

# Characterization of Oligonucleotide Microarray Hybridization

Microarray Fabrication by Light-Directed in situ Synthesis –  
Development of an Automated DNA Microarray Synthesizer,  
Characterization of Single Base Mismatch Discrimination and the  
Position-Dependent Influence of Point Defects on Oligonucleotide  
Duplex Binding Affinities

Von der Universität Bayreuth  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)  
genehmigte Abhandlung

von

Thomas Naiser

geboren in Bayreuth

1. Gutachter: Prof. Dr. Albrecht Ott
2. Gutachter: Prof. Dr. Josef Käs
3. Gutachter: Prof. Dr. Thomas Fischer

Tag der Einreichung: 14.12.2007

Tag des Kolloquiums: 04.07.2008



---

# Abstract

The present thesis focuses on nucleic acid hybridization between free-floating target sequences and complementary end-tethered oligonucleotide probes on the surface of DNA microarrays.

Hybridization experiments were performed on oligonucleotide microarrays (DNA Chips) which were fabricated with an automated synthesis apparatus (developed in the framework of the present thesis). The working principle of the microarray synthesizer is based on a photochemically controlled *in situ* synthesis process [Fod91]. By means of the combinatorial approach up to 25000 different (arbitrary) probe sequences can be fabricated in parallel – starting from nucleotide building blocks (NPPOC-phosphoramidites [Has97]) – directly on the surface of the microarray. Great flexibility with regard to the choice of probe sequences is achieved by use of 'virtual photomasks' [SG99] on the basis of a spatial light modulator (Digital Micromirror Device, DMD<sup>TM</sup>, Texas Instruments Inc.). A microscope projection photolithography system is employed to project the 'virtual masks' (i.e. the photomask images shown on the DMD<sup>TM</sup>) onto the surface of the microarray substrate. Spatially controlled photodeprotection of photolabile NPPOC protective groups (followed by coupling of a further nucleotide building block) enables massively parallel synthesis of DNA probe sequences. In the automated synthesis process microarrays are routinely fabricated over night. Comparable *in situ* synthesis systems are currently operated only at very few institutions around the world.

We first report the application of *phosphorus dendrimer* substrates [LB03] in the *in situ* synthesis of DNA microarrays. With the phosphorus dendrimer functionalization we obtained superior results in regard to sensitivity, surface homogeneity, signal/background-ratio and reusability of the microarrays.

We performed microarray hybridization experiments to investigate the impact of single base defects (deliberately introduced *single base mismatches* and *single base bulges*) on the binding affinity of oligonucleotide duplexes. This is particularly interesting with regard to genotyping microarrays which are increasingly employed as a molecular diagnostics tool for the detection of *single nucleotide polymorphisms* (SNPs).

In a number of experiments we investigated the large influence of the single-defect position [Wic06; Poz06; Nai06b] on duplex binding affinity. The origin of this positional dependence – which is apparently not in agreement with the (two-state) *nearest-neighbor model* – had not been identified so far. We discovered that the influence of the defect position is not restricted to *single base mismatches* but can also be observed for *single base bulge* de-

---

fects. On the basis of the *double-ended zipper model* [Gib59; Kit69] (assuming fluctuating end-domain-opening of the oligonucleotide duplex) we could reproduce the experimentally observed positional influence. Moreover, our theoretical investigations on the zipper model indicate a significant positional influence in regard to the contributions of the individual Watson-Crick *nearest-neighbor* pairs to the Gibbs free energy of oligonucleotide duplex formation. The present work provides for the first time a theoretical approach for the *positional-dependent nearest-neighbor model* (PDNN) of Zhang *et al.* [Zha03].

In the *in situ* synthesis process of DNA microarrays random point-mutations are introduced into the microarray probe sequences. We have shown – experimentally and by means of a numerical model – that synthesis-related defects significantly affect microarray hybridization characteristics.

With regard to single base mismatch discrimination, we discovered significant differences between DNA/DNA- and RNA/DNA hybridization: experimental results indicate an improved discrimination of purine-purine mismatch base pairs in RNA/DNA-duplexes.

For the experimentally observed, unexpectedly high stability of *Group II single bulges* [Zhu99] we provide an explanatory approach on the basis of the *zipper model*.

The selection of appropriate (specific and sensitive) probe sequences is of crucial importance for successful application of DNA microarray technology. Our experimental results confirm previous results [Lue03] which show that only a small fraction (in piecewise sections about 20-30%) of a long cRNA target sequence is available for hybridization with the complementary microarray probes. Reduced binding affinities are assumed to originate from the influence of target secondary structure. Using software tools for *antisense oligonucleotide* design (accounting for target accessibility) we were able to predict efficient microarray probes. We discovered evidence that mechanically stable secondary structures (e.g. double-helical sections) interfere with the microarray surface (sterical hindrance) and thus result in reduced microarray binding affinities.

---

# Kurzzusammenfassung

In der vorliegenden Arbeit wurde die Hybridisierung einzelsträngiger RNA- und DNA-*Target*-Sequenzen mit den für die einzelnen Sequenzen spezifischen Oligonukleotid-Probe-Sequenzen auf der Oberfläche von DNA-Microarrays untersucht.

Die hierbei verwendeten Oligonukleotid-Microarrays wurden mittels eines im Rahmen dieser Arbeit entwickelten Microarray-Synthese-Systems auf der Basis eines automatisierten, photolithographisch kontrollierten Syntheseprozesses [Fod91] hergestellt: Mit Hilfe eines kombinatorischen Verfahrens wurden – ausgehend von chemisch modifizierten NPPOC-Phosphoramidit Basenbausteinen [Has97] – in paralleler Weise bis zu 25000 unterschiedliche (frei wählbare) *Probe*-Sequenzen *in situ* auf dem Microarraysubstrat synthetisiert. Eine hohe Flexibilität hinsichtlich der Auswahl der Probe-Sequenzen wird durch die Verwendung virtueller "Photomasken" [SG99] – auf der Basis eines Mikrospiegelarrays (DMD™ Digital Micromirror Device, Texas Instruments Inc.) – erreicht. Mittels einer Mikroskop-Projektions-Photolithographie-Konfiguration wird das Bild des *Spatial Light Modulators* auf die Substratoberfläche abgebildet, um die Entschützung photolabiler NPPOC-Schutzgruppen – und damit die nachfolgende Ankopplung weiterer Basenbausteine – räumlich kontrolliert zu steuern.

Mit den in unseren Experimenten erstmals bei einer *in situ* Synthese verwendeten *Phosphorus-Dendrimer*-Substraten [LB03] konnten im Vergleich mit anderen Linker/Spacer-Molekülen die besten Resultate in Hinsicht auf Sensitivität, Homogenität, Signal/Untergrund-Verhältnis und Wiederverwendbarkeit, erzielt werden. Mit dem Microarray-Synthesizer können in einem automatisierten Prozess DNA Microarrays mit Tausenden von beliebig wählbaren Probe-Sequenzen praktisch über Nacht hergestellt werden. Vergleichbare Systeme stehen bislang nur wenigen Forschungseinrichtungen zur Verfügung.

Anhand von Hybridisierungsexperimenten wurde untersucht, wie sich (gezielt eingebaute) Einzelbasen-Defekte auf die Bindungsaffinität von Oligonukleotid-Duplexen auswirken. Dies ist in Hinsicht auf die Anwendung von SNP-Microarrays interessant, die zur Detektion von *Single Nucleotide Polymorphismen* – genetisch bedingten Variationen einzelner Basenpaare – in zunehmenden Maße in der molekularen Diagnostik eingesetzt werden.

In einer Reihe von Experimenten lag das Augenmerk auf dem starken Einfluss der Defektposition [Wic06; Poz06; Nai06b] auf die Bindungsaffinität. Die Ursache dieser offensichtlich im Widerspruch zum *two-state nearest-neighbor*-Modell stehenden Positionsabhängigkeit konnte bislang nicht erklärt werden. Unsere Experimente zeigen erstmals, dass die Positionsabhängigkeit nicht nur bei Mismatch-Defekten [Wic06; Poz06; Nai06b], sondern in vergleichbarer Stärke auch bei *single bulge* Defekten auftritt. Auf der Basis eines

---

Zipper-Modells des Oligonukleotid-Duplexes, bei dem eine fluktuierende partielle Denaturierung der Duplexenden angenommen wird (die auch zur vollständigen Dissoziation führen kann), konnte der experimentell beobachtete Positionseinfluss reproduziert werden. Darüber hinaus zeigen unsere theoretischen Untersuchungen (auf der Grundlage des *Zipper* Modells) einen signifikanten Positionseinfluss hinsichtlich der Gewichtung der einzelnen *nearest-neighbor*-Beiträge zur Duplexstabilität auf. Die vorliegende Arbeit liefert damit erstmals einen theoretischen Ansatz für das *positional-dependent nearest-neighbor* Modell (PDNN) von Zhang *et al.* [Zha03].

Verursacht durch Streulicht und andere Einflüsse werden im Verlauf der *in situ* Synthese zufällige Punktmutationen in den Microarray-Probe-Sequenzen generiert. Experimentell und in numerischen Modellen konnte gezeigt werden, dass diese Synthesedefekte maßgeblich die Hybridisierungseigenschaften entsprechender Microarrays beeinflussen.

Eine detaillierte Analyse des Einflusses der einzelnen Mismatch-Basenpaare auf die Bindungsaffinität zeigt hinsichtlich der Mismatch-Diskriminierung signifikante Unterschiede zwischen DNA/DNA- und RNA/DNA-Hybridisierung auf, die wahrscheinlich auf unterschiedliche Duplexstrukturen zurückzuführen sind.

Für die experimentell beobachtete, vergleichsweise hohe Stabilität von *Group II single bulge* [Zhu99] Defekten konnte ein Erklärungsansatz auf der Basis des Zipper-Modells gefunden werden.

Für die Durchführung von Microarrayexperimenten ist die Auswahl geeigneter Probe-Sequenzen mit einer hohen Bindungsaffinität hinsichtlich der dazu komplementären Target-Sequenzen von entscheidender Bedeutung. Wir konnten frühere Resultate [Lue03] bestätigen, wonach – vermutlich durch den Einfluss der Targetsekundärstruktur – nur ein relativ kleiner Teil (abschnittsweise etwa 20 bis 30%) einer mehrere hundert Nukleotide langen cRNA Target-Sequenz für die Hybridisierung mit den Microarray-Probes zur Verfügung steht. Auf der Grundlage eines Software Tools für das Design von Antisense-Oligonukleotiden (Berücksichtigung der Targetsekundärstruktur) konnten die experimentell bestimmten Hybridisierungseffizienzen der Microarray-Probe-Sequenzen reproduziert werden. Darüber hinaus entdeckten wir Hinweise dafür, dass mechanisch stabile Sekundärstrukturen (z.B. doppelhelikale Abschnitte) durch Wechselwirkung mit der Microarray-Oberfläche – aufgrund von sterischer Hinderung der Duplexbildung – die Bindungsaffinität herabsetzen.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Fundamentals</b>	<b>7</b>
2.1	Nucleic Acids . . . . .	7
2.1.1	The Double-Helix Structure . . . . .	8
2.1.2	Stabilizing Interactions . . . . .	9
2.1.3	Differences between DNA and RNA . . . . .	14
2.2	Biological Functions of Nucleic Acids . . . . .	15
2.2.1	The Central Dogma of Molecular Biology . . . . .	15
2.2.2	Genomic DNA . . . . .	16
2.2.3	Genes . . . . .	17
2.2.4	Gene Expression . . . . .	17
2.2.5	Expression Regulation . . . . .	19
2.2.6	Biological Functions of RNA . . . . .	22
2.3	Nucleic Acid Hybridization . . . . .	23
2.3.1	Kinetics of Nucleic Acid Hybridization . . . . .	24
2.3.2	The Nearest-Neighbor Model . . . . .	27
2.3.3	Zipper-Model of the Oligonucleotide Duplex . . . . .	30
2.3.4	Further Models of the DNA Melting Transition . . . . .	34
2.4	Destabilization of Oligonucleotide Duplexes by Point Defects . . . . .	34
2.4.1	Single Base Mismatches . . . . .	35
2.4.2	Single Base Bulges . . . . .	36
2.4.3	Influence of the Defect Position . . . . .	39
2.5	Solid-Phase Synthesis of Nucleic Acids . . . . .	41
2.5.1	Principles of Solid-Phase Chemical Synthesis . . . . .	41
2.5.2	Nucleic Acid Synthesis by the Phosphoramidite Method . . . . .	42
2.6	DNA Microarrays . . . . .	45
2.6.1	Microarray Applications . . . . .	47

2.6.2	The Development of DNA Microarray Technologies . . . . .	49
2.6.3	Characteristics of Microarray Hybridization . . . . .	51
2.6.4	Further Reading . . . . .	53
2.7	DNA Chip Fabrication by Light-Directed <i>In Situ</i> Synthesis . . . . .	54
2.7.1	Photolithographic Control of the Combinatorial Synthesis Process . . . . .	54
2.7.2	”Maskless” Photolithography and Combinatorial Chemistry . . . . .	57
<b>3</b>	<b>Development of the DNA Microarray Synthesizer</b>	<b>61</b>
3.1	Motivation and Overview . . . . .	61
3.2	The Maskless Microprojection Photolithography System (MPLS) . . . . .	63
3.2.1	The UV Light Source . . . . .	63
3.2.2	Digital Mask Projection Using a Digital Micromirror Device . . . . .	65
3.2.3	The Image Projection Optics . . . . .	66
3.2.4	UV-Sensitive Photochromic Films . . . . .	69
3.2.5	Chromatic Correction of the Projection Optical System . . . . .	70
3.2.6	UV Light Intensity and Uniformity of Illumination . . . . .	71
3.2.7	Optical System Performance Testing . . . . .	73
3.2.8	Outlook - Further Possible Applications . . . . .	78
3.3	The Fluidics System . . . . .	80
3.3.1	The Synthesis Cell . . . . .	80
3.3.2	Argon Bubble Trapping . . . . .	83
3.4	Automated Microarray Synthesis . . . . .	84
3.5	Performance of the Microarray Synthesizer . . . . .	84
<b>4</b>	<b>Light-directed <i>in situ</i> Synthesis of DNA Microarrays</b>	<b>87</b>
4.1	Light-Directed <i>in situ</i> Synthesis of DNA Microarrays . . . . .	87
4.2	Preparation of Phosphorus Dendrimer Substrates . . . . .	91
4.3	Noteworthy Characteristics of the Microarrays . . . . .	94
4.3.1	Autofluorescence of the Chip Surface . . . . .	94
4.3.2	Hydrophilicity of DNA Microarray Features . . . . .	95
4.3.3	Hybridization without Detergent - Unspecific Adsorption . . . . .	96
4.3.4	Irreversible Target Adsorption . . . . .	96
4.3.5	Robustness of the Phosphorus Dendrimer Surface Coating . . . . .	97
<b>5</b>	<b>DNA Microarray Analysis</b>	<b>99</b>
5.1	Hybridization Signal Acquisition - Experimental Setup . . . . .	99
5.1.1	The Hybridization Chamber . . . . .	100
5.1.2	Epifluorescence Microscope . . . . .	102

---

5.1.3	Image Acquisition with an EM-CCD Camera . . . . .	103
5.2	Quantitative Analysis of Microarray Hybridization Signals . . . . .	103
5.3	Real-time Monitoring of Microarray Hybridization . . . . .	105
5.3.1	Hybridization Buffer . . . . .	106
5.3.2	Microarray Washing Procedures . . . . .	106
<b>6</b>	<b>Influence of Point-Defects on Oligonucleotide Duplex Binding Affinities</b>	<b>109</b>
6.1	Motivation . . . . .	109
6.2	Conception of the Microarray Hybridization Experiments . . . . .	111
6.3	DNA Microarray Design . . . . .	111
6.3.1	Chip Design - Quantitative Analysis of Hybridization Signals . . . . .	113
6.3.2	Single Base Defect Experiments . . . . .	114
6.4	Hybridization Assays and Image Analysis . . . . .	115
6.4.1	Oligonucleotide Targets . . . . .	115
6.5	Dominant Influence of the Defect Position . . . . .	115
6.6	Mismatch Discrimination in DNA/DNA Duplexes . . . . .	121
6.6.1	Experimental Results . . . . .	121
6.6.2	Discussion . . . . .	125
6.7	Influence of Flanking Base Pairs . . . . .	128
6.8	Mismatch Discrimination in DNA/DNA and RNA/DNA Duplexes . . . . .	132
6.8.1	Outline of the Experiment . . . . .	132
6.8.2	Results . . . . .	133
6.8.3	Discussion . . . . .	135
6.9	Single Base Bulge Defects . . . . .	137
6.9.1	Statistical Analysis . . . . .	137
6.9.2	Discussion . . . . .	140
6.10	Comparison of Single Base Mismatches and Single Base Bulges . . . . .	143
6.11	Binding Affinities of Duplexes Containing Multiple Defects . . . . .	146
6.11.1	Results and Discussion . . . . .	147
<b>7</b>	<b>Modeling the Influence of Point Defects on Duplex Stability</b>	<b>151</b>
7.1	The Double-Ended Zipper Model . . . . .	151
7.2	Stochastic Simulation of Oligonucleotide Duplex Stability . . . . .	153
7.2.1	Stochastic Simulation with the Gillespie Algorithm . . . . .	153
7.2.2	Simulation Results . . . . .	155
7.3	Partition Function Approach of the Double-Ended Zipper Model . . . . .	159
7.3.1	Implementation of the Partition Function Approach (PFA) . . . . .	160

7.3.2	Consideration of Point Defects . . . . .	162
7.3.3	Discussion . . . . .	166
7.4	Relation Between the Hybridization Signal and Duplex Stability . . . . .	171
7.4.1	Heterogeneity of Binding Affinities . . . . .	175
7.4.2	Impact of Random Defects Introduced in the Fabrication Process . . . . .	175
7.5	Approximation of the PFA with a PDNN Model . . . . .	179
<b>8</b>	<b>Microarray Experiments</b>	<b>185</b>
8.1	Investigation of the Influence of Synthesis Defects . . . . .	185
8.1.1	Theoretical Considerations . . . . .	185
8.1.2	Evaluation of the Synthesis Yield . . . . .	187
8.1.3	Impact of Synthesis Defects on Microarray Synthesis Fidelity . . . . .	189
8.2	Temperature Dependence of the Cy3-Fluorescence Intensity . . . . .	192
8.2.1	Experiment . . . . .	192
8.2.2	Results and Discussion . . . . .	193
8.3	Single Molecule Imaging on DNA Microarrays . . . . .	196
8.4	Duplex Melting Characteristics on DNA Microarrays . . . . .	198
8.4.1	Experimental Procedures . . . . .	199
8.4.2	Results and Discussion . . . . .	199
8.5	Target Transport Related Effects . . . . .	203
8.5.1	Experimental Observations . . . . .	203
8.5.2	Discussion . . . . .	204
8.6	Influence of Target Secondary Structure on the Duplex Binding Affinity . . . . .	205
8.6.1	Preparation of the cRNA Target Sequences . . . . .	206
8.6.2	Design of the Tiling Array Experiment . . . . .	208
8.6.3	Microarray Hybridization - Experimental Procedures . . . . .	209
8.6.4	Results . . . . .	212
8.6.5	Consideration of the Target Accessibility . . . . .	217
8.6.6	Hybridization to a Stem-Loop Secondary Structure Target . . . . .	220
8.6.7	Nonspecific Hybridization – Variation of the Wash Stringency . . . . .	222
<b>9</b>	<b>Summary/Zusammenfassung</b>	<b>225</b>
9.1	Summary . . . . .	225
9.2	Zusammenfassung . . . . .	235
<b>A</b>	<b>Experimental Data</b>	<b>261</b>
A.1	Experimental Data . . . . .	262
A.1.1	Comparison: MMs in RNA/DNA and DNA/DNA Duplexes . . . . .	262

---

A.1.2	Single Base Insertion Defect Profiles . . . . .	269
A.1.3	Single Base MMs in DNA/DNA Duplexes - Statistical Analysis . . . . .	272
A.1.4	Single Base Insertions - Statistical Analysis . . . . .	279
<b>B</b>	<b>Supporting Information</b>	<b>283</b>
B.1	The Digital Micromirror Device (DMD <sup>TM</sup> ) . . . . .	284
B.2	Modification of the DLP Video Projector . . . . .	286
B.2.1	Gamma-function of the DLP-Projector . . . . .	286
B.3	Optics of the Microscope Projection Photolithography System . . . . .	288
B.4	Fabrication of the Synthesis Cell . . . . .	290
B.5	Technical Notes on Light-directed DNA Chip Synthesis . . . . .	292
B.5.1	Handling of Phosphoramidite Reagents . . . . .	292
B.5.2	Additional Notes on the Synthesis . . . . .	292
B.6	Technical Notes on Microarray Dendrimer Substrate Preparation . . . . .	293
B.7	Technical Notes on the Synthesizer Control Software <i>DNASyn</i> . . . . .	294
B.7.1	Basic Features . . . . .	295
B.7.2	Communications between the Control PC and the Synthesizer . . . . .	297
B.7.3	Dual Screen Support . . . . .	297
B.7.4	Additional Hints . . . . .	298
B.8	Solenoid Valve Driver for Fluidics Control . . . . .	298
B.9	DNA Microarray Design . . . . .	299
B.10	Microarray-Analysis with <i>ScanRA</i> . . . . .	302
B.11	Temperature Control of the Hybridization Chamber . . . . .	305
B.12	cRNA Secondary Structures . . . . .	306
B.13	3-D Visualization of Nucleic Acid Structures . . . . .	308



# List of Abbreviations

$\delta I_{mp}$	hybridization signal intensity deviation from the mean profile
$T_m$	melting temperature
A	adenine
APTES	3-aminopropyl-triethoxysilane
bp	base pair
C	cytosine
CCD	charge coupled device
cDNA	complementary DNA
Cy3 <sup>TM</sup>	cyanine-3 fluorescent marker
D	duplex
del	deletion
DLP <sup>TM</sup>	digital light processing
DMD <sup>TM</sup>	digital micromirror device
DNA	deoxyribonucleic acid
DOF	depth of focus
EGFP	enhanced green fluorescent protein
EM-CCD	electron-multiplying CCD
Fb	fraction bound
G	guanine
HPLC	high performance liquid chromatography
ins	insertion
IVT	<i>in vitro</i> transcription
MeCN	acetonitrile
MeNPOC	[ $\alpha$ -methyl-2-nitropiperonyl-oxy]carbonyl
MeOH	methanol
MM	mismatch
MPLS	maskless microscope projection lithography system

mRNA	messenger RNA
NA	numerical aperture
NN	nearest neighbor
NPPOC	2-(2-nitrophenyl)propoxycarbonyl
nt	nucleotide
OD	optical density
P	probe
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PDNN	positional dependent nearest neighbor model
PFA	partition function approach
PM	perfect match
PTFE	polytetrafluoroethylene
pur.	purine
pyr.	pyrimidine
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
SLM	spatial light modulator
SNP	single nucleotide polymorphism
T	target
T	thymine
THF	tetrahydrofuran
tRNA	transfer RNA
TSNN	two state nearest neighbor model
U	uracil
UHP	ultra high pressure
UV	ultra violet
WC	Watson-Crick

# Glossary

**defect profile**

this expression has been introduced in the present work for the microarray hybridization signal as a function of defect type and defect position; within a defect profile – based on one particular probe sequence motif – defect type and defect position are varied systematically

**feature**

element of the microarray; small area on the surface of the microarray containing one particular species of microarray probe sequences; the microarray comprises a regular grid arrangement of features

**gene expression**

process in which the genetic information of a gene is converted into a gene product

**hybridization**

binding of two complementary single-stranded nucleic acids to form a double-stranded duplex; hybridization results from sequential base pairing of complementary base pairs

**hybridization signal**

fluorescence intensity of hybridized microarray targets on the surface of the microarray

**immobilization**

tethering of prefabricated nucleic acid probe sequences on the microarray substrate

***in situ* synthesis**

synthesis of probe sequences (from nucleotide building blocks) directly on the substrate of the microarray

**nearest-neighbor model**

the nearest-neighbor model of nucleic acid duplex thermal stability considers hydrogen-bonding and base-stacking interactions; the stacking interactions between directly adjacent (nearest-neighbor) base pairs comprise dispersion forces, electrostatic interactions and hydrophobic interactions

**oligonucleotide**

short nucleic acid strand

**perfect match**

duplex consisting of two completely complementary strands; defect-free duplex

**probe**

a microarray probe is used to detect/identify one specific nucleic acid target sequence; probes are typically oligonucleotide probes (length < 100 nt) or several hundred nt long cDNA sequences; probes are tethered in a regular arrangement (array) – within the microarray features – on the solid support

**probe sequence motif**

in the present work this expression is used for the perfect matching probe sequence that is complementary to the oligonucleotide target sequence employed in a single base defect hybridization experiment. Single base defect probes are derived from the 'probe sequence motif' by substitution, insertion or deletion of a single base. The probe sequence motif may be shorter than the target oligonucleotide used in the experiment. Hybridization signals from the complete set of single base defect probes correspond to the 'defect profile'.

**single base bulge**

defect in a nucleic acid duplex which originates from a surplus unpaired base in one of the two strands; the surplus base can adopt a stacked-in conformation or a looped-out conformation and can result in significant reduction of the binding affinity

**single base mismatch**

defect in a nucleic acid duplex which originates from a non-Watson-Crick base pair; the reduced binding affinities is employed for detection of SNPs and point-mutations

**target**

free nucleic acid sequence whose identity and abundance are to be detected in the microarray assay; for detection target sequences are commonly labeled with fluorescent dyes or with biotin

# Chapter 1

## Introduction

Almost all cells of the human body, regardless of the cell type, contain the same genetic material. However, owing to epigenetic factors (e.g. CpG methylation) the cell types differ in their gene expression – for example, genes which are strongly expressed<sup>1</sup> in one cell type, may not be expressed in others. Knowledge on gene expression is the key for understanding the individual gene functions and the complex interactions between the about 20,000 to 25,000 genes of the human genome.

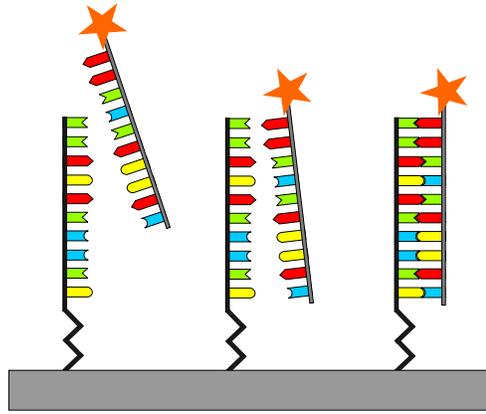
DNA microarrays are a key technology for massively parallel analysis of gene expression. The working principle of DNA microarrays is based on nucleic acid hybridization: sequential Watson-Crick base pairing between the bases of two complementary nucleic acid strands results in the formation of a relatively stable double-helical duplex. Nucleic acid hybridization is highly specific – already a single mismatched (non-Watson-Crick) base pair can significantly reduce the binding affinity [Nel81; Pat82].

The sequence-specific hybridization between complementary strands is employed for the purpose of molecular recognition (Fig. 1.1): surface-tethered single-stranded probes (of known sequences) are employed as sequence-specific scavengers for complementary target sequences in solution. Hybridized target molecules (bound to the surface) can be detected by means of radioactive or fluorescent dye labels.

On DNA microarrays the same detection principle is applied in parallel fashion (Fig. 1.2). Owing to the high specificity of nucleic acid hybridization thousands or even millions of different target sequences can be detected simultaneously. DNA microarrays comprise a regular array of *microarray features*, small areas, each of which is covered with surface-tethered single-stranded DNA probes of a well-known sequence. Individual microarray features (and thus the corresponding probe and target sequences) can be identified by their position on the microarray.

---

<sup>1</sup> Gene expression – the conversion of genetic information into gene products – can be understood as 'gene activity'.



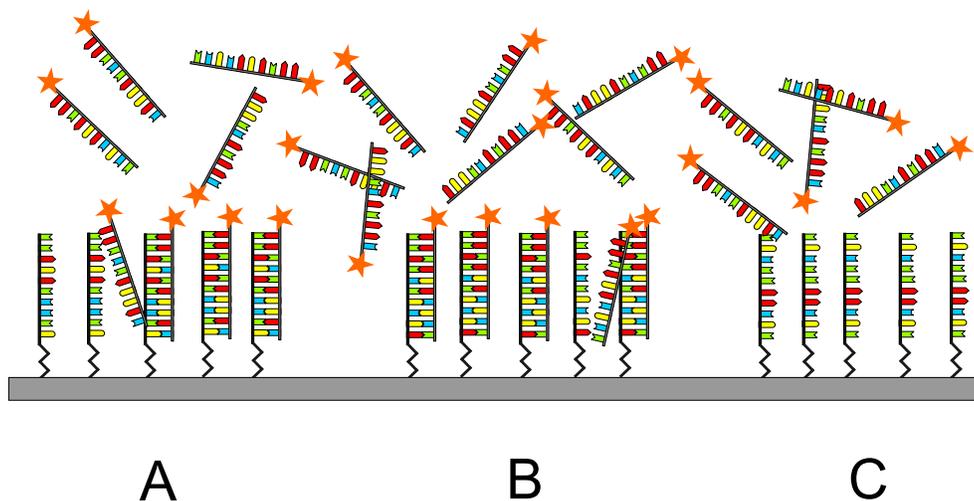
**Figure 1.1:** Nucleic acid hybridization between surface-tethered probe strands and complementary target strands in solution. Nucleic acid hybridization is based on sequential Watson-Crick base pairing between complementary sequences of nucleotides and results in the formation of a relatively stable double-helical nucleic acid duplex. Nucleic acid hybridization is reversible (dissociation is favored by increased temperatures) because the individual binding interactions (hydrogen bonding and base stacking interactions – no covalent bonds involved) between the base pairs are relatively weak. Targets strands are labeled by covalent linkage of a fluorescent dye, or alternatively, by biotinylation.

In a gene expression profiling experiment the messenger RNA (mRNA) sequences (indicators of the individual genes transcriptional activities) are isolated from the biological sample, amplified (if necessary, e.g. by *in vitro* transcription), and labeled for detection. Subsequently the complex mixture of target sequences to be analyzed is applied (in hybridization buffer solution) onto the surface of the microarray. The target strands can freely diffuse around and interact with the surface-tethered microarray probes, until they are captured by a complementary probe and form a stable duplex.

After removal (washing-off) of unhybridized targets, the hybridization signal, which provides information on the quantity of the individual target sequences, is commonly detected by means of fluorescent markers. Comparison of the hybridization signals with the corresponding hybridization signals from a reference sample (by dual-color analysis on the same chip, or by means of two single-channel microarrays) enables identification of genes that have been up- or downregulated. Some commercial platforms enable gene expression profiling on a genome-wide scale.

Genotyping analysis is a further important microarray application: Single nucleotide polymorphisms (SNPs) – sequence variations in which single nucleotides differ between the members of a species<sup>2</sup> (or even between the two alleles in diploid cells) – have a strong influence on the phenotype. SNPs are responsible for the majority of genetic variations

<sup>2</sup> The human genome contains about 3 million SNPs. Thus, about one in a thousand base pairs is subject to this type of inheritable genetic variation.



**Figure 1.2:** Nucleic acid hybridization on the microarray. Three different surface-tethered probe species  $a$ ,  $b$  and  $c$  are located separate from each other within the corresponding *microarray features* A, B and C. A complex mixture of different target sequences is applied to the microarray surface. Driven by diffusion (or active mixing) targets move around and interact with the different probes. If a target meets a complementary probe, a stable duplex can arise. Thus, the target gets captured by the complementary probe. After the hybridization the unbound targets can be removed by washing-off. The remaining hybridization signal (fluorescent signal) of the hybridized probes provides information on the quantities of the individual target sequences. In this example we observe hybridization signals only at features A and B. We conclude that the sample mixture contains targets sequences that are complementary to the probes  $a$  and  $b$ . The sample does not contain targets complementary to probe  $c$ .

within a single species. They are also associated with a predisposition to a variety of diseases. Moreover, SNPs are associated to individuals' response to pathogens, chemicals, drugs, vaccines, and other agents. SNP microarrays make use of the specificity of relatively short 12 to 30mer oligonucleotide probes to detect single mismatched base pairs originating from SNPs [Con83]. Genotyping arrays are a valuable tool in genomics research, pharmaceutical research (with a focus on the individual response to pharmaceutical agents) and now increasingly in medical diagnostics.

Further applications of DNA microarrays include resequencing assays<sup>3</sup> and the identification of pathogens.

Lab-scale fabrication of DNA microarrays on the basis of standard techniques requires considerable technical and financial efforts.<sup>4</sup> To provide a flexible and affordable basis for DNA microarray hybridization experiments we developed a *DNA microarray synthesizer*

<sup>3</sup> Resequencing arrays are used for the search for mutations with respect to a well-known reference sequence. An important application is the identification of (possibly new) virus strains.

<sup>4</sup> These include, for example, the acquisition of a microarray spotting robot (to be operated in a clean room environment) and considerable running expenses for presynthesized microarray probes.

based on the work of Singh-Gasson *et al.* [SG99]. Based on a photochemically controlled *in situ* synthesis process, the DNA probe sequences are synthesized from nucleotide building blocks, directly on the surface of the microarray. The use of expensive chromium photomasks (and associated mask alignment) is circumvented by means of a spatial light modulator (DMD™) obtained from a commercial video projector. Comparable *in situ* synthesis systems are currently operated only at a few institutions around the world.

Even though DNA microarrays have become a well-established technology, the underlying physicochemical principles of DNA microarray hybridization are not yet fully understood [Lev05; Poz06]. For example, an unresolved problem in the application of DNA microarrays is the lack of predictability of the hybridization efficiency of DNA microarray probes. Thermal stability of oligonucleotide duplexes (in solution-phase) is well described by the *nearest-neighbor model* [Cro64; Tin73; Bre86; Fre86], which is accounting for hydrogen bonding and also for base-stacking interactions between adjacent base pairs. Thermodynamic parameters for *nearest-neighbor* doublets of base pairs were derived from solution-phase hybridization experiments [San98]. The *nearest-neighbor model* is widely employed for the prediction of duplex melting characteristics (melting temperatures, Gibbs free energies of duplex formation) – for example, for the design of PCR primers and for the design of DNA microarray probe sequences. The latter application, however, is questionable: on DNA microarrays, due to various surface-effects and fabrication-related effects, there are significant differences with respect to solution-phase hybridization [Hel03; Lev05; Bin06; Poz06]. Moreover, the secondary structure of long target sequences results in a restricted target accessibility. Thus, the binding affinity of individual microarray probes is also governed by the complex target secondary structure [Lue03; Rat05].

In contrast to solution-phase hybridization studies, recent microarray studies [Wic06; Poz06; Nai06b] report a large influence of the position of *single base mismatch* defects on the hybridization signal. A position dependent influence of single base defects is not considered by the (two-state) *nearest-neighbor model*<sup>5</sup> and hasn't been explained so far. According to Pozhitkov *et al.* [Poz06] there is little evidence that microarray hybridization efficiencies can be accurately predicted with software tools on the basis of *nearest-neighbor* thermodynamic parameters derived from solution-phase experiments.

In the experimental part of the present thesis particular interest is on the influence of point defects (*single base mismatches* and *single base bulges*) on microarray binding affinities. We systematically investigate the influence of defect type and defect position on probe-target binding affinities. In the same context we investigate differences between RNA/DNA

---

<sup>5</sup> The *nearest-neighbor model*, on the basis of mismatch base pair *nearest-neighbor* parameters [All97], is also employed for mismatched duplexes [All97; San04]. The model does not consider a position dependent influence, except for the outermost base positions.

and DNA/DNA hybridization. Our theoretical investigation on the influence of point defects on duplex binding affinities is based on a *zipper model* [Gib59; Kit69] of the oligonucleotide duplex.

Further experiments address a variety of poorly understood influences on DNA microarray hybridization. These include:

- random defects in the microarray probes (generated by the *in situ* synthesis process) affect the hybridization characteristics [Job02; Gar02; Hel03]
- the complex secondary structure of long target sequences (widely believed to be a main factor influencing the efficiency of hybridization [Lue03])
- nonspecific cross-hybridization
- diffusion limitation – local depletion of the hybridizing target molecules can result in inhomogeneous hybridization signal intensities and slowed-down hybridization kinetics [Pap06; Dan07]



# Chapter 2

## Fundamentals

### 2.1 Nucleic Acids

For its outstanding role in molecular biology DNA (deoxyribonucleic acid) is often referred to as the "molecule of life". Like a blueprint genomic DNA contains the hereditary information, instructions to grow and sustain all forms of life.

In the higher eucaryotic organisms the genomic DNA is densely packed on chromosomes inside the cell nucleus. Each chromosome comprises a single double-helical DNA molecule. The length of the human chromosomes is varying between  $50 \times 10^6$  and  $250 \times 10^6$  base pairs (corresponding to lengths between 1.7 and 8.5 cm). Overall, stretched end-to-end, the DNA helix contained in a single human diploid cell is about 2 meters long. The information density in the densely packaged nucleus is about  $10^{21}$  bit/cm<sup>3</sup><sup>1</sup> (for comparison: the information storage density on a DVD disc is about  $10^9$  bit/cm<sup>2</sup>).

The biological function of DNA is the safe storage of genetic information. Genomic DNA is basically a read-only memory and in this way comparable to the CD-ROM drive of a computer. Parts of the genome (the genes) are read in the transcription process to produce RNA transcripts of the DNA sequence. RNA is a rather volatile information carrier in the ongoing processing of genetic sequence information. In the above analogy it is therefore comparable with the working memory (RAM) of a computer. However, RNA is more versatile: its not just an information carrier but rather (in form of functional RNA) a crucial part of the translational machinery and involved in regulatory processes.

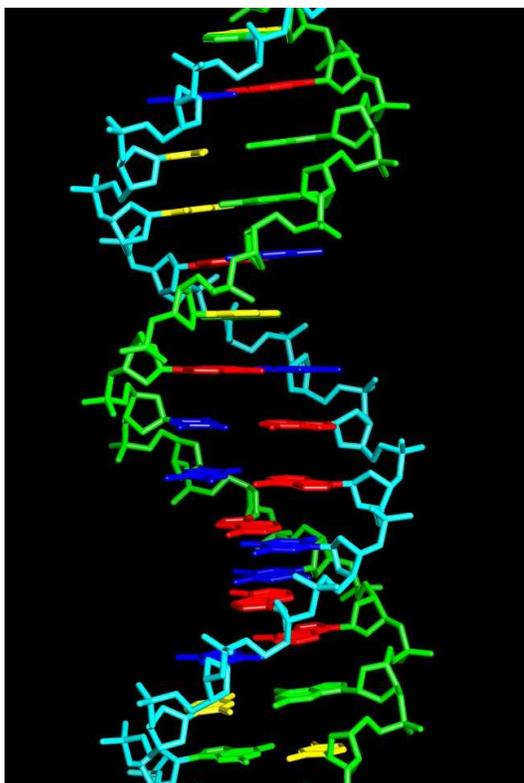
---

<sup>1</sup> The estimate is based on a cell volume of  $8 \mu\text{m}^3$  and a genome size of  $3 \cdot 10^9$  base pairs - which is about the size of the human genome.

### 2.1.1 The Double-Helix Structure of Nucleic Acids

The structure of deoxyribonucleic acid (DNA) was discovered by James Watson and Francis Crick in 1953. A few month earlier, Linus Pauling reported a triple-helix model of the DNA-structure, which assumed that the phosphate groups are arranged in the interior of the helix. The model was based on high resolution electron micrographs showing the DNA as cylindric fibrils with a diameter of 1.5 nm. The wrong assumption of a triple helix originated from the incorrect measurement of a too high packaging density.

Watson and Crick showed that under physiological conditions DNA has indeed a double-helical structure (Fig. 2.1). The hydrophobic bases are located in the center, whereas the hydrophilic phosphate groups are located at the outside of the helix. The discovery of Watson and Crick relies on the work of Rosalind Franklin, whose X-Ray structural analysis of DNA fibres proved that DNA has indeed a helical structure.

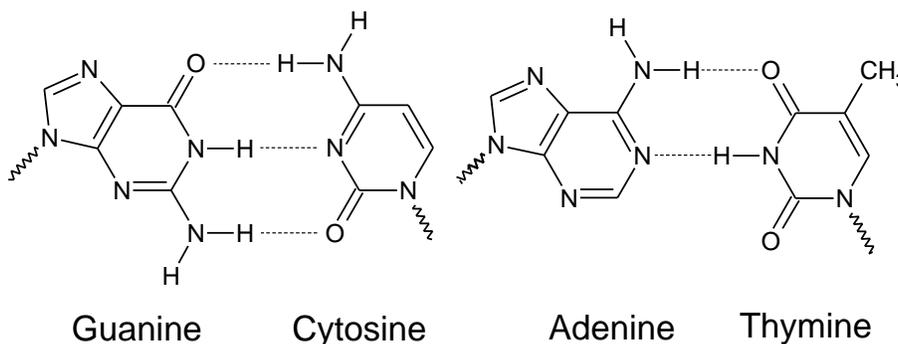


**Figure 2.1:** Watson-Crick model of the DNA double-helix. Canonical (Watson-Crick) base pairs comprise either adenine (red) and thymine (blue), or guanine (green) and cytosine (yellow) bases. The sugar-phosphate-backbones of the two strands (shown in green and cyan) form a right-handed double helix. The ideal B-DNA structure was generated with the *make\_na* web-server which is based on the NAB (Nucleic Acid Builder) by Tom Macke [Mac98]. Image visualization was performed with the UCSF Chimera molecular modeling system [Pet04]. A three-dimensional stereo view of the DNA structure is shown in the appendix, in Fig. B.19.

Another important hint was provided by Erwin Chargaff. According to *Chargaff's rule* the nucleobases A and T, just as the nucleobases C and G always occur with the ratio of about 1:1, independent of the biological origin of the DNA.

### 2.1.2 Nucleic Acid Duplex Structure - Stabilizing Interactions

The DNA double-helix shown in Fig. 2.1 is composed of two complementary single-stranded DNA molecules. It's well-known that the duplex stability originates from inter-strand hydrogen bonding between complementary base pairs A·T and C·G (see Fig. 2.2). However, it is less well-known that a similar degree of stabilization originates from  $\pi$ - $\pi$  interaction between closely-stacked aromatic bases ( $\pi$ -stacking) [Koo01].



**Figure 2.2:** Canonical (Watson-Crick) base pairs A·T and C·G, comprise a bicyclic purine base (adenine or guanine) and a monocyclic pyrimidine base (thymine or cytosine). A·T is stabilized by two, C·G by three hydrogen bonds.

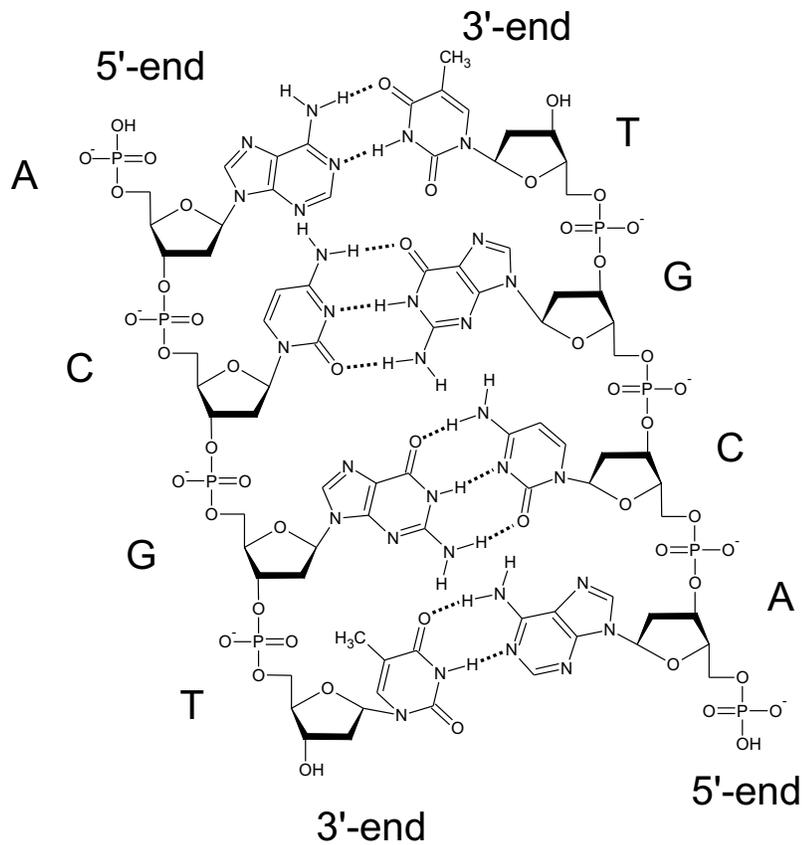
A DNA molecule is basically a flexible polymer-chain<sup>2</sup> made up of nucleotide monomers (as shown in Fig. 2.3).

A nucleotide consists of a heterocyclic base (i.e. adenine, cytosine, guanine or thymine - in RNA the thymine is replaced by uracil) and a pentose sugar-ring (2-deoxyribose in DNA and ribose in RNA), which in conjunction with a phosphate group constitutes the sugar-phosphate-backbone of the DNA molecule. Apart from the nucleobases listed above, further nucleobases occur naturally in RNA (e.g. pseudouridine in transfer-RNA).

Fig. 2.3 shows that subsequent nucleotides are linked via a phosphodiester bond (i.e. over the phosphate group) between the 3'- and 5'-carbons of the deoxyribose sugars. Because

<sup>2</sup> Here one needs to distinguish between the highly flexible single stranded molecule (persistence length values provided in the literature range from  $l_p \simeq 0.5$  nm to 1.3 nm [Koh06]) and the significantly more rigid double-stranded DNA duplex ( $l_p \simeq 45$ -50 nm [Hag88]).

the single bonds of the phosphodiester-linkage enable free rotation of the nucleotides, single stranded nucleic acids have a highly flexible conformation.



**Figure 2.3:** Chemical structure of the DNA/DNA duplex. The duplex is stabilized by hydrogen bonding between complementary base pairs and (though not obvious from this drawing) by base stacking interactions between adjacent base pairs. The two complementary strands have opposite orientations.

