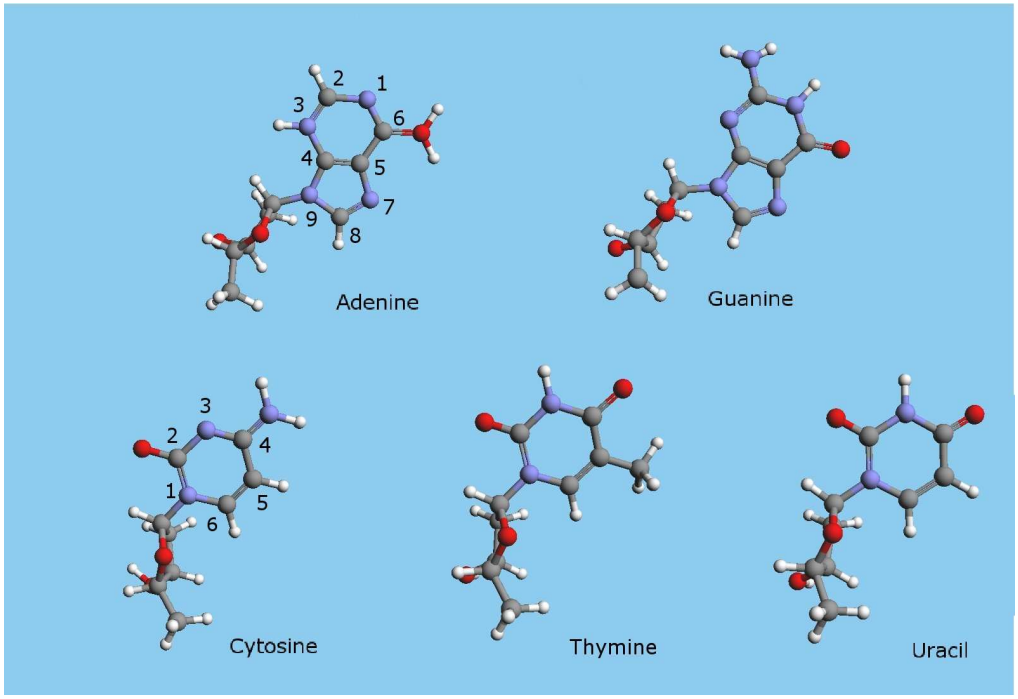


Figure 2.9: The five bases of DNA and RNA. The upper row are the purine bases, the lower the pyrimidines.

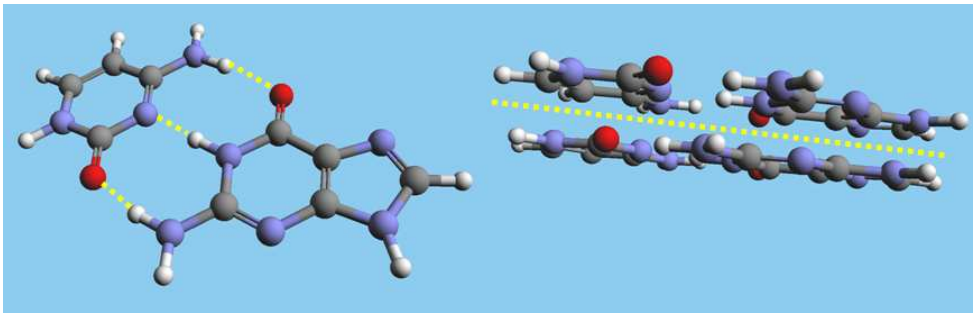


Figure 2.10: Comparison of Watson-Crick basepairing and stacking interaction.

available for modification and interaction with proteins [48, 210, 144, 28, 224]. Fluctuational base pair opening implies disruption of hydrogen bonds between the complementary bases and flipping the base out of the helical stack disrupting two contacts. Obviously the thermal energy is the motor of this structural change. This behavior is described by the Peyrard model [151].

Base stacking

Bases pile up in long stacks like coins in a roll (Figure 2.10). In the solid state bases are found almost exclusively stacked such that one base plane is at the van der Waals distance, ≈ 3.4 Å. In aqueous solutions, such base stacks form as well. There is still an ongoing discussion on the mechanism of stacking [121, 181]. It seems that forces between permanent dipoles are only of minor importance for the stabilization of stacks. Rather it appears that dipole-induced dipole interactions play the major role, with the permanent dipole, predominantly in C=O or C-NH_2 groups, superposed

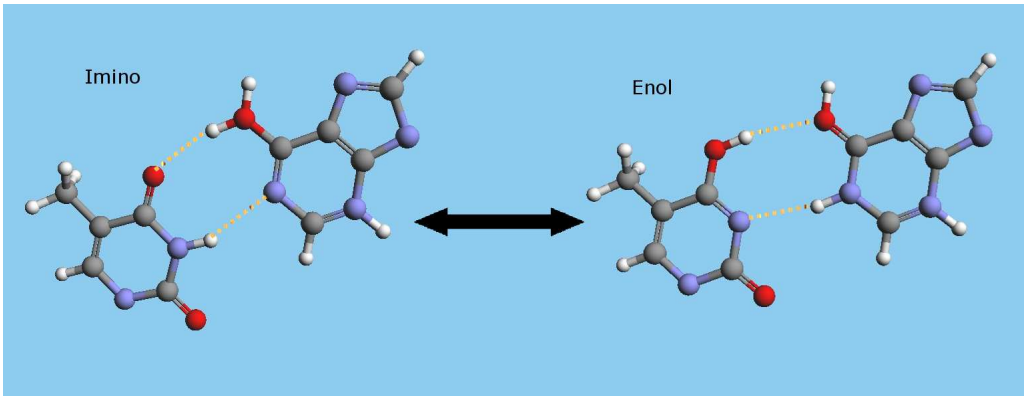


Figure 2.11: Imino and enol form of an AT-basepair. The changes in bonding order are shown, the distances are approximately in scale.

over the π-electronic system of the adjacent base [209]. London dispersion forces and hydrophobic interaction seem to play also an important role in base stacking [5, 170]. While base-pair hydrogen bonding depends on composition, stacking is influenced by composition and sequence. The origin of this energy-sequence correlation is that in alternating purine, pyrimidine sequences the overlap between adjacent bases in a stack is much smaller than the overlap observed in base-paired dimers containing only purine or only pyrimidine bases in one strand. In general, DNA stability is mainly determined by base stacking interactions [224]. While the base stacking in general generates a force

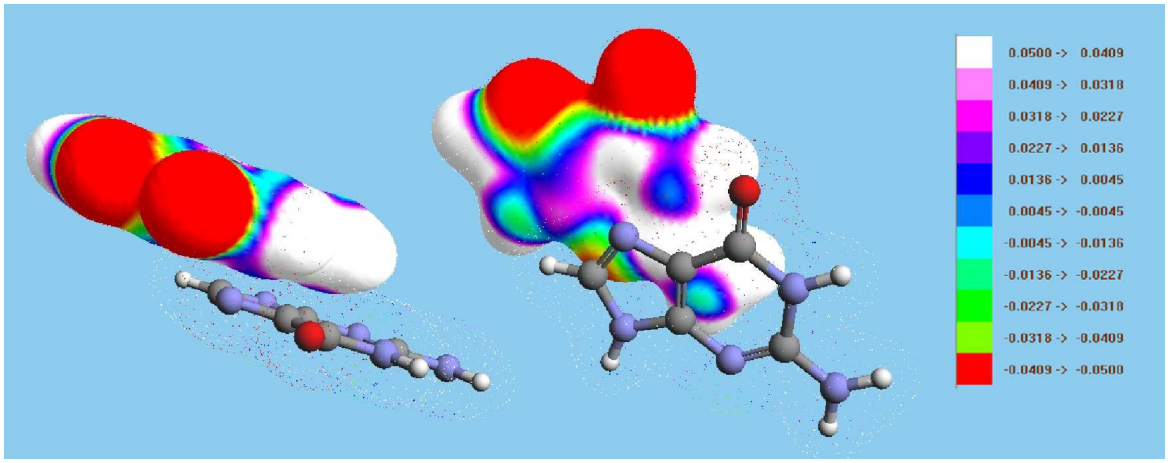


Figure 2.12: Stacking of two guanine bases in two views (distance not in scale). While the hydrophobic forces press the bases together the mostly repellent electrostatic charge distribution aligns them. For visibility the charge distribution of the lower base is represented by points.

which is pressing the bases together, the charge distribution tries to align them. The electrostatic potential of two stacked guanine bases is shown in Figure 2.12, the charge distribution of the five nucleic acid bases is shown in Figure 2.20. The calculation were done with ArgusLab [196, 195] using the ZINDO [91] semi-empirical quantum mechanical algorithm. an experimental approach to study the stacking of nucleic acids is the hybridization of coaxial targets to a single probe [203]. The length of the separated fragments of the interrupted strand must differ sufficiently in order to monitor and analyze the separated transitions. Therefore the melting process of such complexes is non -cooperative consisting of two independent stages. Alternative methods are the use of nicked

structures in solution [224].

2.2.6 The phosphate backbone

A nucleotide is formed by a nucleoside and phosphoric acid. The formation of the phosphate backbone via phosphodiesterbonds is shown in Figure 2.13. The hydroxilgroups of the ribose are forming hydrogen bonds with the surrounding water molecules. Thus the backbone is hydrophilic. DNA is an

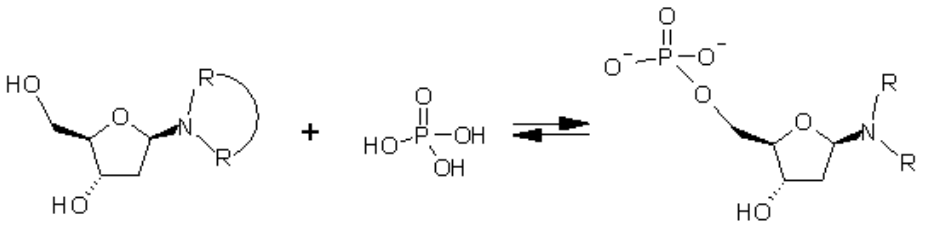


Figure 2.13: Formation of the phophate backbone. The remaining P-OH function is quite acidic and completely ionized in biological systems.

acid, owing to this phosphate groups. The "salt" of DNA is the form in which some of the hydrogen ions have disassociated from the phosphate group. In a sense, a nucleotide is a phosphoester, in which one H of phosphoric acid is replaced by the nucleoside. Denote this "rest" of the molecule by R the dissociation equation becomes

$$RH_2PO_4(l) + 2H_2O(l) \rightleftharpoons RPO_4^{-2}(aq) + 2H_30^+(aq) \tag{2.1}$$

The salt then has a net negative charge, specifically at the oxygens of the phosphates, see Figure 2.4. The phosphodiesterbonds of RNA are hydrolyzed in alkali conditions, while DNA is not affected. The $2'$-hydroxyl group of RNA is the key to the formation of cylcic $2'$, $3'$-monophosphates which are further hydrolyzed.

2.2.7 Structural parameters of nucleic acids

There are mainly three basic conformations for double stranded nucleic acids observed: A-,B-, and Z-DNA. A graphical representation of the three types is shown in Figure 2.15, the key parameters for the three structures are summarized in Table 2.1. The space between the two strands of a nucleic acid helix is forming voids. As the strands are not directly opposite each other, the grooves are unequally sized, see Figure 2.14. The major groove is 22 Å wide and the other, the minor groove, is 12 Å wide. The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove. The $C5'$- methyl group is pointing into the major groove for both strands (Figure 3.12).

A number of rotational and translational parameters (the "Cambridge Accord") have been devised to describe the geometric relations between bases and base pairs, originally defined in 1989 [40]. These definitions are shown in Figure 2.16. In Figure 2.17 the nomenclature for the characterization of backbone torsion angles can be found.

2.2.8 Base opening in RNA and DNA

The DNA bases are protectively embedded in the hydrophobic center of the distinctive double helix. In order to read this code, the bases must be physically exposed by locally separating the strands, e.g.

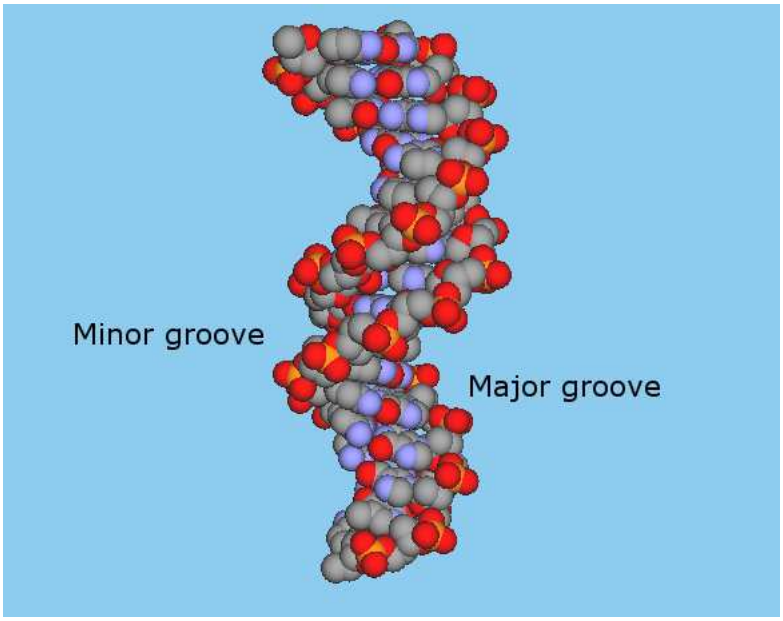


Figure 2.14: Major and minor groove in dsDNA. From [50]: Accesion code 1KBD

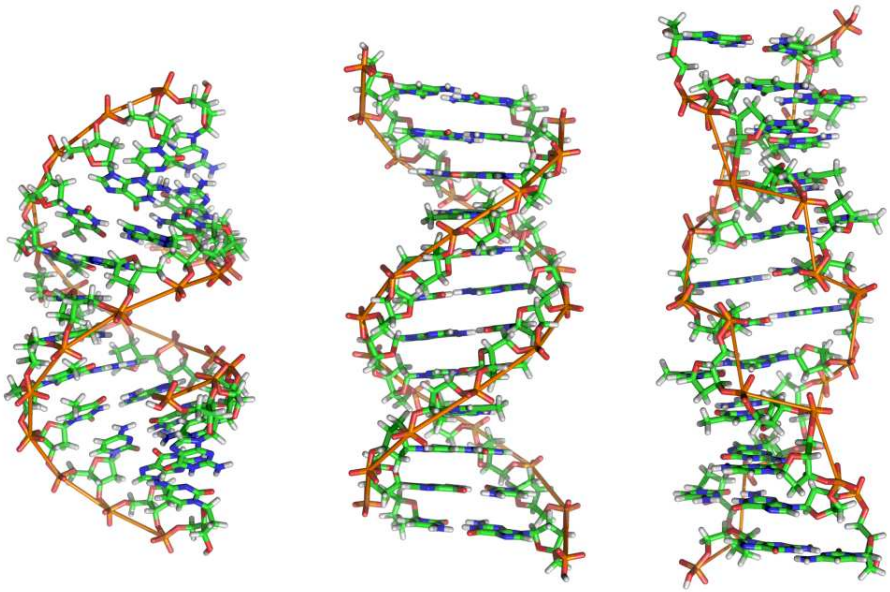


Figure 2.15: The three types of nucleic acid structure, from left to right: A-,B-, and Z-DNA. Source: Wikipedia

at the transcriptional start site (TSS). The denaturation of dsDNA at the right place is essential for initiation of specific gene transcription. While opening sides of DNA are often AT-rich, it is plainly not the case that they always represent the best opening sites in dsDNA [48]. Thermally induced breathing of the two DNA strands is a well-documented phenomenon [210, 8, 130] and it can result in structures significantly different from equilibrium helices. It can give rise to localized DNA melting of 10 bp or more at physiological temperatures, which is similar in size to some transcriptional bubbles [28]. The

Table 2.1: Structural parameters of the three main conformations of nucleic acids

	A-Form	**B-Form**	**Z-Form**
Helix sense	Right handed	Right handed	Left handed
Repeating unit	1 bp	1 bp	2 bp
Rotationbp	33.6°	35.9°	602°
Mean pb/turn	10.7	10.0	12
Inclination of bp	+19.2°	-1.2°	-9°
Rise along axis	2.3Å	3.32Å	3.8Å
Pitchturn of helix	24.6Å	33.2 Å	45.6Å
Propeller twist	+18°	+16°	0°
Glycosyl angle	anti	anti	C: anti, G: syn
Sugar pucker	C3*′*-endo	C2*′*-endo	C: C2*′*-endo, G:C2*′*-exo
Diameter	26Å	20Å	18Å

DNA openings occur spontaneously with thermal energy, demonstrating a crucial role for entropy. Main parameters are the localized sugar-phosphate backbone torsion, base pair hydrogen bonding and adjacent π-electron interactions. Opening of any base pair is coupled to its neighbors, resulting in a cooperative 'bubble' which is observed experimentally [39]. The Peyrard-Bishop nonlinear dynamical model of DNA [151, 38, 229] which has been used to simulate denaturation of short DNA fragments, gives an accurate representation of this phenomenon. Experimental results suggest that the thermal opening profile of DNA may play a role in transcription initiation and other processes. Localized thermal opening was found in double-stranded promoter DNA, but not in non-promoter DNA [28]. Similar to results reported for B-DNA, base opening in dsRNA appears to have an overall preference towards the major groove [27]. Movement of each of the adenine, uracil and cytosine bases into the minor groove is blocked by a high-energy barrier due to severe close contact with neighboring bases. Guanine bases are able to open towards both grooves because of the unique orientation of the base that avoids steric clash along the opening pathway [48], but RNA bases have a substantially smaller major groove opening extent than that of their B-DNA counterparts. A comparison with base opening behavior of A-DNA duplexes suggests that the restriction for RNA results from helix constraint associated with A-form backbone conformation. The reduced opening extent correlates with the RNA duplex stability and is consistent with observed slower imino proton exchange rate in RNA duplexes.

A simple method for the experimental determination of DNA opening is a selective cleavage with S1 glycoprotein nuclease. It is sufficiently sensitive to react with arrested transcriptional elongation complexes which feature a relativey large transcriptional 'bubble'; the physical size of S1 nuclease reduces is ability to detect smaller temporal openings in dsDNA. Thus the bulkiness of S1 nuclease aids in selectively cleaving larger openings over small openings [28].

Time scale of nucleic acid opening

Breathing normal modes are present in all molecules, and are a result of non-zero thermal energy. It is important to distinguish the temporal DNA openings from transcriptional 'bubble' formation. Localized melting is on a very slight scale and, though the probability of opening formation is small, coupled molecular motion occurs on a picosecond to nanosecond timescale. This is distinct from other DNA dynamical features, which occur on a millisecond to microsecond timescale [28]. Experimental data indicate in general AT base pair lifetimes to be in the range of 1-5 ms at 15 ℃ except for AT tracts where lifetimes longer than 100 ms have been observed [48]. For GC base pairs lifetimes about 10 times longer than for AT base pairs usually have been observed, as one might expect from the

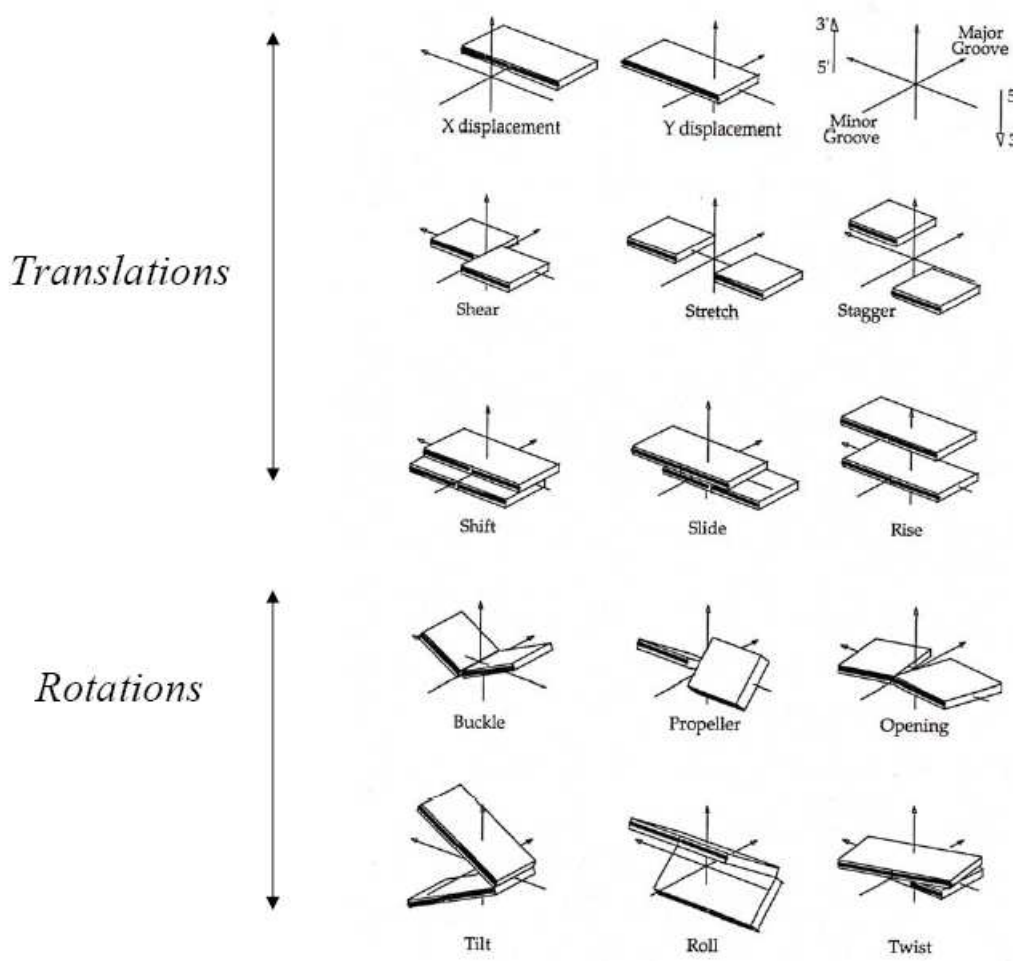


Figure 2.16: Definition and nomenclature of nucleic acid structure parameters according to the "Cambridge Accord" from [40].

higher binding affinity of the GC base pairs. Similar results were found from [102] who estimated for B-DNA A•T and G•C basepair opening probabilities of $10^{-5} - 10^{-6}$ and $10^{-6} - 10^{-7}$. They found that A-tracts of four or more adenines show an order of magnitude slower opening rates. It is interesting that A-tracts are not so sensitive to the reduction of melting temperature due to formamide as well [17]. Sequence-dependent structural features of the DNA double helix have a strong influence on the base pair opening dynamics.

2.2.9 Influence of the C5-methyl group

It is important to remember that the C5-methyl group always points to the major groove side, no matter on which strand of the DNA•DNA duplex the thymine or methylated cytosine is located (Figure 2.18). Investigation on nucleic acid sequences containing series of bases gives insight into the different structural restriction of nuceotides and on base pair dynamics. Sequences of four or more A•T base pairs without a $5' - TA - 3'$ step, so called A-tracts, influence the global properties of DNA by causing curvature of the helix axis if phased with the helical repeat and also influence nucleosome packaging [210]. An example of A-tracts influencing the global properties of DNA is the bending of the helix axis that causes anomalously slow migration in polyacrylamide gels displayed by

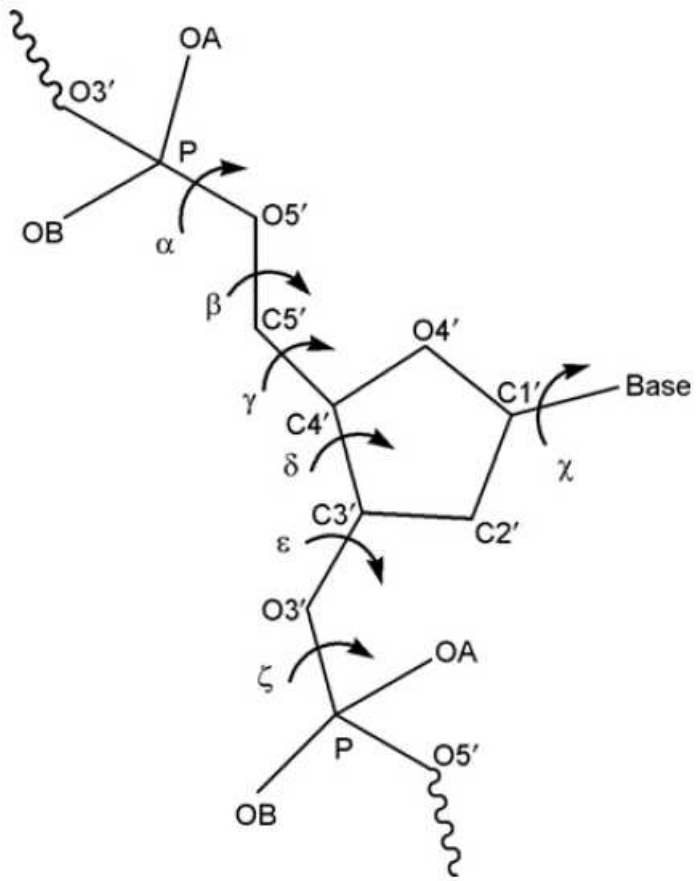


Figure 2.17: The backbone torsion angle in a unit nucleotide. Each rotable bond is indicated by a curved arrow.

Table 2.2: Properties of A and G tract from [48]

A tract	**G tract**
Narrow minor groove	Wide and deep minor groove
Wide and deep major groove	Shallow major groove
Methylated major groove	Unmethylated major groove
No amino group in the minor groove	Amino group in the minor groove
Hydration spine in the minor groove	Poor hydration of the minor groove
Very slow base pair dynamics	Very rapid base pair dynamics

such sequences. It was early observed that slow base pair-opening kinetics were a striking physical property unique to DNA A-tracts [48]. An overview of the structural differences between A and G tracts is shown in Table 2.2. It was found that the base pair opening kinetics in G-tracts, contrary to what has been observed for A-tracts, is unusually fast in particular with high opening rates [48, 119]. According to this study anomalously long base-pair lifetimes are found in DNA tracts of contiguous A•T base pairs. The tract must have a length of at least four, which may include a $5' - AT - 3'$ step but not a $5' - TA - 3'$ step. The lifetime of the A•T base pair at the 5′-A end of the tract is not anomalouly long. The kinetic anomaly is a strong evidence for a *B*-DNA structure. One of many

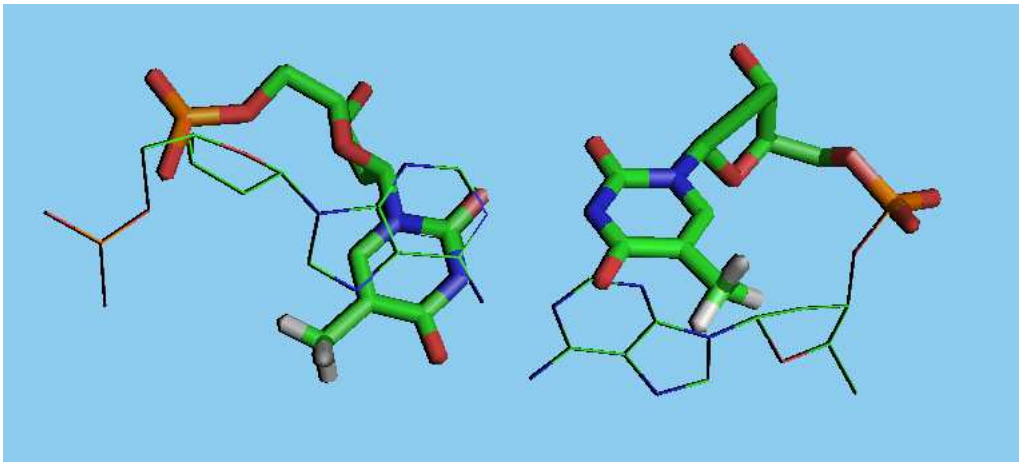


Figure 2.18: Position of the C5-methyl group on an AT stack in an DNA•DNA duplex. The sticks mark the thymine, the lines mark the adenine bases. The C5-methyl groups on both strand point to the major groove.

distinctive features of A- and G-tract DNA is C5′-methylation, which is present on the thymine base but absent on the cytosine base and thus makes the major groove of A- and G-tracts methylated and unmethylated [83, 210]. The influence of the U→ T nor $C5' \rightarrow C5'$m methylation on helix structure or bending of the DNA helix axis is not fully clear [7, 210]. It has been shown that the thymine methyl groups provide the dominant contribution to the hight stability of AT base pairs in A-tracts. There is evidence that the C5′-methyl group occupies a hydrophilic pocket in the major groove of *B*-DNA where it affects the structural and dynamic porperties of the double helix. There is also evidence that the substitution of a methyl group at the C5-position of pyrimidines increases DNA stability via hydrophobic interaction with water molecules. Moreover, methyl substitution increases the molecular polarizability of the pyrimidines, thereby promoting base stacking and this improvement becomes larger for larger propeller twists. There might be other effects contribute to the reduction of base pair breathing by the C5-methyl groups, they are likely to alter the hydration pattern in the major groove which potentially could affect the dynamics. Furthermore, the larger hydrophobicity of the thymine base should favor less solvent-exposed geometries, and this could possible yield a reduced base pair breathing relative to that of the Uracil base. It has been shown that the sugar-phosphate backbone of short pieces of repetitive A-T sequences in DNA form conformation which are very different from *B*-DNA [201, 29, 93]. The methyl group of thymine in an A-T step is placed over the five-member ring of the purine, whereas in the T-A step there was no such interaction. The attractive force is based on the so called CH/π interaction (2.19). This interaction causes an asymmetry, which is not taken into account in existing NN-models. The CH/π interaction is a weak attractive force working between CH groups and π systems [138, 200], the enthalpy of a single Ch/π bond is estimated to be around 1 kcal mol^{-1}. It has been shown that the C5-methyl group and the 2′-OH groups could have nearly the same effect on the stability of the double helix. While both effects have substantial effects on helix stability they can be opposed sometimes [213]. Minor-groove narrowing is often associated with the presence of A-tracts. This effect strongly enhaces the negative potential of the DNA. The nucleosome core particle is a prominent example of this effect [166]. Some of the properties in Table 2.2 do not hold for RNA, containing uracil instead of thymine.

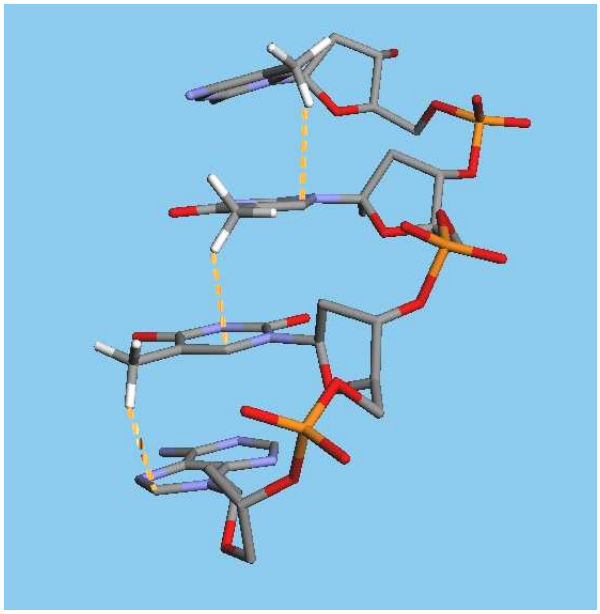


Figure 2.19: Representation of CH/π interaction (dotted line) for the sequence dApTpTpC5m (PDB ID: BDLB73)

2.3 Charge distribution of bases

The electronic charge distribution for each base in its ground and excited states determines its properties. The ground state charge distribution for a molecule is specified by the positions of the nuclei and by the electron density in space around the nuclei. The attraction or repulsion for an external charge can be calculated from the netcharges on the atoms of a molecule. In addition to electrostatic forces one must add London- and van-der-Waals forces. Contrary to Coulomb interactions these forces are due to charge-polarizability and polarizability-polarizability attractions, hydrophobic interaction and short-range electron overlap repulsion. Realistic energy calculations also require the incorporation of solvent effects. The energy of a positive unit charge can be expressed by the simplified Coulomb law [90]:

$$Energy[kJ/mol] = \frac{1389}{\epsilon} \sum \frac{q_i}{r_i} \tag{2.2}$$

or

$$Energy[kcal/mol] = \frac{331,8}{\epsilon} \sum \frac{q_i}{r_i} \tag{2.3}$$

Where q_i is the net charge at each atom (in units of the elementary charge), such as $-0,4$ on a nitrogen or $+0,1$ on a charbon. The static dielectric constant of the medium characterizes how the medium shields the interaction of the charges. Its value is about 80 for water at room temperatur, and it is equal to the square of the refractive index for nonpolar materials ($\epsilon \approx 2 \ldots 4$). In aqueous solutions, ϵ will vary from about 1, if there is no solvent between the charges, to a value approaching 80 at large distances. A common procedure is to use a variable set equal to the magnitude of the distance between charges in Å.

2.3.1 Protonation and ionization of Nucleic Acid bases

Protonation and ionization happens via acceptance or donation of H^+ ions. Similar to the pH-value one defines the pK_a value for the dissociation of a weak acid

$$K_a = \frac{[H^+]\,[A^-]}{[HA]} \tag{2.4}$$

Where $[H^+], [A^-]$ and $[HA]$ are the equilibrium activities. Taking the negative logarithm of the equation gives:

$$pK_a = -\log K_a = -\log \frac{[H^+]\,[A^-]}{[HA]} \tag{2.5}$$

to obtain the Henderson-Hasselbach equation:

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \tag{2.6}$$

Or in the most general form:

$$pH = pK_a + \log \frac{[proton\ acceptor\ or Brønsted\ base]}{[proton\ donor\ or Brønsted\ acid]} \tag{2.7}$$

This equation is important for the understanding of the titration of weak bases and for the all kinds of biological buffers in blood and tissue. Usually, the activity of the hydrogen ions is measured with a pH meter and the ratio of salt and acid activities is replaced by the ratio of molarities. This replacement gives an apparent pK_A that depends on ionic strength. In general the pK_a-values in deoxyribonucleosides are 0,1 to 0,3 pH units larger than ribonucleosides. The addition of a phosphate to form nucleotides raises the nucleoside pK_a values by 0,2 to 0,6 pH units. The negatively charged phosphate attracts the positively charged proton. Under biological conditions ($pH = 7$) the nucleotides are already protonated, compensating the negative charge of the phosphate backbone of the DNA at least partially [209, 90].

2.3.2 Calculation of the ground state charge distribution

Ground state electrostatic potentials for the five bases based on the ZINDO[91] algorithm are shown in Figure 2.20. The calculation was done with Arguslab [196, 195]. There is a remarkable accumulation of negative charge at the major groove side of guanine (the right side of the bases would point to the major groove in this representation).

2.3.3 The interaction with ionic solution

Electrostatic interactions in stiff polyelectrolytes with monovalent counterions and added salt in aqueous solutions are standardly approached via the Poisson-Boltzmann (PB) theory that combines electrostatics with statistical mechanics on a simplified mean-field level [72, 199, 96]. There are several important effects:

- Charge accumulation [198]
- Electrostatic shielding
- Inter-DNA attraction mediated by divalent counterions [158]
- Charge inversion [72]

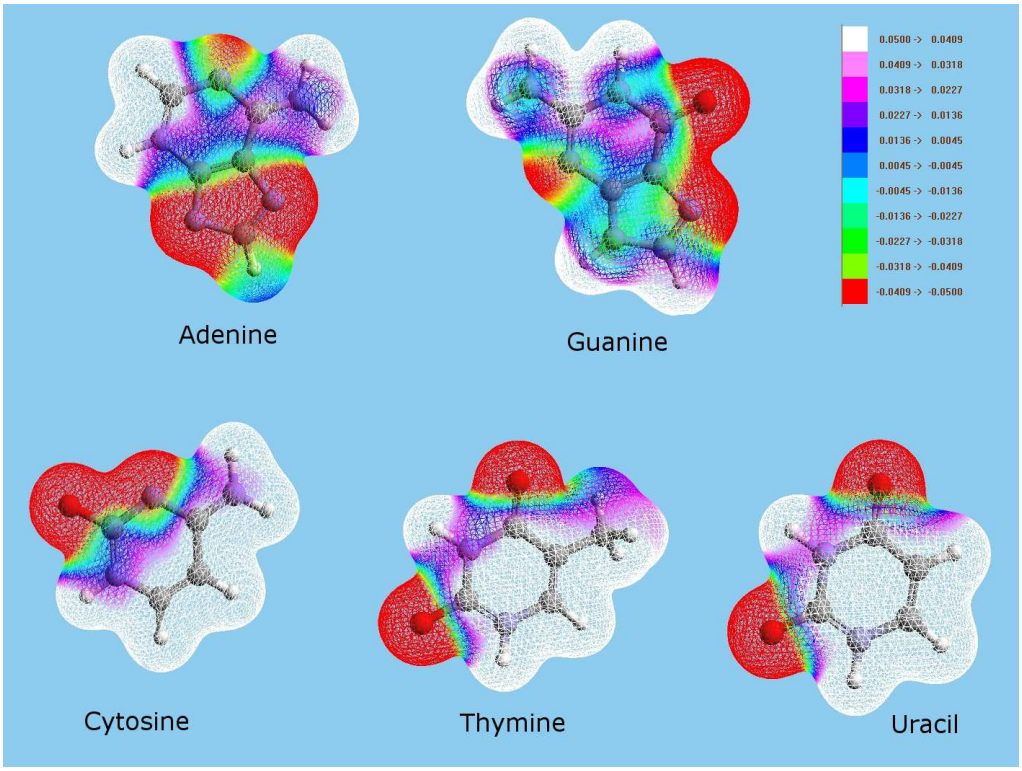


Figure 2.20: The electrostatic potential is mapped on electron density surface. The electron density gives the shape of the surface while the value of the electrostatic potential gives the colors. For all bases the major groove is on the right side in this representation. For the calculation of the ground state electrostatic potential (ESP) and the ground state electron density and Arguslab [196, 195] was used. The single point energy calculation was done with the ZINDO algorithm [91].

Electrostatic forces in electrolytes

Due to the negatively charged phosphorus there is a net charge of -1 every nucleotide in the oligonucleotide strand. This causes several effects on probe and target strands. In an electrostatic field the force of the particles is proportional to their charge q, which is proportional to the number of nucleotides $n_{nucleotides}$:

$$F_{el} = qE \tag{2.8}$$

$$v = \mu E \tag{2.9}$$

$$q \propto n_{nucleotides} \propto l \tag{2.10}$$

Using the Stokes law one gets for the mobility:

$$v = \frac{F_{el}}{6\pi\eta l} = \frac{E}{6\pi\eta} \tag{2.11}$$

Thus the velocity of the polymer is independent of the size. In electrophoresis the faster moving of the shorter molecules comes from the 'wormlike' motion through the gel, which is not described by the Stoke's theory but has been addressed theoretically in the reptation model [230].

Debye length

The Debye length λ is the distance over which significant charge separation can occur in plasmas or in electrolytes:

$$\lambda = \sqrt{\frac{\epsilon_0 \epsilon_r kT}{2N_A e^2 I}} \tag{2.12}$$

A Debye sphere is a volume whose radius is the Debye length, in which there is a sphere of influence, and outside of which charges are screened. The Debye length is of vital importance because of the negative charges of the DNA backbone. It is not fully understood how the charge densities of biopolymers are shielded by electrolytes and water atoms. A rough estimation for the Debye length in saline solutions is (Figure 2.21)

$$\lambda = \frac{0.3nm}{\sqrt{[NaCl]}} \tag{2.13}$$

Thus a 10 mM NaCl solution $[NaCl] = 0,01$ gives a Debye length of about 3 nm. Actual data provide evidence in support of theories that predict an effective diameter of DNA much greater than the Debye screening length, the characteristic distance in a plasma beyond which the electric field of a charged particle is shielded by particles having charges of the opposite sign. It is often assumed that the effective diameter of DNA should correspond to the solvent-dependent Debye length rather than the natural helix diameter. The length for a typical hydrogen bond in water is about 0.2 nm, for the Watson-Crick base pairs the binding lengths is in the range of 0.17 nm. Within our experiments we used NaCl-concentrations of 0.9 M. The Debye shielding length for this conditions is about 0.3 nm according to equation 2.13.

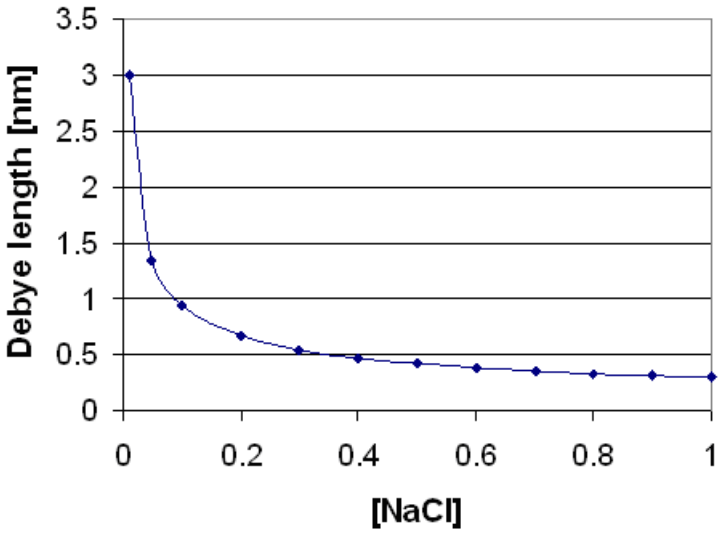


Figure 2.21: Debye screening length vs concentration of NaCl for aqueous solutions. The calculation was done according to eqn. 2.13.

2.4 Physical characterization of nucleic acids

2.4.1 Physics of biopolymers

Charged polymers are ubiquitous in biological systems. Because of their connectivity, stiffness and strong electrostatic interactions, polyelectrolytes show a wide range of complex behavior depending on concentration, added salt and valency of the counterions [199]. DNA is a stiff, highly charged polymer where this complex behavior has been studied in great detail [90]. Electrostatic interactions for a single stiff polyelectrolyte chain such as DNA with monovalent counterions and added salt in aqueous solutions are reasonable well understood and are standardly approached via the Poisson-Boltzmann (PB) theory that combines electrostatics with statistical mechanics of mobile ionic species on a simplified mean-field level. There are several parameters which characterize the physico-chemical behavior of biopolymers:

2.4.2 Parameters characterizing biopolymers

Mean-square end-to-end distance

The distance between the ends of the biopolymer is L. Due to thermal movement there are random fluctuations, thus any measurement always gives an average value. The simplest model for a flexible polymer is the random coil. It consists of N monomers, assumed to be spherical and of diameter b. For this model b is equal to the bond length and the mean-square end-to-end distance

$$\langle L^2 \rangle = b^2 N \tag{2.14}$$

is proportional to the first power of the number of bonds, a characteristic of Gaussian random walk processes. If interactions are long ranged (e.g., by excluded volume or polyelectrolyte repulsion) the chain statistics are no longer Gaussian and

$$\langle L^2 \rangle = b^2 N^{1+\eta} \tag{2.15}$$

where the parameter η is about 0.1 for long, double-stranded DNA. A more realistic model is the wormlike chain, an intermediate between the rigid rod and the random coil, thus taking into account the local stiffness and long-range flexibility of the double helix. An important parameter of a wormlike chain is the persistence length:

Persistance length

Apart from being a charged polymer, DNA is also molecularly rather stiff. DNA in solution is continually changing conformation due to thermal vibration and collisions with water molecules, which makes classical measure of rigidity impossible. Hence the bending stiffness of DNA is measured by the persistence length which is defined (IUPAC):

> The average projection of the end-to-end vector on the tangent to the chain contour at a chain end in the limit of infinite chain length

Or more intuitive:

> The length of an oligonucleotide over which the time-averaged orientation of the polymer becomes uncorrelated by a factor of e, i.e. the length over which correlations in the direction of the tangent are lost.

It can be shown that

$$\langle \cos \Theta \rangle = \exp \left\{ - \left(\frac{L}{p} \right) \right\} \tag{2.16}$$

falls off exponentially with Θ being the angle between a vector that is tangent to the polymer at position 0 and a tangent vector at a distance L away from 0. p is the persistance length. A piece of cooked spaghetti has a persistence length of about 10 cm, dsDNA has a persistance length of about 50 nm (150 bp), depending on the sequence. The variation is largely due to base stacking energies and the residues which extend into the minor and major grooves. DNA differs from many synthetic polymers in that the persistence length is ≈ 25 times larger than the double-helix diameter. Using an intercalating dye is reported to increase the persistence length, *e.g.* [171] found a factor of 1.75 for TOTO-1 concentration of 0.5×10^{-7} M. Under this conditions the natural length of the DNA increases from 16.3 μm to 22 μm (length: ~ 49 kbp). There are still controversies on the persistence length for ssDNA. A good overview about literature data is found in [35]. They suggest a persistent length for ssDNA of 0.75 nm.

Radius of gyration

The radius of gyration is defined:

$$R_g^2 = \frac{1}{N} \sum (\vec{r}_k - \vec{r}_{mean})^2 \tag{2.17}$$

with $\vec{r}_{mean}$ beeing the mean position of the monomers. It can be shown that R_g^2 is proportional to the root mean square distance between the monomers:

$$R_g^2 = \frac{1}{N} \sum (\vec{r}_i - \vec{r}_j)^2 \tag{2.18}$$

For a random coil, which consists of N monomers with the bond length b the radius of gyration is:

$$\langle R_g^2 \rangle = \frac{1}{6} \langle L^2 \rangle = \frac{b^2 N}{6} \tag{2.19}$$

The radius of gyration describes the dimension of a biopolymer and is determined experimentally with static light scattering, small angle neutron and x-ray scattering. The radius of gyration for a 5,9 kbp linear DNA fragment is about $0,213$ μm compared to a length of 2.65 μm [41].

Hydrodynamic Radius

For globular proteins the physical content of the hydrodynamic radius is clear. Using the Stokes-Einstein relation a equivalent can be defined for non-globular polymers with k_b being Boltzmann's constant and the viscosity η:

$$R_H = \frac{k_b T}{6 \pi \eta D} \tag{2.20}$$

with the diffusion constant D. For linear polymers the ratio of the radii R_H and R_g is predicted by the Zimm model [230]

$$\frac{R_g}{R_H} = \frac{8}{3\sqrt{\pi}} \tag{2.21}$$

Particle displacement due to thermal energy

To get a rough estimation about the square mean particle displacement $\langle x^2 \rangle$ of DNA within the time interval t one uses the Einstein-Smoluchowski Formula in one dimension:

$$\langle x^2 \rangle = 2Dt \tag{2.22}$$

Literature data of diffusion coefficients for several types of DNA and different measurement methods are shown in Figure 4.1 p. 72. The big influence of hybridization kinetics from surface treatment and blocking procedure suggests that there is a 2-dimensional diffusion of the DNA along the surface of the Microarray analogous to receptor ligand kinetics at cell surfaces [24]. In two dimension the Einstein-Smoluchowski equation becomes:

$$\langle x^2 \rangle = 4Dt \tag{2.23}$$

and in 3D:

$$\langle x^2 \rangle = 6Dt \tag{2.24}$$

In nucleic acid research, much interest has been focused on diffusion-controlled reactions by the finding that the rate of combination of gene regulatory proteins (such as repressor and polymerase) with their target base sequences is faster than can be accounted for using the above rate constants. This finding has led to the view that the kinetics of events such as repressor-promoter interactions are governed by 3D nonspecific binding to the DNA followed by effectively 1D sliding along the double helix to the target site. A quantitative description of the influence of diffusion on kinetics can be found in Chapter 4.

2.5 Physical chemistry of hybridization

In this section an overview on the relevant physical chemistry of real-time hybridization experiments is given. The main goal of the experiments and calculations presented is to achieve a quantitative description of the transformation from nucleic acid concentrations to digital numbers. Thus it is important to know the factors which are influencing the signal to background ratio on hybridization to solid phase bound oligonucleotides. Literature reports an influence of surface immobilization on hybridization kinetics and thermodynamics [215, 65, 207, 154, 79]. Thus it is necessary for the interpretation of the results to understand the immobilization of the oligonucleotides to the surfaces, especially for the comparison of short and long oligo results [80, 142]. The molecular weight of the nucleic acids was approximated using [188]

$$MW_{Cy3} = 765.95 \tag{2.25}$$

$$MW_{ssRNA} \simeq 320.5\#_{nucleotides} + 159.0 \tag{2.26}$$

$$MW_{ssDNA} \simeq 303.7\#_{nucleotides} + 79.0 \tag{2.27}$$

$$MW_{dsDNA} \simeq 320.5\#_{nucleotides} + 157.9 \tag{2.28}$$

Addition of 159.0 takes into account the MW of a $5'$ triphosphate, 79.0 takes into account the $5'$ monophosphate left by most restriction enzymes. No phosphate is present at the $5'$ end of strands made by primer extension.

2.5.1 Immobilization chemistry on aldehyde surfaces

There are several methods to bind nucleic acid molecules to microarray surface. Within this work mainly aldehyde modified surfaces have been used. An aldehyde is an organic compound containing a

terminal carbonyl group. The word aldehyde is arisen from **al**cohol **dehyd**rogenated. An oxygen atom is double bonded to the carbon at the end of the hydrocarbon chain, generally RCHO. Aldehydes react with amines to form Schiff bases. Aldehyde surfaces form a covaltent binding to the amino groups of nucleic acids via the formation of Schiff bases. A Schiff base (or azomethine) is a functional group that contains a carbon-nitrogen double bond with the nitrogen atom connected to an aryl ar alkyl group, but not the hydrogen. Spotting on aldehyde slides is done with primary amine free buffers. During the drying step, a Schiff base is formed. Though acid labile, Schiff bases can be converted to stable secondary amines (C-N) by reduction, e.g. with sodium borohydride $(NaBH_4)$ (Figure 2.22). Schiff bases are of the general formula $R_1R_2C = N - R_3$ where R_3 is an aryl or alkyl group that make the Schiff base a stable imine. The resulting imine bound overcomes even multiple boiling steps. For immobilized synthetic short oligos a terminal amino modification is often used (3′, 5′ or internal). While the short oligos had an 5′-amino C6 modification the long oligos were unmodified. Unmodified long oligos bind via the amino groups of the bases A,G and C, which gives a binding in a sterical non oriented way. This restricts the accessibility of this bases for the hybridization with the complementary strand. In contrast, good performance of long oligo microarrays with unmodified oligos is reported [211]. The behavior of unmodified long oligo probes was found to be similar to that of amino modified in this study. Thus the requirement for expensive amino modification is in question for real-time applications. Schiff bases are readily hydrolyzed at acid pHs. If reduction is carried out with sodium borohydride, along with Schiff's bases, remaining aldehyde groups will be reduced to primary alcohols $(RCHO \rightarrow RCH_2OH)$. This gives a blocking of the remaining sites and prevents the adsorption of bases during hybridization (Figure 2.22). An immobilization of ds

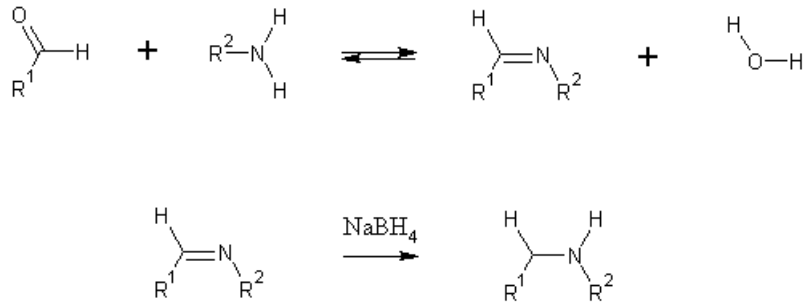


Figure 2.22: Reductive amination scheme: reaction of primary amines with aldehydes to form Schiff base (upper) followed by sodium borohydride reduction (lower). R^1 is the chemistry to the surface, R^2 is either the amino modification or an amino group of the bases.

cDNA with this chemistry was also done but the results varied with the post-processing and the delay between spotting and hybridization. In general the best immobilization chemistry was achieved with 70 mer oligos.

2.5.2 Immobilization of DNA on bare glass

It is possible to immobilize nucleic acid molecules on bare glass substrates using UV crosslinking [176]. While the mechanism is not fully understood experiments suggests an increased immobilization efficiency when using a poly(T)10-Poly(C)10 tagged DNA probe [73]. In general, UV crosslinking seems to increases signal strength but unspecific hybridization as well. In our experiments some probes showed very good results in melting analysis while a lot of spots failed completely.

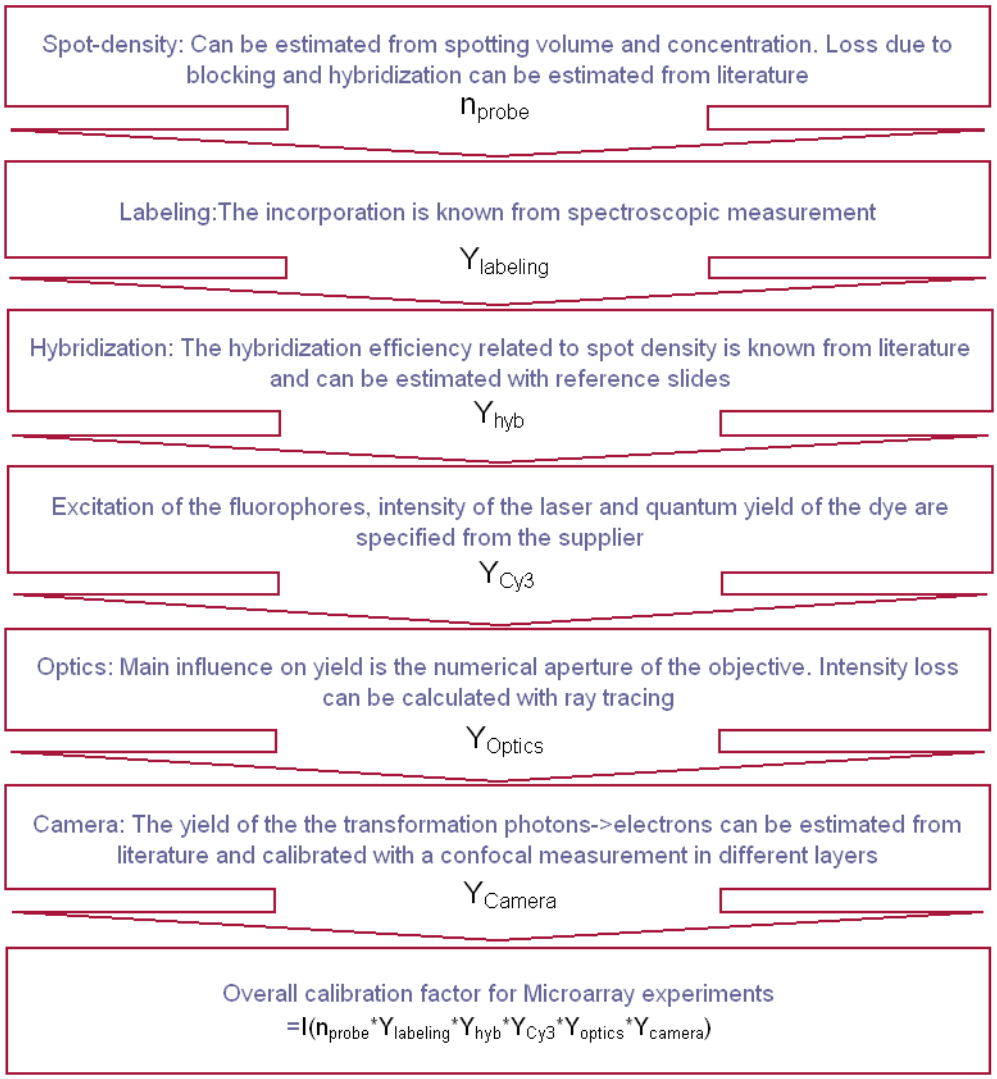


Figure 2.23: Factors influencing the intensity of microarray experiments. While the yield of the optical and electronic components is quite high there is room for improvement in the biochemical parts of the system

2.6 Physics of online hybridization measurement

Based on these experimental and theoretical results an estimation of the upside-potentials of the sensitivity of surface based microarrays is possible. All the calculations within this section are based on results achieved with a short oligo system for methanotroph bacteria and a newly developed actin 'X-chip' for short and long oligo probes. The comparison of the different systems provides valuable insight in the physical chemistry of hybridization. Figure 2.23 presents an overview of the intensity determining factors of a typical microarray experiment.