The sugar-phosphate-backbone also determines the orientation of a DNA strand. According to convention nucleic acid sequences are commonly written in 5'→3' direction (e.g. 5'-ACGGAGGAG-3'). The two strands of double helix are oriented in opposite directions. Due to the negative charge of the phosphate groups, DNA is a strong electrolyte and thus dissolves well in aqueous solution.

The bases are linked to the 1'-carbon atoms of the deoxyribose via single-bonds, thus providing a high degree of conformational freedom. The hydrophobic bases arrange tightly stacked in the center of the helix whereas the hydrophilic phosphate groups of the backbone form the outside of the helix.

## Watson-Crick base pairing

DNA duplexes consist of the complementary base pairs A·T and C·G (Fig. 2.2). These so-called Watson-Crick (or canonical) base pairs comprise a bicyclic purine base (A or G) and a monocyclic pyrimidine base (C or T). A·T base pairs are stabilized by two, C·G base pairs by three hydrogen bonds. Since canonical base pairs have almost the same size they can form a homogenous double-helical structure, independent of the base sequence.

In functional RNA structures (e.g. in ribosomal RNA or transfer-RNA) non-Watson-Crick base pairs are frequently observed. Accommodation of non-Watson-Crick base pairs may result in structural distortion (with respect to the A-form or B-form helix structure). Depending on the nature of the particular mismatch pair, hydrogen bonding may be prevented and/or steric hindrance may occur. As a result, mismatched base pairs (MM base pairs) can significantly reduce the binding affinity of nucleic acid duplexes.

## Base stacking interactions

Stacking interactions between the flat aromatic rings of adjacent base pairs<sup>3</sup> (*nearest-neighbor* base pairs) are of similar importance for duplex stability like hydrogen bonding [Koo01]. The base stacking interaction ( $\pi$ -stacking) comprises:

- Hydrophobic interaction. Tight stacking of the bases is largely due to the hydrophobic effect. The plane faces of the aromatic bases are hydrophobic whereas the small edges are hydrophilic. Water molecules, forming a highly ordered clathrate-like cage around the hydrophobic nucleobases, are released when bases stack upon each other. The entropy increase from the release of water molecules (hydrophobic effect) is one of the main driving forces for the formation of a compact double-helical structure.
- The van-der-Waals interactions (dipole-dipole interactions between induced dipoles) between the closely spaced (stacked), overlapping aromatic rings of the nucleobases stabilize the NN pairs.
- Electrostatic interactions of partial charges can either have a stabilizing or destabilizing effect on a NN pair.

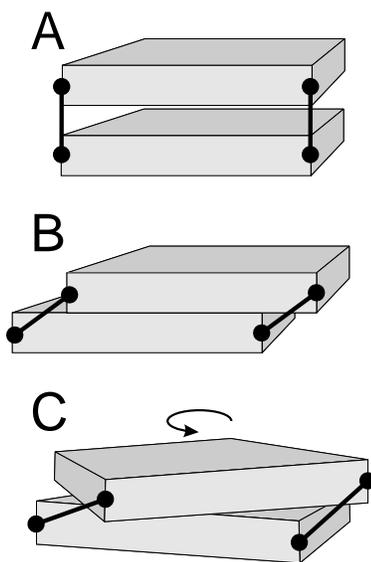
The compact arrangement of the bases in the center of the duplex shields the hydrogen bonds from competing water molecules. This entails a further stabilization of the double helix.

The entropy increase of water molecules released from the clathrate cage around the bases is one of the main driving forces for nucleic acid duplex formation. However, base stack-

---

<sup>3</sup> Base stacking is not restricted to duplexes, but has also been observed in single stranded nucleic acids.

ing results in a decrease of the DNA molecules conformational entropy, since in the duplex conformational degrees of freedom are reduced. At low temperatures binding enthalpy and the entropy increase from released water molecules can balance the decrease of duplex conformational entropy, thus duplex formation is favorable. With increasing temperature the delicate balance between enthalpy and entropy ( $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ ) is shifting towards a positive  $\Delta G^\circ$ , thus duplex formation finally becomes unfavorable.

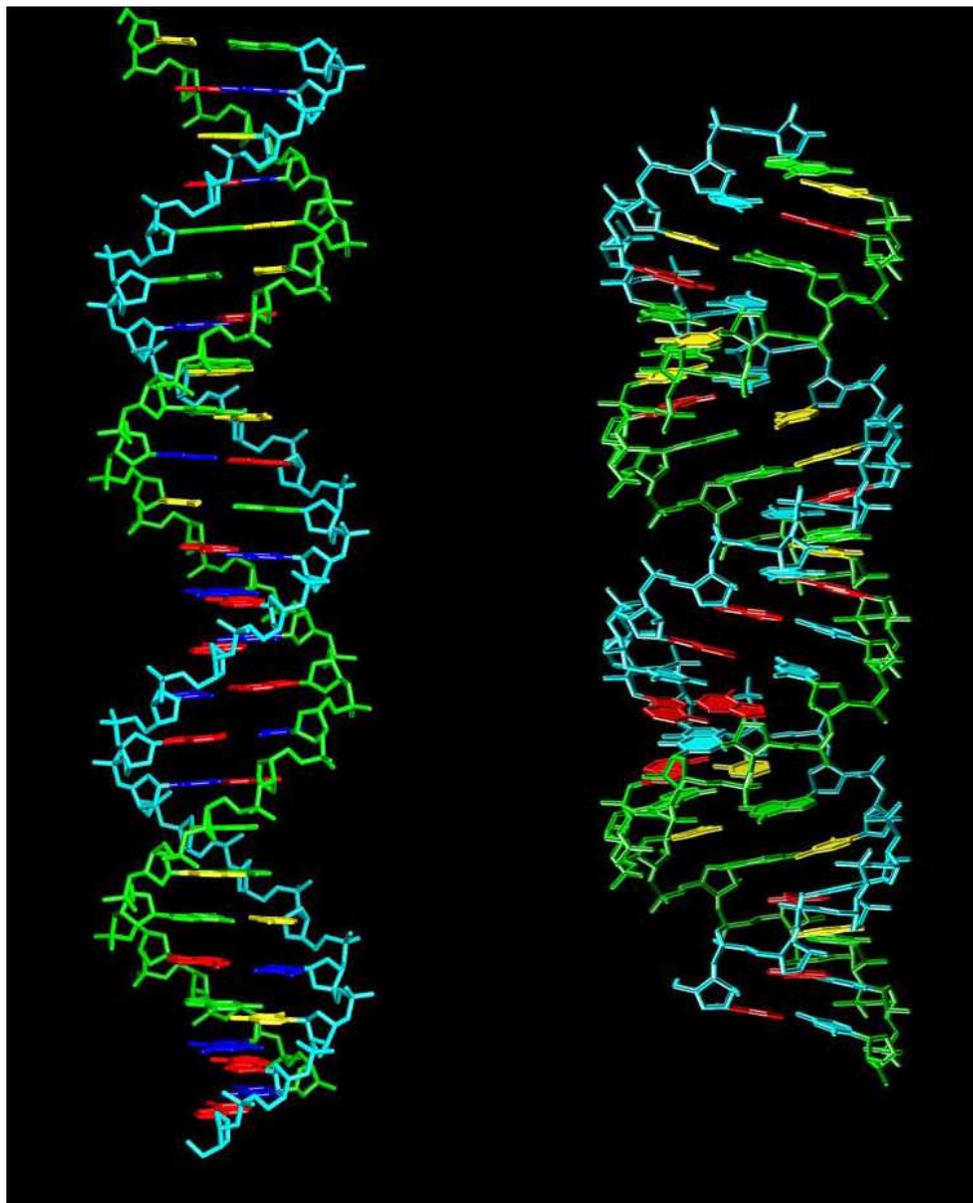


**Figure 2.4:** Origin of the helix twist. In principle the base pairs (drawn as blocks) could arrange as a linear ladder (A). However, stacking interactions favor a closely stacked arrangement of the aromatic rings. Fixed bond lengths between consecutive nucleotides (distance ca. 6 Å) can be accommodated in a skewed ladder structure as shown in (B). Further structural constraints, however, don't allow this skewed ladder, but rather enforce a twisted ladder structure, similar to a spiral-staircase. The twist angle between consecutive base pairs is on average  $36^\circ$  (in B-form DNA duplexes).

The double-helical structure of nucleic acid duplexes originates from compact base stacking. A linear conformation of the nucleotides (as shown in Fig. 2.4A) with a base separation of 0.6 nm due to the hydrophobic effect is less favorable than the compact B-form conformation with a base separation of 0.34 nm. The length difference is compensated by twisting of the nucleotides relative to each other (see Fig. 2.4C). A twist of  $36^\circ$  per base pair results in the helical structure of B-DNA with about 10 base pairs per turn and a pitch of 3.4 nm.

B-DNA (Fig. 2.5 left) is the prevailing helical structure of DNA under physiological conditions. The diameter of the B-DNA helix is about 2 nm.

RNA/RNA duplexes and hybrid duplexes (RNA/DNA duplexes) always adopt an A-form helix structure (Fig. 2.5 right). Due to a smaller twist angle A-RNA has about 11 base pairs per turn and a diameter of 2.6 nm. DNA/DNA duplexes can also adopt an A-form



**Figure 2.5:** Comparison of B-DNA (left) and A-RNA (right) duplex structures. Sequences of both duplexes are identical (except for the substitution of thymine by uracil in RNA). Ideal B-DNA and A-RNA structures were generated with the *make\_na* web-server which is based on the NAB (Nucleic Acid Builder) by Tom Macke [Mac98]. Image visualization was performed with the UCSF Chimera molecular modeling system [Pet04]. More detailed views and stereo-views of nucleic acid structures are included in section B.13.

helix. The A-form of DNA is observed under partially dehydrated conditions (e.g. in an ethanol solution or under dry conditions).

Beyond simple linear duplexes nucleic acids can form highly complex structures like ribosomal RNA (see below). Rothemund [Rot06] recently demonstrated 'bottom-up fabrication' of complex DNA nanostructures: Arbitrary two-dimensional shapes can be created by hybridization-based self-assembly of a set of tailormade oligonucleotide sequences.

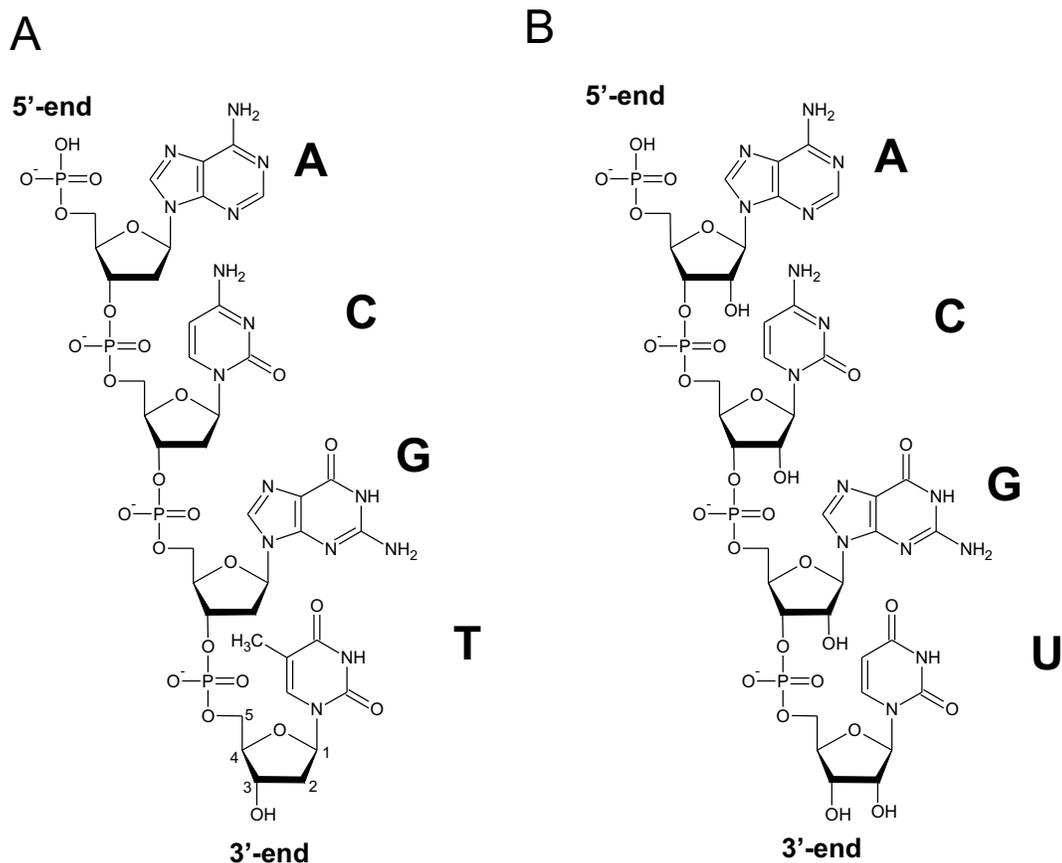
### 2.1.3 Differences between DNA and RNA

As shown in Fig. 2.6 ribonucleic acid RNA looks very similar to DNA, however, there are significant differences in the molecular structure, chemical stability and biological function.

In RNA a hydroxyl-group is attached to the 2'-carbon of the sugar ring, whereas DNA (deoxyribonucleic acid) has a hydrogen-moiety instead (as shown in Fig. 2.6B). Owing to the 2'-hydroxyl-group the conformational freedom of the RNA-duplex is reduced (in comparison with DNA). Therefore, different from DNA, RNA-duplexes can only adopt an A-form duplex structure (see Fig. 2.5). In A-form duplexes, due to a larger lateral offset between stacked bases, the dispersive interaction between the bases is more favorable than in B-form duplexes where the offset between stacked bases is smaller. A-form helices, owing to slightly stronger base-stacking interactions are thermodynamically more stable than B-form helices. The 2'-OH group strongly affects of the chemical stability of RNA. Under alkaline conditions deprotonation of the OH-group can occur. The remaining oxygen can react with the adjacent phosphor atom of the sugar-phosphate-backbone. The subsequent decay of the emerging cyclic phosphate leads to strand breakage. An important biotechnological application is the specific degradation of RNA: under alkaline conditions RNA strands are degraded, whereas DNA strands remain unaffected.

Another important difference between RNA and DNA is the substitution of the DNA-typical base thymine by uracil: basically uracil is favorable, since organisms can produce uracil with less effort than thymine. The use of the base thymine in DNA is related to DNA repair mechanisms, meant to protect the genetic material from mutations. A common mutation caused by chemical action is the desamination of the base cytosine which is thereby converted to uracil. Since DNA repair enzymes can differentiate between DNA-typical thymine and uracil such mutations can be reliably detected and repaired.

The following section discusses the very different biological functions of DNA and RNA.



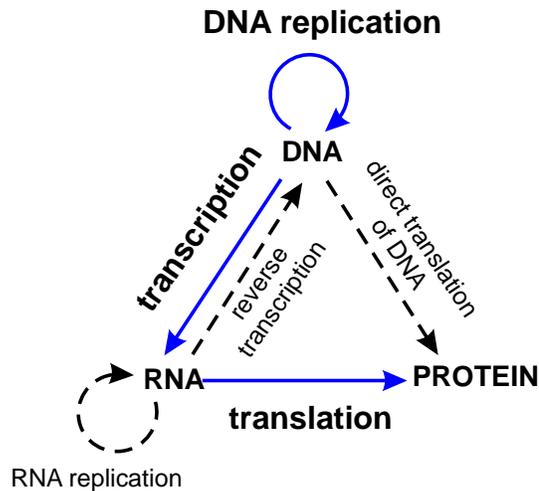
**Figure 2.6:** Structural formulas of DNA (A) and RNA (B) nucleic acid strands. The arrangement of the ribose rings determines the orientation of the strand (here 3'-C at the bottom, 5'-C at the top). Sequences are commonly written from 5'-end to 3'-end (here: 5'-ACGT-3'). The phosphodiester bonds between nucleotides enable free rotation. (B) RNA has a hydroxy-group at the 2'-C of the ribose ring. In RNA the thymine base is substituted by the similar base uracil.

## 2.2 Biological Functions of Nucleic Acids

### 2.2.1 The Central Dogma of Molecular Biology

The *Central Dogma of Molecular Biology* [Cri70] describes the flow of biological sequence information (Fig. 2.7). In the transcription process the genetic information (encoded in DNA sequence) is gene-wise transcribed into messenger RNA (mRNA). The genetic information (carried by the mRNA sequence) is translated via the genetic code into a polypeptide sequence which finally folds into a protein.

The reverse flow of biological information, from the protein back to the genome, is not observed. However, retroviruses can transcribe their RNA-based genetic information into the DNA-based genome of other organisms.



**Figure 2.7:** Francis Crick's *Central Dogma of Molecular Biology*. There are general transfers of biological sequence information (solid arrows) and specialized transfers (dashed arrows). A general transfer of sequence information is from DNA via the transcription process to messenger RNA. mRNA is translated into a polypeptide chain which folds into a protein. Another general transfer is the replication of DNA during cell division. Specialized transfers are related to virus reproduction (e.g. reverse transcription) or have been performed in vitro (e.g. direct translation of DNA sequences into proteins).

## 2.2.2 Genomic DNA

The genomic DNA contains the hereditary information of an organism. Large parts of the genome are arranged as genes, organizational units that are transcribed into one or several gene products. The function of other noncoding parts of the genome, previously termed "junk DNA" is less well understood.

In the simple prokaryotic organisms (e.g. bacteria) the DNA is packaged in ring-shaped chromosomes and plasmids, which are residing in the cytoplasm. In the more complex eukaryotic organisms the DNA is contained in the nucleus, well-separated from the cytoplasm (see Fig. 2.8). Chromosomes contain the DNA in a highly compact, though ordered and accessible form. The double-helical DNA filament contained on a single chromosome can be several centimeters long. Enlarged to a diameter of 2 mm the DNA filament would extend over a length of about 30 km.

In conjunction with a complex of histone proteins, acting as spool around which the DNA double-strand is wound up (roughly two superhelical turns of about 80 base pairs around the cylindrical histone octamer), the DNA forms a nucleosome. Countless nucleosomes condense into an ordered superstructure, forming a chromatin fibre with a diameter of about 30 nm. The chromatin fibre (which via certain domains is connected to the nuclear matrix proteins) forms innumerable loops which compose the structure of the chromosome.

The degree of chromatin condensation is largely determined by the cell-cycle. The chromatin structure, since it determines the accessibility and readability of genes, has a strong influence on gene activity. Transcriptional active regions correlate with an open chromatin structure (euchromatin).

### 2.2.3 Genes

A gene can be understood as a functional unit of the genetic material, which contains the blue print for a gene product. A gene product can be one or several proteins (or subunits of proteins) or a functional RNA, e.g. microRNA (miRNA), ribosomal RNA (rRNA), or transfer RNA (tRNA).<sup>4</sup>

### 2.2.4 Gene Expression

Gene expression can be understood as gene activity. It describes how much of a gene product is produced from each particular gene. The gene activity can be regulated at different stages, e.g. at the transcription initiation, or post-transcriptional, at the mRNA level.

The various cell types of a higher organism all contain the same genetic information. However, the gene activities are different, depending on the requirements of the particular cell function.

#### Transcription

Transcription requires opening of double helix structure first. It is assumed that the reduced duplex stability within the Pribnow-box (comprising the sequence motif TATAAT) supports the opening of the transcription bubble. The transcription bubble extends over about 18 base pairs.

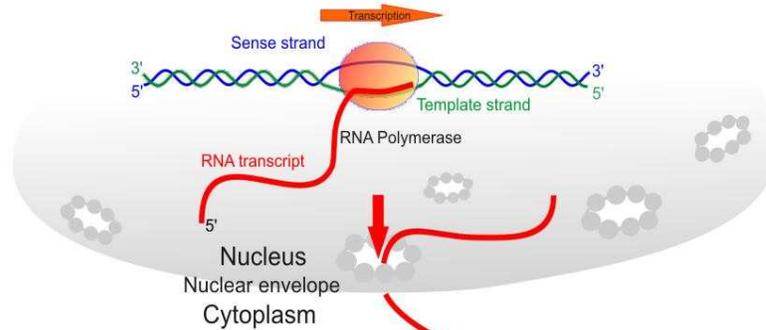
Transcription initiation is followed by the elongation process, in which an RNA-copy of the sense-strand (only the sense-strand encodes the sequence information for the gene product) is transcribed until a terminator sequence at the end of the gene is reached.

During elongation, the holoenzyme slides along the operon from 5' to 3' direction (with respect to sense strand - see Fig. 2.8). The correct nucleotides for the assembly of the mRNA strand are recognized by complementary base pairing with the coding strand. RNA polymerase joins these nucleotides with the growing RNA strand. A proofreading mechanism replaces incorrectly added nucleotides.

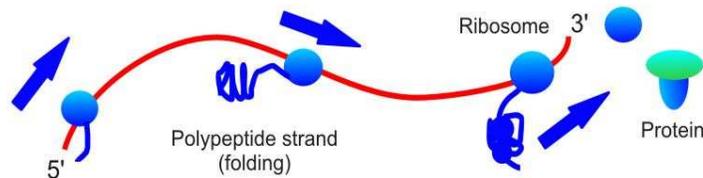
---

<sup>4</sup> These don't serve as templates for the synthesis of polypeptide strands but rather constitute a crucial part of the cells molecular machinery or, like miRNA, are involved in the regulation of the expression of other genes.

## A Transcription



## B Translation



**Figure 2.8:** Transcription (A) and translation (B). RNA polymerase opens a transcription bubble and produces a RNA copy of the sense strand while sliding in 5' to 3' direction (with respect to the sense strand) until transcription termination is encountered. After poly-adenylation (not shown) the mRNA is released from the nucleus through the nuclear pores. Translation of the mRNA sequence into a polypeptide sequence (B) is performed in the cytoplasm. Ribosomes move along the mRNA in 5' to 3' direction, thereby translating the genetic code into a polypeptide sequence. Proteins emerge from folding of the polypeptide chains.

The transcription ends when a *terminator sequence* is encountered. Then the transcription complex comes apart, the transcription bubble collapses and the mRNA strand is released. Still in the nucleus the (eucaryotic) mRNA undergoes poly-adenylation (addition of a poly-A-tail at the 3'-end). By binding the *poly(A)-binding protein* (PABP) the poly-A-tail protects the mRNA from degradation and increases the translation of the mRNA. The poly-A sequence is technically employed for the specific extraction of mRNA sequences with poly-T functionalized magnetic beads.

## Translation

Messenger RNA (mRNA) is used as a template for the synthesis of proteins. Single stranded RNAs similar like polypeptide chains can fold and have the capability to form complex tertiary structures, similar as proteins. Ribosomal RNA (rRNA), the most abundant RNA in cells, is not a simple information carrier like DNA, but rather folds itself into a complex "nanomachine" which is crucial for the synthesis of polypeptide chains.

Like tiny robots ribosomes slide along the mRNA strands (downstream from the 5'- to

the 3'-end) and translate the nucleic acid sequences via the genetic code into polypeptide sequences (Fig. 2.8). The molecular recognition of the codons (base triplets encoding for amino acids) is performed with transfer RNA (tRNA), another functional RNA structure (Fig. 2.10). The anticodon, an exposed base triplet at the end of the anticodon arm of the tRNA, can specifically bind via base pairing<sup>5</sup> to a complementary codon sequence on the mRNA strand. Upon binding the corresponding amino acid which was carried by the tRNA to the site of polypeptide synthesis is attached to the growing polypeptide chain. Subsequently the ribosome moves on to the next codon and simultaneously releases the discharged tRNA.

Translation of an mRNA strand is performed by many ribosomes simultaneously. While the translation process is going on the mRNA strand is degraded by nucleases in the 5' → 3' direction.

### 2.2.5 Expression Regulation

The functions of a cell (e.g. expression of structural and regulatory proteins, differentiation, control of the life cycle, adaption to environmental influences) are largely controlled by gene regulatory networks. Transcription factor proteins (via specific protein-DNA binding) can activate, amplify or inhibit the translation of the targeted gene(s) and thus control the corresponding gene activity.

Post-transcriptional regulatory mechanisms include alternative splicing, RNA silencing, antisense suppression, and the regulation of mRNA stability.

DNA microarrays enable simultaneous investigation of the activity of many genes, on a genome-wide scale. *Gene expression profiles* (which are encoding the complex interactions between genes) are an important tool for the investigation gene regulatory pathways (→ functional genomics). Expression profiling has also emerged as a promising diagnostic tool for identification of cancer types or subtypes, thus enabling a well-directed therapeutic response.

#### Expression regulation at the transcription level

The most prominent regulation mechanism is transcription initiation. In procaryotes essentially only the holoenzyme RNA-polymerase (composed of several subunits) is directly involved in the transcription process. In eucaryotes a large machinery of proteins (including several holoenzymes) needs to form an initiation complex before the transcription can commence.

---

<sup>5</sup> Frequently anticodons contain the relatively unspecifically binding nucleotides inosine or pseudouridine. Unspecific binding accounts for the degeneracy of the genetic code.

In the simple procaryotic organisms (e.g. bacteria) the initiation of transcription is regulated by *activators* and *repressors*. This shall be explained in the following on the example of the regulation of the lactose genes of the bacterium *E. Coli*, which has been investigated by Jacob and Monod [Jac61].

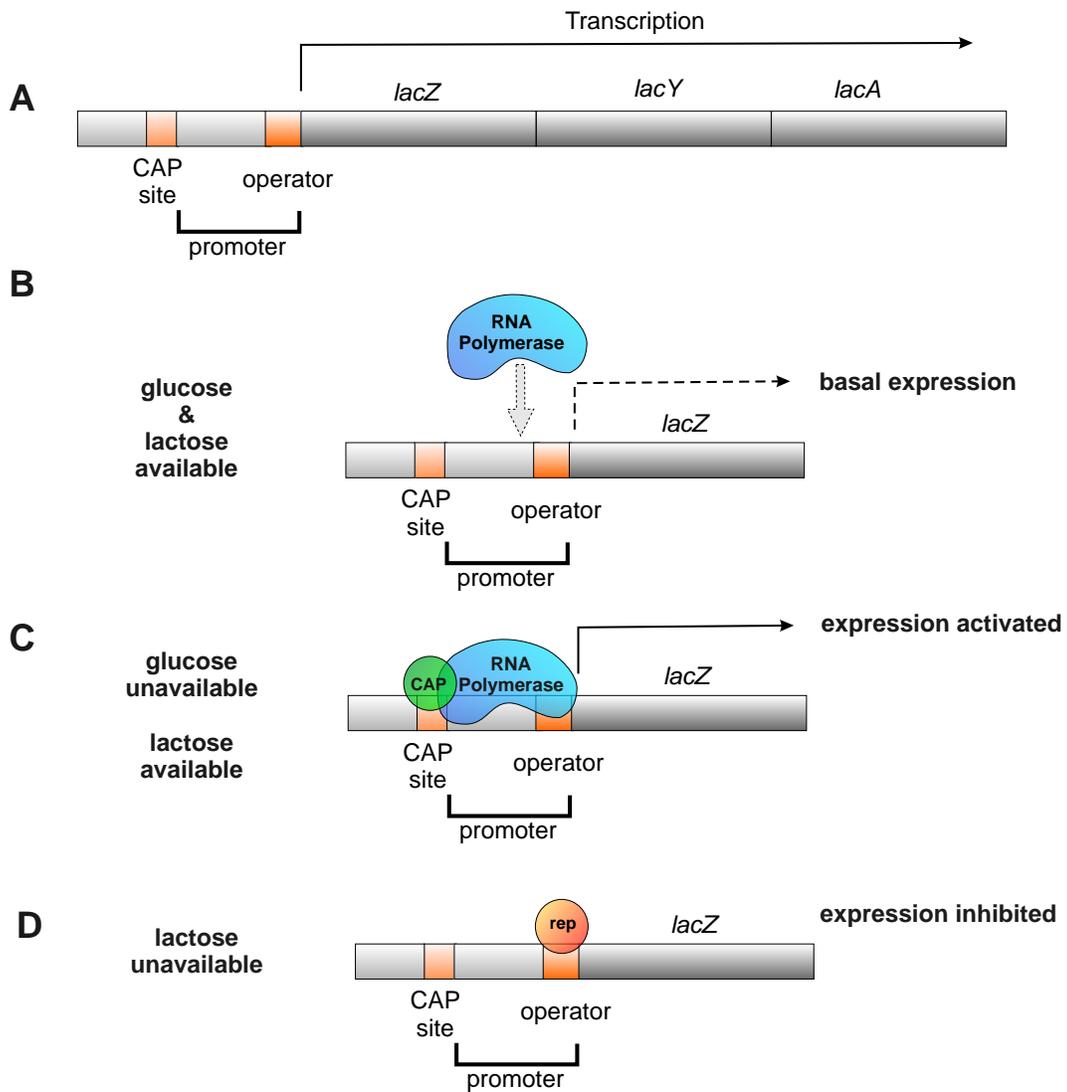
*E. Coli* can digest both food sources - glucose and lactose. To conserve resources the lactose metabolism is only activated if only lactose and no glucose is available. In case both sugars are available *E. Coli* gives preference to glucose since it is the more efficient source of energy. Only if the glucose is depleted and lactose is still present in the medium, *E. Coli* begins to express the gene for the protein  $\beta$ -galactosidase, an enzyme which is required for the digestion of lactose.

The gene for  $\beta$ -galactosidase *lacZ* is combined with two further genes, *lacY* and *lacA* (auxiliary genes, also required for lactose digestion) in a functional unit called *operon* (Fig. 2.9A). The operon is typical for procaryotic organisms. Apart from the coding sequences for the protein(s) the operon contains the promoter sequence. This is recognized by RNA-polymerase and enables binding of the RNA-polymerase to the double-stranded DNA. The promoter contains the Pribnow-box with the sequence motif TATAAT (typical for procaryotes), and the so-called operator. In the *lac* operon the operator is a binding site for a repressor protein. The repressor protein, when bound to the operator site, prevents RNA polymerase from binding to the promoter site (Fig. 2.9D).

Another sequence motif, adjacent to the promotor, serves as specific binding site for the activator protein CAP, which supports the binding of RNA-polymerase to the promoter site (Fig. 2.9C).

The function of the regulatory proteins (activator and repressor) is controlled by the abundance of glucose and lactose, respectively. The activator CAP (a receptor for cyclic AMP) can only bind to CAP binding site (protein-DNA interaction) upon binding to cyclic AMP, which is abundant in the absence of glucose. The *lac* repressor protein can only bind to the operator site if lactose is not available, since the binding affinity of the repressor protein to DNA is significantly decreased by a conformational change induced from the presence of allolactose.

- Glucose and lactose available: In the presence glucose the activator cannot bind to the CAP site. Since the repressor can neither bind, the expression can occur at a low basal level (Fig. 2.9B).
- Lactose available/glucose unavailable: The activator (CAP) can only bind near the promoter site if glucose is not available (Fig. 2.9C). In the presence of lactose only, the activator increases the *lacZ* expression by a factor of about 40 compared to the basal level [Pta02].



**Figure 2.9:** The *lac* operon (A) and the *lac* expression (B-D). The *lac* operon comprises the CAP activator site, the promoter and the genes *lacZ*, *lacY* and *lacA*. The latter are transcribed as a single mRNA. The expression level of the *lac* genes is controlled by the abundance of glucose and lactose, respectively. Activator and repressor proteins which can bind to specific binding sites (protein-DNA interaction), control the binding RNA polymerase. See text for details. Figures were adapted from [Pta02].

- Lactose unavailable: The *lac* repressor can only bind to the operator site if lactose is not available. In this case the repressor is bound to the operator site preventing the binding of the RNA-polymerase, no matter if the activator is bound to the CAP site (Fig. 2.9D). The expression of the *lac* genes is inhibited.

## 2.2.6 Biological Functions of RNA



**Figure 2.10:** Structure of phenylalanine transfer RNA (visualization of 4TNA.pdb [Hin78] with UCSF Chimera). Transfer-RNA is employed in the translation process as a sequence specific vehicle for amino-acids. The anticodon-arm (near the lower edge of the image) contains a unit of 3 nucleotides corresponding to a codon on the mRNA strand. The amino-acid (not shown here) is attached to the acceptor stem (upper right end) with the characteristic CCA 3'-terminal group.

**The biological function of RNA is more versatile than that of DNA:**

- In the process of gene expression messenger RNA (mRNA) is employed as a template for polypeptide synthesis. RNA, unlike DNA, is a volatile information carrier with a rather limited lifetime.
- Micro-RNAs (miRNA) have regulatory functions. Via the RNA interference (RNAi) mechanism they can specifically inhibit the expression of the corresponding target genes.
- Antisense-RNAs (aRNA) have regulatory functions. An aRNA sequence is produced if the noncoding (antisense)-strand of a gene sequence is also being transcribed. Thus the

aRNA is complementary to the mRNA of the particular gene. By base-pairing between the complementary RNA strands the translation of the corresponding polypeptide-sequence is inhibited. In the transgenic *Flavr Savr<sup>TM</sup>* tomato antisense RNA is employed to suppress the expression of an enzyme involved in ethylene production. The significant reduction of ethylene delays the ripening and spoiling of the tomato.

- RNA sequences, similar as polypeptide chains, can fold into complex secondary and tertiary structures. Ribosomal RNA and transfer RNAs (tRNA) are essential parts of the translation machinery (see section 2.2.4).

## 2.3 Nucleic Acid Hybridization

Two complementary (or partially complementary) nucleic acid strands  $S_1$  and  $S_2$  can bind via base pairing and form a stable nucleic acid duplex  $D$ . The double-helical duplex structure is stabilized by hydrogen bonding and base stacking interactions.



The formation of nucleic acid duplexes is commonly called *hybridization* since usually nucleic acid strands from different sources (e.g. DNA probes and RNA targets) are involved. Owing to the non-covalent character of the stabilizing interactions nucleic acid hybridization is reversible: In thermodynamic equilibrium the duplex formation is balanced by duplex dissociation (also called duplex denaturation or melting). Lower temperatures and increased ionic strengths (up to 1 M  $[\text{Na}^+]$ ) favor duplex formation. With increasing temperature or reduced ionic strength of the hybridization buffer solution the duplexes are increasingly destabilized. Depending on the particular duplex sequence, nucleic acid duplexes can have a very distinct melting transition. Owing to the cooperative character of the duplex binding, the fraction of melted duplexes can change from close to 0% to 100% within a temperature range of a few Kelvins.<sup>6</sup>

Only a small fraction of the duplexes is in a partially denatured intermediate state. Therefore, the hybridization/melting transition is frequently described as a two-state transition. An important characteristic of the nucleic acid hybridization is its outstanding sequence specificity. Already a single mismatched base within an oligonucleotide duplex can result in a significantly reduced binding affinity. Molecular recognition by nucleic acid hybridization is employed by nature (e.g. in RNA interference) and by various molecular biology applications:

<sup>6</sup> For oligonucleotide duplexes the width of the melting transition is decreasing with increasing Gibbs free energy of the duplex, thus with increasing duplex length.

- DNA microarrays
- Fluorescent *in situ* hybridization (FISH): sequence specific labeling of mRNA sequences within cells.
- Primer sequences are used as starting points for nucleic acid replication (e.g. in PCR or dideoxy sequencing). For this purpose the primers are hybridized to the template strands.
- Molecular beacon probes: this type of hairpin-shaped nucleic acid probe containing a fluorophore-quencher-pair becomes fluorescent upon hybridization with a complementary target sequence.
- Antisense RNA sequences (sequence-specific silencing of mRNA transcripts)
- RNA interference (sequence-specific silencing of mRNA transcripts)

### 2.3.1 Kinetics of Nucleic Acid Hybridization

The widely used *two-state model* of nucleic acid hybridization assumes that the single stranded species  $S_1$  and  $S_2$  are in equilibrium with the duplexes  $D$ .



Equation 2.2 doesn't describe elementary base pairing processes and is therefore valid only if there are no significantly populated intermediate states. The two-state model is a reasonable approximation, for example, for short linear duplexes. The zipper model of DNA duplex melting transition, which considers individual base pairing and base pair dissociation events, is described in section 2.3.3.

In the following, for simplicity's sake, we assume that are duplexes are not self-complementary and that folding of single stranded species (intrastrand base pairing) can be neglected.

Duplex formation is a second order reaction, whereas the denaturation is a first order reaction.

$$\frac{d[D]}{dt} = -k_-[D] + k_+[S_1][S_2] \quad (2.3)$$

In equilibrium (with  $d[D]/dt = 0$ ) we obtain the equilibrium constant  $K$  (as described by the law of mass action).

$$K = \frac{k_+}{k_-} = \frac{[D]}{[S_1][S_2]} = \frac{[D]}{([S_1]_0 - [D]) \cdot ([S_2]_0 - [D])} \quad (2.4)$$

The Gibbs free energy of duplex formation  $\Delta G_D^\circ$  ( $^\circ$  is referring to standard conditions) is related to the equilibrium constant  $K$  by

$$\Delta G_D^\circ = -R \cdot T \cdot \ln K. \quad (2.5)$$

If the complete temperature dependence of the binding affinity - e.g. from experimentally determined plots of  $1/T_m$  versus  $\ln(C_T/4)$  - is known, the Gibbs free energy  $\Delta G_D^\circ$  can be determined via the van't Hoff equation:

$$\frac{1}{T_m} = \frac{R}{\Delta H_D^\circ} \cdot \ln\left(\frac{C_T}{4}\right) + \frac{\Delta S_D^\circ}{\Delta H_D^\circ} \quad (2.6)$$

$T_m$  is the melting temperature of the duplex - the temperature at which per definition (in thermodynamic equilibrium) 50% of the duplexes are dissociated.  $C_T$  is the total concentration of nucleic acid strands.

From the total enthalpy  $\Delta H_D^\circ$  and entropy changes  $\Delta S_D^\circ$  the Gibbs free energy change  $\Delta G_D^\circ$  of the melting transition can be obtained with

$$\Delta G_D^\circ = \Delta H_D^\circ - T \cdot \Delta S_D^\circ. \quad (2.7)$$

Alternatively  $\Delta H_D^\circ$  and  $\Delta S_D^\circ$  can be predicted from sequence-dependent *nearest-neighbor* thermodynamic parameters (see section 2.3.2).

### Fraction of hybridized duplexes

The fraction of hybridized oligonucleotides  $Fb$  [Koe05] (fraction bound) is a quantity which is directly accessible from experiments (e.g. via the hybridization signal intensity in microarray assays or via the hypochromicity in UV-absorption-based measurements).  $Fb$  can be derived from thermodynamic quantities (e.g. via the equilibrium constant  $K$ ).

$$Fb = \frac{[D]}{\min([S_1]_0, [S_2]_0)} \quad (2.8)$$

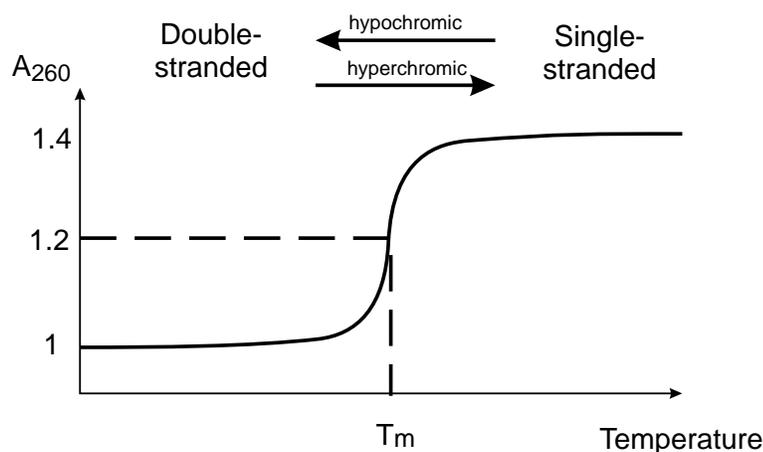
$[S_1]_0$  and  $[S_2]_0$  are the initial concentrations of single-stranded species  $S_1$  and  $S_2$ . How  $T_m$ ,  $\Delta G_D^\circ$  and  $Fb$  are related and influenced by experimental parameters (duplex length, sequence composition, defects, salt concentration, nucleic acid concentration and temperature) is well discussed in [Koe05].

If the fraction bound  $Fb$  is compared to microarray hybridization signals one needs to consider that microarray hybridization is affected by many parameters, which are not accounted for in the simple model described above.

## The duplex melting temperature $T_m$

The melting temperature is defined as the temperature  $T_m$  at the midpoint of the melting transition (in thermodynamic equilibrium) at which 50% of the complementary molecules are either hybridized or dissociated.

For practical applications (e.g. PCR primers and microarray probes) the duplex melting temperature is the most important thermodynamic parameter.



**Figure 2.11:** Melting transition. Duplex melting (denaturation) results in an increase (between 20-40%) of the UV absorbance  $A_{260}$  (hyperchromicity). Vice versa, duplex annealing (renaturation) is accompanied by a decrease of the absorbance (hypochromicity). The melting temperature  $T_m$  is defined by the midpoint of the melting transition.

The standard method for investigation of the nucleic acid melting transition is the measurement of the UV absorbance [App65]. Nucleic acid duplexes due to increased base stacking (with respect to single strands) have a reduced UV absorbance (hypochromicity) at a wavelength of about 260 nm. The melting transition can be observed as an increase in UV absorbance  $A_{260}$  by about 20-40%. One should have in mind that the UV absorbance is related to the fraction of unstacked bases, and not necessarily to the fraction of dissociated duplexes. However, under the assumption of a two-state melting transition (no significant population of partially denatured duplex states), the fraction of melted base pairs is equivalent to the fraction of melted duplexes [Owc05].

A reasonable working approximation for the melting temperature  $T_m$  (applicable for short oligonucleotide duplexes with a length between 5 and 20 bp) is provided by the *Wallace rule*:

$$T_m = 2(n_A + n_T) + 4(n_G + n_C) \quad (2.9)$$

Wallace established the above equation (which provides the melting temperature  $T_m$  in °C) for short ( $l < 18$  bp) membrane-bound oligonucleotide duplexes at 0.9 M  $[\text{Na}^+]$  concen-

tration [Wal79].  $n_i$  are the numbers of the corresponding bases contained in the oligo-sequence.

A more accurate empirical formula for the melting temperature of longer duplexes ( $l > 50$  bp) has been established by Wetmur *et al.* [Wet91]: Like equation 2.9 this equation (2.10-2.12) is based on the GC-content. However, it also considers the salt concentration, the length  $L$  of the oligonucleotides and the increased stability of DNA/RNA and RNA/RNA duplexes.

$$T_m = 81.5 + 16.6 \log ([\text{Na}^+]) + 0.41(\%GC) - 500/L \quad \text{DNA/DNA} \quad (2.10)$$

$$T_m = 78 + 16.6 \log ([\text{Na}^+]) + 0.7(\%GC) - 500/L \quad \text{RNA/RNA} \quad (2.11)$$

$$T_m = 67 + 16.6 \log ([\text{Na}^+]) + 0.8(\%GC) - 500/L \quad \text{DNA/RNA} \quad (2.12)$$

$T_m$ : melting temperature in °C

$L$ : length of the complementary region in bp

$[\text{Na}^+]$ : sodium ion concentration in mol/l

%GC: percentage of GC base pairs

The melting temperature, as will be shown in section 2.3.2, can also be calculated in a thermodynamics approach on the basis of the *nearest-neighbor model*.

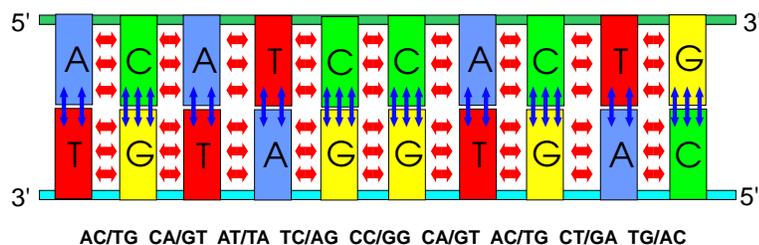
### 2.3.2 The Nearest-Neighbor Model

Nucleic acid duplexes are stabilized by hydrogen bonding and by base stacking interactions between adjacent base pairs (Fig. 2.12). Therefore, nucleic acid duplex stability is not just determined by the base composition (as might be inferred from a stabilization by hydrogen bonds alone), but, considering the stacking interactions, also by from the base sequence [Cro64; Tin73; Bre86; Fre86]. "The stability of the DNA duplex appears to depend primarily on the identity of the nearest-neighbor bases" [Bre86].

Literature describes two different (though equivalent) computational formats for the nearest neighbor model [Owc97]: In the *singlet format*, focusing on individual base pairs (with nearest-neighbor corrections), hydrogen bonding and base stacking interactions are considered separately. In the *doublet format* [Got81] ("doublet" refers to base pair doublets as shown in Fig. 2.12) hydrogen bonding and stacking interactions are combined into a single NN free energy parameter.

In the following we will refer to the doublet format which has been employed throughout this work.

The free energy change  $\Delta G_{37}^{\circ}(\text{total})$  for duplex formation (at a temperature of 37°C) is



**Figure 2.12:** The nearest neighbor model includes hydrogen bonding (inter-strand, blue arrows) and base stacking interactions (intra-strand, red arrows). AT base pairs are stabilized by two hydrogen bonds, CG base pairs by three. The base stacking interaction between adjacent base pairs depends on the identity of the nearest neighbor pair (doublet), e.g. AC/TG (5'-AC-3' paired with 3'-TG-5'). DNA duplex structures thermodynamically can be considered to be the sum of their nearest-neighbor pairwise interactions [Bre86].

calculated with eqn. 2.13:

$$\Delta G_{37}^{\circ}(\text{total}) = \Delta G_{37}^{\circ} \text{init} + \sum \Delta G_{37}^{\circ} \text{NN} + \Delta G_{37}^{\circ} \text{sym} + \Delta G_{37}^{\circ} \text{AT term} \quad (2.13)$$

Helix initiation is considered by the helix initiation free energy  $\Delta G_{37}^{\circ} \text{init}$ . The formation of subsequent base pairs is accounted by summation of the helix propagation free energy parameters (NN free energy parameters)  $\Delta G_{37}^{\circ} \text{NN}$  (see Tab. 2.3.2). For each A·T terminal base pair the term  $\Delta G_{37}^{\circ} \text{AT term}$  is added. Only in case of self-complementary sequences the symmetry correction term  $G_{37}^{\circ} \text{sym}$  is added.

For the duplex shown in Fig. 2.12 equation 2.13 provides

$$\begin{aligned} \Delta G_{37}^{\circ}(\text{total}) &= \\ &= \Delta G_{37}^{\circ} \text{init} + \Delta G_{37}^{\circ} \text{AC/TG} + \Delta G_{37}^{\circ} \text{CA/GT} + \dots + \Delta G_{37}^{\circ} \text{TG/AC} + \Delta G_{37}^{\circ} \text{AT term} \\ &= (1.96 - 1.44 - 1.45 - 0.88 - 1.30 - 1.84 - 1.45 - 1.44 - 1.28 - 1.45 + 0.05) \frac{\text{kcal}}{\text{mol}} \\ &= -10.52 \frac{\text{kcal}}{\text{mol}} \end{aligned}$$

In an analogous way the enthalpy  $\Delta H^{\circ}(\text{total})$  and entropy change  $\Delta S^{\circ}(\text{total})$  for duplex formation can be calculated from unified NN parameters in Table 2.3.2. The Gibbs free energy  $\Delta G^{\circ}$  at temperature T can be determined from tabulated values  $\Delta H^{\circ}$  and  $\Delta S^{\circ}$  with

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}. \quad (2.14)$$

Internal single base mismatches can be accounted for by using MM nearest neighbor parameters established by Allawi *et al.* [All97] (see section 2.4.1).

**Table 2.1: Unified nearest neighbor thermodynamic parameters for DNA Watson-Crick pairs in 1 M NaCl** (adapted from [San04]). The "unified NN parameter set" has been reported by SantaLucia [San98]: SantaLucia found a remarkable agreement between several previous studies on NN thermodynamics of DNA polymers and oligomers. The notation for the NN-pair (base pair doublet) GT/CA (taken as an example) refers to the dinucleotide sequence 5'-GT-3' which is paired with the complementary dimer sequence 3'-CA-5'. The initiation term accounts for the duplex initiation free energy. The symmetry correction has to be considered in case of self-complementary duplexes. For each AT-terminal base pair a penalty term has to be added. Gibbs free energy parameters  $\Delta G_{37}^{\circ}$  are provided for a temperature of 37°C.

NN-pair	$\Delta H^{\circ}$ kcal/mol	$\Delta S^{\circ}$ cal/(K·mol)	$\Delta G_{37}^{\circ}$ kcal/mol
AA/TT	-7.6	-21.3	-1.00
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.58
CA/GT	-8.5	-22.7	-1.45
GT/CA	-8.4	-22.4	-1.44
CT/GA	-7.8	-21.0	-1.28
GA/CT	-8.2	-22.2	-1.30
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.84
Initiation	+0.2	-5.7	+1.96
Terminal AT penalty	+2.2	+6.9	+0.05
Symmetry correction	0.0	-1.4	+0.43

### Melting temperature prediction with the two-state nearest neighbor model

With  $\Delta H^{\circ}(\text{total})$  and  $\Delta S^{\circ}(\text{total})$  we can determine the two-state melting temperature  $T_m$  according to SantaLucia *et al.* [San04] (melting temperature  $T_m$  in °C; ideal gas constant  $R=1.9872$  cal/(K·mol);  $\Delta H^{\circ}$  in kcal/mol;  $\Delta S^{\circ}$  in entropical units (e.u.); total DNA concentration  $C_T$  in mol):

$$T_m = \Delta H^{\circ} \times 1000 / (\Delta S^{\circ} + R \cdot \ln(C_T/4)) - 273.15 \quad (2.15)$$

The above equation for the melting temperature (derived from van't Hoffs equation) is valid for a sodium concentration  $[\text{Na}^+]$  of 1 mol/l and only in the case that the concentrations of complementary strands are equal. Smaller concentrations of  $[\text{Na}^+]$  ions result in reduced screening of the negatively charged phosphate groups. This leads to an increased repulsion of the polyanionic strands and hence results in a reduced duplex stability. An increase of the  $[\text{Na}^+]$  concentration above 1 mol/l doesn't result in a significant increase

of the duplex stability since the electrostatic screening can not be enhanced significantly at higher salt concentrations. Various (empirically derived) salt corrections for the melting temperatures of DNA duplexes like the Schildkraut-Lifson equation (equation 2.16: the uncorrected melting temperature  $T_{m1}$  at the sodium concentration  $[Na^+]_1$  is related to the corrected melting temperature  $T_{m2}$  at the sodium concentration  $[Na^+]_2$ ) are discussed by Owczarzy *et al.* [Owc04].

$$T_{m2}([Na^+]_2) = T_{m1}([Na^+]_1) + 16.6 \log([Na^+]_2/[Na^+]_1) \quad (2.16)$$

According to [San04] the melting temperature prediction with the two-state nearest neighbor model is highly reliable: for a set of 264 sequences (with lengths ranging from 4 to 16 bp) the standard deviation between experimental and predicted melting temperatures was found to be 2.3°C.

Several web-servers for melting temperature calculation (e.g. the DINAMelt server [Mar05]) employ the two-state NN-model or more advanced multi-state models for calculation of duplex melting temperatures and further thermodynamic parameters.

### Positional-dependent nearest neighbor model (PDNN)

Zhang *et al.* [Zha03; Zha07] proposed a positional-dependent nearest neighbor model. In the PDNN model the binding free energy  $\Delta G$  of the  $(n+1)$ mer oligonucleotide duplex is expressed as the weighted sum of all  $n$  nearest neighbor interactions  $\varepsilon(b_k, b_{k+1})$ .

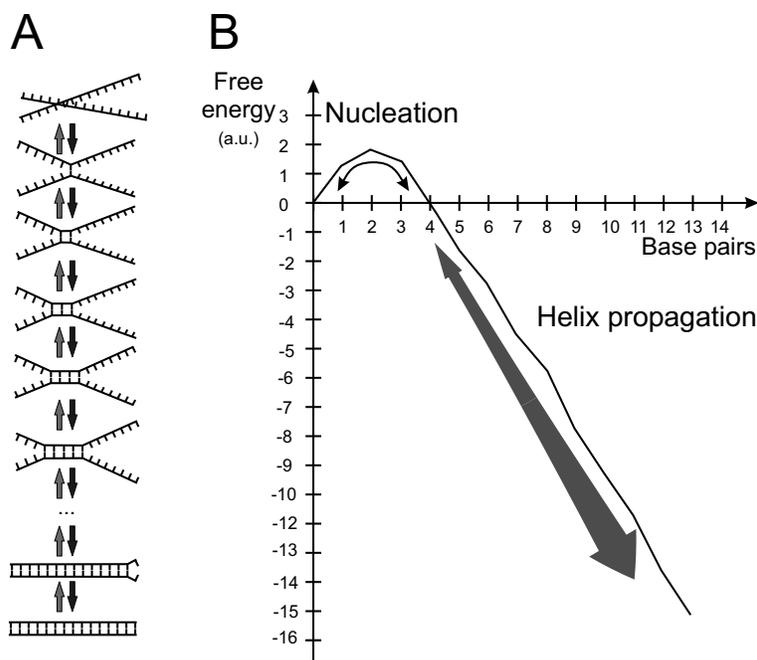
$$\Delta G = \sum_{k=1}^n \omega_k \varepsilon(b_k, b_{k+1}) \quad (2.17)$$

The fitted weight-parameters  $\omega_k$  (determined from fitting microarray hybridization signal data to the expected hybridization signal, which is derived from the model equation 2.17) indicate that the duplex ends contribute less to duplex stability than the center of the duplex, possibly owing to partial unzipping of the duplex ends (*end fraying*).

### 2.3.3 Zipper-Model of the Oligonucleotide Duplex

According to Wetmur and Davidson [Wet68] duplex formation "[...] involves the joining of short, homologous sites on the two strands followed by a fast, reversible zippering reaction [...]". Duplex formation comprises the rate limiting nucleation step (formation of an intermediate duplex, two to three base pairs in length - the nucleation rate depends on the concentrations of the two complementary species) and, if the two strands are complementary, fast helix growth (*helix propagation*) by sequential formation of base pairs [Cra71]

(see Fig. 2.13). The zippering is reversible, so that the duplex can denature by sequential unzipping of the individual base pairs. However the probability for complete strand separation is reduced with increasing duplex length.



**Figure 2.13:** Nucleic acid duplex formation comprises nucleation and rapid helix propagation. (A) Zippering of the duplex by sequential base pair formation. The zippering is reversible, however under hybridization conditions base pairing is favorable over base pair dissociation. (B) Duplex free energy versus duplex length (the sketch is only for demonstration of the principle, and does not directly refer to experimental or theoretical data). The nucleation of the first two base pairs is thermodynamically not favorable (positive free energy) since stacking interactions are stabilizing the base pairs only towards one side (with only one half NN interaction per base pair). Beyond the second base pair every additional (Watson-Crick) base pair increases the duplex stability by adding one full NN-interaction and also due to the cooperative character of the interactions. Provided the strands are complementary the nucleation is followed by rapid helix propagation (zippering). The sketch in (B) was adapted from [Por77].

### The double-ended zipper model

The statistical mechanics of a simple double-ended zipper model of the nucleic acid duplex has been first investigated by Gibbs and DiMarzio [Gib59] in 1959. Their model (which has originally been developed to describe the (polypeptide) alpha-helix to random coil transition), like the Zimm-Bragg model, is basically a linear Ising model and is thus not able to describe a true first order phase transition. However, the sharpness of the transition is increasing with the length of the duplex (number of base pairs) and with the number of

conformations an open link can adopt [Gib59].

In C. Kittels double-ended zipper model [Kit69] the zipper is consisting of  $N$  bonds (corresponding to the base pairs) that can only be opened from the ends. The partition function is determined by summation over the statistical weights of all partially unzipped duplex states. With the partition function the statistical mechanics of the duplex, e.g. the average number of open bonds (corresponding to the degree of partial duplex denaturation) is accessible. Kittel showed that the assumed degeneracy of partially unzipped duplex states (arising from rotational freedom of unpaired nucleotides) - in DNA this degeneracy may be on the order of  $10^4$  - gives rise to a melting transition in the quasi-one-dimensional system.<sup>7</sup> No phase transition can occur in the non-degenerate case (when the number of rotational degrees of freedom equals 1).

Zocchi *et al.* [Zoc03] reported that a zipper-model based on end-domain opening describes well the temperature dependence of the average number of unzipped base pairs determined in UV absorption experiments. However, they also report that their analysis of transition parameters indicates that, apart from end-domain opening, bubble formation is also important for the denaturation process.

Deutsch *et al.* [Deu04] employed the double-ended zipper model for a statistical mechanics based description of microarray hybridization signals.

End-unzipping of the duplex has also been assumed by Ambjörnsson and Metzler [Amb05] for a model to investigate the blinking dynamics of molecular beacons (fluorophore-quencher pair included in a fraying duplex section).

Base pairs at the duplex ends are stabilized by stacking interaction with only one neighboring base pair, whereas base pairs in the interior of the duplex are stacked between two neighboring base pairs. The stabilizing stacking interactions from both sides prevent internal denaturation. Therefore unzipping is (largely) restricted to the duplex ends (*end fraying*) as shown in Fig. 2.14A. Structural constraints arising from the double helix structure may impose further restrictions to internal bubble formation. The influence of the helical structure on duplex stability is, however, not well understood.

## Denaturation bubbles

The above statements, however, do not apply to the denaturation of long duplexes. These denature via the formation of denaturation bubbles in the interior of the duplex (see Fig. 2.14B and C). This is due to several reasons:

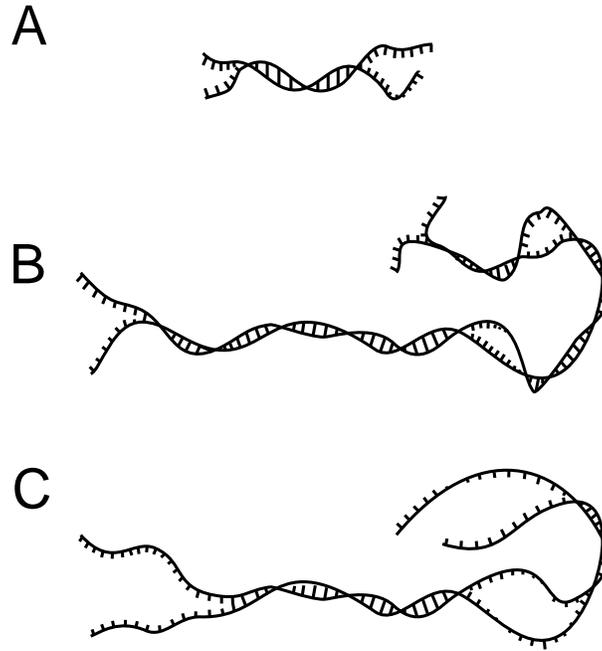
- due to an exponential decrease of the base pair dissociation probability towards the

---

<sup>7</sup> Cuesta and Sanchez [Cue04] discuss why Van Hove's theorem (simply interpreted: "No phase transitions occur in 1D particle systems with short-range pair interactions") doesn't apply to the melting transition of nucleic acid duplexes.

center of the duplex, end-domain opening is restricted to duplex-ends  $\Rightarrow$  thus, long duplexes can only denature via the formation of denaturation bubbles.

- occurrence of relatively weakly bound (AT-rich) subsequences in a long duplex
- increased melting temperatures of long duplexes  $\Rightarrow$  the increased entropy contribution ( $-T\Delta S$ ) results in destabilization of the nearest neighbor interactions  $\Delta G_{NN}$



**Figure 2.14:** Denaturation of short duplexes (A) occurs mainly via end-domain opening. In long duplexes (B) end-domain opening doesn't extend into the middle of the duplex. Rather, denaturation bubbles, forming at weakly bound sections in the interior of the duplexes, propagate and (C) merge with the open end-regions. At increased temperatures denaturation bubble formation leads to dissociation of long duplexes.

The relevance of internal denaturation bubble formation depends on duplex length and, in particular, on the individual sequences (i.e. on the distribution of more/less stable NN pairs). To provide a coarse estimate: for duplexes with  $l < 15$  base pairs end-fraying is expected to be the prevailing mode of nucleic acid denaturation, vice versa, for long and intermediate size duplexes with  $l > 100$  base pairs bubble formation is expected to be relevant or more important than end-domain opening [Blo03].<sup>8</sup> However, Zocchi *et al.* [Zoc03] reported that denaturation bubbles may be relevant also in the denaturation process of short duplexes.

<sup>8</sup> Blossey *et al.* [Blo03]: "On rather short DNA sequences ( $\sim 100$  bp's) the loop entropy contribution is not very important as loops are rare and short and the DNA denatures mainly through unbinding from the edges. A description based on the 1D Ising model with appropriate experimentally determined energy parameters is therefore sufficient [...]."

### 2.3.4 Further Models of the DNA Melting Transition

Further well-established models for the DNA melting transition are the Poland-Scheraga (PS) model [Pol66] and the Peyrard-Bishop-Dauxois (PBD) model [Dau93].

The Poland-Scheraga model describes the helix-coil transition in long polynucleotide duplexes. The duplex comprises alternating double-helical segments and denaturation bubbles. The PS model is essentially a one-dimensional Ising-model. Consideration of the various bubble configurations gives rise to an entropic term. This results in an effective long range interaction, so that in the PS model a phase transition may occur [Blo03].

The PBD model represents a Hamiltonian approach. In the PBD model cooperativity effects - arising from anharmonic nearest neighbor stacking interactions - result in a distinct melting transition.

An overview on theoretical models of the nucleic acid melting transition is provided with reference [Zho06].

Further reading on the DNA melting transition:

- thermal denaturation of DNA [War85] and DNA oligomers [Zoc03]
- DNA breathing dynamics [Amb06]
- zipper models [Kit69; Iva04]
- end-denaturation [Amb05]
- mismatches and bubbles [Zen06]
- thermodynamic properties of DNA sequences [Koe05]
- further related publications [vE06; Eve07]

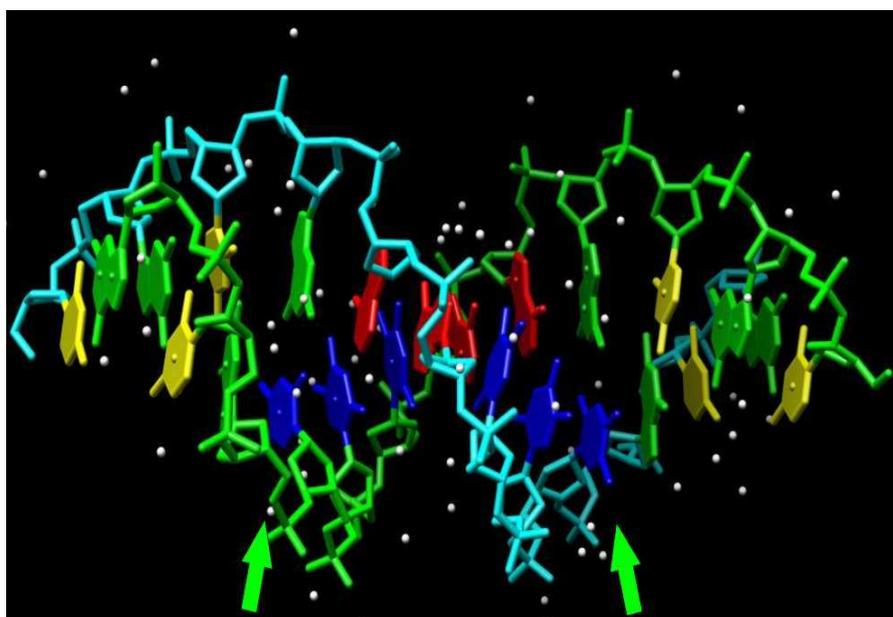
## 2.4 Destabilization of Oligonucleotide Duplexes by Point Defects

A high discrimination capability between similar sequences is important in genotyping applications, where single nucleotide polymorphisms (SNPs), variations of single bases, are the subject of interest. SNPs largely determine genetic individuality, but also disposition to genetically caused diseases or response to medicaments, and are therefore of great interest not only for genetic research but also for medical diagnostics and therapy. SNPs can be detected (using DNA microarrays) by hybridization with short oligonucleotide probes. Already a single mismatching (MM) base pair (owing to the SNP) can result in a significant decrease of duplex stability [Nel81; Pat82; Con83]. The impact of a MM base pair on

duplex binding affinity is determined by the length of the duplex [Koe05], the type of mismatch base pair [All97], the influence of neighboring bases [All97] and by the position of the defect (with respect to the duplex ends) [Wic06; Poz06; Nai06b].

In this study we also investigate single base bulges, another type of point defect, originating from single base insertions and deletions. The insertion of a surplus (unpaired) base into one of the duplex strands results in a small bulge in the regular duplex structure. Similarly a single base deletion creates a bulged base in the opposite strand. Like single base mismatches base bulges can significantly reduce duplex binding affinity.

### 2.4.1 Single Base Mismatches



**Figure 2.15:** Structure of T·G mismatches in a B-DNA duplex (X-ray diffraction data 113D.pdb [Hun87]). Green arrows indicate the T·G mismatches.

Structural investigations (NMR and X-ray studies) have shown that single mismatch base pairs (see Fig. 2.15) mismatches introduce little overall structural distortion on the double helical duplex structure [Hol91; Cog91; Ske93].

#### Consideration of single base mismatches in the nearest neighbor model

The nearest neighbor model has been extended beyond Watson-Crick base pairs to include single base mismatch (MM) defects [All97; San04]. From UV melting experiments Allawi *et al.* [All97] have established a complete database of MM single base MM thermodynamic parameters for DNA/DNA duplexes. The (mostly) destabilizing MM propagation

parameters (a complete table is provided in [San04]) are used for duplex free energy calculations just like the Watson-Crick propagation parameters. Most destabilizing MM nearest-neighbor pairs are AC/TC ( $\Delta G_{37}^{\circ}=1.33$  kcal/mol), TC/AA ( $\Delta G_{37}^{\circ}=1.33$  kcal/mol), TC/AC ( $\Delta G_{37}^{\circ}=1.05$  kcal/mol) and GT/CC ( $\Delta G_{37}^{\circ}=0.98$  kcal/mol). Least destabilizing are GG/CG ( $\Delta G_{37}^{\circ}=-1.11$  kcal/mol) and GT/CG ( $\Delta G_{37}^{\circ}=-0.59$  kcal/mol). An order of DNA/DNA base pair stabilities (based on [All97]) is provided in [San04]:

$$G \cdot C > A \cdot T > G \cdot G > G \cdot T \geq G \cdot A > T \cdot T \geq A \cdot A > T \cdot C \geq A \cdot C \geq C \cdot C$$

The study of Allawi *et al.* [All97] also reveals a strong impact of closing base pairs (the base pairs enclosing the MM base pair) - closing C·G base pairs are more stabilizing than A·T base pairs.

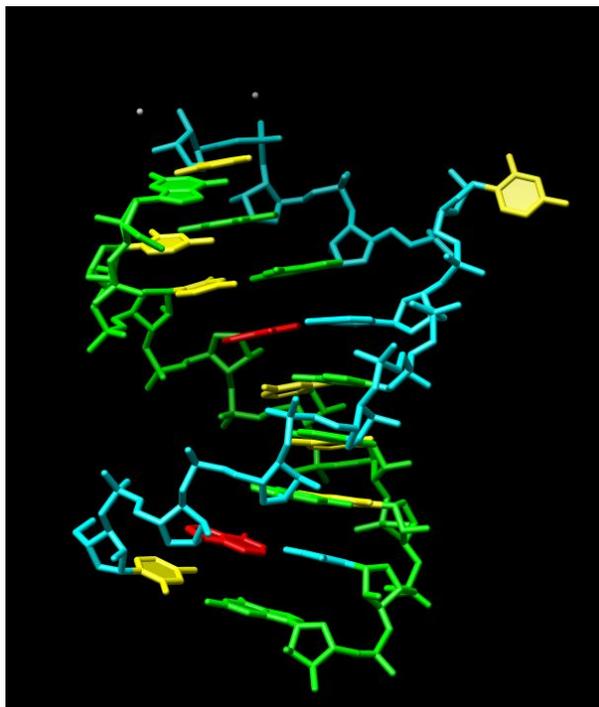
The two-state nearest neighbor model doesn't account for the MM position within the duplex sequence. According to SantaLucia [San04] "[...] with the exception of the terminal and penultimate positions, the thermodynamics of a given mismatch in a given context is independent of its position in a duplex, contrary to common opinion". This, however, is not in agreement with recent observations of a strong influence of defect position on duplex binding affinity [Kie99; Dor03; Wic06; Poz06; Nai06b].

## 2.4.2 Single Base Bulges

Defects originating from insertion or deletion of a base result in *bulged duplexes* as shown in Fig. 2.16. Base bulges are a frequent structural motif in RNA structures e.g. in tRNA and rRNA. It is assumed that bulges may play a role in nucleic acid-protein binding [Wu87]. Single bulged bases can adopt looped out (Fig. 2.16) or intrahelically stacked conformations [Yoo01; Bar06]. According to Woodson and Crothers [Woo88] "[...] evidence from several laboratories suggests that extrahelical purines are generally stacked into the helix, while extrahelical pyrimidines are in equilibrium between stacked and unstacked states [...]" (in this context "extrahelical base" has the meaning "bulged base").

The thermodynamics of bulged duplex was first investigated by Fink and Crothers [Fin72]. They reported a destabilizing free energy (25°C) of 2.8 kcal/mol for a single base bulge. Wartell and coworkers [Ke93; Ke95; Zhu99] investigated the thermodynamics of single base bulges on a larger number of DNA and RNA sequence motifs. The relative stability of bulged RNA duplexes was investigated in temperature gradient gel electrophoresis (TGGE) experiments. For RNA bulges they report an unfavorable free energy (with respect to the bulge-free reference duplex)  $\delta\Delta G_{37}^{\circ}$  between 2.85 and 4.8 kcal/mol.

Wartell and coworkers observed, that the stability of bulged duplexes is increased if the



**Figure 2.16:** Single base bulge (looped out cytosine base, shown in yellow) in a rRNA helix structure (X-ray diffraction data 1DQF.pdb [Sun00]).

bulged base has at least one identical neighboring base. They categorized bulged duplexes - depending on the identity of the bulged base and the duplex sequence - in two groups:

- Group I: the bulged base has no identical neighboring bases
- Group II: the bulged base has at least one identical neighboring base

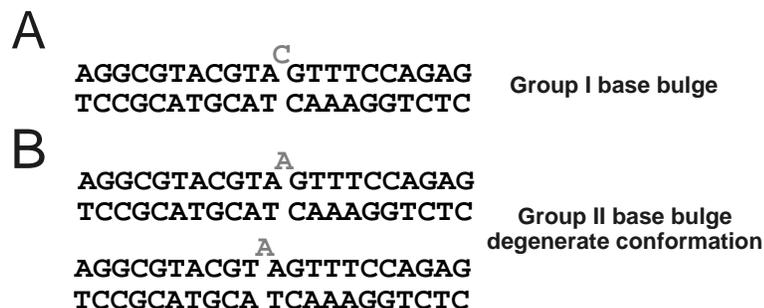
According to [Zhu99] the local average free energy contribution of a DNA base bulge can be expressed as:

$$\Delta G_{37,(XNZ)\cdot(X'-Z')}^{\circ} = 2.72 \text{ kcal/mol} + 0.48\Delta G_{37,(XZ)\cdot(X'Z')}^{\circ} + \delta g \quad (2.18)$$

For the free energy of RNA single bulges a similar relation was derived [Zhu99]:

$$\Delta G_{37,(XNZ)\cdot(X'-Z')}^{\circ} = 3.11 \text{ kcal/mol} + 0.40\Delta G_{37,(XZ)\cdot(X'Z')}^{\circ} + \delta g \quad (2.19)$$

Notation: The unpaired base N is enclosed by the base pairs X·X' and Z·Z'.  $\Delta G_{37,(XZ)\cdot(X'Z')}^{\circ}$  is the stacking energy of the base pair doublet (XZ)·(X'Z'). The stabilizing contribution for degenerate Group II bulges  $\delta g$  is -0.4 kcal/mol for DNA and -0.3 kcal/mol for RNA (in both cases  $\delta g=0$  kcal/mol for Group I bulges).



**Figure 2.17:** Positional degeneracy of base bulges. (A) Group I bulge. The non-degenerate bulged base (C, shown in grey) has no identical neighbor bases. (B) Group II bulge. The bulged base A has an identical neighbor, giving rise to positional degeneracy [Ke95] of the bulge conformation. The increased number of possible bulge conformations (here two rather than only one in A) represents an increase in entropy, resulting in a stabilization of the degenerate Group II bulge with respect to the non-degenerate Group I bulge.

The experimentally observed free energy difference between Group I and degenerate Group II bulges of -0.4 and -0.3 kcal/mol (for DNA and RNA, respectively) is in good agreement with the simplified entropic estimate for a two-position degeneracy of  $-RT \cdot \ln(2) = -0.43$  kcal/mol (at 37°C) [Zhu99].

Znosko *et al.* [Zno02] report an increased stability of pyrimidine single bulges with respect to purine single bulges (0.4 kcal/mol on average). This study, based optical melting experiments (UV absorption) on RNA duplexes, provided different equations (written here in the notation of [Zhu99]) for the bulge free energies of pyrimidines (eqn. 2.20) and purines (eqn. 2.21).

$$\Delta G_{37,(XNZ) \cdot (X'-Z')}^{\circ} = 3.9 \text{ kcal/mol} + 0.10 \Delta G_{37,(XZ) \cdot (X'Z')}^{\circ} + \delta g. \quad (2.20)$$

$$\Delta G_{37,(XNZ) \cdot (X'-Z')}^{\circ} = 3.3 \text{ kcal/mol} - 0.3 \Delta G_{37,(XZ) \cdot (X'Z')}^{\circ} + \delta g \quad (2.21)$$

Here,  $\delta g$  is 0 and -0.8 kcal/mol for Group I and Group II bulges, respectively. The reported stabilization of Group II bulges  $\delta g = -0.8$  kcal/mol is significantly larger than the previously reported stabilization from [Zhu99] ( $\delta g = -0.3$  to  $-0.4$  kcal/mol), thus raises questions about the mechanisms underlying Group II bulge stabilization.

Turner [Tur92] suggested that the stability of a bulged duplex could depend on the proximity of the bulge with respect to the helix end. Znosko *et al.* [Zno02] didn't find evidence for an influence of bulge position on duplex stability.

### 2.4.3 Influence of the Defect Position

Kierzek *et al.* [Kie99] investigated the effect of the position of a single mismatch within short RNA duplexes (optical melting experiments). They observed that "[...] moving the position of the mismatch toward the end of the helix enhances the stability for U·U and A·A mismatches by  $\sim 0.5$  kcal/mol per each position closer to the helix end [...]" For A·A mismatches the observed trend is less obvious than for U·U mismatches, G·G mismatches were found to be insensitive to the position within the helix. Since the study was performed with heptamer duplexes (enabling the comparison of three MM positions) the data base for the observed MM positional influence is rather limited.

Dorris *et al.* [Dor03] observed a similar positional influence for 2-base and 3-base mismatch probes (with respect to cRNA targets) on CodeLink<sup>TM</sup> 3D gel arrays. They also report a strong correlation (including the positional influence) between solution-phase melting temperatures and microarray hybridization signals of the MM duplexes.

Recent microarray studies [Wic06; Poz06; Nai06b; Nai06a] using extensive sets of probe sequences have shown a very distinct influence of mismatch position and bulge position [Nai06a], respectively. The discrimination between MM and PM is significantly more distinct for defects near the center of the duplex than for defects near the duplex ends.

Interestingly, from solution phase hybridization studies (apart from [Kie99] and [Dor03]) an influence of defect position is not been reported. In the nearest neighbor model only terminal and penultimate MM positions are considered to be less destabilizing than MMs in the interior of the duplex [Pey99; San04]. It is not clear whether the positional influence has been overlooked in previous solution-based studies or, if different experimental conditions are the reason, why a distinct positional influence has only been described recently, typically for microarray-based experiments.

**Typical characteristics of studies not reporting an influence of defect position [Ke95; All97; Pey99; Sug00]:**

- mostly solution-phase hybridization
- presynthesized oligonucleotide probes (thus containing a negligible fraction of synthesis defects)
- small probe sets ( $< 100$  probes) investigated
- the defect is typically restricted to one or few positions (commonly in the center of the duplex), no systematical variation of the defect position
- in most studies rather short duplexes  $\leq 10$  bp (little margin for variation of defect position) were employed

- experimental method: measurement of the melting curves by UV absorbance spectroscopy (for an assumed two-state melting transition the measured fraction of dissociated base pairs is equal to the fraction of dissociated duplexes)
- duplex free energies are derived from melting curve analysis

According to [Pey99] binding affinity contributions of mismatches more than three base pairs from the end are independent of the position.<sup>9</sup> [Pey99]: "Consequently, it can be concluded that the nearest-neighbor model is a good approximation for both Watson-Crick pairs and all single mismatches."

**Typical characteristics of studies reporting an influence of defect position [Ura02; Dor03; Wic06; Poz06; Nai06b]:**

- mostly microarrays studies
- microarrays in [Wic06; Poz06; Nai06b] are fabricated by *in situ* synthesis - probes can therefore contain a considerable amount of synthesis defects
- duplex length between 16 and 25 bp
- experimental method (typically): measurement of microarray hybridization signals (mostly fluorescence intensity)
- measurement of binding affinity variations depending on defect type, defect position and closing base pairs. The PM/MM hybridization signal ratio is a direct measure for the MM discrimination.
- microarray studies are favorable for large scale systematic investigations of MM discrimination (improved statistics - many different sequences, "direct comparison" of binding affinities obtained in the same experiment)

The positional influence appears to be most pronounced in [Wic06; Poz06; Nai06b]. However, this may be owing to the fact that the experimental design of these particular studies enables a more systematic and extensive investigation of the position dependence than the other studies.

Experimental results in [Kie99] and [Dor03] indicate that a positional influence is not limited to microarray studies but can be observed in solution-phase hybridization studies as well.

---

<sup>9</sup> This particular study was performed with a relative small set of 51 relatively short 9-12mer duplexes.

## Modeling of the positional influence

Pozhitkov [Poz02] considered mismatch positional influence empirically in an algorithm for finding specific oligonucleotide probes for species identification.

Binder [Bin06] tries to explain the positional influence with a zipper model in which the mismatch affects the base pairing of Watson-Crick base pairs in the duplex section between the MM and the duplex end. Therefore, the impact of a mismatch on duplex stability is getting smaller as its position is closer to the duplex end.

However, the assumed base pair opening probability (as shown in Fig. 10 in [Bin06]) doesn't account for the fact that *end fraying* under hybridization conditions is largely confined to the two [And06] or three [Lei92] outermost base pairs.

Like Binder we use a zipper based model in our analysis, however, we account for the fact that the *end fraying* is largely restricted to the outermost base pairs and that the base pair opening probability is exponentially decreasing towards the center of the duplex. Partial denaturation of inner base pairs is considered as a rare stochastic event.

## 2.5 Solid-Phase Synthesis of Nucleic Acids

In molecular biosciences synthetic nucleic acid sequences are employed in many of applications. For example, as primers for the amplification of DNA sequences by polymerase chain reaction (PCR), as target-specific probe molecules on DNA microarrays (or in *fluorescent in situ hybridization* within biological specimens), or as double-stranded RNAs for gene silencing in *RNA interference* applications.

Synthetic nucleic acid sequences are commonly produced in a *solid-phase synthesis* approach.

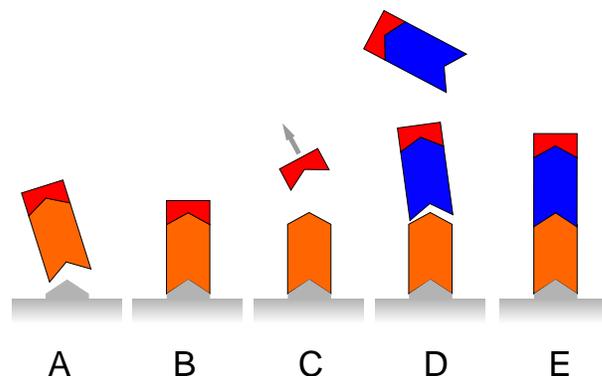
### 2.5.1 Principles of Solid-Phase Chemical Synthesis

Solid-phase synthesis has first been employed for the fabrication of polypeptide sequences<sup>10</sup> [Mer63]. In solid-phase synthesis the polymer-chains to be synthesized are end-tethered to a solid substrate. This enables efficient separation of uncoupled building blocks (in solution) from the surface-tethered synthesis products, after a synthesis step has been completed.

Coupling of monomer building blocks (see Fig. 2.18) is performed via reactive terminal groups. A removable chemical *protection group* prevents uncontrolled polymerization of

---

<sup>10</sup> For the development of the solid-phase polypeptide synthesis R.B. Merrifield received the Nobel Prize in chemistry in 1984.



**Figure 2.18:** Solid phase synthesis. (A) The first monomer (orange) is coupled to the substrate via a suitable substrate functionalization. The protection group (red) prevents further coupling reactions (B) until the a controlled deprotection reaction detaches the protection group (C). In the subsequent coupling reaction another monomer (blue) can bind to the end of growing polymer strand (D). The newly introduced protection group prevents further, uncontrolled, coupling reactions (E).

building blocks. For controlled coupling the protective groups are removed by a chemical (or photochemical) *deprotection* step - prior to addition of the building block.

In a *combinatorial chemistry* approach various types of building blocks are coupled sequentially in an optional order. The combinatorial approach enables a large number of possible products (e.g. DNA or polypeptide sequences). For example, with the four DNA building blocks  $4^{25} \simeq 1.1 \cdot 10^{15}$  different 25mer DNA sequences can be synthesized.

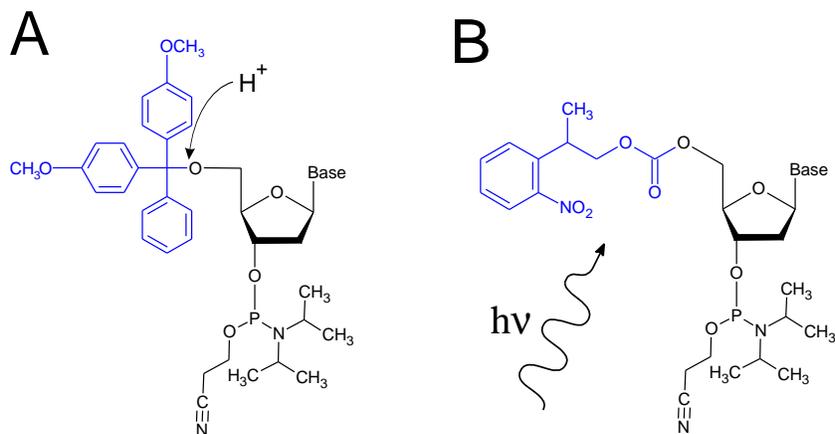
## 2.5.2 Nucleic Acid Synthesis by the Phosphoramidite Method

Nucleic acid sequences are usually synthesized by the phosphoramidite method which has been developed by Caruthers and coworkers [Bea81] in the early 1980s. Nowadays commonly  $\beta$ -cyanoethyl phosphoramidites [Sin84] are in use.

A nucleoside phosphoramidite (as shown in Fig. 2.19) contains a diisopropylamino group on its 3'-phosphate, making it susceptible to nucleophilic attack. It can react with the nucleophilic hydroxyl-group at the 5'-carbon of the (2-deoxy)ribose ring. Since phosphoramidites react with water the coupling has to be carried out under anhydrous conditions.

To prevent uncontrolled coupling of phosphoramidite building blocks a cleavable *protection group* substitutes the 5'-hydroxyl moiety. Under appropriate deprotection conditions this protection group can be removed exposing the 5'-hydroxyl, which can then couple with the 3'-phosphate of another phosphoramidite building block.

For oligonucleotide synthesis commonly dimethoxytrityl (DMT) is used as an acid-labile protection group (Fig. 2.19A). The *deprotection* step is conducted under mildly acidic conditions in 3% trichloroacetic acid.



**Figure 2.19:** Phosphoramidite reagents. A chemical protection group substitutes the 5'-OH moiety of the deoxy-ribose to prevent uncontrolled coupling of the building blocks. (A) The acid-labile dimethoxytrityl-protection group of the widely used DMT phosphoramidites is commonly removed with trichloroacetic acid. It can also be removed for example with an electrochemically generated acid [Mau06] or with photo-generated acid [Gao01]. (B) The photolabile nitrophenyl-propyloxycarbonyl protection group (NPPOC) [Has97] is removed by UV irradiation ( $\lambda \simeq 350 - 380$  nm).

Spatial control can be achieved by light-directed deprotection of photolabile deprotection groups. In the *light-directed in situ synthesis* process [Fod91], under UV irradiation (at  $\lambda = 365$  nm) photoreactive protection groups, e.g. MeNPOC ( $[\alpha$ -methyl-2-nitropiperonyl)-oxy]carbonyl) or NPPOC (2-(2-nitrophenyl)-propoxycarbonyl) (see Figs. 2.19B and 4.1), are cleaved to expose the nucleophilic 5'-hydroxyl coupling-group.

Further chemical protection groups (typically benzoyl and isobutyryl) prevent the primary amines of the nucleobases from being damaged during the synthesis process. These base protection groups, as well as the  $\beta$ -cyanoethyl protection group (at the 3'-phosphate), are removed in the *final deprotection* step, under mildly alkaline conditions.

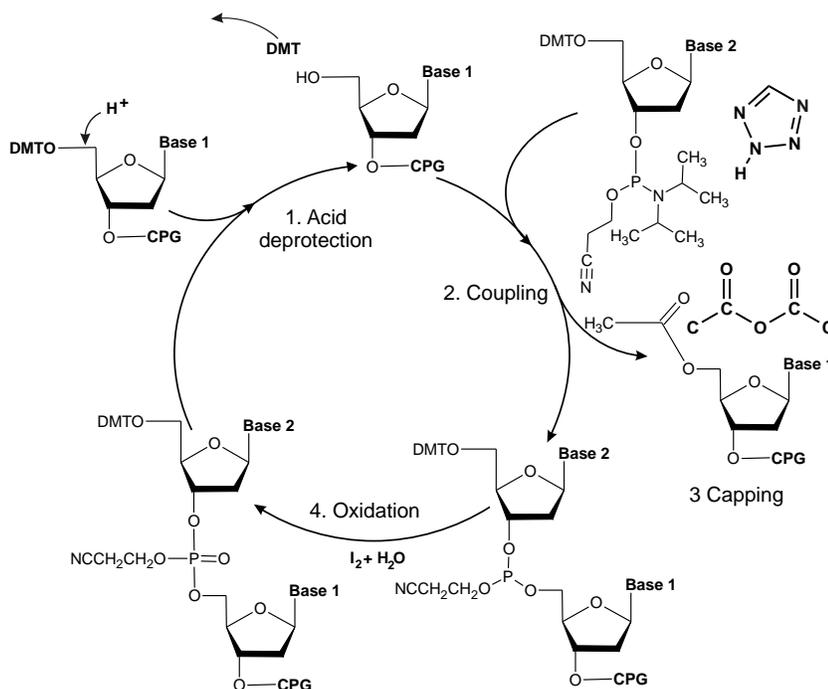
Oligonucleotide synthesis is performed in solid-phase, usually on a functionalized glass surface. Commonly controlled pore glass (CPG) beads are used as substrate material, since their large surface provides a higher yield than a flat glass surface. Substrate functionalization [Bei99; Ben02; LB03] comprises an organic linker/spacer molecule which is covalently bound to the silanized glass surface. Hydroxyl-moieties are required for phosphoramidite coupling.

Steps of the oligonucleotide synthesis cycle (Fig. 2.20) as applied in solid phase oligonucleotide synthesis (phosphoramidite method): the synthesis starts with the coupling of the first nucleotide on the hydroxy-functionalized substrate. The 4-step-cycle is repeated for the addition of each monomer.

- Deprotection (detritylation). Removal of the 5'-DMT group to expose the nucleophilic

5'-hydroxyl is performed under acidic conditions.

- Coupling of the next phosphoramidite building block under anhydrous conditions. Activation of the coupling reaction by the weak acid tetrazole.
- Capping of unreacted hydroxyl groups by acetylation. This prevents the synthesis of strands with single base deletions, since the acetyl group blocks coupling of further monomers. In the final purification (e.g. by HPLC) the significant fraction of truncated strands can be separated easily.
- Oxidation of the phosphite triester linkage with iodine to chemically stabilize the phosphate linkage.



**Figure 2.20:** The phosphoramidite method: oligonucleotide synthesis on controlled-pore glass (CPG) - synthesis cycle. 1. Removal of the acid labile DMT protection group (detritylation) exposes the 5'-hydroxy group. 2. Activation of the 3'-phosphate group (detritylation) exposes the 5'-hydroxy group. 2. Activation of the 3'-phosphate group with tetrazole enables coupling of the new phosphoramidite building block at the deprotected 5'-OH group of the CPG-bound strand. 3. The synthesis of oligonucleotide strands with still unreacted 5'-OH (owing to an incomplete coupling reaction) is blocked by reaction with acetic anhydride. 4. Oxidation of unstable phosphite linkages with iodine results in a more stable phosphate linkage. Washing with pure solvent is required between the steps.

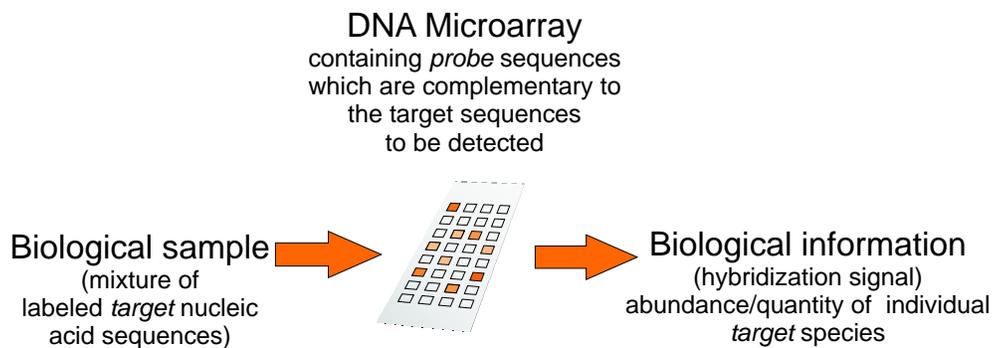
A final deprotection step (using concentrated ammonium hydroxide) is required to remove remaining protection groups (base and phosphate) and to cleave the oligonucleotides from the solid support.

Purification with polyacrylamide gel electrophoresis (PAGE) or high pressure liquid chro-

matography (HPLC) yields the full length oligonucleotides. The coupling efficiency is typically  $> 98\%$ . The yield of full-length oligonucleotides reduces with the power of the length  $n$  (number of bases) of the oligonucleotide sequence. Thus, the yield from a 100mer synthesis (with a coupling efficiency of  $98\%$ ) is  $0.98^{100}=13\%$ . The maximum sequence length for economically reasonable synthesis is about 100 bases. Longer sequences (e.g. for the synthesis of complete genes) can only be produced by ligation of shorter strands. The cost for oligonucleotide synthesis on a  $0.05\ \mu\text{mol}$  scale is currently on the order of 20 cents per base addition. Possible modifications of synthetic oligonucleotides include, for example, fluorescent dye labels, biotin labels, fluorescence quenchers, spacers and various linker chemistries.

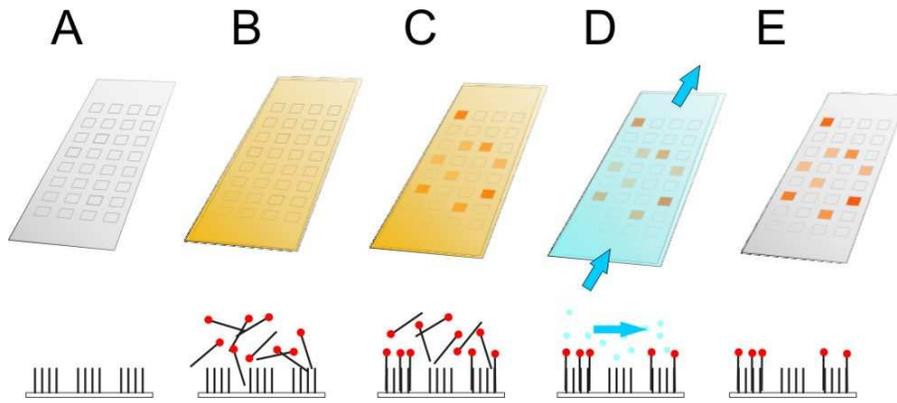
## 2.6 DNA Microarrays

DNA microarrays are hybridization-based sensors for massively parallel analysis of the composition of complex nucleic acid mixtures (Fig. 2.21).