Table 2.3: Calculated number and density of probe molecules on a typical spot. Experimental data are from [207]. Similar values were reported in [187]: $1.4 - 5 \times 10^{11}$ $1/cm^2$ for $5 - 10$ mM oligos and [208]: 6.0×10^{11} $1/cm^2$

Probe-Conc. μmol	**Probe molecules** $1/spot$	**Probe Density** $1/cm^2$		**Mean distances** nm	
		Theory	*Exp.*	*Theory*	*Exp.*
50	2.1×10^{10}	2.7×10^{14}	1.6×10^{12}	0.61	7.9
10	4.2×10^{9}	5.3×10^{13}	2.1×10^{11}	1.4	21.8
2	8.4×10^{8}	1.1×10^{13}	–	3.1	–

2.6.1 Probe density on microarrays

The amount of spotted nucleic acids can be calculated from the concentration of the oligos and the volume of hybridization buffer.

$$n_{probe} = C_{oligo} \times V_{spotting} \tag{2.29}$$

The spotting volume and the spot diameter of the used system is 0.7 nl and 100 μm. This equals

$$50 \times 10^{-6} \times 0.7 \times 10^{-9}\ mol = 3.5 \times 10^{-14}\ mol \equiv 2.1 \times 10^{10}\ molecules \tag{2.30}$$

thus within our oligo systems there are theoretically about 2×10^{10} molecules on a spot after spotting. It is important to know the average distance between nucleotides, which is calculated simply

$$d_{nucleotide} = \left(\frac{n_{nucleotides}}{A_{spot}} \right)^{-1/2} \tag{2.31}$$

A_{spot} is the area of the probe spot. Comparing the densities with the average length of the oligonucleotides gives a hint on interaction and steric hindrance of the oligos. Values for the number of oligos on the probe spot and the density of probe molecules for different concentrations are found in Table 2.3. Assuming a monomer size of $b \approx 0.6$ nm and 28 bp length a simple calculation for the mean-square end-to-end distance gives

$$\sqrt{\langle L^2 \rangle} = b\sqrt{N} = 3.17\ nm \tag{2.32}$$

while the radius of gyration is

$$\sqrt{\langle R_g^2 \rangle} = \sqrt{\frac{b^2 N}{6}} = 1.3\ nm \tag{2.33}$$

assuming a freely moving random coil in three dimension. Comparing this values with Table 2.3 indicates a decrease in hybridization efficiency due to probe interaction. There is a loss of probe molecules during blocking. From the known spotting concentrations and the measured concentrations on the surface in [207, 165, 150] it seems that only about 1 % of the probe oligos are bound covalently. The experimental data are also shown in Table 2.3.

2.6.2 Target and dye concentration

Measurement of the target concentration and the dye incorporation is done on the Nanodrop absorption spectrometer. Typical results are shown in Tab 2.4. It is important to note that within our short oligo system with target amplification the number of target molecules in solution is about $10...100$ times the number of probe molecules on a single spot. In microarray systems with no amplified target (e.g. gene expression arrays) the number of target molecules can be within the range of the

Table 2.4: Number and density of target and dye molecules within the hybridization solution. The numbers are based on RNA strands of 800 bp lengths

Target	**RNA-Conc.**		**Cy3-Conc.**	**Ratio**	**No of Molecules**	
	$ng/\mu l$	$pmol/\mu l$	$pmol/\mu l$		$RNA/\mu l$	$Cy3/\mu l$
5hm22	42.0	0.23	3.1	41.0	1.38×10^{11}	1.86×10^{12}
rpb46	61.6	0.34	3.6	51.7	2.02×10^{11}	2.16×10^{12}
or2	61.2	0.34	4.2	44.0	18.6×10^{11}	2.52×10^{12}

number of probe molecules. In this case there is a serious depletion of the target concentration during hybridization which can influence the results. Another important factor for the intensity is the ratio between unlabeled and labeled nucleotides, the incorporation rate. Using Klenow labeling incorporation rates of about $40...50$ for DNA targets were achieved. In general the incorporation rate is between $30...100$. In case of end-labeled PCR-primers there is one dye molecule per target strand, which makes this method well suited for reference purposes.

2.6.3 Hybridization efficiency

Some results on hybridization efficiency from SPR-based measurement are shown in Figure 2.24 (from [165]). For low probe densities the efficiency is close to 1, every probe is hybridized by a target molecule. Based on this results an estimated hybridization efficiency of about 0.1 for a probe density of $3 \times 10^{12}\ 1/cm^2$ was used.

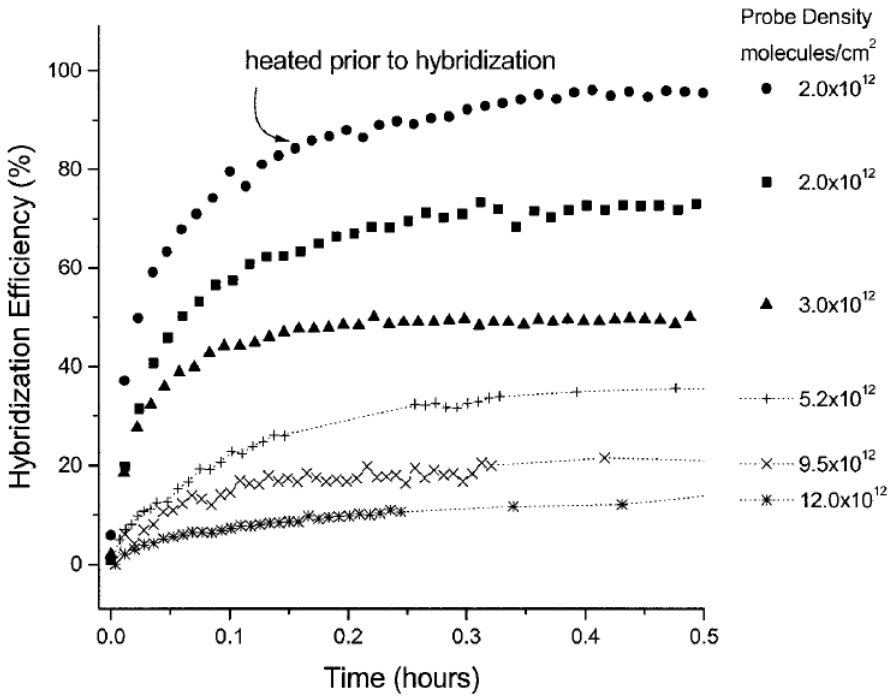


Figure 2.24: Target hybridization kinetics as a function of probe density. Data are from SPR-experiments [165].

Table 2.5: Absorption for different Cy3 concentrations. It can be seen that even for the highest concentration only 0.58 percent of the incoming light is absorbed.

Cy3-Conc. $1/cm^2$	**Absorption** $1/Mol$	I_0/I $-$	$I_0/I-1$ ppm
1×10^{12}	2.50×10^{-5}	1.000058	57.56
1×10^{13}	2.50×10^{-4}	1.000576	575.8
1×10^{14}	2.50×10^{-3}	1.005773	5773

2.6.4 Absorption of the incoming photons

To calculate the number of photons emitted from the laser, the equation

$$E = \hbar n\nu = P_{Laser}t \tag{2.34}$$

is used. With a laser frequency of $5.64\times10^{14}\ Hz$ (wavelength $532\ nm$) and a power of $10\ mW = 0.01W$ ($\hbar = 6.6\times10^{-34}\ Js$) the number of incoming photons is:

$$n = \frac{P_{Laser}}{\hbar\nu} = \frac{0.02}{6.6\times10^{-34}5.64\times10^{14}} = 2.68\times10^{16}\ photons/s \tag{2.35}$$

The power of the laser is spread over the area of the slide, which is $7.5\times2.5 = 18.75\ cm^2$ Thus for a scan duration of 4 minutes on a Tecan LS scanneer the density of photons is about

$$\frac{2.68\times10^{16}\times60\times4}{18.75} = 3.43\times10^{17}\ photons/cm^2 \tag{2.36}$$

A potential source of error in this calculation is saturation. If the rate of excitation f_{ex} is greater than the reciprocal of the lifetime of the excited state τ_F

$$f_{ex} \geq 1/\tau_F \tag{2.37}$$

then the linear relationship between excitation energy and emitted photons is not valid anymore. Typical values for the lifetime of Cy3 and Cy5 are $0.3\ ns$ and $1\ ns$. With this small lifetime and absorption rates (see below) saturation is not expected. The absorption A is calculated from the Beer-Lambert law with the molar extinction coefficient $\epsilon(\lambda_x) = 150000\ 1/(Mol\ cm)$ and the surface concentration C_s:

$$A = \log\frac{I_0}{I} = \epsilon(\lambda_x)\ C_s * 100 \tag{2.38}$$

for $\epsilon(\lambda_x)\ C\ L \ll 1$. The factor 100 comes from the different concentration units.

Calculated values of the absorption are found in Table 2.5. For a typical microarray probe concentrations of $2\times10^{12}\ 1/cm^2$ there is only an absorption of about 0.05% of the incoming photons; most of the laser power is wasted.

2.6.5 Emission intensity of the fluorophores

To get the number of emitted photons this result has to be multiplied with the quantum yield of the dye

$$n_{emitted} = 2.3\times10^8\times0.15 = 3.5\times10^7 photons/s \tag{2.39}$$

The fluorescence intensity $F(\lambda_x,\lambda_m)$ at a particular excitation (λ_x) and emission wavelength λ_m will depend on the absorption $I_A(\lambda_x)$ and the quantum yield $\phi(\lambda_m)$:

$$F(\lambda_x,\lambda_m) = I_A(\lambda_x)\phi(\lambda_m) \tag{2.40}$$

For this labeling a quantum yield of ≥ 0.15 is specified. For our calculation we are assuming isotrop emission of fluorophores.

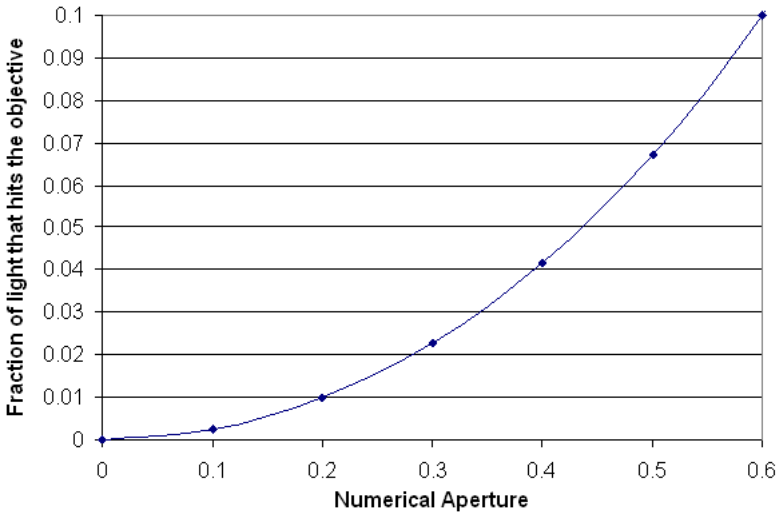


Figure 2.25: Intensity of an objective as a function of the numerical aperture

2.6.6 Influence of photobleaching

For all experiments the influence of photobleaching was checked by a series of 10 image acquisitions. For the HybLive system no significant influence was visible.

2.6.7 Influence of the optics

The most important parameter of the objective is the numerical aperture .

$$NA = n \sin \theta \tag{2.41}$$

where n is the index of refraction (1.0 for air) and θ is the half-angle of the maximum cone of light that can enter or exit the lens. The fraction of emitted light that enters the objective is

$$F_{NA} = \frac{2\pi(1 - \cos \arcsin \frac{NA}{n})}{4\pi} = \frac{1 - \cos \arcsin \frac{NA}{n}}{2} \tag{2.42}$$

Figure 2.25 gives a graphical representation of this dependency. The transparency of the filter is $\geq 90\%$, thus the influence of the filter is neglected. For a typical absorption and emmission spectra see Figure 2.26. For the efficiency of the emission filter an overall value of approximately 0.5 was estimated [47] and Figure 2.26.

2.6.8 Confocal online measurement

Measurement with small aperture pinholes or confocal optics increases the signal to background ratio. To evaluate the effect of an annular apertures on the detection of online hybridization the formulas for the point spread function PSF [116, 115] along the optical axis:

$$|h\,(u,0)\,|^2 = \frac{1}{(1-\epsilon^2)^2} \left\{ \frac{\sin\left[\left(1-\epsilon^2\right) u/4\right]}{u/4} \right\}^2 \tag{2.43}$$

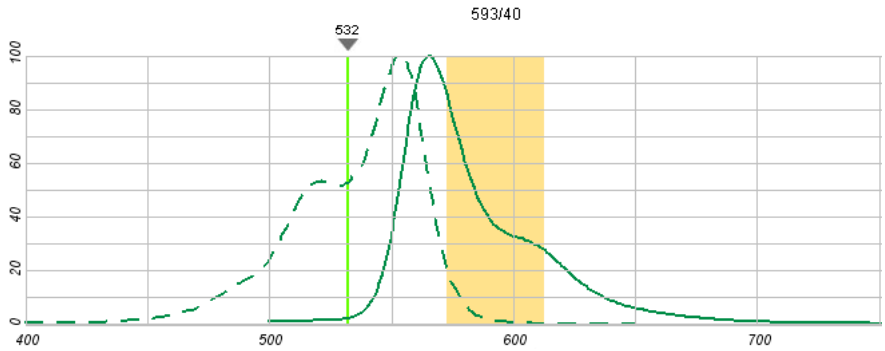


Figure 2.26: Excitation and emission spectra of Cy3 with Laser and filter frequencies

where $J1\,(v)$ is the first-order Bessel function of the first kind, ϵ is the obstruction ratio describing the central fraction of the diameter that is opaque, ranges from 0 to 1 and u and v are generalized optical coordinates:

$$u = nk_0 z \sin^2 \alpha \tag{2.44}$$

$$v = nk_0 \sqrt{(x^2+y^2)} \sin \alpha \tag{2.45}$$

The PSF 2.6.8 in the focal plane is described by:

$$|h\,(0,v)\,|^2 = \frac{1}{(1-\epsilon^2)^2}\left[\frac{2J_1\,(v)}{v} - \epsilon^2 \frac{2J_1\,(\epsilon v)}{\epsilon v}\right]^2 \tag{2.46}$$

for verification a measurement of the fluorescent image on different focal positions was done and the curves were approximated by PSF and exponential curve the results can be found in Figure 2.27. Du to the non-ideal optics the fit with the gaussian function is a better representation of the experimental results.

2.6.9 Efficiency of an EM-CCD camera

The transformation of photons into electrical charge is done in a CCD camera. During this process only a fraction of photons produce photoelectrons. This fraction is given by the quantum efficiency

$$QE = \frac{electrons/sec}{photons/sec} \tag{2.47}$$

Typical values for CCD-sensors without EMV are about $0.4...0.7$. The sensitivity of the Tecan LS Scanner of this specified with $\leq 0.1\ Fluorophoreequivalent/\mu m^2$. Essentially, the EM-CCD is an image sensor that is capable of detecting single-photon events due to the electron multiplying (EM) structure built into the chip. The EM-gain can be increased linearly. The performance of photomultipliers is similar. The overall performance is limited by the Signal to Noise ratio which is not critical compared to other parameters of typical microarray experiments.

2.6.10 Quantitiative description of online measurment

The results are summarized in Table 2.6. Overall there are quite good yields for existing microarray systems. Main potential for improvement is the low absorption of incoming photons. Possibilities for improvement are reflective coated and gel (3d) slides. The low absorption coefficient is known challenge in confocal microscopy as well.

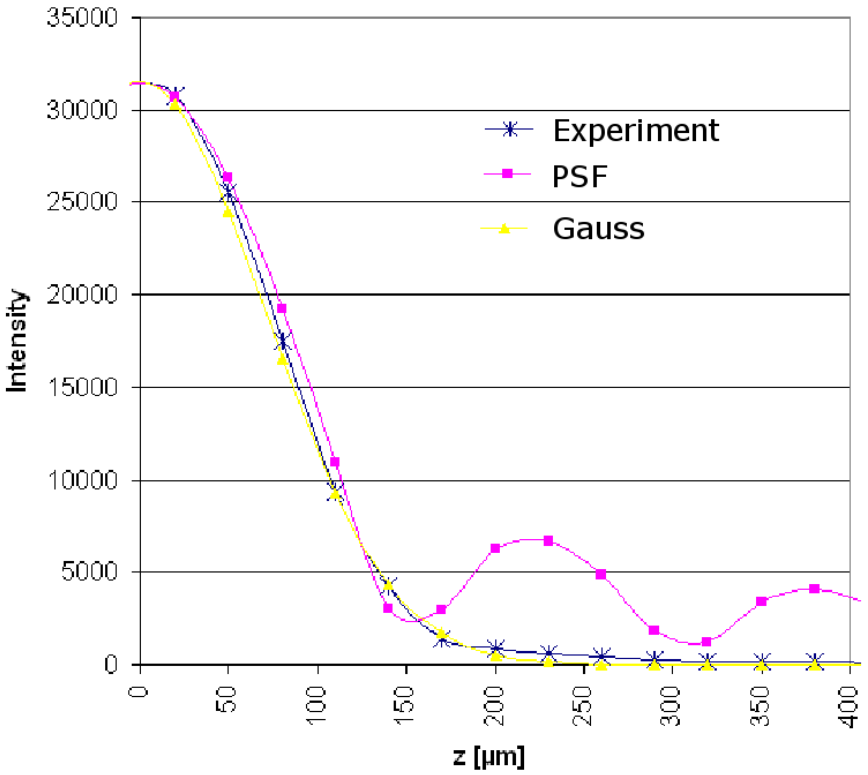


Figure 2.27: Measurement of PSF and comparison with an theoretical fit according to eqn. (2.46) and a gaussian fit.

Table 2.6: Summary of the factors influencing a microarray experiment. The numbers are rough estimation for typical microarray experiments using spotted short oligo probes and amplified targets

Parameter	**Yield-value**	**Potential for improvement**
Immobilization	$\approx 1\%$	process chemistry
Hybridization efficiency	0.1...1	agitation, process chemistry
Absorption of dye	$\ll 1\%$	Reflective coating, gel chips
Quantum yield of Cy3	4%...40%	Dye development, labeling
Excitation filter	≈ 0.5	–
Objective NA (Tecan)	0.6	High NA objectives
Filter	≥ 0.9	–
Camera yield	0.4...0.7	–

2.7 Summary

This chapter provides a short overview on nucleic acid structure and on the technological and physical basics of real-time measurement on microarrays. The structural differences of DNA and RNA are important for the understanding of kinetics and thermodynamics of competitive hybridization. Especially the C5-methyl group in thymine and methylated cytosine has a strong influence on the properties of nucleic acids. It is interesting that the C5 methyl group is one of the two differences between DNA and RNA as well as methylation of the C5-atom of cytosine is the most prominent

effect in epigenetic. I turned out that technological aspects of online-hybridization can be estimated with simple models, which is vital for the extraction of kinetic and thermodynamic parameters from the complex results of real-time data acquisition. The effort necessary for the understanding and optimization of the molecular biological part of biosensors is often underestimated in interdisciplinary projects. Most of the incoming photonic energy is lost due to the low absorption. Thus the potential for improvement of sensitivity and specificity is mainly in the biochemical part while the optical and signal amplification part of the system are already quite optimized.

Chapter 3

Thermodynamics of hybridization

3.1 Introduction

The main structural motif for natural nucleic acids is the double helix with Watson-Crick base pairs. Cellular DNA is almost exclusively in this form. Known RNA structure is more than 50% double helical due to secondary structure [90]. Accurate prediction of DNA thermal denaturation is important for several molecular biological techniques and for the understanding of partial opening of DNA ("breathing"), the key process for biological specificity. A good overview on different methods for the calculation of thermodynamic parameters can be found in [145]. Melting experiments based on UV-absorption measurement have been a useful way to measure the thermodynamic properties of RNA and DNA structures under different conditions. HRM (high resolution melting) analysis is a RT-PCR based technique for the analysis of gene fragments in solution with a very high specificity for hetero- and homo duplexes [85]. For short oligos used approximation for the transition from the duplex helix state to single strands is the two-state model. In this model the nucleic acid strands are assumed to be either double stranded or random coils. In reality single strands do not form unique structures, but rather ensembles of partially stacked helices, where the degree of stacking is strongly dependent on temperature and sequence [106]. Oligonucleotide strand geometry and electrostatic considerations indicate that single strands exist as partially ordered structure rather than random coil. On surfaces melting is not only determined by thermodynamics of the nucleic acids but from several effects related to the immobilized probes:

- Steric hindrance
- Entropic restrictions
- Electrostatic repulsion due to negatively charged backbone
- Surface effects of blocking, crosslinking, adsorption effects
- Influence of spacer, spotting parameters and labeling effects
- Diffusion-transport effects slow down the hybridization reaction

There is an ongoing discussion about the differences between surface bound and solution thermodynamic parameters [86]. In early studies of gel pad microarrays a linear relationship between microarray hybridization and free energy $\Delta G_{\mu array}$ and the corresponding free energy in solution ΔG_{sol} was

found [58]. Recently [216] a similar relationship was observed on self-spotted codelink activated slides. Other studies on Affymetrix Genechips [55] found that hybridization in solution is different from microarray hybridization which occurs in the vicinity of a solid surface and the value of ΔG obtained from hybridization in solution may not adequately describe hybridization to surface bound oligonucleotides. But due to the experimental difficulties there are only few studies dealing with melting analysis on microarrays. The comparison with HRM based results is not straightforward, high salt hybridization buffers are not compatible with intercalating dyes used in qPCR.

3.2 Parameters influencing melting temperature

The melting temperature of an oligonucleotide duplex T_m refers to the temperature at which the oligonucleotide is 50% annealed. Due to the extreme cooperativity of DNA hybridization and melting, this means for short oligo duplexes that 50% of the molecules are single-stranded (ss) while 50% of the molecules are in the double-stranded form. This two-state based simplified model of T_m seems to hold true for most short oligo sequences. It is interesting to note that melting analysis on solid phase immobilized oligos is a two state process because it only indicates the complete denaturation of the target molecules, the partial opening of several bases is not visible within this experimental setups.
There are several factors influencing T_m:

3.2.1 Oligonucleotide length and sequence

The summary of hydrogen bonds and stacking interactions. In general the longer the length of the nucleotide and the higher the content of G•C basepairs, the higher is the melting temperature. Due to the high cooperativity there are several nonlinear effects, e.g. tracts of repetitive basepairs show different behavior than expected from Nearest Neighbor models

3.2.2 Salt concentration

Monovalent cations (e.g. sodium ions) interact electrostatically with nucleic acids (mainly at the phosphate groups) so that the electrostatic repulsion between the two strands of the duplex decreases with increasing salt concentrations, i.e. higher salt concentrations increase the stability of the hybrid. Low sodium concentrations affect the T_m, as well as the renaturation rate, drastically. Sodium ion (Na^+) concentrations above 0.4 M only slightly affect the rate of renaturation and the melting temperature. Free divalent cations strongly stabilize duplex DNA. An overview on the mechanism of dielectric screening is given in 2.3.3, page 25.

3.2.3 Oligonucleotide concentration

Higher concentrations favor duplex formation and increase T_m. Concentration is influencing the entropic contribution of the free energy. Melting experiments with different target concentrations allow a separation of the thermodynamic parameters, a description is given in 3.3.1, page 46.

3.2.4 Denaturating Agents

Formamide has for years been the organic solvent of choice. It reduces the melting temperature of DNA-DNA and DNA-RNA duplexes in a linear fashion by 0.6 to 0.72 ℃ for each percent formamide in solution hybridization. It seems that formamide is forming more stable hydrogen bonds with the partially opened bases than the opposite base [17]. For DNA•DNA duplexes it was found that the

denaturating effect of formamide is 0.65 ℃$/(\%formamide)$. For RNA•DNA duplexes there was no clear-cut relation visible.

3.2.5 Influence of metal ions on nucleic acid hybridization

Metal Ions have major effects on double helix stability and structure, see Figure 3.1. Ions will raise the melting temperature if they bind more strongly to the double-helical form than to the single-stranded form; such binding may be specific or nonspecific. While monovalent cations except for Ag^+ generally appear to operate by the nonspecific polyelectrolyte mechanism the divalent cations show evidence of site-binding [90]. There is a general rule of thumb to calculate the Ionic equivalent of $[Mg^{++}]$:

$$[Na^+]_{equivalent} = [MonovalentCations] + 120 \times \sqrt{[Mg^{++}] - [dNTP]} \tag{3.1}$$

A higher concentration of Mg^{++} has a big influence on the shape of the melting curve (Figure 3.1). This appears to be due to the increased effects of counterion condensation and residual ion atmosphere screening for divalent relative to univalent cations, and to site binding of $[MG^{++}]$ ions to phosphates in the double-stranded form. The broadening of the transition at half saturation, $r = 0.5$ seems to be related to site binding. From Figure 3.2 it can be seen that the influence of the other

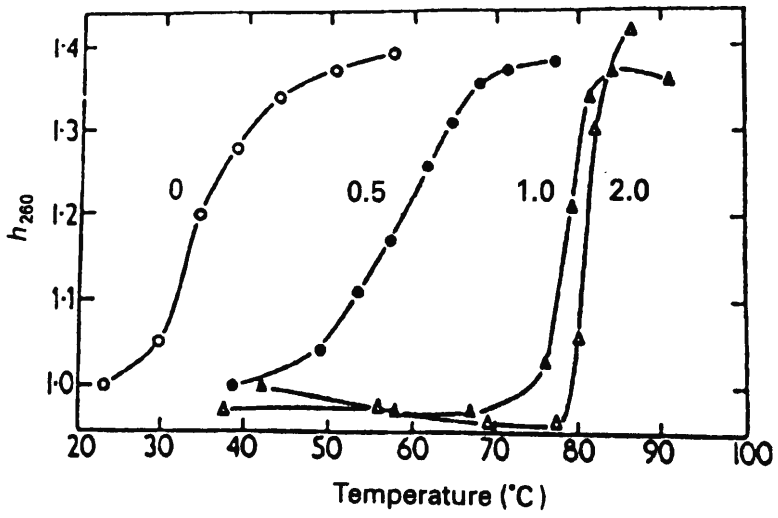


Figure 3.1: Influence of Mg^{2+}-concentration on thermal denaturation of *Bacillus megaterium* DNA, at pH 7 and ionic strength 3×10^{-4} (from [90]). Curves are labeled with the number of equivalents of Mg^{2+} per equivalent of DNA phophate.

divalent metal ions on the melting temperature is much more complicated. For the calculation of T_m with simple formulas see 3.2.6. Variations of the metal ion content provide a tool for the shifting and an enhancement of the melting transitions. The use of hybridization solutions with different content of metal ions can also improve the understanding of nucleic acid hybridization.

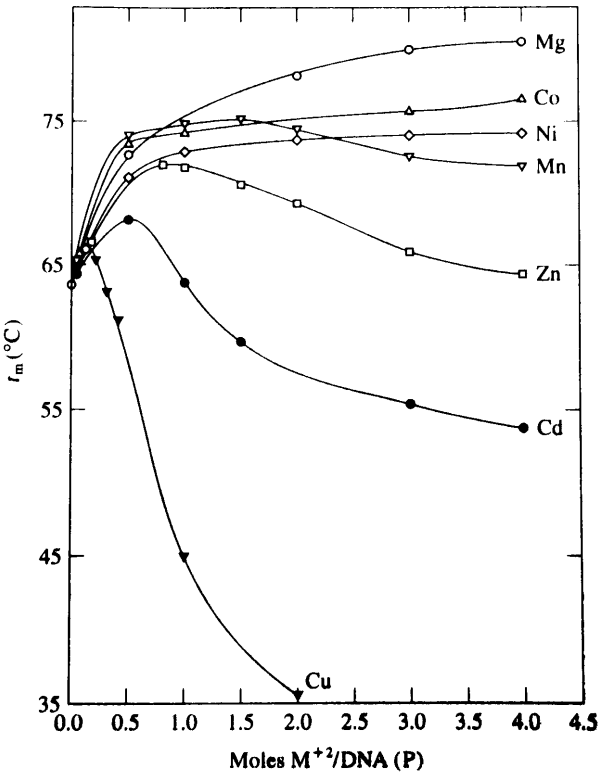


Figure 3.2: Variations of T_m for solutions of DNA as a function of divalent metal ion concentration. From [90].

3.2.6 Simple formulas for T_m

For polynucleotides there are several methods for an estimation of the melting temperature. Basic T_m calculations were performed according to [128]:

$$T_m = 100.5 + 4.0 \times \frac{yG + zC - 16.4}{wA + xT + yG + zC} - \frac{820}{wA + xT + yG + zC} + 16.6 \log[Na^+] \tag{3.2}$$

where x,y,w and z are the number of the bases of T, G, A and C. The influence of formamide on melting temperature of $DNA \bullet DNA$ duplexes can be calculated according to 3.2.4. More accurate formulas, which takes into account different types of nucleic acids are [219]:

- **DNA duplexes:**

$$T_m^{DNA} = 81.5 + 16.6 \log_{10} \left\{ \frac{[Na^+]}{1.0 + 0.7[Na^+]} \right\} + 0.41\,(\%_{G+C}) - \frac{500}{D} - P \tag{3.3}$$

  with D being the number of base pairs and P the percent mismatching. Similar formulas exist for:

- **RNA duplexes:**

$$T_m^{RNA} = 78 + 16.6 \log_{10} \left\{ \frac{[Na^+]}{1.0 + 0.7[Na^+]} \right\} + 0.70\,(\%_{G+C}) - \frac{500}{D} - P \tag{3.4}$$

and for

- **RNA•DNA hybrids:**

$$T_m^{hybrid} = 67 + 16.6 \log_{10} \left\{ \frac{[Na^+]}{1.0 + 0.7[Na^+]} \right\} + 0.80 \, (\%_{G+C}) - \frac{500}{D} - P \qquad (3.5)$$

These simple equations do not describe the melting temperature sufficient. It does not give a quantitative description of the asymmetry of hybrids, DNA•RNA hybrids change their thermodynamic behavior when the sequence of the DNA and the RNA strand are exchanged. A correlation between the results of this simple algorithm with the results of [231] can be found in Figure 3.3. While the correlation between the two methods is sufficient, the correlation to experimental values was for both methods: $R^2 < 0.7$. There are several different parameters sets, some widely used can be found in

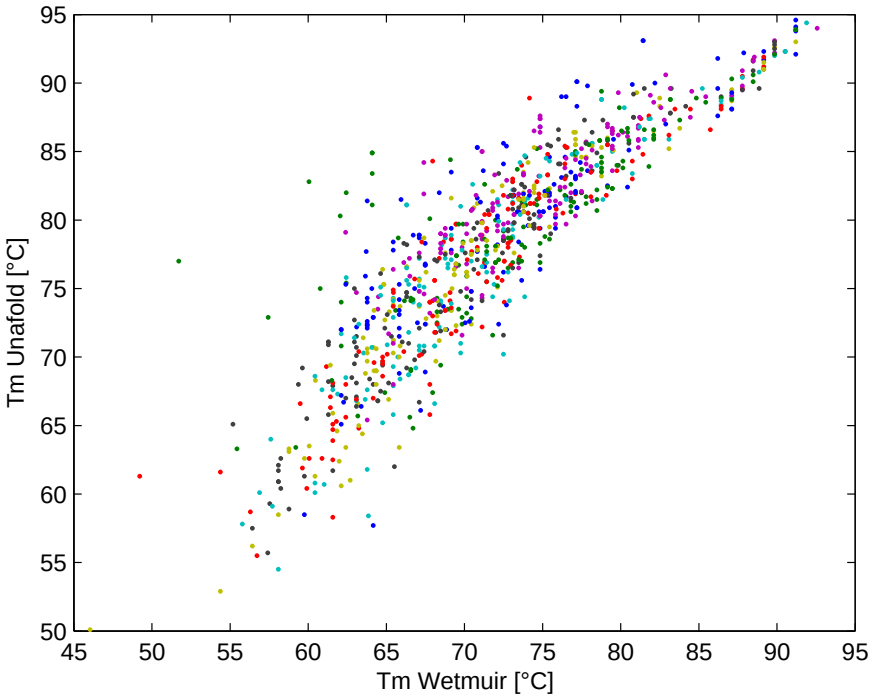


Figure 3.3: Correlation of melting temperature for the Actin X-chip and the long oligo targets calculated with two different methods described in the text [127, 219]. The different colors correspond to different Actin targets. The correlation coefficient (R^2) is 0.80.

[186, 184, 185, 174]. It is interesting to note that there is no universal set of NN-parameters and that there is still no agreed model for the calculation of the thermodynamic properties of oligonucleotides.

3.3 Thermodynamic of Hybridization

3.3.1 Theory of the two-state model

For a non-selfcomplementary mixture of two Nucleic Acid strands with the initial concentrations $[A_{ss}]_0$, $[B_{ss}]_0$

$$A_{ss} \; + \; B_{ss} \; \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \; (A \bullet B)_{ds} \tag{3.6}$$

$$[A_{ss}]_0 = [B_s ss]_0 = C_T/2 \tag{3.7}$$

the reaction constant K is [90]

$$K = \frac{[AB]}{[A_{ss}]\,[B_{ss}]} = \frac{2\alpha}{(1-\alpha)^2\, C_T} \tag{3.8}$$

Here α is the fraction of the total strand concentration C_T that is in duplex. Definition of the melting Temperature T_m where 50% of the targets are in duplex gives $\alpha = 0.5$. For non-selfcomplementary targets this reduces 3.8 to

$$(K)_{T_m} = 4/C_T \tag{3.9}$$

i.e. the concentration of the duplexes is 0.25 times the concentration of each single-strand nucleic acid. Substituting these results into the equation that relates the free energy ΔG with the reaction constant K and the entropy ΔS^0 and enthalpy ΔH^0

$$\Delta G^0 = -RT \ln K = \Delta H^0 - T\Delta S^0 \tag{3.10}$$

and rearranging lead to

$$\frac{1}{T_m} = \frac{R \ln (C_T/4)}{\Delta H^0} + \frac{\Delta S^0}{\Delta H^0} \tag{3.11}$$

Thus a plot $1/T_m$ versus $\ln (C_T)$ should be linear, and enthalpy and entropy can be determined from slope and intercept. Using this equation one gets for the melting temperature in ℃ [126]:

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R \ln (C_T/f)} - 273.15 \tag{3.12}$$

with $f = 4$ for non-selfcomplementary bimolecular targets. 3.11 holds if the concentrations of the two non-selfcomplementary strands are equal. In many applications one strand is in large excess, *i.e.* the number of probemolecules on a spot compared to the targets in the solution $[B]_0 \gg [A]_0$. In this case, when T_m is defined as the temperature where half of the less concentrated sequence is bound

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R \ln (C_B - 0.5 C_A)} - 273.15 \tag{3.13}$$

For more than two target

$$[A_{ss}]_0 = [B_{ss}]_0 = [C_{ss}]_0 = C_T/2 \tag{3.14}$$

strands similar transformations can be done [90] and 3.11 becomes in general for a mixture of N target molecules of the same concentration is

$$\frac{1}{T_m} = \frac{(N-1)\,R}{\Delta H^0} \ln C_T + \frac{\left[\Delta S^0 - (N-1)\,R \ln 2N\right]}{\Delta H^0} \tag{3.15}$$

3.3.2 Temperature dependence of thermodynamic properties

Different methods for the solution based measurement of enthalpy changes in Nucleic Acid hybridization gives significant differences. The possible interpretation is that the helix-to-coil melting is a non-two-state transition and the difference in hydration between the duplex-stranded groups and single-stranded groups results in a heat capacity increase [222]. It should be noted that the two-state character of hybridization decreases with oligo length. The change in standard enthalpy change and equilibrium constant is described in the van't Hoff equation

$$\frac{d\ln K}{dT} = \frac{\Delta H}{RT^2} \tag{3.16}$$

There are several methods for the extraction of the ΔH and ΔS. One is to plot the reciprocal of the melting temperature T_m^{-1}, vs. $\ln(C_t/4)$ using eqn. 3.16

$$\frac{1}{T_m} = \frac{R}{\Delta H}\ln\frac{C_T}{4} + \frac{\Delta S}{\Delta H} \tag{3.17}$$

C_T is the total species concentration and R is the gas constant $1.987\ calK^{-1}mol^{-1}$. Another method is to fit the shape of the melting curves by using nonlinear least-squares program, or use a physical modeling, see chapter 4. The enthalpy and entropy changes as a function of temperature are given by [222]:

$$\begin{aligned}\Delta H(T_m) &= \Delta H^0 + \int_{T^0}^{T} \Delta C_{p,H} dT \\ &= \Delta H^0 + \Delta C_{p,H}(T_m - T^0) \end{aligned} \tag{3.18}$$

$$\begin{aligned}\Delta S(T_m) &= \Delta S^0 + \int_{T^0}^{T} \Delta C_{p,S} d\ln T \\ &= \Delta S^0 + \Delta C_{p,H}\ln(T_m/T^0) \end{aligned} \tag{3.19}$$

where $\Delta H(T_m)$ and $\Delta S(T_m)$ are the enthalpy and entropy changes at the temperature of interest, ΔH^0 and ΔS^0 are the enthalpy and entropy changes in the reference state, T^0 is the reference temperature, $\Delta C_{p,H}$ is the heat capacity change from linear regression of the enthalpy change with respect to the melting temperature ($\Delta C_{p,H} = d\Delta H/dT_m$), and $\Delta C_{p,S}$ is the heat capacity change in entropy derived from a linear regression of the entropy change with respect to the logarithmic scale of the melting temperature ($\Delta C_{p,S} = d\Delta S/d\ln T_m$). While in principle the two methods should give equivalent results experiments showed significant differences [167]. It is known from investigation on proteins that the change in heat capacity is the net sum of the positive contribution from the exposure of nonpolar groups and the negative contribution from exposure of polar groups. When the double strand is melting into the coiled single strands, the difference in hydration between the different strands results in an increase of the heat capacity. The free energy change and melting Temperature T_m are often used to characterize the stability of base pairing. In contrast to the clear temperature-dependence of the enthalpy and entropy changes, they are relatively insensitive to the heat capacity change. This suggests that the free energy change determined by $\Delta C_p = 0$ would be a more accurate parameter than either the individual enthalpy change or entropy change.

Critical remarks on the two state model

Using 3.11 for extraction of the enthalpy H^0 and the entropy S^0 is a questionable method because the T_m for different concentrations is less than 10 °C. The intercept with the y-axes is at infinite

temperature. This is a known problem of Arrhenius-like plots [32]. The analysis is based on the relation between free energy and reaction constant:

$$\Delta G^0 = -RT \ln K = \Delta H^0 - T\Delta S^0 \tag{3.20}$$

Hybridization and dissociation of DNA is assumed to be a two state process in this model, in reality this process is more complex. Experimental methods like UV-spectroscopy or HRM Analysis show shoulders in the melting curves which indicate melting domains. The signal is assumed to approach an asymptotic value, presumably as a result of the finite number of probe sites available. The linear dependence of signal strength on c is implicitly assumed in most statistical models. Typical values for the densities of probe molecules can be found in Tab 2.3, page 33 and Tab 2.4, page 33.

With a typical target concentration of $100\ nM$ the number of ss nucleic acid molecules is about $6 \times 10^{10}\ 1/\mu l$. According to [84] this is within the range of the number of probes on a typical spot. Saturation of the intensity was observed at high target concentrations of ≥ 1 nM. Literature

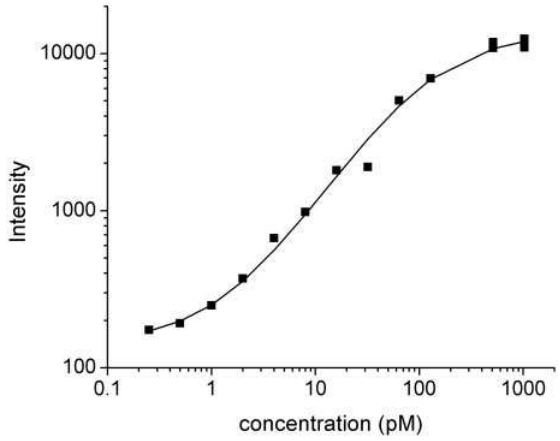


Figure 3.4: Intensity as a function of target concentration from [84].

data suggest for oligo systems with low probe density a efficiency close to 1 whereas at a high probe density the efficiencies drop and are nearly independent of immobilization. In Figure 2.24 results from SPR measurement are shown. For this SPR based experimenta hybridization efficiency of about 0.1 for probe density of $3 \times 10^{12}\ 1/cm^2$ was found.

Assuming that the observed fluorescence intensity scales linearly with the number of hybridized probe•target duplexes, one can model the observed intensity by considering the binding and unbinding reactions

$$probe_{ss} + target_{ss} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} duplex, \tag{3.21}$$

where k_a and k_d are the respective association and dissociation rate constants for the reaction, and the concentrations Equilibrium thermodynamics gives a relation between the observed intensity and the free energy ΔG:

$$Intensity \propto exp\frac{\Delta G}{RT}. \tag{3.22}$$

From experiments it is known that the dependence of ΔG from temperature is much weaker than predicted from equilibrium thermodynamics [84, 82].

3.3.3 Solution based results

Most of the early experiments on thermodynamics of nucleic acids were based on UV-spectroscopy. The development of intercalating dyes and RT-PCR based HRM technologies opened the door for parallel investigation on larger probe-sets [220]. Compared to online microarray experiments the buffers are very different. Intercalating dyes are not working in the high salt hybridization buffer (e.g. 6xSSC). For our application we used either commercial PCR buffer or 0.5x TBE buffer. A typical value for the target concentration within our experiments is 200 nM. 3.5 gives a typical result.

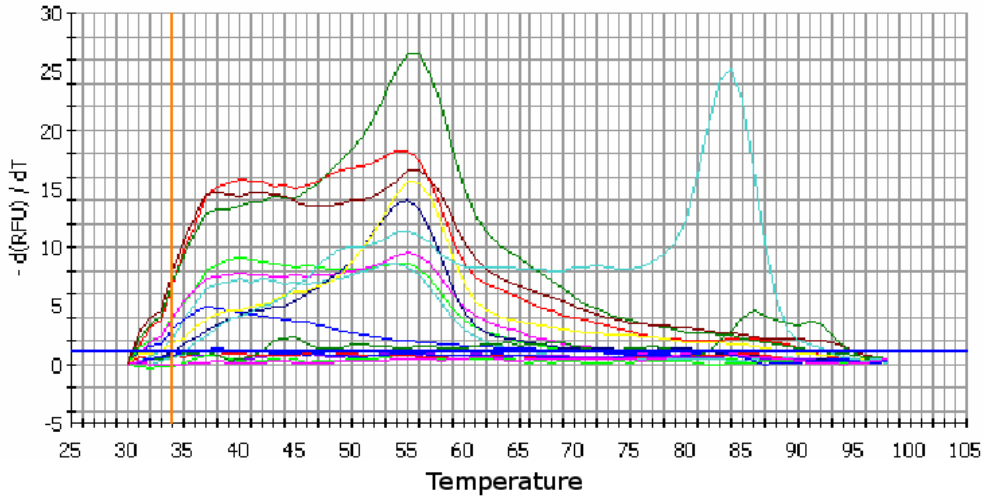


Figure 3.5: Melting curves for different oligo lengths. The peak at 55℃ corresponds to melting of the PCR-primer to the Actin fragments, the peak at 85℃ is the denaturation of the actin fragment (795bp).

3.4 The Nearest Neighbor model

3.4.1 Parameters of the NN model

For short double-strand DNA duplex the relative stability depends primarily on the identity of the nearest neighbor bases [173, 219]. The validity of this approach is a length of approximately 50 bp. The sequence dependent stability is considered in terms of basepair doublets. In dsDNA (and dsRNA) there are 10 unique internal nearest-neighbor symbols, in $5' - 3'$ direction:

AT•AT, TA•TA, AA•TT, CA•TG, TC•GA, CT•AG, CG•CG GC•GC, GG•CC

For RNA•DNA duplexes there are 16 parameters not taking into account end effects and initiation terms. A full set of NN parameters should also consider the different behavior of terminal bases. If E denotes the ends, the four end interaction ($5' - 3'$) are

TE•EA, CE•EG, AE•ET, GE•EC

With these interactions, the number of possible NN parameters increases to 14. Some authors postulate the use of initiation (nucleation) parameters. For melting analysis of DNA denaturation these terms are obviously not relevant. Based on the assumption that each base pair has a neighbor

on the left side as well as on the right side there are two mathematical constraints which limit the number of independent parameters

$$\begin{aligned} N_{AA\bullet TT} + N_{AT\bullet AT} + N_{AC\bullet GT} + N_{AG\bullet CT} + N_{AE\bullet ET} &= \\ N_{AA\bullet TT} + N_{TA\bullet TA} + N_{CA\bullet TG} + N_{GA\bullet TC} + N_{EA\bullet TE} & \end{aligned} \tag{3.23}$$

$$\begin{aligned} N_{GA\bullet TC} + N_{GT\bullet AC} + N_{GC\bullet GC} + N_{GG\bullet CC} + N_{GE\bullet EC} &= \\ N_{AG\bullet CT} + N_{TG\bullet CA} + N_{CG\bullet CG} + N_{GG\bullet CC} + N_{EG\bullet CE} & \end{aligned} \tag{3.24}$$

Therefore only $10 - 2 = 8$ independent NN parameters without ends and $14 - 2 = 12$ parameters including end effects are possible. An overview on the NN parameters for DNA•DNA, RNA•DNA and RNA•RNA duplexes is shown in 3.6. The NN parameters from different authors are shown

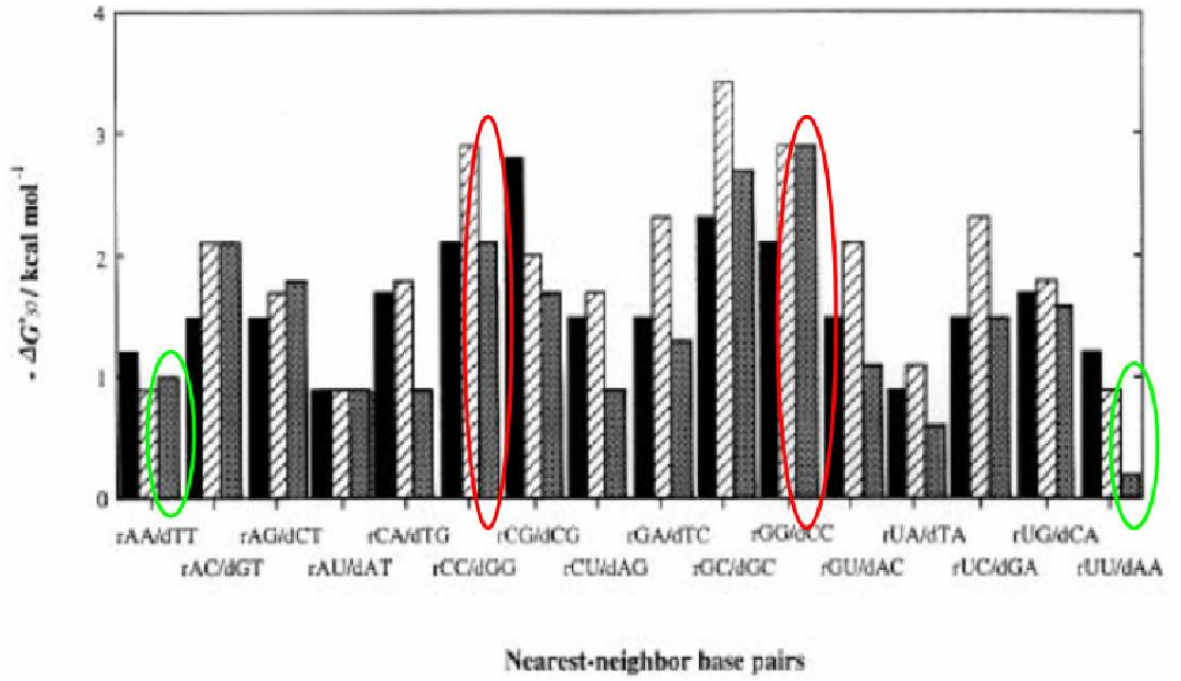


Figure 3.6: Representation of the free energy (ΔG) for DNA•DNA (black), RNA•RNA (stripes) and DNA•RNA duplexes from [184]. Marked are highly asymmetric base pairs in DNA•RNA hybrids.

in Table 3.1. A comparison of enthalpy, entropy and free energy of NN parameters from different authors is found in Table 3.2 In general a description of nucleic acid thermodynamic by means of NN based models is not fully sufficient. Actual research is done on next nearest neighbor models [70] and advanced models like Peyrard-Bishop Hamiltonians [152].

3.4.2 Gibbs free energy

The relative stability and temperature-dependent behavior of each NN interaction can be characterized by Gibbs free energy ΔG^0, enthalpy ΔH^0, and entropy ΔS^0, which are related by the standard thermodynamic relationship

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \tag{3.25}$$

Table 3.1: Comparison of DNA•DNA Free Energy ΔG in $kcal/mol$ from different authors [18, 173, 21, 184, 69, 205]

	Gotoh 1981	**Vologodskii 1984**	**Breslauer1986**	**Blake 1998**	**Benight 1997**	**SantaLucia1996**	**Sugimoto**	**Unified**
AA•TT	-0.43	-0.89	-1.66	-0.67	-0.93	-1.02	-1.2	-1
AT•TA	-0.27	-0.81	-1.19	-0.62	-0.83	-0.73	-0.9	-0.88
TA•AT	-0.22	-0.76	-0.76	-0.7	-0.7	-0.6	-0.9	-0.58
CA•GT	-0.97	-1.37	-1.8	-1.19	-1.26	-1.38	-1.7	-1.45
GT•CA	-0.98	-1.35	-1.13	-1.28	-1.52	-1.43	-1.5	-1.44
CT•GA	-0.83	-1.16	-1.35	-1.17	-1.03	-1.16	-1.5	-1.28
GA•CT	-0.93	-1.25	-1.41	-1.12	-1.56	-1.46	-1.5	-1.3
CG•GC	-1.70	-1.99	-3.28	-1.87	-1.65	-2.09	-2.8	-2.17
GC•CG	-1.64	-1.96	-2.82	-1.85	-2.44	-2.28	-2.3	-2.24
GG•CC	-1.22	-1.64	-2.75	-1.55	-1.67	-1.77	-2.1	-1.84
Average	-0.92	-1.32	-1.82	-1.2	-1.36	-1.39	-1.64	-1.42
Init. term G•C	NA	NA	2.6	NA	NA	0.91	1.7	0.98
Init. term A•T	NA	NA	2.6	NA	NA	1.11	1.7	1.03
[Sodium] M	0.0195	0.195	1	0.075	0.115	1	1	1
Rank	8	8	11	8	9	10	11	12

Gibbs free energy describes the potential of a reaction to occur spontaneously; enthalpy provides the amount of heat released from or absorbed by the system; and entropy measures the randomness or disorder of a system. The corresponding parameters for RNA•DNA duplexes in 1 m NaCl Buffer are shown in Table 3.3. The Gibbs free energy of a double-stranded molecule given by $x = a_1 \dots a_n$, with reverse complementary strand $\bar{a}_n \dots \bar{a}_1$, is calculated as

$$\Delta G^0 = \Delta g_i + \Delta g_s + \sum_{i=1}^{n-1} \Delta G^0 \left(a_i a_{i+1} \bullet \bar{a}_i \bar{a}_{i+1} \right) \tag{3.26}$$

where Δg_i denotes the helix-initiation energy and Δg_s is the symmetry correction. The enthalpy of a double-stranded molecule is similarly computed

$$\Delta H^0 = \Delta h_i + \sum_{i=1}^{n-1} \Delta H^0 \left(a_i a_{i+1} \bullet \bar{a}_i \bar{a}_{i+1} \right) \tag{3.27}$$

where Δh_i denotes the helix initiation enthalpy. The entropy of a short stretch of dsDNA can either be computed by using 3.25.

3.4.3 Comparison of NN results with simple formulas

A correlation of the T_m values calculated with Eqn. 3.3 and with [127] is shown in 3.7. Both methods give similar results for the melting temperature, the correlation coefficients is $R^2 = 0.986$. The gain due to the complicated NN-model is quite small.