**Figure 2.21:** Microarray principle. A biological sample containing a mixture of nucleic acid sequences to be analyzed is applied - dissolved in a buffer solution - on the microarray surface. The labeled *target* strands hybridize to the complementary *probe* sequences to which they have a highly specific binding affinity. The hybridization signal (typically the fluorescence intensity of the labeled targets), which is related to the abundance of hybridized targets, is acquired by CCD or confocal microarray scanners. The microarray analysis is performed with many probe-target species in parallel. Arranged in a regular array of *microarray features*, *probes* can be identified by their position on the microarray. The hybridization signal of a microarray feature provides a semi-quantitative measure for the abundance of the corresponding *target* species.

The well-known *probe* sequences - which are arranged in a regular array structure - are end-tethered on the planar microarray surface (Fig. 2.22). The *target* mixture to be analyzed (identification or quantification of individual *target* species) is applied in solution onto the microarray surface. Targets can freely diffuse in the hybridization buffer solution



**Figure 2.22:** Microarray hybridization assay. (A) Microarray with surface tethered *probe* sequences (lower images). The individual *probe* sequences can be identified by the position of the corresponding feature on the regular microarray grid. (B) Application of the hybridization solution. (C) Labeled *targets* can freely diffuse over the microarray surface until they hybridize with a complementary probe. (D) In the washing step after the hybridization unbound *targets* are washed away. Microarray analysis (E) - quantification of the surface bound *targets* is performed by measurement of the fluorescence intensity. By the position of a feature the *probe* sequence can be identified. The hybridization signal of a microarray feature depends on the *target* abundance, but also on the particular *probe-target* binding affinity.

until they are captured by a complementary *probe* sequence.

After the hybridization, which, owing to the slow diffusion process, typically requires several hours, unbound *targets* can be washed off easily, whereas hybridized *targets* remain bound on the the microarray surface. A molecular marker (usually a fluorescent dye label, or an antibody-specific molecule like biotin) is used for identification and quantification of the hybridized *target* molecules.

The particular *probe* sequence species are restricted to small areas commonly called features or spots. The arrangement of these features as a regular grid ("array") enables identification of the features. By the feature position the hybridization signal can be assigned to the corresponding *probe* sequence.

In principle DNA microarrays could be used for measurement of individual *target* concentrations. However, owing to many factors affecting the measurements (e.g. poor predictability of individual *probe-target* affinities, cross hybridization) microarrays are only semi-quantitative. This, however, is mainly limiting the comparability between different *probe-target* pairs. For individual *probes* the hybridization signal difference measured for two different nucleic acid target samples corresponds to the change in *target* concentrations. In gene expression analysis (see below) this so-called *fold-change* describes how many times the expression signal for a given transcript is increased or decreased with respect to the control.

### 2.6.1 Microarray Applications

**Gene expression profiling.** The experimental design of gene expression assays is usually based on the comparison of the gene expression levels of a biological sample of interest with that of a reference sample (see Fig. 2.23).

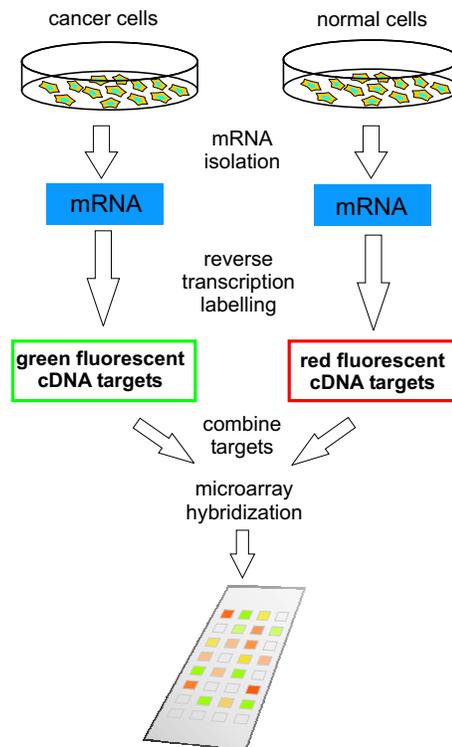
Example: From a batch of yeast cells grown under controlled conditions, one sub-sample is exposed to a heat shock - the condition to be investigated, while another sub-sample is employed as a reference sample. The comparison of the expression profiles from both samples highlights the differences in gene expression, and thus enables identification of genes involved in stress response.

Practical applications: functional genomics - investigation of gene functions, pathology (e.g. for identification of cancer-types), pharmacogenomics (investigation of the individual drug response), toxicity tests

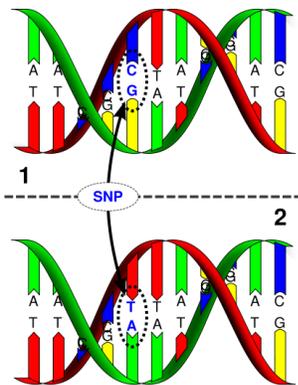
**Genotyping assays.** Single nucleotide polymorphisms (SNPs - see Fig. 2.24) – sequence variations in which single nucleotides differ between the members of a species (or even between the two alleles in diploid cells) – have a strong influence on the phenotype. SNPs are responsible for the majority of genetic variations within a single species. SNPs are associated with a predisposition to a variety of diseases. Other SNPs are associated to individuals' response to pathogens, chemicals, drugs, vaccines, and other agents. SNP microarrays employ probe sequences specific to known SNP sites. SNP arrays make use of the specificity of relatively short 12 to 30mer oligonucleotide probes to detect single mismatched base pairs originating from SNPs [Con83]. Genotyping arrays are a valuable tool in genomics research, pharmaceutical research (with a focus on the individual response to pharmaceutical agents) and increasingly in medical diagnostics.

**Resequencing.** Resequencing assays enable identification of genetic differences with respect to a well-known reference genome. This enables, for example, discrimination between closely related virus or bacteria strains and identification of previously unknown strains [Won04].

**Pathogen detection/identification.** Microarrays comprising pathogen-specific probe sequences enable fast detection and identification of viruses and bacteria (see above). Microarrays can be employed for large scale pathogen screening (medical diagnostics, food safety, biodefense applications).



**Figure 2.23:** Dual color microarray experiment. In this example the expression profile of cancer cells is compared to a reference sample of normal cells. Complex mixtures of messenger RNAs (mRNAs) are isolated from each sample and fluorescently labeled via reverse transcription labeling. The targets from the cancer cell are labeled with a green fluorescent dye, whereas the targets from the reference sample are labeled with a red fluorescent dye. The targets are combined and hybridized on the same microarray. Analysis and comparison of the two color-channels enables identification of up- and down-regulated genes. (Adapted from Wikipedia: [http://en.wikipedia.org/wiki/DNA\\_microarray](http://en.wikipedia.org/wiki/DNA_microarray))



**Figure 2.24:** Single Nucleotide Polymorphisms (SNPs) are genetic variations of single base pairs between members of the same species, or even between the two copies of a chromosome pair. The DNA strand in 1 differs from the DNA strand in 2 by a single base pair. Genotyping assays enable high-throughput screening for single nucleotide polymorphisms. (Source: Wikipedia, [http://en.wikipedia.org/wiki/Single\\_nucleotide\\_polymorphism](http://en.wikipedia.org/wiki/Single_nucleotide_polymorphism))

## 2.6.2 The Development of DNA Microarray Technologies

An early method (1975) for the analysis of complex nucleic acid mixtures is the *Southern blot* [Sou75]. Thereby the mixture of unidentified DNA fragments (*targets*) is separated by gel electrophoresis, transferred and immobilized on a flexible nylon membrane. For identification of the *targets* radioactively or chemically labeled *probes* (with well-known sequences) are incubated with the membrane, thus enabling hybridization with the complementary *target* sequences.

The so-called *dot blot* is a similar technique, in which the (unseparated) *target* sample is directly applied onto the membrane as "dots". After fixation the identification of the *target* sequences is performed by hybridization with a labeled *probe* sequence (or a mixture of labeled *probes*).

Miniaturization and parallelization have evolved the *dot blot* into the high throughput *macroarray* technique. With the help of automated methods several thousand millimeter-sized nucleic acid *spots* can be immobilized on a nylon membrane (typically 10 to 20 cm in size). Here, different from the blotting techniques described above, the known *probe* sequences (e.g. cDNA or synthetic oligonucleotide *probes*) are immobilized on the solid substrate, whereas the *targets* are applied in hybridization solution. Autoradiographic analysis and the large quantity of *probe* material provide a high sensitivity. However, radioactive labeling with  $^{32}P$  or  $^{33}P$  (requiring precautious handling) and the need for large quantities of *probe* and *target* material are serious disadvantages of the *macroarray* technique.

By using rigid substrates rather than flexible nylon membranes, a significant miniaturization was achieved, giving rise to *DNA microarray* technology. Microarrays are commonly produced on chemically functionalized glass substrates - frequently a microscope slide format is employed. The use of glass substrates, which, unlike nylon-membranes, have low auto-fluorescence, enables highly sensitive detection of fluorescently labeled *targets*.

Different types of DNA microarrays have been developed in several independent approaches:

- In 1995 Schena *et al.* [Sch95] reported the first gene expression assay on a printed microarray. They employed a *contact printing technique* for deposition of tiny spots (about 0.1-0.2 mm in diam.) of nucleic acids *probes* (cDNA *probes*) on a chemically functionalized glass substrate. This now widely-used technique is also known as *spotting*. The spotting solution with the prefabricated nucleic acid *probes* is deposited on the surface by a pin. A capillary gap at the tip of the pin releases a small (and reproducible) amount of the spotting solution when the pin is touching the substrate surface. Chemical functionalization of the substrate (e.g. with amino-, epoxy- or aldehyde-groups) and the *probe* molecules (e.g. by attachment of an amino group)

enable fixation (immobilization) of the *probes*. cDNA microarrays are mainly used in gene expression assays. Apart from cDNA and PCR products, presynthesized oligonucleotide probes can be immobilized on microarrays ( $\rightarrow$  *oligonucleotide microarray*). Microarray robots (*arrayers*) are commonly employed for a fully automated fabrication process.

- Already several years earlier Fodor *et al.* [Fod91] developed a photolithographically controlled combinatorial chemistry approach for the fabrication of *high-density oligonucleotide microarrays*.<sup>11</sup> Owing to the similarity of the photolithographic fabrication process with semiconductor fabrication techniques, these microarrays are commonly called *DNA chips*. Unlike the spotting approach the light-directed *in situ* synthesis approach doesn't require prefabricated *probes* for deposition. The *probe* molecules (DNA oligonucleotides) are fabricated *in situ*, i.e. nucleotide by nucleotide, on the microarray substrate. The massively parallel synthesis of up to a million different sequences on the same chip is directed by UV light exposure. In the combinatorial synthesis process chrome masks provide a sequence specific exposure scheme (spatially restricted to the particular microarray features) to control the sequence of nucleotide couplings for each *probe* sequence individually.
- Ink-jet techniques (based on piezoelectric deposition) are used for *in situ* synthesis of microarrays [Bla96] (by deposition of phosphoramidites) and also for *spotting* of presynthesized DNA [Sch98].
- A rather novel technique is the electrochemical *in situ* synthesis of DNA microarrays [Mau06]. Thereby nucleic acid coupling is controlled by acid generation on a CMOS addressable electrode array.

Depending on the type of probes employed DNA microarrays (not to be confused with other types of microarrays, e.g. protein microarrays) can be categorized into two groups:

### cDNA microarrays

This type of microarray comprises immobilized cDNA probes or PCR products. Owing to the availability of cDNA and PCR products from biological sources, cDNA arrays (*microarrays* and *macroarrays*) are frequently prepared by biological labs. Since the long *probe* sequences (typically one hundred to several hundred nt long) are not suitable for discrimination between similar sequences (e.g. for the identification of single base MMs) the application of cDNA microarrays is restricted to gene expression profiling.

---

<sup>11</sup> "High density" refers to a high density of microarray features (up to one million per  $\text{cm}^{-2}$ )

## Oligonucleotide microarrays

Oligonucleotide microarrays comprise synthetically fabricated *probe* sequences which are typically between 15 and 100 bases long. Unlike cDNA microarrays, oligonucleotide microarrays enable discrimination of very similar genes belonging to the same gene family. Short oligonucleotides (< 30 nt) owing to their high discrimination capability are used for genotyping and resequencing applications. Long oligonucleotide probes (~60 nt) have the advantage of providing a high sensitivity for the detection of low abundance transcripts. Oligonucleotide microarrays are fabricated by immobilization (*spotting*) of presynthesized oligonucleotides, or by *in situ* synthesis.

A detailed overview of microarray types and fabrication methods is provided in [Gao04].

### 2.6.3 Characteristics of Microarray Hybridization

Literature reports a large discrepancy between hybridization characteristics in bulk solution and on the microarray surface [Hel03; Bin06; Poz06]. Since NN thermodynamic parameters were determined in solution-phase experiments - under "ideal hybridization conditions", the nearest-neighbor model doesn't necessarily perform satisfactory for the prediction microarray binding affinities.

According to Bhanot *et al.* [Bha03], the loss of translational energy and entropy of microarray-bound probes (with respect to hybridization of free strands in bulk solution), and the constraint that targets can approach the probes only from one half-space, is independent of the sequence. Thus, with respect to bulk-solution, hybridization equilibrium constants, equilibrium constants for microarray hybridization are multiplied by the same sequence-independent factor. The difference between solution-phase and surface-phase hybridization is of little consequence for specificity and sensitivity when equilibrium is achieved. However, hybridization kinetics (which is different for surface- and solution-phase hybridization) has a pronounced effect on specificity and sensitivity [Bha03].

Levicky and Horgan [Lev05] reviewed physicochemical aspects of DNA microarray hybridization. In particular they discussed differences with respect to solution-phase hybridization.

On DNA microarrays (with respect to solution hybridization) melting temperatures [Hel03] are significantly reduced. Additionally, significantly broadened hybridization isotherms (deviating from Langmuir-type characteristics) [Bin06] are observed. Moreover, on DNA microarrays a strong influence of the position of single base MMs on duplex binding affinities [Wic06; Poz06; Nai06b] is observed.

An "ideal microarray" (in terms of specificity and target quantization) would require the following characteristics:

- target-specific hybridization: i.e. *probes* hybridize only with the complementary *target* species
- there is no intra-strand base pairing leading to formation of *probe* or *target* secondary structures.
- all *probe-target* pairs have approximately the same binding affinity
- there is a simple (e.g. linear) relation between the hybridization signal and the concentration of the corresponding *target* sequence

Real microarrays deviate from the "ideal microarray" (above) in several aspects:

- complex *target* mixtures give rise to competitive hybridization processes [Poz06] (unspecific *target/target* and *probe/target* cross hybridization)
- the use of long relatively long *target* sequences (typically between 100 and several hundred nt long) results in *target* secondary structure formation and increased potential for *cross hybridization*. Both processes compete with the specific *probe-target* hybridization. *Target* secondary structure can prevent *probe-target* hybridization, thus leading to false negatives. Unspecific cross-hybridization can result in false positives.
- surface effects (e.g. electrostatic effects and sterical hindrance) can (in conjunction with the varying length/secondary structure of individual *targets*) affect the quantitiveness of the measurement (binding affinity is a function of the amount of hybridized *targets*, *target* length and *target* structure)
- *probes* are confined to a small area on the microarray ( $\Rightarrow$  diffusion-limitation effects)
- synthesis defects (originating from *in situ* synthesis) affect binding affinities
- labeling of the *target* sequences (e.g. with large fluorescent dye molecules like Cy3 attached at random positions) may affect binding affinities

### Microarray hybridization - a diffusion driven process

Microarrays are often fabricated on microscope slides with dimensions of about 75 mm  $\times$  25 mm. The hybridization solution (ten to several hundred  $\mu$ l) is inserted into the gap between the microarray and a cover glass, thus forming a thin film with a thickness of 20-100  $\mu$ m. This is better illustrated by the following comparison in which the microarray is assumed to be enlarged to the size of a football field. On this scale, the liquid film corresponds to a puddle between 2 and 10 cm deep. The size of a microarray features may be visualized by a soccer ball.

Hybridization in such a configuration is a slow process since diffusion is the dominating transport mechanism for the targets. The hybridization of a *target* with the corresponding *probe* is usually limited by the slow diffusion process [Pap06]. With the Einstein-

Schmoluchowski relation we find that the average distance a target molecule (with a molecular diffusion coefficient of  $10^{-11}$  m<sup>2</sup>/s [Pap06]) is traveling in an overnight hybridization is approximately 1 mm. In microarray assays, by diffusive transport alone, the equilibrium can't be reached on a reasonable time scale.

Novel chaotic micromixing techniques, e.g. based on surface acoustic waves (SAW) [Toe03], can very efficiently generate microagitation in the capillary gap and thus overcome the diffusion limitation.

By scaling down the dimensions of the microarray ( $\Rightarrow$  increased ratio between the diffusion coefficient and the microarray surface) the hybridization equilibrium can be reached on a realistic time scale [Dan07].

#### 2.6.4 Further Reading on the Technical and Physical Principles of DNA Microarrays

- Sensitivity, specificity, cross hybridization ( $\rightarrow$  detection of false positives) [Bha03; Bin06]
- Point defects (mismatches [Dod77; Wal79; Nel81; All97], base bulges [Ke95; Zhu99; Zno02]), influence of defect position [Dor03; Wic06; Poz06], discrimination capability [Ura03; Lee04]
- Secondary structure of *probes* and *targets* ( $\rightarrow$  detection of false negatives) [Lue03; San04]
- Microarray fabrication (immobilization, *in situ* synthesis) [Sch99; Sch02; Gao04]
- Quality of the *probe* sequences - synthesis defects [Gar02; Job02; Ric04; Bin06]  $\rightarrow$  heterogeneity of binding affinities
- Target preparation [Sch02] (*target* length, fluorescent labeling, composition of the *target* mixture, type of nucleic acid *target* - DNA or RNA)
- Competitive effects [Bin06]
- Surface density of the *probes* [Pet02; Wat00; Lev05] (steric hindrance [Hal05], electrostatic repulsion [Vai02; Bin06])
- Attachment of the *probes* [Sch02], linker/spacer [Bei99], linear and dendrimeric linkers [Cam06]
- various hybridization parameters [Sch02] (e.g. ionic strength, temperature, pH, blocking reagents) [Koe05]
- Washing characteristics [Poz07]
- Microarray size [Dan07], diffusion-limited *target* transport [Pap06], mixing [Gut05; Toe03]

## 2.7 DNA Chip Fabrication by Light-Directed *In Situ* Synthesis

Light-directed *in situ* synthesis of DNA microarrays was developed around 1990 by Fodor and coworkers [Fod91]. Short (typically  $\leq 25$ mer) oligonucleotide *probe* sequences are synthesized nucleotide by nucleotide on the surface of the microarray. Spatially addressable photo-deprotection enables a massive parallel synthesis of arbitrary DNA *probe* sequences on a single microarray.

Light-directed *in situ* synthesis is basically a solid-phase synthesis process (see section 2.5.1), requiring phosphoramidite reagents with photo-labile protection groups. Spatially controlled photo-deprotection is achieved with a photolithographic process and the use of phosphoramidite reagents with photolabile protection groups. *Probe* sequence information and microarray geometry is encoded in the photomasks.

Today, commercial high density oligonucleotide microarrays (fabricated with high resolution photomasks) have up to 6.5 million *probes* (with a feature size of 5  $\mu\text{m}$ ). Light-directed *in situ* synthesis can also be employed for the synthesis of polypeptide sequences [Fod91] ( $\rightarrow$  protein microarrays) or other combinatorial chemistries.

### 2.7.1 Photolithographic Control of the Combinatorial Synthesis Process

For parallel synthesis of different *probe* sequences spatial control of the phosphoramidite coupling reaction is required. This is achieved by a spatially controlled photo-deprotection of the photolabile 5'-protection group (chemical structure shown in Fig. 2.19). The photocleavage generates a hydroxy-group at the 5'-ends of the exposed sequences and thus determines where on the microarray (i.e. at which microarray features/probe sequences - see Fig. 2.25 h and k) the next phosphoramidite building block (provided in the subsequent coupling step) will elongate the sequence.

The fabrication of a microarray comprising arbitrary N-mer sequences requires  $4 \times N$  deprotection/coupling steps. It is necessary to provide all coupling alternatives (X=A, C, G and T) in each "nucleotide layer". Thus, the light-directed combinatorial synthesis comprises a series of  $4 \times N$  photo-deprotection  $D_{X_i}$  and associated nucleotide coupling

steps  $C_{X_i}$ :

$$\begin{aligned}
 &1. D_{A_1}/C_{A_1} \rightarrow D_{C_1}/C_{C_1} \rightarrow D_{G_1}/C_{G_1} \rightarrow D_{T_1}/C_{T_1} \\
 &2. D_{A_2}/C_{A_2} \rightarrow D_{C_2}/C_{C_2} \rightarrow D_{G_2}/C_{G_2} \rightarrow D_{T_2}/C_{T_2} \\
 &\quad \Rightarrow \dots \\
 &N. D_{A_N}/C_{A_N} \rightarrow D_{C_N}/C_{C_N} \rightarrow D_{G_N}/C_{G_N} \rightarrow D_{T_N}/C_{T_N}
 \end{aligned}$$

Spatially controlled photo-deprotection is shown in more detail in Fig. 2.25 where each *probe* strand symbolizes an individually addressable microarray feature (whereas in reality each feature comprises millions of identical *probes*). Assuming a stepwise coupling efficiency  $f_c$  the yield  $Y = f_c^N$  of *probes* which are free of synthesis defects is decreasing with the power of the *probe* length  $N$ . *Probes* containing defects cannot be repaired or removed as in common solid-phase synthesis (capping, truncation, HPLC separation). Synthesis defects (i.e. single base mismatches, insertions and deletions) will therefore affect microarray hybridization [Job02].

The length of the microarray-*probes* is determined by the application. Shorter 15-25mer probes provide a high discrimination capability between PM and MM and are therefore suitable for SNP detection and resequencing assays. Longer *probes* are less discriminative but rather more sensitive (increased binding affinity), and are therefore favorable for detection of low abundance mRNAs in expression profiling applications. Typically *probes* on high density oligonucleotide microarrays have a length of  $\leq 25$  nt, however, the fabrication/application of arrays with longer 40-60 nt *probes* has also been reported.

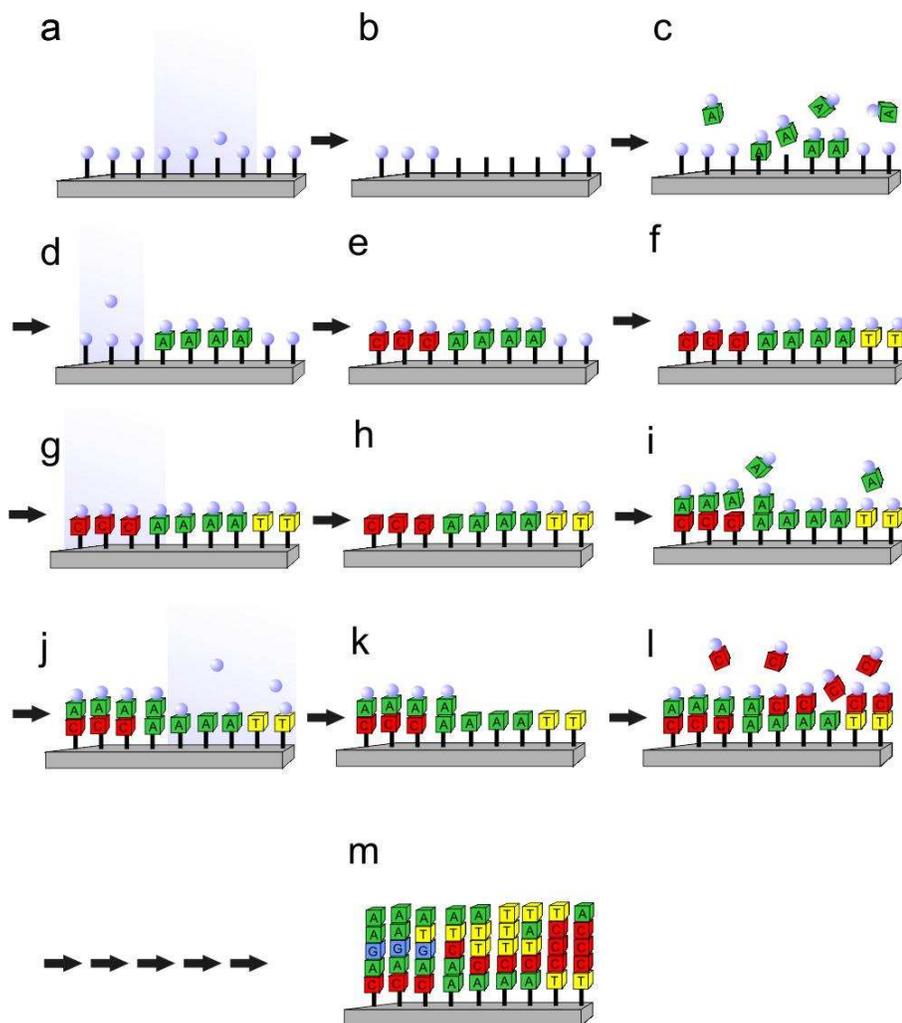
### The synthesis cycle

For light-directed *in situ* synthesis photolabile phosphoramidite reagents,  $\alpha$ -methyl-6-nitropiperonyloxycarbonyl (MeNPOC) [Pea94; McG97] or [2-(2-nitrophenyl)-propyloxycarbonyl]-2'-deoxynucleoside (NPPOC) phosphoramidites [Has97] are used.

The MeNPOC-chemistry (employed in the fabrication of Affymetrix GeneChips<sup>®</sup>) has a stepwise yield of 92 to 94% [McG97]. Significantly better coupling yields have been reported for NPPOC phosphoramidites [Bei99]. Nuwaysir *et al.* [Nuw02] reported stepwise chemical yields between 96 and 98%.<sup>12</sup>

Use of NPPOC phosphoramidite reagents (chemical structure shown in Fig. 4.1) has been reported in [Has97; Bei99; Nuw02; Bau03; Wol04; Woe06]. MeNPOC phosphoramidites reagents were used in [Pea94; McG97; SG99; Lue02].

<sup>12</sup>Stepwise synthesis yields of NPPOC phosphoramidites according to [Nuw02]: NPPOC-A(tac) 96%, NPPOC-C(ibu) 99%, NPPOC-G(ipac) 97%, NPPOC-T 98%



**Figure 2.25:** Light-directed *in situ* synthesis of DNA microarrays [Fod91]. The spatially controlled combinatorial chemistry approach enables parallel synthesis of arbitrary *probe* sequences. In the (non-optimized) coupling scheme shown here, the *probe* sequences are synthesized "layer by layer": to cover all coupling-alternatives, in each "nucleotide layer" phosphoramidite-couplings are performed in the order A, C, G, T. In the above series of images each *probe* strand symbolizes an individually addressable microarray feature (whereas in reality each feature contains millions of *probes*). Synthesis of the first nucleotide layer (a-f). (a) The substrate is initially functionalized with photo-labile protection groups (depicted as blue balls). Spatially controlled UV exposure (use of photomasks) is restricted to those feature areas where phosphoramidite building blocks are to be attached in the subsequent coupling step. (b) Photo-cleavage of the protection groups created hydroxyl-moieties, which are the binding sites for the subsequent adenosine-phosphoramidite coupling step (c). In the coupling step only one building block can attach to each deprotected strand. Further couplings are prevented by new protection groups (imported with the building blocks). (d) Photo-deprotection of those probes which require cytosine at the first base position. (e) Coupling of cytosine-phosphoramidite. The first nucleotide layer is completed after deprotection and coupling of G nucleotides (not shown) and T nucleotides (f). The second layer is synthesized upon the first layer (g-l). The deprotection/coupling scheme is continued until the final length of the oligonucleotide probes is reached (m).

Owing to the the 5'-attachment of the NPPOC protection groups the *in situ* synthesis is performed in 3'→5' direction. Therefore the probes are typically 3'-tethered at the microarray surface. However, 5'-tethered *probes* can be synthesized (in 5'→3' direction) with modified phosphoramidite reagents (carrying 3'-NPPOC protection groups) [Alb03]. 5'-tethered microarray probes are, unlike 3'-tethered probes, available for enzymatic modification.

In the following we refer to the 5'-NPPOC phosphoramidite chemistry (Fig. 4.1) which has been employed in this work.

The synthesis cycle in the light-directed synthesis process (Fig. 4.2) is very similar to the scheme employed for oligonucleotide synthesis on CPG-supports (shown in Fig. 2.20). Exposure with UV light ( $\lambda=350-380$  nm) induces photo-deprotection and enables coupling of the next phosphoramidite building block. A capping step (as shown in Fig. 2.20), resulting in truncated strands rather than in strands containing single base MMs, is of a rather limited value in microarray synthesis (truncated strands cannot be removed) and is therefore omitted. Coupling and oxidation steps are performed in the same way as in the oligonucleotide synthesis on CPG-supports.

Photo-deprotection of NPPOC results in short-lived intermediate states. According to [Wal01] an *aci*-nitro intermediate is in acid-base equilibrium with its anion. The unstable anion can fragment, thus resulting in the desired deprotection reaction (complete removal of the NPPOC group). However, via a competing reaction pathway, the *aci*-nitro intermediate can also form a nitroso product, which is not removed from the phosphoramidite residue, thus preventing photo-deprotection.

To promote the desired reaction pathway the photo-deprotection needs to be performed in a solvent providing sufficient proton acceptors. Therefore the basicity of solvent acetonitrile is increased by addition of a mild base (e.g. piperidine [Bei99]).

### 2.7.2 Combination of "Maskless" Digital Photolithography and Combinatorial Chemistry

Light-directed *in situ* synthesis of DNA microarrays with high resolution photomasks has been developed and is employed on an industrial scale by Affymetrix Inc.. High costs for chromium masks, considerable technical effort for the mask alignment, and the lack of flexibility (a new set of photomasks is required for each new microarray design) have so far prevented lab-scale application of the photomask-based fabrication technique.

The use of computer-controlled spatial light modulators as "virtual photomasks" can circumvent the limitations related to the use chromium photomasks and thus provide great flexibility for custom microarray fabrication.

Cerrina and coworkers (University of Wisconsin) developed a "maskless" fabrication technique based on pattern projection with a digital micromirror array [SG99]. In their setup a *digital micromirror device* (DMD<sup>TM</sup>, Texas Instruments, see Appendix B.1), a spatial light modulator as commonly used in video projectors, is employed for "virtual photomask" projection. Unlike LCD spatial light modulators, DMDs, owing to robust microelectromechanically controlled mirrors, are suitable for the projection of UV light. About one million individually controlled tiltable micromirrors (corresponding to the pixels on conventional display devices) can either reflect incoming light towards the projection plane, or in a different direction into a light trap (→ Digital Light Processing, DLP<sup>TM</sup>, Texas Instruments).

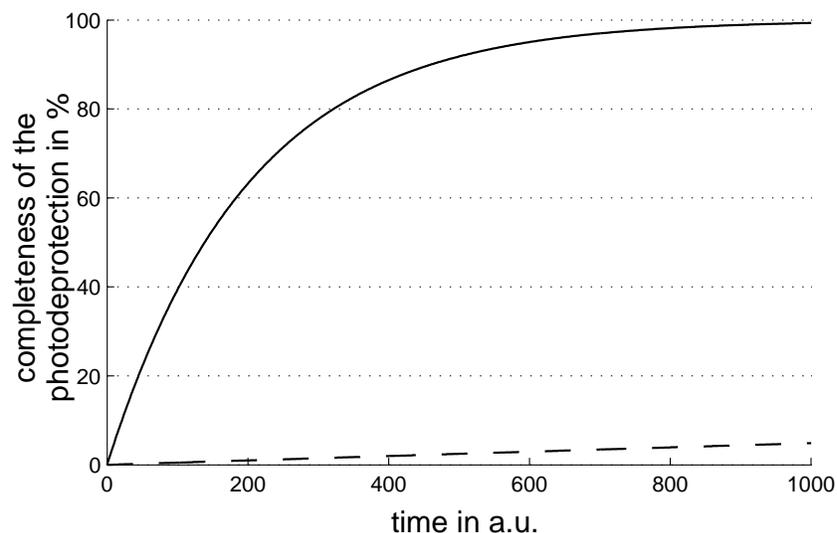
The virtual masks are generated by a personal computer and displayed as "black and white" images on the DMD (basically in the same way as on a computer screen). Projection of the DMD onto the microarray glass substrate is performed with a reflective *Offner relay* 1:1 imaging system. Mask alignment during the  $\sim 100$  exposures with different mask patterns is inherent to the system. However, considerable image drift due to thermal expansion of optical components has been reported to occur during the several hours lasting synthesis process [Ric04]. The use of an image-locking system to counter image-drifting considerably improved the quality of the DNA probes synthesized [Ric04].

### Image contrast

Image contrast is crucial for the quality of the DNA sequences synthesized with light-directed *in situ* synthesis: Garland *et al.* [Gar02] point out that an assumed contrast ratio of 400:1 (for spatial light modulator based synthesis process), gives rise to a considerable amount of synthesis defects (65% of the products of a 20mer synthesis due to random insertions are actually 21mers or longer).

The photo-deprotection reaction needs to be driven close to completeness to prevent single base deletions. However, the ratio between exposure and stray light (background) intensity has to be maximized to prevent unwanted deprotection by stray light, which can produce random base insertions.

Unlike for example photoresist, the photo-deprotection reaction has a linear response to UV intensity. Even worse, the deprotection of an exposed feature is only asymptotically increasing towards completion (Fig. 2.26), whereas simultaneously in unexposed features, induced by stray light, the fraction of erroneously deprotected groups increases almost linearly. Furthermore, in the combinatorial synthesis process, owing to the alternative deprotection and coupling steps, the exposure to stray light is 3 times longer than the actual photo-deprotection step.



**Figure 2.26:** Time course of the photo-deprotection in a microarray feature under exposure (solid line) in comparison to another microarray feature which is exposed to stray light (originating, for example, from a neighboring feature under exposure) only (dashed-line). Under stray light intensity - in this example we assumed a contrast ratio 1:100 - the deprotection rate is reduced by the corresponding factor. Owing to the almost linear increase of the fraction of erroneously deprotected groups the fraction of synthesis errors (random insertions) is distinctly larger than the contrast ratio ( $I_{stray}/I_{exposure}$ ) might suggest: after an exposure of  $t=800$  a.u. the fraction of deprotected groups has reached about 97%. In the same time approx. 4% of the protection groups of the unexposed features have been deprotected by stray light.

In the photolithographically controlled fabrication process stray light originating from a feature under UV exposure will affect mainly neighboring features that are currently not under exposure. The local contrast [Kim04] between neighboring features in particular matters: It is determined by the microarray design (features size and feature spacing), diffraction of the light at the edges of the features and optical flare (caused by reflections within the UV optical system).

Considering the large fraction of probes containing single base defects (MMs, insertions and deletions) it is almost surprising that DNA microarrays produced by an *in situ* synthesis process perform well.

### Implementations of DMD-based maskless *in situ* synthesis systems

The maskless array synthesizer (MAS) developed by the Wisconsin group has been commercialized by NimbleGen Systems Inc., which is now using the MAS technology for fabrication of customized microarrays.

The similar DMD-based "Digital Optical Chemistry System" [Lue02; Lue03] has been developed by the Garner Lab (University of Texas Southwestern Medical Center).

Another system developed by Xeotron Corp. uses photo-generated acid for the deprotection of standard acid-labile phosphoramidites [Gao01]. In the individual cells of the microreactor acid is generated by UV exposure which is controlled by a digital light processing system.

The Geniom<sup>®</sup> system (Febit biotech GmbH, Heidelberg) is marketed as an integrated system for customized DNA microarray synthesis and analysis [Bau03]. Light-directed *in situ* synthesis and microarray analysis are performed in small-volume microfluidics channels.

# Chapter 3

## Development of the DNA Microarray Synthesizer

### 3.1 Motivation and Overview

Compared with the immobilization of presynthesized probes the "maskless" light-directed *in situ* synthesis of DNA microarrays is a highly advanced and flexible technique. Since probe sequences don't need to be synthesized one-by-one the highly parallel *in situ* synthesis technique is significantly less labor intensive and more cost-efficient than traditional microarray fabrication methods. Short turnaround times from microarray design to application enable a fast evolution of experiments. So far, due to relatively large technical requirements (and also for intellectual property reasons) light-directed *in situ* synthesis has not become a standard technique for lab-scale fabrication of DNA microarrays.

Commercial microarrays usually appear to the user as a "black box technology". The relatively high costs still limit the widespread application of microarray technologies.

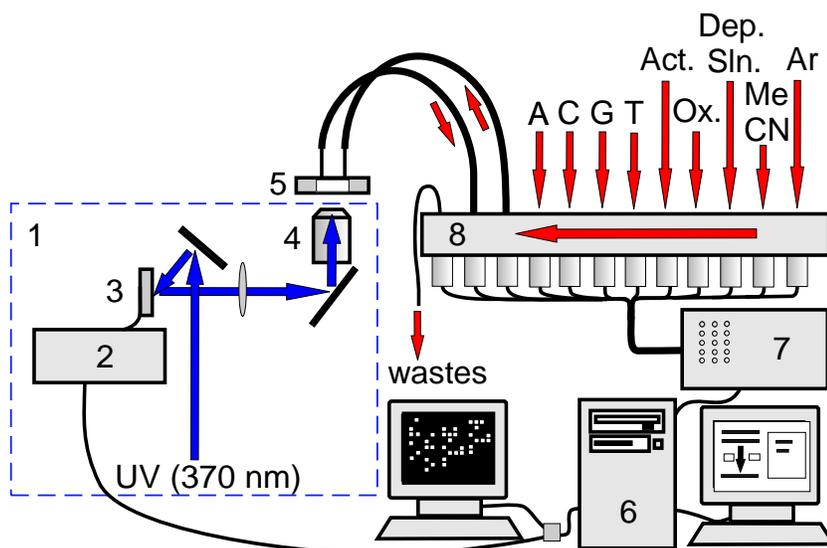
Aside from Affymetrix Inc. which uses a chrome-mask based technique for industrial-scale fabrication of DNA Chips, NimbleGen Systems Inc. and Febit biotech GmbH employ the light-directed synthesis process. NimbleGen uses a DMD-based (maskless) technique for fabrication of customized chip designs. Febits DMD-based Geniom<sup>®</sup> system is an integrated platform for customized microarray fabrication and analysis.

Academic research on light directed *in situ* synthesis is performed in the groups of F. Cerina (University of Wisconsin, Madison) and H.R. Garner (Southwestern Medical Center, University of Texas, Dallas).

Our demand for customized (though still affordable) DNA microarrays required the development of a DMD-based *microarray synthesizer* system, similar as described by Singh-Gasson *et al.* [SG99].

The synthesis apparatus (Fig. 3.1) was constructed from the following components:

- DLP™ video projector A+K AstroBeam 540 (Anders+Kern), SXGA resolution (1024×768 pixels)
- Inverted microscope Axiovert 135 (Zeiss)
- DNA Synthesizer ABI 381A (Applied Biosystems)
- Personal Computer (Pentium III, 1.4 GHz, 512 MB RAM, dual-head graphics card)



**Figure 3.1:** Schematic of the DNA Chip synthesizer system. (1) The maskless microscope projection lithography system comprises the DLP video projector (2) with the Digital Micromirror Device (DMD) (3). It also includes the microscope optics (4) for UV pattern projection onto the substrate surface inside the synthesis cell (5). Mask projection is controlled by the synthesis control software running on a personal computer (6). The software simultaneously controls the fluidics system via the solenoid valve driver (7). The valve block (8) of an ABI 381 DNA synthesizer is employed for reagent delivery.

The video projector and the microscope have been integrated into the "Maskless Microprojection Photolithography System" [Nai06b]. The valve block of the DNA synthesizer constitutes the main component of the fluidics system. A personal computer (running the Java-based DNA synthesizer control software *DNASyn* - see section B.7) is used for synchronized fluidics control and "virtual photomask" projection, thus enabling a fully automated synthesis.

## 3.2 The Maskless Microprojection Photolithography System (MPLS)

The basic idea of MPLS is to use a spatial light modulator for displaying computer generated "virtual photomasks" in the focal plane of a microscope. For that purpose we make use of a DMD spatial light modulator and its driver electronics, both obtained from a secondhand commercial video projector (Anders+Kern AstroBeam 540 - which is similar in construction with the DAVIS DL X10). The optics of a Zeiss Axiovert 135 inverted microscope is employed in a reverse optical path for image projection: Using a  $5\times$  (0.25 NA) Fluar microscope objective (Zeiss), the image of the DMD - which is located in the intermediate image plane - is scaled down to  $3.5\text{ mm}\times 2.6\text{ mm}$ . As a light source for visible and near UV wavelengths we employ a 250 W Ultra High Pressure (UHP) mercury arc lamp also from a video projector (Optoma EP758). The hardware described in the following sections is illustrated in Fig. 3.2. Components shortcuts refer to annotations on Fig. 3.2(b).

### 3.2.1 The UV Light Source

Ultra High Pressure (UHP) mercury arc lamps [Der05] are optimized for use in video projection systems.

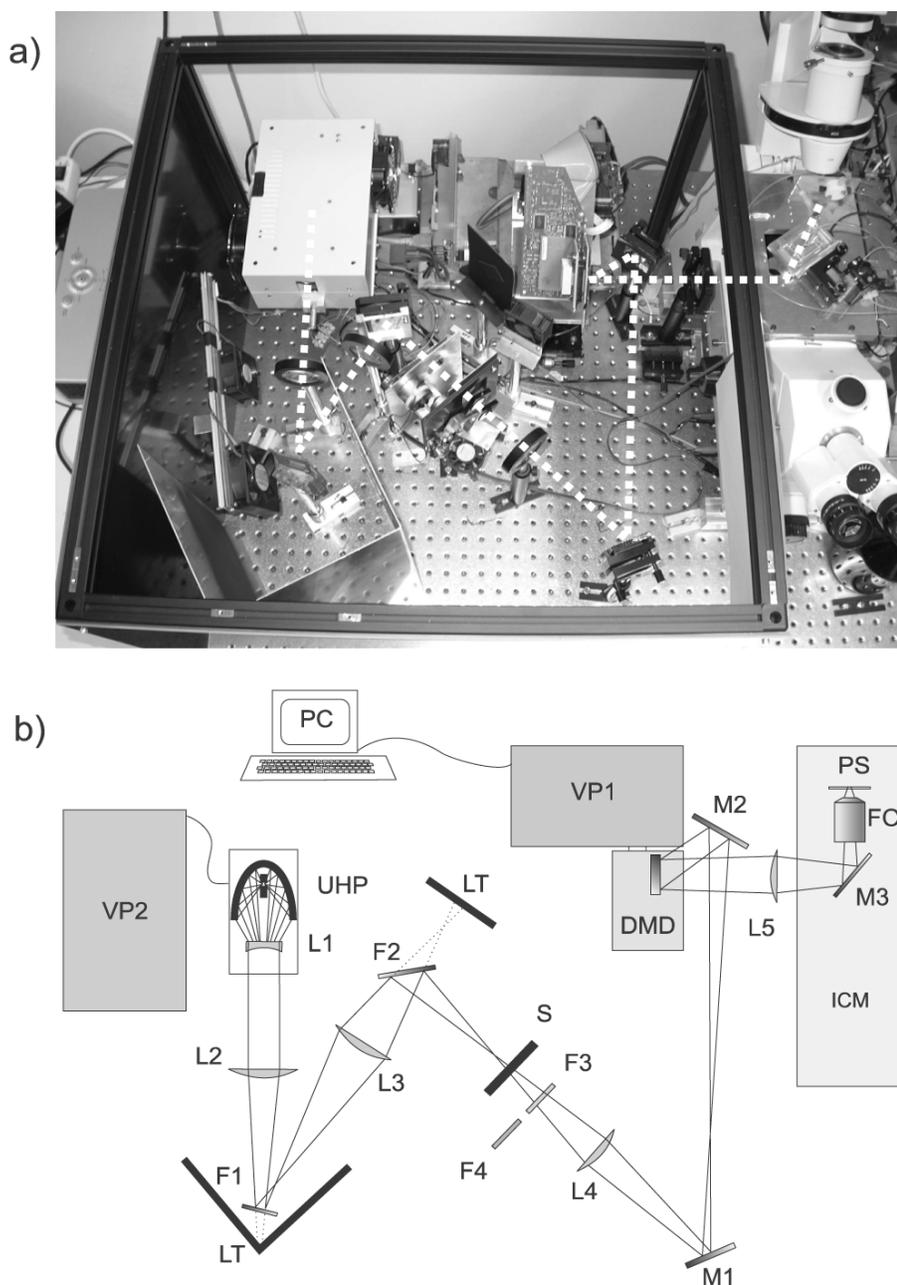
In our lithography setup the use of a UHP lamp has several advantages over a conventional mercury arc lamp. Compared to the latter, the UHP lamp has a smaller and brighter arc region (with an arc gap of typically 1-1.5 mm), resulting in an increased light throughput. A short arc length is very important for low *etendue*<sup>1</sup> projection systems with a small display size (like the DMD) [Der05]. Only a fraction of the light emitted from the lamp can be transmitted through the optical system. For a high efficiency the luminance ( $\text{cd}/\text{m}^2$ ) of the arc should be as high as possible.

Further advantages of the UHP lamp are a high arc stability and a very long life time of up to several thousand hours.

The operation pressure of up to 300 bar causes considerable line broadening - resulting in an almost continuous spectrum. This is advantageous for video display applications, but has to be considered (e.g. in the band with of interference filters, chromatic aberration) in photolithographic applications.

Initially we used the 120 W UHP lamp of the AstroBeam projector. To increase the light intensity (in order to reduce exposure times) we replaced this lamp by a more powerful

<sup>1</sup> The *etendue* - also called *optical invariant* - describes the capability of an optical system to conduct light.



**Figure 3.2:** (a) Photograph of the maskless microscope projection photolithography system (top view). Along the optical path (dotted white line): UHP lamp housing, UV cold mirrors, shutter, band pass filters (green and UV), DMD and driver electronics, tube lens, microscope (Zeiss Axiovert 135) and the reaction cell, which is mounted onto the sample holder.

(b) Drawing of the lithography system: Ultra High Pressure lamp (UHP) powered by video projector (VP2), plano-concave silica lens (L1), plano-convex lens (L2), UV cold mirror F1, light trap (LT), plano-convex lens (L3), UV cold mirror (F2), shutter (S), bandpass filters for UV (F3) and green (F4) illumination, plano-convex lens (L4), fold mirrors (M1 and M2), DMD and driver electronics of the AstroBeam projector (VP1), tube lens (L5), infinity corrected microscope (ICM), mirror/beamsplitter-assembly (M3), 5× (0.25 NA) Fluar microscope objective (FO), substrate to be patterned (PS). Technical details are provided in Appendix B.3.

250 W UHP lamp of another video projector (Optoma EP 758). Since UHP lamps require specialized power supplies (integrated in the video projector) the Optoma projector is now employed as a lamp power supply.

Due to the requirement for high UV transmission we couldn't use the highly optimized optics<sup>2</sup> of the video projector. For the photolithography system a new UV illumination optics had to be designed: the lamp module for the Optoma EP758 projector was built into an air cooled housing and connected via an extension cable to the lamp driver of the Optoma projector (VP2). The arc of the 250 W UHP lamp is located at the inner focal point of the elliptical lamp reflector. To efficiently collimate the strongly divergent beam, a plano-concave diffraction lens (L1) (f=50 mm, 25.4 mm diam., fused silica) is placed between the lamp window and the outer focal point of the reflector.

Efficient filtering of the near UV wavelength band required for the photo-deprotection reaction proved to be difficult owing to the high thermal load. Filtering is therefore performed in several steps: A dichroic filter from the Optoma lamp module (F1) (originally designed as a UV protection filter) is employed as a UV cold mirror to cut down the visible light intensity to about 10 percent. UV light below 400 nm is efficiently reflected.<sup>3</sup> Infrared radiation is filtered using another UV cold mirror (Oriel) (F2). Finally a band pass interference filter (F3) (bk-370-35-B, Interferenzoptik Elektronik GmbH) is used for selecting the wavelength band in the mercury i-line region ( $\lambda = 365$  nm) required for the photo-deprotection reaction. Taking into account that the mercury i-line is considerably broadened due the high operation pressure of the lamp, we had to use a relatively wide band pass filter (peak transmission  $T_{max} = 60\%$  at 370 nm, FWHM: 33 nm) to achieve a sufficiently high UV transmission. Use of a broadband filter (color glass UG-5, Schott, transmission between 230 and 430 nm and above 650 nm,  $T_{max} = 90\%$  at 350 nm ) would result in severe chromatic aberration.

### 3.2.2 Digital Mask Projection Using a Digital Micromirror Device

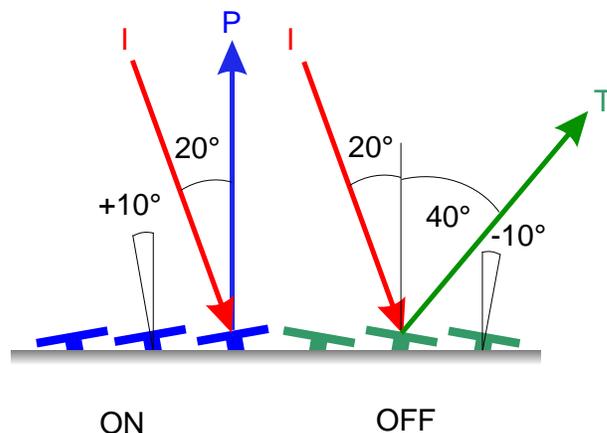
The DMD is a spatial light modulator commonly used for image generation in DLP video projection systems (for technical details on DMD technology see section B.1). In our setup we use a DMD with XGA resolution containing  $1024 \times 768 = 786432$  square mirrors (16  $\mu\text{m}$  in size with a pitch of 17  $\mu\text{m}$ ) that can be tilted by an angle of  $+10^\circ$  or  $-10^\circ$  relative to the

---

<sup>2</sup> Optimized for high light throughput and uniformity of illumination.

<sup>3</sup> The transmission spectrum of the dichroic filter shows a distinct cutoff at 415 nm - from 420 to 700 nm the transmission is  $\geq 90\%$ . The reflectivity in the i-line range couldn't be measured with the spectrophotometer available. However, a simple experiment with a 100 mW UV-LED (Nichia NCCU033) shows that the UV reflectivity is (coarsely estimated) between 60 and 80%.

normal axis of the chip. The two positions are referred to as on- and off-state: Mirrors in the on-state reflect the incident light perpendicular to the DMD surface into the projection optical system, whereas mirrors in the off-state reflect light at an angle of  $40^\circ$  relative to the DMD normal axis into a light trap (Figure 3.3).



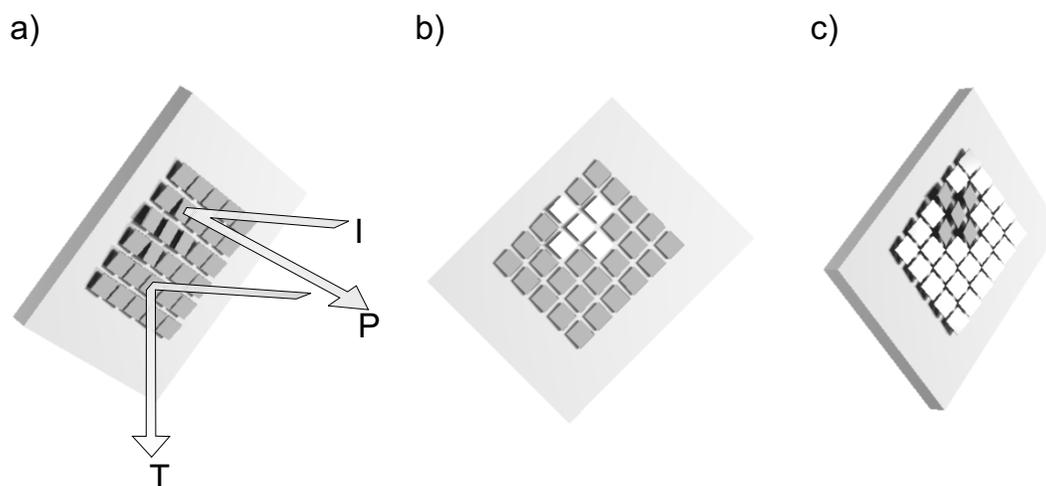
**Figure 3.3:** Spatial light modulation with a Digital Micromirror Device. Mirrors in the on-state (blue) reflect the incident light (I) in a direction normal to the DMD surface into the projection optics (P). Mirrors in the off-state (green) reflect the light under an angle of  $40^\circ$  with respect to the normal axis into a light trap (T).

The DMD is oriented perpendicular to the optical axis of the projection system. The micromirrors tilt around their diagonal axis. We have rotated the DMD by  $45^\circ$  around the optical axis, so that the incident beam and the reflected beam lie both in the horizontal plane of the setup (Figure 3.4).

Technical details on the modification of the DLP video projector are provided in Appendix B.2. For better accessibility of the micromirror array the DMD board had to be removed from the projector chassis and reconnected to the driver board via a 148 pin extension cable. Because the driver electronics of the projector remains unchanged, all sorts of video signals can be used to control the image display. Connection to a PC with a dual-head graphics card proved to be useful, as one screen can be used for control purposes (e.g. for running the DNA synthesis control program which automates and coordinates photolithographic pattern display and the fluidics system) while the other one is reserved for pattern display.

### 3.2.3 The Image Projection Optics

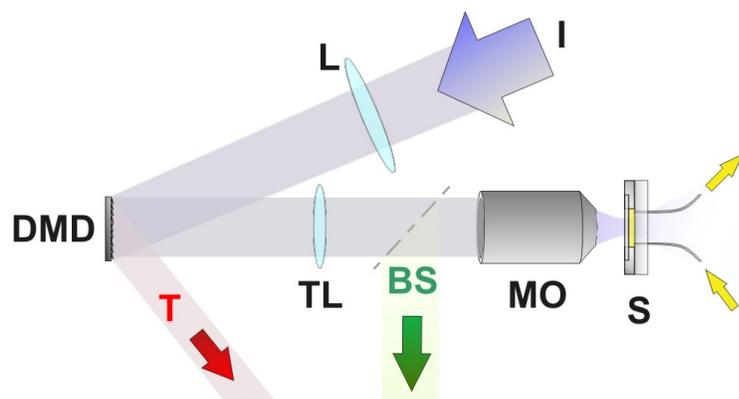
To reduce the microarray size to a few  $\text{mm}^2$  we opted for a microscope projection approach. Reduced dimensions of the microarray are beneficial for microarray hybridization due to reduced diffusion times [Dan07] and reduced material requirements (synthesis reagents



**Figure 3.4:** Rotated DMD arrangement in the maskless microscope projection lithography setup. The DMD is rotated by  $45^\circ$  around the optical axis, so that the tilting axis of the mirrors is vertical. The incident beam and the reflected beam lie both in the horizontal plane of the setup. (Left image) From the mirrors in on-position (arranged as a "X") the incoming light (I) is reflected towards the projection optics (P). Mirrors in the off-position reflect the light into a light trap (T). (Center image) View of the DMD from the projection optics. (Right image) View of the DMD from the light trap.

and nucleic acid sample size). By reducing the image area, the illumination intensity is increased by a similar factor: a 250 W UHP lamp does suffice in order to keep the time required for optical deprotection in a reasonable relationship to the total turnover time of the chip synthesis. Use of the microscope also provides superior control of the image focusing and mechanical stability. Image drift occurring from thermal expansion of the optical parts has previously been described as serious problem in the light-directed synthesis process, requiring active control of focusing, e.g. by means of an image-locking technique [Ric04]. An important aspect in the design of the lithography system is image contrast. In light-directed microarray synthesis stray light is much more critical than for example with photoresist. Photoresist, having a strong nonlinear exposure characteristics, doesn't respond to small stray light intensities below a threshold value. In microarray synthesis there is no threshold and stray light induced errors can accumulate over many exposure steps. Within the total exposure time of about two hours, stray light causes base insertion errors, affecting most of the synthesized DNA strands.

The whole synthesis process involves about 80 exposures with different mask patterns, it extends over about 6.5 hours. Mask alignment requires thermal and mechanical stability. To make use of the maximum pixel resolution of the setup (which is  $3.5 \mu\text{m}$  with a  $5\times$  microscope objective) no movements caused by vibrations, tension release, or thermal expansion larger than about  $1 \mu\text{m}$  (in the front focal plane of the objective) can be tolerated.



**Figure 3.5:** The projection optics system. Incident light (I) (filtered - either near UV for the exposure or green for focusing); light trap (T); tube lens (TL); beam splitter (BS); microscope objective (MO); synthesis cell (S).

The micromirror array (DMD) is placed in the image plane (located outside the microscope frame) of the inverted microscope. With infinity corrected microscope objectives, a tube lens (TL) is necessary to project the image of the DMD to infinity. The adjustment of the distance between DMD and tube lens, which does not exactly equal the nominal focal length of 164.5 mm (as specified by the manufacturer), is crucial for the calibration of the setup, as explained later (in Sec. 3.2.5).

A movable half mirror/half beamsplitter optical element (BS), located at the position of the microscope's fluorescence filter block, is used to reflect the light into the objective back aperture. Using the beamsplitter part, the light reflected from the surface of the microarray substrate can be coupled into the microscope. This is employed for exact focusing and direct observation of the projected image through the eyepiece. For photopatterning, the mirror part is used (exchange is achieved by sliding the plate by hand). In principle for this purpose a dichroic beamsplitter (reflection of UV light and reduced reflection of visible light) could be used. However, the use of a beamsplitter plate (for photo-protection) turned out to be problematic since even a small amount of reflection at the backside of the plate can produce ghost images, and thus significantly affect the image contrast.

Among several objectives (MO) tested, we found the Zeiss Fluar 5 $\times$ (0.25 NA) as most suitable for DNA chip fabrication, particularly for its superior UV transmittance and its large back aperture allowing for efficient light collection. Over a working distance of 12.5 mm the image of the DMD is projected onto the DNA synthesis substrate - a chemically functionalized glass surface - inside the synthesis cell (S).

A 10 $\times$ (0.30 NA) Plan Neofluar and a 20 $\times$ (0.5 NA) Plan Neofluar objective (Zeiss) were successfully used to further reduce the image size. Diminished contrast makes these ob-

jectives less suitable for light directed microarray fabrication. However, patterning of photoresist - having lower requirements on contrast - should be simple with these higher magnification objectives.

At a wavelength of 365 nm the diffraction limit of the  $5\times(0.25 \text{ NA})$  Fluar objective is  $R = \lambda/(2 \cdot \text{NA}) = 0.73 \text{ }\mu\text{m}$ . However, a significantly larger distance between adjacent features is necessary to achieve a sufficient local contrast for the light-directed fabrication process.

Reflective objectives have the advantage of a high UV transmission and are not subject to chromatic aberrations. We therefore tested image projection with a  $15\times(0.28 \text{ NA})$  Schwarzschild type reflective objective (Ealing). However, a satisfactory image contrast over the whole field couldn't be achieved. Also, in the given optical system, owing to a narrow back aperture, the light throughput through the reflective objective is very limited.

### 3.2.4 Fabrication and Application of UV-Sensitive Photochromic Films

For evaluation of the imaging quality a fast and simple method for generating patterns upon UV exposure is required. Photographic films and photoresist turned out to be not very useful due to difficult handling and processing efforts. Therefore we have developed a UV-sensitive film based on the photochromic dye spiropyran. Spiropyran undergoes a structural change when exposed to UV-light. This results in a strongly increased light absorption in the visible range.

#### Preparation of photochromic films:

We dissolved 10 mg of spiropyran dye (1',3'-dihydro-1',3',3'-trimethyl-6-nitrospiro [2H-1-benzopyran-2,2'-(2H)-indole], Aldrich, Cat.: 27,361-9) in 1 ml of PMMA photoresist (E-beam resist PMMA 200 k; AR-P 641.04, Allresist GmbH, Strausberg, Germany) and spincoated a thin film (thickness about 1  $\mu\text{m}$ ) onto a microscope slide. Other resists - we also tried with MicroChem PMMA and MicroChem SU-8 50 - work equally well. The photoresist is used as a carrier material only. After spincoating, and brief heating on a hot plate (1 minute at 100°C) the slides are ready for use.

We found these photochromic films to be a well-suited imaging material. Unlike with photoresist or photographic material no developing or other processing is required. Under UV exposure the film changes from transparent to an almost opaque purple. With the intensities we usually apply (50-100  $\text{mW}/\text{cm}^2$ ) this happens within seconds. The process can be reversed by heating or by illumination with bright light (at visible wavelengths).

Unless the spiropyran has been bleached with high irradiation doses, the films can be reused several times.

For a small exposure dose the optical density increases almost linearly with the dose of UV light. For larger doses  $D$  the optical density  $OD$  approaches saturation.

$$OD = OD_{sat}(1 - \exp(-\text{const.} \cdot D)) \quad (3.1)$$

Upon very high exposure, photodegradation of the photochromic dye (bleaching) results in reduced  $OD$  values. Since the linear exposure characteristics of the spiropyran dye are very similar to that of NPPOC phosphoramidite reagents, spiropyran films are a very useful tool for testing and evaluation of the UV optical system.

### 3.2.5 Chromatic Correction of the Projection Optical System

Since the depth of focus  $DOF = \lambda/NA^2$  is only about  $6 \mu\text{m}$  for the  $5 \times (0.25 \text{ NA})$  Fluar objective (at  $\lambda = 365 \text{ nm}$ ), it is necessary to perform proper focusing each time a new patterning substrate is mounted on the sample holder. The focus range providing optimum contrast is even smaller than the depth of focus, thus perfect focusing of the pattern onto the surface is crucial. It can be achieved by observing the back reflection of the projected image (from the patterning surface) through the microscope eyepiece. This is easy to perform with visible light, but rather difficult with UV light.

If the back-reflected image of the pattern is perfectly focussed in visible (green) light, this usually is not true for UV at the same time. This is owing to chromatic aberration. Longitudinal chromatic aberration causes an axial focus shift usually resulting in a completely blurred image in UV. In the following we describe a method for the correction of this longitudinal chromatic aberration, so that focusing of the near UV image can be performed by observation (through the eyepiece) and focus adjustment under green light illumination.

Using photochromic films as a control for the quality of the projected UV pattern, we found that the chromatic aberrations can be compensated by fine-adjustment of the distance  $d$  between the DMD and the tube lens (see Fig. 3.2). The distance  $d$  is roughly the nominal focal length of the tube lens of  $164.5 \text{ mm}$ . After focusing with green light, the film is exposed with a control pattern in UV and subsequently inspected on a light microscope. The distance  $d$  now can be adjusted iteratively until the patterns imaged on the spiropyran slide indicate perfect focusing. Just a small deviation of a few millimeters from the nominal focal length of the tube lens is necessary for chromatic correction. The tolerance of  $d$ , within which a good correction is achieved, is only a few tenths of a millimeter wide. Once the

chromatic correction procedure has been accomplished, focusing can always be performed under illumination with green light.

Caution! The above optical adjustment depends on the eye focal length of the experimenter who performed the adjustment. In daily use of the microarray synthesizer, when focusing on the microarray substrate is performed, deviating eye focal lengths (near/far sightedness) of other personnel using the equipment do matter and need to be accounted for.

### 3.2.6 UV Light Intensity and Uniformity of Illumination

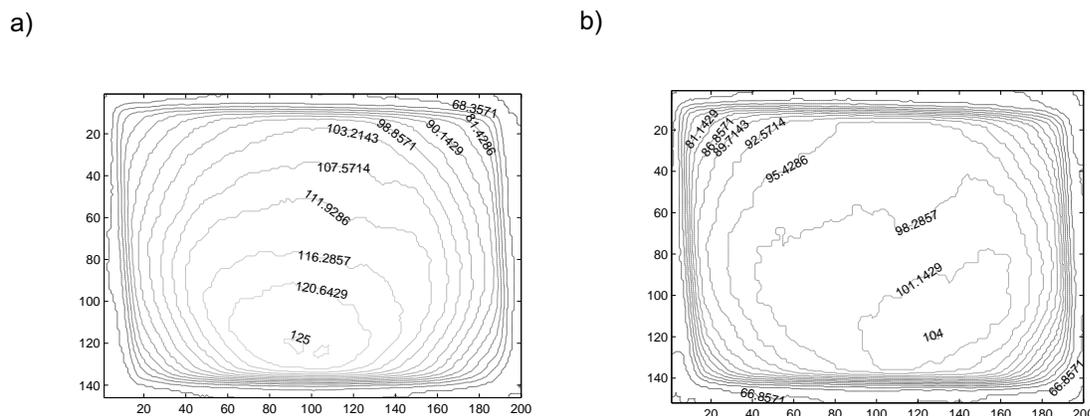
For measuring the intensity at the image plane we used a laser power sensor (PS10Q, Coherent Inc.). The thermopile sensor was placed in the focal plane of the microscope objective. To measure the mean intensity, a completely white image was displayed on the DMD. With the measured total power of 7.8 mW we determined the intensity in the image plane as 87 mW/cm<sup>2</sup>.

To study the uniformity of the illumination we projected the image onto a screen. The intensity was measured at different regions of the projected image. An asymmetric large scale deviation with a peak intensity of about 140% of the mean intensity is observed. This is due to the configuration of the illumination system: The UHP lamp's arc gap is oriented parallel to the optical axis, providing a very inhomogeneous illumination profile. For this reason in a video projection system an integrator element, e.g. an integrator rod (which is a light guide with a rectangular cross section) or a fly-eye lens array is employed to generate a very uniform illumination. Using the integrator rod of the AstroBeam projector turned out to be not feasible as the glass rod absorbs most of the UV light.

We decided to flatten the illumination profile by using only a small homogeneous section of the light cone for illuminating the DMD. This way we sacrifice about 80% of the light. Nevertheless, the remaining 20% of light allow photo-deprotection to be performed in a reasonable time. Alternatively, if such parts were available, a quartz integrator rod or an integrator plate (fly-eye lens array [Sun05]) could be used to achieve significantly higher light intensities.

To attain a more uniform illumination we employ the DMD for intensity leveling, similar as described by Huebschman *et al.* [Hue04]. For this purpose we have created an "intensity leveling mask". The black and white images (to be used as a photomasks) can easily be leveled to reduce intensity variations to about  $\pm 10\%$  by pixelwise multiplication with this mask. To generate the intensity leveling mask, a fully illuminated image (all mirrors in the on-state) is projected on the screen (as described above, still without using the microscope objective) and photographed with a Nikon Coolpix 4500 digital camera. Deskewing the

raw image using standard image processing software results in a  $1024 \times 768$  pixel image, which finally has to be inverted and adjusted in brightness and contrast. The leveling mask is then projected onto the screen and a photometer is used to measure uniformity of illumination. In an iterative way image brightness and contrast are adjusted to achieve a uniform intensity within most of the image area. Contour plots of the light intensity before and after intensity leveling are shown in Fig. 3.6. Only in the outermost corners of the image (comprising about 10% of the total image area) the intensity is reduced to about 50% of the mean intensity. This is due to vignetting: Light reflected from the corners of the DMD, which are located close to the edge of the entrance pupil, is partially blocked by the apertures of the tube lens respectively the microscope objective. Applying intensity leveling we achieved a mean light intensity of  $76 \text{ mW/cm}^2$ .



**Figure 3.6:** Uniformity of illumination. (a) Intensity contour map before intensity leveling. (b) After intensity leveling. Using the tube lens, the image of the DMD was projected onto a screen, without the microscope objective in place, and photographed with a digital camera. Vignetting from the microscope objective is neglected here but this effect is small compared to vignetting of the tube lens.

The intensity values mentioned above were achieved using an interference filter with a FWHM of 33 nm and a maximum transmission of 60% at a center wave length of 370 nm. Using a narrow i-line filter (FWHM 12 nm at a center wavelength of 365 nm; 35% maximum transmission) provided significantly lower intensities (about one ninth of the intensity achieved with the broad filter). The demand for a wide filter can be explained by the strong line broadening due to the high operation pressure of the UHP mercury arc lamp.

### 3.2.7 Optical System Performance Testing with UV-Sensitive Photochromic Films

Light-directed synthesis of DNA microarrays requires that the image is projected onto a substrate inside an inert reaction chamber, so that reactions can take place under a moisture free argon atmosphere. The synthesis substrate, a 0.17 mm thickness microscope cover glass, is forming the window of the reaction cell. Hence the image has to be projected onto the inner face of the window. For image focusing (see Sec. 3.2.5) we use the small fraction of green light which is reflected back from the imaging surface into the microscope. Applying a similar approach for contrast measurement is not practicable because the outer face of the cover glass contributes to back-reflection as well. Multiple reflections in the microscope system (e.g. from a beamsplitter) may degrade the image contrast further.

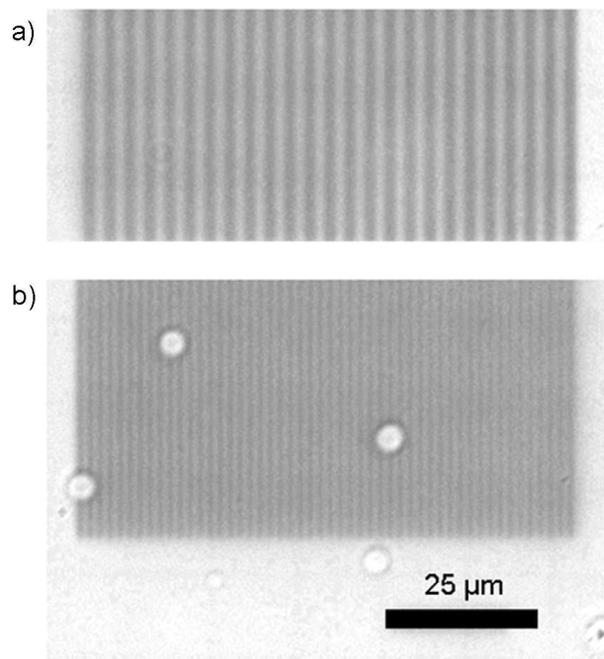
Contrast ratios of 1:3000 (as can be found in product specifications of video projection systems) usually refer to the full-on/full-off contrast obtained by comparing the intensities of completely black respectively white images. On our setup (placing a photometer into the focal plane of the microscope objective) we measured a full-on/full-off contrast of about 3400:1. This means that the DMD chip with the mirrors in the off-position reflects only about 0.03% of the exposure intensity onto the imaging substrate. This means that the amount of light scattered by the DMD housing and by the mirrors in the off-position is negligible.

Much more relevant for DNA microarray synthesis is the local contrast [Kim04] between neighboring features. The local contrast is diminished by light-scattering and diffraction from mirrors in the on-state, but also by optical aberrations, which cause distortions to the point spread function. It also depends on the feature geometry (i.e. feature size and feature spacing). Reflections within the imaging optics cause flare. This could possibly be improved by using UV anti-reflection coated optical surfaces (DMD window, tube lens).

The patterns used for microarray synthesis typically have an array structure with a pitch of 17  $\mu\text{m}$  or less. To obtain an estimate of the stray light induced error rate we have measured the image contrast at high spatial frequencies.

We found that the UV-sensitive films we already used for adjustment of the UV optics (see section 3.2.5) are very well suited for testing the performance of the photolithography system. For visual inspection of the patterns we used an optical microscope (Olympus IX81) equipped with an automated X-Y translational stage and with a high resolution CCD camera (C9100 EM-CCD, Hamamatsu Photonics).

Patterns of regularly spaced line pairs (a pair comprises a black and a white bar of equal width), were imaged onto photochromic film (Fig. 3.7). The spatial frequency of the pattern was varied between 14 and 70 line pairs per millimeter (lp/mm). Using an exposure



**Figure 3.7:** Line patterns on photochromic film produced with a 20× (0.5 NA) Plan Neofluar objective. (a) Linewidth 1.7  $\mu\text{m}$  (corresponding to a double row of micromirrors) (b) Linewidth 0.85  $\mu\text{m}$  (corresponding to one single line of micromirrors).

time scalebar (like in Fig. 3.9, optical density versus exposure time<sup>4</sup>) allowed us to quantify the stray light intensity in the unexposed lines. To obtain the relative stray light intensity (in percent of the exposure intensity - see Table 3.1) the optical density was compared to the exposure scale. The ratio between the exposure time and the equivalent (stray light) exposure time is a measure for the contrast between exposed and unexposed lines.<sup>5</sup> The

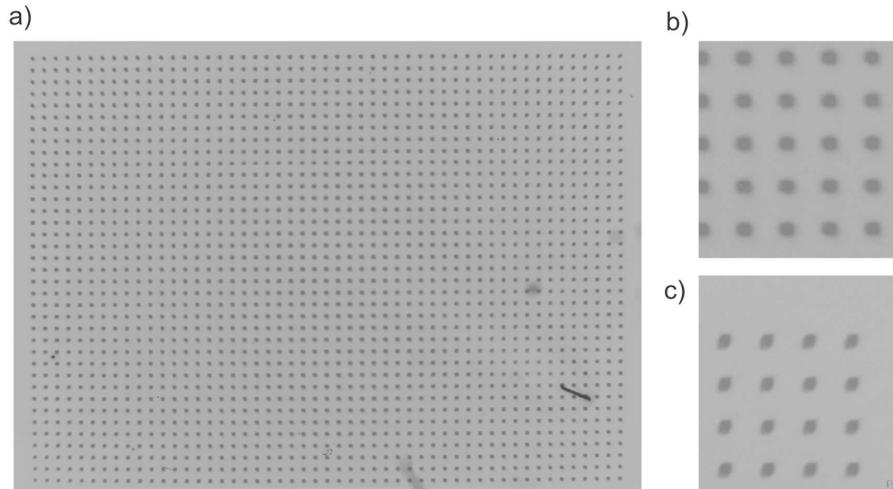
spatial frequency (line pairs/mm)	relative stray light intensity (percent)
70	10
35	5.5
28	3.2
21	2.2
14	0.5

**Table 3.1:** Relative stray light intensities versus the spatial frequency of the line pattern. Stray light intensities were measured at the center of the unexposed lines.

<sup>4</sup> For a direct comparison the scalebar was imaged onto the substrate - next to the line patterns.

<sup>5</sup> Imaging contrast is best described by the *modulation*  $m = (I_{max} - I_{min}) / (I_{max} + I_{min})$ . However, here we cannot determine the *modulation* since the saturation of the optical density doesn't allow a measurement of the equivalent exposure time of the exposed lines (corresponding to  $I_{max}$ ), which, due to loss of light into unexposed lines, is smaller than the actual exposure time.

stray light intensity increases towards higher spatial frequencies, and is therefore setting an upper limit for the feature density in the light-directed microarray synthesis.



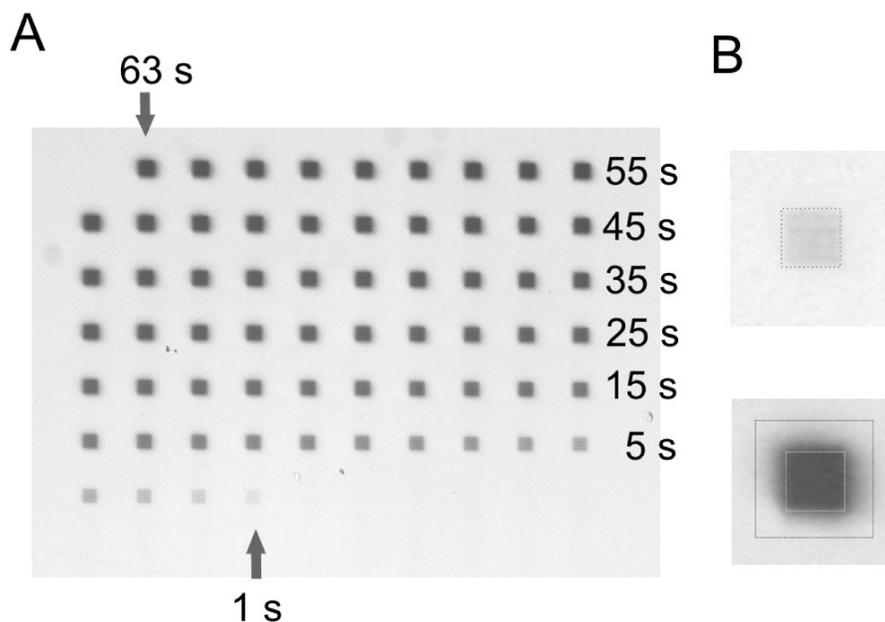
**Figure 3.8:** Stray light due to curvature of field. (a) A test pattern of  $4 \times 4$  pixel squares (pitch 20 pixels) covering the whole DMD area imaged onto photochromic film. The montage of several micrographs covering an area of about  $3.5 \times 2.6$  mm, doesn't show a macroscopic image distortion. (b) Close-up view of the center region. The response of the photochromic material is asymptotically saturating in the center of the features. Stray light produces a halo around the features. (c) Close-up of the upper right corner of the imaging field. The square features are radially distorted.

To demonstrate the effects of optical aberrations on the imaging performance, a pattern comprising of  $4 \times 4$  pixel features (with a pitch of 20 pixels) was imaged onto photochromic material. A radial distortion of the square features is recognizable in Fig. 3.8(c), which was taken at the upper right corner of the imaging field (Fig. 3.8(a)). As non-corrected curvature of field is supposed to be responsible for the distortion, we tried to improve image quality using a plan-corrected microscope objective. This, as well as using a narrower band pass filter to reduce chromatic aberrations didn't significantly improve imaging quality. The increased number of lens elements in the plano-corrected objective (with respect to the Fluor objective) significantly reduced the UV intensity. Another possible source for contrast impairment is the illumination system, which has been designed for a high light throughput. It may be possible to improve the optical aberrations, if this constraint is relaxed.

We found that using higher magnification objectives is possible. Using a  $20 \times$  (0.5 NA) Plan NeoFluar (Zeiss), the total image size is reduced to  $0.87 \times 0.65$  mm<sup>2</sup>. As shown in Fig. 3.7, spatial frequencies of 588 lp/mm (line width 0.85  $\mu$ m) can clearly be resolved on the photochromic film. Due to reduced depth of focus, and non-corrected field curvature, this resolution can only be achieved in the center of the imaging area. At high spatial

frequencies the measured contrast is additionally reduced due to the modulation transfer function of the inspection microscope optics. Therefore the contrast observed in Fig. 3.7 represents a lower limit.

Fig. 3.9 shows an exposure scale.  $4 \times 4$  pixel features<sup>6</sup> were illuminated with exposures ranging from 1 to 63 s. The optical density of the stray light halo around the features is compared with the exposure scale to determine the local contrast ratio. In an approx. 2 pixel wide region around the exposed features the local contrast is significantly impaired (Fig. 3.9 B). Near the outer edge of this region the stray light intensity is estimated to be around 2 %.

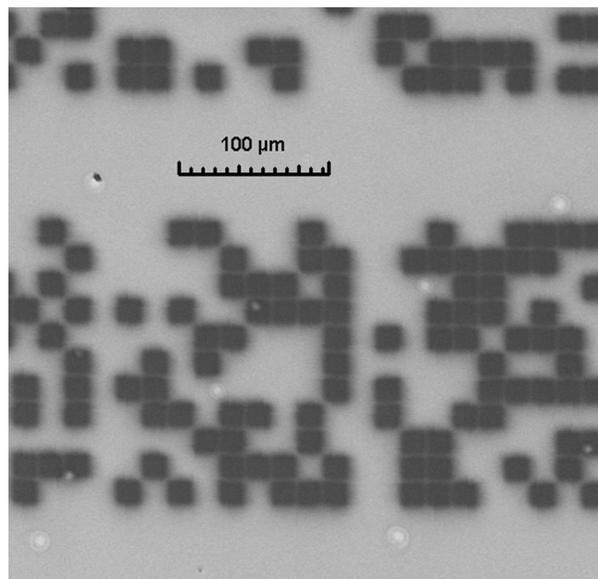


**Figure 3.9:** The magnitude of the local stray light intensity can be estimated from this pattern of  $4 \times 4$  pixel features (size  $14 \mu\text{m}$ ) from an exposure on spiropyran photochromic film. (A) The exposure was varied between 1 and 63 seconds. (B) Enlarged view of two features corresponding to exposures of 2 s (top) and 63 s (bottom). The stray light is concentrated in an asymmetric halo around the features. The dashed boxes show the directly exposed feature area (inner box) and the surrounding area (outer box) that is affected by stray light.

To determine the stray light as it occurs during typical DNA microarray fabrication, a synthesis mask pattern is projected onto the photochromic film (Fig. 3.10) with a fill factor (fraction of illuminated features) of about 25%. We measured that (on average) the stray light intensity reaching the center of unexposed features is on the order of 0.5% of the exposure intensity. For unexposed features completely surrounded by exposed features we have measured a stray light intensity of 1.5%. These values were measured in the center of

<sup>6</sup> "Pixel" corresponds to pixel in the "virtual photolithography mask" image. Upon projection with the DMD each pixel of the mask image corresponds to an individual micromirror.

the unexposed features - at the edges of the features the stray light intensity can be significantly higher.



**Figure 3.10:** A synthesis mask pattern (4×4 pixel features, 1 pixel separation gap), as used in light-directed synthesis of microarrays, is imaged onto photochromic film. The stray light affecting unexposed features is determined by the density of exposed features nearby. Fig. 3.11 shows a similar pattern (same dimensions) which was imaged onto photoresist.

The photolithography-related parameters of the MPLS DNA synthesis apparatus are summarized in Table 3.2.

Imaging optics	5×0.25NA Fluar objective (Zeiss)
Exposure wavelength	370±17 nm
Exposure intensity	76 mW/cm <sup>2</sup>
Pixel resolution	1024×768 (XGA)
Pixel size	3.5 μm
Size of field	3.5 mm×2.6 mm
Drift stability	<1 μm over 6 hours
Time required for synthesis of a 25mer chip	ca. 6.5 hours
max. number of microarray features	25000 (4×4 pixel per feature and 1 pixel space)
Useful area for DNA synthesis	ca. 80 percent of the DMD imaging field
Reagent consumption for a 25mer synthesis	30-40 mg of each NPPOC-phosphoramidite

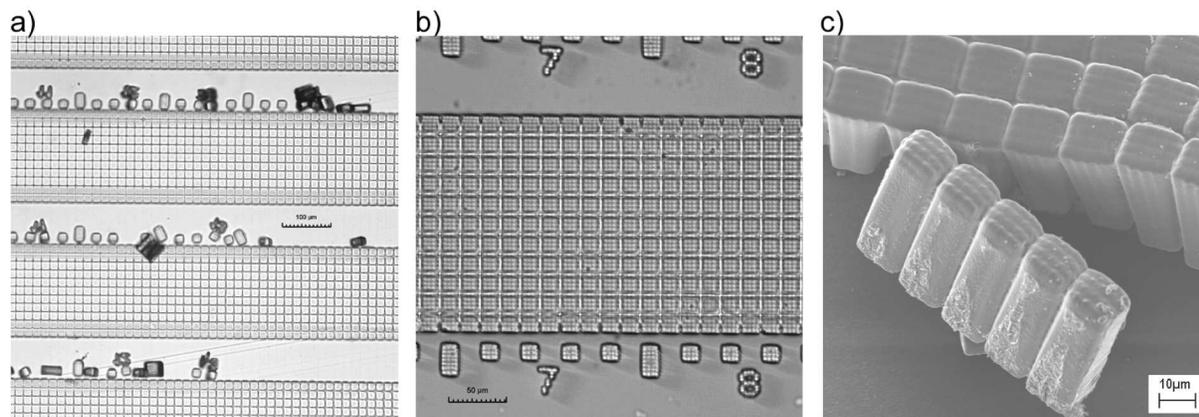
**Table 3.2:** Technical parameters of the MPLS photolithography setup

### 3.2.8 Outlook - Further Possible Applications for the Maskless Microprojection Lithography System

#### Photoresist Patterning

To demonstrate the application of MPLS as a highly flexible photolithography system, we have produced microstructures in SU-8 photoresist (Microchem Corp.). The SU-8 50 negative resist was spincoated on microscope slides at 3000 rpm, resulting in a film thickness of 40  $\mu\text{m}$ . For the experiment a pattern of tightly spaced square-features 14  $\mu\text{m}$  in size, separated by 3.5  $\mu\text{m}$  gaps was used. Processing was accomplished according to the processing guidelines of the manufacturer.

The photoresist, which is sensitive between 350 and 400 nm has been exposed to UV light for different times, ranging from 5 s to 25 s. After postbaking and developing the resulting microstructures were imaged using a research microscope as described above. We found 15 s to be an appropriate exposure time. Fig. 3.11 illustrates the high quality of the resulting microstructures. As one can see in Fig. 3.11(b), features as small as 3.5  $\mu\text{m}$  (the line width of the number "7") have been reproduced very well. The aspect ratio of the structures, as can be seen in Fig. 3.11, is about 1:10.



**Figure 3.11:** MPLS-generated pattern in SU-8 photoresist - film thickness: 40  $\mu\text{m}$ . (a) Each block corresponds to 4 $\times$ 4 Pixels, the separation gap is one pixel wide. Some of the letters are lying sideways on the surface, demonstrating that an aspect ratio of 1:10 is achievable with MPLS (scalebar 100  $\mu\text{m}$ ). (b) Micrograph of the same pattern as in (a). Even small features like the number "seven" with a line width of only 3.5  $\mu\text{m}$  (corresponding to a single micromirror, the pixelation is clearly visible) are reproduced in the photoresist (scalebar 50  $\mu\text{m}$ ). (c) Electron micrograph of the photoresist structures. Due to relatively poor surface adhesion of the photoresist the structures have partly detached from the glass surface (scalebar 10  $\mu\text{m}$ ).

### Optoelectronic Tweezers

Chiou *et al.* [Chi05] described a DMD based microprojection setup for massively parallel manipulation of single cells and microparticles. The image of a DMD spatial light modulator is projected on a photoconductive layer to create light-patterned electrodes. Dielectrophoretic forces resulting from the interaction of induced dipoles in the particles with the nonuniform electric field can be employed for particle manipulation. By variation of the mask geometry (dynamic masks) particles can be moved along with the "virtual electrodes". Parallel manipulation of up to 15,000 (largely independent) particle traps has been demonstrated.

### Projection Micro-Stereolithography

Sun *et al.* [Sun05] report a DMD-based method for the fabrication of 3D microstructures. In the fabrication process the image of the DMD is projected onto the surface of a UV curable resin. Three-dimensional objects (with a smallest feature size of 0.6  $\mu\text{m}$ ) are constructed layer by layer.

Possible applications of this technique are the fabrication of photonic crystals [Che07] and microstructured 3D scaffolds employed as substrates for tissue engineering [Lu06].

## 3.3 The Fluidics System

The modified valve block of a commercial DNA synthesizer (Applied Biosystems ABI 381A) constitutes the main component of the fluidics system (Fig. 3.12). A microcontroller-operated solenoid valve driver (technical details described in Appendix B.8) enables control of the fluidics system via the RS232-interface of the control PC.

To ensure water- and oxygen-free conditions the microarray synthesis is performed in an air tight flow cell (Fig. 3.13) under an inert argon atmosphere. Argon gas pressure is employed to drive the reagent transport. A detailed schematic of the fluidics system is shown in Fig. 3.12.

### 3.3.1 The Synthesis Cell

#### Technical requirements

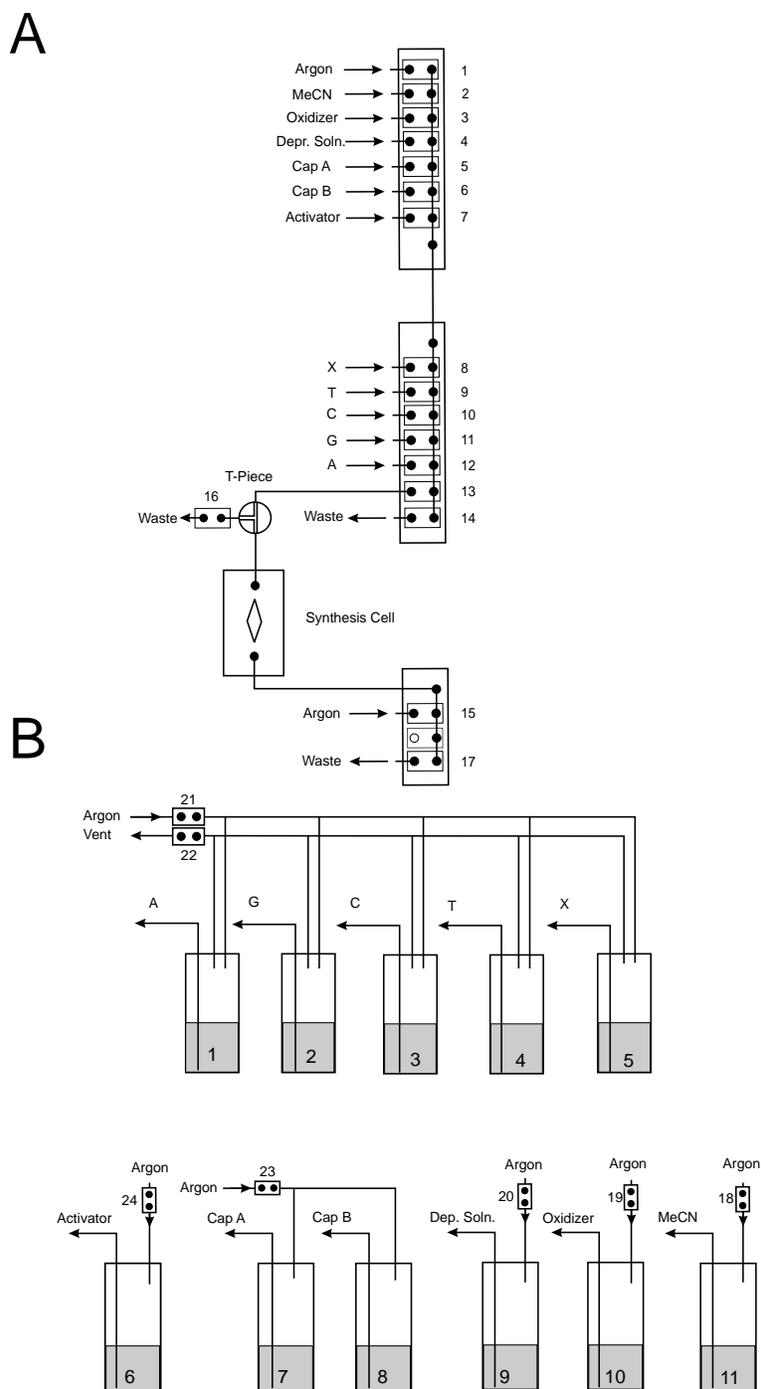
- resistance to the aggressive solvents MeCN, THF and pyridine
- use of chemically inert materials (not affecting DNA synthesis)
- tight sealing (no seeping of reagents below the gasket) is necessary to enable complete exchange of reagents (e.g. to avoid contamination with water left over from the previous reaction step)
- negligible dead volume (required for fast and complete exchange of reagents)
- the 0.17 mm microarray substrate must constitute a window of the cell
- prevention of gas bubble sticking at the edges of the cell volume
- light reflection and scattering must be avoided

#### Implementation

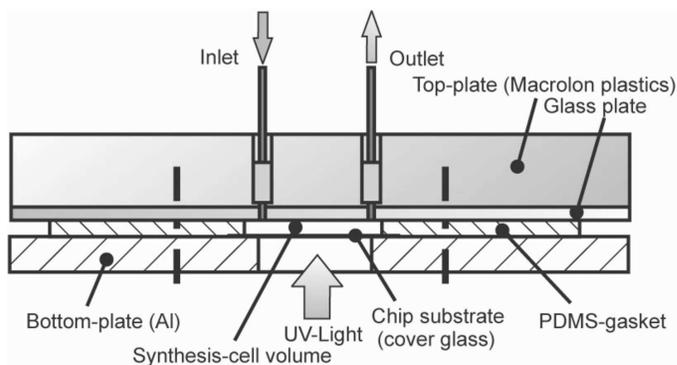
The cell volume is formed by a streamlined cutout (shown for example in Fig. 3.15) in an approx. 1 mm thick sheet of polydimethylsiloxane (PDMS) silicone rubber. PDMS is used for its chemical inertness and sealing capability.<sup>7</sup> The fragile microarray substrate (diam. 22 mm round cover glass) can be reliably sealed with little force, thus without the risk of fracture. The DNA chip substrate is employed as an optical window (Fig. 3.14). Photomasks are projected with a microscope objective onto the inner surface of the substrate, where the DNA probes are synthesized.