Table 3.2: Nearest Neighbor parameters from [173]

NN pair	**Δ H** ($kcal/mol$)			**Δ S** ($cal/K\ mol$)			**ΔG** ($kcal/mol$)		
	Bre68	**SL96**	**Sug1996**	**Bre68**	**SL1996**	**Sug96**	**Bre68**	**SL1996**	**Sug96**
AA•TT	-9.1	-8.4	-8.0	-24.0	-23.6	-21.9	-1.9	-1.02	-1.2
AT•TA	-8.6	-6.5	-5.6	-23.9	-18.8	-15.2	-1.5	-0.73	-0.9
TA•AT	-6.0	-6.3	-6.6	-16.9	-18.5	-18.4	-0.9	-0.60	-0.9
CA•GT	-5.8	-7.4	-8.2	-12.9	-19.3	-21.0	-1.9	-1.38	-1.7
GT•CA	-6.5	-8.6	-9.4	-17.3	-23.0	-25.5	-1.3	-1.43	-1.5
CT•GA	-7.8	-6.1	-6.6	-20.8	-16.1	-16.4	-1.6	-1.16	-1.5
GA•CT	-5.6	-7.7	-8.8	-13.5	-20.3	-23.5	-1.6	-1.46	-1.5
CG•GC	-11.9	-10.1	-11.8	-27.8	-25.5	-29.0	-3.6	-2.09	-2.8
GC•CG	-11.1	-11.1	-10.5	-26.7	-28.4	-26.4	-3.1	-2.28	-2.3
GG•CC	-11.0	-6.7	-10.9	-26.6	-15.6	-28.4	-3.1	-1.77	-2.1
Any G•C pair	0.0	0.0	+0.6	-16.77	-5.9	-9.0	+5.0	+1.8	+3.4
Only A •T pair	0.0	0.0	+0.6	-20.13	-9.0	.9.0	+6.0	+2.8	+3.4
Symmetry	0.0	0.0	0.0	-1.34	-1.4	-1.4	+0.4	+0.4	+0.4
5 ′ -terminal-T•A-3′ bp	0.0	+0.4	0.0	0.0	0.0	0.0	0.0	+0.4	0.0

Table 3.3: Gibbs free energy ΔG^0, enthalpy ΔH^0 and entropy ΔS^0 for RNA•DNA nearest neigbor parameters from [186] in 1 M NaCl Buffer

Sequence	$\Delta H^0/kcalmol^{-1}$	$\Delta S^0/calmol^{-1}K^{-1}$	$\Delta G^0/kcalmol^{-1}$
rAA•dTT	-7.8	-21.9	-1.0
rAC•dTG	-5.9	-12.3	-2.1
rAG•dTC	-9.1	-23.5	-1.8
rAU•dTA	-8.3	-23.9	-0.9
rCA•dGT	-9.0	-26.1	-0.9
rCC•dGG	-9.3	-23.2	-2.1
rCG•dGC	-16.3	-47.1	-1.7
rCU•dGA	-7.0	-19.7	-0.9
rGA•dCT	-5.5	13.5	-1.3
rGC•dCG	-8.0	-17.1	-2.7
rGG•dCC	-12.8	-31.9	-2.9
rGU•dCA	-7.8	-21.6	-1.1
rUA•dAT	-7.8	-23.2	-0.6
rUC•dAG	-8.6	-22.9	-1.5
rUG•dAC	-10.4	-28.4	-1.6
rUU•dAA	-11.5	-36.4	-0.2
Initiation	1.9	-3.9	3.1

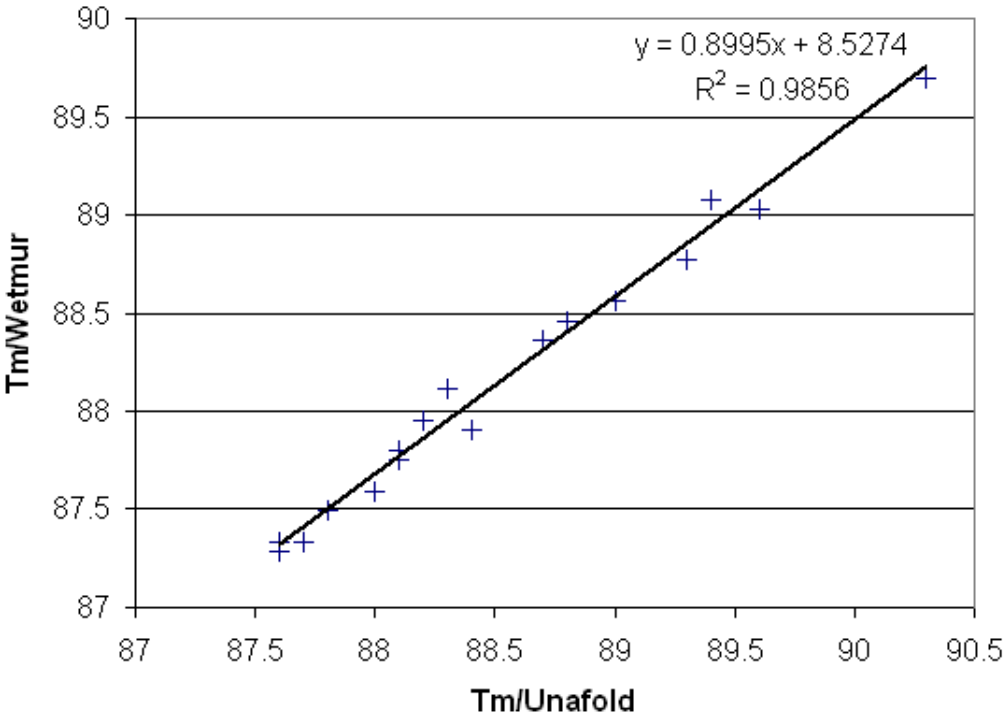


Figure 3.7: Correlation between different calculation methods for T_m. The values for the x-axis are calculated with UNAFold, for the y-axis with the Eqn. 3.3. The UNAFold parameters were $C_T = 0.000001$ M, $[Na^+] = 0.01$ M, $[Mg^{++}] = 0.0025$ M, for the calculation according to 3.3 the cumulative ionic concentration was set to 0.2.

3.4.4 Extraction of NN parameters from experimental data

Thermodynamic parameters from endpoint measurement

There are several ways of extracting dinucleotide parameters from the experimental data [86]. One can either fit the Langmuir equations 3.38, or for experiments at sufficiently low concentrations one could consider that far from chemical saturation, i.e. when only a small fraction of surface sequences is hybridized i.e. $\left(ce^{-\Delta G/RT} \ll 1\right)$ one can neglect the nominator in 3.38 to get:

$$I \simeq Ace^{-\Delta G/RT} \tag{3.28}$$

The parameters could be extracted either from an experiment at fixed concentration c, by comparing the intensities of different probe sequences, or from experiments at different concentration by analyzing the intensities of identical probe sequences over a concentration range. The overall scale factor A changes from experiment to experiment. A possible solution is to use interslide normalization or use relative intensities. But obviously the process conditions are strongly influencing the absolute intensity. This is an inherent disadvantage compared to real-time measurement.

Thermodynamic parameters from melting analysis

Based on the two state model and assuming that the thermodynamic parameters are temperature independent it is possible to deduce the van't Hoff enthalpy ΔH_{vH} [21, 26]:

$$\Delta H_{vH} = 6RT_{T_m}^2 \left(\frac{d\theta}{dT}\right)_{T=T_m} \tag{3.29}$$

with θ beeing the fraction of remaining duplexes. The entropy ΔS^0 and the free energy at temperature T ΔG_T^0 of DNA hybridization can be calculated:

$$\Delta S^0 = \frac{\Delta H_{vH}}{T_m} - R\ln\left(\frac{c_a^{ss}c_b^{ss}}{c^{ds}}\right)_{T=T_m} \tag{3.30}$$

$$\Delta G_T^0 = \Delta H_{vH} - T\Delta S^0 \tag{3.31}$$

In a similar approach to calculate dissociation enthalpy ΔH_{sol}^0 and entropy ΔS_{sol}^0 from solution melting experiments a fitting of the melting curve can be done [25]:

$$\Delta G_{sol}^0 = -RT\ln\left(\frac{\theta(T)^2}{1-\theta(T)}c_{ds}\right) \tag{3.32}$$

$$\Delta G_{sol}^0 = \Delta H_{sol}^0 - T\Delta S_{sol}^0 \tag{3.33}$$

Some results for the thermodynamic parameters of our long oligo test system can be found in Table 3.4. the calculation was done with [231].

3.4.5 The influence of mismatches

Mismatches within DNA-duplexes have important biological consequences. They have to be detected during replication and corrected from the post-replicational repair system [218, 110] as well as mutations during cell lifetime have to initiate adoptosis if they can not be corrected [172, 6]. The post-replicative repair system is apparently specialized to process those X•Y mismatches that occur comparatively frequently as the result of errors made by the DNA polymerase, i.e. purine•pyrimidine

Table 3.4: Calculated thermodynamic paramters of the long oligo probes for *solananceas* based on the algorithm [231].

	ΔH $[kcal/mol]$			ΔS $[cal/mol]$		
	1Potato	**2Pepper**	**5Tobacco**	**1Potato**	**2Pepper**	**5Tobacco**
01-032	-487.2	-439.5	-443.4	-1301.1	-1184.9	-1188.2
01-277	-479.6	-408.9	-410.3	-1294.6	-1112.4	-1106.8
01-583	-469.4	-405.2	-412.3	-1280.7	-1117.3	-1133.2
02-120	-447.7	-477.3	-433.9	-1221.4	-1294.5	-1191.3
02-714	-414.2	-473.8	-393.6	-1124.2	-1275.2	-1077.4
05-103	-453.9	-417.4	-477.2	-1226.2	-1125.6	-1292.1
05-387	-419.7	-429.0	-484.8	-1137.9	-1161.4	-1304.5
05-399	-386.4	-405.1	-479.7	-1051.2	-1095.8	-1285.4
05-543	-419.3	-389.1	-477.4	-1133.4	-1057.0	-1281.6

and certain purine•purine mismatches While the nearest neighbor models limit the stacking interactions to the adjacent base pair it is known that a X•Y mismatch within palindromic NN basepairs but not within a palindromic oligo (e.g. d(TTXAC)•d(GTYAA)) has thermodynamic properties that differ from the corresponding Y•X mismatch (e.g [218]). This quite old knowledge clearly suggests that the interaction is not limited to nearest neighbor basepairs. Despite this fact there are still many publications about nearest-neighbor based mismatch parameters [185, 110, 43, 86] and only few which try to point out this lack [155, 157]. [185] reported for the order of mismatch stability of DNA•RNA hybrids: rG•dT $\gg$ rU•dG $\approx$ rG•dG $>$ rA•dG $\approx$ rG•dA $\approx$ rA•dC $>$ rA•dA $\approx$ rU•dT $\approx$ rU•dC $>$ rC•dA $\approx$ rC•dT. In general the longer the oligo the smaller the influene of an SNP. There is an ongoing discussion on the reliability of surface based hybridization with short oligo systems [157]. Real time melting analysis gives a tool for an accurate characterization of the physical chemistry of different hybridization systems. Good publications on this item are [163, 136].

3.4.6 Influence of probe length on melting analysis

In general longer oligos have a higher degree of cooperativity. Thus the melting transition is sharper the longer the probe molecule. Sensitivity is increasing and specificity is decreasing with the oligo length. For SNP detection with standard endpoint based microarray systems short oligo probes have to be used. A good overview on the performance of microarray systems are heatmaps of all probe target combinations. Figure 3.10 gives the heatmap for our short oligo system, Figure 3.8 for the long oligo system and Figure 3.9 for a probe system based on [111]. The corresponding mismatch tables are found in Appendix A. It can be seen that with melting analysis the specificity of a 60 mer system is in the range of short oligo systems. Due to the more pronounced melting transition the long oligo system are well suited for analysis based on T_m.

3.5 Thermodynamic changes during melting

3.5.1 Introduction

When comparing the van't Hoff enthalpy derived from UV measurements with the calorimetric enthalpy it was often found that the two quantities disagreed sometimes up to a factor of 2 [222]. There are several interpretation, that the helix-to-coil melting is a non-two-state prcess and the difference in hydration between the duplex-stranded groups and single-strand groups results in a heat capacity increase. In general short oligonucleotide sequences the melting behaves in a two

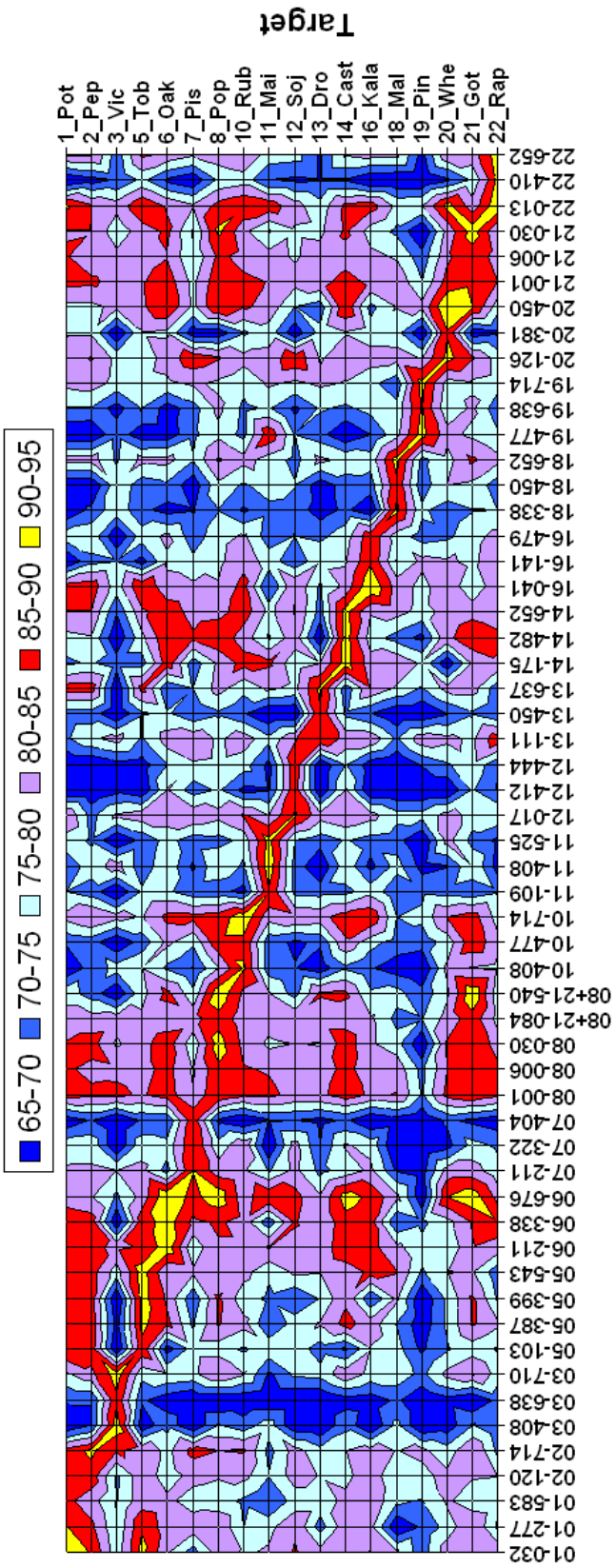


Figure 3.8: Heatmap of T_m for long oligos. The y-axis gives the target name. T_m was calculated using [231], the influence of formamide is not included within this algorithm.

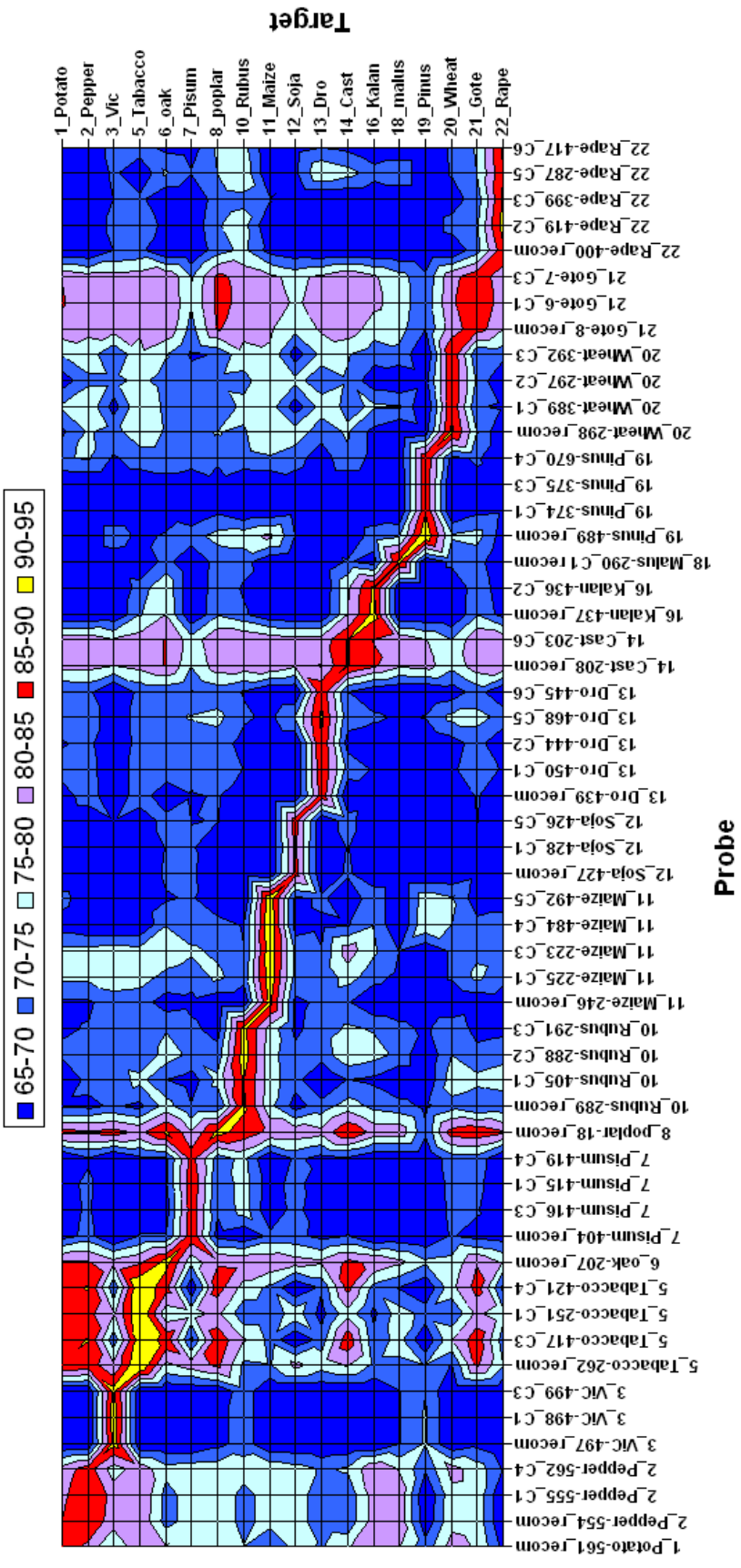


Figure 3.9: Heatmap of T_m for long oligos designed according to [111]. The high specificity is visible. T_m was calculated using [231], the influence of formamide is not included within this algorithm.

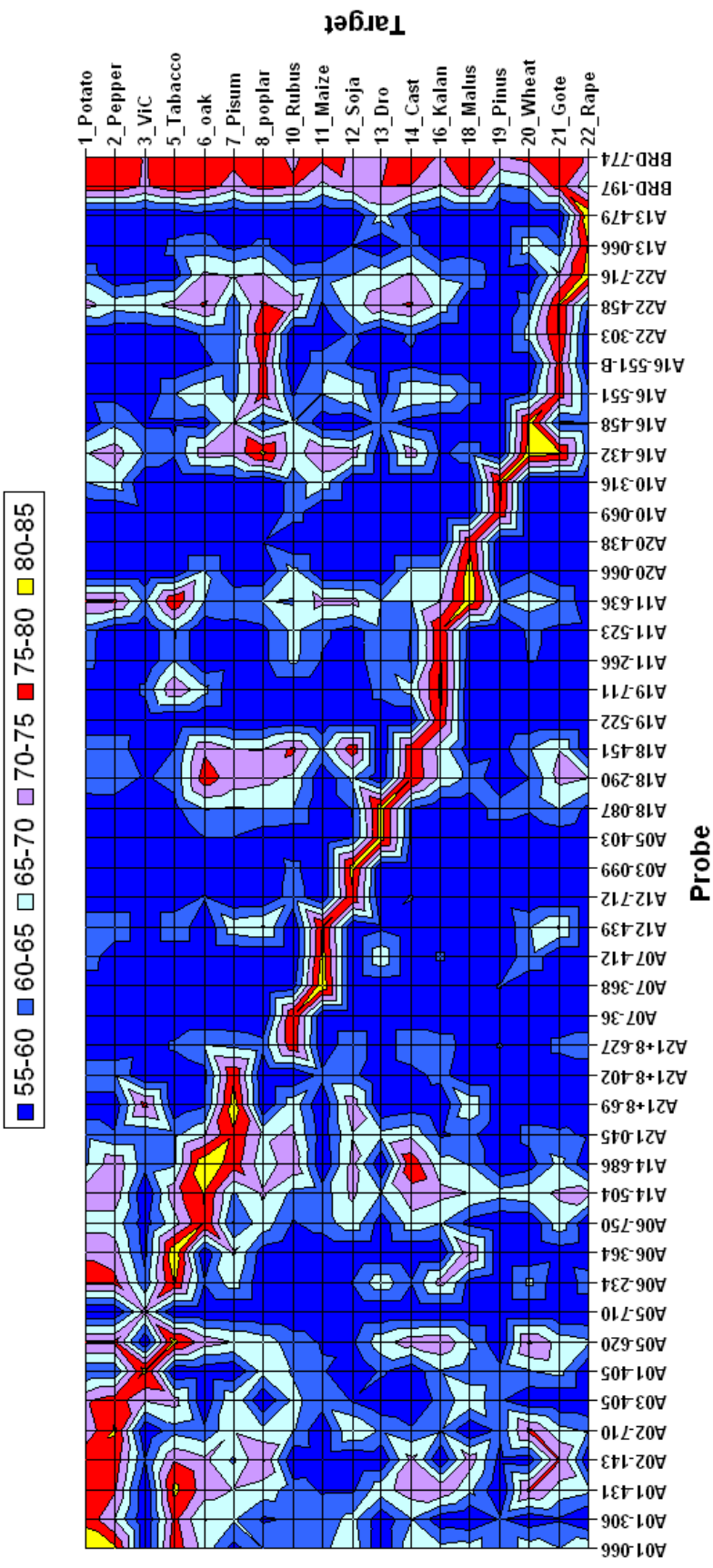


Figure 3.10: Heatmap of the T_m for the short oligo system. T_m was calculated using [231], the influence of formamide is not included within this algorithm.

state transitions while for longer oligonucleotide sequences the duplex formation often behaves as a non-two-state transition due to the self assembled population of single strands.

Good overviews on the changes of heat capacity and thermodynamic parameters during melting are [167, 168, 23, 189]. There are some fluctuation effects during melting which are still not fully understood [94, 135].

3.5.2 Calorimetric parameters of DNA melting

The changes in thermodynamic parameters can be extracted from nearest neighbor based algorithms, e.g. [231]. The thermodynamic data for our actin based long oligo system can be found in Table 3.4.

The extraction of thermodynamic parameters of DNA with Differential Scanning Calorimetry An example for an experimental result for the change in heat capacity is given in Figure 3.11.

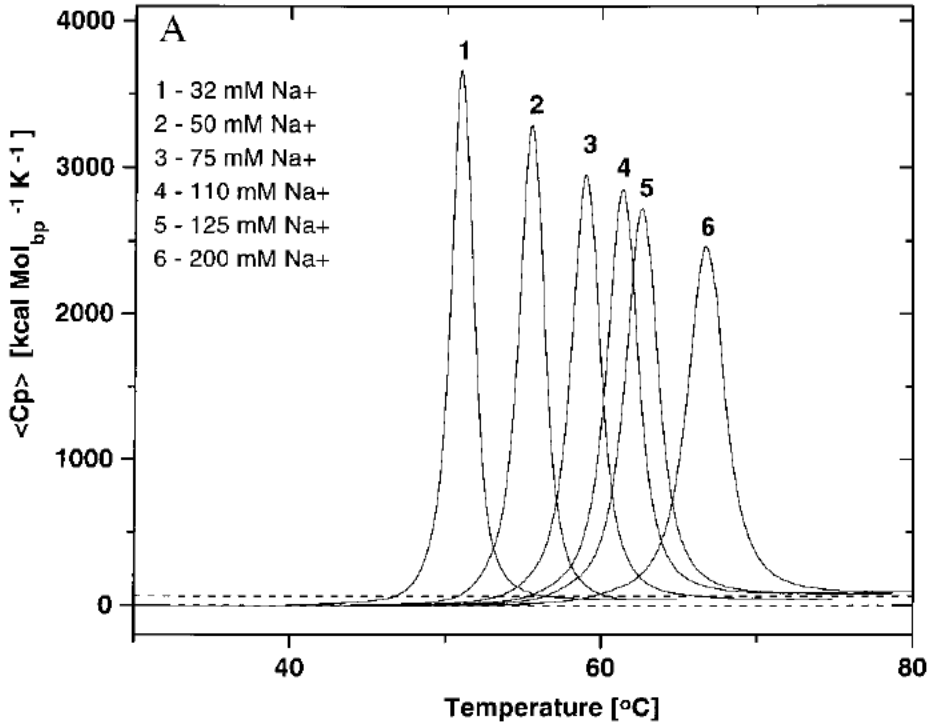


Figure 3.11: Experimental data for heat capacity from Diferential Scanning Calorimetry experiments [23]. The y-axes is not correct labeled it is 'cal/mol/K'.

3.6 Thermodynamics of DNA•RNA hybrids

There are several publications doing investigations on the thermodynamic analysis of DNA•RNA hybrids [76, 75, 202, 184, 185]. Without the mathematical constraints shown in section 3.4.1 there are 16 instead of 10 NN parameters necessary due to the asymmetry of the two strands. A comparison of the physical parameters of DNA•DNA, DNA•RNA and RNA•RNA duplexes is given in [67]. RNA•RNA duplexes are the most stable. If DNA•DNA or the corresponding DNA•RNA duplexes has a higher thermodynamic stability depends on the sequence, see Figure 3.6 p.50. Generally it was found in solution experimetns that the duplex initiation process for DNA•DNA, DNA•RNA and RNA•RNA duplexes seems to be energetically similar [184].

3.6.1 Structure and thermodynamics of DNA•RNA duplexes

RNA duplexes form A-type conformation characterized by $C3' - endo$ sugar puckers in aqueous solution while DNA duplexes adopt a B-form with $C2' - endo$ Figure 2.5 p.13 [209]. The axial rise and the rotation from one residue to the next are: 0.26 nm and 32.7° for A-RNA and 0.34 nm and 36° for B-DNA [209]. An overview on the structure of DNA•RNA hybrids is given in chapter 2.

Wobble base pairs

A wobble base pair is a non-Watson-Crick base pairing between two nucleotides. The four main wobble base pairs are G•U, G•G, C•A, A•A, A•G and some with the base inosine: inosine-uracil, inosine-adenine, and inosine-cytosine I•U, I•A and I•C [218, 90]. The thermodynamic stability of a wobble base pair is comparable to that of a Watson-Crick base pair. Wobble base pairs are fundamental in RNA secondary structure and are critical for the proper translation of the genetic code. Investigation on dU•dG and dT•dG indicated an increase in stability due the methyl group of the thymine. Figure 3.12 shows the molecular structure of the methyl group on the $C5'$ position. Methyl groups are supposed to stabilize by enhancing stacking as well as hydrophobic interaction [185, 137]. It is improtant to note that the methyl-group on the $C5'$ position of cytosine is the key

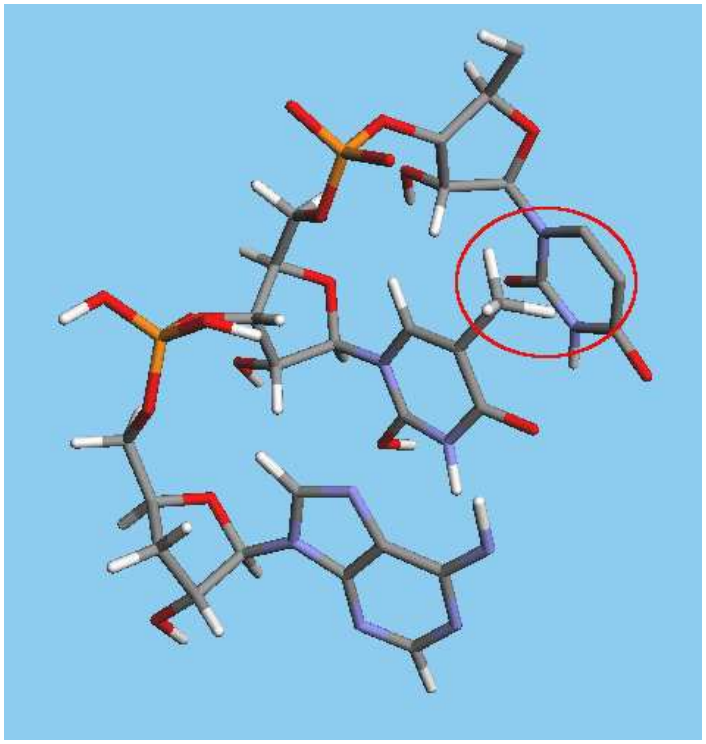


Figure 3.12: Position of the C5-methyl group of thymine within a A-RNA duplex (red selection).

effect of epigenetics.

3.7 Thermodynamic of surface adsorption

3.7.1 The Langmuir theory of adsorption

This theory was devised by Irving Langmuir, 1881-1957, an American industrial chemist who won the 1932 Nobel Prize in chemistry for his work on surface chemistry [132]. This theory assumes the process:

$$A \; + \; surface \; site \; \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \; A_{adsorbed} \tag{3.34}$$

where A is a solution particle which can form a single layer of molecules on the surface sites. k_a and k_d are the reaction rates of absorption and desorption. This surface monolayer model is a commonly used approximation for the surface hybridization on microarrays. Let Θ represent the fraction of the surface sites that are occupied by adsorbed A molecules. The adsorption is assumed to be an elementary process thus the rate of adsorption is proportional to the concentration of A in the fluid phase and to $1-\Theta$, the fraction of surface sites available for adsorption:

$$rate\ of\ adsorption = k_a[A](1-\Theta) \tag{3.35}$$

where $[A]$ is the concentration of A in the gas or liquid phase and k_a is a rate constant that can depend on temperature but not on $[A]$ or Θ. The desorption is also assumed to be an elementary process

$$rate\ of\ desorption = k_d\Theta \tag{3.36}$$

where k_d is the rate constant of desorption. At equilibrium, the rate of desorption equals the rate of adsorption:

$$k_d\Theta = k_a[A]1-\Theta \tag{3.37}$$

This equation can be solved for Θ to give the *Langmuir isotherm*:

$$\Theta = \frac{k_a[A]}{k_d + k_a[A]} = \frac{K[A]}{1+K[A]} \tag{3.38}$$

where K is an equilibrium constant given by

$$K = \frac{k_a}{k_d} \tag{3.39}$$

Thus a *Langmuir isotherm* gives a relation between the surface concentration and the fraction of occupied surface binding sites under the assumption of constant target concentration. This assumption will be reviewed critically in the next section.

3.7.2 Is Microarray hybridization a Langmuir process?

One assumption for the Langmuir behavior is the constant concentration of the target molecules within the solution. It can be seen from Table 2.3 and Table 2.4 that the number of target and probe molecules is within the same range for targets in higher concentrations. Thus the assumption of constant concentration is not valid, especially for low expressed genes in microarray experiments without preaplification. Thus for an accurate quantitative prediction a diffusion-reaction based modeling is necessary. Simulation results for the depletion of target concentration during hybridization are shown in 4.4. A test on Langmuir like behavior can be done using

$$\frac{1}{\Theta} = \frac{1+K[A]}{K[A]} = \frac{1}{K[A]} + 1. \tag{3.40}$$

Plotting of $1/\Theta$ vs. $1/[A]$ should give a linear function proportional to $1/K$ with an intercept equal to unity if a Langmuir isotherm is obeyed.

3.8 Publication: Solid phase high resolution melting for improved microarray specificity

S. Krainer, S. Fluch, M. Stierschneider , L. Bodrossy

Submitted to *Journal of Biophysical Chemistry*

3.8.1 Abstract

The rather complex processing of state-of-the-art microarray systems introduces a range of process parameters influencing the quality of the assay results and is limiting the direct application of hybridization analysis for integrated Lab-on-a-Chip systems. A possible solution is real-time data acquisition during hybridization as well as denaturation to gain more reliable information about kinetic and thermodynamic properties of probe-target duplexes. To test the potential of high resolution melting analysis in biochip applications, a 60 mer Actin based cross-species microarray was developed and evaluated. With this experimental setup it was possible to show that it is possible to detect SNP using long oligo microarray system in combination with solid phase high resolution melting. Compared to state-of-the-art endpoint based microarray systems, the method provided better reproducibility, smaller influence of process parameters, detection of multiple target hybridization and better discrimination between specific and nonspecific hybridization. Using this system it was possible to discriminate even closely related species showing the potential for the analysis of mixed samples e.g. for food or microbial diagnostics. Microarray based melting analysis provides valuable data to improve the understanding and modeling of nucleic acid hybridization to surface bound oligos thus elucidating the physico-chemical basics for next generation of miniaturized biosensors independent of the detection principle.

3.8.2 Introduction

While microarrays are widely used for collecting biological data, there is still limited understanding of the possibilities and limits of the existing technology. In addition to binding energy there are several factors influencing the final detected fluorescent signal intensity of probe-target duplexes e.g. surface probe density, microarray surface composition, target labeling and secondary structure [165, 157]. The analysis of endpoint based intensity data without the careful discrimination between specific and nonspecific binding events can therefore lead to incorrect assignment of basepair interactions and thus false positive signals. Excluding nonspecific hybridization to a given probe is still a challenge in microarray probe design and data analysis [214]. There is an ongoing discussion about possible improvements based on algorithms for probe design and determination of unspecific background intensity and cross hybridization [11, 124, 156, 169]. Another approach is to adjust duplex stability by chemical modification of probes, rather than confining oneself in the available sequence space. Oligonucleotides that give duplex stabilities independent of sequence (isostable DNA) can be expected to be superior hybridization [1]. But the development and verification of these methods also suffer from the experimental verification on microarrays. This lack in understanding of hybridization to surface bound oligos is one of the limiting steps in the development of integrated hybridization based Lab-on-a-Chip assays. Cross hybridization on DNA microarrays was proved using high-throughput sequencing after target capture. Only 40 to 75 % of the captured DNA fragments mapped to expected genomic loci [2]. The rest of the target bound to the DNA probes was due to unspecific hybridization. For a better understanding of nucleic acid hybridization and dissociation to solid phase bound oligo nucleotides, a systematic analysis of kinetic and thermodynamic behavior of probe-target duplexes is necessary. While the analysis of melting behavior of double stranded nucleic acids is already used

for solution based high resolution melting analysis (HRM) [36, 85, 221] and for advanced surface based methods with small probe numbers like surface plasmon resonance (SPR) [74, 223, 225] or dynamic allele-specific hybridization (DASH) [87, 183], there are still only few experimental data available for standard microarray based melting analysis [97, 125]. This can be attributed to the experimental and bioinformatics difficulties of real time data acquisition using microarray formats [49, 77] To evaluate the biological and physico-chemical relevance of melting analysis the Actin X-chip was designed. Actin was chosen because it is one of the most conserved housekeeping genes in eukaryotes [63, 160]. Cross-species comparison of this gene provides a set of sequences with a broad range of homology levels. For the probes we chose 60 mers being a good compromise between specificity and sensitivity [43, 197, 211, 217]. While there are quite a lot of algorithms for the calculation of the thermodynamic parameters of short oligo probes [107, 127] there is still no accepted model for the prediction of melting parameters for surface bound long oligos (i.e. 60 mers). Thus there is still the need for basic research on thermodynamic properties and potential for specificity improvement of long oligo assays. For endpoint long oligo systems thresholds of sequence similarity for detectable cross hybridization events have been reported to range from 75 $\%$ [95] to 87 $\%$ [197] of minimum similarity. This equals 8 to 15 MM for 60 mer oligos. Improving the specificity of long oligo microarrays would open the possibility for high sensitivity and specificity microarrays. Within this work, real time melting analysis as a tool for the discrimination of specific and nonspecific hybridization and for improvement of microarray data quality was evaluated using the Genewave HybLive real-time platform [66]. Overall we did 40 experiments, starting with basic measurements for reproducibility and hybridization optimization up to mixtures of labeled targets.

3.8.3 Materials and Methods

Isolation of the Actin gene fragments

In angiosperms, actina are encoded by a relatively large, diverse, and dispersed multigene family comprising 8-40 genes per species. As in most other eukaryotes, the multiple copies of actin genes in angiosperms are thought to allow a more diverse pattern of gene regulation, rather than direct the production of a large amount of actin. Relative to other nuclear genes, actin genes are highly conserved in both size and sequence. However, relative to mammalian actins, angiosperm actin genes show a greater sequence divergence [131]. Within this work actin gene fragments from 18 different plant species have been extracted, PCR amplified, cloned and sequenced. PCR amplification was performed on genomic DNA (gDNA) from the following species potato (*Solanum tuberosum* L.) , bell pepper (*Capsicum annuum* L.), faba bean (*Vicia faba*), tobacco (*Nicotiana tabacum*), truffle oak (*Quercus robur* L.), garden pea (*Pisum sativum* L.), black poplar (*Populus nigra* L.), european raspberry (*Rubus idaeus*), maize (*Zea mays* L.), soybean (*Glycine max*), sundew (*Drosera rotundifolia* L.), european chestnut (*Castanea sativa Mill.*), devil's backbone (*Kalanchoe daigremontiana*), apple (*Malus x domestica*), scots pine (*Pinus sylvestries* L.), bread wheat (*Triticum aestivum* L.), tree-of-heaven (*Ailanthus altissima*), oilseed rape (*Brassica napus*) using the degenerate primers ActinForward: 5′-ACT GGG ATG AYA TGG AGA AG-3′ and ActinReverse 5′-AYC CTC CAA TCC AGA CAC TG-3′, PCR reaction of 50 l volume, consisting of 10x PCR buffer (KTAQ, Kbioscience buffer B), 20 mM dNTP each, 5 U Taq polymerase (KTAQ, Kbioscience), 4 μM of each primer and 20-50 ng genomic DNA as template, were performed in a Dyade thermocycler (MJResearch Dyade). Amplification conditions were: 95 ℃ for 15 min activation, then 34 cycles of: 30 s at 95 ℃, 30 s at annealing temperature 55 ℃, 1 min at 72 ℃, followed by a final elongation step of 10 min at 72 ℃. To ensure that only a single fragment of the actin gene family is used in the consecutive steps [129, 133] the resulting PCR fragments were cloned using TA cloning kit (Clonetech) following manufacturers instructions. Plasmid DNA was extracted from transformed *E.coli* using

standard alkalyine lysis protocol. Per fragment one clone was picked, the insert was PCR amplified and sequenced (AGOWA GmbH). The sequences were submitted to NCBI GenBank: accession no. GQ339765-GQ339782. After sequencing, a phylogenetic tree was generated using ClustalW [195] to choose 6 out of the 18 gene representatives for further experimental steps. Two phylogenetically close groups with high intra-group sequence homology and limited similarity to other actin genes of the test set were selected as targets:

- Closely related *solanaceae*: potato-pepper-tobacco
- Closely related: oak-poplar
- Orthogonal probe with little similarity to other probes: pinus (*pinaceae*)

Probe design and preparation of the slides

Based on the sequences of the 18 actin gene sequences, probe design was carried out using the ARB phylogenetic software package [120]. The synthesis of the 54 unmodified desalted oligos was done by Sigma-Aldrich. A mismatch table for the 20 probes versus the 6 species which were used in this study are shown in Tab. 1.

Each oligo was dissolved in 10 μl 50% DMSO to a probe concentration of 50 μM and spotted in three technical replicates to provide measurements of intra-microarray reproducibility. In addition probes for the 6 selected targets were spotted in three additional concentrations (5.6, 16.7 and 150 μM). With this setup the influence of the probe concentration on hybridization intensity and melting point was investigated. Spotting was performed on Genewave Amplislides® with Aldehyde coating using SMP3 stealth pins on a Omnigrid 100 platform. Processing of the spotted slides was carried out using the borohydride method according to [19]. Dried slides were stored at room temperature in the dark until use.

RNA target preparation

Preparation of the RNA targets was carried out as described earlier [19]. In brief, actin fragments were amplified from plasmid DNA by PCR (same protocol as above) using a modified reverse primer, containing the T7 promoter site: 5′-TAATACGACTCACTATAG-[ActinReverse]-3′. The resulting PCR product was used as template for *in vitro* transcription as follows: 8 μl purified PCR product (50 ng/ μl), 4 μl 5x T7 RNA polymerase buffer, 2 μl DTT (100 mM), 0.5 μl RNAsin (40 U/μl) (Promega), 1 μl each of ATP, CTP, GTP (10 mM), 0.5 μl UTP (10 mM), 1 μl T7 RNA polymerase (40 U/μl) (Gibco BRL) and 1 μl Cy3 or Cy5-UTP (5 mM) were mixed in a 1.5 ml microcentrifuge tube and incubated at 37 °C for 4 h. RNA was purified immediately using the Quiagen RNeasy kit according to manufacturer's instructions. Purified RNA was eluted into 50 μl dH2O and fragmented in a chemical procedure using alkaline conditions, elevated temperature and Zn(II)-ions [19]. PCR yield and dye incorporation rate (degree of labeling, DoL) measurement was done using a ND-100 Spectrophotometer (NanoDrop, Wilmington, DE).

Hybridization and melting temperature measurement

Hybridization was carried out using the Genewave HybLive® hybridization and scanning workstation [125]. The hybridization buffer consisted of 6x SSC, 1x Denhardt's reagent, 0.1 % SDS and 45 % formamide. For all hybridizations we used stock solutions to minimize the influence of chemical batches and variation in preparation. The standard target concentration for hybridizations was set to 1 nM. To reduce the influence of secondary structures, the hybridization solution was incubated

at 95 ℃ for 5 min and cooled on ice immediately before loading on the microarray. During the 12h hybridization time the temperature was set to 42 ℃ and the agitation was on. One picture/min was taken for the first hour of hybridization and 1 picture/30 min afterwards. After hybridization slides were washed for 5 min in 6x SSC followed by a high stringency wash for 5 min with 0.2x SSC, both at the hybridization temperature of 42 ℃ in the HybLive® equipment. The recorded signal intensity after the washing was referred to as end-point measurement data. Before melting, 15 exposures (1 per minute) at hybridization temperature were done to ensure equilibrium conditions and to evaluate the amount of photobleaching. Within our measurements we did not see a significant impact. The melting analysis was done in hybridization solution ramping the temperature from 42 ℃ to 85 ℃ at a rate of 1 ℃/min with image acquisition every 20 s. For the "low quality" experiments, hybridized slides were refurbished by denaturing for 15 min at 95 ℃ in dH2O and rinsing for 5 min afterwards in dH2O.

Data processing

Segmentation and quantification of the generated .tif files was done using the Genewave HybLive® software package using the "irregular shape" segmentation algorithm, all parameters were set to default values. Quantification was based on the average of the three technical replicates for each spot. Normalization and analysis of the melting curves was done with the HybLive® software package. T_m was defined as the temperature at which 50 % of the probe-target duplexes were denaturated. For most of the analysis we skipped the 5.6 μM probes due to the low signal intensities. For the automated analysis of T_m, the generation of mismatch (MM) tables and the correlation between different experiments self written Matlab scripts were used.

3.8.4 Results and discussion

Assessment of experimental quality

To ensure the experimental quality of the data, several repeated experiments were done, including dye swap experiments and measurements with refurbished slides ("low quality"- measurement). The corresponding correlation coefficients (R^2) of endpoint measurement were 0.93 for the repeated measurement, 0.92 for the dye swap and 0.605 for the low quality slides. For the melting analysis irrespective of the treatment, all correlation coefficients were $<$ 0.98, see Figure 3.13. This shows that the reproducibility of melting analysis experiments is much better than that of an intensity based assay. A typical result for a set of melting curves can be found in 3.14. The influence of the probe concentration (150 μM, 50μM and 16.7 μM) is negligible even though SPR-based publications report an influence of probe density on thermodynamic parameters due to electrostatic repulsion [223, 225]. The fact that T_m was independent of probe concentration in our experiments might be due to the higher ionic concentration (6xSSC equals 900 mM NaCl) in our experimental setup as compared to SPR measurements. An estimation of the Debye shielding radius (describing the range of electrostatic forces), can be estimated as follows [41].

$$\lambda_D = 0.3nm/\sqrt{[NaCl]} \tag{3.41}$$

Our hybridization buffer, containing 0.9 M NaCl gives a quite small Debye shielding radius of $\lambda_D = 0.3$ nm, which is much smaller than the DNA helix diameter of about 2 nm which gives the distance of the negatively charged phosphate groups. The average probe to probe distance on the spot is about 15nm, assuming a surface density of $\sim 5 \times 10^{11}\ molecules/cm^2$ [187]. This indicates that the negatively charged backbone of the DNA is shielded effectively within distances much smaller than the relevant geometric dimensions. Melting point analysis reduces the influence of probe concentration,

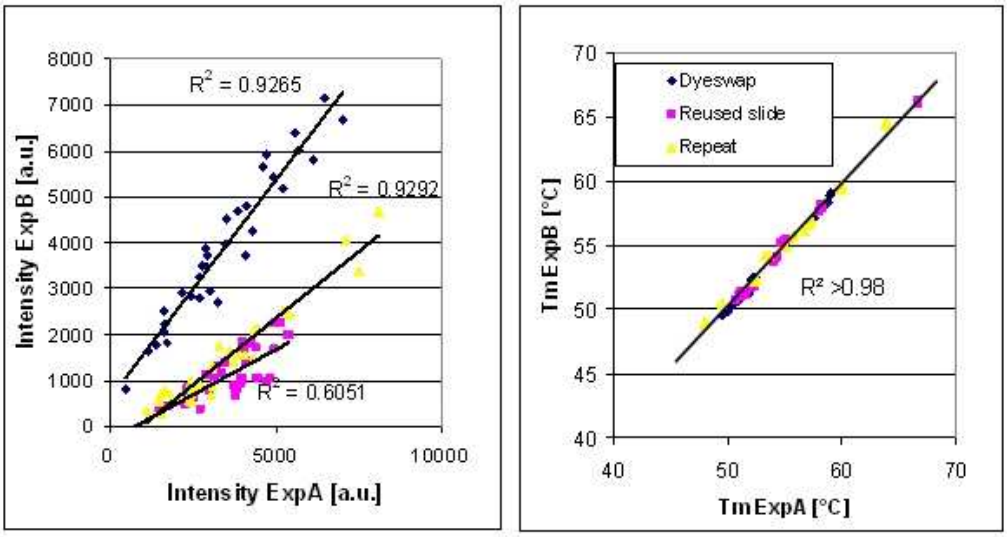


Figure 3.13: Correlation between intensity based measurement (left) and melting analysis (rigth) with the same target. The blue diamonds give the correlation for repeated measurement, the red squares between a standard experiment an a "low quality" experiment with a reused slide and the yellow triangles mark a dye swap experiment. Due to background subtraction there is an offset of the regression line.

spotting artifacts, etc. on the results as compared to standard intensity based readouts of microarrays where these factors obviously influence the results [165]. This is related to the principle of measuring the change fluorescent intensity during melting transitions instead of an absolute value as being done in endpoint systems.

Detection of unspecific hybridization

One of the main problems of state of the art microarrays are false positive results due to nonspecific hybridization. Theoretically it is possible to minimize this effect by isothermal probe design but there is still no widely accepted model for melting point prediction of surface bound long oligos especially with mismatches or when working with different types of nucleic acids [157]. Figure 3.15 gives an extreme example for the melting curve of nonspecific hybridization to a 3 mismatch (MM) probe 02x120 as compared to specific hybridization to perfect match (PM) probes 05x399 and 05x543. The 3 spotting concentrations (150 μM, 50 μM, 16.7 μM) of the MM probe showed hybridization intensities equal or higher than PM probes of the same concentration. In endpoint based microarray this would be a clear false positive signal. It is obviously possible to detect the nonspecific hybridization with melting analysis via the large difference in T_m between the MM versus the PM probes. While it would be possible to improve the specificity of probe 02x120 for endpoint detection by applying more stringent conditions, this would negatively influence the overall performance of the entire probe set. In general it is hardly possible to find an optimal hybridization temperature for each individual probe within a large probe set without knowledge of the exact T_m. Interestingly while probe 02x120 was found to be a "bad" probe for intensity based experiments, it was a quite specific probe when melting analysis was applied. For PM hybridization using the PM target pepper, the melting temperature of 02x120 is about 58 ℃ which is a T_m shift of 8 ℃ as compared to the 3 MM hybridization using tobacco as a target (T_m 50℃). In all out measurements, the correlation between end point measurement intensities and T_m values was found to be poor, even for PM probe-target

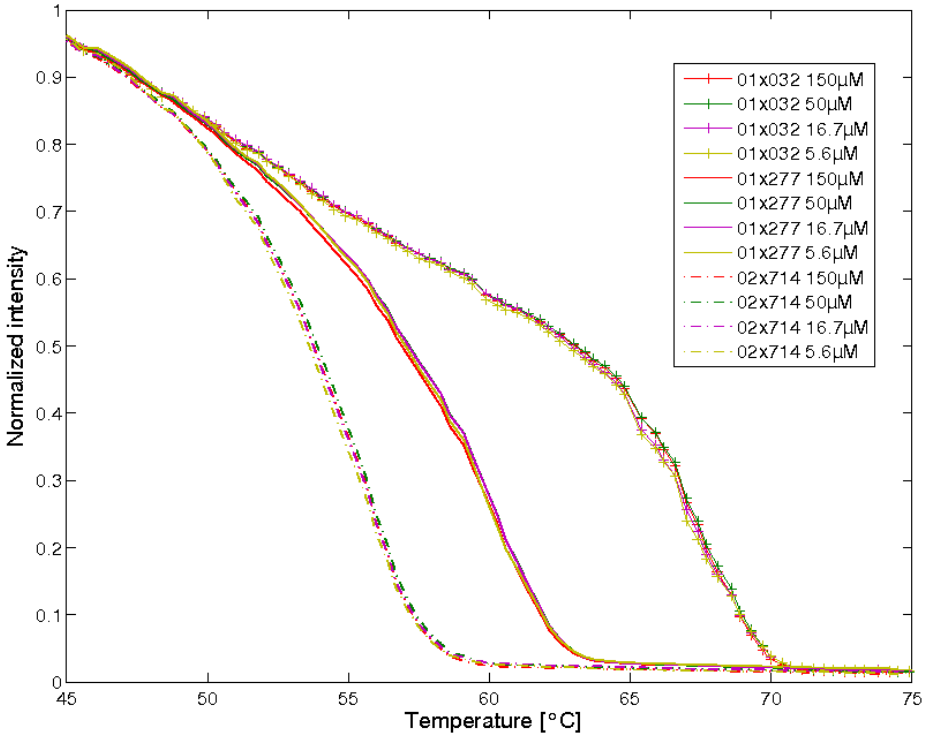


Figure 3.14: Typical normalized melting curves for three different probes. For every probe three different spotting concentrations are shown (16.5 μM; 50 μM; 150 μM. Target was potato (1). There is no significant influence of the spotted probe concentration on the T_m.

pairs (data not shown). To investigate if it is possible to detect cross hybridization based on the shape of the melting curve, results from single target experiments were compared to multiple target experiments where two targets were labeled with the same dye and co-hybridized to the same array. 3.16 shows typical results. For 2 MM probes there is an indication of cross hybridization with a second target which can be seen by the less pronounced melting curve as compared to the single target hybridization, while for the 4 MM case there is a clearly visible shoulder in the melting curve indicating 2 melting events at 2 distinct temperatures for the two hybridized sequences. This shape of the melting curves is similar to solution based high resolution melting analysis (HRM) results [221]. In HRM analysis, the formation of melting domains gives local denaturation inside the longer double stranded nucleic acid molecule where the intercalating dye is then being released. These shoulders are also generated by single target hybridization when melting domains are present. In microarray based melting analysis using labeled target molecules, this type of curves is only generated when one bound target is released by denaturation while the another target with higher T_m still remains bound to the probe. The visibility of this effect is dependent on the relative intensities, concentrations and differences in T_m of the respective targets. Specificity of melting analysis For an assessment of the specificity we used the relative differences of T_m in hybridization experiments with phylogenetically closely related pairs of targets. 3.17 shows the shift in T_m versus number of MM for the closely related solanaceae potato, pepper and tobacco (without any normalization/calibration to reference probes). These investigations were done not only for PM-MM target-probe combinations but also for targets with different numbers of mismatches. The close grouping of the data points for the same probes with different probe concentrations indicates a high accuracy of the measurement. Within our experimental setup it was possible to discriminate 2 MM even with non optimal probes with

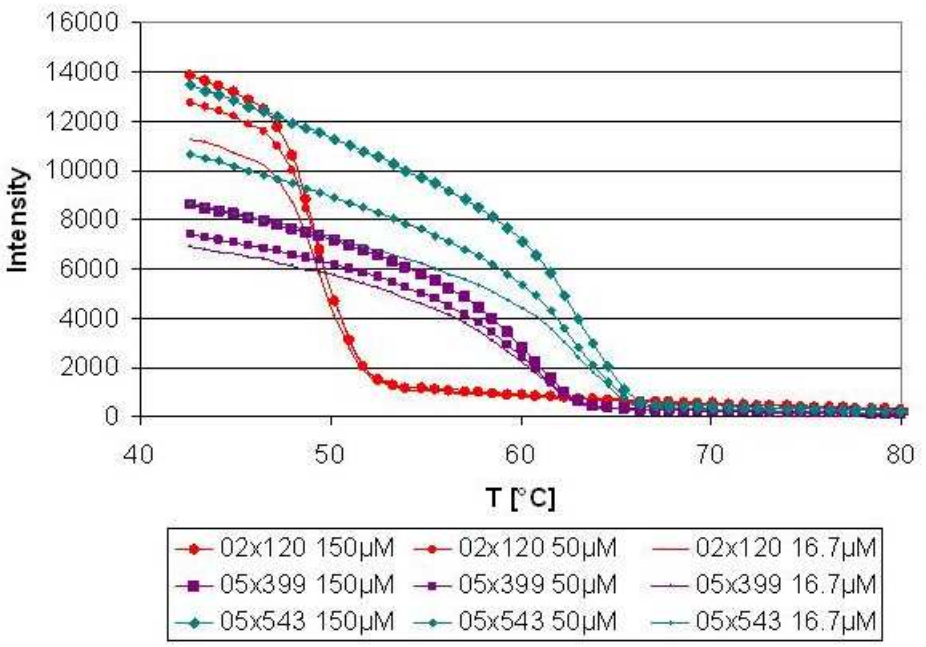


Figure 3.15: Melting analysis raw data (not normalized) of target 5 (tobacco), displaying perfect match to probes 05x399 (purple squares) and 05x543 (green diamonds) and 3 mismatches to probe 02x120 (red circles). While intensity based analysis under this conditions would give a false positive signal on probe 02x120 (value at 42 °C), melting curve analysis gives a clear discrimination of PM to MM. The different spotting concentrations demonstrate a high reproducibility of the experiment.

mismatches in terminal or near terminal position. It is known from endpoint analysis that mismatches at or close to the end of the probe have very little influence on specificity [136, 163, 89]. We did a simple correction on the representation of the T_m versus the number of effective mismatches: mismatches located less than 7 bp from the terminal end were not considered. The inset of Fig. 5 gives a representation of the difference in T_m versus these corrected (effective) mismatches There is a clear correlation between the effective number of mismatches and the change in T_m ($R^2 = 0.69$). The high significance, the high specificity and the high intra- and interslide reproducibility of the data indicate that our results are determined by inherent physico-chemical behavior of the nucleic acids.

3.8.5 Conclusion

Within this work an evaluation of microarray based melting analysis for long oligo probe sets was done. Very high reproducibility of melting point measurements ($R^2 > 0.98$) including dye swap experiments and artificially generated low quality measurements was found. False positive endpoint results caused by strong unspecific hybridization could be excluded with melting analysis due to the low T_m of the respective duplexes. Two nucleotide mismatches were reliably discriminated with 60 mer oligo probes based on the shift of T_m as compared to PM probe-target combinations without normalization or calibration to reference spots. As literature data did not suggest this high degree of specificity [197, 95], our probe set was not optimized for SNP detection. Thus, there is still potential for further improvement of specificity with melting analysis using probes designed for SNP detection. The shape of the melting curve provides information about cross hybridization caused by multiple

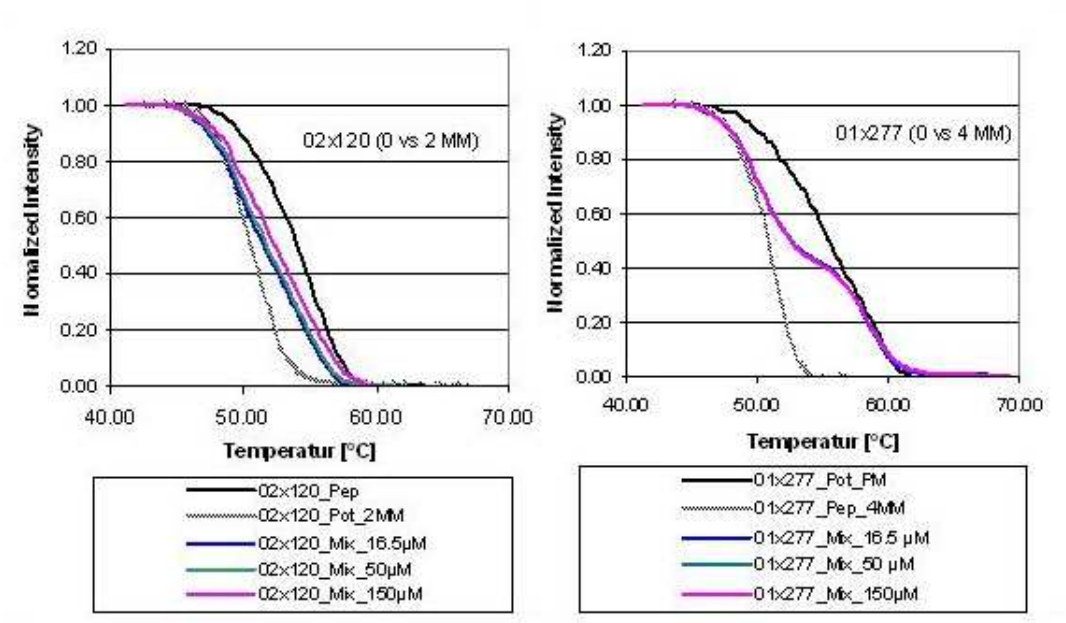


Figure 3.16: Normalized melting curves for a single target and a mixture of two closely related targets with (left) 0 vs 2MM and (right) 0 vs 4 MM. For the 2 MM there is a significant change in the slope of the melting curve. For the 4 MM probe there is a clear indication for multiple target hybridization due to the shoulder in the melting curve. For visibility we did show only one spotting concentration 50 μM for the single target melting curve but three (16.5 μM, 50 μM, 150 μm) for the melting curve of multiple targets to show the quality of the measurement.

targets with high similarity to the same probe. It was shown that for cross hybridization with 2 MM the slope of the melting curve changed significantly, for 4 MM a clear shoulder was visible in the melting curve. The limitations of microarray experiments in specificity as compared to solution based HRM analysis should diminish with melting analysis based methods. On-chip melting analysis provides a tool for the characterization of a large number of probes and can help to improve specificity for endpoint based assays using more accurate values for T_m as key parameter for probe design. Not only in solutions based assays but also with solid phase bound oligo probes melting analysis has the potential to generate large amounts of high quality thermodynamic data for different nucleic acid types and probe lengths. These data sets could serve as basis for the development and verification of new algorithms for the prediction of thermodynamic parameters of nucleic acid hybridization enabling the development of future miniaturized Biochip based systems with reduced process steps. Systematic investigation of larger probe-target sets and their thermodynamic analysis together with advanced bioinformatics will help get a better physico-chemical understanding of one of the most elementary processes in biology, the hybridization of nucleic acids.

3.8.6 Acknowledgement

The authors would like to thank Ildiko Matusikova and Friederike Trognitz for the support in generating the actin set of sequences and Julius Trognitz for the programming of the Matlab scripts.

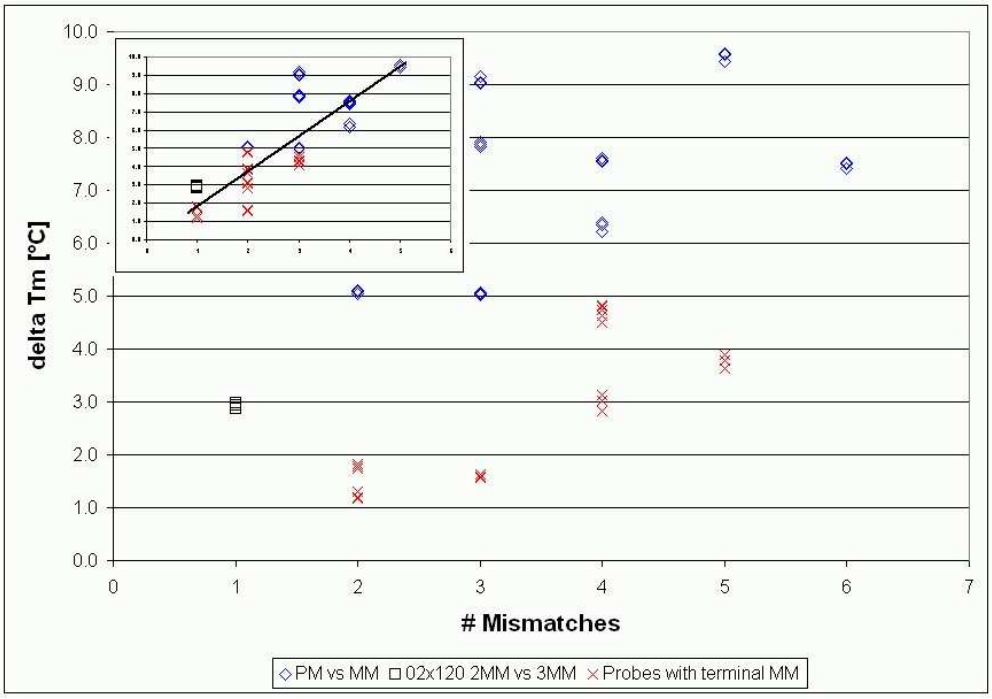


Figure 3.17: Shift of melting temperatures compared to the number of mismatches (MM). Some of the probe-target combinations had terminal mismatches which decrease specificity (red crosses). Well designed probes with mismatches close to the center showed very high specificity. We used the data of three different probe concentrations, for most of the probes there is no visible difference in melting temperature. The inset shows the same diagram when using effective mismatches thus ignoring mismatches close to the terminal end of the probe (less than 7 bases distance to the end). The correlation coefficient for the effective mismatches is $R = 0.69$.

Chapter 4

Kinetic of Nucleic Acid Hybridization

4.1 Diffusion of nucleic acid molecules

The thermal energy of nucleic acid molecules causes the Brownian motion and is the 'motor' of hybridization. Diffusion is an important factor in gene transcription and protein synthesis, thus the first investigations of diffusion were done in vitro with bacterial DNA and RNA. Quantitative understanding of the transport phenomena is vital for the modeling of biological processes. On microarrays the sensing of DNA occurs on the liquid-solid interface and two processes determine the rate of target capture. mass transport of the target to the sensing surface and kinetics of surface hybridization [13]. Especially for low abundant targets lateral diffusion of targets to the sensing spot and vertical diffusion of targets to the surface become rate limiting. Due to the non-equilibrium process a depletion zone is created above the spot which lead to experimental bias. Agitation during hybridization can help to overcome this limitations. Due to the transport limitations and competitive hybridization time-to-equilibrium can be in the range of 70 hours [147] Cross hybridization and non-equilibrium conditions have been implicated as mechanisms affecting the interpretation of microarray results [78, 13].

Besides being of interest in biophysics, DNA is now well established as a useful model system for studying basic polymer physics phenomena. DNA replication yields a homogeneous sample of molecules of exactly the same length and DNA topology can be precisely controlled.

4.1.1 Influence of DNA length on mobility

Theoretical models of the diffusion properties of rod shape biopolymers predict a power scaling law for the dependency between the polymer length and the diffusion constant.

$$D \propto L^{-\alpha} \tag{4.1}$$

In [41] the diffusion coefficient for DNA molecules in the range of 6 kbp to 290 kbp is measured by tracking the Brownian motion. A scaling law

$$D \propto L^{-0.571} \tag{4.2}$$

was found for linear DNA. The observed scaling is theoretically expected for 'flexible' molecules [171] which have lengths much longer than the persistence length (about 150 bp for native DNA). For shorter, 'semiflexible' molecules, the length dependence is expected to increase, a scaling exponent

Table 4.1: Diffusion coefficient for nucleic acids from different measurement

Author	**Viscosity**$[cP]$	**Basepairs**	**T**$[K]$	**Scaling exp**
Robertson	1.2	6 kbp...290 kbp	297	-0.571
Smith	0.95	4.3 kbp...310 kbp	297	-0.608
Lucaks	1	21 bp...6000 bp	297	-0.2

of $\alpha = 0.72$ for 21 to 6000 bp nucleotides is reported in [34]. A table with an overview on the results can be found in Table 4.1 and in Figure 4.1.

An estimation for the diffusion coefficient D of DNA in m^2/s for microarray relevant DNA fragments is [34]:

$$D = 4.9 * 10^{-10} * (n_{nucleotides})^{-0.72} \tag{4.3}$$

An overview on the results of these studies and a fit based on this parameters is shown in Figure 4.1.

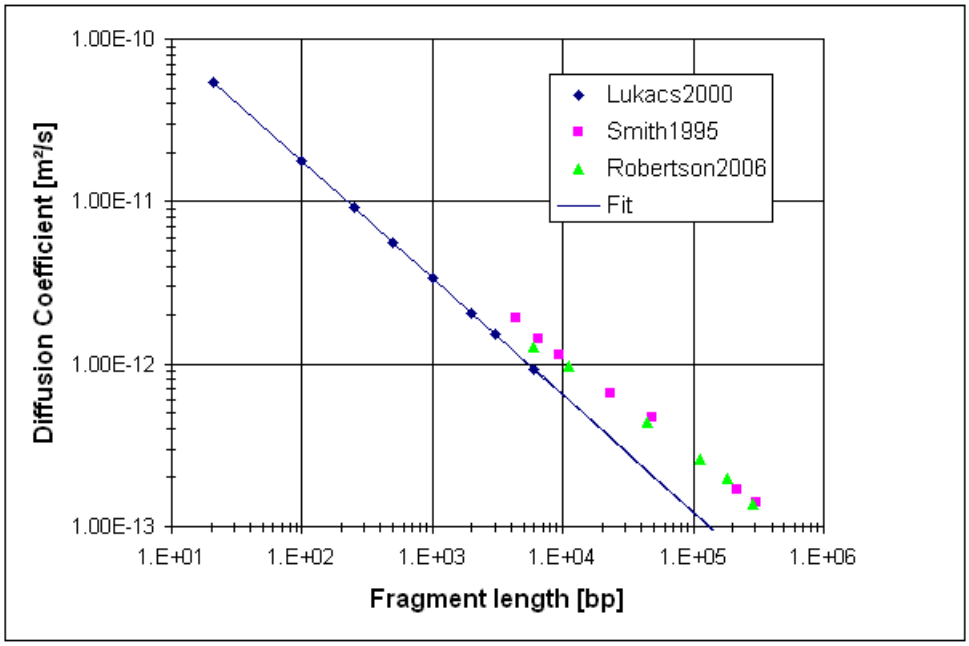


Figure 4.1: Diffusion coefficient for different DNA length from literature data [41, 171, 34] and fit according to Equation 4.3. The quality of the fit in the range of interest (20-1000 bp) is very good. There is an offset between the different studies, which might also be due to the different experimental conditions.

4.1.2 Estimation of the diffusion coefficient of DNA

As shown in Section 2.4.2 the diffusion coefficient is related with the average mean-square displacement

$$\langle x^2 \rangle = 2Dt \tag{4.4}$$

for each dimension. Thus for the 3d case

$$\sqrt{x^2} = \sqrt{6Dt} \tag{4.5}$$

It is interesting to note that due to this square root dependency diffusion is quite effective for short distances while for longer distances the average displacement is slowing down. Assuming a diffusion coefficient of $D = 1 \times 10^{-11}\ m/s^2$, the average time of DNA molecules for traveling characteristic distances is

- 6 μm (cell nucleus diameter): $t = 6 \times 10^{-1}\ s \approx 1$ s
- 70 μm (typical focal depth of microarray scanner): $t = 81.6\ s \geq 1$ min
- 500 μm (z-component of a microarray): $t = 4.16 \times 10^{3}\ s \approx 1$ h
- 5 cm (y-component of a microarray): $t = 4.16 \times 10^{7}\ s \geq 1$ year

A representation of the average displacement for different length of DNA fragments is shown in Figure 4.2. Obviously the assumption of constant concentration is not fulfilled for microarray experiments without effective agitation. Comparing the result with [52] gives a quite good agreement for

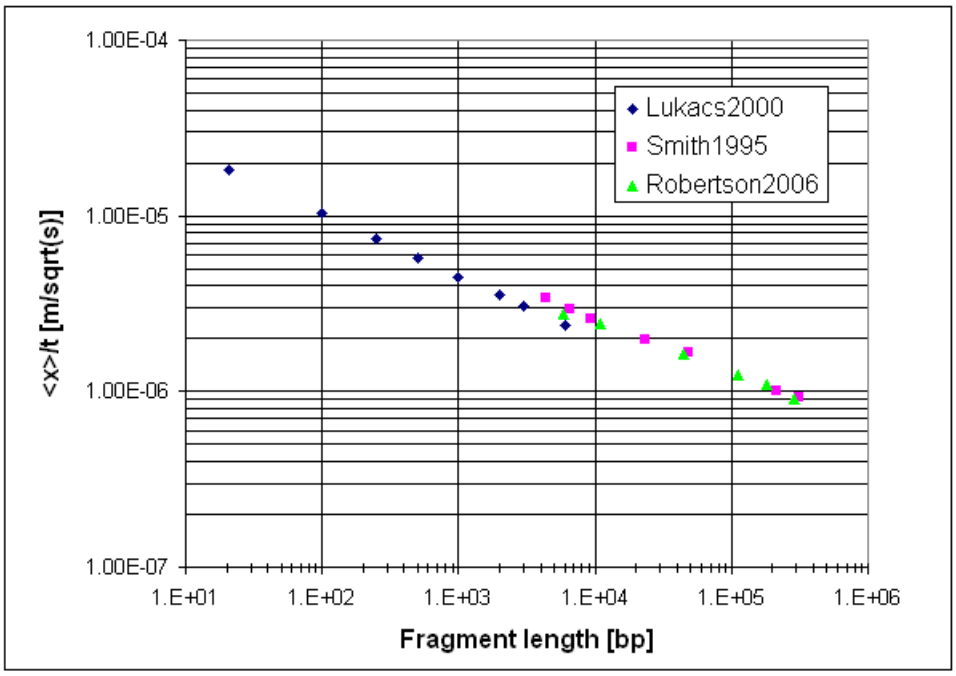


Figure 4.2: Average displacement for different duplex length calculated from Figure 4.1.

the value of $D = 0.9 \ldots 1.9 \times 10^{-11}\ m/s^2$ for an average length of $500 - 600$ bp.

4.1.3 Diffusion coefficient in solution and in cytoplasm

The measurement of drift velocities for different fragment length gives a picture about the mobility and the nature of dissipative forces in cyto- and nucleoplasm. It is important to note that non Einstein-Stokes forces give a different scaling law than Equation 4.1. In [34] these investigation have been done by means of spot photobleaching: The diffusion of larger DNA fragments in cytoplasm became remarkably slowed, with little or no diffusion for DNA $\geq 2000\ bp$. In nucleus, DNA fragments of all sizes were nearly immobile on a distance scale of $\approx 1 \mu m$. The immobilization of DNA in the nucleus is probably because of extensive DNA binding to nuclear components, including the positively

charged histones. An important result to this experiment comes from [159], which found that the nucleoplasm is filled by a cage-like network consisting of SatB1 proteins, which bind preferable with DNA regions with low melting temperatures. The SatB1 fixes many genes on specific chromosomal loci. These proteins seem to support the transfer of the DNA from the 'spaghetti-like' interphase to the well known condensed metaphase-chromosomes during cell cycle. The drift mobility of DNA fragments in cytoplasm is also thought to be an important determinant of the efficiency of DNA delivery in gene therapy and antisense oligonucleotide therapy. They showed a reduction of DNA diffusive rates in cytoplasm as DNA size increased beyond 1000 bp. The slowing of DNA diffusion might be due to a combination of binding and crowding effects. This findings indicate that diffusion of DNA can be a significant rate-limiting barrier in the cellular processing of plasmids and large DNA fragments.

4.1.4 Temperature behavior of diffusion

The temperature dependency of the diffusion coefficient according to Arrhenius equation is:

$$D = D_0 e^{-\frac{E_A}{RT}} \tag{4.6}$$

E_A is the activation energy (considered to be temperature independent) and R is the gas constant $(8.3144 J/K \times mol)$. In contradiction to hybridization, analysis of the hairpin opening/closing of 40 base oligonucleotides in [35] showed non-Arrhenius kinetics. There is a strong influence of secondary structure effects on hybridization kinetics. On surface bound oligonucleotides the main influence on kinetics comes from the transport-diffusion limitations.

4.2 Physics of diffusion-reaction systems

The hybridization of nucleic acid targets to surface immobilized oligonucleotides is an interplay of transport and reaction phenomena. While the time constants for solution based hybridization are in the range of seconds to minutes, hybridization to surface immobilized oligonucleotides takes hours. Depending on agitation, temperature, hybridization buffer and surface effects different limitations are important.

4.2.1 Is DNA hybridization diffusion or reaction controlled?

Diffusion theory is important for understanding the rates of bimolecular reactions in solution. If the activation energy of the reaction is fairly small, the rate may be limited by the diffusion encounter frequency of the reactants. An estimation of the diffusion constant using the Stokes-Einstein relation is:

$$D = \frac{2N_{av}kT}{3000\eta}(1+\frac{1}{\alpha})1+\alpha \tag{4.7}$$

$$\alpha \equiv -r_A/R_B \tag{4.8}$$

If the two reactants are of equal size then $\alpha = 1, R = N_{av}k$ and $k = 8RT/3000\eta$. This gives for water at 20 ℃ a reaction rate of $6,5 \times 10^9\ mol^{-1}sec^{-1}$ [90]. There are numerous refinements to this simple theory. If only one fraction of the surface of A and/or B is reactive, then Equation 4.7 must be reduced by a factor related to the product of the reactive fraction of each surface and to the rotational diffusion coefficients of the partners. In case of electrically charged reactants or if they interact by other intermolecular forces, their approach will produce attractions or repulsions that will

accelerate or slow the reaction. Several methods deal with the calculation of the variable dielectric medium representing the polymer and solvent

It has been demonstrated that the rate of combination of gene regulatory proteins with their target base sequences is faster than can be accounted using the above rate constants. This has led to the view, that the kinetics of events such as repressor-promotor interaction are governed by 3D nonspecific binding to the DNA followed by effectively 1D sliding along the doule helix to the target site. This 'reduction of dimensionality' approach is also used for the description of receptor-ligand interactions on cell surfaces [4]. The effect of surface diffusion on hybridization kinetics is easily visible. Within our experiments a dependency of the initial hybridization kinetics from blocking procedure was observed. This effect is not included within this version of the algorithm. Early determination of hybridization reaction rates for short fragments in solution [209] showed that rates of duplex formation for homo- and heterocomplexes are practically diffusion controlled. The dissociation rates are decreasing with binding energy. During Hybridization to surface immobilized oligonucleotides a depletions of the hybridization solution is happening. A comparison of a typical experimental result with simulated kinetics is shown in 4.3. There are two hybridization regimes visible, immediately after the start of hybridization nucleic acids from the surrounding of the spot are hybridizing.This is important especially for low expressed genes and a large number of complementary spots. After several minutes the hybridization is slowing down.

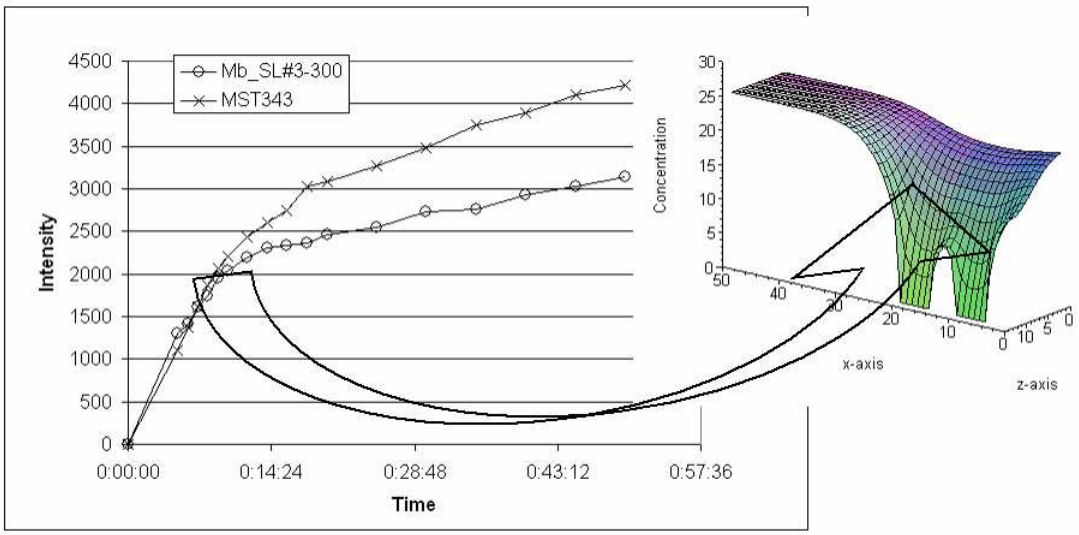


Figure 4.3: Depletion of hybridization solution during the hybridization process. The left diagram gives the hybridization kinetics of two probes from the methanotroph array [19]. The right figure gives the simulation result of the Crank-Nicolson algorithm. The depletion area is good visible.

4.3 Solution based results for hybridization

In the absence of secondary structure, two stretches of bases, termed nucleation regions, cooperate with one another by a looping mechanism to nucleate hybridization. At temperatures well below (T_m) hybridization occurred by a collision mechanism characterized by a positive energy of activation (E_a). Increasing temperature yields increased reaction rate. The rate limiting step is the collision of two strands to form a small complementary double helix. Once formed, the small double helix was

postulated to rapidly propagate by a zipping mechanism to complete double helix formation [134]. In contrast, at temperatures near T_m hybridization was shown to take place by a nucleation mechanism characterized by a negative activation energy in which two single strands were in rapid equilibrium with a small double helix. If sufficiently stable, the small duplex was postulated to rapidly propagate by a zipping mechanism to complete double helix formation [123]. The formation and dissociation of the double helix is described in solution

$$A_{ss}^{D} + B_{ss}^{D} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} (A \bullet B)_{ds}^{D} \tag{4.9}$$

The reaction can proceed in both directions and this is reversible. When the reaction reaches the equilibrium, both forward and reverse reaction rates are equal. The forward rate constant k_f depends on DNA length, sequence contest and salt concentration [90, 227]

$$k_f = \frac{k'_N \sqrt{L_s}}{N} \tag{4.10}$$

where L_s is the length of the shortest strand participating in the duplex formation; N is the total number of base pairs present; and k'_N is the nucleation rate constant, estimated to be $(4.35 \log_{10}[Na^+] + 3.5) \times 10^5$ for $0.2M \leq [Na^+] \leq 4.0M$ The reverse rate constant k_r is very sensitive to DNA length and sequence

$$k_r = k_f e^{\Delta G^0 / RT} \tag{4.11}$$

With the gas constant R and the hybridization temperature T. The rate of formation of this second order reaction is

$$\frac{d[B]}{dt} = k_f[A][B] - k_r[C] \tag{4.12}$$

When the concentration of one single-stranded reactant is larger and does not change, hybridization is pseudo-first order and integration yields

$$1 - \frac{[A]}{[A_0]} = (1 - K')(1 - e^{-(k_f C_0 + k_r)t}) \tag{4.13}$$

where

$$K' = \frac{k_2}{C_0 + k_f + k_r} \tag{4.14}$$

and $[A_0]$ is the total concentration of $A1$ $1 - \frac{[A]}{[A_0]}$ is the fraction of hybrid at the time t. When k_2 is negligible the equation reduces to the well-known $C_0 t$ Equation 4.15 which assumes that hybridization is an irreversible process:

$$1 - \frac{[A]}{[A_0]} = 1 - e^{-(k_1 C_0)t} \tag{4.15}$$

When $k_r t$ is large Equation 4.13 reduces to the Langmuir expression for equilibrium binding:

$$1 - \frac{[A]}{[A_0]} = (1 - K') = \frac{K_{eq} C_0}{K_{eq} C_0 + 1} \tag{4.16}$$

with $K_{eq} = k_f / k_r$. Causes for increased dissociation rates might be increased temperature, electrostatic repulsion, deletions or mutations.

4.4 Numerical solution of diffusion-reaction equation

4.4.1 Introduction

Goal of this section is to find a mathematical description for the modeling of hybridization to surface immobilized oligonucleotides. In Figure 4.3 and Section 5.3 some results for hybridization kinetics are shown. From results without agitation (Figure 4.3) it is visible that there are two regimes: A steep increase for the first at the beginning of the hybridization and a much slower increase afterwards. There are several factors which can slow down the reaction kinetics:

- Electrostatic repulsion due to accumulation of negative charge on the spot.
- Diffusion limitation due to depletion of target molecules.
- Inreasing backward (=denaturation) reaction with higher target coverage.
- Competitive hybridization.

All this effects have been the subject of research. For our long oligo microarray system it was found that the influence of electrostatic repulsion is negligible, The melting Temperature T_m was not influenced from the probe spotting concentration, see Section 3.8. An significant influence of increased denaturation was not observed. Prior to melting analysis 10 image acquisitions in washing solution have been made and there was no intensity reduction visible. The results of a long term experiment in 0.2xSSC is visible in

The influence of competitive hybridization is described in Section 6.5. While giving a lot of information on the physical-chemistry of nucleic acid hybridization there is no influence on transport properties. Thus it was not included within the numerical model. Nevertheless, a more advanced algorithm could give insight into the complex transport-reaction behavior of competitive nucleic acid hybridization.

4.4.2 The diffusion equation

Exact solutions of the diffusion equation

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{4.17}$$

are restricted to simple geometries and to constant diffusion properties. For numerical solutions there are several methods:

1. Finite differences method (FDM)
2. Finite elements method (FEM)
3. Monte Carlo algorithms
4. Random Walk algorithms

Due to the possibility of the implementation of derivative (Neumann) boundary condition and the pseudo 2-dimensional character of the system an algorithm based on FDM was implemented. Crank-Nicolson algorithm is known to provide stable solutions to diffusion and heat equation. The realization was done in Maple.

4.4.3 Finite-difference algorithm for diffusion equation

The range of space and time coordinates is divided into equal intervals δx, δy and δt, so that the x,y,t continuum is covered by a grid of rectangles. The indexes i and j denotes space (x, y), the upper index n time (t) coordinates. Capital letters denote the continuous value of the concentration, small letter the discrete value of the FEM mesh. The calculation in space are explicitly done for the x-coordinates, the y-coordinates are analogous. Using Taylor's series in the t direction but keeping x constant gives

$$c_{i,j,n+1} = c_{i,j,n} + \delta t \left(\frac{\partial c}{\partial t}\right)_{i,j,n} + \frac{1}{2}(\delta t)^2 \left(\frac{\partial^2 c}{\partial t^2}\right)_{i,j,n} \tag{4.18}$$

from which an approximate expression for the first order derivation follows immediately

$$\left(\frac{\partial c}{\partial t}\right)_{i,j,n} = \frac{c_{i,j,n+1} - c_{i,j,n}}{\delta t} + O(\delta t) \tag{4.19}$$

where $O(\delta t)$ signifies that the leading term to have been neglected is of the order of δt: Similarly, by applying Taylor's series in the x direction, keeping y and t constant

$$c_{i+1,j,n} = c_{i,j,n} + \delta x \left(\frac{\partial c}{\partial x}\right)_{i,j,n} + \frac{1}{2}(\delta x)^2 \left(\frac{\partial^2 c}{\partial x^2}\right)_{i,j,n} \tag{4.20}$$

$$c_{i-1,j,n} = c_{i,j,n} - \delta x \left(\frac{\partial c}{\partial x}\right)_{i,j,n} + \frac{1}{2}(\delta x)^2 \left(\frac{\partial^2 c}{\partial x^2}\right)_{i,j,n} \tag{4.21}$$

On adding we find an approximation for the second order derivation in x-direction

$$\left(\frac{\partial^2 c}{\partial x^2}\right)_{i,j,n} = \frac{c_{i+1,j,n} - 2c_{i,j,n} + c_{i-1,j,n}}{(\delta x)^2} + O(\delta x^2) \tag{4.22}$$

By substituting in the diffusion Equation 4.17 and neglecting the error terms we find after slight re-arrangement

$$c_{i,j+1,n} = c_{i,n,j} + D\frac{\delta t}{(\delta x)^2}(c_{i-1,j,n} - 2c_{i,j,n} + c_{i+1,j,n}) \tag{4.23}$$

and with $a \equiv D\delta t/(\delta x)^2$

$$c_{i,j+1,n} = (1 - 2a)c_{i,j,n} + a(c_{i-1,j,n} - 2 + c_{i+1,j,n}) \tag{4.24}$$

one gets an explicit expression for the calculation of the concentrations at the time $j + 1$ from the concentration values at j. Starting from an given initial distribution it is possible with this equation to solve the discrete equations for the transient behavior of the system. This simple algorithm is called FTCS, FT for 'forward time' and CS for 'centered space'. The stability condition for this algorithm is

$$2D\frac{\delta t}{(\delta x)^2} \leq 1 \tag{4.25}$$

Using the characteristic time τ necessary for the diffusive movement across δx

$$\tau = \frac{(\delta x)^2}{2D} \tag{4.26}$$

the necessary condition for stability of this algorithm becomes

$$\delta t \leq \tau \tag{4.27}$$

Reduction of the spatial grid size δx yields a second order decrease of τ. Thus this simple and stable algorithm is not efficient.

4.4.4 Crank-Nicolson algorithm in one dimension

The Crank-Nicolson algorithm overcomes this weakness by replacing $\delta^2 c/\delta x^2$ with the mean of its finite-difference representation on the nth and $(n+1)$th time row. The discretization of the 1d diffusion equation according to Crank-Nicolson is

$$\frac{(c_{i,n+1} - c_{i,n})}{\delta t} = \tag{4.28}$$

$$= \frac{D}{2(\delta x)^2}\left[(c_{i+1,n+1} - 2c_{i+1,n+1} + c_{i,n+1}) + (c_{i,n} - 2c_{i-1,n} + c_{i,n})\right]$$

and with

$$a \equiv D\delta t/2(\delta x)^2 \tag{4.29}$$

one gets a system of equation for the Crank-Nicolson coefficients

$$-ac_{i+1,n+1} + (1+2a)c_{i,n+1} - ac_{i-1,n+1} = ac_{i+1,n} + (1-2a)c_{i,n} + ac_{i-1,n} \tag{4.30}$$

This tridiagonal problem can be solved efficiently by using tridiagonal matrix algorithms.

4.4.5 Crank-Nicolson algorithm in two dimension

To separate the spatial from the time coordinate the time coordinate is denoted with the index n, i.e.

$$c_{i,j,n} \tag{4.31}$$

is n the discrete time coordinate while i and j are the discrete spatial coordinates. With this notation the Crank-Nicolson method for the diffusion equation in two spatial dimensions becomes

$$\frac{c_{i,j,n+1} - c_{i,j,n}}{\delta t} = \frac{1}{2}D\left\{\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right)_{i,j,n} + \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right)_{i,j,n+1}\right\} \tag{4.32}$$

with

$$\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right)_{i,j,n} =$$

$$= \frac{c_{i+1,j,n} - 2c_{i,j,n} + c_{i-1,j,n}}{(\delta x)^2} + \frac{c_{i,j+1,n} - 2c_{i,j,n} + c_{i,j-1,n}}{(\delta y)^2} \tag{4.33}$$

Solving this equation gives a system of banddiagonal matriy rather than tridiagonal ones.

4.4.6 Crank-Nicolson in cylinder coordinates

While it is very time- and memory consuming to solve the diffusion equation 4.17 in 3 spatial coordinates it is possible to reduce the amount of calculation by using the inherent symmetry of the problem for a symmetric setup. Escpecially for basic investigations on hybridization kinetics and the influence of experimental parameters this is a sufficient solution. By considering the transformation into cylinder coordinates

$$x = r\cos\theta \tag{4.34}$$

$$y = r\sin\theta \tag{4.35}$$

$$z = z \tag{4.36}$$

the equation for diffusion in cylinder coordinates is obtained

$$\frac{\partial C}{\partial t} = \frac{1}{r}\left\{\frac{\partial}{\partial r}\left(rD\frac{\partial C}{\partial r}\right) + \frac{\partial}{\partial \theta}\left(\frac{D}{r}\frac{\partial C}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(rD\frac{\partial C}{\partial z}\right)\right\} \tag{4.37}$$

Using radial symmetry and the fact that D is a scalar constant

$$\frac{\partial C}{\partial \theta} = 0 \tag{4.38}$$

one gets after some rearrangement for Equation 4.17 in cylinder coordinates

$$\frac{\partial C}{\partial t} = D\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial z^2}\right) \tag{4.39}$$

for numerical solution it is easier to evaluate the first expression

$$\frac{\partial C}{\partial t} = D\left(\frac{1}{r}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2} + \frac{\partial^2 C}{\partial z^2}\right) \tag{4.40}$$

compared to Equation 4.32 shows that the second order term are equal for the coordinates r and z while there is an additional first order term

$$\frac{1}{r}\frac{\partial C}{\partial r} \tag{4.41}$$

The finite-difference approximations corresponding for this equation are

$$\frac{\partial^2 C}{\partial r^2} \approxeq \frac{c_{i+1,j,n} - 2c_{i,j,n} + c_{i-1,j,n}}{(\delta R)^2} \tag{4.42}$$

$$\frac{1}{r}\frac{\partial C}{\partial r} \approxeq \frac{c_{i+1,j,n} - c_{i-1,j,n}}{2i(\delta R)^2} \tag{4.43}$$

$$\frac{\partial^2 C}{\partial z^2} \approxeq \frac{c_{i,j+1,n} - 2c_{i,j,n} + c_{i,j-1,n}}{(\delta z)^2} \tag{4.44}$$

with

$$r \approxeq i\delta R \tag{4.45}$$

δR is the mesh spacing for r, i is the coordinate in r-direction. Using Equation 4.32 a Crank-Nicolson solution of the diffusion equation in cylinder coordinates can be derived:

$$\frac{c_{i,j,n+1} - c_{i,j,n}}{\delta t} =$$

$$\frac{1}{2}D\left\{\left(\frac{\partial^2 C}{\partial r^2} + \frac{1}{R}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2}\right)_{i,j,n} + \left(\frac{\partial^2 C}{\partial r^2} + \frac{1}{R}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2}\right)_{i,j,n+1}\right\} \tag{4.46}$$

with

$$\begin{aligned}\left(\frac{\partial^2 C}{\partial r^2} + \frac{1}{r}\frac{\partial C}{\partial r} + \frac{\partial^2 C}{\partial r^2}\right)_{i,j,n} \approxeq\ & \frac{c_{i+1,j,n} - 2c_{i,j,n} + c_{i-1,j,n}}{(\delta R)^2} \\ & + \frac{c_{i+1,j,n} - c_{i-1,j,n}}{2i(\delta R)^2} \\ & + \frac{c_{i,j+1,n} - 2c_{i,j,n} + c_{i,j-1,n}}{(\delta z)^2}\end{aligned} \tag{4.47}$$

It is visible that the second derivative spatial terms are similar compared to Equation 4.32. There is an additional first derivative term in the Crank-Nicolson equation of the diffusion equation in cylinder coordinates. Thus it is necessary to include the calculation of the first derivative in the algorithm. For basic investigation the 2d-solution is sufficient.

4.4.7 Modeling of denaturation

With the Crank-Nicolson based modeling it was possible to give a quantitative explanation many experimental observations. Hybridization kinetics slows down remarkable due to depletion effects of target concentration, see Figure 4.3. Simulation results showed that in narrow structures the high concentration of the hybridization sites on the probe spots gives an significant depletion of target molecules in the region above the spot. After several minutes the target molecules have to diffuse sideways to the spot, which slows down the overall hybridization reaction. From our experimental setup we could not give any influence of nucleic acid hybridization kinetics on overall reaction rate to the surface bound probes. But it was possible to give an accurate value for the target concentration after some reference measurement, see Appendix **??**. One possibility to learn more about the reaction constants of nucleic acids is competitive hybridization, see 6. Figure 4.4 gives a typical 2d result of

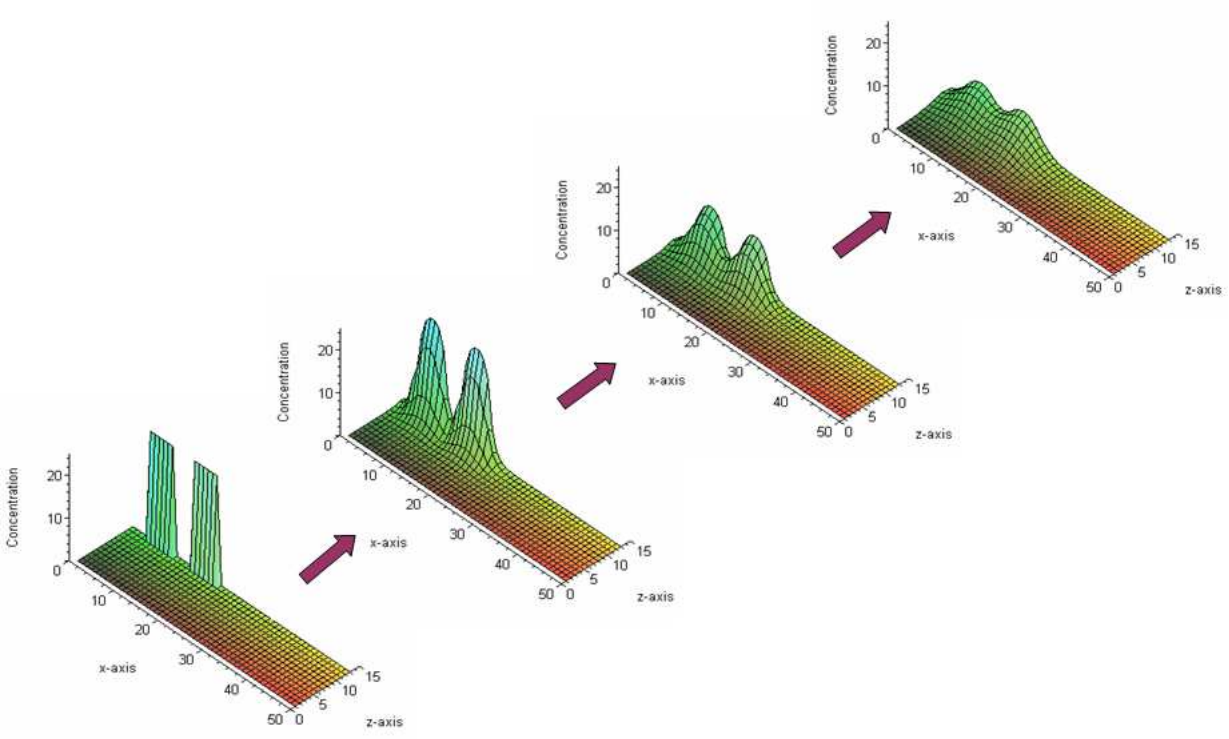


Figure 4.4: Typical results for denaturation kinetics for two model spots in 2d. The time difference between the snapshots is 150 s. It is visible that without agitation there is a region of high background intensity forming. This prevents an accurate quantification of the melting curve. Thus an effective agitation was necessary within our experimental setup.

denaturation kinetics. Due to the transport limitation of diffusion the background intensity of the spot is increasing, the fluorescent targets are not disappearing fast enough. This effect hides the denaturation of the target molecules from the spot, thus preventing the measurement of an accurate melting curve. With this setup an effective agitation would be necessary to speed up the transport

of the denaturated and fluorophore labeled molecules. Another possibility to overcome this problem would be an advanced labeling concept with intercalating fluorophores or FRET based methods. Slides with interference of evanescence or SPR based amplification of surface based fluorphores could also improve the visibility of the melting transition. Nevertheless overcoming the transport-diffusion limitations is a key issue for high resolution surface-based melting analysis.

4.5 Conclusion

For an interpretation of the results of real-time measurement an understanding of transport limitation is crucial. With the Crank-Nicolson based algorithm it was possible to show that without efficient evanescent illumination (or other effects which enhance the surface intensity compared to the background) the specificity improvement of melting analysis is limited compared to endpoint analysis. The usage of slides with interference based enhancement of fluorescent intensity is another possibility. The target depletion is visible in the solutions of the algorithm and is observed in the kinetics of real time experiments without agitation. There are several different approaches to numerical solutions of diffusion-reaction models, for the understanding of the transport properties of surface based melting analysis this algorithm was used due to limitation in computational power.

Chapter 5

Algorithms and modeling tools

5.1 DNA hybridization as a Fermi-Dirac system

5.1.1 Introduction

FermiDirac Statistics describes the energies of single particles in a system comprising many identical particles that obey the Pauli exclusion principle. It is named after Enrico Fermi and Paul Dirac, who each discovered it independently [54, 46]. The FDS gives the average occupation of a single-particle state for a system of noninteracting identical fermions.

$$\overline{n}_i = \frac{1}{e^{(\epsilon_i - \mu)/kT} + 1} \tag{5.1}$$

where k is the Boltzmann constant, T the absolute temperature, ϵ_i is the energy of the single particle state i, and μ is the chemical potential. Another assumption for FDS is that the system is large enough so that adding one more fermion to the system ha negligible effects on μ. The average number of fermion can be found by multiplying the FDS distribution $\overline{n}_i$ in Equation 5.1 b¡ the degeneracy g_i (i.e. the number of states with the energy ϵ_i

$$\begin{aligned} \overline{n}(\epsilon_i) &= g_i \overline{n}_i \\ &= \frac{g_i}{e^{(\epsilon_i - \mu)/kT} + 1} \end{aligned} \tag{5.2}$$

Before the introduction of FDS the understanding of electron behavior was difficult due to seemingly contradictory phenomena. It was difficult to understand why electron in a metal can move freely to conduct electric current while their contribution to the specific heat is negligible. It was believed that each electron contributed to specific heat $1/2kT$. Contrary to current opinion, Fermi did not introduce his quantum statistics in order to explain metal conductivity. His aim was to obtain a correct derivation of Sackur and Tetrode's Formula which gives the entropy of a perfect gas:

$$S = k_B \ln \left[\left(\frac{2\pi^{3N/2}}{(3N/2 - 1)!} \right) (2mE)^{3N/2} V^N \frac{1}{N!} (2\pi\hbar)^{-3N/2} \right] \tag{5.3}$$

for the entropy of a perfect monoatomic gas [54, 10]. In this representation E is the energy of the N molecules and $N!^{-1}$ reflects the fact that gas molecules are indistinguishable, Dirac discovered this distribution independently [46]. FDS was applied in different field of physics like the collapse of

Table 5.1: Comparison of the hybridization to surface bound oligos with free electron model

DNA hybridization	**G tract**
No interaction between the target molecules	No electron-electron interaction
Two state model	
All probe sites have the same energy	Unmethylated major groove
Screening of the surface charges due to ionic buffer	Effective mass due to periodic interaction
Two state model	Electronic bands with discrete occupation number
Very slow base pair dynamics	Very rapid base pair dynamics

a white dwarf [59] or the theory of electrons in metals [180]. Some research was done to derive the partition function for physical adsorption of a quantum gas to a solid surface in two dimensions [99] and for the distribution function of adsorption of gases to homogeneous solid surfaces [146] based on FDS. Recently the application of the FDS occupation distribution to individual adsorption sites proved to be a powerful description to predict adsorptive equilibria [206]. [212] gives a good overview about modern methods for the description of protein-DNA interactions. In a quantum mechanical notation the solution of the Schrödinger equation for a system of fermions is antisymmetric

$$\psi(\ldots, x_i, \ldots, x_j, \ldots; t) \equiv -\psi(\ldots, x_j, \ldots, x_i, \ldots; t) \tag{5.4}$$

It follows immediately that all the quantum numbers have to be different, otherwise the function would be equal to their negative value. [33, 122] give an explanation of the linear temperature behavior of the Cy-dyes.

5.1.2 Fermi-Dirac algorithm

While it is not so obvious for solution based melting experiments there is a clear analogy between surface bound oligo hybridization and solid-sate physic Fermi-Dirac applications. In Sommerfeld model [180] the valence electrons are assumed to be completely detached from the ionic molecules, forming an ideal electron gas with no electron-electron interaction. In reality the electron-electron interaction is weak because of the electrostatic shielding effect. A quantum mechanical treatment of this model gives that an unbound electron moves in a periodic potential as a free electron in vacuum with an effective mass. In DNA hybridization to surface bound single stranded probe molecules the single stranded target moves freely in the solution without mutual interaction. It is interesting to note that the extremely high charge of every DNA-strand (1 e^- for every nucleotide) is also shielded by counterions. For a rough numerical estimation of the shielding radius with formula 2.13 from [41] one gets for our 0.9 M solution a Debye length of 0.32 nm. This result is supported by our finding that there is no measurable dependency of the T_m from probe concentration within our experimental conditions [19]. Without shielding the higher probe density should give larger electrostatic repulsion and thus a lower melting temperature [223, 118].

5.1.3 Experimental data

It is interesting that the good description of the melting transition allows the extrapolation of experimental measurement below T_m to higher temperatures. This opens the possibility to extract relevant information from very few experimental results. Figure 5.3 shows the result for several amplitudes and melting temperatures.

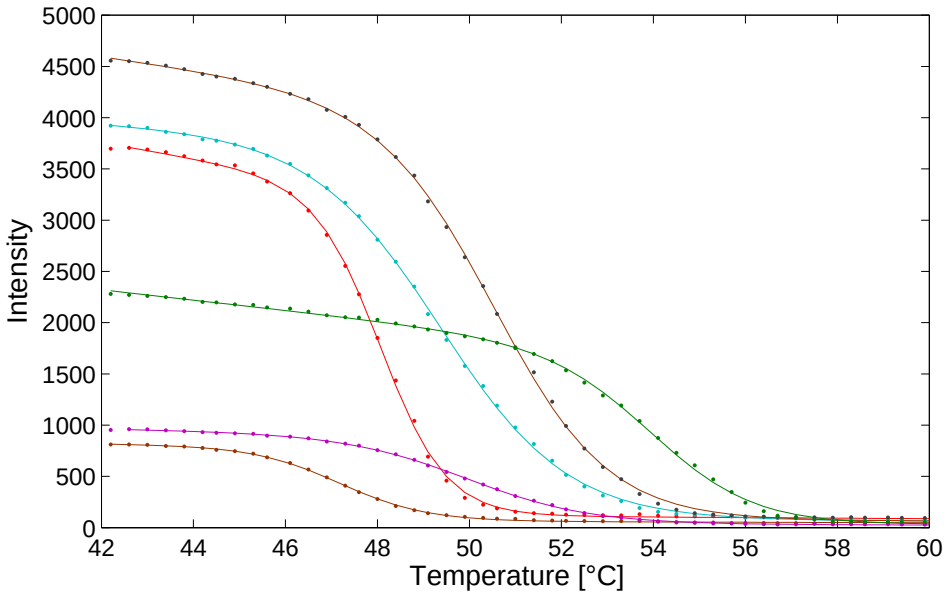


Figure 5.1: Comparison of experimental melting curves (points) and the result of the Fermi-Dirac based fit (solid line). The different amplitudes and melting points are very good approximated.

Table 5.2: Parameters for Modeling in ChipCheckII

Target Concentration	$1 * 10^{-9}$
Probe Concentration	$4 * 10^{-15}$
Volume	$2 * 10^{-4}$ l
Temperature	42 ℃
Salt concentration	$[Na^+] = 0.9$ M; $[Mg^{++}]$=0 M
Type of duplexes formed	DNA-DNA
ID in ChipCheckII	KG5507

5.2 Modeling with ChipCheckII

While there are several algorithms available for the prediction of hybridization results, ChipCheckII [179] includes the target depletion due to hybridization and is compatible to existing NN-algorithms [127]. For the used parameters see Table 5.2. A partition function based approach can be found in [45].

The algorithm is based on the assumption:

$$probe_i + target_j \underset{k_{ij}^d}{\overset{k_{ij}^h}{\rightleftharpoons}} probe_i \bullet target_j \tag{5.5}$$

A description of the reaction equilibrium constant K_{ij} according to the law of mass action would be [179]

$$K_{ij} = \frac{x_{ij}}{x_i c_j} \tag{5.6}$$

with the equilibrium target concentration c_j and the mole fraction of single stranded probes x_i. From 5.6 it follows:

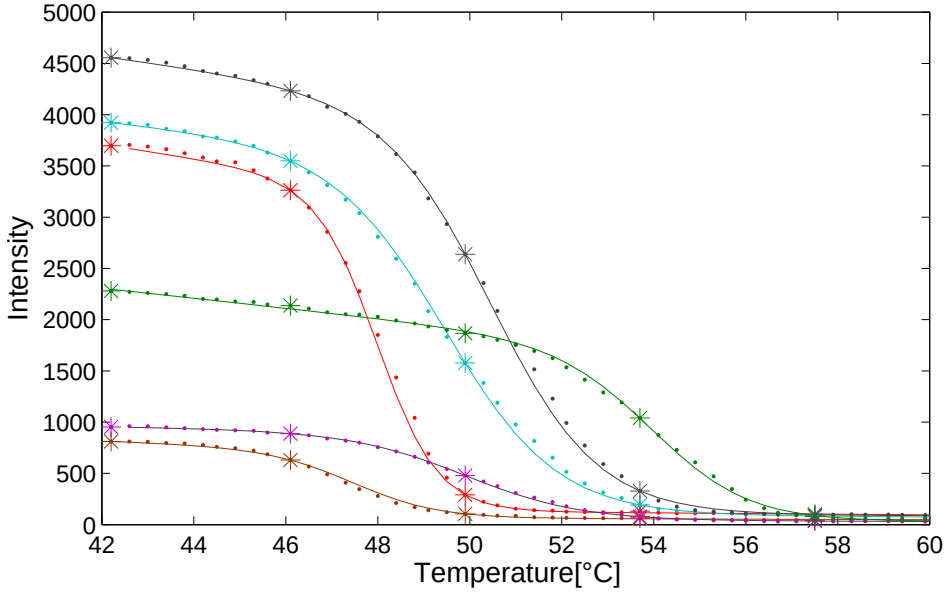


Figure 5.2: Comparison of experimental melting curves (points) and the result of the Fermi-Dirac based fit with a reduced number of experimental values (solid line). Only the experimental values which are denoted with a cross have been used for the fit. The quality of the fit is still very good.

$$x_{ij} = K_{ij} x_i c_j \tag{5.7}$$

Here it is visible that this approach does not give an Langmuir like dependency of c_j but a linear. A Langmuir like equation for x_{ij} would be

$$x_{ij} = \frac{K_{ij} x_i c_j}{1 + K_{ij} c_j} \tag{5.8}$$

Nevertheless the following equations of [179] are valid:

$$\sum_j x_{ij} + x_i = 1 \tag{5.9}$$

which is representing the mass balance for probe P_i, the sum of all hybridized and single stranded probes is one. It is important to note that x_i and x_{ij} has to be normalized with the total number of probe strands n_i to get the number in moles. The mass balance for the target T_j is represented via

$$\sum_i n_i x_{ij} + V c_j = V c_j^0 \tag{5.10}$$

with the initial concentration c_j^0. Assuming that the free binding energy is

$$\Delta G_{ij}^0 = \Delta H_{ij}^0 - T \Delta S^0 ij = -RT \ln K_{ij} \tag{5.11}$$

it is possible to calculate K_{ij} from thermodynamic data, either they are obtained by calculation or experiment:

$$K_{ij} = \exp\left(-\frac{\Delta G_{ij}^0}{RT}\right) \tag{5.12}$$

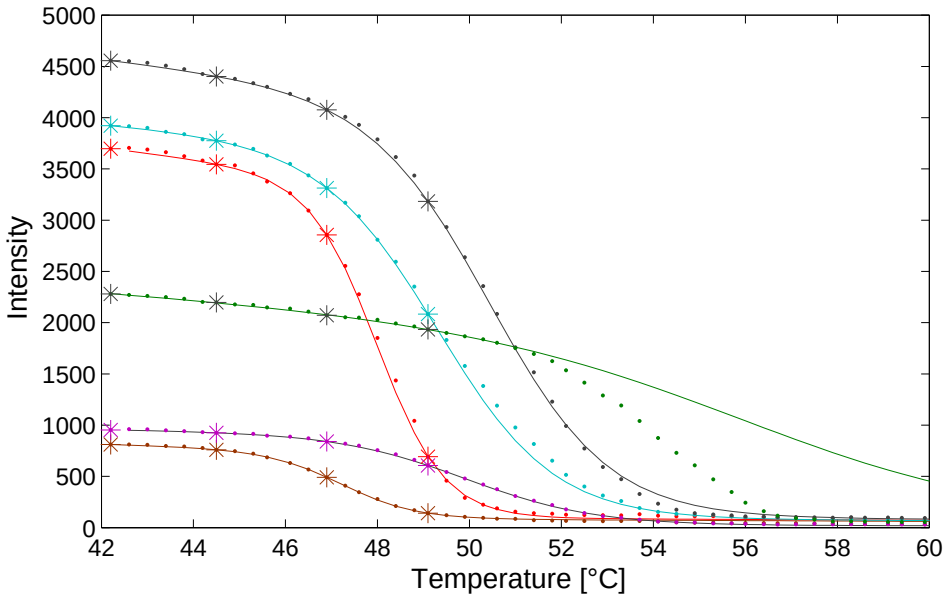


Figure 5.3: Comparison of experimental melting curves (points) and the result of the Fermi-Dirac based extrapolation with a reduced number of experimental values (points) taken at lower temperature. Only the experimental values which are denoted with a cross have been used for the fit. The quality of the extrapolation is quite good. Obviously it is not possible to extrapolate the green curve du to the missing experimental values during melting.