---

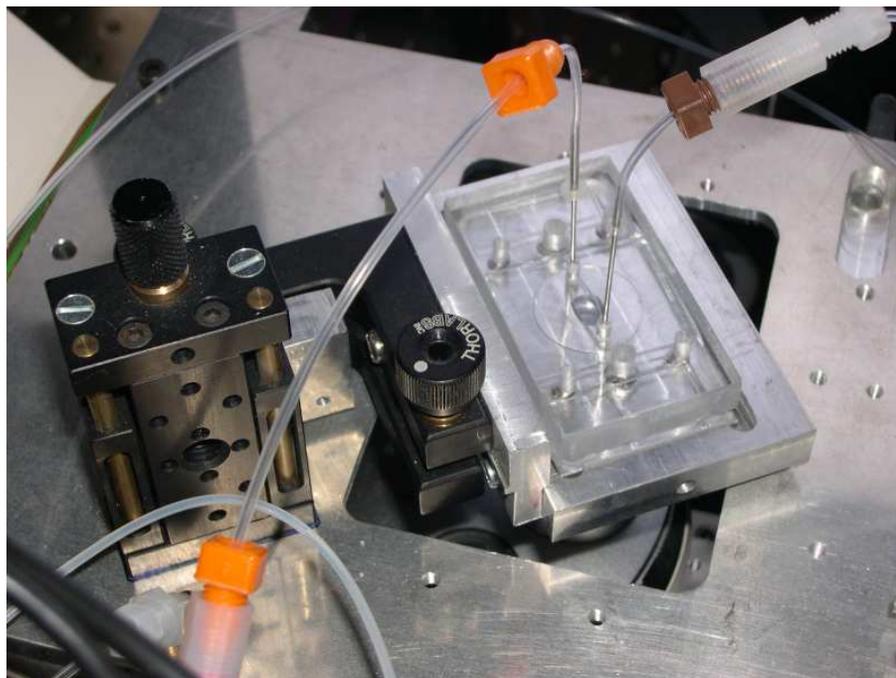
<sup>7</sup> Even though PDMS appears to be chemically inert, we observed (reversible) solvent swelling of the PDMS gasket upon exposure to tetrahydrofuran and pyridine. To prevent excessive deformation of the synthesis volume, oxidation and capping steps should not be longer than necessary. An alternative THF- (and water-) free oxidizer solution (enabling phosphoramidite synthesis on PDMS surfaces) has been described in [Moo05].



**Figure 3.12:** Schematic of the fluidics system. The valve block has been adopted from a commercial oligonucleotide synthesizer. The synthesis cell has replaced the synthesis column employed in standard oligonucleotide synthesis. Valve numbers correspond to those used in the synthesis control software. (A) Valve block. (B) Reagent storage bottles. Argon gas pressure (via valves 1, 15, 18-21, 23-24) is employed to drive the reagent transport.



**Figure 3.13:** Schematic of the synthesis cell. Syringe needles form the in- and outlet of the flow cell. A PDMS gasket with a streamlined cutout forms the synthesis cell volume which is sandwiched between the chip substrate and a glass plate. The assembly is placed on an inverted microscope. UV light from the objective is entering the cell through the chip substrate. Mask patterns are projected onto the inner face of the substrate, where the *in situ* synthesis takes place.



**Figure 3.14:** The synthesis cell on the microscope. The assembly is mounted on a precision-adjustable aluminium support. The microarray substrate (round cover glass) is located above the microscope objective. Use of transparent materials (polycarbonate, PDMS and glass) simplifies handling and enables visual control.

To prevent the attachment of gas bubbles (argon gas employed to drive the fluidics system tends to form bubbles upon pressure relief) at the edges of the PDMS cell, we put effort in making a cell with very smooth edges. This is achieved using a sharp-edged punching tool (see appendix B.4) fabricated (electrical discharge machining) by the mechanics workshop of the university. Smooth surfaces also improve the reagent exchange between consecutive synthesis steps. For the purpose of chemical inertness the upper side of the synthesis cell consists of a glass microscopy slide which is glued onto a 10 mm thick block of UV absorbing Makrolon<sup>®</sup> plastics. Back-reflection (and back-scattering) of UV light from the interfaces is reduced (by index matching) with a thin layer of PDMS employed as glue. In- and outlet are formed by syringe needles which are connected to the valveblock via PTFE tubing (Fig. 3.14). More detailed information on the construction of the synthesis cell is provided in appendix B.4.

The design of the cell is optimized for DNA *in situ* synthesis with light-directed photodeprotection. It enables a very small reagent consumption of ca. 40 mg of each NPPOC-phosphoramidite for a 25mer synthesis [Nai06b].

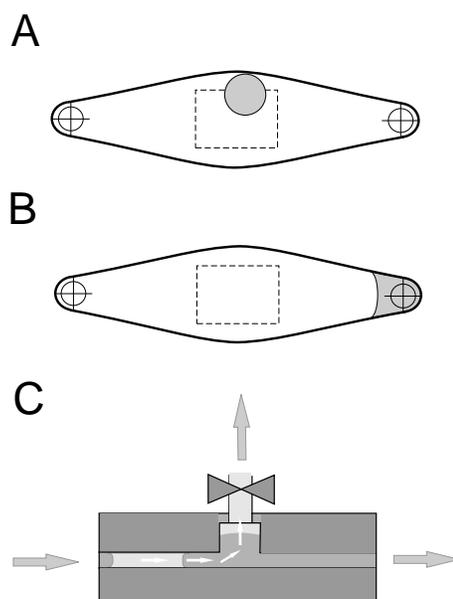
### 3.3.2 Argon Bubble Trapping

The occasional formation of argon bubbles, owing to pressure relief during the reagent transport towards the synthesis cell (the solvent MeCN is saturated with argon) represented a serious problem for the microarray synthesis. Bubbles which have become trapped in the synthesis volume (Fig. 3.15A) do locally increase the stray light intensity during the UV exposure or affect synthesis reactions (coupling etc.) since the substrate surface beneath the bubble is not covered by the reagents.

The "argon bubble problem" has been resolved with a cleverly devised technique:

- Large bubbles are captured by a T-piece bubble trap (Fig. 3.15C) which is integrated in the inlet line.
- Small bubbles (<2 mm diam.), owing to the increased channel width in the synthesis area, have the tendency to get stuck in the synthesis volume (Fig. 3.15A). By employing a short suction pulse the small bubbles are pushed into the inlet region of the synthesis cell (Fig. 3.15B), where they get reliably trapped.

This method of bubble catching is highly reliable. In the critical steps of the synthesis process the occurrence of bubbles within the synthesis area is prevented almost completely.



**Figure 3.15:** Bubble trapping techniques. Top views of the synthesis cell volume (A) and (B) – in- and outlet holes are shown at the right and the left end of the streamlined cell volume. During reagent supply small argon bubbles can get into the synthesis cell (A). They can be removed from the synthesis area (dashed box) by applying a short suction pulse. Bubbles are moved into the narrow inlet region of the chamber (B), where they are trapped due to a more favorable surface energy. (C) Larger bubbles (too large to get trapped in the inlet region) are captured in a ”T-piece bubble trap” before they reach the synthesis cell. Venting of the accumulated gas is achieved by occasionally opening the valve to the venting line.

### 3.4 Automated Microarray Synthesis - Controller-Hard- and Software

The original ABI 381A DNA synthesizer control hardware has been substituted by a personal computer based controller. Fully automated light-directed *in situ* synthesis is performed with the synthesizer control software *DNASyn*, which is described in detail in appendix B.7. *DNASyn* integrates fluidics control (via an external microcontroller-based solenoid valve driver - technical details are provided in appendix B.8) with the ”virtual photolithography mask” projection.

### 3.5 Performance of the Microarray Synthesizer

An affordable microarray synthesizer system for lab-scale fabrication of DNA microarrays has been developed from the following widely available components:

- Oligonucleotide synthesizer (second-hand)

- DLP video projector (second-hand)
- Inverted research microscope (second-hand)
- Personal computer
- Microcontroller-based solenoid valve driver (home-built)
- Optical components: Optical table, filters, lenses, mirrors

The highly flexible microarray synthesis system enables massively parallel *in situ* synthesis of almost arbitrary probe sequences. New microarray designs can be developed within hours and automatically synthesized overnight. The synthesis of a 25mer microarray requires about 6.5 hours (plus 1.5 hours for the final deprotection step outside the synthesis apparatus). With our microscope-projection-lithography setup the size of the microarrays has been reduced to a total area of  $< 10 \text{ mm}^2$ . Owing to the miniaturization, the costs for synthesis reagents (NPPOC-phosphoramidites, MeCN, activator, oxidizer, ethanol) are about 50 Euros per microarray synthesis. Moreover, the small area of the microarray enables hybridization with a very small amount of target solution (in principle less than  $10 \mu\text{l}$  are required). The high stability of our microscope-projection-lithography setup (with respect to image drifting originating from thermal expansion etc.) is beneficial for the quality of the synthesized DNA probes.

In principle each micromirror-pixel (in total  $1024 \times 768$ ) could be used to synthesize a microarray feature. However, the need for a high local contrast and expected difficulties with the image analysis of the small densely-packed features (image distortions etc.) require the use of composite features consisting of  $5 \times 5$  DMD pixels ( $4 \times 4$  pixel feature area plus 1 pixel separation gap). In the corners of the synthesis area (i.e. the imaging field defined by the DMD chip) DNA probe quality is suffering from vignetting ( $\rightarrow$ reduced exposure intensity) and uncorrected curvature of field ( $\rightarrow$ reduced local contrast). For quantitative investigations of *probe-target* binding affinities, a maximum number of about 25000 microarray features is currently achievable.



# Chapter 4

## Light-directed *in situ* Synthesis of DNA Microarrays

### 4.1 Light-Directed *in situ* Synthesis of DNA Microarrays

#### Reagents

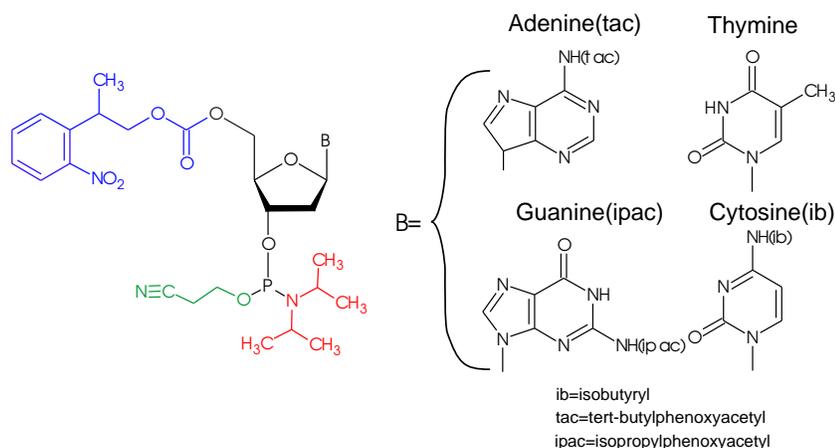
RayDite™ 3'-phosphoramidites NPPOC-dA(tac), NPPOC-dC(ib), NPPOC-dG(ipac) and NPPOC-dT (see Fig. 4.1) carrying photolabile 5'-nitrophenylpropyloxycarbonyl protective groups were purchased from Sigma-Proligo (Hamburg, Germany).

Acetonitrile (ROTISOLV® for DNA synthesis, water < 10 ppm, Carl Roth GmbH, Germany); Activator42™, 0.25 M (Proligo®); iodine based oxidizer (part no. 401732, Applied Biosystems); Trap-Pak™ molecular sieve bags (Applied Biosystems); water-free argon ( $\leq 0.5$  ppm H<sub>2</sub>O)

Photo-deprotection is carried out in a mildly basic (deprotection) solution of 25 mM piperidine (99%, Aldrich) in waterfree acetonitrile. Alternatively, the use of dimethylsulfoxide (DMSO) has been reported [Woe06].

Final base deprotection is performed (at room temperature for about 90 minutes) in a 1:1 mixture of ethylenediamine (analytical grade, Fluka) and ethanol (analytical grade, VWR, Germany).

UV glue (Norland optical adhesive 60, Edmund optics) is used to fix the chip onto a stainless steel support.



**Figure 4.1:** 5'-[2-(2-Nitrophenyl)-propyloxycarbonyl]-2'-deoxynucleoside phosphoramidites. Similar as nucleosides, nucleoside phosphoramidites comprise nucleobases and deoxyribose sugar. Additionally, phosphoramidites contain a phosphorus group, which, when chemically activated, can react with the hydroxy group of a growing (deprotected) oligonucleotide strand. This coupling reaction creates the phosphate group in the sugar-phosphate backbone. Various *protection groups* enable a controlled synthesis of oligonucleotide chains without the risk of unwanted side reactions. The photolabile NPPOC group (blue) substitutes the 5'-hydroxyl of the pentose ring. Its removal (deprotection) enables coupling of another building block. The phosphorus group is protected by a diisopropylamino group (red) ( $\rightarrow$  phosphoramidite) and a 2-cyanoethyl protection group (green). Further protection groups (*ib,tac, ipac*) are necessary to prevent side reactions of the exocyclic amine groups of the nucleobases during the *in situ* synthesis process. All protection groups are removed at the end of the synthesis.

## Preparation of the Microarray Synthesis

Light-directed *in situ* synthesis was performed with NPPOC-phosphoramidites [Has97; Bei99; Nuw02; Nai06b] which differ from the commonly used acid-labile DMT-protected phosphoramidites by the photo-cleavable 5'-nitrophenylpropyloxycarbonyl protection group (NPPOC).

Phosphoramidite reagents are highly sensitive to water. To minimize contamination with water, NPPOC-phosphoramidite solutions - 40 mM in water-free MeCN - are prepared only immediately before the start of the synthesis. Deprotection solution, oxidizer<sup>1</sup> and activator are more stable and can remain on the synthesizer for prolonged times. Contamination with water is particularly critical for the phosphoramidite/MeCN solution contained in the storage bottles. Once degradation due to a small amount of water has started, the phosphoramidites undergo autocatalytic degradation [Kro04]. To minimize water contamination in critical reaction steps, molecular sieve bags (Trap-Pak<sup>TM</sup>) are added to the MeCN storage bottle and to the activator storage bottle. Further hints on phosphoramidite han-

<sup>1</sup> The oxidizer solution itself contains a considerable amount of water (several percent)

ding procedures are provided in appendix B.5.1.

The preparation of the automated synthesis should be performed with the synthesis script *PrepSyn.prg*, which is executed by the controller software *DNASyn*. The *PrepSyn*-script includes the preparation of the phosphoramidite solutions, priming of the reagent supply lines and checklist functionality (installation of the synthesis cell, optics, argon pressure, valve function, reagent availability).

### The automated synthesis cycle

An initial photoreactive monolayer is created by coupling of NPPOC-dT-phosphoramidite to the hydroxyl-groups of the dendrimer functionalized substrate. The synthesis cycle, to be repeated  $4 \times 25 = 100$  times for the synthesis of a microarray with 25mer probes<sup>2</sup>, comprises phosphoramidite coupling, phosphite triester bond oxidation, and photo-deprotection.

- Phosphoramidite coupling is carried out for one minute with a 1:1 mixture of 40 mM NPPOC-amidite solution in water-free MeCN and activator solution (Activator42<sup>TM</sup>, 0.25M)
- A iodine based oxidizer solution (ABI) is employed for about 40 s (after every fifth coupling step) to oxidize unstable phosphite triester bonds, thus to form stable phosphotriester linkages
- The photo-deprotection step (exposure dose 7 J/cm<sup>2</sup> at  $\lambda = 370$  nm) is performed in a 25 mM solution of piperidine (Sigma-Aldrich) in MeCN. Piperidine [Bei99] provides the mildly basic conditions necessary for photocleavage of the NPPOC protection group [Wol04].

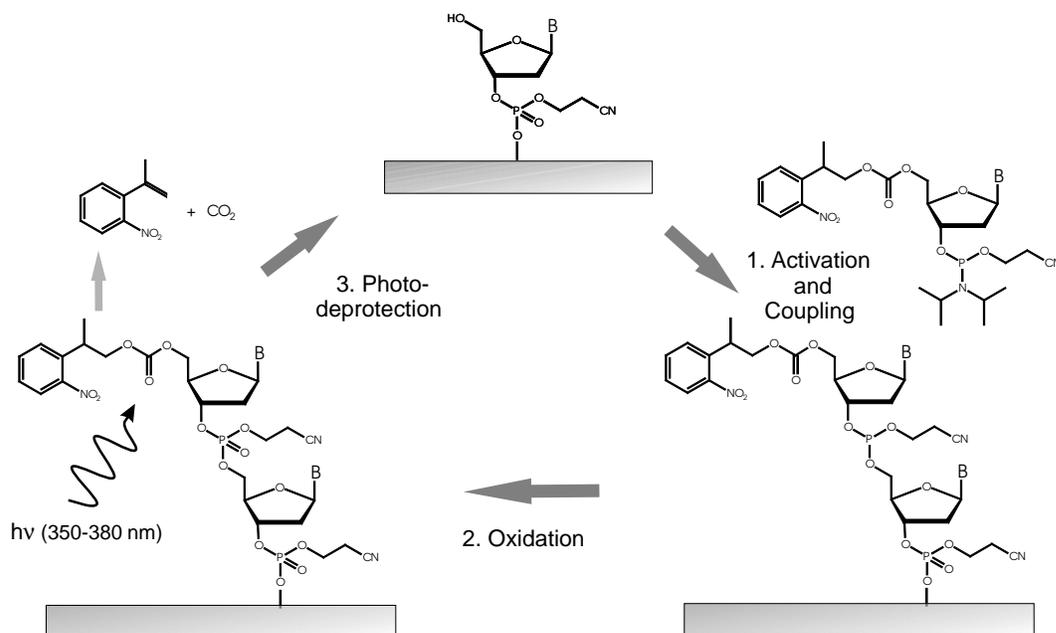
Between the individual reaction steps extensive washing of the valve block and of the synthesis cell/supply line is performed. It is, for example, absolutely necessary to remove trace amount of water (from previous oxidation steps) from the fluidics system prior to the next coupling reaction. Alternating rinsing with pure MeCN and flushing with argon gas is very efficient to remove remaining reagents from the previous reaction step. However, it is important that solid residues are not allowed to dry on the substrate surface.

The final coupling step is followed by complete photo-deprotection of the whole microarray, to remove all remaining NPPOC protection groups, and by a final oxidation step.

Capping of unreacted binding sites by acetylation is commonly employed in oligonucleotide synthesis to prevent the synthesis of strands containing point defects. Because of the rather limited benefits of a capping in light-directed microarray fabrication (see section 2.7.1) we do not apply capping in our DNA Chip synthesis scheme.

---

<sup>2</sup> In practice, owing to mask optimization, only about 80 cycles are required.



**Figure 4.2:** NPPOC synthesis cycle. 1. Activation and coupling of a phosphoramidite building block. 2. Oxidation of unstable phosphite bonds with iodine based oxidizer. In our synthesis scheme oxidation is performed after every fifth coupling step. 3. Photo-deprotection under UV-irradiation results in detachment of the NPPOC group, exposing a hydroxy group at the 5'-C.

Oxidation of unstable phosphite triester bonds into more stable phosphotriester linkages is necessary to prevent strand breakage during the synthesis process. However, it is not necessary to perform oxidation after every coupling step. To accelerate the synthesis process and also to reduce the of import of water (a main ingredient of the the oxidizer solution) into the synthesis cell, the oxidation step is performed after every fifth coupling step only. After finishing the synthesis cycle the microarray substrate is removed from the synthesis cell. In the final deprotection step (performed in a closed glass beaker) the base protection groups are removed in a 1:1 mixture of ethylenediamine and ethanol (for about 90 minutes at ambient temperature) [Nuw02]. Subsequently the substrates are washed with ethanol (analytical grade) and water, and dried under a stream of nitrogen. Storage in 50 ml Falcon tubes helps to prevent scratches on the microarray surface.

The fabrication of a 25mer DNA chip, involving about 80 coupling steps, requires about 8 hours (including 90 minutes for the final base deprotection).

To provide mechanical stability the microarray (on the 20 mm diam. cover glass) is fixed on a stainless steel support (microcopy slide format plate  $76 \times 25 \times 2$  mm). To avoid deposition of adhesive fumes (caution: don't use cyanoacrylate superglue) a UV curable glue (Norland optical adhesive 60) is used to glue the microarray above the 10 mm diam. window in the center of the plate. Use of an i-line (365 nm) UV lamp is recommended.

Microarrays can be safely stored (for several months) at 4°C in 50 ml Falcon tubes.

## 4.2 Preparation of Phosphorus Dendrimer Functionalized Microarray Substrates

### Reagents

All reagents are used as purchased without further purification. Unless specified otherwise aqueous solutions are prepared with Milli-Q water (18.2 MΩcm).

Diam. 20 mm round cover glasses (Menzel-Gläser, Braunschweig, Germany); Deconex 11 UNIVERSAL (Borer Chemie AG, Zuchwil, Switzerland); (3-aminopropyl)-triethoxysilane (APTES) (Sigma-Aldrich); ethanol analytical grade (VWR, Germany); 1,2-dichloroethane (Cat. No. 6837.1, Carl Roth GmbH, Germany); phosphorous dendrimers with aldehyde moieties cyclotriphosphazene-PMMH-96 (Cat. No. 552097, Aldrich); potassium hydroxide (Carl Roth GmbH); sodium borohydride (99.99 %, Sigma-Aldrich).

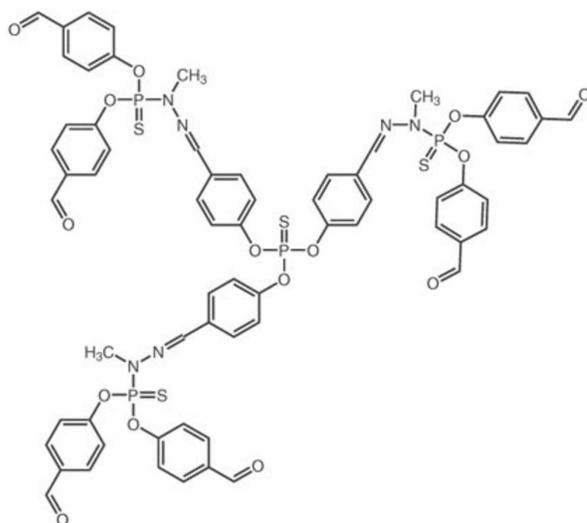
### Surface functionalization

Microscope optics, employed for projection lithography onto the synthesis substrate, requires 0.17 mm thin cover glasses to be used as synthesis substrates. Improved image quality, reduced UV absorption, and also reduced autofluorescence of the glass substrate, later at the analysis stage, are beneficial compared to the commonly used 1 mm thick microscope slide format.

Dendrimer-functionalized substrates on the basis of phosphorus dendrimers (PD) - (chemical name: cyclotriphosphazene-phenoxyethyl(methylhydrazono) dendrimers) were prepared according to LeBerre *et al.* [LB03]. To adapt the substrate chemistry to the needs of the *in situ* synthesis process the aldehyde groups are reduced to hydroxyl groups in the final step.

### **The chemical functionalization (adapted from [LB03])**

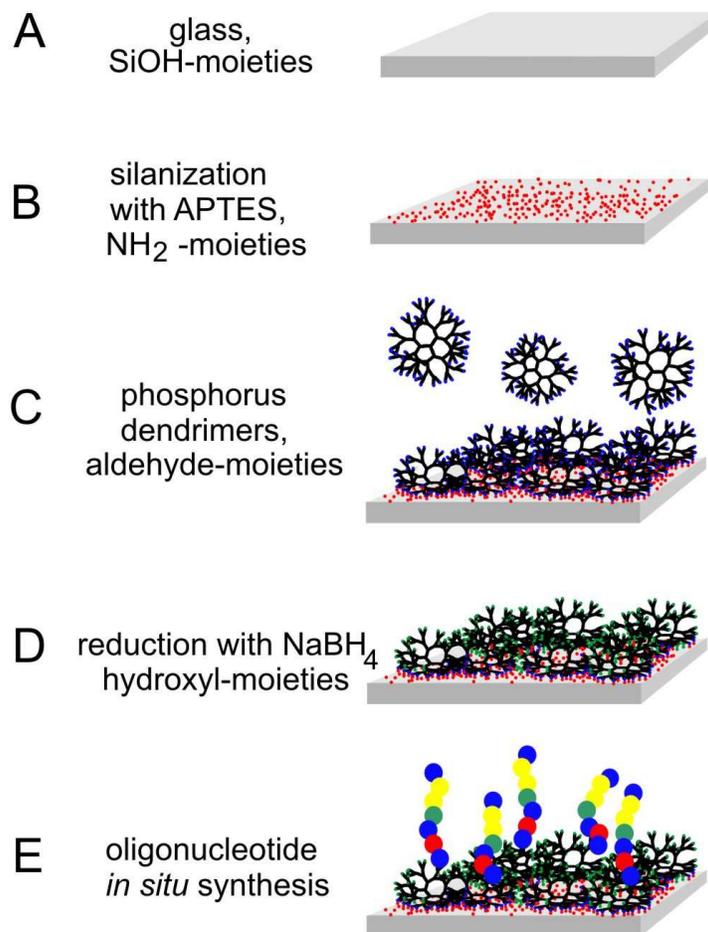
20 mm round cover glasses are used as substrates as these have the advantage of being mechanically more robust (in respect to loads applied for sealing the synthesis chamber) than squared or rectangular ones. The cover glasses were sonicated for 30 minutes in detergent solution (5% Deconex) and rinsed with MilliQ-filtered water. After drying under a stream of nitrogen, a laboratory plasma cleaner (air plasma) is used for 10 minutes to remove organic decontaminants and to activate the surface for subsequent silanization. Immediately after plasma treatment the slides are silanized with a 10% (v/v) so-



**Figure 4.3:** Phosphorus dendrimer (cyclotriphosphazene-phenoxy methyl(methylhydrazono) dendrimer) - generation 1.5 - with aldehyde functionalization. Substrate preparation is performed with generation 4.5 dendrimers carrying 96 aldehyde moieties.

lution of 3-aminopropyltriethoxysilane (APTES) in 95:5 ethanol/water. The silanization is performed at room temperature under gentle agitation for 12 hours. Subsequently the slides are rinsed two times with analytical grade ethanol and once with water (MilliQ). After drying under a stream of nitrogen the slides are baked for 3 hours at 120°C. Prior to dendrimer coupling the slides are activated in an aqueous solution of KOH (8%) for 5 minutes. The activation is followed by rinsing with MilliQ filtered water (3 times) and drying under a stream of nitrogen. Aldehyde-functionalized phosphorus dendrimers (cyclotriphosphazene-PMMH-96) are dissolved in dichloroethane (0.1% w/v). At room temperature, under gentle agitation dendrimer molecules are allowed to couple to the aminosilane surface for 7 hours. Aldehyde-amine condensation results in the formation of imine bonds. The dendrimer solution can be stored and reused several times. Dendrimer-functionalized surfaces are rinsed with dichloroethane and ethanol (two times) and dried with nitrogen.

For use in the *in situ* synthesis process (coupling of phosphoramidites) the aldehyde moieties of the dendrimers are reduced to hydroxyl groups. Reduction is performed in an aqueous solution of sodium borohydride (0.35%) for 3 hours (at room temperature, under gentle agitation). Reduction with sodium borohydride also reduces the unstable imine to more stable amine. After rinsing with MilliQ-water the slides are ready for use. Long term storage over one year at 4°C (under air atmosphere) doesn't affect the substrates. Additional information on substrate preparation is provided in appendix B.6.



**Figure 4.4:** Dendrimer substrate preparation. (A) Cleaning and plasma-activation exposes silanol groups at the glass surface. (B) Silanization with 3-aminopropyltriethoxysilane (APTES) comprises hydrolysis of the alkoxy groups, condensation (due to hydrogen bonding of the silanol groups) and curing at 120 °C to establish a covalent linkage. (C) The aldehyde moieties (blue) of the phosphorus dendrimers react with the NH<sub>2</sub> groups (red) of the aminosilane to form imine bonds. (D) Sodium borohydride is used as a reducing agent to convert unstable imine bonds to stable amine bonds. Along with these the remaining aldehyde groups at the upper side of the dendrimers are reduced to hydroxyl groups (green) - the coupling sites for phosphoramidite oligonucleotide synthesis (E).

### Experiments with other surface functionalizations

Several substrate functionalizations have been tested in the early development stage of the microarray synthesis process. Hydroxy-functionalized<sup>3</sup> polyamidoamine (PAMAM) dendrimers [Ben02] and polyethylene glycol brushes (PEG on epoxysilane - GPTS) [Pie00] didn't provide satisfactory results (strong background fluorescence and low stability, respectively).

<sup>3</sup> Protocol according to [Ben02]. Hydroxyl moieties were created by linkage of aminopentanol.

More satisfactory results have been achieved with a monohydroxysilane functionalization [Gao01; Ric04]. To increase the distance between the probe and the glass surface Singh-Gasson *et al.* [SG99] inserted additionally a hexaethylene glycol-linker.

#### Preparation of monohydroxysilane slides (adapted from [Gao01]):

Cover glasses were sonicated 20 minutes in 5% Deconex solution and washed several times with water. Activation of the surface in a plasma cleaner (air plasma) for 10 minutes. Silanization with N-(3-triethoxysilylpropyl)-4-hydroxybutyramide (ABCR GmbH) 1% (v/v) in 95:5 (v/v) ethanol/water at room temperature, overnight. Washing with ethanol. Curing for 1 h at 120 °C.

Unter identical hybridization conditions microarrays produced on PAMAM substrates provided the largest hybridization signals.

$$I_{\text{PAMAM}} > I_{\text{PD}} > I_{\text{Monohyd.}}$$

However, unlike phosphorus dendrimer (PD) and monohydroxysilane, PAMAM substrates showed a strong background fluorescence and were highly attractive for deposition fluorescent particulates contained in the hybridization solution.

The hybridization signal on phosphorus dendrimer slides is (roughly estimated - no systematic experiments performed) two to three times higher than on monohydroxysilane slides. Also, the phosphorus dendrimer chips are more stable than the monohydroxysilane slides (repeated washing and hybridization steps) and can therefore be reused more often. Compared to other substrates tested, the phosphorus dendrimer-microarray surfaces look very homogeneous and unspecific adsorption of nucleic acid and particulates is negligible. These observations are in accordance with the [LB03]. Our experiments demonstrate that phosphorus dendrimer functionalized surfaces are a favorable substrate not only for immobilization techniques, but also for the *in situ* synthesis of DNA microarrays.

## 4.3 Noteworthy Characteristics of the Microarrays

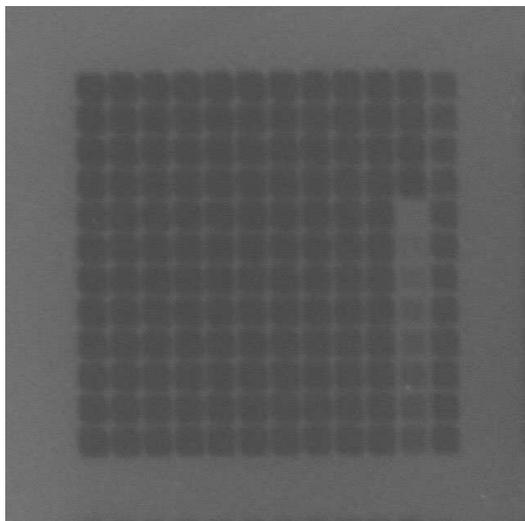
### 4.3.1 Autofluorescence of the Chip Surface

The microarrays fabricated in the *in situ* synthesis process show an autofluorescence under blue excitation with the Olympus U-MNB2 narrow blue excitation (470-490 nm) filter set<sup>4</sup> (Fig. 4.5). The green fluorescence emission is largely restricted to the areas between the

---

<sup>4</sup> With the U-MWG2 (510-550 nm excitation) mirror unit, which is used for imaging the Cy3 hybridization signal, the autofluorescence is barely noticeable, and doesn't affect microarray analysis.

microarray features. Within the feature areas (containing DNA probes) the autofluorescence of the chip is significantly smaller. Increasing probe length correlates with reduced autofluorescence intensity. This may indicate that the autofluorescence originates from the



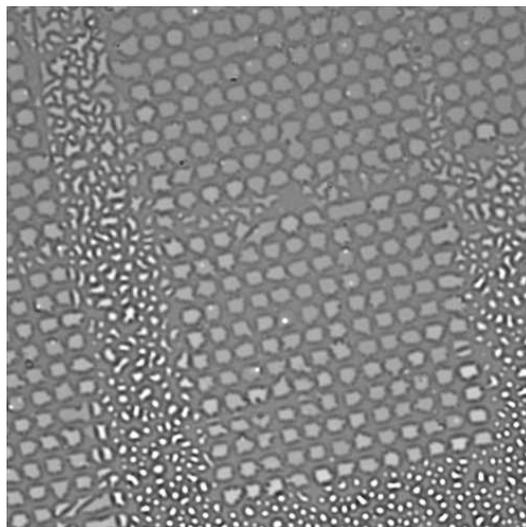
**Figure 4.5:** The fluorescence micrograph of the feature block demonstrates that the DNA probes (mostly 20mers) quench the substrate autofluorescence. In the two rightmost feature columns the probe length is incrementally increased from 1 to 20 nucleotides. The image demonstrates that the fluorescence quenching depends on the amount of DNA material covering the surface.

phosphorus dendrimer substrate and is quenched by the overlying DNA probes. However, the pure dendrimer substrate itself initially doesn't show fluorescence. The fluorescence is restricted to the area of the substrate that has been in contact with synthesis reagents inside the reaction chamber.

Incomplete final deprotection (i.e. base protection groups *tac*, *ib* and *ipac* are not completely removed) results in strong fluorescence of the microarray features (with intensities inverse to those shown in Fig. 4.5).

### 4.3.2 Hydrophilicity of DNA Microarray Features

Oligonucleotide probes render the microarray surface hydrophilic. This can be employed to make the microarray visible (for alignment etc.). As shown in Fig. 4.6 water vapor condenses on the cold microarray surface. Tiny droplets form a milky haze on the hydrophobic substrate. The hydrophilic area covered by DNA is completely wetted with a water film (see Fig. 4.6) and therefore appears clear.



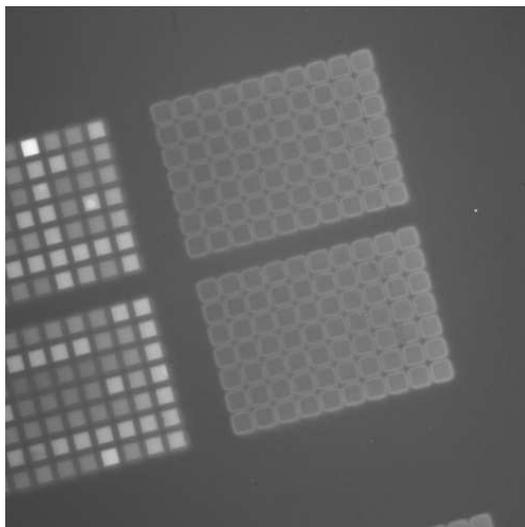
**Figure 4.6:** Wetting characteristics of the microarray surface (microscope image). Hydrophilic features (size about 20  $\mu\text{m}$ ) are covered by a closed thin film of water. Regions between feature blocks are covered with tiny droplets.

### 4.3.3 Hybridization without Detergent - Unspecific Adsorption

Omittance of the surfactant (Tween-20<sup>TM</sup> or SDS) resulted in very strong surface adsorption on the entire microarray surface - also in the regions where no probes have been synthesized. Subsequent addition of 0.01% Tween-20 on the same microarray resulted in probe-specific hybridization. Hybridization in this particular experiment was performed with MES hybridization buffer at room temperature.

### 4.3.4 Irreversible Target Adsorption

In MES hybridization buffer at temperatures  $> 55^{\circ}\text{C}$  targets tend to bind irreversibly to the substrate surface, making a reuse of the microarrays impossible. The fluorescence intensity is particularly high between the features (see Fig. 4.7). This suggests that targets which have dissociated from the probes are captured by reactive groups at the substrate surface adjacent to the features. The problem seems to be related to the use of the MES hybridization buffer at high temperatures ( $> 55^{\circ}\text{C}$ ). Using 5 $\times$ SSPE buffer instead, we do not observe this characteristics. However, we found that often (even at high temperatures of  $70^{\circ}\text{C}$ ) the hybridization signals can not be completely removed. This problem, which has also been reported by Hu *et al.* [Hu05], could be owing to stable duplexes which do not completely dissociate at the temperatures applied. It is also possible that hybridized targets have an increased probability for bonding to unblocked reactive sites at the microarray surface. In

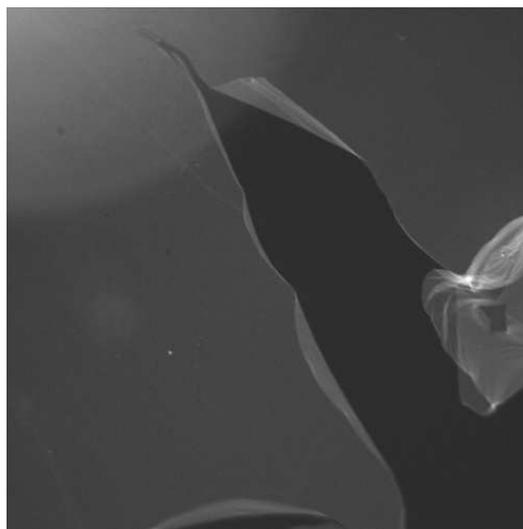


**Figure 4.7:** Fluorescence micrograph of irreversible adsorption. The feature blocks in the center of the image have undergone dissociation in pure MES hybridization buffer. At temperatures of about  $60^{\circ}\text{C}$  rather than to detach from the surface the fluorescently labeled targets have irreversibly bound to the microarray surface. The brightest signal is visible in the gaps between the features. At the left edge of the image another feature block (with another sequence motif) is shown, which has been hybridized after dissociation conditions have been applied (thus demonstrating that the other probes on the microarray maintained their hybridization capability).

either case the targets can be removed completely if RNA targets are used rather than DNA targets. An alkaline stripping procedure (sodium hydroxide) will selectively degrade RNA targets (into nucleotides), whereas DNA probes remain unaffected [Hu05].

#### 4.3.5 Robustness of the Phosphorus Dendrimer Surface Coating

Fig. 4.8 demonstrates that the phosphorus dendrimer functionalization (section 4.2) forms a stable network on the glass surface. Parts of the dendrimer coating (autofluorescence under blue excitation) have come off the surface after harsh treatment with an unsuitable stripping buffer. The robust closed-film structure shown in Fig. 4.8 is rather unexpected since the chemistry of the surface-functionalization would rather suggest a monomolecular layer of unconnected dendrimer molecules. However, dendrimers bound to the aminosilane layer possibly form a densely interwoven network. It is further possible, that the functionalization with the aminosilane APTES results in the formation of a stable multi-layer film.



**Figure 4.8:** Fluorescence micrograph of the phosphorus dendrimer substrate. Use of an unsuitable stripping buffer (10 minutes in boiling 0.1 M  $\text{Na}_2\text{CO}_3$  solution) revealed the stable network structure of the surface coating. It appears that the phosphorus dendrimer network remained intact, even though the coating is completely detached from the glass surface.

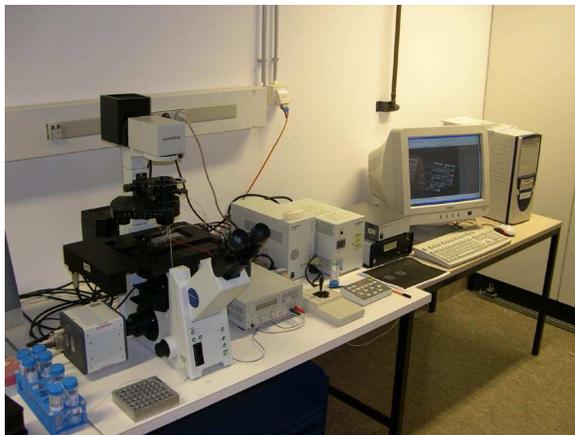
# Chapter 5

## DNA Microarray Analysis

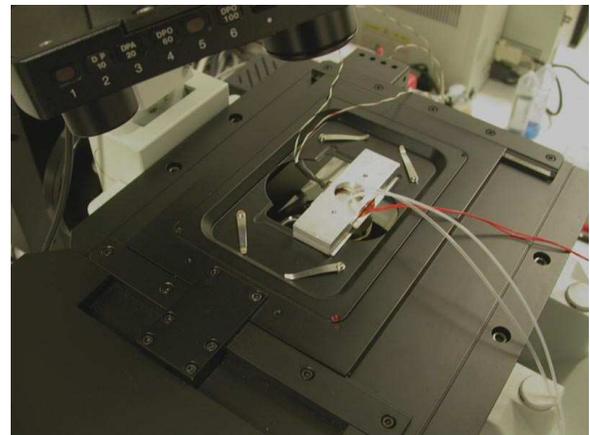
### 5.1 Hybridization Signal Acquisition - Experimental Setup

Microarray hybridization assays were performed in a temperature-controlled hybridization chamber. The design of the flow-through type chamber is similar to that of the synthesis cell (see section 3.3.1). Installation on an epifluorescence microscope setup enables real time monitoring of the hybridization signal. A sensitive electron multiplying CCD-camera (EMCCD) is used for image acquisition.

A



B

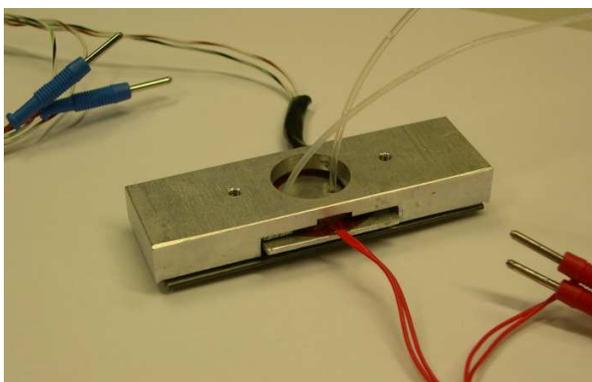
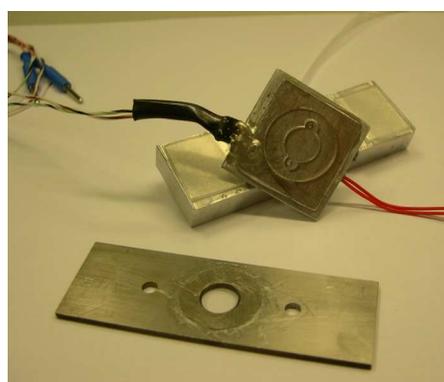


**Figure 5.1:** Microarray analysis setup. (A) Motorized fluorescence microscope with EMCCD-camera (bottom left). (B) Hybridization chamber on the XY-stage of the microscope.

### 5.1.1 The Hybridization Chamber

Design considerations:

- realtime monitoring (requires a window into the sealed chamber and low background fluorescence)
- reagent exchange (e.g. to replace the hybridization buffer by a washing solution)
- high mechanical stability to minimize defocusing and xy-drifting of the image upon thermal expansion
- temperature control

**A****B**

**Figure 5.2:** Hybridization chamber assembly (A). Inlet/outlet tubes enter from the top. The microarray is located at the bottom. Part (B) shows the microarray on its stainless steel support (lying in the front) and the stainless steel top plate (leaning against the brace) with the PDMS gasket. The microarray slide is pressed against the aluminium brace with two fastening screws.

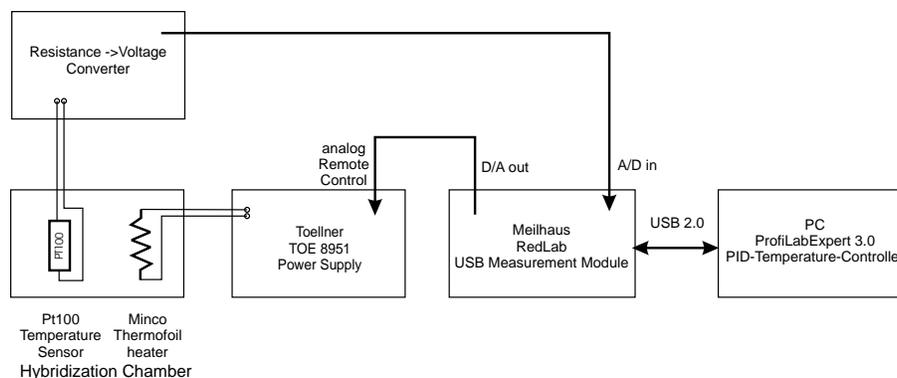
The hybridization chamber is made from a 1.5 mm thick PDMS gasket. The chamber volume (about 120  $\mu\text{l}$ ) is formed by a 10 mm diam. hole (cut from a sheet of PDMS with a punching tool). Circular cut-offs at the inlet and outlet openings (see Fig. 5.2B) prevent sticking of air bubbles inside the chamber volume.

The microarray with its stainless steel support constitutes the bottom side of the hybridization volume. This configuration, using the chip substrate as window, enables observation of the hybridization signal with an inverted microscope. A stainless steel plate forms the upper side of the hybridization volume. Stainless steel is used because it is resistant to the hybridization buffer (no salt corrosion). Also important, since the steel plate is in the background of the microscope field of view: the steel plate isn't fluorescent and doesn't adsorb nucleic acid targets.

A flexible Thermofoil<sup>TM</sup> heater (Minco) (with a 15 mm diam. opening in the center - for

inlet/outlet tubes) is glued onto the upper side of the steel plate. Temperature is measured with a platinum resistor (Pt-100) which is fixed with thermal adhesive at the edge of the steel plate. In- and outlet tubes (located at opposing ends of the chamber volume) penetrate the steel plate from the top side (see Fig. 5.2A). To avoid dead volume and corrosion and to achieve reliable sealing, the PFA tubes are directly connected to the plate by press-fitting.<sup>1</sup>

## Temperature Control



**Figure 5.3:** Control of the hybridization temperature. Heating of the hybridization chamber is performed with a Minco Thermofoil™ heater which is in thermal contact with the hybridization solution via a corrosion resistant stainless steel plate. The temperature is measured with a Pt-100 sensor (in thermal contact with the stainless steel plate). The resistance is converted into a voltage signal that is proportional to the temperature. The voltage is read by an A/D input channel of the RedLab USB measurement module. A software based PID-temperature controller (run as a PC application generated with ProfiLabExpert 3.0 - see Fig. B.16) by comparing the actual temperature and the set temperature, determines the control voltage (output via the RedLab D/A output) that is used to operate the remote controlled heater power supply.