5.3 Publication: A model based algorithm for automated analysis of real-time kinetic and thermodynamic analysis on microarrays

S. Krainer, S. Fluch, M. Stierschneider, L. Bodrossy

Submitted to *Algorithms in molecular biology*

5.3.1 Abstract

Background

Real-time data acquisition of microarray hybridization kinetics and melting analysis has the potential to improve microarray technology in terms of quantitative analysis and specificity. Furthermore it generates valuable data for the understanding of nucleic acid hybridization. While there is progress in solving existing experimental challenges, an automatable algorithm for extraction of the relevant kinetic and thermodynamic parameters from the huge data sets generated by sequential time resolved measurements is still lacking. Here we describe an algorithm for automated analysis of kinetic and thermodynamic datasets based on a nonlinear regression to simple physico-chemical models.

Results

The verification of the newly developed algorithm was done with a set of experiments including short and long oligo systems, different types of nucleic acids (DNA, RNA and LNA) and competitive hybridization. With the new algorithm it is possible to reduce the kinetic data to 3 parameters and the

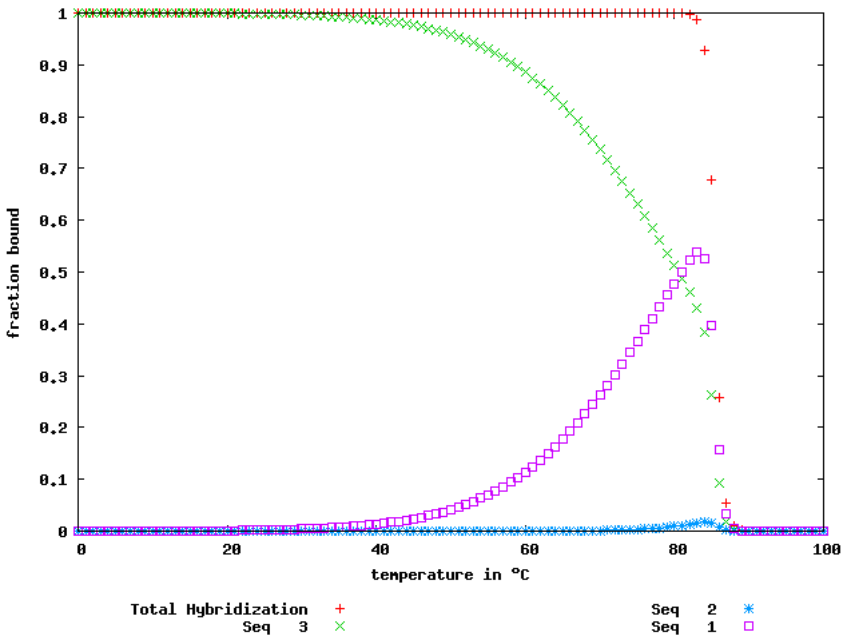


Figure 5.4: Fraction of bound probes versus hybridization temperature. Seq 1 and Seq 2 are the MM targets to 05x103 Potato and Pepper, Seq 3 the PM (Tobacco)

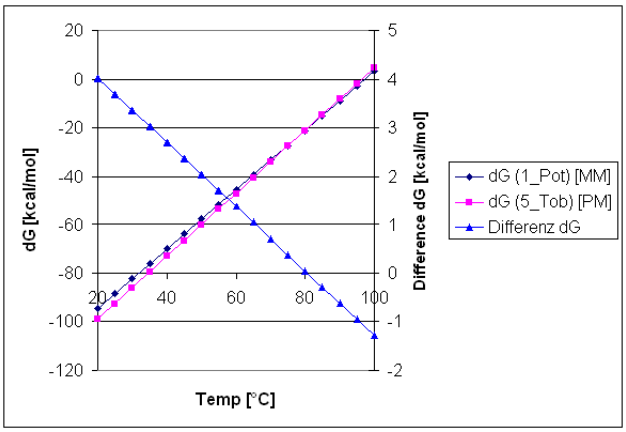


Figure 5.5: Free Energy dG for a perfect match and a mismatch probe target combination and the difference.It can be seen that above 80° the MM target has a more negative free energy (= stronger binding). This is due to the higher contribution of the entropic part, the entropy of the MM is smaller than of the PM (1226 vs 1292)

melting analysis data to 6 parameters. These parameters are characterizing either physico-chemical parameters of the probe-target combination or characteristic properties of the experimental system.

Conclusion

We could show that it is possible to reduce the huge amount of data obtained from kinetic and melting analysis to a small number of relevant parameters. The number of necessary experimental data is reduced significantly and an automated quality control gives the opportunity to discriminate between

specific and nonspecific hybridization. Together with experimental improvements with respect to throughput of microarray based real-time analysis this could open the door for the next generation of high-specificity microarray systems for quality critical applications.

5.3.2 Background

The currently used microarray technology, where signal intensity after stringent washing is a measure for the number of hybridized targets is already in a mature state. There is still an ongoing discussion on quality improvements of microarray experiments [49, 100, 157]. In general the measured intensity is [101]

$$I = I_{Sp}(c) + I_{Bg} \tag{5.13}$$

where $I_{Sp}(c)$ is the specific signal due to hybridization of the complementary probe-target duplex and I_{Bg} is the background due to nonspecific effects. I_{Sp} depends on the concentration c of the complementary strand in solution (target). For hybridization to surface immobilized nucleic acids there are several factors influencing the unspecific term I_{Bg} e.g. cross hybridization, optical effects, microarray surface composition and intra- and interarray process variations [150]. The discrimination of specific and nonspecific hybridization of similar targets is still a challenge in microarray probe design and data analysis [71, 111]. Further the development of advanced model based algorithms for background determination is still a matter of research [62, 86, 101, 143]. Currently the design of the probes is based on the assumption that melting temperature is related to the nearest neighbor (NN) enthalpy change ΔH^0 and the sum entropy changes ΔS^0 via

$$T_m = \frac{\Delta H^0}{R \ln (C_t/4) + \Delta S^0} \tag{5.14}$$

where R is the gas constant and C_t is the total molar concentration of strands. Experimental melting curves for the extraction of the NN parameters are generated from solution based UV absorption experiments. From the slope and intercept of plots vs it is possible to evaluate thermodynamic parameters from these solution based melting experiments. In contrast, the thermodynamics and kinetics of duplex formation on surface-immobilized DNA is influenced by many aditional factors and there is still no widely accepted description [157]. Microarray based real-time data acquisition during hybridization and melting [14, 22, 98, 125] tries to overcome these problems by recording the change of specific intensity signal $I_Sp(c)$ of 5.14 during the hybridization and denaturation relative to the unspecific signal I_Bg. Thus it gives the possibility to monitor melting of nucleic acid duplexes within the environment of the application. This way the fact that hybridization kinetics can provide information for the discrimination between specific and unspecific hybridization and for quantification of targets [13, 37, 97] can be exploited. The multiple measurements of intensity changes opens the door for advanced signal processing and noise reduction techniques, model-based fitting can solve the inherent weakness of single sampling measurement. The potentials and limitations of real-time microarrays are reviewed in [22]. Melting analysis is already being used for solution based high resolution melting (HRM) [36, 85, 221] and for advanced surface based methods with small probe numbers in SPR [223, 225], DASH [87, 183] or single molecule measurements [74]. Microarray based real-time analysis creates one image for each time/temperature point which results typically in 50-100 uncompressed .tif result files [125] for complete kinetic and thermodynamic analysis. Due to the noisy signal of surface based hybridization experiments low pass filtering and derivative based algorithms, which are used in HRM, are not suitable for microarray based methods. Low-pass filtering has the further disadvantage of increasing the number of necessary measurement for a given accuracy. In this paper we report on the development of a physical model based nonlinear regression algorithm for real-time microarray data analysis and the implementation within a Matlab programmed GUI.

Based on the experimental data of 40 experiments we used equations of physico-chemical systems with similar behavior as fit function for the nonlinear regression putting the focus on high-density microarray applications. The paper is organized as follows. In Methods we describe the physico-chemical background and the implementation of the algorithm. A short overview on design and preparation of the microarray system for different types of nucleic acids is presented as well because we found the optimization of the probe-target system as a crucial point in real-time hybridization experiments. In Results the application of the algorithm to 40 experiments with different nucleic acids is presented including the physico-chemical interpretation of the parameters used in this algorithm.

5.3.3 Methods

Extraction and purification of the actin gene

See 3.8.3

Probe design and preparation of the slides

See 3.8.3

Target preparation

See 3.8.3

Hybridization

See 3.8.3

Data preprocessing

Segmentation and quantification of the .tif files was done with the HybLive software. The "irregular shape" segmentation algorithm was used, all parameters were set to default values. Quantification was based on the average of the three technical replicates for every probe. Input files for the algorithms were comma separated value files.

Hybridization isotherms

Several promising approaches exist for a numerical description of hybridization kinetics of non agitated microarrays [64, 103]. Agitation is used to improve stringency and to overcome transport-diffusion limitation microarray systems [81, 3]. Currently it is not possible to give a numerical description of these complex hydrodynamic transport processes. Thus nonlinear regression to physico-chemical model systems was used. The intensity of kinetic measurements in microarray experiments is not necessarily monotonically increasing. In case of competitive hybridization a target with a high reaction rate but lower binding energy is replaced during hybridization process with targets of a higher binding energy [15]. Thus the equations of the kinetics of chemical reaction with an intermediate product have been used. It is characterized by the number n_{ss} of single stranded target in solution, n_{us} for the unspecific target hybridized to the probe and n_s for the specific target. For competitive hybridization the two constants k_h and k_c give the reaction rates of the two consecutive steps of hybridization and competitive displacement

$$n_{ss} \underset{0}{\overset{k_h}{\rightleftharpoons}} n_{us} \underset{0}{\overset{k_c}{\rightleftharpoons}} n_s \tag{5.15}$$

The rate laws of this reaction with the target concentration C_h^T and C_c^T are

$$\frac{dn_{ss}}{dt} = -C_h^T k_h n_{ss} \tag{5.16}$$

$$\frac{dn_{us}}{dt} = C_h^T k_h n_{ss} - C_c^T k_c n_{us} \tag{5.17}$$

The solution of these equations for the competitive replaced target n_{us} with the initial concentration n_{ss}^0 is

$$I_{us} \propto n_{us} = n_{ss}^0 \frac{C_h^T k_h}{C_c^T k_c - C_h^T k_h} \left(e^{-C_h^T k_h t} - e^{-C_c^T k_c t} \right) \tag{5.18}$$

I_{us} is the intensity of the target with the higher reaction rate and the lower binding affinity. In case of non competitive hybridization $k_c = 0$ the formula is still valid. In geneneral bi-exponential fits were reported to be a good approximation for the kinetic of surface hybridization reactions [92, 74, 190].

Analysis of melting curves

There are several algorithms for predicting melting temperature and thermodynamic parameters in solution [16, 127, 173, 179]. Moreover, due to the influence of surface related effects on nucleic acid thermodynamics like steric hindrance, entropic restrictions, electrostatic repulsion and the high probe density there is no accepted numerical description for the complex melting transition of surface bound oligos. To develop an automated algorithm we evaluated two approaches: a derivative based analysis similar to HRM analysis used in solution based assays [85, 161] with two different filtering algorithms and a physical model based fitting. To improve the specificity compared to that of existing endpoint based microarray systems it is necessary to achieve an accuracy for melting point analysis < 1 ℃. In general low pass filtering of experimental data increases the number of necessary measurements for a given accuracy. To evaluate the limits of different derivative based algorithms, a simple digital lowpass filtering ("moving average") with different window size and a Savitzky-Golay [175] based smoothing filter was used. Savitzky-Golay is a powerful technique for filtering complex noisy functions in signal processing because it can be adjusted with several parameters to preserve features of the signal function. Reference values for the melting temperature were the values of the Hyblive normalization method. The fitting of the melting curves was based on a convolution of linear functions for the description of the temperature behavior of the dye and an exponential distribution function.

$$I(T) = \frac{k_a (k_{dye} - T)}{e^{k_b (T - k_T m)} + 1} - k_{HT} + k_{const} \tag{5.19}$$

with the intensity $I(t)$ and the fitting parameters parameters k_a, k_b, k_{dye}, k_{Ht}, k_{Tm} and k_{const}. k_a is related to the amount of hybridized target and k_b to the slope of the melting curve at melting temperature (Tm). The larger k_b the steeper is the melting transition. In general, k_b is larger for longer oligos. k_{dye} considers the linear temperature behavior of the dye [122]. The term $-k_{HT}+k_{const}$ is a first order correction considering the remaining fluorescent intensity at temperatures above the melting of the nucleic acids due to unspecific binding of fluorophores to the probe spots. The whole nonlinear behavior of the system is described by the exponential function with the two parameters k_{Tm} and k_b. This term is similar to Fermi-Dirac distribution:

$$\langle n(E_i) \rangle = \frac{1}{e^{\beta(E_i - \mu)} + 1} \tag{5.20}$$

which gives the distribution of single particles in a many particle system of identical fermions. The usefulness of the Fermi-Dirac statistic for the description of surface adsorption is already proven [146, 206]. Another advantage of this method is that the parameter k_{Tm}, which is adequate to the single-particle state energy in the Fermi-Dirac statistics, gives the melting temperature straightforward.

Theoretical results and sequence information

Two NN algorithms, UNAFOLD [127] and Meltsim [16] have been used. The influence of formamide and of surface effects is not included in these algorithms, thus we did only relative comparisons of the melting temperatures. Additionally a T_m calculation based on a simple formula [219] was done. The results have been generated and stored automatically based on fasta files and can be used for correlation analysis with a graphical user interface (GUI). Mismatch tables including position and type of mismatch were created automatically with Matlab and stored within the GUI for user-friendly generation of correlations with experimental data. Further the important correlation of sequence information (content of different bases, type and position of mismatches,. . .) with the experimental and theoretical data is automated within the GUI. The results are written into a Matlab data structure for easy accessibility.

5.3.4 Results and discussion

A flow diagram of the implementation of the algorithm is shown in Figure 5.14.

We did 10 initial measurements which generated the basic data set. The algorithm was applied to the 40 following experiments. The analysis of a typical experiment with data from more than 100 .tif files need about 4 min on a standard PC (Intel Pentium 1,6 GHz, 2 GB RAM). A binning according to the the Goodness of Fit (GoF) parameter was done. Melting analysis showed extreme good correlation between different measurements, irregularities during the experiment were immediately visible. Within our measurements no significant impact of photobleaching was observed. The fitting was done based on "mean-background" values of the intensity. For the nonlinear regression the "Trust-region" method was used, This method is suited for solving nonlinear problems efficiently and is more robust compared to e.g. "Gauss-Newton". The lower and upper boundaries for hybridization kinetics are: $k_h\,[0,10]$, k_c : $[0,10]$, A : $[0,10]$ and for melting curve analysis k_{Cy} : $[45,1000]$, k_{HT} : $[0,0.1]$, k_a : $[0,1]$, k_b : $[0.2,2]$, k_{const} : $[0,1]$, k_{Tm} : $[45,85]$. The starting values of the algorithm for the kinetic analysis are: $k_h^0 = 0.01$, $k_c^0 = 0$, $A^0 = 1$ and for melting curve analysis: $k_{Cy}^0 = 100$, $k_{HT}^0 = 0$, $k_a^0 = 1$, $k_b^0 = 1$, $k_{const}^0 = 0$, $k_{Tm}^0 = 50$. The parameter for k_{Tm}^0 was adjusted to the expected melting temperature for long and short oligo systems. The goodness of fit was monitored via Coefficient of Determination (CoD). This gives a good automatable quality control of the experiment and of the fitting. Typical threshold values are 0.996 for kinetic fit and 0.998 for melting point analysis. In general we found the correlation of the theoretical and experimental values for T_m in the range of $R^2 = 0.7$ (data not shown). There was no significant difference between the NN algorithms and the simple formula.

5.3.5 Kinetic analyis

Figure 5.6 gives an example of the application of the algorithm for long and short oligo microarrays. The proposed fit function gives a good representation of the kinetics for both types of nucleic acids. The kinetic behavior of long and short oligo hybridization kinetics can be described by the bi-exponential function with a very good accuracy. It is interesting to see that in spite of the effective agitation equilibrium was not reached after 18 hours for low target concentration of 0.5 nM. For higher target concentrations (3.6 nM) equilibrium was reached after about 5 hours without competitive hybridization. In case of competitive hybridization, where signal intensity decreases after a specific maximum, equilibrium was never reached during our experiments (up to 18 hours hybridization time). This is an interesting result indicating the limitations of endpoint microarray analysis. In case of closely related target mixtures there is a significant change of signal ratios after overnight hybridization. One main focus of the development of the algorithm was to assess the potential for the

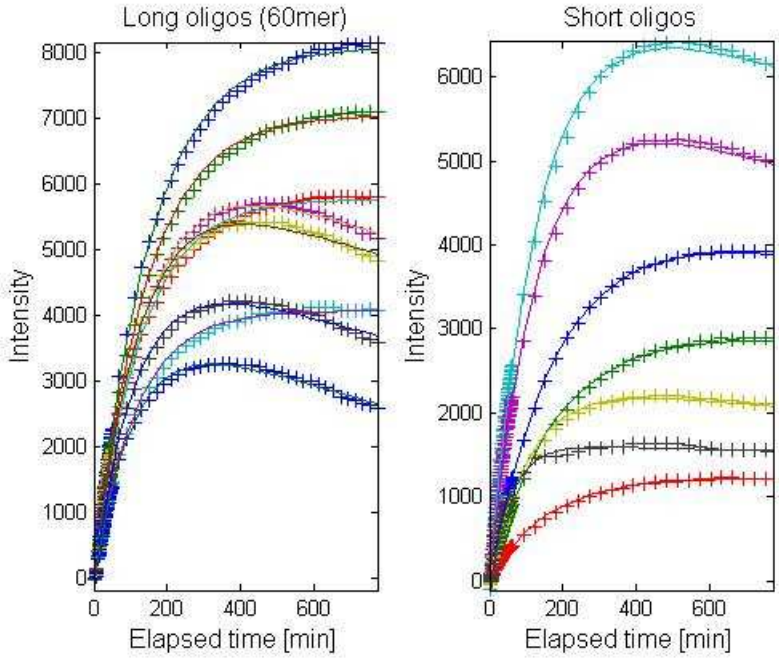


Figure 5.6: Application of the proposed algorithm to long (left) and short (right) oligo hybridization kinetics. The crosses are the experimetnal values, the lines gie the resutls of the proposed algorithm. For the first hour the interval for the image acquisition was one min afterwards 30 min.

reduction of experimental data acquisition. Reduced data sets were generated and used for fitting. The full dataset was used for verification. The results for the theoretical limit of 3 experimental values and the initial point (0,0) is shown in Figure 5.7. Even in this case the GoF is quite high, thus proving the validity of the model for a numerical description of nucleic acid hybridization kinetics. Obviously it is not possible to do an accurate quality control with this reduced data set without reference curves. Automated comparison of the kinetic parameters to reference measurement is a fast and simple quality control of the hybridization and for quantification of competitive hybridization in huge data sets. Due to the concentration dependency of competitive hybridization this can be used for quantification of target concentration. Figure 5.8 gives a histogram of the quality parameter CoD. Low quality spots can be easily separated from the standard measurements. Some authors propose a more complex model with 3 species including low affinity background hybridization [13]. Within our stringent experimental setup we did not see significant influence of unspecific background (salmon sperm DNA in 100x target concentration) nor the necessity for a more complex model even when using up to 8 closely related targets in similar concentration or in unspecific background (salmon sperm DNA 100x concentration). Nevertheless, it is important to note that it is not an analytical model for the complex hybridization and displacement reaction but a physical model based fitting algorithm.

Analysis of melting curve

The determination of the melting temperatures needs a high accuracy and reproducibility to improve upon the specificity achievable via mature endpoint microarrays. The starting point was a comparison of the derivative based method and the model based algorithm. Correlation of the two methods with manually adjusted reference values from standard Genewave normalization software are shown in

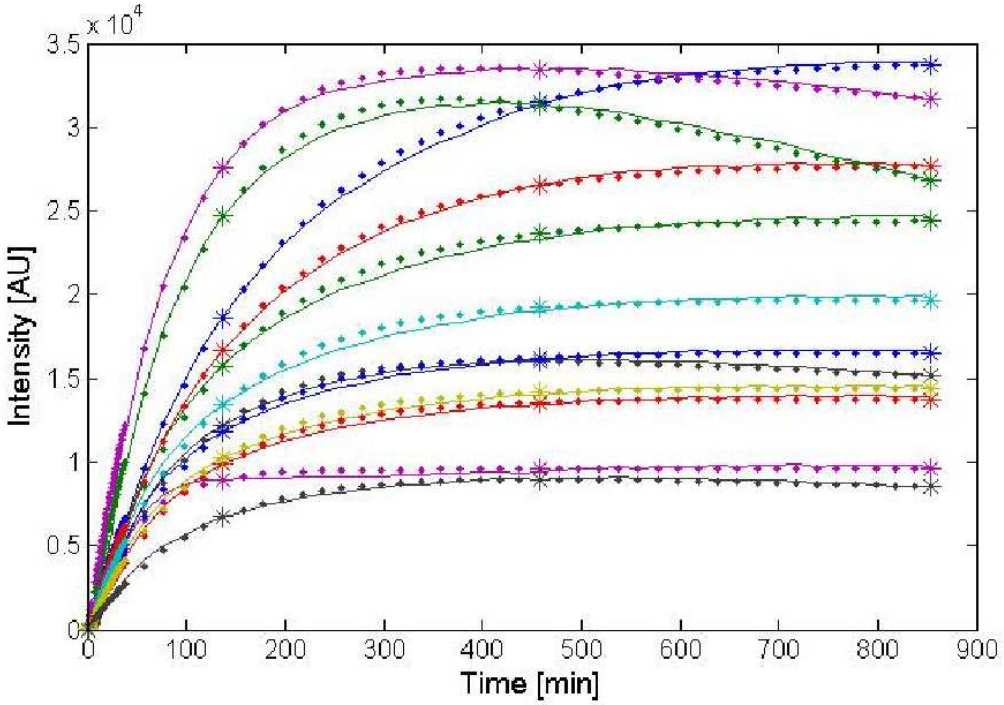


Figure 5.7: Applicatin of the algorithm using only three datapoints, marked with stars, and the initial value (0,0). The points show all measured data, the fit is given by the solid line.

Figure 5.9. The model based values gives a better correlation than derivative based values. No improvement of Savitzky-Golay filtering in comparison to simple low-pass filtering was found. This might be due to the quite simple shape of the melting curve. For different hybridization experiments the parameters of Savitzky-Golay filtering had to be modified manually and there is no potential for reduction of experimental data acquisition. With the model based algorithm it was possible to reproduce the results of the standard software without any manual adjustment. Figure 5.10 gives examples for the application of the algorithm to normalized melting curves. The temperature behavior of the different target systems (RNA•DNA hybrids and DNA•DNA duplexes) is according to the model. The normalized melting curves for short oligo systems and the fit functions are shown in 5.11. The quality of the description is comparable to that of long oligo systems. To test the potential for inter- and extrapolation datasets with reduced number of data points were created as well. Figure 5.12 gives a typical result. Only every third measurement point was used up to melting temperature. No experimental data above the melting temperature have been used. For this analysis a selection of probes with similar melting temperature has to be done. To take into account the remaining intensity of unspecific bound fluorophores at high temperatures the final intensity was set to 100 for all experiments. The algorithm gives a good prognosis to higher temperatures which could give another advantage of the algorithm for melting analysis. The higher the experimental temperature the lower is the signal to background ratio. The high potential in reduction in experimental data opens the door for cheaper and faster microarray based melting analysis. The good results for extrapolation suggest a fundamental relation between the melting of surface bound nucleic acids and Fermi-Dirac statistics. Another interesting opportunity of this algorithm is the automated detection of multiple target hybridization to a single spot. In endpoint based mutation screening the discrimination between different alleles in heterocygote samples and nonspecific hybridization of a single target is still challenging. In case of melting analysis the shape of

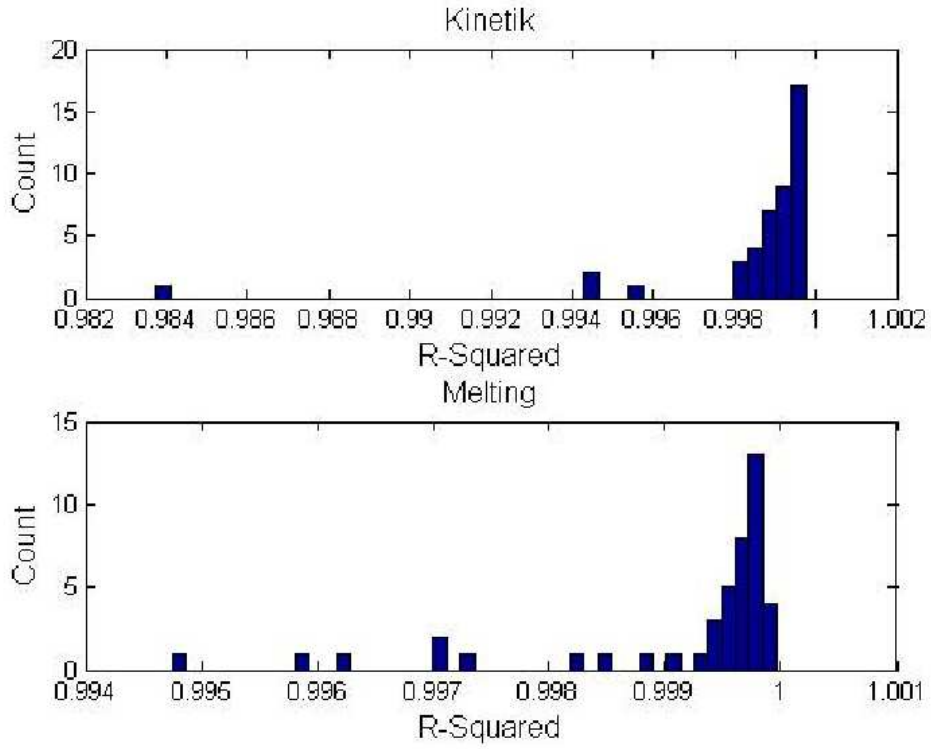


Figure 5.8: Histogram of CoD for kinetic analysis (top) and melting analysis (bottom). It is easy to set a threshold for quality monitoring. It is important to note the very high value for the Coefficient of determination.

the melting curve is changed due to the overlaid melting processes of the different targets. Thus for multiple target melting the parameter is smaller. A histogram of the distribution of k_b for single and multiple target hybridization is shown in 5.13. The histogram is based on the results of 6 experiments and shows the robustness of this quality control parameter. With 60 mer probes the 2 mismatches yield a difference of about 4 °Cfor single target hybridization between the PM and the MM. In case of two targets the slope of the melting curve is decreased and the corresponding parameter k_b is clearly shifted to smaller values. Thus automated discrimination between single and multiple target hybridization is possible.

5.3.6 Conclusion

Analysis of microarray based real-time analysis is an unsolved challenge. Within this work we developed an algorithm which uses standard comma separated data file and fasta files as input. We could reduce the complex information to 3 parameters for the kinetic and 6 parameters for the melting analysis. The algorithm was evaluated with 40 experiments including short and long oligo probes and DNA and RNA targets. The automated analysis of the hybridization kinetic provides a valuable tool for the assessment of competitive hybridization and in combination with reference curves it is possible to use real-time assays for quantitative analysis of target concentration. Fully automated analysis of melting curves for all 40 experiments was possible. The long oligo probes are a good compromise of sensitivity and specificity with the potential for SNP detection in melting analysis. Using melting analysis for high density arrays the possibility for straight forward data processing and automated quality control is a big step towards simple, userfriendly data processing. Thresholds for

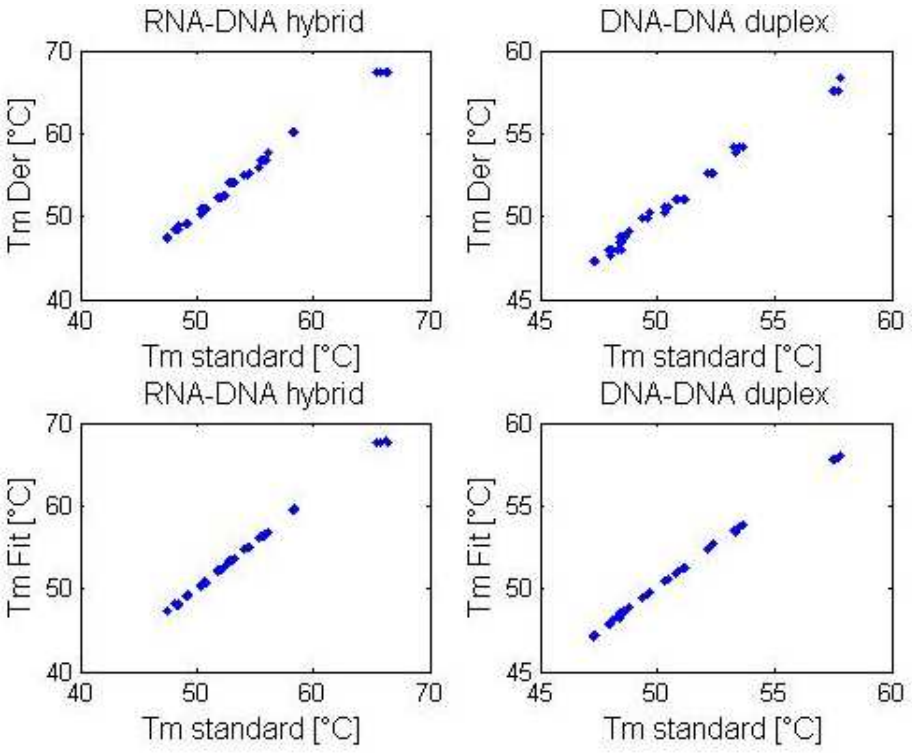


Figure 5.9: Comparison of a derivative based method and the model based algorithm with the reference values for the melting temperature. The upper row gives the correlation between the derivate based method and the standard value, the lower for the model based value.

the CoD and allow an automated binning of the experimental data, the detection of multiple target hybridization and measurement artifacts. We were able to show that with the proposed model based algorithm the number of measurements can be decreased dramatically and that it is even possible to extrapolate the melting curves from measurements done at temperatures below the melting temperature to higher temperatures. The reduced number of process steps for real-time measurement in combination with the high reproducibility and the robust algorithm opens possibilities for integrated real-time Lab-on-Chip systems. Furthermore the possibility of analyzing thousands of probe-target combination with this fast and accurate method will support probe design for standard microarrays and generate valuable data to understand the most elementary process in biology, the hybridization of nucleic acids.

5.3.7 Competing interests

The authors declare that they have no competing interests

5.3.8 Author's contribution

SF, LB and GH planned and supervised the research. LB did the design of the short and long oligo probes. MS designed and prepared the microarray. HT wrote the code for data analysis. SK carried out the molecular biological work for the extraction, amplification and labeling of the targets, the experiments, the conceptual work on the algorithm and wrote the paper.

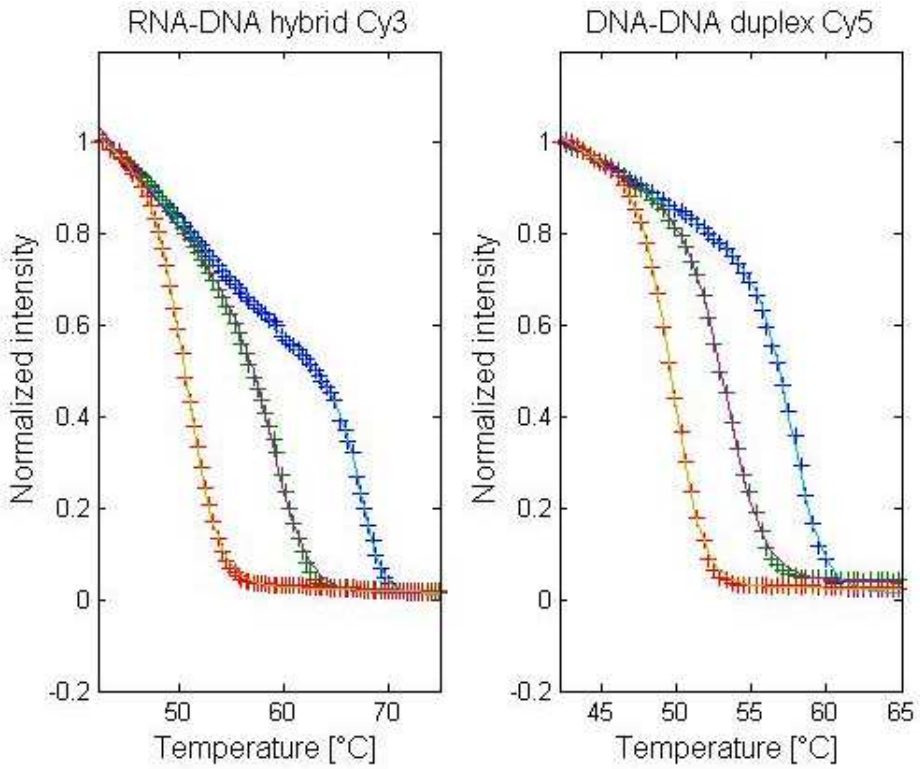


Figure 5.10: Normalized melting curves for RNA•DNA hybrids (left) and DNA•DNA duplexes (right).The normalized representation gives a good picture of the temperature behavior of the system. Crosses indicate experimental data, the lines indicate the fit functions.

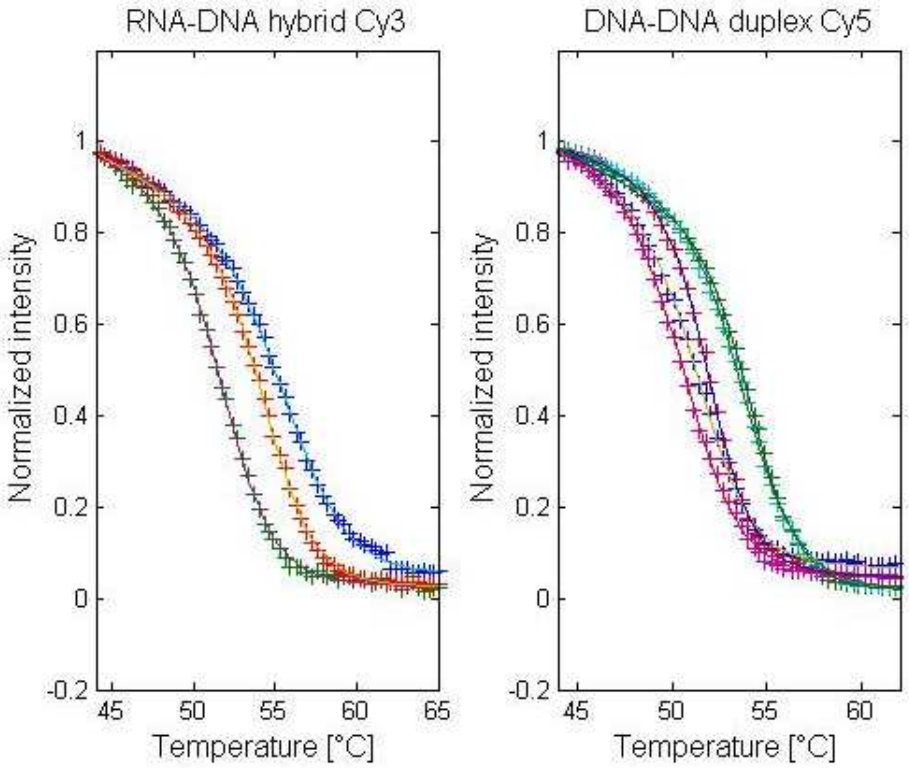


Figure 5.11: Examples for RNA•DNA hybrids (left) and DNA •DNA duplexes (right) for short oligo systems. Crosses indicate experimental values, the lines are the fit functions.

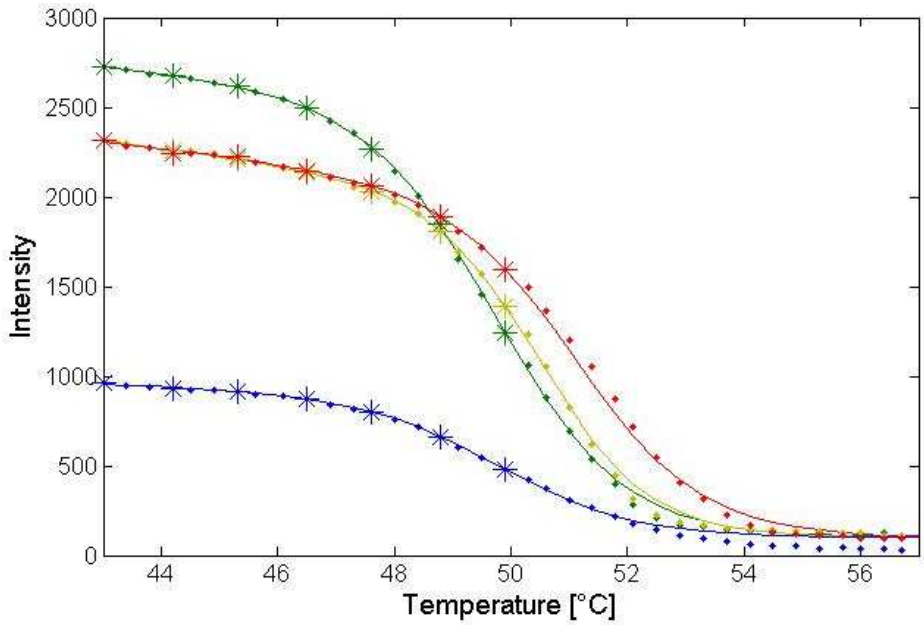


Figure 5.12: Application of the algorithm to a reduced data set. The stars indicate data points used for algorithm. The solid line gives the fitted function, the points denote experimental values. For high temperatures the value of the fit function were set to 100. For this representation a selection of probes based on the melting temperature was done and not normalized data were used for better visibility.

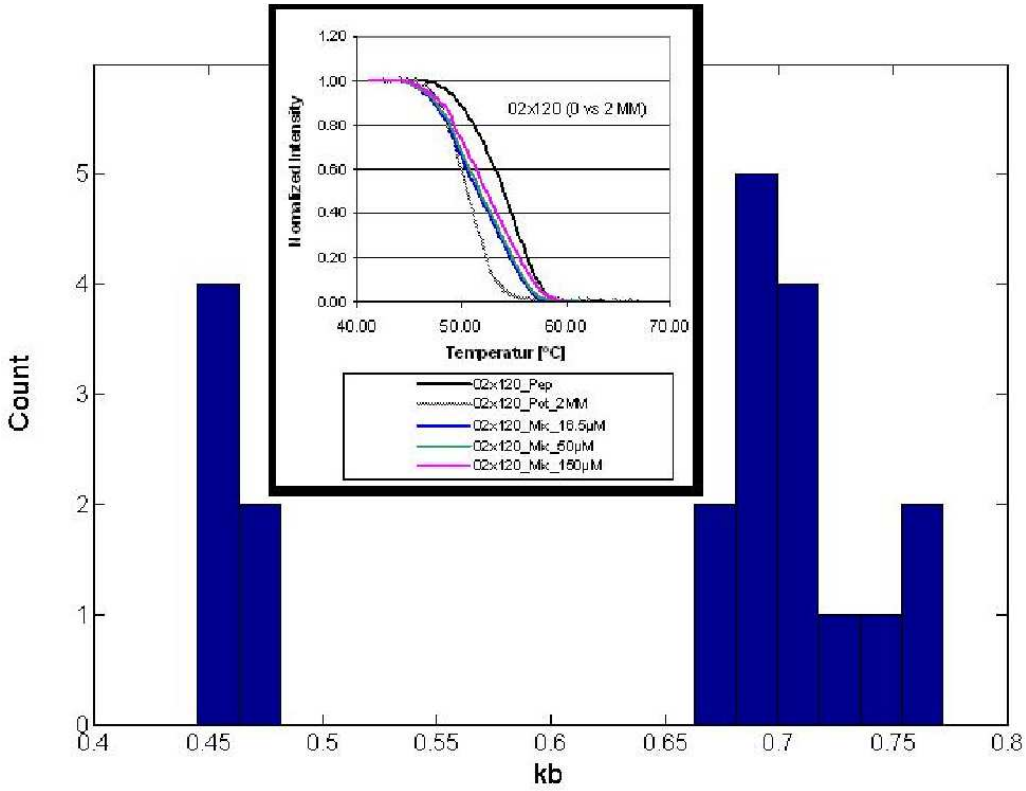


Figure 5.13: Histogram of the parameter k_B for single and multiple target hybridization. The right peak denotes the single target hybridization to the probe, the left a hybridization of a mixture of a perfect match and a 2MM target. The inset shows the normalized melting curves. Data are from 6 experiments.

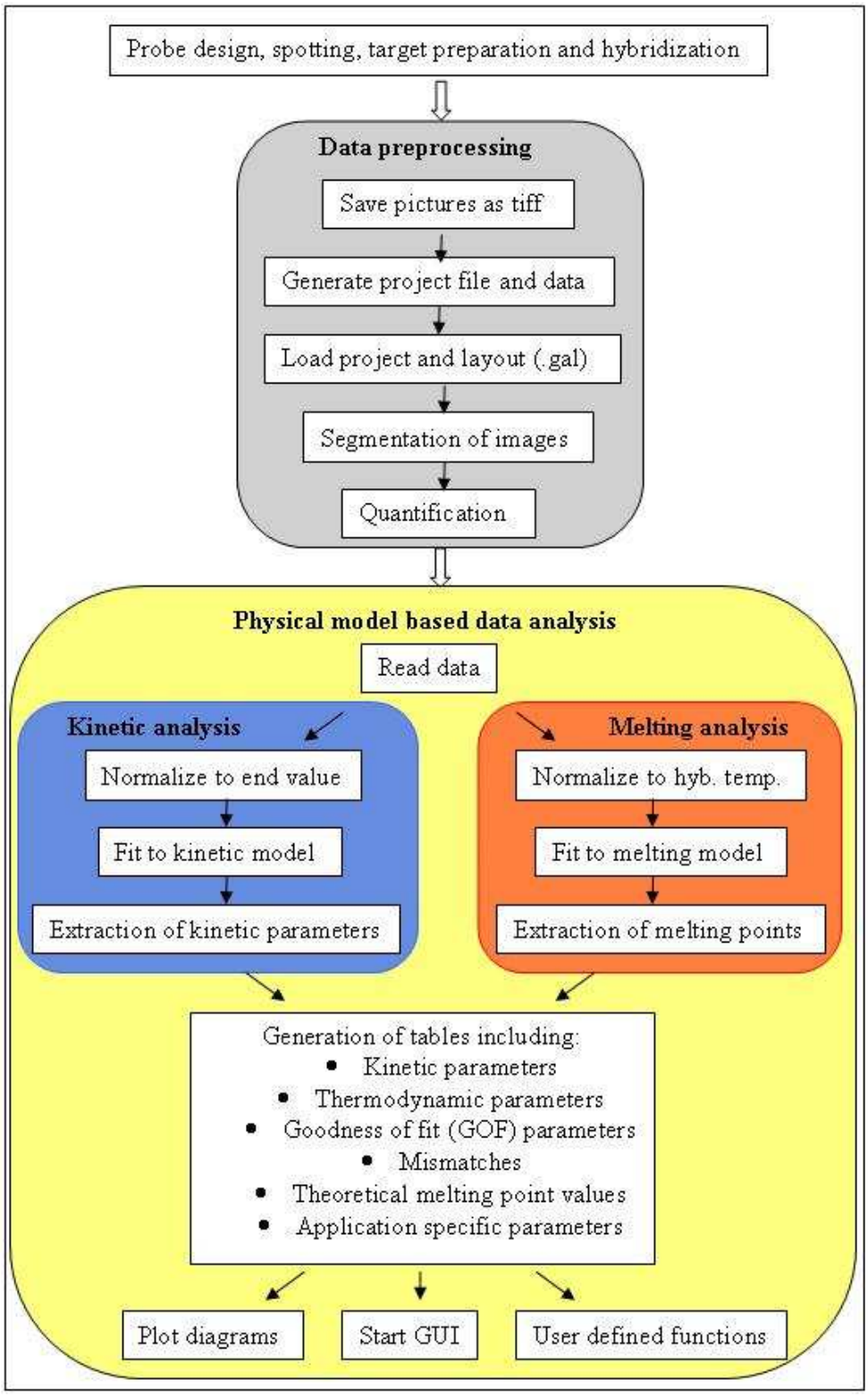


Figure 5.14: : Flow diagram of real-time microarray analysis. The gray shaded block at the left site describes the overall flow. The yellow block on the right gives the flow for the proposed algorithm for the automated extraction of kinetic and thermodynamic parameters.

Chapter 6

Competitive hybridization of DNA and RNA

6.1 Introduction

Competitive hybridization of different nucleic acid molecules introduces a lot of new effects to hybridization kinetics and thermodynamics [15, 68, 164]. In case of hybridization to surface immobilized probes the characteristics of kinetics changes from mainly transport limited to reaction limited. The dominant parameter is the reaction constant of the dissociation of the less specific nucleic acid molecule. It has been observed that DNA on glass microarrays is forming multi-stranded DNA structures that could not be accounted for by classical Watson-Crick base pairing [178]. These structures were observed with a number of heterogeneous sequences, including both, purine and pyrimidine bases, with a shared sequence identity between ssDNA and one of the duplex strands. The replacement mechanism in such cases occurs through the formation of a single-strand branch migration complex [108, 164]. Progressive migration of the branch point in this complex results in the eventual replacement of the origin strand. Due to the possibility of parallel analysis of large number of probe target combination microarray based real-time analysis is a valuable tool for the investigation of this basic effect.

6.2 The influence of target concentration

Competitive hybridization is a complex interplay of enthalpy, entropy and concentration of the different targets. From

$$\Delta G = \Delta H - T\Delta S \tag{6.1}$$

it is visible that even in this equation the ratio of the ΔG for different targets is changing with temperature. The target concentration is influencing the entropic part. It is known that during hybridization there is a significant depletion in the vicinity of the spot, especially for low abundant targets [179]. Experimental results for the background depletion for target concentrations of 0.5 nM and 3.6 nM are shown in Figure 6.1. While the maximum and minimum in the normalized representation is similar for both concentrations, the lower concentration is recovering while the higher is nearly constant at the depleted value. For targets with different ratios of ΔH and ΔG there is a changing intensity pattern for different temperatures. A theoretical prediction of rate of

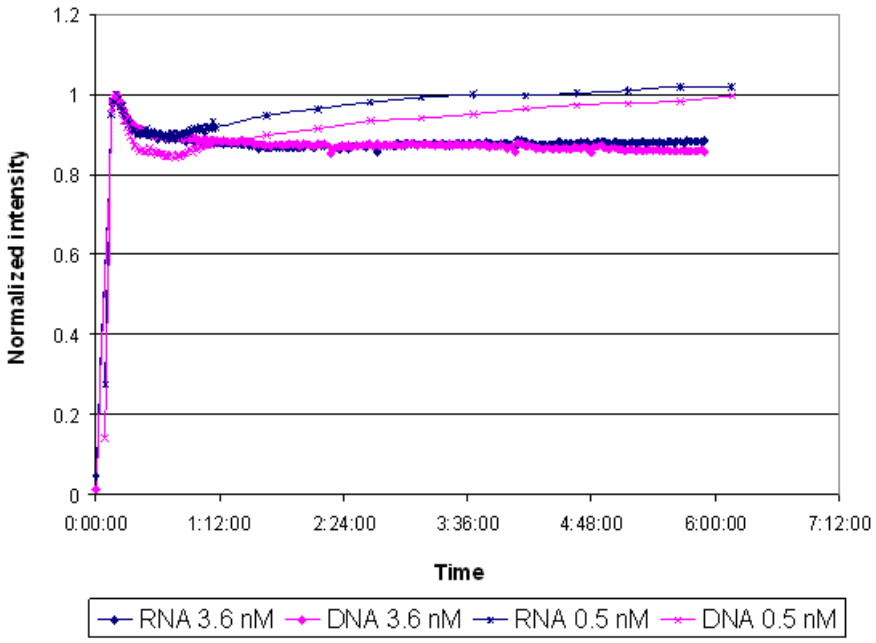


Figure 6.1: Depletion of the background intensity during hybridization. It is counterintuitive that the lower concentration recover while the 3.6 nM concentration stays nearly constant. From the absolute values of the intensity a accurate estimation of the concentration ratio is possible (7.6 compared to the spectroscopic value of 7.2).

coverage with different target concentrations can be seen in Eqn. 3.38. For the calculation we used a slightly modified formula:

$$\Theta = \frac{abK[A]}{1 + aK[A]} \tag{6.2}$$

with the fitting parameters $a = 0.01, b = 100$ and K was set to 7.3×10^{12} [56]. It is also interesting to compare the theoretical result with Eqn. (3.4). Using a logarithmic axis it can be seen that at low concentrations the intensity is higher than predicted by the Langmuir model. This might be due to unspecific hybridization. It can also be seen from Eqn. 3.38 that the ratio between foreground and background intensity decreases with increasing target concentration. This may lead to segmentation bias and cause the observed increase of background intensity for low abundant targets.

6.3 Kinetics of DNA•DNA competitive Hybridization

DNA•DNA competitive hybridization is on of the key processes in microarray hybridization and is determining time-to-equilibrium. Systematic investigation are challenging because of the big influence of mismatches on thermodynamics and secondary structure. In general it is assumed that kinetics on long oligonucleotide based microarrays is 'more irreversible' than on short oligo arrays. We found a significant difference in hybridization kinetics for probes including a few LNA targets, a typical result can be found in Figure 6.3, the corresponding alignment of the two target sequences is shown in Figure 6.4. Pinus has 5 MM position for the probe 2Pep730x798vsPot which are all standard (DNA) nucleotides. It is interesting that the competitive replacement of the LNA probes is higher than with standard *deoxy*-nucleotides. Intensity measurement in Figure 6.3 shows that the intensity is lower on the probes with LNA nucleotides, the highest concentration on the LNA spots has a lower intensity

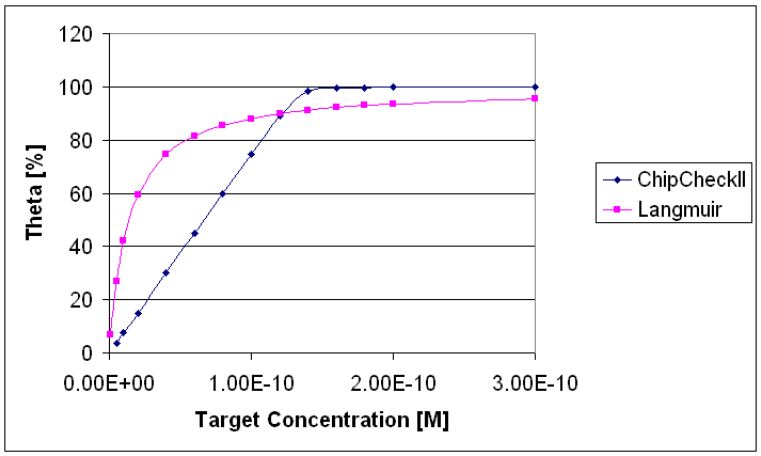


Figure 6.2: Rate of coverage versus target concentration predicted from Langmuir isotherm and linear approximation of ChipCheckII.

than the lowest concentration on the DNA probes. In our experiments the immobilization efficiency of targets with LNA incorporates was lower than the comparing DNA probes. This behavior gives another example where melting curve measurement gives a more reliable information than intensity based analysis.

6.4 Kinetics of DNA•RNA competitive Hybridization

6.4.1 Hybridization of RNA to double-stranded DNA

Hybrid duplexes comprising a DNA and an RNA strand occur in several important biological processes. They are intermediates in transcription, in DNA replication, and in the synthesis of retroviral cDNA by reverse transcription. They are substrates for the enzyme RNAseH which catalyzes the hydrolysis of RNA only when it is present in an RNA•DNA hybrid duplex and not when it is part of an RNA•RNA duplex. RNAseH shows little if any sequence specificity and it seems that the enzyme recognizes a difference in conformation between RNA•DNA hybrid and RNA•RNA duplexes [114, 140]. The origin of the differences in free energy of DNA and RNA duplexes was rationalized in term of the chemical differences arising from the 5-methyl group of T in the DNA and the $2'-$OH-group of the ribose in RNA [214]. The solution conformation of DNA•RNA hybrids of mixed sequence is neither the A-form of typical of duplex RNA nor the B-form typical of duplex DNA, though the overall conformation is closer to the A-form than to the B-form. It was shown that the rR•dY hybrids are more like RNA (i.e., A-form) whereas dR•rY is intermediate between the DNA and RNA duplexes and that this difference is recognized by RNAseH [114, 140]. Experiments with different base compositions suggested that there is a continuum of structures between the A- and B-forms [76]. It has been reported that the thermodynamic stability of hybrid duplexes in which the DNA stand consists of purines and the RNA strand consists of pyrimidines (dR•rY) is less than the corresponding duplex containing a purine RNA strand and a pyrimidine DNA strand (rR•dY) [104]. Solution based studies found that the rate of R-loop formation is maximal at the temperature at which half of the duplex DNA is irreversible converted to ssDNA (T_m) and falls precipitously a few degrees above or below that temperature. This maximal rate is similar to the rate of hybridization of RNA to ssDNA under the same conditions. In prehybridized dsDNA it was found that at temperatures above T_m the rate is proportional to the RNA concentration, at temperatures below T_m the rate of R-loop formation

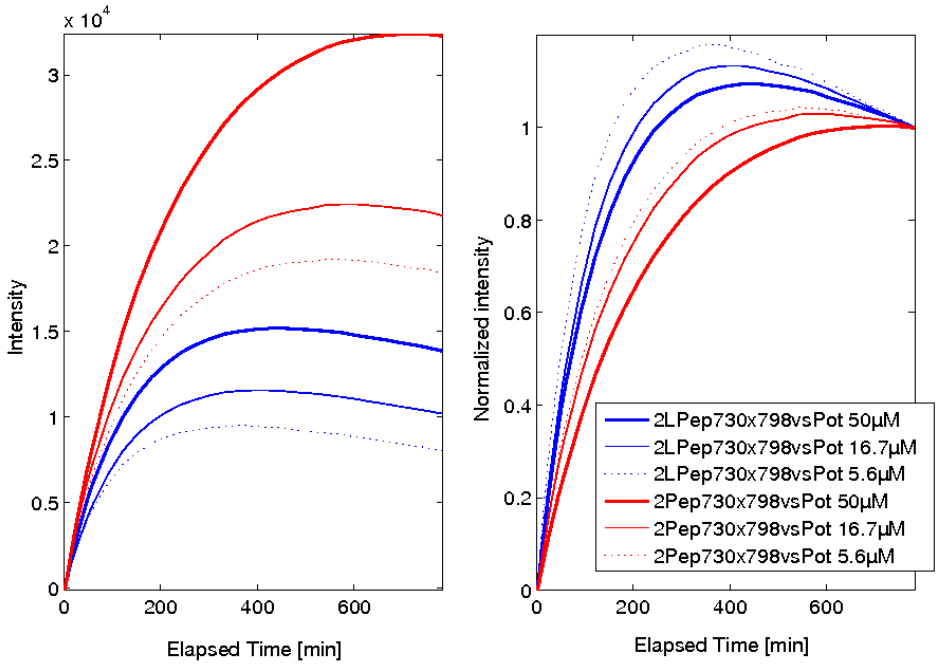


Figure 6.3: Competitive displacement of DNA target Pin from target Pot. The upper figure gives the kinetic, the lower the normalized kinetic for Exp. 25 (F532 RDNA Pot; F635 DNA Pin). It is interesting to note that the competitive hybridization is concentration dependent and that on the LNA target the replacement is stronger even the melting temperature is higher Figure 6.5.