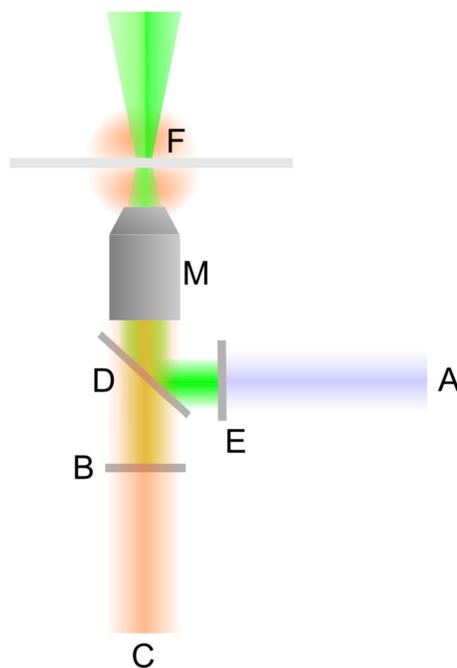
Temperature is measured with a Pt-100 resistor and converted into a temperature-proportional voltage signal. A USB measurement module (ME-Redlab, Meilhaus) is employed for signal acquisition with a personal computer. A software-based PID-controller (see appendix Fig. B.16) designed with ProfiLab-Expert 3.0 (ABACOM Electronics-Software) enables user-defined temperature profiles and temperature-recording. The heating power for the foil heater is provided by a remote-controlled power supply (TOE 8951, Toellner Electronic Instrumente GmbH) which is controlled via the D/A-output of the USB-module. A test with a calibrated Pt-100 resistor - brought in thermal contact with the outside of the microarray substrate - showed that the temperature at the microarray surface is controlled

<sup>1</sup> The tubes - outer diam. 1.2 mm int. 0.8 mm were drawn through the 1 mm diam. inlet/outlet mounting holes in the steel plate (→ stable press-fit-connection).

with an accuracy of approximately  $\pm 1^\circ\text{C}$ . The temperature can be held constant within a variation of  $< 0.2^\circ\text{C}$ .

### 5.1.2 Epifluorescence Microscope

The microarray hybridization signal is acquired by epifluorescence microscopy (principle shown in Fig. 5.4). Realtime monitoring of the hybridization signal is performed with an



**Figure 5.4:** Epifluorescence microscopy. The light of a bright mercury arc lamp (A) is filtered by the excitation filter (E) and reflected by the dichroic mirror (D) through the microscope objective (M) onto the fluorescently labeled sample (F). Fluorescent dye molecules absorb the excitation light, and enter an excited electronic state. Due to the Stokes shift the emitted light has a longer wavelength than the excitation light. The Cyanine 3 (Cy3) dye used throughout this study has a peak absorption at 550 nm (green) and shows yellow to orange fluorescence emission (with a peak at 570 nm). A fraction of the fluorescence signal (emitted in all directions) is collected by the microscope objective M and transmitted through the dichroic mirror (D). The barrier filter (B) passes only the fluorescence light to the camera (C).

Olympus IX81 inverted research microscope (Fig.5.1). The fluorescence of Cy3 labeled targets is imaged using an UPlanApo 10 $\times$ 0.40 NA microscope objective (Olympus) and the U-MWG 2 filter set (Olympus).

### 5.1.3 Image Acquisition with an EM-CCD Camera

High resolution image acquisition was performed with a sensitive Hamamatsu EM-CCD C9100-02 electron multiplying camera.

#### Camera specifications:

- Peltier cooling:  $-50^{\circ}\text{C}$
- Gain factor: 800
- Read-out noise:  $<1$  electron r.m.s. at high gain mode
- Dynamic range: 14 bit
- Full resolution:  $1000 \times 1000$  pixels

Camera and microscope (shutter, filter, exposure, focus, XY-stage etc. ) were controlled by the SimplePCI (Compix Inc.) image acquisition software.

#### Shading correction

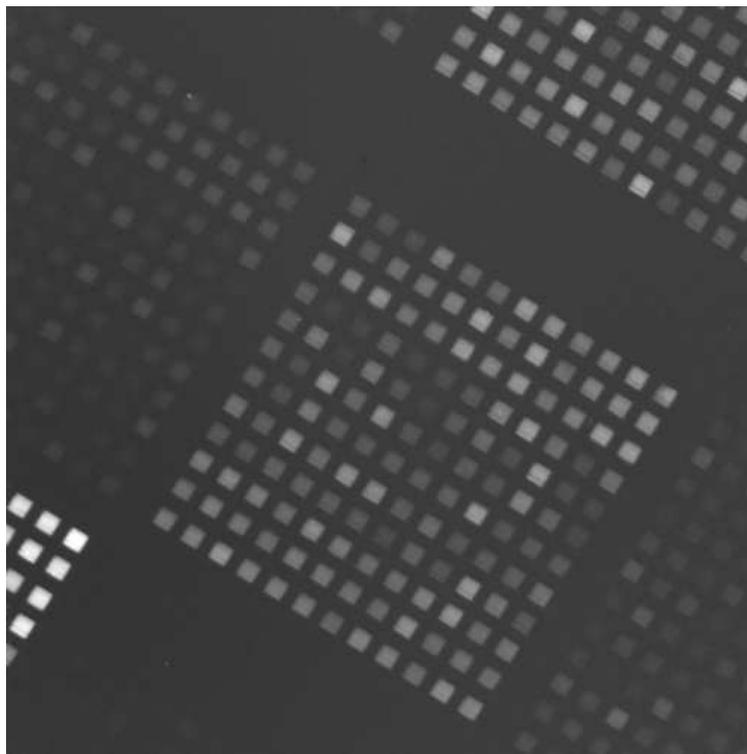
Uneven fluorescence excitation and fluorescence collection, owing to vignetting (larger blockage of off-axis light rays) yield fluorescence micrographs that are brighter at the center and darker at the edges. Intensity gradients due to shading can be a significant source of error for quantitative analysis of hybridization signals.

Shading correction (using the SimplePCI setting *Ratio shade correction*) is therefore performed by dividing the specimen image (microarray) through a fluorescence reference image, which is acquired by imaging a uniformly fluorescent surface. As described by Model *et al.* [Mod01] spatially uniform fluorescence is obtained from a thin layer of fluorescent dye (e.g. 20  $\mu\text{l}$  hybridization solution with 100 nM of Cy3 labeled targets) sandwiched between a microscopy slide and a cover glass.

## 5.2 Quantitative Analysis of Microarray Hybridization Signals

Fluorescence micrographs of the hybridization signal are saved as 16-bit grayscale TIFF images. Shading correction is performed during image acquisition.

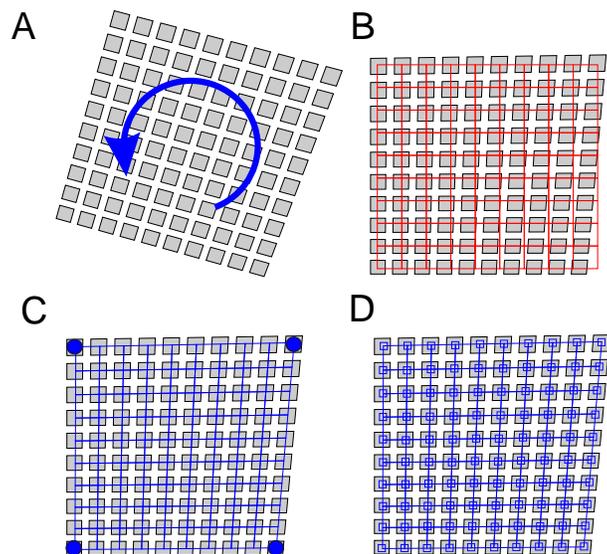
Quantization of feature intensities is carried out with the Java program *ScanRA* (technical details in appendix B.10). The software (which was developed as part of this thesis) enables automatic analysis of microarray feature intensities. To define feature positions a



**Figure 5.5:** Raw hybridization signals as imaged with the Hamamatsu EM-CCD camera (original resolution of the image  $1000 \times 1000$  pixel - size reduced to  $500 \times 500$ ). For the image acquisition the 16mer-microarray remained in the hybridization solution (1 nM Cy3-end-labeled RNA oligonucleotide target). The hybridization temperature was  $30^\circ\text{C}$ .

*readout grid* (Fig. 5.6C) is placed on the microarray image. Then the program integrates pixel intensity values over the *integration boxes* located at the grid points in center of the features. The size of the integration boxes should be chosen to prevent integration over feature boundaries (Fig. 5.6D). The exact placement of the readout grid requires rotation of the image, so that the microarray grid is approx. aligned with the screen axis. Considering small image distortions an orthogonal grid of evenly spaced points is not suitable to determine features positions (Fig. 5.6A). Rather, a quadrilateral grid (defined by the four corner points) is suitable to account for first order distortions of the microarray image.

Microarray hybridization signals (16-bit intensity values) are averaged over the integration boxes to provide a 16-bit mean intensity value. The standard deviation of the pixel intensity values provides information about the homogeneity of the individual microarray feature intensities. Large standard deviations can indicate defects (e.g. fluorescent particles or scratches on the microarray surface) or bad alignment of the readout grid. Average brightness, standard deviation of the feature brightness and the position of the individual features are saved in comma-separated value (CSV) format.



**Figure 5.6:** Microarray analysis with *ScanRA*. Readout of the hybridization signal intensities of a feature block. (A) Rotation of the image. (B) An orthogonal readout grid doesn't match all feature positions exactly if the array is slightly distorted. (C) A quadrilateral readout grid defined by the four corner points is a good first order approximation for small image distortions. (D) Integration boxes (blue) are located at the grid points in the center of the features.

Time series of fluorescence micrographs can be analyzed in batch mode – the readout grid needs to be defined only once. To account for drifting of the image owing to thermal expansion of the hybridization chamber (if the temperature has been varied significantly), the position of the first (upper left) corner point has to be provided manually for about five images. The drift offsets of the other images are determined by linear interpolation.

### 5.3 Real-time Monitoring of Microarray Hybridization

Microarray hybridization is usually followed by one or several washing steps to remove unhybridized targets (see Fig. 2.22). Washing is necessary for the detection of small hybridization signals since these are otherwise not visible within the fluorescent background of the hybridization solution. This is typically the case for expression profiling experiments where thousands of different nucleic acid targets comprise the hybridization solution.

However, in most of the experiments performed this study only a single target species is contained in the hybridization solution. At a target concentration of 1 nM the concentrated fluorescence of the hybridized targets (surface-bound in the microscope focal plane) can be well-distinguished from the background fluorescence of the hybridization solution.

This enables real-time acquisition of the hybridization signal in the hybridization buffer solution.

### 5.3.1 Hybridization Buffer

A minimalist hybridization buffer comprises salt (commonly NaCl - 0.2 to 1 M dissolved in water) to reduce electrostatic repulsion between negatively charged nucleic acid strands, a buffer reagent to maintain a pH between 6 and 8, and a surfactant to prevent unspecific adsorption of targets on the microarray surface.

The following hybridization buffer - based on 5×SSPE (saline sodium phosphate - EDTA) - has been used in most experiments.

#### **5×SSPE based hybridization buffer:**

- 5× SSPE (nuclease-free water, 0.75 M NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 5 mM EDTA)
- add 0.01% (v/v) Tween-20
- adjust to pH 7.4 with NaOH

Addition of the surfactant Tween-20<sup>TM</sup> (polyoxyethylene (20) sorbitan monolaurate) is essential to prevent a strong unspecific adsorption of targets on the microarray surface. Use of 0.1 % sodium dodecyl sulfate (SDS), rather than Tween-20, turned out to be less suitable, since SDS tends to precipitate as fluorescent crystals at temperatures below 30°C. The chelating agent ethylenediaminetetraacetic acid (EDTA) is added to inhibit nuclease activity.

The widely-used MES (2-(N-morpholino)ethanesulfonic acid) hybridization buffer [Nuw02] works equally well, however, at temperatures above 60°C strong irreversible adsorption (Fig. 4.3.4) of the fluorescent targets is observed (→ reuse of the microarray not possible).

#### **MES hybridization buffer:**

- 50 mM MES, 0.5 M NaCl, 10 mM EDTA
- add 0.01% (v/v) Tween-20

### 5.3.2 Microarray Washing Procedures

Using standard microarray washing procedures salt residues on the small microarray features can be a serious problem. For prevention of salt residues the microarray washing can be finalized with the following procedure (performed at room temperature):

- Washing in ethanol/water (50:50) for 1 minute. Salts dissolve in the water. The ethanol content prevents dissociation of the duplexes.
- Washing in 95% ethanol for 1 minute
- Drying under a stream of nitrogen

Alternatively, washing can be performed within the hybridization chamber. The hybridization solution and weakly-bound targets are flushed away by washing buffers. Microarray analysis in solution (in a low stringency buffer) circumvents the problems related to salt residues on the dry microarray surface.



# Chapter 6

## Influence of Single Base Defects on Oligonucleotide Duplex Binding Affinities - Microarray Experiments

### 6.1 Motivation

Oligonucleotide microarrays are increasingly used for genotyping and resequencing applications. The *discrimination* capability<sup>1</sup> of short (< 30 nt) oligonucleotide probes is used for the detection of *single nucleotide polymorphisms*<sup>2</sup> (SNPs) and gene mutations.

SNPs can be detected by hybridization with short oligonucleotide probes (typically 12 to 30 nucleotides long). Already a single mismatched base pair can result in a significant decrease of duplex binding affinity [Wal79].

According to [Sug86; Sug00; San04] the Gibbs free energy for the formation of mismatched duplexes can be established on the basis of the *nearest neighbor model* (see section 2.3.2). Appropriate nearest neighbor parameters for combined hydrogen bonding and stacking interactions between the MM base pair and neighboring base pairs have been determined by Allawi *et al.* [All97]. However, the current thermodynamic models of oligonucleotide duplex stability, based on these parameters do not describe the dominant influence of defect position that has been observed in recently published DNA microarray

---

<sup>1</sup> The ability to discriminate between a perfect match (PM) and mismatch duplex (MM) can be used, for instance, to discriminate between the wild-type and a mutant gene.

<sup>2</sup> Variation of a single base pair (point mutation) in the genome, occurring in at least 1% of a population. SNPs largely determine genetic individuality, but also the individual susceptibility to gene-related diseases, and are therefore of great interest not only for genetic research but also for medical diagnostics.

studies [Poz06; Wic06; Nai06b]. Oligonucleotide hybridization is by far not yet understood. This is particularly true for surface-bound hybridization on DNA microarrays and for the hybridization of mismatched duplexes.<sup>3</sup> Detailed knowledge about the underlying physical process is still lacking.<sup>4</sup>

The development of the microarray synthesizer motivated an investigation of the impact of the point defects originating from the *in situ* synthesis process. Our initially technically motivated interest led us to a comprehensive investigation of the impact of point defects on oligonucleotide binding affinities on DNA microarrays.

Previous related studies on the impact of single base MM defects were based on relatively complex target mixtures of biological origin, i.e. fluorescently labeled PCR-products (up to 600 bp long) [Wic06] and RNA targets amplified via *in vitro* transcription from PCR-products (originating from ribosomal RNA) [Poz06], respectively.

The hybridization affinity of individual probe-target pairs is affected by many factors. Therefore, in this study, complications originating from target secondary structure, steric hindrance, labeling effects, cross hybridization, competitive effects, and influences related to target preparation (e.g. bias in nucleic acid amplification) have largely been avoided by using rather short (20-37mer) synthetically fabricated oligonucleotide targets (see Tab. 6.1). DNA and RNA oligos were end-labeled with Cyanine 3 (Cy3<sup>TM</sup>) fluorescent markers. To minimize competitive hybridization and cross-hybridization effects the hybridization assays were performed with only one target species at a time.

In previous studies [Wic06; Poz06] defect positional influence has been investigated statistically (as an average characteristics of many different mismatched duplexes). Here, however, we have focused on the position-dependent impact of single-base defects in individual sequence motifs.

---

<sup>3</sup> Tautz and coworkers in [Poz06]: "We also examined the effects of single-base pair mismatch (MM) (all possible types and positions) on signal intensities of duplexes. We found that the MM effects differ from those that were predicted from solution-based hybridizations. These results recommend against the application of probe design software tools that use thermodynamic parameters to assess probe quality for species identification. Our results imply that the thermodynamic properties of oligonucleotide hybridization are by far not yet understood."

<sup>4</sup> Zhang *et al.* [Zha07]: "DNA/DNA duplex formation is the basic mechanism that is used in genome tiling arrays and SNP arrays manufactured by Affymetrix. However, detailed knowledge of the physical process is still lacking."

## 6.2 Conception of the Microarray Hybridization Experiments

The impact of single base defects on oligonucleotide binding affinities was systematically investigated with DNA microarrays comprising large sets of deliberately point-mutated probe sequences. Within each probe set the individual probe sequences were derived from a common *probe sequence motif* by substitution, insertion or deletion of single nucleotides at a systematically varied position (see Fig. 6.1). Hybridization was performed with fluorescently labeled target oligonucleotides, which were complementary (thus perfectly matching) to the corresponding probe sequence motif. The DNA and RNA oligonucleotide target sequences (Tab. 6.1) were chosen (using the UNAFold software for nucleic acid structure prediction) to avoid stable secondary structures interfering with the hybridization to the microarray-tethered probes.

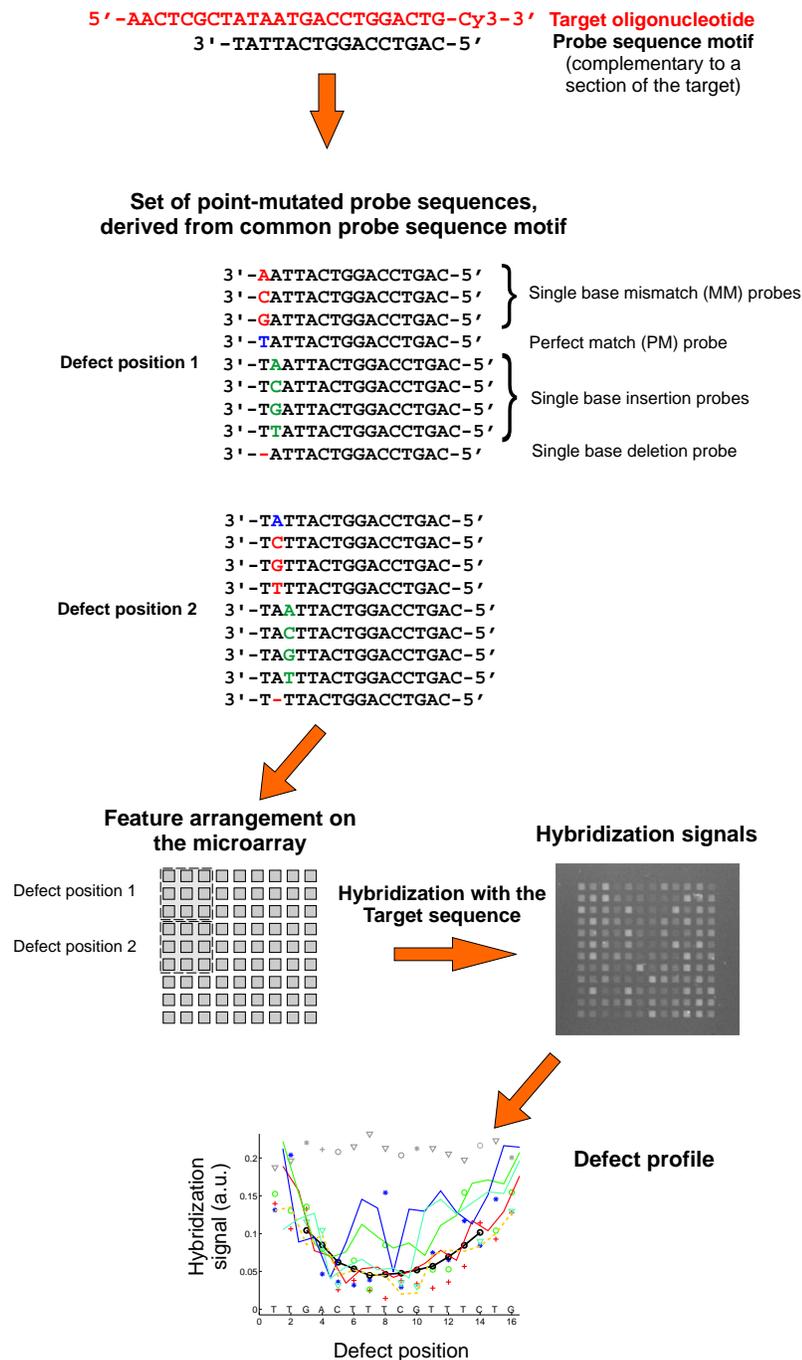
Sets of probe sequences containing single base variations (base substitutions, insertions and deletions) with respect to a common sequence motif were generated with a MatLab code. Upon hybridization with the target sequence these variations of the probe sequences give rise to destabilizing point defects in the duplex structure (i.e. single base mismatches, insertions and deletions).

In each probe set the defect position is shifted (in increments of one base position) from the 3'-end to the 5'-end of the probe sequence motif (see Fig. 6.1). For each defect position the probe set comprises 3 MM probes (the perfect matching base is substituted by one of the 3 remaining bases, thus resulting a mismatched base pair) and 1 PM probe, which is used for quality control (e.g. to identify gradients). Additionally, four single base insertions and a single base deletion probe were generated by insertion of a surplus base, or by deletion of a base, respectively.

The high flexibility gained from DNA microarray *in situ* synthesis and the excellent spot homogeneity - in comparison to spotted microarrays - simplifies a comprehensive comparative analysis and provides the capability to detect subtle differences of the probe affinities. As will be shown below the wealth of data contained in the *defect profiles* (hybridization signal as a function of defect type and defect position) for the individual sequence motifs provides new insight into the molecular mechanisms determining oligonucleotide duplex binding affinity on DNA microarrays.

## 6.3 DNA Microarray Design

The individual experiments performed with different microarray designs focus on the



**Figure 6.1:** Design of the experiment: a comprehensive set of point-mutated *probes* is derived from a common *probe sequence motif* which is complementary to the *target* sequence (*probe* sequences are shown for the first two defect positions only). For each defect position these include 3 single base mismatches (MMs - shown in red), 4 single base insertions (green), one single base deletion (red) and one perfectly matching (PM) control *probe* (blue). To enhance quantitative analysis, *probe* sets are arranged on the microarray as a compact *feature block*. Hybridization signal intensities from hybridization with the *target* sequence are plotted versus defect position. The *defect profile* shows relative binding affinities (i.e. the discrimination between the defect hybridization signal and the corresponding PM hybridization signal) as a function of defect type and defect position.

- extraction of the defect positional dependence,
- comparison of the binding affinities of different defect types, and on the
- identification of further influential parameters.

The individual chip designs employed differ in selection and spatial arrangement of the probe sequences.

### 6.3.1 Microarray Design Considerations for Quantitative Analysis of Hybridization Affinities

Several factors affect quantitative analysis of microarray hybridization signals: Spatial variations of the photo-deprotection intensity and optical aberrations affecting the imaging contrast can result in gradients of the probe DNA quality (as indicated in Fig. 6.2B). Depending on their position on the microarray, probes contain a varying degree of random synthesis errors. The corners of the rectangular synthesis area are most affected, since here the UV exposure dose, due to vignetting, is significantly smaller than in the center of the synthesis area.

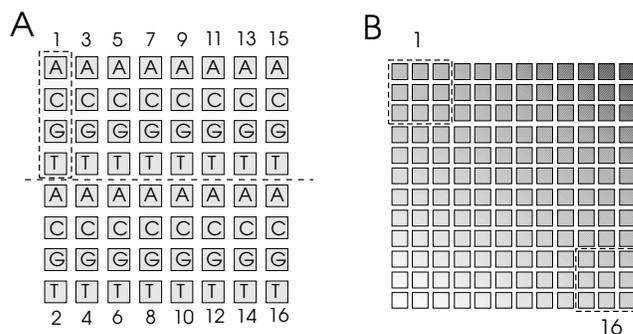
Gradients on the fluorescence intensity also arise from optical vignetting in the fluorescence microscope. This is largely compensated by shading correction (see section 5.1.3). To minimize impairments by gradients, probes for which hybridization signals are to be compared directly were arranged in closely spaced *feature blocks* (as shown in Figs. 6.1 and 6.2).

Local target depletion during hybridization (see section 8.5) can likewise result in position-dependent gradients of the hybridization signal intensity. In feature blocks with identical (or very similar) probe sequences, owing to the competition of the probes for the same pool of targets, features in the center of the block (surrounded by 8 competing features) - under unfavorable hybridization conditions [Pap06] - can have smaller hybridization signals than equivalent features at the edges of the feature block.

Control features (comprising perfect matching probes) which are evenly distributed over the feature block, are employed to indicate hybridization signal gradients: the variation of the PM signals (e.g. in Fig. 6.4A) shows the magnitude of feature-position dependent bias. Usually the impairment of the hybridization signal by such gradients is relatively small, resulting in variation of the control-probe intensities which is typically smaller than 5-10% of the PM hybridization signal intensity. However, if the hybridization kinetics is very fast - thus incoming targets are preferentially captured by the probes at the edge of the feature block - spatial variations of the hybridization signal of up to 50% of the PM intensity can occur [Pap06]. Unfavorable conditions affecting quantitative measurement are avoided by

using relatively short probes (rather 16mers than 25mers), using sequences with moderate binding affinities, and by application of sufficiently stringent hybridization conditions.

### 6.3.2 Single Base Defect Experiments



**Figure 6.2:** Microarray feature arrangement (**A**) for the single base mismatch experiment (compare with Fig. 6.3) and (**B**) for the direct comparison of various defect types. In **A** the feature block comprises 16 MM positions. The substitution base is either A, C, G or T. Depending on the *probe sequence motif* the substitutions will result in one PM and three MM probes. The design in **B** includes one PM, three MM, four single base insertion and one single base deletion probe four each of the 16 defect positions. The 9 probes belonging to each position are randomly arranged in a  $3 \times 3$  matrix (depicted by dashed boxes for defect positions 1 and 16). In this arrangement, as shown in (**B**), the gradient-related variation within the closely spaced  $3 \times 3$  feature group (belonging to a particular defect position) is significantly smaller than the variation between features (belonging to different defect positions) which are located further apart.

#### Single base mismatches

To investigate the positional dependence of single base mismatches and the impact of the mismatch type, we designed microarrays containing comprehensive sets of MM probes derived from a series of twenty-five 16mer probe sequence motifs. As described above, position and type of the mismatch base pair were systematically varied, allowing us later to distinguish between the dominating positional dependence and other influential factors. The features are arranged in groups of four, corresponding to the four possible substituent bases (A, C, G and T) at a particular base position. A group comprises three mismatch probes plus one perfect match probe used for control. Sixteen of these feature groups (one for each base position) are arranged in a square feature block comprising in total 64 features (Figs. 6.3 and 6.2A).

## Single base bulges

Probes containing single base insertions and deletions, owing to an unpaired unpaired nucleotide form bulged duplexes (see Fig. 2.16) with reduced stability. A comprehensive study on the impact of single base insertions was performed. The experiment comprised about 1000 single base insertion probes (insertion base type and position systematically varied) derived from twelve 20 to 25mer probe sequence motifs. The feature arrangement is similar to that in Fig. 6.2A.

## Direct comparison of single base MMs and single base bulges

Probe sets were derived from 16mer probe sequence motifs, complementary to the targets in Table 6.1. For each of the 16 possible defect positions a subset of 9 probes (comprising four single base insertions, one base deletion, three MMs and one PM probe) has been created. To prevent that regular arrangement of the defect types can create a systematic bias on measurement (e.g. due to increased target depletion near the PM probes), the subsets of 9 probes were randomly arranged in  $3 \times 3$  matrices as shown in Fig. 6.2B.

# 6.4 Hybridization Assays and Image Analysis

## 6.4.1 Oligonucleotide Targets

DNA and RNA target oligonucleotides (Tab. 6.1) were synthesized by MWG Biotech AG (Ebersberg, Germany) and by IBA Nucleic Acids Synthesis (Göttingen, Germany). 5'-Cy3 markers were attached in the final coupling step of the oligonucleotide synthesis via coupling of Cy3-phosphoramidite. The 3'-Cy3 modifications were produced postsynthetically by linkage of amino-reactive NHS-esters.

Gibbs free energies  $\Delta G_{37}^{\circ}$  and melting temperatures  $T_m$  of the PM-duplexes (predicted with the DINAMelt server - two-state hybridization) are provided in Tab. 6.2. Target secondary structure could not be avoided completely - in particular for the longer sequences and for the more stable RNA sequences. Possible target oligonucleotide secondary structure (loop and hairpin formation) was investigated with the DINAMelt Server [Mar05] (see Tab. 6.2).

## 6.5 Dominant Influence of the Defect Position

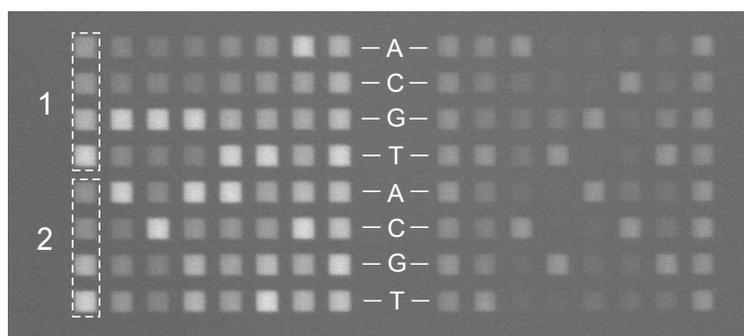
The "defect profile" plots (plots of the normalized hybridization signal vs. defect position - e.g. in Figs. 6.4 and 6.15) show that the dominant parameter determining oligonucleotide probe-target-affinity - on the microarray surface - is the position of the defect.

**Table 6.1:** Fluorescently labeled DNA and RNA target oligonucleotides

Name		Target sequence (5'→3')	Label	Length (nt)
URA	DNA	ACTACAACTTAGAGTGCAG... ...CAGAGGGGAGTGGAATTC	5'-Cy3	38
NIE	DNA	ACTCGCAAGCACCACCCTATCA	3'-Cy3	22
LBE	DNA	GTGATGCTTGTATGGAGGAA... ...TACTGCGATT	3'-Cy3	30
PET	DNA	ACATCAGTGCCTGTGTACTAGGAC	3'-Cy3	24
BEI	DNA	ACGGAAGTAAAGCAAAGAC	3'-Cy3	20
COM	DNA	AACTCGCTATAATGACCTGGACTG	5'-Cy3	24
NCO	DNA	TAGTGGGAGTTGTTAGTGATGTGA	3'-Cy3	24
PET	RNA	ACAUCAGUGCCUGUGUACUAGGACA	5'-Cy3	25
LBE	RNA	GUGAUGCUUGUAUGGAGGAA ...UACUGCGAUUCGAU	5'-Cy3	34
COM	RNA	AACUCGCUAUAUGACCGGACUG	5'-Cy3	24

**Table 6.2:** Gibbs free energies and melting temperatures of PM duplexes and target secondary structures (DINAMelt server [Mar05]), T=37°C, [Na<sup>+</sup>]= M, strand concentration 1 nM. The targets COM (DNA) and NCO don't form relevant secondary structures. For RNA/DNA duplexes no data on duplex stability is available (NDA).

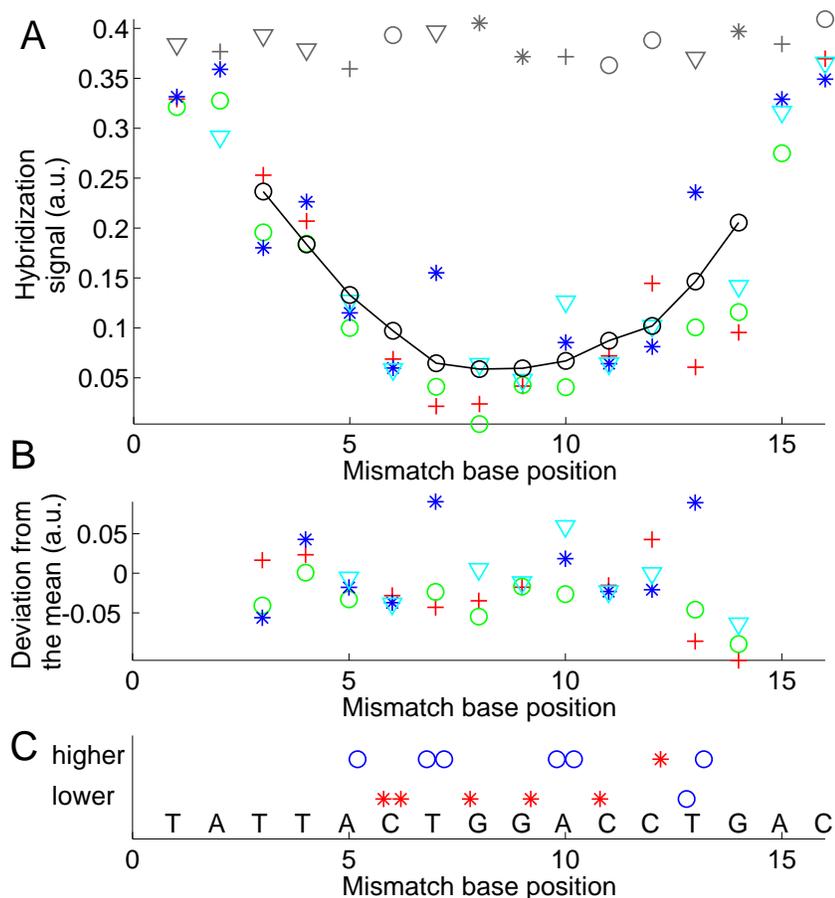
Target name	Duplex type	PM duplex		Target secondary structure	
		$\Delta G_{37}^{\circ}$ in kcal/mol	$T_m$ in °C	$\Delta G_{37}^{\circ}$ in kcal/mol	$T_m$ in °C
URA	DNA/DNA	-48.1	77.5	-0.1	40.1
NIE	DNA/DNA	-29.2	67.1	0.5	27.6
LBE	DNA/DNA	-36.6	70.7	-1.16	45.3
LBE	RNA/DNA	NDA	NDA	-7.1	63.0
PET	DNA/DNA	-29.6	66.1	-1.23	54.5
PET	RNA/DNA	NDA	NDA	-1.23	54.5
BEI	DNA/DNA	-24.2	59.6	0.08	35.1
COM	DNA/DNA	-28.7	64.5	-	-
COM	RNA/DNA	NDA	NDA	-0.1	37.4
NCO	DNA/DNA	-28.7	65.1	-	-



**Figure 6.3:** Fluorescence micrograph of two neighboring feature blocks in the 16mer mismatch experiment. The shading-corrected image shows two feature blocks corresponding to two different 16mer probe sequence motifs (3'-TTGAGCGATATTAAGT-5' to the left, and 3'-TATTACTGGACCTGAC-5' to the right) both hybridizing with the fluorescently labeled target sequence COM (5'-Cy3-AACTCGCTATAATGACCTGGACTG-3'). The different hybridization signal intensities of the two feature blocks are owing to different binding affinities of the two probe sequence motifs. The feature size is 21  $\mu\text{m}$ . Each feature block comprises all single base mismatches that can occur in the corresponding probe sequence motif. Groups of four features (as indicated by the marked groups 1 and 2) correspond to each one of the 16 possible mismatch base positions. As indicated by the letters between the feature blocks the uppermost row of features in each group corresponds to an A base at the corresponding base position, followed by probes with C, G and T (see also Fig. 6.2). The brightest feature within each group corresponds to the perfect matching probe. Nonhybridized targets in the hybridization solution contribute to the background intensity between the features. The "mismatch defect profile" for the probe sequence motif 3'-TATTACTGGACCTGAC-5' is shown in Fig. 6.4.

Defects near the duplex ends are distinctly less destabilizing than defects in the center of the duplex. As shown in Fig. 6.4 the hybridization signals of the individual mismatch probes are lined-up along the trough-like "mean profile" curve (*solid black line*). A parabolic fit can provide a reasonable approximation for the average position dependence obtained from a large number of different sequence motifs (as shown in [Wic06; Poz06]). The discrimination between PM and MM hybridization signals is largest if the defect is located in the middle of the duplex. For 16mer duplexes (as shown in Fig. 6.4) a single base mismatch (MM) in the center typically yields 0-40% of the perfect match (PM) hybridization signal, whereas at the duplex ends defects have significantly less impact on the hybridization signal.

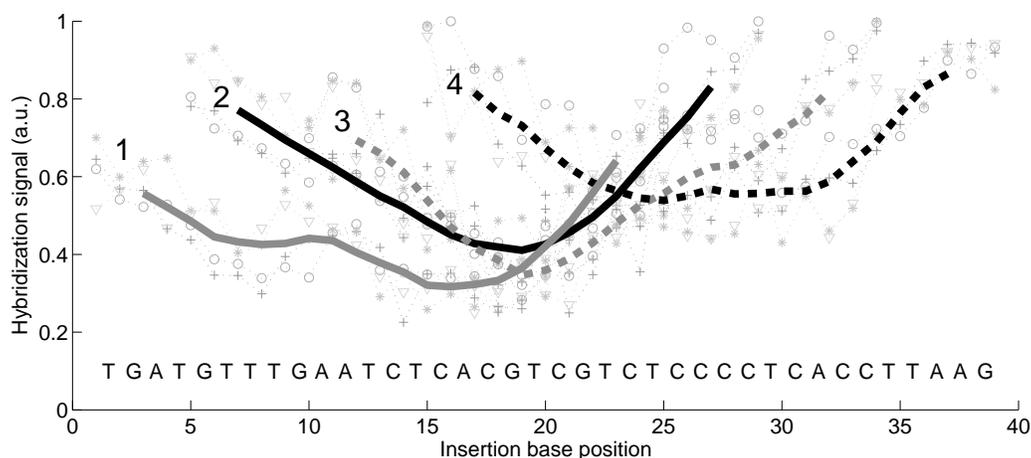
The discrimination between PM and point-mutated probes depends on the stability of the particular probe sequence motif: The more stable 25mer probes (shown in Fig. 6.15) are less discriminative than the shorter 16mer probes (Figs. 6.4A and 6.19). Reduced discrimination is also observed (see Fig. 6.5) for sequences which are stabilized by a high CG-content.



**Figure 6.4:** The mismatch defect profile (A) (hybridization signal versus defect base position) was obtained from the analysis of the hybridization signals of the feature block shown in the right part of Figure 6.3. The probe sequence motif 3'-TATTACTGGACCTGAC-5' is complementary to the target oligonucleotide COM. The different types of base substitutions are highlighted by different markers (A red crosses; C green circles; G blue stars; T cyan triangles). The black line indicates the mean profile (moving average of all mismatch hybridization signals over positions  $p - 2$  to  $p + 2$ ). PM probes (grey symbols) are used as a control to detect systematic bias (gradient effects) on the hybridization signal. The variation of the PM probe intensities also provides an estimate for the error of the measurement. Bias-related deviations between distant features, owing to gradient effects, are expected to be larger than the errors between the compactly arranged features corresponding to the same defect position. (B) Deviation profile. The strong position dependent component of the hybridization signal was eliminated by subtraction of the mean profile. In the following the hybridization signal deviation from the mean profile is referred to as  $\delta I_{mp}$ . (C) Comparison of mean mismatch hybridization signals (average of the three mismatch hybridization signals at a particular defect position) at the sites of C·G base pairs to mean MM hybridization signals at the site of adjacent A·T base pairs. A marker (red star: A·T; blue circle C·G) is set in the upper row if the hybridization signals of the mismatches at the corresponding site is higher than that at the adjacent site; otherwise a marker is set in the lower row. We noticed that mismatched base pairs substituting a C·G base pair usually have systematically lower hybridization signals than mismatches substituting a neighboring A·T base pair.

The positional influence observed in the mean profiles is largely determined by the defect-to-end distance, but is superimposed by a sequence dependent contribution. The variation of the shapes of the mean insertion profiles in Fig. 6.5 indicates that the impact of a defect is affected by the stability of the local sequence environment (i.e. not only by the next nearest neighbor base pairs). We discovered that single base bulge defects, originating from single base insertions (Fig. 6.15) and deletions (Fig. 6.19) - within the individual defect profiles - display the same positional dependence as single base mismatch defects. An attempt to explain the origin of defect positional influence is made in section 7.

To investigate other factors influencing oligonucleotide duplex binding affinity (e.g. defect type and defect neighborhood) the dominating positional influence needs to be eliminated. Design (selection and arrangement of probes) and analysis of our experiments enable separation of the different influential factors.



**Figure 6.5:** The impact of defects is affected by the local sequence environment. Normalized single base insertion profiles (hybridization signal plotted versus the insertion base position) of four 25mer probe sequence motifs complementary to the same target sequence (URA - shown below). The probe motifs 1 to 4 hybridize to different sections of the target oligonucleotide. Mean profiles (bold lines) were obtained from the moving average of the particular insertion profiles (individual hybridization signals are shown as faint grey symbols - profile 4 is shown in detail in Figure 6.15A). The mean profiles 1 to 3 have a distinct minimum between base positions 15 to 20. The stabilizing CG-rich region between base positions 15 and 33 is the reason for the reduced MM discrimination in profile 4.

## Discussion

We observe a dominating influence of the defect position on duplex binding affinity. Defects located in the center of the oligonucleotide duplexes are significantly more destabilizing than defects at the ends.

Strong influence of MM position has been reported previously mainly by other microarray based studies, but also from hybridization experiments in solution.

- From optical melting studies (on 7mer RNA/RNA duplexes in solution) Kierzek *et al.* [Kie99] report a 0.5 kcal/mol stabilization increment per each base position that the defect is closer to the helix end. A positional influence was observed for U·U and A·A, whereas the G·G mismatch stability was largely unaffected by the position.
- Dorris *et al.* [Dor03] found a similar positional influence for 2-base MM and 3-base MM probes on CodeLink 3D gel arrays. They also report a strong correlation (including the positional influence) between solution-phase melting temperatures and microarray hybridization signals of the MM duplexes.
- More recently Wick *et al.* [Wic06] and Pozhitkov *et al.* [Poz06] reported a strong influence of the defect position on the binding affinity of single base MM duplexes on DNA microarrays.

In accordance with [Poz06] we have identified MM position (relative to the duplex ends) as the strongest influential factor on the hybridization signal, when compared to MM-type (determined by the mismatch base pair  $X \cdot Y$ ) and nearest neighbors.<sup>5</sup>

To our knowledge only two studies ([Kie99] and [Dor03]) report a defect positional influence for hybridization in solution. This may partly be due to the unavailability of a large number of appropriate probes for a systematic study. So far the strong positional influence, mostly observed in microarray experiments, is unexplained.

The observation of a strong position dependence is in conflict with the two-state nearest-neighbor model of DNA duplex thermal stability, where the thermodynamics of internal mismatches is treated as independent of the MM position [San04]. Also, oligonucleotide duplex stability prediction software (based on a multi-state model) underestimates the MM positional influence when compared to microarray hybridization assays [Wic06].

For single base bulge defects we observed a very similar position dependence as for single base mismatches. Also, the magnitudes of the impacts of the MMs and base bulges on the hybridization signal are very similar (apart from the relative high binding affinity of Group II bulges). This consistency suggests a common origin of the positional influence, expected to be independent of the defect type.

Sterical crowding at the surface as discussed by Peterson *et al.* [Pet02] could possibly introduce a positional dependence on the hybridization signals of defect probes. Reduced accessibility of the probes surface-bound 3'-ends can in principle decrease the impact of

---

<sup>5</sup> According to the *nearest-neighbor model* the flanking base pairs towards both sides of the mismatched base pair  $X \cdot Y$  – just like the mismatched base pair itself – determine the base stacking interactions.

defects located near the 3'-end, and thus result in increased hybridization signals of the corresponding probes. This, however, runs contrary to the largely symmetrical intensity profiles observed (Fig. 6.4) and therefore does not provide a satisfactory explanation for the influence of defect position.

Focusing on individual probe sequence motifs, we observe, that the positional influence is not simply a function of the defect-to-end distance: it rather has a sequence-dependent contribution. This indicates that the mismatch discrimination could be affected by the stability of the nearest neighbor pairs between the defect and the proximate duplex end.

The observed influence of the duplex sequence and the symmetry of the defect positional influence with respect to both duplex ends suggest that end-domain opening (i.e. sequential unzipping of the double-helix from the duplex ends) is the key mechanism for understanding the influence of defect position on duplex stability.

## 6.6 Mismatch Discrimination in DNA/DNA Duplexes

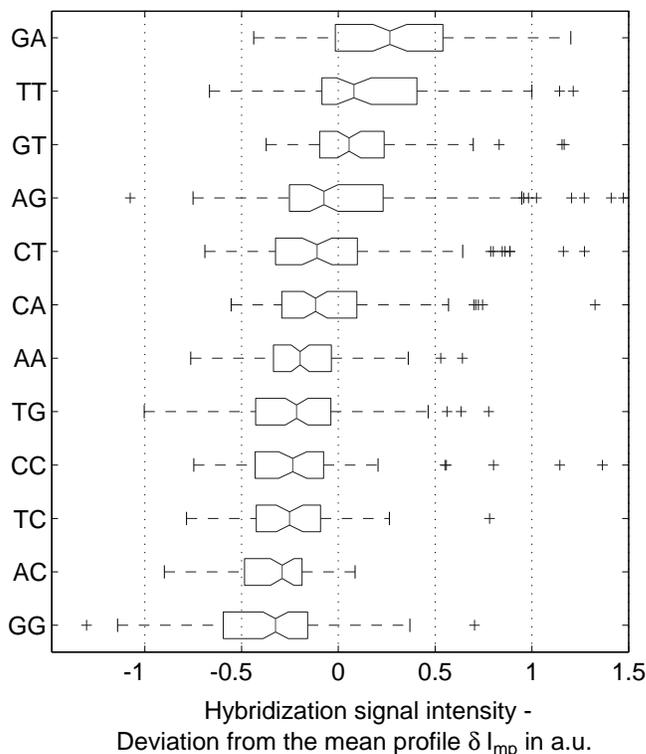
### 6.6.1 Experimental Results

For statistical analysis of MM type and nearest-neighbor influences the superimposed positional influence needs to be eliminated. This is achieved by subtraction of the (moving average) mean profile. In the following the hybridization signal deviation from the mean profile is referred to as  $\delta I_{mp}$ . The resulting position-independent defect profile (for simplicity we keep using the expression "defect profile") comprising defect-type and flanking base pair influences only, is shown in Fig. 6.4B.