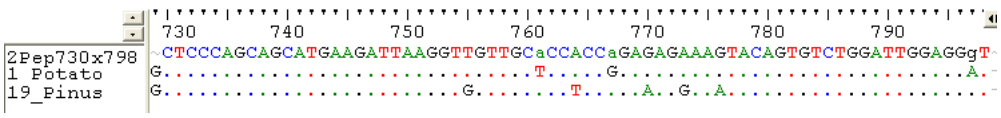


Figure 6.4: Alignment for the DNA•DNA competitive hybridization of Pot and Pin to probe 2Pep730x798vsPot. The lower case letters denote the LNA bases in 2LPep730x798vsPot.

is less dependent upon the RNA concentration [193]. In general the DNA hybridization has higher reaction rates than the RNA hybridization. Thus DNA is binding to the oligos first, but is replaced when bound not stringent enough. The displacement of the DNA strand by an RNA strand can be seen in Figure 6.6. There is a clear reduction in intensity and a corresponding rise of the RNA signals. It is important to note that this only happens at spots with DNA melting temperatures close to the hybridization temperature. Due to the competitive kinetics RNA equilibrium is not reached after 13 hours. It was found that non specific bound DNA is replaced while perfect matched probe-target duplexes were not affected by the displacement kinetics, thus the competitive displacement of DNA by RNA increases the stringency of the hybridization and thus the specificity of the assay.

6.4.2 Strand displacement kinetics

The displacement of DNA strands with an RNA strands of the same sequence is described by the chemical equation:

$$A^D_{ds} \bullet B^D_{ds} + B^R_{ss} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} A^D_{ds} \bullet B^R_{ds} + B^D_{ss} \tag{6.3}$$

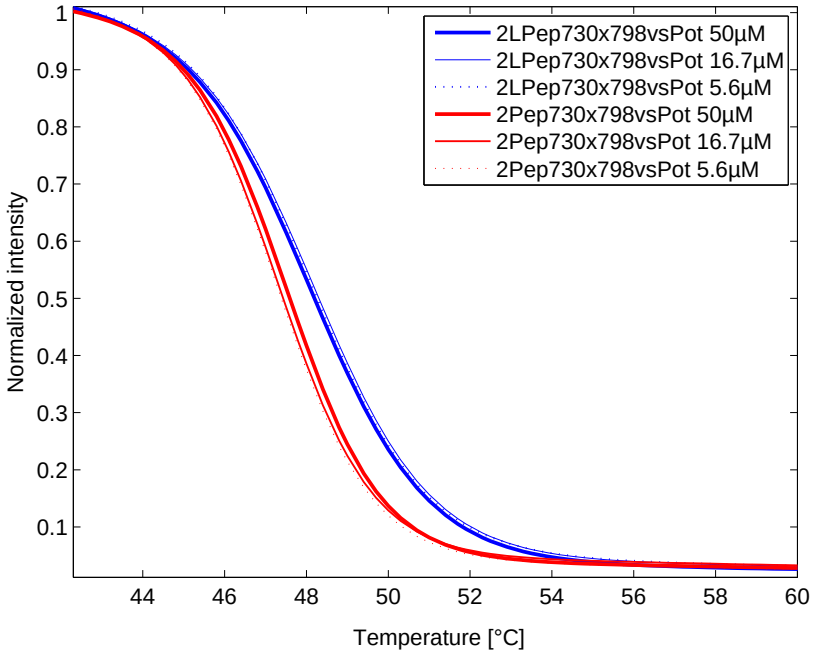


Figure 6.5: Melting curves for probe 2Pep730x798vsPot and Pinus target. According to theory the melting temperatures are higher for the DNA•LNA hybrid. There is no concentration dependency visible. The T_m of the Pot•probe duplex is about 3℃ higher than the replaced Pin•probe duplex (data not shown).

This second order-reaction can be described by the differential equation

$$\begin{aligned} \frac{d\left[A^D_{ds} \bullet B^R_{ds}\right]}{dt} &= \frac{d\left[B^D_{ss}\right]}{dt} = \\ &= k_f \left[A^D_{ds} \bullet B^D_{ds}\right]\left[B^R_{ss}\right] - k_r \left[A^D_{ds} \bullet B^R_{ds}\right]\left[B^D_{ss}\right] \end{aligned} \tag{6.4}$$

The concentrations of A^D_{ds}•B^D_{ds} and B^R_{ss} at the time t depends on its initial concentration $\left[A^D_{ds} \bullet B^D_{ds}\right]_0$, $\left[B^R_{ss}\right]$ and the concentration of $A^D_{ds} \bullet B^R_{ds}$

$$\left[A^D_{ds} \bullet B^D_{ds}\right] = \left[A^D_{ds} \bullet B^D_{ds}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right] \tag{6.5}$$

$$\left[B^R_{ss}\right] = \left[B^R_{ss}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right] \tag{6.6}$$

Assuming appropriate hybridization conditions, the dissociation rate constant $k_r$of the reverse reaction is negligible. In this way we obtain

$$\frac{d\left[A^D_{ds} \bullet B^R_{ds}\right]}{dt} = k_f \left(\left[A^D_{ds} \bullet B^D_{ds}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right]\right)\left(\left[B^R_{ss}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right]\right) \tag{6.7}$$

Separation of the variables gives

$$\int \frac{d\left[A^D_{ds} \bullet B^R_{ds}\right]}{\left(\left[A^D_{ds} \bullet B^D_{ds}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right]\right)\left(\left[B^R_{ss}\right]_0 - \left[A^D_{ds} \bullet B^R_{ds}\right]\right)} = k_f \int dt \tag{6.8}$$

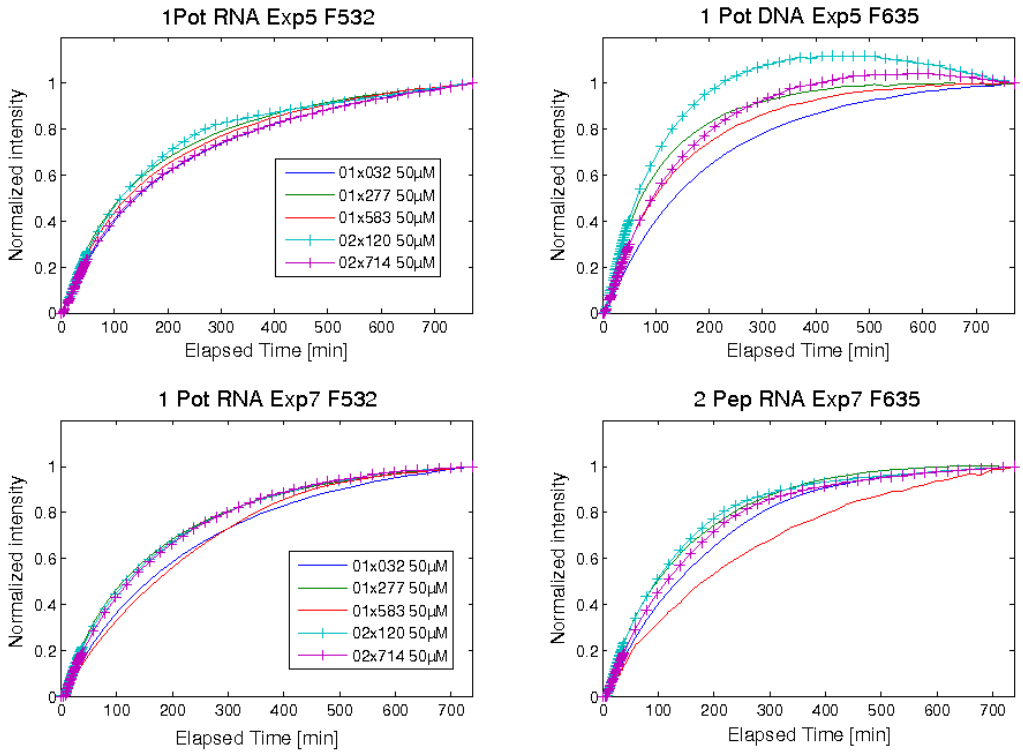


Figure 6.6: Competitive displacement between PM and MM probes. The upper row gives the results of Exp. 5 (Left figure: F532 RNA Pot; right side: F635 DNA Pot), the lower row gives the results of Exp 7 (Left figure: F532 RNA Pot; right side: F635 RNA Pep). The crosses indicates MM targets for Pot. Competitive hybridization between closely related RNA molecules did not show this displacement kinetics (lower row).

Integration yields the concentration of the DNA•RNA hybrid $\left[A^D_{ds} \bullet B^R_{ds}\right]$ at the time t

$$\left[A^D_{ds} \bullet B^R_{ds}\right] = \frac{\left[A^D_{ds} \bullet B^D_{ds}\right]_0 \left[B^R_{ss}\right]_0 \left(1 - e^{\left(\left[B^R_{ss}\right]_0 - \left[A^D_{ds} \bullet B^D_{ds}\right]_0\right) k_f t}\right)}{\left[A^D_{ds} \bullet B^D_{ds}\right]_0 - \left[B^R_{ss}\right]_0 e^{\left(\left[B^R_{ss}\right]_0 - \left[A^D_{ds} \bullet B^D_{ds}\right]_0\right) k_f t}} \tag{6.9}$$

This equation shows that at large time instant t the concentration of the product $\left[A^D_{ds} \bullet B^R_{ds}\right]$ tends towards the concentration of the reactant, either $A^D_{ds} \bullet B^D_{ds}$ or B^R_{ss}, depending on which of the initial concentration is lower.

Another possibility is a numerical description similar to chemical reactions with an intermediate product. Assuming the reactant is the single stranded oligo, the $DNA \bullet DNA$ is the intermediate product and that the $RNA \bullet DNA$ hybrid is the final product one gets for the numbers of each reactant assuming that there is no reverse reaction

$$n_{ss} \underset{0}{\overset{k_a}{\rightleftharpoons}} n^{DNA}_{ds} \underset{0}{\overset{k_b}{\rightleftharpoons}} n^{RNA}_{ds} \tag{6.10}$$

This lead to the equation for the observed intensity

$$I = A \frac{k_a}{k_b - k_a} \left(e^{k_a t} - e^{k_b t}\right) \tag{6.11}$$

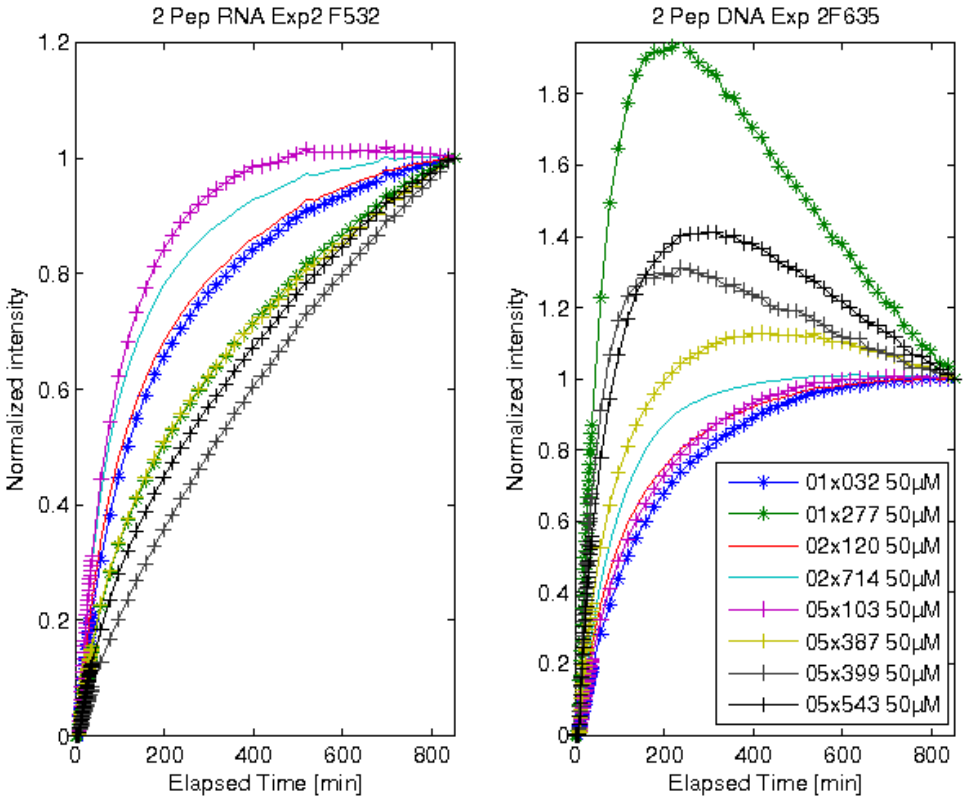


Figure 6.7: Competitive displacement between PM and MM probes. While there is a general trend that mismatch probes show higher degree of competitive displacement (indicated by crosses and asterisks) it is not a general rule.

This equation was used for the physical model based fitting algorithm. We found this algorithm sufficient for the fitting of our results.

6.5 Publication: Competitive hybridization of different nucleic acids

S. Krainer, S. Fluch, M. Stierschneider, L. Bodrossy

Submitted to *Nucleic Acid Research*

6.5.1 Abstract

Microarray based real time data acquisition of hybridization kinetics and melting analysis provides opportunity for high density analysis of genetic data and for basic research on physico-chemical properties of nucleic acid hybridization. While single target hybridization kinetics is dominated by diffusion transport limitations, competitive hybridization gives insight into elementary interactions of nucleic acids. A model system was developed for systematic investigations of competitive hybridization kinetics between different types of nucleic acids. ssDNA and RNA targets of the same sequence were prepared and hybridized competitively to surface immobilized DNA probes with and without LNA nucleotides in critical positions. This reduces the differences of the competing target

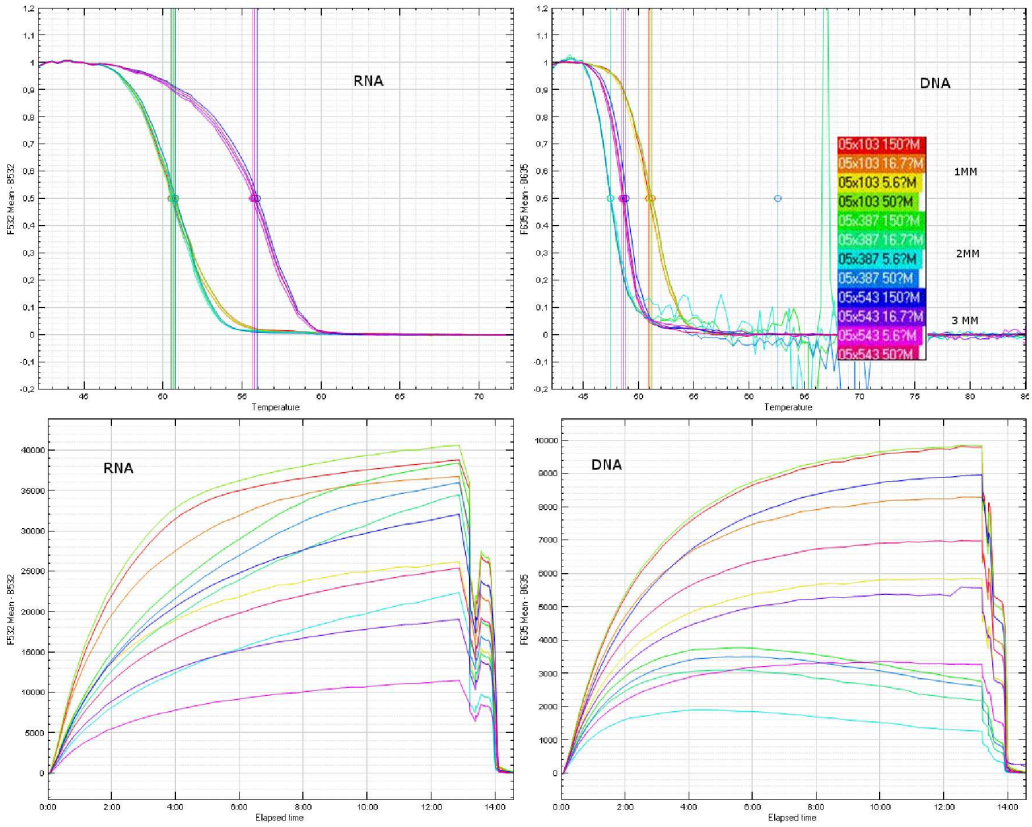


Figure 6.8: Melting curve analysis (upper row) and hybridization kinetics (lower row) of DNA vs. RNA. The lines are averages of intensities of three technical replicates.

molecules to the sugar conformation and the exchange of the two bases thymine and uracil. It was found that in many cases the DNA target strand of the initially formed DNA•DNA duplex is replaced by the RNA target. Melting analysis of the duplexes showed that there was a critical melting temperature for the DNA•DNA duplexes which defines the upper limit for this replacement kinetics. Comparison of competitive hybridization for probes with and without LNA nucleotides in mismatch position showed a significantly higher degree of competitive displacement on the probes with LNA molecules incorporated. This offers an explanation for the reported higher specificity of probes with LNA nucleotides incorporated. It is demonstrated that the effect of competitive hybridization can be used for specificity improvement and for quantitative analysis of target concentration.

6.5.2 Introduction

Competitive hybridization is known to have a major influence on microarray results [15, 77, 228]. In case of multiple target hybridization time-to-equilibrium is not determined by association and dissociation constants of single-component system. Instead, the dissociation rate constant of non-specific targets emerges as the dominant parameter. At low target concentration, time-to-equilibrium can be in the order of hundreds of hours [12]. Due to the single point measurement characteristics of standard endpoint based microarray systems it is obviously not possible to detect equilibrium conditions. A comprehensive investigation on competitive hybridization of closely related targets is challenging because of the influence of base mismatch type and position on thermodynamics, hydrogen bonding and secondary structure [14]. To overcome these problems a model system of

RNA and ssDNA of the same sequence was designed. The application of isosequential ssDNA and RNA targets reduces the difference between the two competing molecules to sugar conformation and the exchange of the two bases thymine and uracil, i.e. the $2' - hydroxyl$ group in RNA ribose and the C5 methyl group in thymine. Due to these differences RNA•DNA hybrids are in general more stable than DNA•DNA duplexes [27, 213, 222]. A full set of nearest neighbor (NN) parameters of DNA•RNA duplexes including mismatches is still not available, with only limited parameter sets reported [9, 185]. RNA•DNA hybrids have essential functions in many biological systems e.g. DNA replication, transcription, the replication of retroviruses and the various functions of noncoding RNA [182]. A comprehensive review on the importance of RNA•DNA hybrids can be found in [177]. Real-time monitoring of hybridization isotherms is known to improve the discrimination of specific and unspecific hybridization and gives the opportunity for quantitative analysis [22, 37]. The widespread use of this technology has been so far prevented by the experimental and bioinformatics challenges. Recent developments in real-time microarray analysis [97, 125] opened the possibility to analyze large numbers of probe-target combinations on microarray formats, thus circumventing the limitations associated with the more complex methods of kinetic and thermodynamic investigations, e.g. SPR, TIRF, DASH [92, 87, 113, 183, 190, 223], waveguide based methods (Bishop et al., 2007a; Brandenburg et al., 2009) and advanced methods for measuring hybridization of single molecules [74]. Besides microarray quality improvement this also provides opportunity for large scale basic research on nucleic acid hybridization behavior to surface bound probe molecules.

Locked nucleic acids (LNA)

LNA monomers contain a modified ribose moiety in which the 2′O and 4′C are linked by a methylene bridge (2′-O,4′-C-methylene-β-D-ribofuranosyl), locking the sugar in the C3′-endo conformation of RNA [20, 148, 204]. The incorporation of LNA nucleotides with reduced conformational flexibility in DNA and RNA sequences increase the stability of the corresponding duplexes due to local organization of the phosphate backbone [20]. LNA resemble natural nucleic acids with respect to Watson-Crick base pairing and the potential of LNA containing oligonucleotides is their ability to mediate high affinity pairing with complementary RNA or DNA strands, with equal or often superior sequence specificity than their natural equivalent [149, 226]. Generally, sequence specificity and binding affinity of DNA•DNA, RNA•DNA and/or RNA•RNA interactions negatively correlate with each other, i.e. as the affinity for the chosen target sequence increases, the likelihood of association with closely related but nontarget sequences also increases [42, 117]. Studies on short oligos showed that in some cases the introduction of a few LNA nucleotides gives a better specificity than using full length LNA oligomers [30]. When using LNA nucleotides only in the mismatch position the best discrimination is achieved when the LNA nucleotide is a purine [153, 226]. The combination of 70 mers with a few LNA nucleotides incorporated gives a high affinity and sensitivity hybridization system and is well suited for investigation of the influence of single LNA nucleotides on the kinetic and thermodynamic behavior of DNA strands.

6.5.3 Methods

The investigations have been done with the 'Actin X-chip'. Actin was chosen because it is one of the most conserved housekeeping genes in eukaryotes with potential for species discrimination and food diagnostics [63, 160]. Cross-species comparison of this gene provides a set of sequences with a broad range of homology levels and biological relevance. DNA probes were synthesized with and without LNA nucleotides incorporated. Probes were hybridized with DNA and RNA targets. Long oligos were chosen because of their higher sensitivity and cooperativity. The longer the sequence in general the more pronounced it the melting transition, which makes long oligos well suited for real

time analysis. Most of the investigations was done on 60 mer probes, due to the high specificity of melting analysis 70 mer oligos have been designed and synthesized for the LNA investigations.

Extraction and purification of the actin gene

See 3.8.3

Probe design and preparation of the slides

See 3.8.3

Target preparation

See 3.8.3

Hybridization

See 3.8.3

Algorithms for data analysis

Segmentation and quantification of the generated .tif files was done with the Genewave HybLive software package using the 'irregular shape' segmentation algorithm; all parameters were set to default values. Quantification was based on the average of the three technical replicates for each probe. For most of the analysis the 5.6 μM probes were skipped due to the low signal intensities.

6.5.4 Theoretical models

Structural differences between DNA and RNA

There are two differences between DNA and RNA structure. The first is the existence of the 2′-deoxyribose in the former and ribose in the later, the second is the presence of thymine (T) instead of uracil (U). Due to the 2′-hydroxil group the preferred conformation of RNA is C3′-endo. Thus RNA•RNA duplexes prefer the A-form while the typical conformation of DNA•DNA duplexes is the B-form. DNA•RNA hybrid conformation is in between the A and the B forms, depending on base composition [53, 60, 75, 162, 177]. The only structural variance between the T and U bases in natural occurring DNA and RNA is the existence of a methyl group in the 5-position of the pyrimidine ring in T. Methylation of the C5-atom has an impact on DNA thermodynamic and kinetics via an attractive interaction between the methyl group and the neighboring base, the so called CH/π-interaction [7, 138, 201, 213, 210]. The CH/π interaction is a weak attractive force acting between CH groups and π-system of aromatic rings [31, 138]. Several studies investigated the influence of chemical modifications on nucleic acid stability and structure, especially on the influence of methylation of C5 position of cytosine [44, 192]. The stabilizing influence of the methyl group of thymine is visible via the NN parameters for ΔG for dTT•dAA (-1.2 kcal/mol) versus rUU•dAA (-0.2 kcal/mol), where the DNA•DNA sequence is much more stable [186, 184]. Propynil and methyl substitution on the C5 atom increase the thermal stability of siRNA duplexes or DNA•RNA hybrids containing them [60, 192]. While it is known that neither of the U→ T or C→ 5m-C C5 methylation have a significant effect on the helix structure or bending of the DNA helix axis it has been shown that the thymine methyl groups provide the dominant contribution to the high stability of AT base pairs in A-tracts [210]. It has been shown that the 5-methyl group in an ApT dinucleotide favorably interacts with the base π-ring before it whereas in TpA there was no such interaction [201]. Base

pair opening ('breathing') of nucleic acid is the key mechanism to understand the specificity and kinetics of hybridization. Investigation on imino proton exchange provides insights in duplex stability and dynamics [48, 88]. Base opening in RNA appears to have an overall preference towards the major groove, similar to results previously reported for B-DNA but RNA bases are found to have a substantially smaller major groove opening extent that that of their B-DNA counterparts [144]. The reduced opening extent correlates with the RNA duplex stability and is consistent with observed slower imino proton exchange rates in RNA duplexes [27]. Another effect may also contribute to the reduction of base pair breathing by the C5 methyl groups. Besides improving the stacking energies between the base pairs the methyl groups are also likely to alter the hydration pattern in the major groove [210].

Modeling of competitive hybridization kinetics

Modeling of the data was done according to [15]. The equations were reduced to a two component model. It is characterized by the number n_{ss} of single stranded DNA target in solution, n_{us} for the unspecific target hybridized to the probe and n_s for the specific target:

$$n_{ss} \underset{0}{\overset{k_h}{\rightleftharpoons}} n_{us} \underset{0}{\overset{k_c}{\rightleftharpoons}} n_s \tag{6.12}$$

The two constants k_h and k_c are the reaction rates of the two consecutive steps of hybridization and competitive displacement. The rate laws of this reaction with the target concentration C_h^T and C_c^T for the hybridizing and competitive targets are:

$$\frac{dn_{ss}}{dt} = -C_h^T k_h n_{ss} \tag{6.13}$$

$$\frac{dn_{us}}{dt} = C_h^T k_h n_{ss} - C_c^T k_c n_{us} \tag{6.14}$$

The solution of these equations for the competitive replaced target n_{us} with the initial concentration n_{ss}^0 is:

$$I_{us} \propto n_{us} = n_{ss}^0 \frac{C_h^T k_h}{C_c^T k_c - C_h^T k_h} \left(e^{-C_h^T k_h t} - e^{-C_c^T k_c t} \right) \tag{6.15}$$

I_{us} is the intensity of the target with the higher reaction rate and the lower binding affinity. In case of non competitive hybridization $k_c = 0$ and the formula is still valid. Bi-exponential fits were reported to be a good approximation for the kinetic of surface hybridization reactions in general [92, 74, 190]. For the calculation of the timepoint of maximum signal (t_{max}), after some basic transformations the following equation is obtained:

$$t_{max} = \frac{\ln\left(\frac{C_h^T k_h}{C_c^T k_c}\right)}{C_h^T k_h - C_c^T k_c} \tag{6.16}$$

This equation is very useful for the estimation of target concentration. Changing the concentration of both nucleic acids by a factor of x gives a $1/x$ dependency of t_{max}. This simple model holds for short oligo arrays as well (data not shown). The goodness of fit (GOF) was monitored with R^2. For GOF we used a threshold of $R^2 = 0.995$. This gives a good automatable quality control. Low stringency hybridization conditions possibly raise the necessity for a 3 species model taking into account the low affinity background [13] or a more complex Langmuir based model [190]. Within our experimental setup this was not necessary. The algorithm was realized in Matlab™.

6.5.5 Results and discussion

Competitive kinetics of hybridization between DNA and RNA targets

Similar to DNA-DNA displacement [14, 112, 228] competitive hybridization kinetics was visible on many probes. Figure 6.9 shows a typical result for long and short oligos including different probe

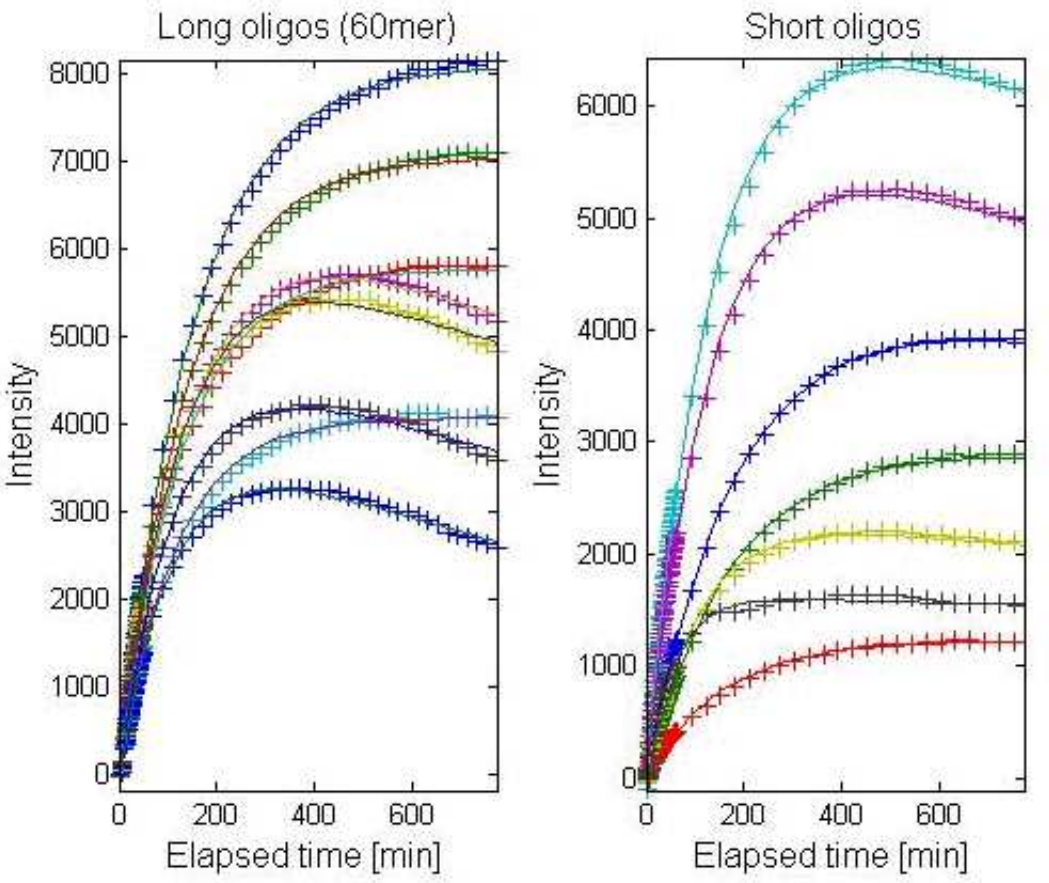


Figure 6.9: Examples of competitive hybridization for long oligos (left) and short oligos (right). The diagrams give the intensity of the DNA channel minus background intensity. The crosses indicate experimental values, the lines the model based fit of the data. In both cases equilibrium is not reached after 12h. The good quality of the model based fit for the different hybridization characteristics is visible.

concentration. In this figure the probes were chosen to give a good overview on possible intensities and displacement characteristics. The goodness of the fit (solid line) is visible. The influence of the probe length on hybridization kinetics was small, which proves the diffusion-transport limitation of hybridization to surface immobilized probes [64, 103]. Figure 2 shows the dependency of competitive hybridization kinetics from target concentration for two different target concentrations. The timescale in the two experiments shown are different because in the low concentration case competitive hybridization started only after 17 hours. It is visible that in this experiment only MM probes show competitive displacement. While competitive kinetic is not limited in general to mismatch probes, it was found to occur much more frequently there. This might be due to thermodynamic reasons as well as due to the higher opening rates ('breathing') of mismatch sequences. In an analogous endpoint based microarray experiment 02x714 would be a clear false positive result; the intensity

of the MM probe 02-714 is higher than that of two PM probes 05-387 and 05-399. Competitive hybridization is only affecting the mismatch probes in this case and has not reached equilibrium in the low concentration experiment. Extrapolation of the competitive kinetics suggests that the intensity will reduce further thus correcting the high unspecific hybridization of the MM targets. It is visible from the results of 6.10 that it is not possible to quantify the target concentration based on

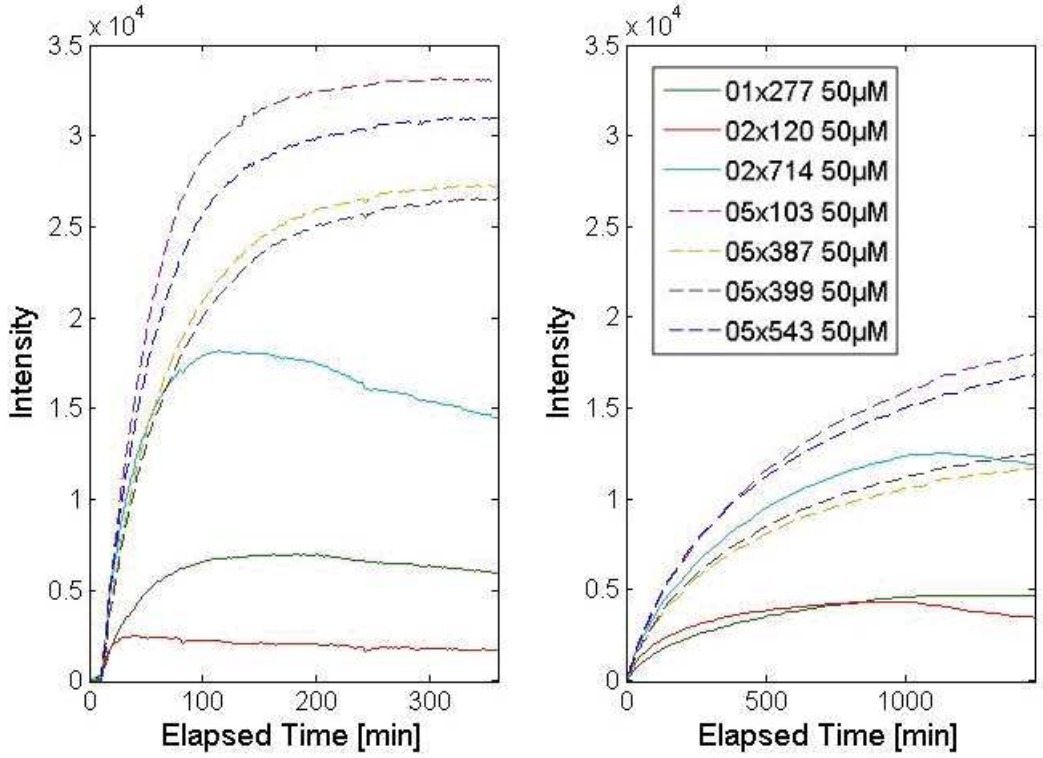


Figure 6.10: Competitive kinetics for different target concentration. Cy5 labeled DNA and Cy3 labelled RNA targets were co-hybridized to the array. The diagrams show the background corrected intensity of the DNA channel (Cy5). The legend is valid for both diagrams. In the left figure the target concentration is 3.6 nM, in the right figure it is 0.5 nM. The scaling of the y-axes is the same for both diagrams, the x-axes have different time scales due to the different reaction rates of the two experiments. Perfect match (PM) probes are represented by dashed lines, MM probes (Potato,Pepper) with solid lines.

endpoint intensities. Using the equation 6.15 for the calculation for the modeling of the hybridization kinetics of the two experiments gives 7.32 ± 1.22 (SD) for the ratio of the concentration C_x^T. This is a quite good result compared to the measured ratio of 7.2. Another possibility for an estimation for single target concentration is the fluorescence of the background. Using the maximum background intensity value during hybridization for the two experiments a ratio of 7.60 was calculated. The effect of the probe concentration on the reaction rate was negligible in our experimental setup, i.e. in the concentration range applied the normalized kinetics was not dependent on the spotting concentrations. It was shown that limiting the DNA probe density allows the evaluation of kinetic effects without intermolecular interaction between neighboring binding sites [190]. Together with the effective agitation this allows for an accurate estimation of the target concentration based on equation 6.15. Melting analysis was done in order to better understand competitive hybridization. Figure 6.11 shows the correlation between the melting temperature of the DNA•DNA duplex and

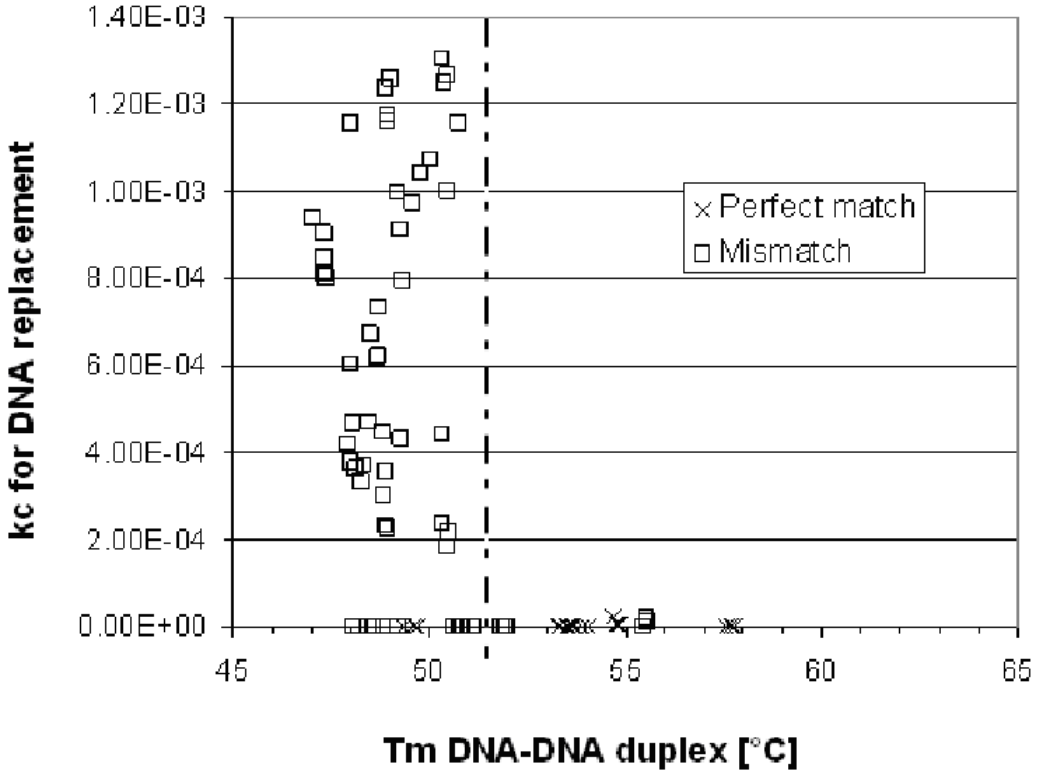


Figure 6.11: Competitive hybridization k_c versus melting temperature of the DNA•DNA duplex. Competitive hybridization above 51°Cwas not observed in these experiments.

the displacement parameter k_c for a target concentration of 3.5 nM from three different experiments for PM and MM probes (targets: pepper, potato, tobacco). There was no competitive hybridization observed for DNA•DNA duplexes with $T_m > 51$°C. This seems to be a critical temperature for the local opening of the DNA•DNA duplex. Nevertheless, in these three experiments several PM as well as MM probes were found which did not show competitive hybridization with melting temperatures of the DNA•DNA duplex < 51°C. No correlation was found to a thermodynamic parameter (e.g. DNA•RNA melting temperature or sequence related effects). A linear regression for melting temperature differences of DNA•RNA and DNA•DNA to the base composition of the three target molecules of Figure 6.11 was done in order to prove the model of nucleic acid interaction. Mismatches were not taken into account. The function

$$T_m^{DNA \bullet RNA} - T_m^{DNA \bullet DNA} \approx 0.036[A] + 0.29[C] + 0.73[G] - 0.31[T(u)] \qquad (6.17)$$

with [A],[C], [G] and [T(U)] being the number of corresponding bases was found to be a good approximation for these three experiments ($R^2 = 0.826$) Figure 6.12. While this equation obviously does not replace complex algorithms for the determination of T_m, it gives an estimation of the influence of the base content on the T_m-difference between DNA•RNA and DNA•DNA for closely related targets. The nucleotide with the largest influence on $T_m(RNA \bullet DNA) - T_m(RNA \bullet DNA)$ is guanine (G) (regression coefficient is 0.73). The higher the guanine content (in the target strand), the higher the difference in T_m is between DNA•RNA and DNA•DNA thermodynamic parameters. In general the hybrid rG•dC has a higher thermodynamic stability than rC•dG. Second

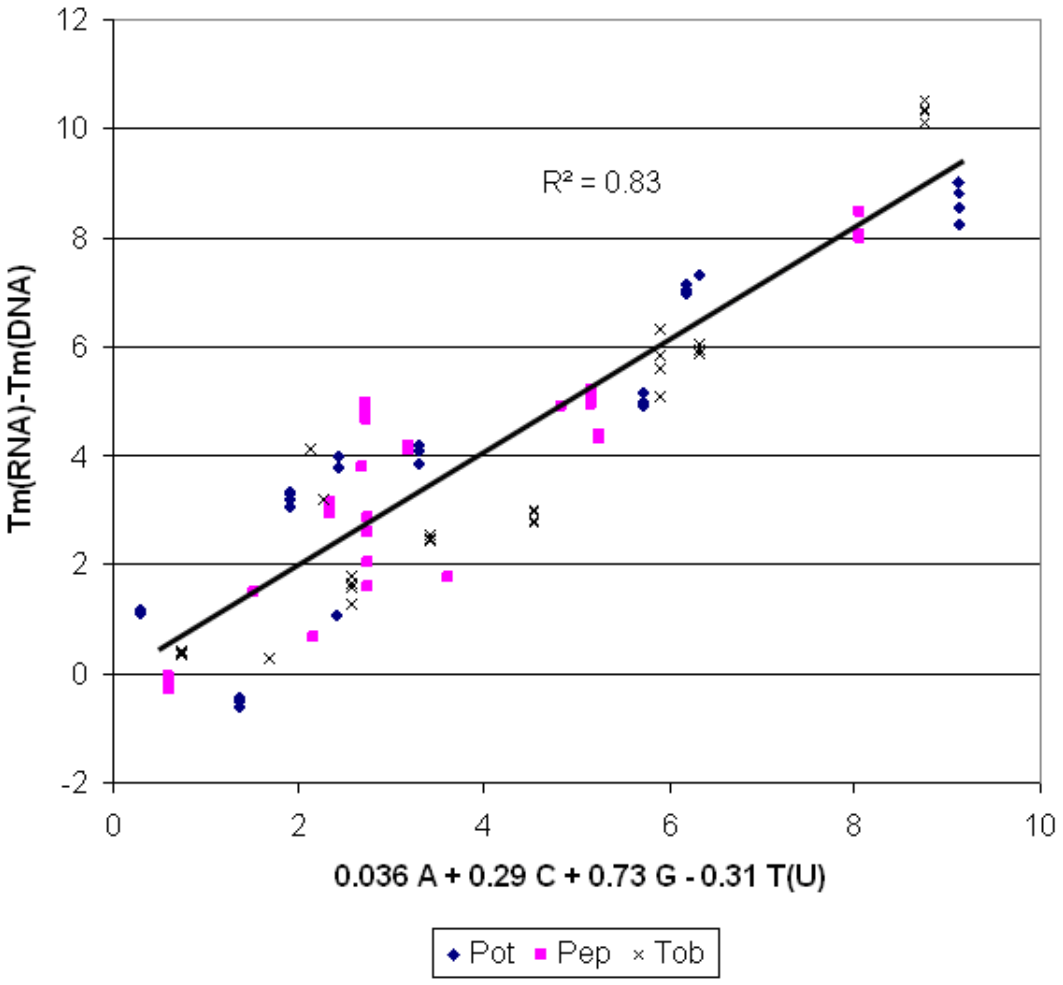


Figure 6.12: Difference in RNA and DNA melting temperatures versus linear regression function for base composition. With this correlation it is possible to estimate the influence of structural effects on the thermodynamic stability of nucleic acid duplexes. Thymine(Uracil) has a negative coefficient which is related to the CH/π? interaction of the methyl group present on thymine but missing on uracil.

largest influence has C, regression coefficient is 0.29. This is in line with the observation that a higher GC-content favors A-DNA which is the preferred structure of RNA [185]. The negative coefficient of thymine is supporting the theory of CH/π interaction. Target strands with a high uracil content in the RNA strand (missing the thymine C5-methyl group present on the DNA analogue) have a lower melting temperature than the corresponding DNA•DNA strands. rU•dA has a lower thermal stability than dT•dA. The results are in agreement with solution based nearest neighbor coefficients [88, 186, 184, 213]. Nevertheless, the calculation of T_m according to [16, 127] had however poor correlation; R^2 was in the range of 0.65. The influence of formamide is not described in the models. Our results showed a linear dependency of 0.63°CT_m decrease for every % formamide for DNA•DNA duplexes, while there was no linear relation for the DNA•RNA hybrids visible. The phenomenon of competitive displacement of one strand of DNA by RNA is related to the formation of R-loops in solution [105, 194]. R-loops are nucleic acid configurations where one strand of the dsDNA is replaced by RNA and are thought to prime replication in bacterial or viral genomic DNA [139], as

well as mammalian mitochondrial DNA [108, 109]. From solution based experiments it is known that R-loops are forming at temperatures close to the melting temperature of the DNA•DNA duplex and with high concentrations of a denaturant e.g. formamide [105, 194]. While the detection of R-loops is quite difficult in solution based experiments and the number of data on various probe-target combinations is limited, microarray based real-time analysis is a valuable tool for the systematic investigation of this basic nucleic acid behavior on a large number of probe target combination.

Competitive kinetics on probes with LNA nucleotides

For the investigation of competitive hybridization of DNA and RNA targets on DNA probes with and without LNA nucleotides a target concentrations of 3.6 nM was used. Analogous to standard microarray systems where a possible definition of specificity is via the intensity ratio of perfect match and mismatch targets the ratio between RNA and DNA target intensity was used as a parameter to quantify the higher affinity of targets with higher binding energy. This parameter was compared for the LNA and DNA probes. A typical result for 70 mer probes is 1.36 ± 0.18, thus on average the LNA spots have a 36% higher 'specificity' for the RNA targets compared to DNA. Surprisingly the competitive displacement of the DNA target was much more pronounced on the spots with LNA nucleotides incorporated (a typical example is shown in Figure 6.13. For all concentrations the competitive kinetic is significantly higher on the LNA spots. A slight probe concentration effect was visible, the smaller the density of the surface immobilized probes, the larger the competitive displacement. This is according to theory: increasing the ratio of target/probe concentration is increasing the competitive displacement due to saturation effects. For statistical analysis a nonlinear regression according to equation 6.15 was done and the ratio of the parameter k_c (rate constant for competitive displacement) was compared for probes with and without LNA nucleotides. For all the representation of k_c a threshold of 1e-6 was defined. The ratio of the displacement parameter k_c for LNA and non LNA spots is shown in Figure 6.14. For the mismatch targets potato and tobacco the competitive displacement is significantly higher on the LNA spots compared to the isosequential RNA probes. The probes have been designed to discriminate pepper against this two closely related mismatch targets with LNA nucleotides on mismatch positions. There is a clear trend that the corresponding probes show increased displacement. The same analysis was done for parameter hybridization parameter k_h (rate constant for initial hybridization) for reference purposes. No significant difference between LNA and non LNA probes was found (data not shown). Thus the hybridization kinetics for single target hybridization is not affected from the LNA nucleotides. The higher competitive hybridization seems to play a central role in the higher specificity of LNA based probes. There is an ongoing discussion about the specificity improvement due to LNA probes. Our results clearly show that a few LNA nucleotides in 70mer DNA probes change the competitive kinetic behavior significantly. The melting analysis of the 70 mers showed a significant increase in melting temperature for LNA•RNA especially for PM probes. LNA•DNA duplexes showed only minor changes. Figure 6.15 shows an overview on the melting analysis results. The big difference in T_m for PM DNA•RNA duplexes is notable, e.g. for 2LPep543vsPot it is almost 8°Cwith 4 LNA nucleotides (one in terminal position). The differences between the probes showed a good correlation of RNA purines content ($R^2 = 0.74$, data not shown).

6.5.6 Conclusion

Within this study we investigated the elementary interplay of different types of nucleic acids with surface bound DNA and DNA-LNA hybrid oligonucleotides on microarray formats. Using real time analysis and a model system based on the actin gene it was possible to relate competitive hybridization of RNA and DNA targets of the same sequence to thermodynamic parameters. Competitive

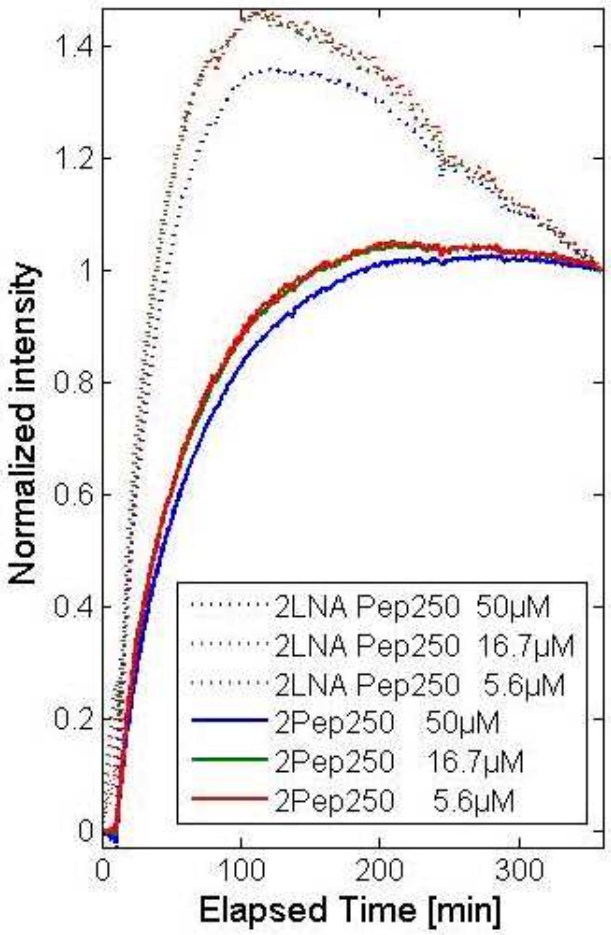


Figure 6.13: Typical result for the different normalized hybridization kinetics on probes with and w/o LNA. Target was Tobacco with 3.6 nM concentration. The corresponding melting temperatures are for Exp6: DNA 50.1 vs 49.9; RNA 56.2 vs 56.6. The difference in kinetics is not visible on the thermodynamic data.

hybridization is limited by the melting temperature of the DNA•DNA duplex. Partial opening (breathing) of the duplex might induce the exchange of the nucleic acid strands. Further research is necessary for a full understanding and quantification of this effect. At low target concentrations competitive hybridization has a large influence on microarray results. With a reference measurement it was possible to quantify the target concentration of our experiments based on the different kinetics. It was shown that 1. . . 4 LNA nucleotides in 70mer DNA probes have a significant influence on hybridization kinetics. The probes with LNA nucleotides showed in general a higher competitive displacement. The design of the probes with LNA nucleotides in purine mismatch positions proved successful within our experimental setup. While a few LNA nucleotides in the probe change the melting temperature for the RNA target significantly, there was only a minor effect for the DNA targets. The hybridization kinetics and thermodynamics to solid phase immobilized oligonucleotides is still not fully understood. The results of this study are important for the understanding of the limitation of specificity on microarrays. Further work on larger probe-target sets and modified nucleic acids might help to improve our understanding of the most elementary process in biology, the hybridization of nucleic acids.

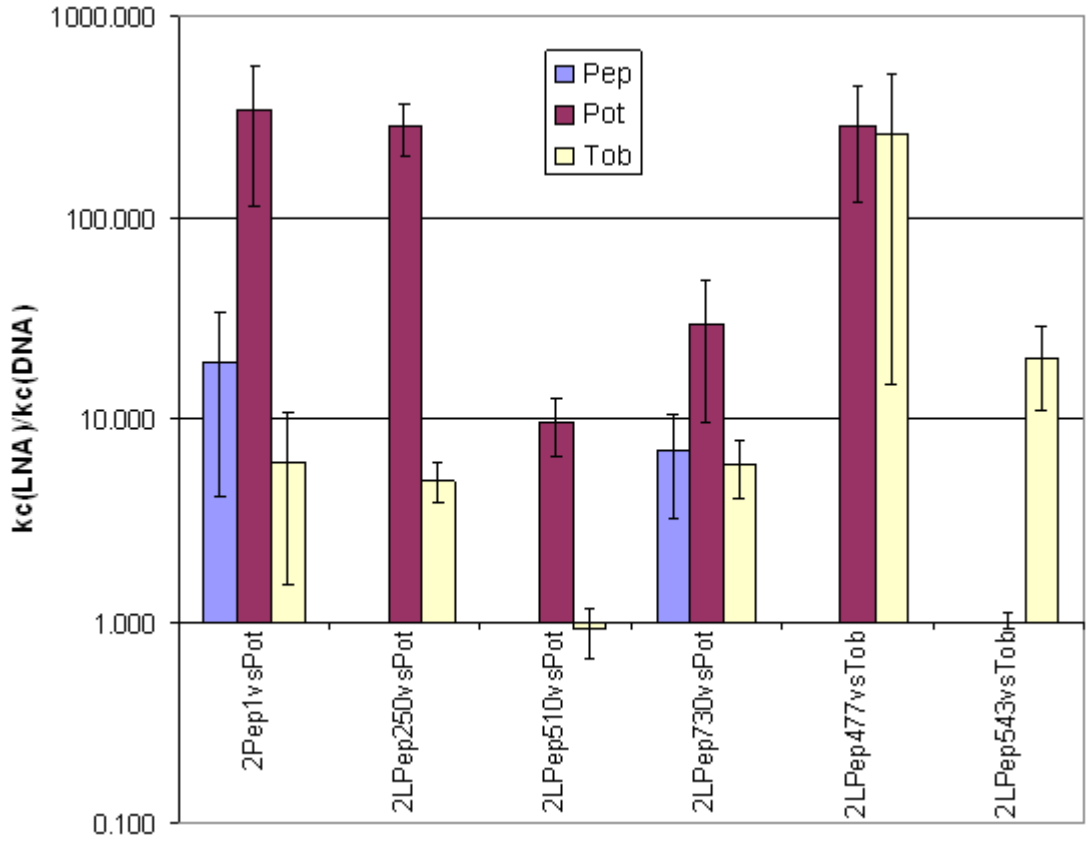


Figure 6.14: Ratio of competitive displacement for probes with and w/o LNA nucleotides. The higher the value, the more target is replaced from the probe. In general the kinetics of the mismatch targets changed more than that of the PM probes. The probes were designed to distinguish pepper against potato or tobacco. There is a trend visible that the corresponding mismatch target changes the most when using LNA nucleotides.

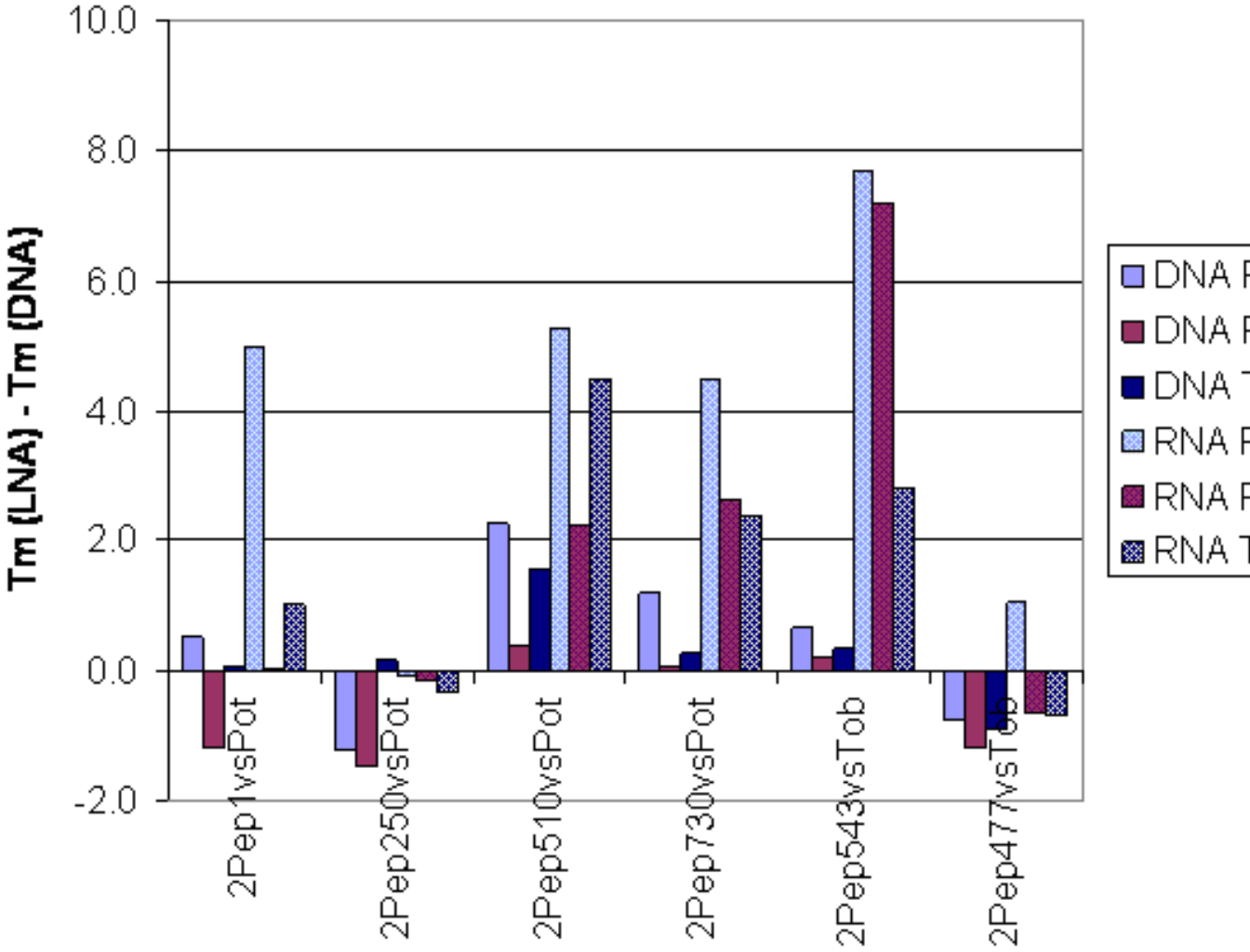


Figure 6.15: Difference in melting temperature for probes with and without LNA nuclecotides. While in case of DNA targets the change is not very significant, RNA targets show a large increase in T_m, especially for the PM target, e.g. for 2LPep 543x522vsTob 4 LNA nucleotides (one in terminal position) induce a T_m change of 7.5℃

Chapter 7

Discussion

7.1 Physics of Real-time measurement of hybridization isotherms and melting analysis on solid-phase bound nucleic acids

Based on the results of the development of the evanescent prototype a modeling of the whole signal transfer function from the probe and target concentration to the digital output of the electronic signal path was done. The major challenge of evanescent illumination is the coupling of the light into the glass slide. An alternative to the evanescent optics would be the use of confocal or nearly confocal optics. Due to the homogeneous background and the multiple measurement signal processing techniques can improve the signal to noise ratio significantly. Intercalating or FRET-based fluorophores could improve the signal to background ratio further. Nevertheless, if there is demand for higher sensitivity will be ruled from the applications. Potential for improvement of the signal to background ratio is increased spotting density (matrix surfaces), surface enhancing slide technologies or advanced labeling techniques.

7.2 Hybridization isotherms of surface based nucleic acid hybridization

Kinetics of solid phase bound nucleic acid hybridization is transport-diffusion limited. The results of the numeric analysis and basic experiments on transport-diffusion limitations showed that with an optical setup without effective agitation the melting analysis needs has to be performed very slowly. Interestingly the hybridization kinetics is not much accelerated when using agitation. In case of competitive hybridization the replacement of the unspecific target became the dominant parameter and time-to-equilibrium can be hundreds of hours. While about half of the end point signal is reached after one hour, competitive displacement can change the signal ratios afterwards significantly. Kinetic effects have a strong dependency on target concentration, thus they can be used for determination of the target concentration.

7.3 High resolution melting analysis of surface bound nucleic acid

Within this work an evaluation of microarray based melting analysis for long oligo probe sets was done. Very high reproducibility of melting point measurements ($R^2 \geq 0.98$) including dye swap

experiments and artificially generated low quality measurements was found. False positive endpoint results caused by strong unspecific hybridization could be excluded with melting analysis due to the low T_m of the respective duplexes. Two nucleotide mismatches were reliably discriminated with 60 mer oligo probes based on the shift of T_m as compared to PM probe-target combinations without normalization or calibration to reference spots. As literature data did not suggest this high degree of specificity (Kane et al., 2000; Tiquia et al., 2004), our probe set was not optimized for SNP detection. Thus, there is still upside potential for specificity of Microarray based high-resolution-metling. The shape of the melting curve provides information about cross hybridization caused by multiple targets with high similarity to the same probe. It was shown that for cross hybridization with 2 MM the slope of the melting curve changed significantly, for 4 MM a clear shoulder was visible in the melting curve. The limitations of microarray experiments in specificity as compared to solution based HRM analysis should diminish with melting analysis based methods. On-chip melting analysis provides a tool for the characterization of a large number of probes and can help to improve specificity for endpoint based assays using more accurate values for T_m as key parameter for probe design. In contrast to solutions based assays with solid phase bound oligo probes melting analysis has the potential to generate large amounts of high quality thermodynamic data for different nucleic acid types and probe lengths. These data sets could serve as basis for the development and verification of new algorithms for the prediction of thermodynamic parameters of nucleic acid hybridization enabling more accurate probe design and the development of future miniaturized surface based systems with reduced processing steps. Systematic investigation of larger probe-target sets and their thermodynamic analysis together with advanced bioinformatics will help get a better physico-chemical understanding of one of the most elementary processes in biology, the hybridization of nucleic acids.

7.4 Competitive hybridization of RNA•DNA hybrids and DNA•DNA duplexes

Using real time analysis and a model system based on the actin gene it was possible to relate competitive hybridization of RNA and DNA targets of the same sequence to the melting temperature of the DNA•DNA duplex. Competitive hybridization only occurs close to the melting temperature of the probe-target duplex. Thus it is an inherent problem of microarray analysis, hybridization temperature should be as close to the melting temperature as possible for maximum stringency. The ratios between specific and non specific hybridization are changing significantly, even after 24 hours.

7.5 Influence of the C5-methyl group

Comparing the melting temperatures of RNA•DNA hybrids and DNA•DNA duplexes it was found that for gene sequences with great homology there is a good correlation to the base composition. The higher the guanine content (or the GC content in general) in the RNA strand the higher is the difference in T_m between RNA•DNA hybrid and DNA•DNA duplex. The higher the content of uracil in the RNA strand the smaller the difference between the two. This seems to be related to the CH/π interaction of the methyl group in thymine. This simple correlation does not consider any stacking effects and is obviously not replacing NN-models. But to our knowledge it is the first time that this elementary interplay of different types of nucleic acids was investigated on surface bound oligonucleotides on microarray formats.

7.6 Outlook

Real-time data acquisition during hybridization and denaturation has proven to be a reliable method for the analysis of biological samples. During this work we could establish cooperations with Infineon Austria and Sony Japan with the goal of a joint development in the field of biosensors. With the help of competent industrial partners it will be possible to make real-time measurement to a accurate and robust technology for Lab-on-chip systems. We could show that it is possible to do detect SNPs even with long oligo systems. The development of detection systems for polymorphic gene regions and pathogenic microorganisms will be another focus in future research. While nowadays there is a strong focus on next-generation sequencing technologies there will be a strong need for reliable and cheap SNP detection including preamplification from biological samples. The line between sequencing technologies and andvanced real-time hybridization technologies (e.g. primer extension, minisequencing) is blurred. In addition to sequence information it is possible with real-time hybridization measurement to learn about the collective behavior of bases. There is still no discussion about systematic errors in sequencing technologies. This could be a critical point for SNP detection. A systematic investigation of a large number of probe-target combinations will allow the extraction of parameters for testing the existing Nearest-Neighbor models for DNA•DNA, DNA•RNA and RNA•RNA duplexes. Especially the method of competitive hybridization of isosequential nucleic acids with different $2'$ groups and C5-methyl group could give valuable insight into the influence of the different molecule groups on kinetics and thermodynamics of nucleic acids.

Bibliography

[1] Carolin Ahlborn, Karsten Siegmund, and Clemens Richert. Isostable dna. *Journal of the American Chemical Society*, 129(49):15218–15232, 2007. [cited at p. 62]

[2] Thomas J. Albert, Michael N. Molla, Donna M. Muzny, Lynne Nazareth, David Wheeler, Xingzhi Song, Todd A. Richmond, Chris M. Middle, Matthew J. Rodesch, Charles J. Packard, George M. Weinstock, and Richard A. Gibbs. Direct selection of human genomic loci by microarray hybridization. *Nat Meth*, 4(11):903–905, November 2007. [cited at p. 62]

[3] Toegl Andreas, Kirchner Roland, Gauer Christoph, and Wixforth Achim. Enhancing results of microarray hybridizations through microagitation. *J Biomol Tech*, 14(3):197–204, 2003. [cited at p. 90]

[4] D. Axelrod and M.D. Wang. Reduction-of-dimensionality kinetics at reaction-limited cell surface receptors. *Biophysical Journal*, 66(3, Part 1):588 – 600, 1994. [cited at p. 75]

[5] Calladine C.R. Drew H. Luisi B. *Understanding DNA. The Molecule and How It Works*. Academic Press, 2004. [cited at p. 10, 17]

[6] Matthias Bahr, Valerie Gabelica, Anton Granzhan, Marie-Paule Teulade-Fichou, and Elmar Weinhold. Selective recognition of pyrimidine-pyrimidine DNA mismatches by distance-constrained macrocyclic bis-intercalators. *Nucl. Acids Res.*, 36(15):5000–5012, 2008. [cited at p. 54]

[7] Christian Bailly, Stephen Crow, Andrew Minnock, and Michael J. Waring. Demethylation of thymine residues affects dna cleavage by endonucleases but not sequence recognition by drugs. *Journal of Molecular Biology*, 291(3):561 – 573, 1999. [cited at p. 23, 112]

[8] S. K. Banik, T. Ambjornsson, and R. Metzler. Stochastic approach to dna breathing dynamics. *EPL (Europhysics Letters)*, 71(5):852–858, 2005. [cited at p. 19]

[9] Flavia Barone, Luciano Cellai, Mirella Matzeu, Filomena Mazzei, and Francesco Pedone. Dna, rna and hybrid rna-dna oligomers of identical sequence: structural and dynamic differences. *Biophysical Chemistry*, 86(1):37–47, 2000. [cited at p. 111]

[10] L. Belloni. On fermi's route to fermi-dirac statistics. *Eur. J. Phys.*, 15:102–109, 1994. [cited at p. 83]

[11] H. Binder and S. Preibisch. Specific and nonspecific hybridization of oligonucleotide probes on microarrays. *Biophysical Journal*, 89(1):337–352, 2005. [cited at p. 62]

[12] J. Bishop, S. Blair, and A. M. Chagovetz. A competitive kinetic model of nucleic acid surface hybridization in the presence of point mutants. *Biophysical Journal*, 90(3):831–840, 2006. [cited at p. 110]

[13] J. Bishop, A. M. Chagovetz, and S. Blair. Kinetics of multiplex hybridization: Mechanisms and implications. *Biophysical Journal*, 94(5):1726–1734, 2008. [cited at p. 71, 89, 93, 113]

[14] J. Bishop, C. Wilson, A. M. Chagovetz, and S. Blair. Real-time optical detection of competitive surface hybridization on microarrays. *Proc SPIE*, 6430(1):643002–643006, 2007. [cited at p. 89, 110, 114]

[15] J. Bishop, C. Wilson, A.M. Chagovetz, and S. Blair. Competitive displacement of dna during surface hybridization. *Biophysical Journal*, 92(1):L10 – L12, 2007. [cited at p. 90, 103, 110, 113]

[16] RD Blake, JW Bizzaro, JD Blake, GR Day, SG Delcourt, J Knowles, KA Marx, and Jr SantaLucia, J. Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics*, 15(5):370–375, 1999. [cited at p. 91, 92, 117]

[17] RD Blake and SG Delcourt. Thermodynamic effects of formamide on DNA stability. *Nucl. Acids Res.*, 24(11):2095–2103, 1996. [cited at p. 21, 42]

[18] RD Blake and SG Delcourt. Thermal stability of DNA. *Nucl. Acids Res.*, 26(14):3323–3332, 1998. [cited at p. 51]

[19] L Bodrossy, N Stralis-Pavese, JC Murrell, S Radajewski, A Weilharter, and A Sessitsch. Development and validation of a diagnostic microbial microarray for methanotrophs. *ENVIRONMENTAL MICROBIOLOGY*, 5(7):566–582, JUL 2003. [cited at p. 64, 75, 84]

[20] Dwaine A. Braasch and David R. Corey. Locked nucleic acid (lna): fine-tuning the recognition of dna and rna. *Chemistry & Biology*, 8(1):1–7, 2001. [cited at p. 111]

[21] K J Breslauer, R Frank, H Blocker, and L A Marky. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 83(11):3746–3750, 1986. [cited at p. 51, 54]

[22] Alexander Chagovetz and Steve Blair. Real-time dna microarrays: reality check. *Biochemical Society Transactions*, 037(2):471–475, 2009. 10.1042/BST0370471. [cited at p. 89, 111]

[23] Tigran V. Chalikian, Jens Voelker, G. Eric Plum, and Kenneth J. Breslauer. A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14):7853–7858, 1999. [cited at p. 59]

[24] V Chan, D.J. Graves, and S.E. McKenzie. The biophysics of dna hybridization with immobilized oligonucleotide probes. *Biophysical Journal*, 69(6):2243 – 2255, 1995. [cited at p. 30]

[25] Chunlai Chen, Wenjuan Wang, Jing Ge, and Xin Sheng Zhao. Kinetics and thermodynamics of DNA hybridization on gold nanoparticles. *Nucl. Acids Res.*, 37(11):3756–3765, 2009. [cited at p. 54]

[26] Chunlai Chen, Wenjuan Wang, Zhang Wang, Fang Wei, and Xin Sheng Zhao. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucl. Acids Res.*, 35(9):2875–2884, 2007. [cited at p. 54]

[27] Y. Z. Chen, V. Mohan, and R. H. Griffey. Base opening in rna and dna duplexes: Implication for rna stability. *Phys. Rev. E*, 61(5):5640–5645, May 2000. [cited at p. 20, 111, 113]

[28] Chu H. Choi, George Kalosakas, Kim O. Rasmussen, Makoto Hiromura, Alan R. Bishop, and Anny Usheva. DNA dynamically directs its own transcription initiation. *Nucl. Acids Res.*, 32(4):1584–1590, 2004. [cited at p. 16, 19, 20]

[29] Shan-Ho Chou, Ko-Hsin Chin, and Andrew H.-J. Wang. Unusual DNA duplex and hairpin motifs. *Nucl. Acids Res.*, 31(10):2461–2474, 2003. [cited at p. 23]

[30] Ulla Christensen. Thermodynamic and kinetic characterization of duplex formation between 2′-o, 4′c -methylene-modified oligoribonucleotides, dna and rna. *Bioscience Reports*, 27(6):327–333, 2007. [cited at p. 111]

[31] Valentino R. Cooper, Timo Thonhauser, Aaron Puzder, Elsebeth Schr+der, Bengt I. Lundqvist, and David C. Langreth. Stacking interactions and the twist of dna. *Journal of the American Chemical Society*, 130(4):1304–1308, 2007. doi: 10.1021/ja0761941. [cited at p. 112]

[32] Athel Cornish-Bowden. Enthalpy-entropy compensatio: a phantom phenomenon. *Journal of Biosciences*, 27(2):121–126, 2002. [cited at p. 48]

[33] G. Cosa, K.S. Focsaneanu, J.R.N. McLean, J.P. McNamee, and J.C. Scaiano. Photophysical properties of fluorescent dna-dyes bound to single- and double-stranded dna in aqueous buffered solution. *Photochemistry and Photobiology*, 73(6):585–599, 2007. [cited at p. 84]

[34] Lukacs L.G. Haggie P. Seksek O. Lechardeur D. Size-dependent dna mobility in cytoplasm and nucleus. *Proc. Natl. Acad. Sci. U.S.A.*, 275:1625–1629, 2000. [cited at p. 72, 73]

[35] Wallace M.I. Ying L. Balasubamanian S. Klenerman D. Non-arrhenius kinetics for the loop closure of a dna hairpin. *Proc. Natl. Acad. Sci. U.S.A.*, 98:5584–5589, 2000. [cited at p. 29, 74]

[36] Christina Dahl and Per Guldberg. High-Resolution Melting for Accurate Assessment of DNA Methylation. *Clin Chem*, 53(11):1877–1878, 2007. [cited at p. 63, 89]

[37] Hongyue Dai, Michael Meyer, Sergey Stepaniants, Michael Ziman, and Roland Stoughton. Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Research*, 30(16):e86, 2002. 10.1093/nar/gnf085. [cited at p. 89, 111]

[38] Thierry Dauxois, Michel Peyrard, and A. R. Bishop. Dynamics and thermodynamics of a nonlinear model for dna denaturation. *Phys. Rev. E*, 47(1):684–695, Jan 1993. [cited at p. 20]

[39] Thierry Dauxois, Michel Peyrard, and A. R. Bishop. Entropy-driven dna denaturation. *Phys. Rev. E*, 47(1):R44–R47, Jan 1993. [cited at p. 20]

[40] RE DDickerson. Definition and nomenclature of nucleic acid structure parameters. *Journal of biomolecular structure & dynamics*, 6(4):627–634, Feb 1989. [cited at p. 18, 21]

[41] Robertson R.M. Laib S. Smith D.E. Diffusion of isolated dna molecules: Dependence on length and topology. *Proc. Natl. Sci. U.S.A*, 103:7310–7314, 2006. [cited at p. 29, 65, 71, 72, 84]

[42] Vadim V. Demidov and Maxim D. Frank-Kamenetskii. Two sides of the coin: affinity and specificity of nucleic acid interactions. *Trends in Biochemical Sciences*, 29(2):62–71, 2004. [cited at p. 111]

[43] Ye Deng, Zhili He, Joy Van Nostrand, and Jizhong Zhou. Design and analysis of mismatch probes for long oligonucleotide microarrays. *BMC Genomics*, 9(1):491–496, 2008. [cited at p. 55, 63]

[44] Scott J. Diede, Jamie Guenthoer, Linda N. Geng, Sarah E. Mahoney, Michael Marotta, James M. Olson, Hisashi Tanaka, and Stephen J. Tapscott. Dna methylation of developmental genes in pediatric medulloblastomas identified by denaturation analysis of methylation differences. *Proceedings of the National Academy of Sciences*, 107(1):234–239, 2010. 10.1073/pnas.0907606106. [cited at p. 112]

[45] Roumen A. Dimitrov and Michael Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophysical Journal*, 87(1):215–226, 2004. [cited at p. 85]

[46] P. A. M. Dirac. On the Theory of Quantum Mechanics. *Proceedings of the Royal Society of London. Series A*, 112(762):661–677, 1926. [cited at p. 83]

[47] James M. Dixon, Masahiko Taniguchi, and Jonathan S. Lindsey. Photochemcad 2: A refined program with accompanying spectral databases for photochemical calculations. *Photochemistry and Photobiology*, 81(1):212–213, 2009. doi: 10.1562/2004-11-06-TSN-361.1. [cited at p. 36]

[48] Utz Dornberger, Mikael Leijon, and Hartmut Fritzsche. High Base Pair Opening Rates in Tracts of GC Base Pairs. *Journal of Biological Chemistry*, 274(11):6957–6962, 1999. [cited at p. 16, 19, 20, 22, 113]

[49] Martin Dufva, Jesper Petersen, and Lena Poulsen. Increasing the specificity and function of dna microarrays by processing arrays at different stringencies. *Analytical and Bioanalytical Chemistry*, 395(3):669–677, 2009. [cited at p. 63, 89]

[50] Karen J. Edwards, David G. Brown, Neil Spink, Jane V. Skelly, and Stephen Neidle. Molecular structure of the b-dna dodecamer d(cgcaaatttgcg)2 an examination of propeller twist and minor-groove water structure at 22resolution. *Journal of Molecular Biology*, 226(4):1161 – 1173, 1992. [cited at p. 19]

[51] M Ellemann Nielsen, KM an Petersen, Hakansson AE, J Wemgel, and PJ Jacobsen. α-l-lna (α-l-*ribo* configured locked nucleic acid) recognition of dna:an nmr spectroscopic study. *Chemistry - A European Journal*, 8(13):3001–3009, 2002. [cited at p. 13]

[52] Borden J.R. Paredes C.J. Papoutsakis E.T. Diffusion, mixing and associated dye effects in dna-microarray hybridizations. *Biophys. J.*, 89:3277–3284, 2005. [cited at p. 73]

[53] Oleg Yu Fedoroff, Miguel Salazar, and Brian R. Reid. Structure of a dna : Rna hybrid duplex : Why rnase h does not cleave pure rna. *Journal of Molecular Biology*, 233(3):509–523, 1993. [cited at p. 112]

[54] E. Fermi. Sulla quantizzazione del gas perfetto monoatomico. *Rend. Accad. Naz. Lincei*, 3:145–149, 1926. [cited at p. 83]

[55] Alessandro Ferrantini, Joke Allemeersch, Paul Van Hummelen, and Enrico Carlon. Thermodynamic scaling behavior in genechips. *BMC Bioinformatics*, 10(1):3, 2009. [cited at p. 42]

[56] Daniel J. Fish, M. Todd Horne, Greg P. Brewood, Jim P. Goodarzi, Saba Alemayehu, Ashwini Bhandiwad, Robert P. Searles, and Albert S. Benight. DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison. *Nucl. Acids Res.*, 35(21):7197–7208, 2007. [cited at p. 104]

[57] Jrg Fohrer, Mirko Hennig, and Teresa Carlomagno. Influence of the 2'-hydroxyl group conformation on the stability of a-form helices in rna. *Journal of Molecular Biology*, 356(2):280 – 287, 2006. [cited at p. 11]

[58] AV Fotin, AL Drobyshev, DY Proudnikov, AN Perov, and AD Mirzabekov. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucl. Acids Res.*, 26(6):1515–1521, 1998. [cited at p. 42]

[59] R. H. Fowler. On dense matter. *Monthly Notices of the Royal Astronomical Society*, 87:114–122, 1926. [cited at p. 84]

[60] SM Freier and KH Altmann. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucl. Acids Res.*, 25(22):4429–4443, 1997. [cited at p. 12, 112]

[61] Miriam Frieden, Signe M. Christensen, Nikolaj D. Mikkelsen, Christoph Rosenbohm, Charlotte A. Thrue, Majken Westergaard, Henrik F. Hansen, Henrik Orum, and Troels Koch. Expanding the design horizon of antisense oligonucleotides with alpha-L-LNA. *Nucl. Acids Res.*, 31(21):6365–6372, 2003. [cited at p. 13, 14, 15]

[62] Chikara Furusawa, Naoaki Ono, Shingo Suzuki, Tomoharu Agata, Hiroshi Shimizu, and Tetsuya Yomo. Model-based analysis of non-specific binding for background correction of high-density oligonucleotide microarrays. *Bioinformatics*, 25(1):36–41, 2009. 10.1093/bioinformatics/btn570. [cited at p. 89]

[63] Goodson V.H. Hawse F.W. Molecular evolution of the actin family. *Journal of Cell Science*, 115:2619–2622, 2002. [cited at p. 63, 111]

[64] Chetan Gadgil, Andrew Yeckel, Jeffrey J. Derby, and Wei-Shou Hu. A diffusion-reaction model for dna microarray assays. *Journal of Biotechnology*, 114(1-2):31 – 45, 2004. [cited at p. 90, 114]

[65] Yang Gao, Lauren K. Wolf, and Rosina M. Georgiadis. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucl. Acids Res.*, 34(11):3370–3377, 2006. [cited at p. 30]

[66] Genewave. Hyblive™. Product data sheet, Genewave, 2008. http://www.genewave.com/products.php?id=8. [cited at p. 9, 63]

[67] Bonifacio G.F., T. Brown, G.L. Conn, and A.N.; Lane. Comparison of the electrophoretic and hydrodynamic properties of dna and rna oligonucleotide duplexes. *Biophysical Journal*, 73(3):1532–1538, 1997. [cited at p. 59]

[68] Marc Glazer, Jacqueline A. Fidanza, Glenn H. McGall, Mark O. Trulson, Jon E. Forman, Audrey Suseno, and Curtis W. Frank. Kinetics of oligonucleotide hybridization to photolithographically patterned dna arrays. *Analytical Biochemistry*, 358(2):225 – 238, 2006. [cited at p. 103]

[69] O Gotoh and Y Tagashira. Stabilities of nearest-neighbor doublets in double-helical dna determined by fitting calculated melting profiles to observed profiles. *Biopolymers*, 20(5):1033–1042, 1981. [cited at p. 51]

[70] Donald Gray, Carla Gray, Byong-Hoon Yoo, and Tzu-Fang Lou. Antisense dna parameters derived from next-nearest-neighbor analysis of experimental data. *BMC Bioinformatics*, 11(1):252, 2010. [cited at p. 50]

[71] David Gresham, Bo Curry, Alexandra Ward, D. Benjamin Gordon, Leonardo Brizuela, Leonid Kruglyak, and David Botstein. Optimized detection of sequence variation in heterozygous genomes using dna microarrays with isothermal-melting probes. *Proceedings of the National Academy of Sciences*, 107(4):1482–1487, 2010. 10.1073/pnas.0913883107. [cited at p. 89]

[72] A. Yu. Grosberg, T. T. Nguyen, and B. I. Shklovskii. Colloquium: The physics of charge inversion in chemical and biological systems. *Rev. Mod. Phys.*, 74(2):329–345, Apr 2002. [cited at p. 25]

[73] H. Gudnason, M. Dufva, D.D. Bang, and A Wolff. An inexpensive and simple method for thermally stable immobilization of dna on an unmodified glass surface: Uv linking of poly(t)10-poly(c)10tagged dna probes. *BioTechniques*, 45(3):261–271, 2008. [cited at p. 31]

[74] Anders Gunnarsson, Peter Jonsson, Vladimir P. Zhdanov, and Fredrik Hook. Kinetic and thermodynamic characterization of single-mismatch discrimination using single-molecule imaging. *Nucleic Acids Research*, 37(14):e99, 2009. 10.1093/nar/gkp487. [cited at p. 63, 89, 91, 111, 113]

[75] Jeffrey I. Gyi, Andrew N. Lane, Graeme L. Conn, and Tom Brown. Solution structures of dna-rna hybrids with purine-rich and pyrimidine-rich strands: Comparison with the homologous dna and rna duplexes. *Biochemistry*, 37:73–80, 1998. [cited at p. 12, 59, 112]

[76] J.I. Gyi, G.L. Conn, A.N. Lane, and T. Brown. Comparison of the thermodynamic stabilities and solution conformations of dna-rna hybrids containing purine-rich and pyrimidine-rich strands with dna and rna duplexes. *Biochemistry*, 35(38):12538–12548, 1996. [cited at p. 59, 105]

[77] A Halperin, A Buhot, and E B Zhulina. On the hybridization isotherms of dna microarrays: the langmuir model and its extensions. *Journal of Physics: Condensed Matter*, 18(18):S463–S490, 2006. [cited at p. 63, 110]

[78] A. Halperin, A. Buhot, and E.B. Zhulina. Sensitivity, specificity, and the hybridization isotherms of dna chips. *Biophysical Journal*, 86(2):718 – 730, 2004. [cited at p. 71]

[79] A Halperin, A Buhot, and EB Zhulina. Hybridization at a surface: The role of spacers in dna microarrays. *Langmuir*, 22(26):11290–11304, 2006. [cited at p. 30]

[80] Zhili He, Liyou Wu, Matthew W. Fields, and Jizhong Zhou. Use of Microarrays with Different Probe Sizes for Monitoring Gene Expression. *Appl. Environ. Microbiol.*, 71(9):5154–5162, 2005. [cited at p. 30]

[81] Rudolf Heer, Moritz Eggeling, Joerg Schotter, Christa Nhammer, Rudolf Pichler, Markus Mansfeld, and Hubert Brueckl. Acceleration of incubation processes in dna bio chips by magnetic particles. *Journal of Magnetism and Magnetic Materials*, 311(1):244 – 248, 2007. Proceedings of the Sixth International Conference on the Scientific and Clinical Applications of Magnetic Carriers - SCAMC-06. [cited at p. 90]

[82] T. Heim, L.-C. Tranchevent, E. Carlon, and G. T. Barkema. Physical-chemistry-based analysis of affymetrix microarray data. *The Journal of Physical Chemistry B*, 110(45):22786–22795, 2006. [cited at p. 48]

[83] U Heinemann and M Hahn. C-C-A-G-G-C-m5C-T-G-G. Helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G. *Journal of Biological Chemistry*, 267(11):7332–7341, 1992. [cited at p. 23]

[84] G. A. Held, G. Grinstein, and Y. Tu. Modeling of DNA microarray data by using physical properties of hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7575–7580, 2003. [cited at p. 48]

[85] Mark G. Herrmann, Jacob D. Durtschi, L. Kathryn Bromley, Carl T. Wittwer, and Karl V. Voelkerding. Amplicon DNA Melting Analysis for Mutation Scanning and Genotyping: Cross-Platform Comparison of Instruments and Dyes. *Clin Chem*, 52(3):494–503, 2006. [cited at p. 41, 63, 89, 91]

[86] J. Hooyberghs, P. Van Hummelen, and E. Carlon. The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters. *Nucl. Acids Res.*, 37(7):e53–, 2009. [cited at p. 41, 54, 55, 89]

[87] W.Mathias Howell, Magnus Jobs, Ulf Gyllensten, and Anthony J. Brookes. Dynamic allele-specific hybridization. *Nat Biotech*, 17(1):87, 88 1999. [cited at p. 63, 89, 111]

[88] Yuegao Huang, Congju Chen, and Irina M. Russu. Dynamics and stability of individual base pairs in two homologous rna•dna hybrids. *Biochemistry*, 48(18):3988–3997, 2009. doi: 10.1021/bi900070f. [cited at p. 113, 117]

[89] TR Hughes, M Mao, AR Jones, J Burchard, MJ Marton, KW Shannon, SM Lefkowitz, M Ziman, JM Schelter, MR Meyer, S Kobayashi, C Davis, H Dai, YD He, SB Stephaniants, G Cavet, WL Walker, A West, E Coffey, DD Shoemaker, R Stoughton, AP Blanchard, SH Friend, and PS Linsley. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotech*, 19:342–347, 2001. [cited at p. 68]

[90] Bloomfield A.V. Crothers D.M. Tinocco I. *Nucleic Acids: Structure, Properties and Functions*. University Science Books, 2000. [cited at p. 10, 12, 15, 24, 25, 28, 41, 43, 44, 46, 60, 74, 76]

[91] An intermediate neglect of differential overlap technique for spectroscopy of transition-metal complexes. Zerner michael c. and loew gilda h. and kirchner robert f. and mueller-westerhoff ulrick t. *J. Am. Chem. Soc.*, 102(2):589–599, 1980. [cited at p. 17, 25, 26]

[92] Fiche J.B., Buhot A., Calemczuk R., and Livache T. Temperature effects on dna chip experiments from surface plasmon resonance imaging: Isotherms and melting curves. *Biophysical Journal*, 92(3):935–946, 2007. [cited at p. 91, 111, 113]

[93] Pous Joan, Urpi Lourdes, Juan A Subirana, Gouyette Catherine, Jorge Navaza, and J.Lourdes Campos. Stabilization by extra-helical thymines of a dna duplex with hoogsteen base pairs. *Journal of the American Chemical Society*, 130(21):6755–6760, 2008. [cited at p. 23]

[94] Marc Joyeux, Sahin Buyukdagli, and Michaël Sanrey. $1f$ fluctuations of dna temperature at thermal denaturation. *Phys. Rev. E*, 75(6):061914, Jun 2007. [cited at p. 59]

[95] Michael D. Kane, Timothy A. Jatkoe, Craig R. Stumpf, Jia Lu, Jeffrey D. Thomas, and Steven J. Madore. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucl. Acids Res.*, 28(22):4552–4557, 2000. [cited at p. 63, 68]

[96] Yoichi Katsumoto, Shinji Omori, Daisuke Yamamoto, Akio Yasuda, and Koji Asami. Dielectric dispersion of short single-stranded dna in aqueous solutions with and without added salt. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 75(1):011911, 2007. [cited at p. 25]

[97] E. B. Khomyakova, E. V. Dreval, A. A. Papin, and F. P. Soussaline. Osa microarray reader instrumentation and its applications for registration of on-chip real-time reactions. *IRBM*, 28(5-6):230–234, 2008. [cited at p. 63, 89, 111]

[98] E. B. Khomyakova, A. Mikhei, Steinhauser M.-C., L. Dauphinot, S. Cohen-Kaminsky, J. Rossier, F. Soussaline, and M.-C. Potier. On-chip hybridization kinetics for optimization of gene expression experiments. *BioTechniques*, 44:109–117, 2008. [cited at p. 89]

[99] Un Sik Kim, R. Schmidt, Mu Shik Jhon, and Henry Eyring. Physical Adsorption of the Quantum Gas. *Proceedings of the National Academy of Sciences of the United States of America*, 69(7):1690–1692, 1972. [cited at p. 84]

[100] Hinanit Koltai and Carmiya Weingarten-Baror. Specificity of dna microarray hybridization: characterization, effectors and approaches for data correction. *Nucleic Acids Research*, 36(7):2395–2405, 2008. 10.1093/nar/gkn087. [cited at p. 89]

[101] K. Myriam Kroll, Gerard Barkema, and Enrico Carlon. Linear model for fast background subtraction in oligonucleotide microarrays. *Algorithms for Molecular Biology*, 4(1):15, 2009. [cited at p. 89]

[102] Andrew Krueger, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. Sequence-dependent base-pair opening in dna double helix. *Biophysical Journal*, 90(9):3091 – 3099, 2006. [cited at p. 21]

[103] Wlad Kusnezow, Yana V. Syagailo, Sven Ruffer, Nina Baudenstiel, Christoph Gauer, Jorg D. Hoheisel, David Wild, and Igor Goychuk. Optimal Design of Microarray Immunoassays to Compensate for Kinetic Limitations: Theory and Experiment . *Mol Cell Proteomics*, 5(9):1681–1696, 2006. [cited at p. 90, 114]

[104] R Landgraf, KS Ramamurthi, and DS Sigman. Kinetics of spontaneous displacement of RNA from heteroduplexes by DNA. *Nucl. Acids Res.*, 24(16):3246–3252, 1996. [cited at p. 105]

[105] Ralf Landgraf, Chi hong Chen, and David S. Sigman. Double stranded scission of dna directed through sequence-specific r-loop formation. *Nucleic Acids Research*, 23(17):3524–3530, 1995. 10.1093/nar/23.17.3524. [cited at p. 117, 118]

[106] Andrew N. Lane and Terence C. Jenkins. Thermodynamics of nucleic acids and their interactions with ligands. *Quarterly Reviews of Biophysics*, 33(3):255–306, 2000. [cited at p. 41]

[107] Nicolas Le Novere. Melting, computing the melting temperature of nucleic acid duplex. *Bioinformatics*, 17(12):1226–1227, 2001. 10.1093/bioinformatics/17.12.1226. [cited at p. 63]

[108] Daniel Y Lee and Clayton David A. Properties of a primer rna-dna hybrid at the mouse mitochondrial dna leading-strand origin of replication. *Proc Natl Acad Sci U S A*, 73(7):2294–2298, 1996. [cited at p. 103, 118]

[109] Daniel Y Lee and Clayton David A. Initiation of mitochondrial dna replication by transcription and r-loop processing. *Journal of Biological Chemistry*, 273:30614–30621, 1998. [cited at p. 118]

[110] Inhan Lee, Alan A. Dombkowski, and Brian D. Athey. Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. *Nucl. Acids Res.*, 32(2):681–690, 2004. [cited at p. 54, 55]

[111] German Gaston Leparc, Thomas Tuchler, Gerald Striedner, Karl Bayer, Peter Sykacek, Ivo L. Hofacker, and David P. Kreil. Model-based probe set optimization for high-performance microarrays. *Nucleic Acids Research*, 37(3):e18, 2009. 10.1093/nar/gkn1001. [cited at p. 55, 57, 89]

[112] Shuzhao Li, Alex Pozhitkov, and Marius Brouwer. A competitive hybridization model predicts probe signal intensity on high density DNA microarrays. *Nucl. Acids Res.*, page gkn740, 2008. [cited at p. 114]

[113] Thorsten Liebermann, Wolfgang Knoll, Peter Sluka, and Rupert Herrmann. Complement hybridization from solution to surface-attached probe-oligonucleotides observed by surface-plasmon-field-enhanced fluorescence spectroscopy. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 169(1-3):337–350, 2000. [cited at p. 111]

[114] Walt F. Lima and Stanley T. Crooke. Binding affinity and specificity of escherichia coli rnase h1: Impact on the kinetics of catalysis of antisense oligonucleotiderna hybrids. *Biochemistry*, 36(2):390–398, 1997. PMID: 9003192. [cited at p. 105]

[115] Steffen Lindek, Christoph Cremer, and Ernst H. K. Stelzer. Confocal theta fluorescence microscopy with annular apertures. *Appl. Opt.*, 35(1):126–130, 1996. [cited at p. 36]

[116] E H Linfoot and E Wolf. Diffraction images in systems with an annular aperture. *Proceedings of the Physical Society. Section B*, 66(2):145–149, 1953. [cited at p. 36]

[117] Aleksey Lomakin and Maxim D. Frank-Kamenetskii. A theoretical analysis of specificity of nucleic acid interactions with oligonucleotides and peptide nucleic acids (pnas). *Journal of Molecular Biology*, 276(1):57–70, 1998. [cited at p. 111]

[118] M S Loth and B I Shklovskii. Non-mean-field screening by multivalent counterions. *Journal of Physics: Condensed Matter*, 21(42):424104 (8pp), 2009. [cited at p. 84]

[119] Leroy Jean Louis, Charretier Eric, Kochoyan Michel, and Gueron Maurice. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to dna curvature. *Biochemistry*, 27(25):8894–8898, 2002. [cited at p. 22]

[120] W Ludwig, O Strunk, R Westram, L Richter, H Meier, I Yadhukumar, A Buchner, T Lai, S Steppi, G Jobb, W Forster, I Brettske, S Gerber, AW Ginhart, O Gross, S Grumann, S Hermann, R Jost, A Konig, T Liss, R Lussmann, M May, B Nonhoff, B Reichel, R Strehlow, A Stamatakis, N Stuckmann, A Vilbig, M Lenke, T Ludwig, A Bode, and KH Schleifer. ARB: a software environment for sequence data. *Nucl. Acids Res.*, 32(4):1363–1371, 2004. [cited at p. 64]

[121] R Luo, HSR Gillson, MJ Potter, and Gilson MK. The physical basis of nucleic acid base stacking in water. *Biophysical Journal*, 80:140–148, 2001. [cited at p. 16]

[122] Sanborn M.E. Connolly B.K. Gurunathan K. Levitus M. Fluorescence properties and photophysics of the sulfoindocyanine cy3 linked covalently to dna. *J. Phys. Chem. B*, 111(37), 2007. [cited at p. 84, 91]

[123] Mehrdad Majlessi and Michael M. Becker. Formation of the double helix: a mutational study. *Nucl. Acids Res.*, 36(9):2981–2989, 2008. [cited at p. 76]

[124] Luisa A. Marcelino, Vadim Backman, Andres Donaldson, Claudia Steadman, Janelle R. Thompson, Sarah Pacocha Preheim, Cynthia Lien, Eelin Lim, Daniele Veneziano, and Martin F. Polz. Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. *Proceedings of the National Academy of Sciences*, 103(37):13629–13634, 2006. [cited at p. 62]

[125] Yann Marcy, Cousin P.-Y., M Rattier, G.Escalier G. Cerovic, G. Bna, M. Guron, L McDonagh, F. Boulaire le, Bnisty H., C. Weisbuch, and Avarre J.C. Innovative integrated system for real-time measurement of hybridization and melting on standard format microarrays. *BioTechniques*, 44(7):913–920, 2008. [cited at p. 9, 63, 64, 89, 111]

[126] Nicholas R. Markham and Michael Zuker. DINAMelt web server for nucleic acid melting prediction. *Nucl. Acids Res.*, 33(suppl 2):W577–581, 2005. [cited at p. 46]

[127] Nicholas R. Markham and Michael Zuker. Unafold. *Springer Protocols*, 453:3–31, May 2008. [cited at p. v, 45, 51, 63, 85, 91, 92, 117]

[128] J Marmur and P Doty. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J Mol Biol*, 5:109–118, 1962. [cited at p. 44]

[129] E. C. McKinney and R. B. Meagher. Members of the Arabidopsis Actin Gene Family Are Widely Dispersed in the Genome. *Genetics*, 149(2):663–675, 1998. [cited at p. 63]

[130] Ralf Metzler, Tobias Ambjornsson, Andreas Hanke, and Hans C Fogedby. Single dna denaturation and bubble dynamics. *Journal of Physics: Condensed Matter*, 21(3):034111 (14pp), 2009. [cited at p. 19]

[131] M Moniz de Sa and G Drouin. Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol*, 13(9):1198–1212, 1996. [cited at p. 63]

[132] Robert G. Mortimer. *Physical Chemistry*. Elsvier Academic Press, 3rd edition, 2008. [cited at p. 61]

[133] Jean Muller, Andre Mehlen, Guillaume Vetter, Mikalai Yatskou, Arnaud Muller, Frederic Chalmel, Olivier Poch, Evelyne Friederich, and Laurent Vallar. Design and evaluation of actichip, a thematic microarray for the study of the actin cytoskeleton. *BMC Genomics*, 8(1):294, 2007. [cited at p. 63]

[134] R. Murugan. Revised theory on dna renaturation kinetics and its experimental verification. *Biochemical and Biophysical Research Communications*, 293(2):870 – 873, 2002. [cited at p. 76]

[135] K. S. Nagapriya, A. K. Raychaudhuri, and Dipankar Chatterji. Direct observation of large temperature fluctuations during dna thermal denaturation. *Phys. Rev. Lett.*, 96(3):038102, Jan 2006. [cited at p. 59]

[136] Thomas Naiser, Oliver Ehler, Jona Kayser, Timo Mai, Wolfgang Michel, and Albrecht Ott. Impact of point-mutations on the hybridization affinity of surface-bound dna/dna and rna/dna oligonucleotide-duplexes: Comparison of single base mismatches and base bulges. *BMC Biotechnology*, 8(1):48, 2008. [cited at p. 55, 68]

[137] S Nakano, M Fujimoto, H Hara, and N Sugimoto. Nucleic acid duplex stability: influence of base composition on cation effects. *Nucl. Acids Res.*, 27(14):2957–2965, 1999. [cited at p. 60]

[138] Motohiro Nishio, M Hirota, and Yoji Umezawa. *The CH/π Interaction. Evidence, Nature, and Consequences*. Wiley VCH, New York, 1998. http://www.tim.hi-ho.ne.jp/dionisio. [cited at p. 23, 112]

[139] Nancy G Nossal, Kathleen C Dudas, and Kenneth N Kreuzer. Bacteriophage t4 proteins replicate plasmids with a preformed r loop at the t4 ori(uvsy) replication origin in vitro. *Molecular Cell*, 7:31–41, 2007. [cited at p. 117]

[140] Marcin Nowotny, Susana M. Cerritelli, Rodolfo Ghirlando, Sergei A. Gaidamakov, Robert J. Crouch, and Wei Yang. Specific recognition of rna/dna hybrid and enhancement of human rnase h1 activity by hbd. *EMBO J*, 27(7):1172–1181, 2008. 10.1038/emboj.2008.44. [cited at p. 105]

[141] Agnes Noy, Alberto Perez, Manuel M+rquez, F. Javier Luque, and Modesto Orozco. Structure, recognition properties, and flexibility of the dna-rna hybrid. *Journal of the American Chemical Society*, 127(13):4910–4920, 2005. doi: 10.1021/ja043293v. [cited at p. 10]

[142] Y Okahata, M Kawase, K Niikura, F Ohtake, H Furusawa, and Yasuhito Ebara. Kinetic measurements of dna hybridization on an oligonucleotide-immobilized 27-mhz quartz crystal microbalance. *Analytical Chemistry*, 70(7):1288–1296, 1998. [cited at p. 30]

[143] Naoaki Ono, Shingo Suzuki, Chikara Furusawa, Tomoharu Agata, Akiko Kashiwagi, Hiroshi Shimizu, and Tetsuya Yomo. An improved physico-chemical model of hybridization on high-density oligonucleotide microarrays. *Bioinformatics*, 24(10):1278–1285, 2008. [cited at p. 89]

[144] Yongping Pan and Jr MacKerell, Alexander D. Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucl. Acids Res.*, 31(24):7131–7140, 2003. [cited at p. 16, 113]

[145] Alejandro Panjkovich and Francisco Melo. Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, 21(6):711–722, 2005. [cited at p. 41]

[146] Soo-Jin Park and Jean-Baptiste Donnet. Evaluation of the distribution function of adsorption site energies based on the fermi-dirac's law in a monolayer. *Journal of Colloid and Interface Science*, 200(1):46 – 51, 1998. [cited at p. 84, 91]

[147] Michael Petersen, Kent Bondensgaard, Jesper Wengel, and Jens Peter Jacobsen. Locked nucleic acid (lna) recognition of rna: Nmr solution structures of lna:rna hybrids. *Journal of the American Chemical Society*, 124(21):5974–5982, 2002. [cited at p. 14, 71]

[148] Michael Petersen, Anders E. Hakansson, Jesper Wengel, and Jens Peter Jacobsen. $\alpha-l-$lna ($\alpha-l-ribo$ configured locked nucleic acid) recognition of rna. a study by nmr spectroscopy and molecular dynamics simulations. *Journal of the American Chemical Society*, 123(30):7431–7432, 2001. [cited at p. 111]

[149] Michael Petersen and Jesper Wengel. Lna: a versatile tool for therapeutics and genomics. *Trends in Biotechnology*, 21(2):74–81, 2003. [cited at p. 111]

[150] AW Peterson, Wolf LK, and Georgiadis RM. Hybridization of mismatched or partially matched dna at surfaces. *Journal of the American Chemical Society*, 124(49):14601–14607, 2002. [cited at p. 33, 89]

[151] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for dna denaturation. *Phys. Rev. Lett.*, 62(23):2755–2758, Jun 1989. [cited at p. 16, 20]

[152] Michel Peyrard. Nonlinear dynamics and statistical physics of dna. *Nonlinearity*, 17(2):R1–R40, 2004. [cited at p. 50]

[153] XianYu Piao, Ying Yan, Jing Yan, and YiFu Guan. Enhanced recognition of non-complementary hybridization by single-lna-modified oligonucleotide probes. *Analytical and Bioanalytical Chemistry*, 394(6):1637–1643, 2009. [cited at p. 111]

[154] Lena Poulsen, Martin Jensen Soe, Detlef Snakenborg, Lisbeth Birk Moller, and Martin Dufva. Multi-stringency wash of partially hybridized 60-mer probes reveals that the stringency along the probe decreases with distance from the microarray surface. *Nucl. Acids Res.*, 36(20):e132–, 2008. [cited at p. 30]

[155] Alex Pozhitkov, Peter A. Noble, Tomislav Domazet-Loso, Arne W. Nolte, Rainer Sonnenberg, Peer Staehler, Markus Beier, and Diethard Tautz. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucl. Acids Res.*, 34(9):e66–, 2006. [cited at p. 55]

[156] Alex E. Pozhitkov, Georg Nies, Barbara Kleinhenz, Diethard Tautz, and Peter A. Noble. Simultaneous quantification of multiple nucleic acid targets in complex rrna mixtures using high density microarrays and nonspecific hybridization as a source of information. *Journal of Microbiological Methods*, 75(1):92 – 102, 2008. [cited at p. 62]

[157] Alex E. Pozhitkov, Diethard Tautz, and Peter A. Noble. Oligonucleotide microarrays: widely applied poorly understood. *Brief Funct Genomic Proteomic*, page elm014, 2007. [cited at p. 55, 62, 66, 89]

[158] Xiangyun Qiu, Kurt Andresen, Lisa W. Kwok, Jessica S. Lamb, Hye Yoon Park, and Lois Pollack. Inter-dna attraction mediated by divalent counterions. *Physical Review Letters*, 99(3):038104, 2007. [cited at p. 25]

[159] Meera Ramakrishnan, Wen-Man Liu, Patricia A. DiCroce, Aleza Posner, Jian Zheng, Terumi Kohwi-Shigematsu, and Theodore G. Krontiris. Modulated Binding of SATB1, a Matrix Attachment Region Protein, to the AT-Rich Sequence Flanking the Major Breakpoint Region of BCL2. *Mol. Cell. Biol.*, 20(3):868–877, 2000. [cited at p. 74]

[160] Hightower C.R. Meagher R.B. The molecular evolution of actin. *Genetics*, 114:315–332, September 1986. [cited at p. 63, 111]

[161] Gudrun H Reed, Jana O Kent, and Carl T Wittwer. High-resolution dna melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics*, 8(6):597–608, 2007. [cited at p. 91]

[162] Jinsong Ren, Xiaogang Qu, Nanibhushan Dattagupta, and Jonathan B. Chaires. Molecular recognition of a rna•dna hybrid structure. *Journal of the American Chemical Society*, 123(27):6742–6743, 2001. doi: 10.1021/ja015649y. [cited at p. 112]

[163] Catriona Rennie, Harry Noyes, Stephen Kemp, Helen Hulme, Andy Brass, and David Hoyle. Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays. *BMC Genomics*, 9(1):317, 2008. [cited at p. 55, 68]

[164] Luis P. Reynaldo, Alexander V. Vologodskii, Bruce P. Neri, and Victor I. Lyamichev. The kinetics of oligonucleotide replacements. *Journal of Molecular Biology*, 297(2):511–520, 2000. [cited at p. 103]

[165] Peterson W.A. Richardson J.H. Georgiadis R.M. The effect of surface probe density on dna hybridization. *Nucleic Acids Res.*, 29(24):5163–5168, 2001. [cited at p. 13, 33, 34, 62, 66]

[166] Remo Rohs, Sean M West, Alona Sosinsky, Peng Liu, Richard S Mann, and Barry Honig. The role of dna shape in protein-dna recognition. *Nature*, 461(7268):1248–1253, 2009. [cited at p. 23]

[167] Ioulia Rouzina and Victor A. Bloomfield. Heat capacity effects on the melting of dna.1. general aspects. *Biophysical Journal*, 77(6):3242–3251, 1999. [cited at p. 47, 59]

[168] Ioulia Rouzina and Victor A. Bloomfield. Heat capacity effects on the melting of dna.2. analysis of nearest-neighbor base pair effects. *Biophysical Journal*, 77(6):3252–3255, 1999. [cited at p. 59]

[169] Rebecca A. Rule, Alex E. Pozhitkov, and Peter A. Noble. Use of hidden correlations in short oligonucleotide array data are insufficient for accurate quantification of nucleic acid targets in complex target mixtures. *Journal of Microbiological Methods*, 76(2):188 – 195, 2009. [cited at p. 62]

[170] Neidle S. *Principles of Nucleic Acids Structure*. Academic Press, 2007. [cited at p. 10, 17]

[171] Smith D.E. Perkins T.T. Chu S. Dynamical scaling of dna diffusion coefficients. *Macromolecules*, 29:1372–1373, 1996. [cited at p. 29, 71, 72]

[172] Jr Salsbury, Freddie R., Jill E. Clodfelter, Michael B. Gentry, Thomas Hollis, and Karin Drotschmann Scarpinato. The molecular mechanism of DNA damage recognition by MutS homologs and its consequences for cell death response. *Nucl. Acids Res.*, 34(8):2173–2185, 2006. [cited at p. 54]

[173] John SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighborthermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95(4):1460–1465, 1998. [cited at p. vii, 49, 51, 52, 91]

[174] John SantaLucia, Allawi H. T., and Senerviratne P. A. Improved Nearest-neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry*, 35(11):3555–3562, 1996. [cited at p. 45]

[175] Abraham Savitzky and M.J.E. ; Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1964. [cited at p. 91]

[176] T. Schueller, A. Nykytenko, A. Csaki, R. Moeller, W. Fritzsche, and J. Popp. Uv cross-linking of unmodified dna on glass surfaces. *Analytical and Bioanalytical Chemistry*, 395(4):1097–1105, 2009. [cited at p. 31]

[177] Nicholas N. Shaw and Dev P. Arya. Recognition of the unique structure of dna•rna hybrids. *Biochimie*, 90(7):1026–1039, 2008. [cited at p. 111, 112]

[178] Shirley J. Shi, Alicia Scheffer, Erik Bjeldanes, Mark A. Reynolds, and Lyle J. Arnold. DNA exhibits multi-stranded binding recognition on glass microarrays. *Nucl. Acids Res.*, 29(20):4251–4256, 2001. [cited at p. 103]

[179] Karsten H Siegmund, Ulrich E Steiner, and Clemens Richert. Chipcheck–a program predicting total hybridization equilibria for dna binding to small oligonucleotide microarrays. *J Chem Inf Comput Sci*, 43:2153–2162, 2003. [cited at p. 85, 86, 91, 103]

[180] A Sommerfeld. Zur elektronentheorie der metalle. *Naturwissenschaften*, 15(41):824–832, 1927. [cited at p. 84]

[181] J Sponer, KE Riley, and P Hobza. Nature and magnitude of aromatic stacking of nucleic acid bases. *PHYSICAL CHEMISTRY CHEMICAL PHYSICS*, 10(19):2595–2610, 2008. [cited at p. 16]

[182] Gisela Storz. An expanding universe of noncoding rnas. *Science*, 296(5571):1260–1263, 2002. 10.1126/science.1072249. [cited at p. 111]

[183] Linda Stromqvist Meuzelaar, Katie Hopkins, Ernesto Liebana, and Anthony J. Brookes. DNA Diagnostics by Surface-Bound Melt-Curve Reactions. *J Mol Diagn*, 9(1):30–41, 2007. [cited at p. 63, 89, 111]

[184] N Sugimoto, S Nakano, M Yoneyama, and K Honda. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucl. Acids Res.*, 24(22):4501–4505, 1996. [cited at p. 45, 50, 51, 59, 112, 117]

[185] Naoki Sugimoto, Mariko Nakano, and Shu-ichi Nakano. Thermodynamics-structure relationship of single mismatches in rna/dna duplexes. *Biochemistry*, 39(37):11270–11281, 2000. [cited at p. 45, 55, 59, 60, 111, 117]

[186] Naoki Sugimoto, Shu ichi Nakano, Misa Katoh, Akiko Matsumura, Hiroyuki Nakamuta, Tatsuo Ohmichi, Mari Yoneyama, and Muneo Sasaki. Thermodynamic parameters to predict stability of rna/dna hybrid duplexes. *Biochemistry*, 34(35):11211–11216, 1995. [cited at p. vii, 45, 53, 112, 117]

[187] Michel W. Mai T. Naiser T. and Ott A. Optical study of dna surface hybridization reveals dna surface density as a key parameter for microarray hybridization kinetics. *Biophysical Journal*, 92(3):999–1004, Februar 2007. [cited at p. 33, 65]

[188] Molecular Biology Tables. Nucleotide molecular weights. Technical report, Applied Biosystems, 2008. http://www.ambion.com/techlib/append/na_mw_tables.html. [cited at p. 30]

[189] Jennifer C. Takach, Peter J. Mikulecky, and Andrew L. Feig. Salt-dependent heat capacity changes for rna duplex formation. *Journal of the American Chemical Society*, 126(21):6530–6531, 2004. [cited at p. 59]

[190] Keiko Tawa and Wolfgang Knoll. Mismatching base-pair dependence of the kinetics of dna-dna hybridization studied by surface plasmon fluorescence spectroscopy. *Nucleic Acids Research*, 32(8):2372–2377, 2004. 10.1093/nar/gkh572. [cited at p. 91, 111, 113, 115]

[191] Tecan. Lasercheck. Ls reloaded, Tecan, 2005. [cited at p. 9]

[192] Montserrat Terrazas and Eric T. Kool. RNA major groove modifications improve siRNA stability and biological activity. *Nucl. Acids Res.*, 37(2):346–353, 2009. [cited at p. 112]

[193] M Thomas, R L White, and R W Davis. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences of the United States of America*, 73(7):2294–2298, 1976. [cited at p. 106]

[194] M. Thomas, R. L. White, and R. W. Davis. Hybridization of rna to double-stranded dna: formation of r-loops. *Proceedings of the National Academy of Sciences of the United States of America*, 73(7):2294–2298, 1976. [cited at p. 117, 118]

[195] Julie D. Thompson, Toby J. Gibson, and Des G. Higgins. Multiple sequence alignment using clustalw and clustalx. *Curr Protoc Bioinformatics*, Chapter 2:Unit 2.3, Aug 2002. [cited at p. 17, 25, 26, 64]

[196] Mark A. Thompson. Arguslab 4.0. Technical report, Planaria Software LLC, Seattle, WA, 1996. http://www.arguslab.com. [cited at p. 17, 25, 26]

[197] S.M. Tiquia, L. Wu, S.C. Chong, S. Passovets, D. Xu, and J. Xu, Y.and Zhou. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques*, 36(4), 2004. [cited at p. 63, 68]

[198] S. Tomić, S. Dolanski Babić, T. Vuletić, S. Krča, D. Ivanković, L. Griparić, and R. Podgornik. Dielectric relaxation of dna aqueous solutions. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 75(2):021905, 2007. [cited at p. 25]

[199] S. Tomić, T. Vuletić, S. Dolanski Babić, S. Krča, D. Ivanković, L. Griparić, and R. Podgornik. Screening and fundamental length scales in semidilute na-dna aqueous solutions. *Physical Review Letters*, 97(9):098303, 2006. [cited at p. 25, 28]

[200] S Tsuzuki and A Fujji. Nature and physical origin of ch/π interaction: significant difference from conventional hydrogen bonds. *Phys. Chem. Chem. Phys.*, 10:2584–2594, 2008. [cited at p. 23]

[201] Yoji Umezawa and Motohiro Nishio. Thymine-methyl/π interaction implicated in the sequence-dependent deformability of DNA. *Nucl. Acids Res.*, 30(10):2183–2192, 2002. [cited at p. 23, 112]

[202] Hidetoshi Urakawa, Said El Fantroussi, Hauke Smidt, James C. Smoot, Erik H. Tribou, John J. Kelly, Peter A. Noble, and David A. Stahl. Optimization of Single-Base-Pair Mismatch Discrimination in Oligonucleotide Microarrays. *Appl. Environ. Microbiol.*, 69(5):2848–2856, 2003. [cited at p. 59]

[203] Vadim A. Vasiliskov, Dmitry V. Prokopenko, and Andrei D. Mirzabekov. Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucl. Acids Res.*, 29(11):2303–2313, 2001. [cited at p. 17]

[204] Birte Vester and Jesper Wengel. Lna (locked nucleic acid): High-affinity targeting of complementary rna and dna. *Biochemistry*, 43(42):13233–13241, 2004. [cited at p. 13, 111]

[205] Alexander V. Vologodskii, B R Amirikyan, Y L Lyubchenko, and Maxim Frank-Kamenetskii. Allowance for heterogeneous stacking in the dna helix-coil transition theory. *Journal of Biomolecular Structure Dynamics*, 2(1):131–148, 1984. [cited at p. 51]

[206] Ulrich von Gemmingen. The fermi-dirac concept for isotherms of adsorbents with heterogeneous surfaces. *Chemical Engineering Science*, 60(19):5198 – 5205, 2005. [cited at p. 84, 91]

[207] Gong P. Harbers G. M. Grainger D. W. Multi-technique comparison of immobilized and hybridized olionucleotide surface density on commercial amine-reactive microarray slides. *Anal. Chem.*, 78(7):2342–2351, April 2006. Overview about the amount of DNA after the blocking. [cited at p. 30, 33]

[208] Pourmand N. Karhanek M. Persson H. H. J. Webb C. D. Lee T. H. Zahradnikova A. Davis R. W. Direct electrical detection of dna synthesis. *Proc. Natl. Acac. Sci.*, 103(17):6466–6470, April 2006. [cited at p. 33]

[209] Saenger W. *Principles of Nucleic Acid Structure*. Springer New York, 1984. [cited at p. 10, 12, 14, 15, 17, 25, 60, 75]

[210] Sebastian Waermlaender, Judit E. Sponer, J Sponer, and Mikael Leijon. The Influence of the Thymine C5 Methyl Group on Spontaneous Base Pair Breathing in DNA. *Journal of Biological Chemistry*, 277(32):28491–28497, 2002. [cited at p. 16, 19, 21, 23, 112, 113]

[211] Hong-Ying Wang, Renae Malek, Anne Kwitek, Andrew Greene, Truong Luu, Babak Behbahani, Bryan Frank, John Quackenbush, and Norman Lee. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biology*, 4(1):R5, 2003. [cited at p. 31, 63]

[212] Junbai Wang and Morigen. Bayespi - a new model to study protein-dna interactions: a case study of condition-specific protein binding parameters for yeast transcription factors. *BMC Bioinformatics*, 10(1):345, 2009. [cited at p. 84]

[213] Shaohui Wang and Eric T. Kool. Origins of the large differences in stability of dna and rna helixes: C-5 methyl and 2'-hydroxyl effects. *Biochemistry*, 34(12):4125–4132, 1995. [cited at p. 23, 111, 112, 117]

[214] Yulei Wang, Catalin Barbacioru, Fiona Hyland, Wenming Xiao, Kathryn Hunkapiller, Julie Blake, Frances Chan, Carolyn Gonzalez, Lu Zhang, and Raymond Samaha. Large scale real-time pcr validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics*, 7(1):59, 2006. [cited at p. 62, 105]

[215] HJ Watts, D Yeung, and H Parkes. Real-time detection and quantification of dna hybridization by an optical biosensor. *Analytical Chemistry*, 67(23):4283–4289, 1995. [cited at p. 30]

[216] Stefan Weckx, Enrico Carlon, Luc De Vuyst, and Paul Van Hummelen. Thermodynamic behavior of short oligonucleotides in microarray hybridizations can be described using gibbs free energy in a nearest-neighbor model. *The Journal of Physical Chemistry B*, 111(48):13583–13590, 2007. [cited at p. 42]

[217] Hairong Wei, Pei F. Kuan, Shulan Tian, Chuhu Yang, Jeff Nie, Srikumar Sengupta, Victor Ruotti, Gudrun A. Jonsdottir, Sunduz Keles, James A. Thomson, and Ron Stewart. A study of the relationships between oligonucleotide properties and hybridization signal intensities from nimblegen microarray datasets. *Nucl. Acids Res.*, pages gkn133+, April 2008. [cited at p. 63]

[218] Heinz Werntges, Gerhard Steger, Detlev Riesner, and Hans-Joachim Fritz. Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies. *Nucl. Acids Res.*, 14(9):3773–3790, 1986. [cited at p. 54, 55, 60]

[219] James G. Wetmur and Jaques Fresco. Dna probes: Application of the principles of nucleic acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology*, 26(3), 1991. [cited at p. v, 44, 45, 49, 92]

[220] Carl T. Wittwer, Gudrun H. Reed, Cameron N. Gundry, Joshua G. Vandersteen, and Robert J. Pryor. High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. *Clin Chem*, 49(6):853–860, 2003. [cited at p. 49]

[221] Tomasz K. Wojdacz and Alexander Dobrovic. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucl. Acids Res.*, 35(6):e41–, 2007. [cited at p. 63, 67, 89]

[222] Peng Wu, Shi-ichi Nakano, and Naoki Sugimoto. Temperature dependence of thermodynamic properties for dna/dna and rna/dna duplex formation. *European Journal of Biochemistry*, 269(12):2821–2830, 2002. [cited at p. 47, 55, 111]

[223] Fei Xu, August M. Pellino, and Wolfgang Knoll. Electrostatic repulsion and steric hindrance effects of surface probe density on deoxyribonucleic acid (dna)/peptide nucleic acid (pna) hybridization. *Thin Solid Films*, 516(23):8634 – 8639, 2008. [cited at p. 63, 65, 84, 89, 111]

[224] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucl. Acids Res.*, 34(2):564–574, 2006. [cited at p. 16, 17, 18]

[225] Danfeng Yao, Junyoung Kim, Fang Yu, Peter E. Nielsen, Eva-Kathrin Sinner, and Wolfgang Knoll. Surface density dependence of pcr amplicon hybridization on pna/dna probe layers. *Biophysical Journal*, 88(4):2745 – 2751, 2005. [cited at p. 63, 65, 89]

[226] Yong You, Bernardo G. Moreira, Mark A. Behlke, and Richard Owczarzy. Design of lna probes that improve mismatch discrimination. *Nucleic Acids Research*, 34(8):e60, 2006. 10.1093/nar/gkl175. [cited at p. 111]

[227] Ignatova Z., Martinez-Perez I., and Zimmermann K.H. *DNA Computing Models*. Springer Science+Business Media, 2008. [cited at p. 76]

[228] Y. Zhang, D. A. Hammer, and Graves D.J. Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophysical Journal*, 89:2950–2959, 2005. [cited at p. 110, 114]

[229] Yong-li Zhang, Wei-Mou Zheng, Ji-Xing Liu, and Y. Z. Chen. Theory of dna melting based on the peyrard-bishop model. *Phys. Rev. E*, 56(6):7100–7115, Dec 1997. [cited at p. 20]

[230] Bruno H. Zimm and Oscar Lumpkin. Reptation of a polymer chain in an irregular matrix: diffusion and electrophoresis. *Macromolecules*, 26(1):226–234, 1993. [cited at p. 26, 29]

[231] Michael Zuker. http://dinamelt.bioinfo.rpi.edu/twostate.php. [cited at p. 45, 54, 55, 56, 57, 58, 59]

Appendices

Appendix A

Probe and target sequences

A.1 Phylogenetic tree of actin genes

For the phylogenetic tree according to UPGMA (Unweighted Pair Group Method with Arithmetic Mean) is shown in Figure A.1.

A.2 Sequences of the Actin X-chip

A.2.1 Sequences of the Actin genes

The strand complementary to the primer design is given.

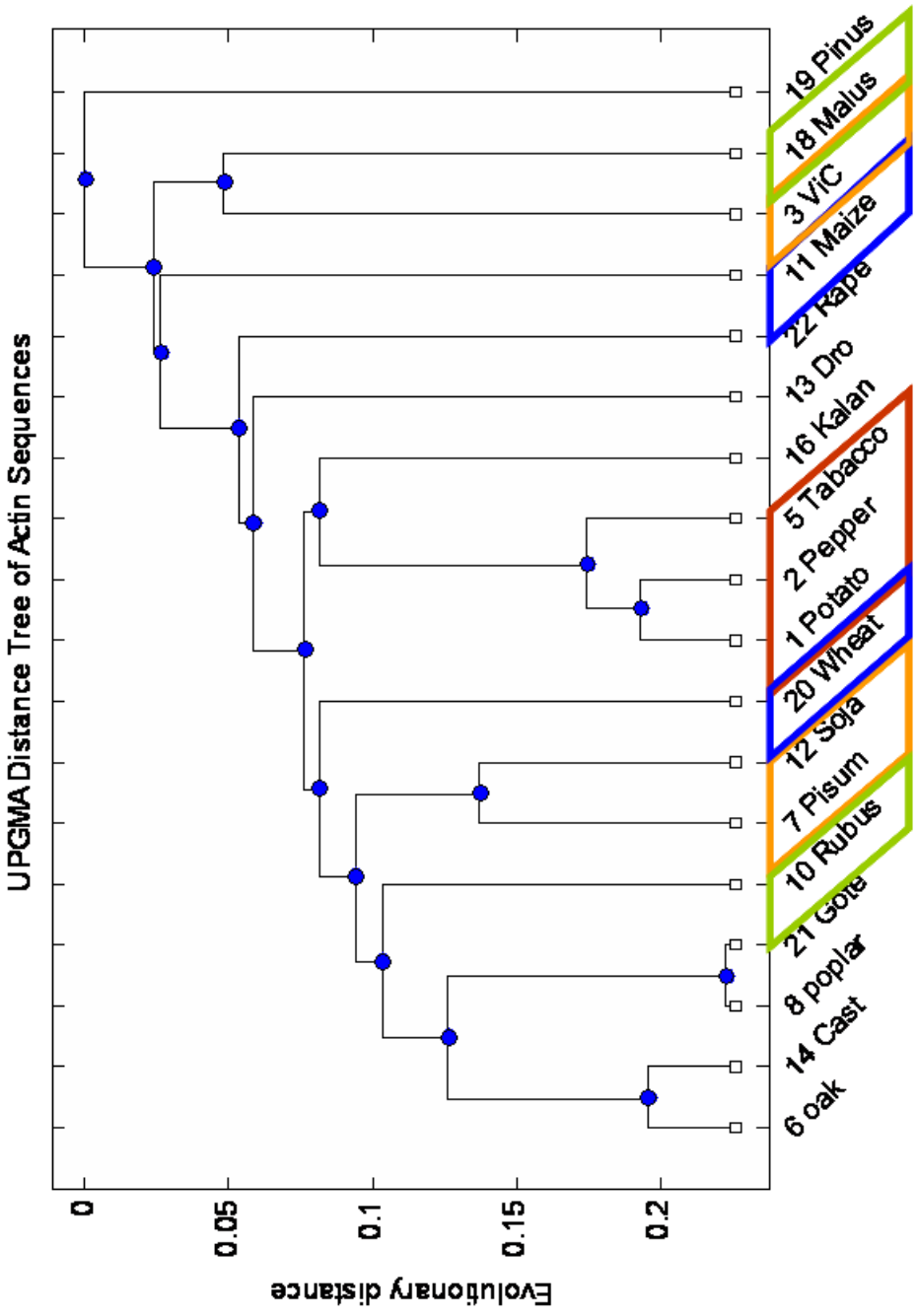
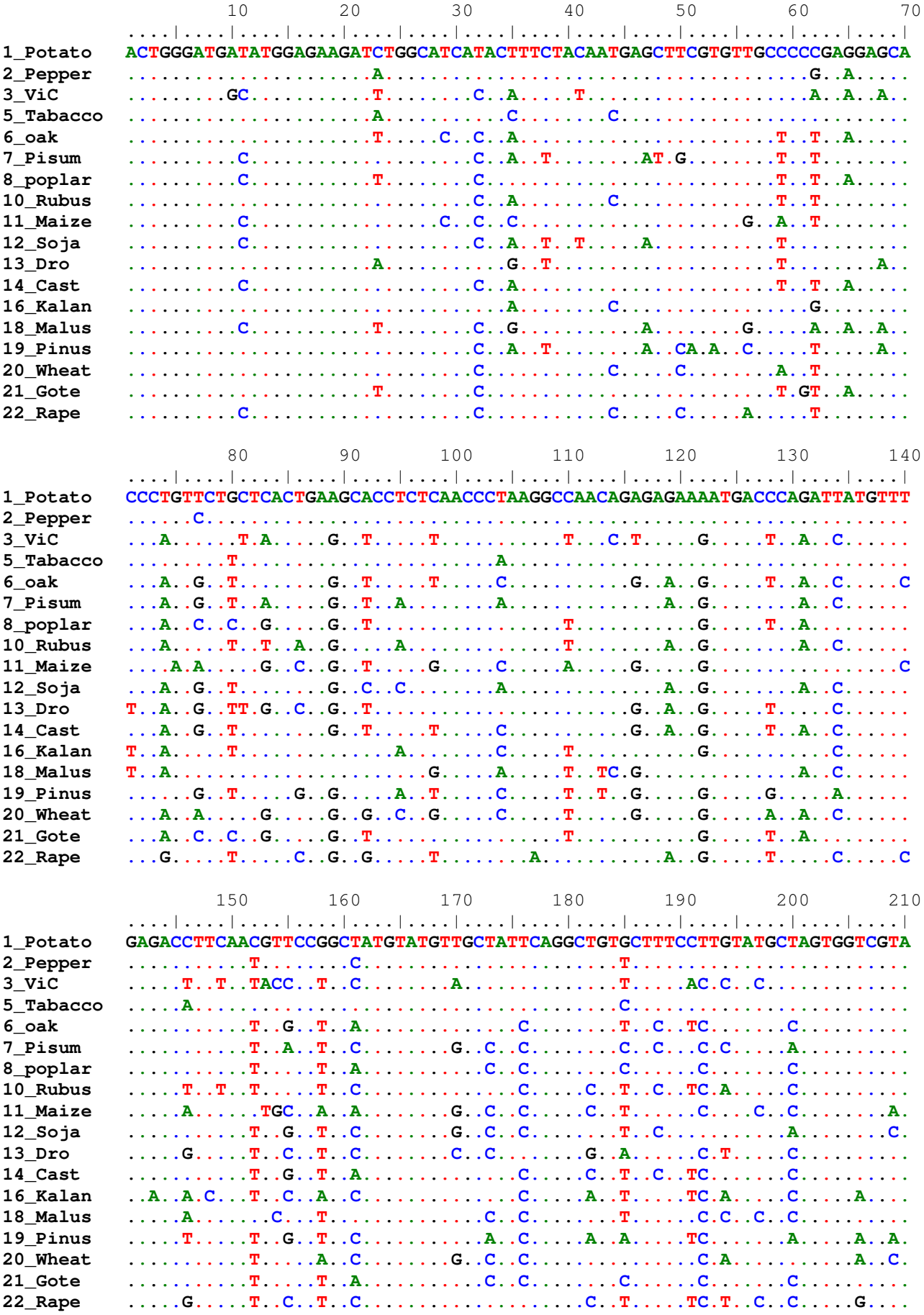
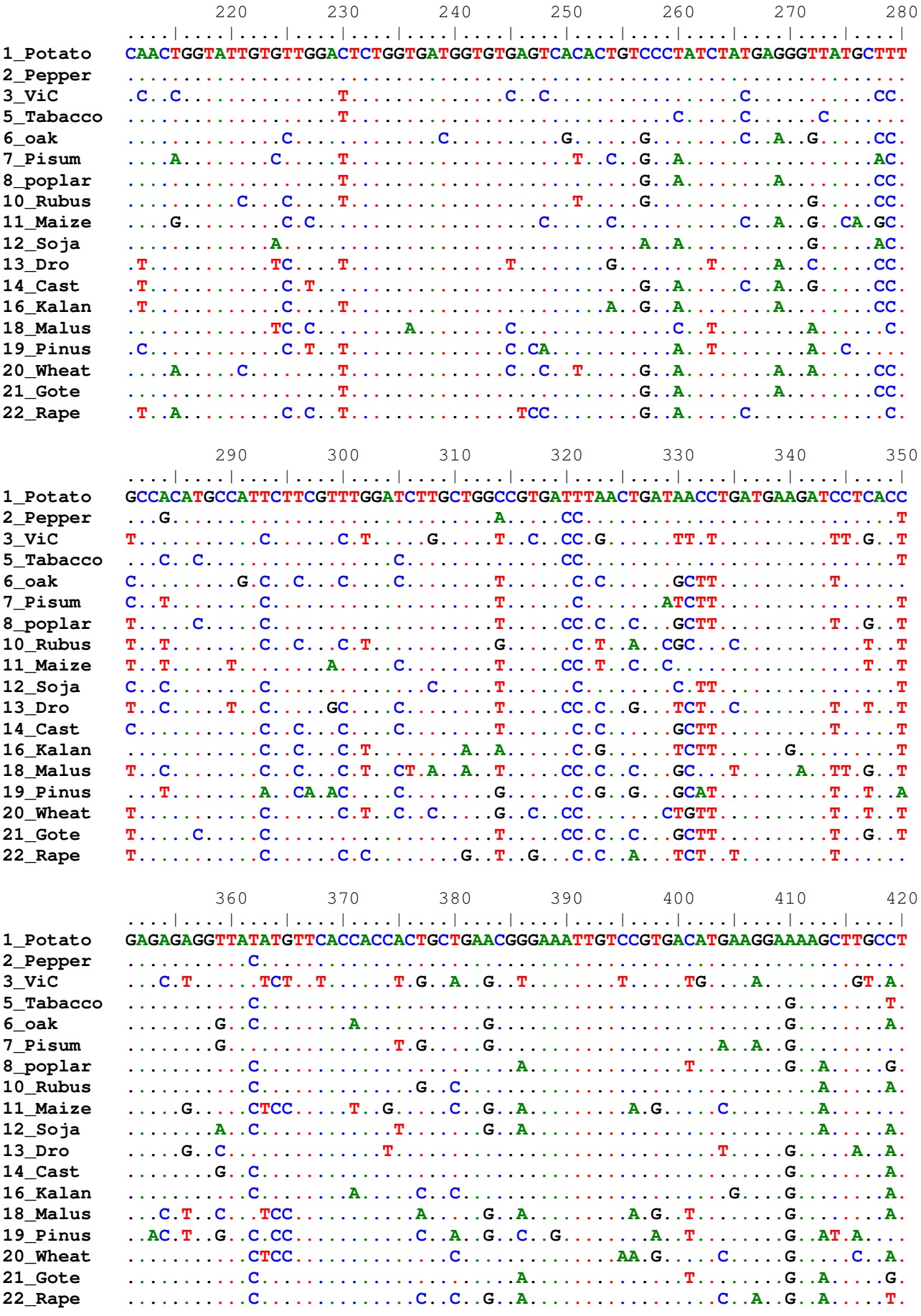
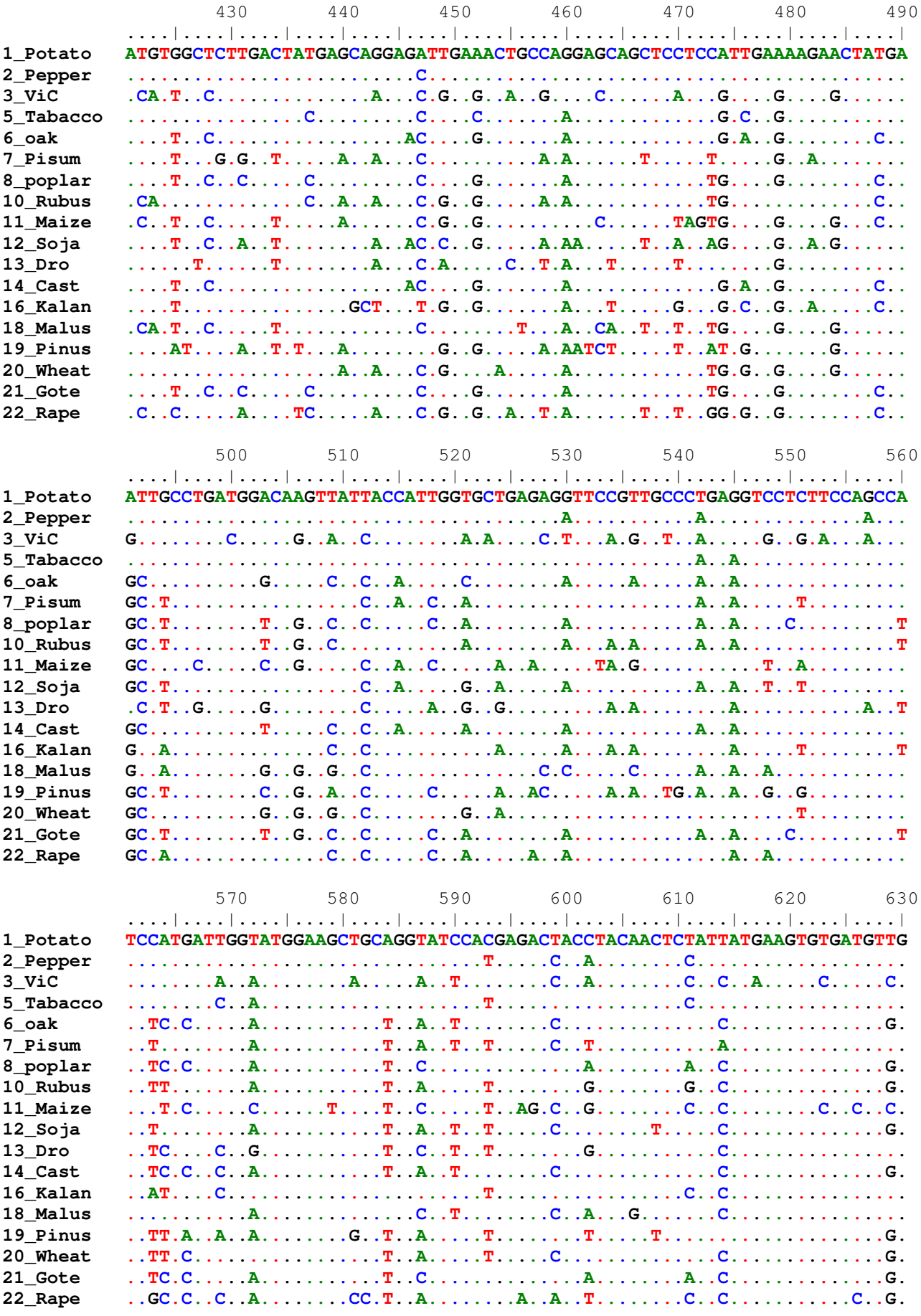
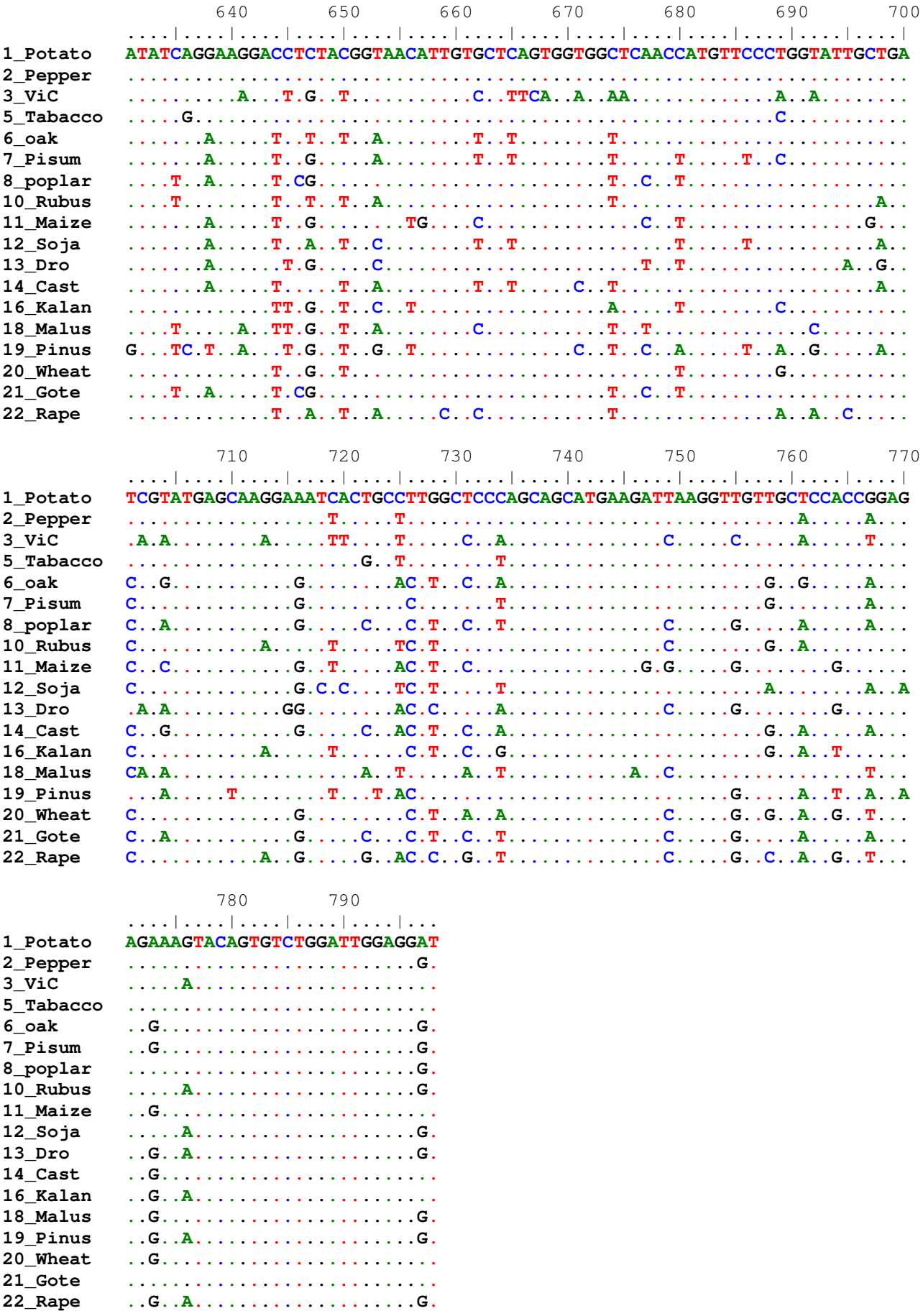


Figure A.1: Phylogenetc tree of the actin genes according to UPGMA (Unweighted Pair Group Method with Arithmetic Mean). The colored boxes denote taxonomic similarity. While some closely related plants are clustered well (*solanaceas*, red box), the overall correlation between genotype and phenotype is not so strong.


10 20 30 40 50 60 70
1_Potato ACTGGGATGATATGGAGAAGATCTGGCATCATACTTTCTACAATGAGCTTCGTGTTGCCCCCGAGGAGCA
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
80 90 100 110 120 130 140
1_Potato CCCTGTTCTGCTCACTGAAGCACCTCTCAACCCTAAGGCCAACAGAGAGAAAATGACCCAGATTATGTTT
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
150 160 170 180 190 200 210
1_Potato GAGACCTTCAACGTTCCGGCTATGTATGTTGCTATTCAGGCTGTGCTTTCCTTGTATGCTAGTGGTCGTA
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape


220 230 240 250 260 270 280
1_Potato CAACTGGTATTGTGTTGGACTCTGGTGATGGTGTGAGTCACACTGTCCCTATCTATGAGGGTTATGCTTT
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
290 300 310 320 330 340 350
1_Potato GCCACATGCCATTCTTCGTTTGGATCTTGCTGGCCGTGATTTAACTGATAACCTGATGAAGATCCTCACC
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
360 370 380 390 400 410 420
1_Potato GAGAGAGGTTATATGTTCACCACCACTGCTGAACGGGAAATTGTCCGTGACATGAAGGAAAAGCTTGCCT
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape


430 440 450 460 470 480 490
1_Potato ATGTGGCTCTTGACTATGAGCAGGAGATTGAAACTGCCAGGAGCAGCTCCTCCATTGAAAAGAACTATGA
2_Pepper .....C.....
3_ViC .CA.T..C.....A...C.G..G..A..G....C.....A...G....G....G.....
5_Tabacco .....C.....C....C.....A.....G.C..G.....
6_oak ....T..C.....AC....G.....A.....G.A..G.....C..
7_Pisum ....T...G.G..T....A..A...C.....A.A.....T.....T.....G..A.....
8_poplar ....T..C..C.....C.....C....G.....A.....TG....G.....C..
10_Rubus .CA.....C..A..A...C.G..G....A.A.....TG.....C..
11_Maize .C..T..C.....T.....A.....C.G..G.....C.....TAGTG....G....G...C..
12_Soja ....T..C..A..T.....A..AC.C..G.....A.AA.....T..A..AG....G..A.G.....
13_Dro ......T.....T.....A...C.A.....C..T.A...T.....T.....G.....
14_Cast ....T..C.....AC....G.....A.....G.A..G.....C..
16_Kalan ....T.....GCT...T.G..G.....A...T.....G...G.C..G..A.....C..
18_Malus .CA.T..C.....T.....C.....T...A..CA..T..T..TG....G....G.....
19_Pinus ....AT....A..T.T..A.....G..G....A.AATCT.....T..AT.G.....G.....
20_Wheat .....A..A...C.G...A.....A.....TG.G..G....G.....
21_Gote ....T..C..C.....C.....C....G.....A.....TG....G.....C..
22_Rape .C..C.....A....TC.....A...C.G..G..A..T.A.....T..T..GG.G..G.....C..
500 510 520 530 540 550 560
1_Potato ATTGCCTGATGGACAAGTTATTACCATTGGTGCTGAGAGGTTCCGTTGCCCTGAGGTCCTCTTCCAGCCA
2_Pepper .....A.....A.....A...
3_ViC G.....C.....G..A..C.....A.A....C.T...A.G..T..A.....G..G.A...A...
5_Tabacco .....A..A.....
6_oak GC.....G.....C..C..A.....C.....A.....A.....A..A.....
7_Pisum GC.T.....C..A..C..A.....A..A.....T.....
8_poplar GC.T.....T..G..C..C.....C..A.....A.....A..A....C.....T
10_Rubus GC.T.....T..G..C.....A.....A...A.A.....A..A.....T
11_Maize GC....C.....C..G.....C..A..C.....A..A.....TA.G.....T..A.....
12_Soja GC.T.....C..A.....G..A.....A.....A..A..T..T.....
13_Dro .C.T..G.....G.....C.....A..G..G.....A.A.....A.....A..T
14_Cast GC.....T.....C..C..A.....A.....A.....A..A.....
16_Kalan G..A.....C..C.....A.....A...A.A.....A.....T.....T
18_Malus G..A.....G..G..G..C.....C.C.....C.....A..A..A.....
19_Pinus GC.T.....C..G..A..C.....C.....A..AC.....A.A..TG.A..A..G..G.....
20_Wheat GC.....G..G..G..C.....G..A.....T.....
21_Gote GC.T.....T..G..C..C.....C..A.....A.....A..A....C.....T
22_Rape GC.A.....C..C.....C..A.....A..A.....A..A.....
570 580 590 600 610 620 630
1_Potato TCCATGATTGGTATGGAAGCTGCAGGTATCCACGAGACTACCTACAACTCTATTATGAAGTGTGATGTTG
2_Pepper .....T.....C..A.....C.....
3_ViC .....A..A.....A.....A..T.....C..A.....C..C..A.....C.....C.
5_Tabacco .....C..A.....T.....C.....
6_oak ..TC.C.....A.....T..A..T.....C.....C.....G.
7_Pisum ..T.....A.....T..A..T..T.....C..T.....A.....
8_poplar ..TC.C.....A.....T..C.....A.....A..C.....G.
10_Rubus ..TT.....A.....T..A.....T.....G.....G..C.....G.
11_Maize ...T.C.....C.....T....T..C.....T..AG.C..G.....C..C.....C..C..C.
12_Soja ..T.....A.....T..A..T..T.....C.....T.....C.....G.
13_Dro ..TC....C..G.....T..C..T..T.....G.....C.....
14_Cast ..TC.C..C..A.....T..A..T.....C.....C.....G.
16_Kalan ..AT....C.....T.....C..C.....
18_Malus .....A.....C..T.....C..A...G.....C.....
19_Pinus ..TT.A..A..A.....G..T..A.....T.....T.....T.....G.
20_Wheat ..TT.C.....T..A.....T.....C.....C.....G.
21_Gote ..TC.C.....A.....T..C.....A.....A..C.....G.
22_Rape ..GC.C..C..A.....CC.T..A.....A..A..T.....C..C.....C..G.


640 650 660 670 680 690 700
1_Potato ATATCAGGAAGGACCTCTACGGTAACATTGTGCTCAGTGGTGGCTCAACCATGTTCCCTGGTATTGCTGA
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
710 720 730 740 750 760 770
1_Potato TCGTATGAGCAAGGAAATCACTGCCTTGGCTCCCAGCAGCATGAAGATTAAGGTTGTTGCTCCACCGGAG
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape
780 790
1_Potato AGAAAGTACAGTGTCTGGATTGGAGGAT
2_Pepper
3_ViC
5_Tabacco
6_oak
7_Pisum
8_poplar
10_Rubus
11_Maize
12_Soja
13_Dro
14_Cast
16_Kalan
18_Malus
19_Pinus
20_Wheat
21_Gote
22_Rape

A.2.2 Sequences of the PCR primers

> M13-Forward
GTA AAA CGA CGG CCA G

> M13-Reverse
CAG GAA ACA GCT ATG AC

> Actin-Forward
ACT GGG ATG AYA TGG AGA AG

> Actin-Reverse
AYC CTC CAA TCC AGA CAC TG

In this representation the sequence matches to the Actin-Forward-Primer and to the reverse complement of the Actin-Reverse-Primers. "Y" is a degenerate primer for C and T.

A.2.3 Sequences of the oligo probes

The sequences of the oligo probes can be found in Table A.1, Table A.2 and in Table A.3.

Table A.1: Sequences of the short oligo probes. The probes named 'BRD' are universal.

Probe	Sequence
A01-066	AGCACCCTGTTCTGCTCACTGA
A01-306	TTGCTGGCCGTGATTTAACTGATAACC
A01-405	AGGAGAAGCTTGCTTATGTGGCTCTTG
A01-431	GACTATGAGCAGGAGATTGAAACTGCC
A02-143	ACCTTCAATGTTCCGGCCATGT
A02-710	GCAAGGAAATTACTGCTTTGGCTCC
A03-099	ACCCTAAGGCTAACCGTGAGAAGATG
A03-405	AGGAGAAGCTTGCTTATGTGGCTCTTG
A05-403	GAAAGAAAAGCTGTCATACATTG
A05-620	TGTGATGTTGATATCGGGAAGGACCT
A05-710	AAGGAAATCACGGCTTTGGCTCCTA
A06-234	GTGACGGTGTGAGTCGCACT
A06-364	GTTCACAACCACTGCTGAGCGG
A06-750	AGGTTGTGGCGCCACCAGAG
A07-36	TTTACAATGAATTGCGTGTTGCTCCTGA
A07-368	ACCACCTCGGCTGAGCGG
A07-412	GCTTGCCTATGTTGCTGTGGATTATGA
A10-069	ACCCAGTTCTTCTTACAGAGGCAC
A10-316	TGATCTTACAGACGCCCTCATGAAGA
A11-266	GAAGGGTACACGCTTCCTCATGCTA
A11-523	AGAAAGGTTTAGGTGCCCTGAGGT
A11-636	GAAAGGATCTGTACGGTAATGTTGTCCT
A12-439	GCAAGAACTCGAGACTGCAAAAAGCA
A12-712	GGAGACCCCTGCTCTTGCTC
A13-066	AACATCCAGTGCTTTTGACCGAGG
A13-479	AAGAACTATGAACTTCCGGATGGGCAA
A14-504	AAGTCATCACAATTGGAGCTGAGAGA
A14-686	CCTGGTATTGCAGACCGGATGA
A16-432	ACTATGAGGCTGAGTTGGAGACTG
A16-458	AAGAGTAGCTCGTCCGTCGAGAAA
A16-551	TTCCAGCCTTCATTGATCGGTATGGA
A16-551-B	TTCCAGCCTTCATTGATCGGTATGGAA
A18-087	AAGCACCTCTGAACCCAAAGGC
A18-290	ATCCTCCGTCTTGACTTAGCAGGTC
A18-451	AACTTCCAAGACAAGTTCTTCTGTTGAG
A19-522	CAGAACGGTTCAGATGTGCAGAAGTG
A19-711	AGGAAATTACTTCACTGGCTCCCAGC
A20-066	AGCACCCAGTACTGCTGACTGAGG
A20-438	AACAAGAGCTGGAAAATGCCAAGAGC
A21+8-402	TGAAGGAGAAACTTGCGTATGTTGC
A21+8-627	TGGATATTAGAAAGGATCCGTACGGT
A21+8-69	ACCCAGTCCTCCTGACTGAGG
A21-045	AGCTTCGTGTTGCTCGTGAAGA
A22-303	ATCTTGCGGGTCGGGATCTCAC
A22-458	AAGAGCAGTTCTTCGGTGGAGAAG
A22-716	ATCACGGCACTCGCGCCT
BRD-197	GCTAGTGGTCGTACAACTGGTATTGT
BRD-774	AGTACAGTGTCTGGATTGGAGG

Table A.2: Sequences of the long oligo probes

Probe	Sequence
01-032	ACTTTCTACAATGAGCTTCGTGTTGCCCCCGAGGAGCACCCTGTTCTGCTCACTGAAGCA
01-277	TTTGCCACATGCCATTCTTCGTTTGGATCTTGCTGGCCGTGATTTAACTGATAACCTGAT
01-583	AGGTATCCACGAGACTACCTACAACTCTATTATGAAGTGTGATGTTGATATCAGGAAGGA
02-120	AAATGACCCAGATTATGTTTGAGACCTTCAATGTTCCGGCCATGTATGTTGCTATTCAGG
02-714	AAATTACTGCTTTGGCTCCCAGCAGCATGAAGATTAAGGTTGTTGCACCACCAGAGAGAA
03-408	AAAAGCTGTCATACATTGCCCTTGACTATGAGCAAGAGCTGGAGACAGCGAGGACCAGCT
03-638	AAAGACTTGTATGGTAACATTGTCCTTTCAGGAGGAACAACCATGTTCCCAGGAATTGCT
03-710	AAAGAAATTTCTGCTTTGGCCCCAAGCAGCATGAAGATCAAGGTCGTTGCACCACCTGAG
05-103	AAAGGCCAACAGAGAGAAAATGACCCAGATTATGTTTGAGACATTCAACGTTCCGGCTAT
05-387	AAATTGTCCGTGACATGAAGGAGAAGCTTGCTTATGTGGCTCTTGACTACGAGCAGGAGC
05-399	ACATGAAGGAGAAGCTTGCTTATGTGGCTCTTGACTACGAGCAGGAGCTTGACACTGCCA
05-543	AAGTCCTCTTCCAGCCATCCATGATCGGAATGGAAGCTGCAGGTATCCATGAGACTACCT
06-211	AACTGGTATTGTGCTGGACTCTGGTGACGGTGTGAGTCGCACTGTGCCTATCTACGAAGG
06-338	AAGATTCTCACCGAGAGAGGGTACATGTTCACAACCACTGCTGAGCGGGAAATTGTCCGT
06-676	AACCATGTTCCCTGGTATTGCTGACCGGATGAGCAAGGAGATCACTGCACTTGCCCCAAG
07-211	AACAGGTATTGTCTTGGATTCTGGTGATGGTGTGAGTCATACCGTGCCAATCTATGAGGG
07-322	AACTGAATCTTTGATGAAGATCCTCACTGAGAGAGGGTATATGTTCACCACCTCGGCTGA
07-404	AAAGAGAAGCTTGCCTATGTTGCTGTGGATTATGAACAAGAGCTTGAAACTGCAAAGAGC
08-001	ACTGGGATGACATGGAGAAGATTTGGCATCACACTTTCTACAATGAGCTTCGTGTTGCTC
08-006	ATGACATGGAGAAGATTTGGCATCACACTTTCTACAATGAGCTTCGTGTTGCTCCTGAAG
08-030	ACACTTTCTACAATGAGCTTCGTGTTGCTCCTGAAGAGCACCCAGTCCTCCTGACTGAGG
08+21-084	CTGAGGCTCCTCTCAACCCTAAGGCTAACAGAGAGAAGATGACTCAAATTATGTTTGAGA
08+21-540	CAGAAGTCCCCTTCCAGCCTTCTCTCATTGGAATGGAAGCTGCTGGCATCCACGAGACTA
10-408	AAAAACTTGCATACATGGCTCTTGACTACGAACAAGAGCTGGAGACTGCAAAGAGCAGCT
10-477	AAAAGAACTACGAGCTTCCTGATGGTCAGGTCATTACCATTGGAGCTGAGAGATTCAGAT
10-714	AAATTACTGCTCTTGCTCCCAGCAGCATGAAGATCAAGGTTGTGGCACCACCGGAGAGAA
11-109	AAACAGGGAGAAGATGACCCAGATTATGTTCGAGACATTCAACTGCCCAGCAATGTATGT
11-408	AAAAACTTGCCTACGTTGCCCTTGATTATGAACAGGAGCTGGAGACTGCCAGGACCAGCT
11-525	AAAGGTTTAGGTGCCCTGAGGTTCTATTCCAGCCATCCTTCATTGGCATGGAATCTGCTG

Continued on next page

Table A.2: Sequences of the long oligo probes (continued from previous page)

Probe	Sequence
12-017	AAGATCTGGCATCACACATTTTATAATGAACTTCGTGTTGCTCCCGAGGAGCACCCAGTG
12-412	ACTTGCATATGTTGCCCTAGATTATGAGCAAGAACTCGAGACTGCAAAAAGCAGTTCATC
12-444	AACTCGAGACTGCAAAAAGCAGTTCATCAGTTGAGAAAAGCTATGAGCTTCCTGATGGAC
13-111	ACAGGGAAAAGATGACTCAGATCATGTTTGAGACGTTCAATGTCCCTGCCATGTATGTCG
13-450	AAACCGCTAAGAGTAGCTCTTCCATTGAGAAGAACTATGAACTTCCGGATGGGCAAGTTA
13-637	AAAGGACTTGTACGGCAACATTGTGCTCAGTGGTGGCTCTACTATGTTCCCTGGTATAGC
14-175	CCAGGCCGTTCTCTCTCTGTATGCCAGTGGTCGTACTACTGGTATTGTGCTTGACTCTGG
14-482	AACTACGAGCTGCCTGATGGTCAAGTCATCACAATTGGAGCTGAGAGATTCCGTTGCCCA
14-652	AAACATTGTTCTTAGTGGCGGTTCAACCATGTTCCCTGGTATTGCAGACCGGATGAGCAA
16-041	AACGAGCTTCGTGTTGCCCCGGAGGAGCATCCAGTTCTTCTCACTGAAGCACCACTCAAC
16-141	AAACATCCAATGTCCCAGCCATGTATGTTGCTATCCAGGCAGTTCTTTCTCTATATGCCA
16-479	AAAAACTACGAGTTACCTGATGGACAAGTCATCACCATTGGTGCAGAGAGATTCAGATGC
18-338	AAAATTTTGACTGAGCGTGGCTATTCCTTCACCACCACAGCTGAGCGAGAAATTGTCAGG
18-450	AAACTTCCAAGACAAGTTCTTCTGTTGAGAAGAGCTATGAGTTACCTGATGGGCAGGTGA
18-652	AAACATTGTCCTCAGTGGTGGTTCTACCATGTTCCCTGGCATTGCTGACAGAATGAGCAA
19-477	AAAAGAGCTATGAGCTTCCTGATGGCCAGGTAATCACCATCGGTGCAGAACGGTTCAGAT
19-638	AAAGACTTGTATGGGAATATTGTGCTCAGTGGCGGTTCCACAATGTTTCCAGGGATTGCA
19-714	AAATTACTTCACTGGCTCCCAGCAGCATGAAGATTAAGGTGGTTGCACCTCCAGAAAGGA
20-126	CACAAATCATGTTTGAGACCTTCAATGTTCCAGCCATGTATGTGGCCATCCAGGCTGTGC
20-381	AACGGGAAATTGTAAGGGACATCAAGGAGAAGCTCGCATATGTGGCTCTTGACTATGAAC
20-450	AAAATGCCAAGAGCAGCTCCTCTGTGGAGAAGAGCTATGAGCTGCCTGATGGGCAGGTGA
21-001	ACTGGGATGATATGGAGAAGATTTGGCATCACACTTTCTACAATGAGCTTCGTGTTGCTC
21-006	ATGATATGGAGAAGATTTGGCATCACACTTTCTACAATGAGCTTCGTGTTGCTCGTGAAG
21-030	ACACTTTCTACAATGAGCTTCGTGTTGCTCGTGAAGAGCACCCAGTCCTCCTGACTGAGG
22-013	GGAGAAGATCTGGCATCACACTTTCTACAACGAGCTCCGTGTAGCCCCTGAGGAGCACCC
22-410	AAACTTGCTTACGTCGCTCTAGACTTCGAGCAAGAGCTGGAGACAGCTAAGAGCAGTTCT
22-652	AAACATCGTCCTCAGTGGTGGTTCAACCATGTTCCCAGGAATCGCTGACCGTATGAGCAA

A.3 Similarity matrix of the Actin genes

The similarity matrices based on nucleic acid sequence is shown in Table A.4, for amino acids in Table A.5.

Table A.3: Sequences of the 70 mer LNA oligo probes. The lower case letters denote LNA positions.

Probe	Sequence
LNA2Pep1x70	ACTGGGATGATATGGAGAAGATaTGGCATCATACTTTCTACAATGAGCTTCGTGTTGCCCCgGAaGAGCA
LNA2Pep250x319	ACACTGTCCCTATCTATGAGGGTTATGCTTTGCCgCATGCCATTCTTCGTTTGGATCTTGCTGGaCGTGA
LNA2Pep510x579	ATTACCATTGGTGCTGAGAGaTTCCGTTGCCCaGAGGTCCTCTTCCAaCCATCCATGATTGGTATGGAAG
LNA2Pep730x798	CTCCCAGCAGCATGAAGATTAAGGTTGTTGCaCCACCaGAGAGAAAGTACAGTGTCTGGATTGGAGGgT
LNA2Pep477x546	GAaAAGAACTATGAATTGCCTGATGGACAAGTTATTACCATTGGTGCTGAGAGaTTCCGTTGCCCAGAgG
LNA2Pep543x522	ACTGCCAgGAGCAGCTCCTCCaTtGAaAAGAACTATGAATTGCCTGATGGACAAGTTATTACCATTGGTG

Table A.4: Actin mismatch table based on nucleic acid sequences.

	Pot	**Pep**	**Bea**	**Tob**	**Oak**	**Pea**	**Pop**	**Bla**	**Mai**	**Soy**	**Sun**	**Cas**	**Kal**	**Mal**	**Pin**	**Whe**	**Ail**
Pepper	26																
Bean	150	139															
Tobacco	38	42	155														
Oak	111	109	150	114													
Pea	109	109	161	112	93												
Poplar	107	99	147	104	79	97											
Blackberry	111	105	146	113	95	106	91										
Maize	141	140	175	146	139	151	131	146									
Soye	115	111	160	117	90	67	105	110	151								
Sundew	124	124	164	125	119	121	116	124	150	131							
Castanea	110	108	147	108	24	89	72	91	140	86	118						
Kalanchoe	109	108	157	105	107	122	116	93	156	127	127	101					
Malus	132	130	130	130	129	141	121	138	152	143	142	130	149				
Pine	164	162	180	167	156	154	157	144	171	155	159	150	154	164			
Wheat	118	120	149	124	109	110	101	105	124	119	128	105	110	138	155		
Ailanthus	106	100	148	103	80	100	3	92	132	108	117	73	115	124	158	100	
Rape	141	141	155	134	114	129	114	112	152	137	139	106	123	153	163	129	117

Table A.5: Actin mismatch table based on amino acid sequences.

	Pot	**Pep**	**Bea**	**Tob**	**Oak**	**Pea**	**Pop**	**Bla**	**Mai**	**Soy**	**Sun**	**Cas**	**Kal**	**Mal**	**Pin**	**Whe**	**Ail**
Pepper	1																
Bean	17	16															
Tobacco	6	5	19														
Oak	7	6	19	7													
Pea	7	6	18	9	9												
Poplar	7	6	19	7	4	9											
Blackberry	6	5	17	6	3	8	3										
Maize	15	14	20	17	17	18	17	16									
Soye	8	7	15	8	8	8	8	7	17								
Sundew	7	6	20	9	7	6	7	6	17	10							
Castanea	5	4	17	5	2	7	2	1	15	6	5						
Kalanchoe	9	8	21	9	7	10	7	6	19	10	8	5					
Malus	11	10	14	11	10	13	10	8	15	10	13	8	13				
Pine	13	13	24	15	12	16	12	11	22	14	13	10	15	15			
Wheat	9	8	16	9	8	9	8	7	12	8	8	6	10	9	13		
Ailanthus	8	7	20	8	5	10	1	4	18	9	8	3	8	11	13	9	
Rape	8	7	19	8	6	7	6	5	16	9	5	4	7	12	12	7	7