In the following we use the notation of the mismatch base pair  $X\cdot Y$  consisting of the mismatched base  $X$  in the probe sequence and the base  $Y$  in the target sequence. In our experiments the systematic variation was restricted to the bases  $X$  in the microarray probe sequences. Since we had only a limited set of fluorescently labeled target oligonucleotides available (see Tab. 6.1) - the target sequences with the bases  $Y$  remained unchanged.

To investigate how the particular MM-types  $X\cdot Y$  affect duplex stability we measured probe-target-affinities for 25 different probe sequence motifs (distributed over three different microarrays). The PM hybridization signals of the 16mer probe sequence motifs display a strong variation (up to a factor of 20). The absolute hybridization signals from different probe sets are therefore not directly comparable. However, since the relative intensities (of the various MM probes) within the probe sets are largely unaffected by this variation, we can normalize the "position-independent defect profiles" by division by their standard

deviation. The resulting database comprises normalized hybridization signals (with the positional influence eliminated) from about 1000 different single MM probe sequences. The large database enables categorization of the hybridization signals according to the mismatch-type.



**Figure 6.6:** Mismatch-type dependent impact of single base MMs on the binding affinities of DNA/DNA oligonucleotide duplexes. Box-whisker plot representation of the hybridization signal distributions for the individual mismatch types, arranged according to the median values (depicted by the vertical line at the notch). More exactly, rather than hybridization signal intensities the plot shows the normalized hybridization signal deviations from the mean profile  $\delta I_{mp}$  - see Fig. 6.4B. Boxes indicate the interquartile range (from the 25th to 75th percentile) containing 50% of the data. Whiskers extend to a maximum value of 1.5 times the interquartile range from the boxes ends. Values beyond are classified as outliers. If the notches of two boxes do not overlap the medians values differ significantly with a 95 percent confidence. Histograms of the hybridization signal distributions are shown in Fig. A.11. The largest discrimination between PM and MM hybridization signals is observed for those mismatches where C·G base pairs are affected by the mispaired base (i.e. T·G, C·C, T·C, A·C, G·G). An exception is A·G. The positive tails of this and other distributions seem to originate from stabilizing C·G base pairs next to the defect.

The boxplot representation of this data in Fig. 6.6 demonstrates that MM-types affecting C·G base pairs (i.e. A·C, C·C, T·C and A·G, G·G, T·G) have consistently lower median hybridization signal values than those MM-types affecting A·T base pairs (A·A, C·A, G·A and C·T, G·T, T·T). In other words, the MM discrimination is systematically increased for

MM defects affecting C·G base pairs.

This explains the obvious differences between the distributions for the MMs A·C/C·A, A·G/G·A, T·C/C·T and T·G/G·T that have also been observed by [Poz06]. Although the mismatch types  $X·Y$  and  $Y·X$  could be thought to be equivalent (because the bases involved are the same), they result in different PM/MM hybridization signal ratios, depending on the type of PM-base pair (A·T or C·G) affected by the mismatch. For example, the impact of the MM A·C affecting an A·T base pair is (on average) smaller than the impact of the MM C·A affecting a C·G base pair. Therefore, the ratio of PM to MM hybridization signal intensities (i.e. the mismatch discrimination) is larger for the mismatch C·A than for the mismatch A·C :

$$I_{A·T}/I_{A·C} < I_{C·G}/I_{C·A}$$

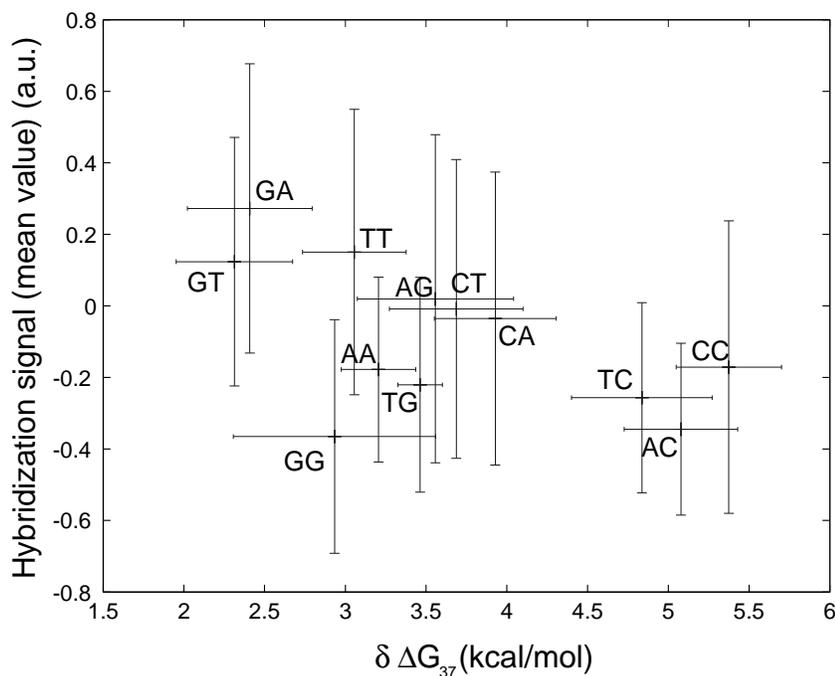
This can also be seen in the individual MM defect profiles: For a fixed sequence motif we compared the impact of MMs affecting a C·G base pair (average hybridization signal value calculated from the 3 MM probes - shown in Fig. 6.4B - of the corresponding defect position) to the impact of MMs affecting a directly adjacent A·T base pair. The analysis shown in Fig. 6.4C demonstrates that mismatches which substitute C·G base pairs are significantly more discriminating (MM hybridization signal about 5 to 10% with respect to the PM hybridization signal) than mismatches affecting a neighboring A·T base pair. The above results on DNA/DNA mismatches are in good agreement with [Wic06].

Fig. 6.7 demonstrates that the defect-type related deviations  $\delta I_{mp}$  from the mean MM profiles are correlated with the predicted Gibbs free energy differences  $\delta\Delta G_{37}^{\circ}$  between the MM and the PM duplexes<sup>6</sup>: The hybridization signal intensity is (with several exceptions) gradually decreasing with increasing  $\delta\Delta G_{37}^{\circ}$ . Thus, the experimentally observed MM discrimination on DNA microarrays is correlated with the free energy difference between MM and PM duplexes (calculated from nearest-neighbor free energy parameters [San98; All97]). A similar result has been reported in [Wic06].

The large discriminations for A·A, T·G and in particular G·G mismatches (as shown in Fig. 6.7) are not in agreement with the above established relation. In agreement with our results Wick *et al.* [Wic06] found A·A mismatches to be more destabilizing than 'predicted' by  $\delta\Delta G_{37}^{\circ}$ . Pozhitkov *et al.* [Poz06], in agreement with our study, reported G·G mismatches to be among the least stable MM defects.<sup>7</sup>

<sup>6</sup>  $\delta\Delta G_{37}^{\circ} = \Delta G_{37MM}^{\circ} - \Delta G_{37PM}^{\circ}$  was determined from MM nearest neighbor thermodynamic parameters [All97] for a temperature of 37°C

<sup>7</sup> However, in [Poz06] hybridization was performed with RNA targets (rather than DNA targets). Differences between RNA/DNA and DNA/DNA hybridization are discussed in section 6.8.



**Figure 6.7:** Correlation between MM-type-related impact on the hybridization signal intensity (mean values of the experimentally determined distributions of hybridization signal deviations  $\delta I_{mp}$  from Fig. 6.6) and predicted Gibbs free energy increments  $\delta \Delta G_{37}^{\circ}$  between MM and corresponding PM duplexes.  $\delta \Delta G_{37}^{\circ}$  was calculated from mismatch NN-parameters [All97]. Error bars account for flanking base pair related variation of  $\delta \Delta G_{37}^{\circ}$  (see Fig. 6.10). The experimentally measured MM discrimination for G·G, A·A and T·G is larger than predicted by the nearest-neighbor parameters. The influence of flanking base pairs (on both sides of the MM base pair) is considered in Fig. 6.11B.

## 6.6.2 Discussion

We observed that single-base MMs introduced at the site of a C·G base pair result in a larger decrease of the hybridization signal (with respect to the PM hybridization signal) than MM defects affecting A·T base pairs. The same applies for single base deletions (see Figs. 6.19 and 6.20). These experimental results, in accordance with nearest-neighbor thermodynamic parameters for Watson-Crick base pairs [San04], mainly reflect the increased base stacking and hydrogen bonding interactions of C·G base pairs. The effect largely determines the impact (i.e. the discrimination) of the different MM types  $X\cdot Y$  with respect to the perfect match hybridization signal of DNA/DNA duplexes. Our results for MMs in DNA/DNA duplexes (with the exception of the MM base pair G·G, which is the most destabilizing MM in our study) are in good agreement with Wick *et al.* [Wic06].

Differences between mismatch types  $X\cdot Y$  and  $Y\cdot X$  (a similar observation has been reported by [Poz06] for RNA/DNA hybrid duplexes) originate from the normalization with the corresponding PM hybridization signals. Again, the reason is that defects substituting the more stable C·G base pairs are more discriminating than defects affecting A·T base pairs. Remarkably, this increased discrimination has not been observed for RNA/DNA [Sug00; Poz06] and RNA/RNA [Sch06] hybridization (discussion below).

### Comparison of our experimental results to previous work

This section discusses our experimental results in the context of previous work. Main differences between the various studies discussed in the following are:

- hybridization on the microarray surface or in solution-phase
- hybridization of DNA/DNA, RNA/DNA or RNA/RNA duplexes
- lengths of the probe and target sequences (e.g. oligonucleotides, PCR products)

We emphasize the very good agreement between our MM stability order (Fig. 6.8e) and that of Wick *et al.* [Wic06] (see Fig. 6.8d). The only major difference is seen for the MM-pair G·G, which is the least stable in our study. In contrast Wick and Sugimoto [Sug00] found G·G to be relatively stable. Interestingly, Pozhitkov *et al.* [Poz06] in accordance with our results report G·G to be among the least stable MMs.

Wick *et al.* investigated the impact of single base MMs on the hybridization signal of DNA/DNA duplexes (Wick - Fig. 5a:  $\log_2(\text{PM/MM})$  values). The microarrays with 18-20mer probes were fabricated *in situ* by Xeotron (Houston, TX). Targets (PCR-products) were internally labeled with aminoallyl-dUTP. The hybridization was performed with a buffer consisting of  $6\times$ SSPE, 25% formamide.

Tautz and coworkers [Poz06] performed a similar microarray study with 20mer oligonu-

a) Solution hybridization DNA/RNA (Sugimoto *et al.* 2000)

**T·G>>G·U≈G·G>G·A≈A·G≈C·A>A·A≈T·U≈C·U>A·C≈T·C**

b) Microarray hybridization DNA/RNA (Pozhitkov *et al.* 2006)

**T·G≈T·U≈T·C>G·U≈A·C≈C·C≈C·U≈A·A≈A·G>C·A>G·G≈G·A**

c) Gene silencing RNA/RNA (Schwarz *et al.* 2006)

Silencing efficiency depends on the single base mismatch between  
the mRNA and siRNA sequences

**C·A>U·G≈C·U>U·U>A·C≈G·U≈C·C>U·C>G·A>G·G>A·A>A·G**

d) Microarray hybridization DNA/DNA (Wick *et al.* 2006)

**G·T≈G·A>T·T>G·G>T·G>A·G≈C·T≈A·A>C·A>A·C>C·C≈T·C**

e) Microarray hybridization DNA/DNA (this study)

**G·A>T·T>G·T>A·G>C·T≈C·A>A·A≈T·G≈C·C≈T·C>A·C>G·G**

**Figure 6.8:** Stability orders of MM-types  $X·Y$  for hybridization in solution (a) and on microarrays (b,d,e). In the microarray experiments (b,d and e) MM binding affinities have been normalized with the corresponding PM binding affinity, whereas the orders a) and c) reflect the absolute impact of the MM pairs on duplex binding affinity. Here (in series b,d and e) the probe base  $X$  is on the left and the target base  $Y$  is on the right. The efficiency of mRNA silencing (RNAi) (c) is determined by the stability of A-form RNA/RNA duplexes between the *RISC*-bound *guide strand* and the complementary mRNA [Sch06]. The left base  $X$  is part of the guide strand (position 10) and the right base  $Y$  is part of the mRNA. Apart from the base pair  $X·Y$  the mRNA and siRNA sequences remained fixed. In (a) to (c) purine bases are highlighted in blue. In (d) and (e) mismatches with respect to a perfect matching C·G base pair are highlighted in red. The MM stability order in (d) was extracted from the plot of  $\log_2(\text{PM/MM})$  hybridization signal values in *Fig. 5a* in [Wic06]. Further details on the individual stability orders are provided in the text.

cleotide microarrays fabricated by light-directed *in situ* synthesis with the Geniom One<sup>®</sup> instrument (Febit GmbH, Heidelberg). Similar as in our study they analyzed normalized hybridization signal intensities. The order of mismatch stabilities in [Poz06] (see Fig. 6.8b) indicates that purine-purine MMs (i.e. A·G, G·A and G·G) result in larger duplex destabilization than pyrimidine-pyrimidine pairs. According to [Poz06] the steric clash between the large double-ringed purine bases may cause an unfavorable distortion of the helix geometry, and thus result in increased duplex destabilization.

An important difference between the experiments described in [Poz06] and our hybridization experiments is the use of RNA targets. Therefore the results in [Poz06] refer to RNA/DNA hybridization rather than to DNA/DNA hybridization.

Remarkably, we found no significant correlation between the MM stability orders obtained from our DNA/DNA hybridization experiments (see Fig. 6.8e) and that reported for RNA/DNA hybridization in [Poz06]. An interesting question is whether the differences originate from different duplex structures - the B-form helix in our DNA/DNA study and

the A-form helix for the RNA/DNA duplexes in [Poz06].

Further differences with *respect* to our study: The cRNA transcripts originating from ribosomal RNA are significantly longer than the oligonucleotide targets employed in our study. In [Poz06] targets were labeled internally with Alexa Fluor-dUTP - different from the Cy3 fluorescent end-labels used in our experiments. Hybridization was performed in  $5\times$ SSC buffer. After hybridization the microarrays were washed to remove unbound target strands, whereas in our study the hybridization signal was acquired while the microarray was still immersed in the hybridization buffer.

A further study on the impact of MM stabilities in RNA/DNA duplexes, in solution rather than on a microarray surface, has been published by Sugimoto *et al.* [Sug00]. Pozhitkov *et al.* point out significant differences between their results and the results of Sugimoto *et al.* (Fig. 6.8a): In particular the destabilizing effect of purine-purine MMs described in [Poz06] is not observed by Sugimoto *et al.*.

The reported discrepancies between stability orders from Sugimoto (Fig. 6.8a) and Tautz [Poz06] could be interpreted that there are significant differences between hybridization in solution and on the microarray surface. However, the stability order in [Sug00] refers to  $\Delta G_{37}$  values of mismatched trinucleotide duplexes (i.e. to absolute stability parameters) whereas [Poz06; Wic06] and our study employ for each individual MM type the corresponding PM binding affinity as a reference level. Therefore the comparability of the RNA/DNA stability order in [Sug00] with the other studies discussed (our study, [Wic06] and [Poz06]) is somewhat limited.

Recent work on the impact of single base MMs in RNA-interference based gene silencing experiments [Sch06] is very interesting in the context of our study, since here the sequence recognition is based on base-pairing between the *guide strand* (a single RNA strand which is bound to the *RISC* complex) and a complementary mRNA.

Like nucleic acid hybridization RNA-interference (RNAi) is highly specific and can discriminate single base mutations. This is particularly interesting since several genetic disorders have been identified in which a point mutation affects only one allele of a gene, whereas the other (wild-type) allele is fully functional. Disease is caused by toxic properties of mutated protein products. Allele-specific gene silencing of the mutated gene is currently being investigated as a promising approach for gene therapy of dominantly inherited diseases (e.g. forms of Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis and Huntington's disease [RL06]).

For an application as a potential gene therapy it is important to understand how the guide strand has to be selected so that the mutant allele is silenced, whereas the wild-type allele of the gene remains functional. Schwarz *et al.* (see *table 5b* in [Sch06]) have shown that among all MM-types incorporated at position 10 of the guide strand (except for the point-

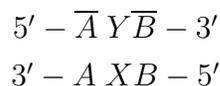
mutations the sequence of the guide strand was preserved) purine-purine MMs resulted in the least silencing of gene activity, whereas U·G, C·U and U·U mismatches resulted in a very efficient gene silencing (see Fig. 6.8c).<sup>8</sup> "A favored model is that purine-purine mismatches disrupt RISC activity by preventing the formation of a conventional A-form helix between the guide strand and the target mRNA, a structural requirement for RISC-mediated cleavage" [RL06]. Interestingly, the reported reduced stability of purine-purine mismatches is in good agreement with the findings of Pozhitkov *et al.*. However, the inferred RNA/RNA MM stability order in Fig. 6.8c, like that in in [Sug00], is not normalized with respect to the corresponding PM stabilities, but rather reflects the absolute impact of the MM base pairs in a given duplex sequence.

Differences between MM discrimination in DNA/DNA hybridization and RNA/DNA hybridization are not surprising since DNA/DNA duplexes (under the experimental conditions employed) occur as B-form helices, whereas RNA/DNA and RNA/RNA duplexes commonly occur as A-form helices (see Fig. 2.5).

The apparent discrepancy between the stability orders in the studies discussed above (see Fig. 6.8) motivated a systematic comparison of single base MM discrimination in DNA/DNA and RNA/DNA duplexes (see section 6.8).

## 6.7 Influence of Flanking Base Pairs on Single Base Mismatch Binding Affinities in DNA/DNA Microarray Hybridization

Due to stacking interactions the destabilizing impact of a mismatch defect not only depends on the MM base pair  $X·Y$ , but also on the flanking Watson-Crick base pairs  $A·\bar{A}$  and  $B·\bar{B}$  on both sides of the defect. [Alk82; Sug86].

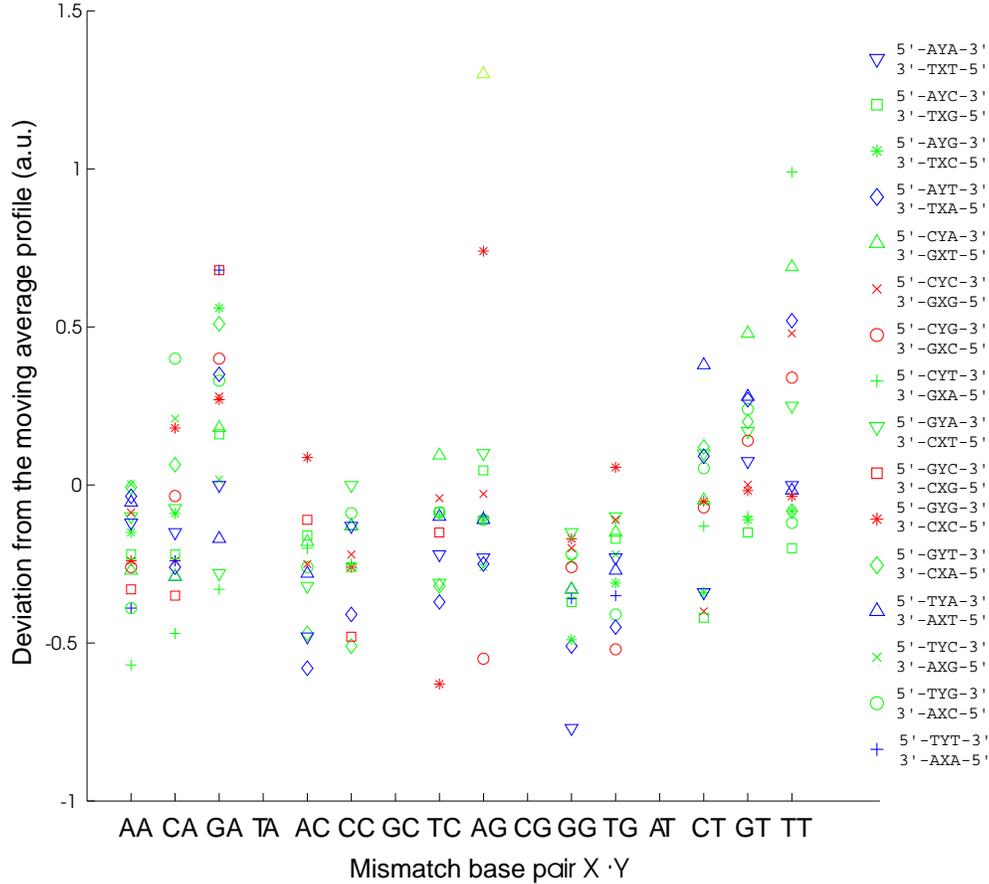


For a systematic study of the next-nearest-neighbor influence the mismatch hybridization signal data was categorized not only according to the the mismatch type (as discussed in section 6.6), but also according to the flanking base pairs at both sides of the mismatched base pair.

---

<sup>8</sup> [Sch06]: "Mismatches to be well accommodated in an A-form RNA/RNA helix (pyrimidine:pyrimidine, pyrimidine:purine, or purine:pyrimidine) displayed intermediate levels of discrimination, whereas purine:purine mismatches, expected either to destabilize the helix or to promote a stable, but nonhelical, conformation, silenced the reporter least."

There are 16 neighborhood classes (combinations of  $A \cdot \bar{A}$  and  $B \cdot \bar{B}$ ) for each of the 12 mismatch types  $X \cdot Y$ .

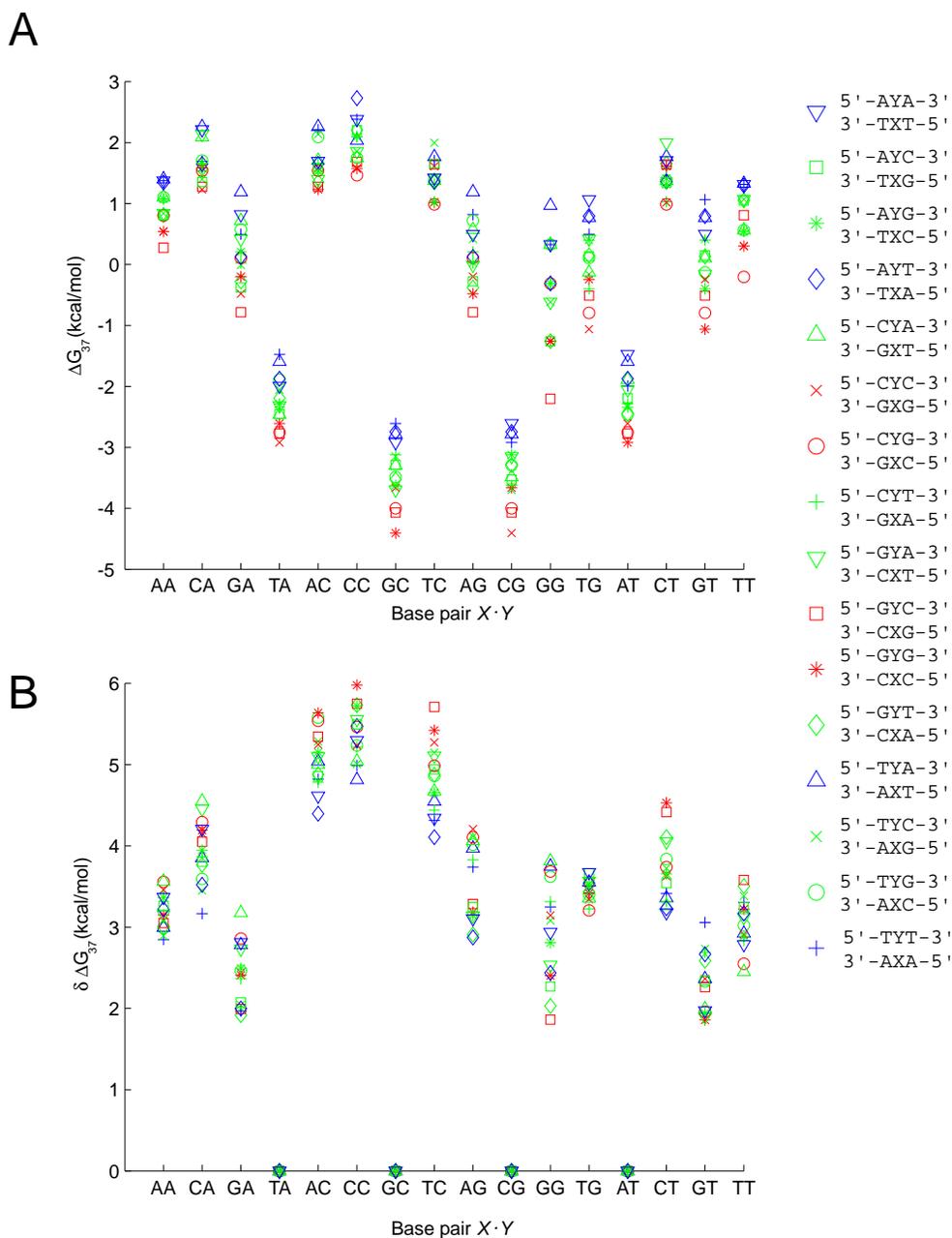


**Figure 6.9:** Distribution of the median hybridization signal values (deviation from the moving average profiles) of the various MM neighborhoods classes (see legend) as shown in Figs. A.12 - A.22. Red symbols denote C·G neighbors only, blue symbols denote A·T neighbors only. Green symbols correspond to mixed neighbors. The maximum value of about 1.3 a.u. for A·G MMs (green up-pointing triangle) is probably an outlier (only a single measurement was available for that particular MM class), whereas the value of 0.74 (red star) is based on 10 measurements. The significance of individual data points (which can be affected by lack of experimental data) can be evaluated from the corresponding histograms in Figs. A.12 - A.22.

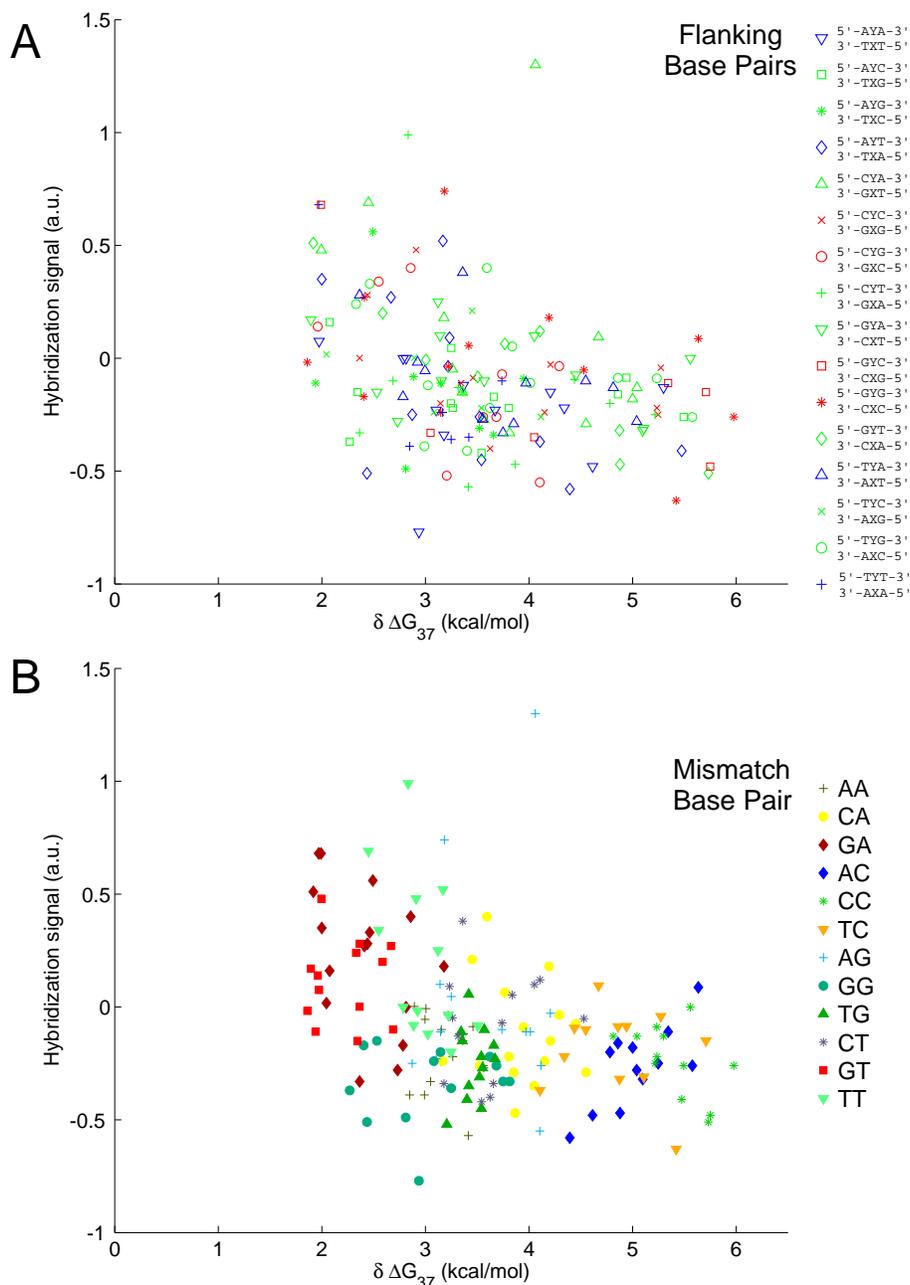
Splitting of the experimental data into 192 subsets (see Figs. A.12 - A.22) results in a relatively small statistical base for the individual MM classes ( $\rightarrow$  large statistical errors and sequence dependent bias). The significance of individual data points can be evaluated from the corresponding histograms in Figs. A.12 - A.22.

The median values of the neighborhood-dependent MM hybridization signal<sup>9</sup> distributions are shown in Fig. 6.9. To investigate if the experimentally observed influence of flank-

<sup>9</sup> normalized hybridization signals, positional influence eliminated



**Figure 6.10:** Influence of flanking base pairs on MM duplex stability. (A) Gibbs free energies  $\Delta G_{37}^{\circ}$  of mismatched and perfect-matching DNA/DNA trinucleotide duplexes were calculated from MM nearest-neighbor parameters [All97]. C·G flanking base pairs (red markers) are consistently stabilizing, whereas A·T flanking base pairs (blue markers) have a destabilizing influence. (B) Gibbs free energy increments  $\delta\Delta G_{37}^{\circ}$  between MM and corresponding PM duplexes. In the two-state nearest-neighbor model the discrimination between single base MM and PM duplexes only depends on the identity of the affected trinucleotide sequence (MM base pair and flanking base pairs).  $\delta\Delta G_{37}^{\circ}$  does not depend on the rest of the duplex sequence or on the position of the defect (unless the defect is located at a terminal position).



**Figure 6.11:** Comparison of MM hybridization signals (normalized with respect to PM hybridization signals - thus representing a measure for MM discrimination) with predicted Gibbs free energy increments  $\delta \Delta G_{37}^{\circ}$ . Hybridization signals (as shown in Fig. 6.9) are categorized according to MM base pair type and according to flanking base pairs. Each data point represents the median value of a distribution of hybridization signals (in detail shown in Figs. A.12 to A.22). We observe a significant correlation between the MM hybridization signal and the predicted Gibbs free energy increment  $\delta \Delta G_{37}^{\circ}$ . Part (A) highlights the influence of flanking base pairs on MM discrimination. Flanking A·T base pairs on both sides of the defect (blue symbols) result (on average) in smaller hybridization signals than C·G-only (red symbols) or mixed flanking base pairs (green symbols). However, the influence of flanking base pairs is little consistent compared with the influence of the MM base pair type, which is highlighted in (B): The discrimination of G·G, A·A and T·G mismatches is larger than predicted by MM nearest-neighbor parameters from [All97] and larger than in a similar experiment in [Wic06].

ing base pairs on binding affinities is in agreement with MM nearest-neighbor parameters [All97], we compared our experimental data (Fig. 6.9) to predicted free energy increments between MM and PM duplexes (Fig. 6.10): The MM nearest-neighbor parameters from [All97] predict a stabilizing influence of C·G flanking base pairs. Fig. 6.10A shows a consistently increased stability of those duplexes with C·G next nearest neighbors only, whereas a systematically decreased stability is seen for duplexes with A·T nearest neighbors only. For the predicted difference  $\delta\Delta G_{37}^{\circ}$  between PM and MM free energies - which is expected to be reflected in the experimentally determined MM discrimination - this consistency is somewhat reduced (see Fig. 6.10B). The comparison of  $\delta\Delta G_{37}^{\circ}$  with experimentally determined hybridization signals in Fig. 6.11A confirms a significant influence of flanking base pairs. On average, flanking A·T base pairs result in smaller hybridization signals than C·G or mixed flanking base pairs. However, the influence of the MM-base pairs  $X \cdot Y$  on the MM binding affinity (see Fig. 6.11B) is distinctly more consistent than the influence of flanking base pair types. A larger scale investigation of flanking base pair influence (based on a much larger set of oligonucleotide target sequences/probe sequence motifs) would be necessary to increase the statistical significance of the above results.

## 6.8 Mismatch Discrimination in DNA/DNA and RNA/DNA Duplexes - a Direct Comparison

To investigate if the above results from DNA/DNA hybridization also apply to hybridization of RNA/DNA duplexes we performed a direct comparison between DNA/DNA hybridization and RNA/DNA hybridization (employing DNA targets and equivalent RNA target sequences - see Tab. 6.1 ) on the same microarray.

### 6.8.1 Outline of the Experiment

The experiment is basically identical with the experiments described in section 6.6. Hybridization assays are conducted with fluorescently labeled DNA targets and corresponding RNA target sequences (Table 6.1). To avoid fabrication-related variation of the hybridization signals the DNA and RNA hybridization assays were performed on the same chip, first with RNA target oligonucleotides and - after regeneration of the microarray with NaOH (selective degradation of RNA targets) - with the corresponding DNA targets.

Three different microarrays were fabricated, each one focussing on one particular target sequence (*COM*, *PET* and *LBE*). The individual microarrays comprise single base MM and insertion probes ( $\rightarrow$ single base bulges) for 6 different probe sequence motifs (probing

different 16 to 20mer subsequences of the target sequence).

Two replicates of each feature block provide a test for the reproducibility of the measurement. The subsets of data obtained from the individual microarrays were analyzed independently to check the consistency of the observed results: apart from small sequence-related biases the three microarrays provided basically the same results. Hybridization was performed with 1 nM target solutions in 5×SSPE (0.01% Tween-20<sup>TM</sup>). Hybridization temperatures were 30°C for *PET* and *LBE* and 40°C for *COM* (for the target sequence *COM* the temperature had to be increased to 40°C since local depletion led to inhomogeneous hybridization - see section 8.5).

## 6.8.2 Results

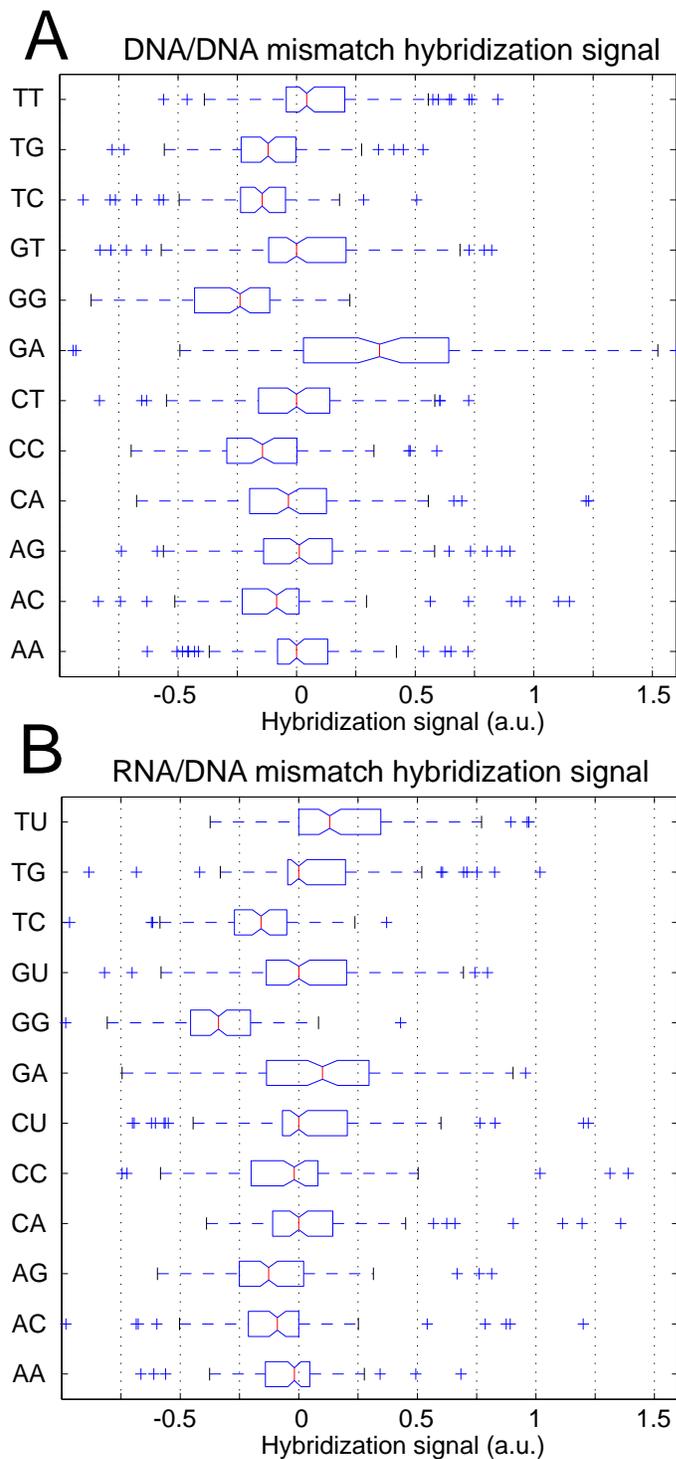
The influence of the defect position is very similar for the DNA/DNA and the RNA/DNA binding affinities (see Fig. A.1). However, there are small, though reproducible differences, as the comparison between replicate feature blocks (see Figs. A.2 - A.7) shows. For single base bulges no defect type specific differences between RNA/DNA and DNA/DNA hybridization were found.

We observed that under equivalent hybridization conditions the hybridization signal from RNA targets is on average about 1.3 times brighter than that of the corresponding DNA targets. This is anticipated: RNA targets have a slightly larger binding affinity than DNA targets since stacking interactions are stronger in A-form RNA/RNA and RNA/DNA duplexes than in B-DNA duplexes.<sup>10</sup>

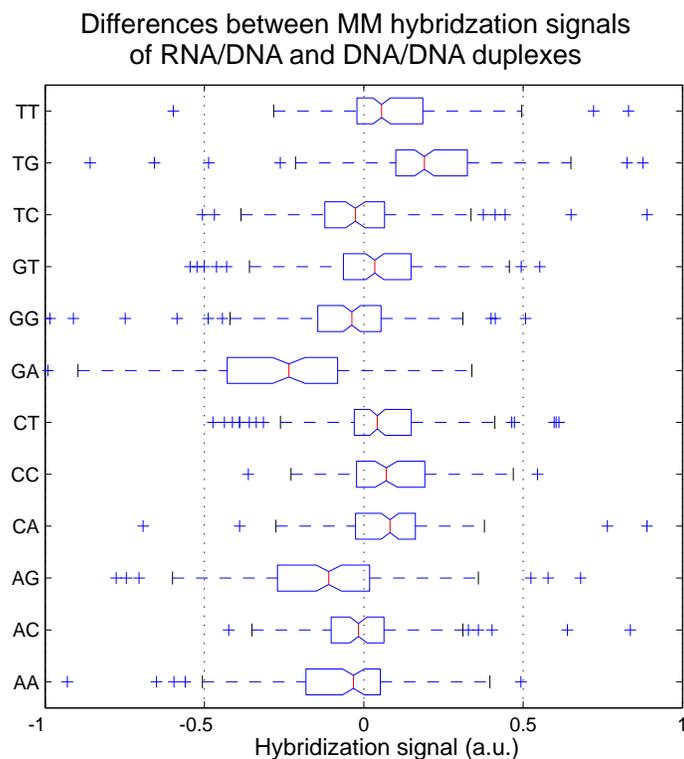
### Differences between MM stabilities in DNA/DNA and RNA/DNA duplexes

The MM discrimination in RNA/DNA duplexes (Fig. 6.12B) is very similar to that in DNA/DNA duplexes (Fig. 6.12A). However, a closer look reveals systematic differences between DNA/DNA and RNA/DNA hybridization. A statistical analysis (Figs. 6.12 and 6.14) revealed that purine-purine MMs are less stable in RNA/DNA duplexes (Fig. 6.14c) than in DNA/DNA duplexes (Fig. 6.14b). Three independent experiments (performed on different microarrays and with different probe/target sequences) provided the same trends. The decrease of purine-purine MM stabilities becomes obvious in the ranking order of differences between RNA/DNA and DNA/DNA MM stabilities (Fig. 6.14d). The largest differences between RNA/DNA and DNA/DNA MMs are observed for the MM-types G·A and A·G (which are more stable in DNA/DNA duplexes) and, with reversed sign, for the MM-type T·G, which is significantly more stable in RNA/DNA duplexes.

<sup>10</sup>Binding affinities: RNA/RNA > RNA/DNA > DNA/DNA



**Figure 6.12:** Comparison of DNA/DNA and RNA/DNA mismatch hybridization signals - statistical analysis. (A) MM-type related influence in DNA/DNA oligonucleotide duplexes. The positional influence was eliminated by subtraction of the moving average MM profile. Subsequent normalization was performed by division through the mean hybridization signal of the particular MM profile. (B) MM-type related influence in RNA/DNA oligonucleotide duplexes.



**Figure 6.13:** Differences between RNA/DNA and DNA/DNA MM binding affinities. Largest differences between RNA/DNA and DNA/DNA have been found for the MM-types T·G, G·A and A·G.

### 6.8.3 Discussion

Our investigation on the impact of MM-types in DNA/DNA oligonucleotide duplexes revealed that single base mismatches substituting C·G base pairs are more destabilizing than mismatches substituting A·T base pairs.

However, this seemingly plausible result (shown in Fig. 6.6) is not in general agreement with previous work [Sug00; Wic06; Poz06; Sch06] on the influence of the MM type on binding affinities.

Our direct comparison (“direct” in the sense of using the same probe sequences on the same microarray) between DNA/DNA and RNA/DNA hybridization on microarrays reveals - for RNA/DNA duplexes - an increased destabilization of purine-purine mismatches, with respect to other MM types. However, we did not observe such a distinct impact of purine-purine MMs as reported in [Poz06] and [Sch06]. Rather the MM stability order was very similar to that for DNA/DNA hybridization.

From MM stability orders in Figs. 6.14c and 6.14b (and Fig. 6.8e) we infer that the stability of MMs in RNA/DNA duplexes is determined by two factors:

- In RNA/DNA duplexes purine-purine MMs tend to be more destabilizing (with respect

- a) DNA/DNA hybridization (large data set)  
**G·A>T·T≥G·T>A·G≥C·T≈C·A>A·A≈T·G≈C·C≈T·C≥A·C≥G·G**
- b) DNA/DNA hybridization (small data set for direct comparison with RNA/DNA hybridization)  
**G·A>T·T>A·G≈C·T≈G·T≈A·A>C·A>A·C≥T·G≥T·C≈C·C>G·G**
- c) RNA/DNA hybridization (small data set - equivalent to the DNA/DNA dataset in b)  
**T·U≥G·A>T·G≈G·U≈C·U≈C·A≈C·C≈A·A>A·C≥A·G≥T·C>G·G**
- d) Difference between RNA/DNA and DNA/RNA hybridization signals. Uracil is treated as thymine. (TG to GT positive; AC to GA negative)  
**T·G>C·A≥C·C≥T·T>C·T≈G·T>A·C≈T·C≥A·A≈G·G>A·G>G·A**

**Figure 6.14:** Ranking orders of DNA/DNA MM stabilities in comparison with that of RNA/DNA MMs. (a) For comparison the DNA/DNA MM stability order from an independent experiment (Fig. 6.8) is shown here again. (b) As anticipated the ranking order for DNA/DNA MMs obtained from the smaller data set which is used for the direct comparison between DNA/DNA and RNA/DNA hybridization (Fig. 6.12A) is very similar. The ranking order for RNA/DNA mismatch stabilities (c) (extracted from Fig. 6.12B) reveals significant differences with respect to (b). In part (d) MM-types are ordered according to the hybridization signal differences between RNA/DNA and DNA/DNA MMs (extracted from Fig. 6.12 A and B). Purine bases are highlighted in blue.

to other MM-types) than purine-purine MMs in DNA/DNA duplexes.

- The influence of the "affected base pair" - the base pair which has been substituted by the MM base pair - is the other factor that determines the impact of the MM type. In the experiments the PM hybridization signal is used as a reference value for the reduction of the hybridization signal due the MM defect. In agreement with [Wic06] we observed that MMs affecting C·G base pairs are more discriminating than MMs affecting A·T base pairs.

In the order of RNA/DNA mismatch stabilities (Fig. 6.14c) the latter effect is superimposed by the destabilizing effect of purine-purine MMs, whereas in DNA/DNA duplexes (Fig. 6.14b - our results - in agreement with [Wic06] - see Fig. 6.8d) an increased destabilization of purine-purine MMs is not observed.

An explanation for the observed differences between DNA/DNA and RNA/DNA binding affinities is, that purine-purine MMs cause larger steric hindrance in the A-form hybrid duplexes than in the B-form DNA/DNA duplexes.

In this study, like in [Poz02], a destabilizing impact of purine-purine MMs was observed in RNA/DNA hybridization. However, we found only a slightly increased destabilization with respect to the corresponding purine-purine MMs in DNA/DNA duplexes, whereas [Poz02] and [Sch06] reported that purine-purine MMs - in absolute terms - are the most

discriminating MMs with respect to other MM-types.<sup>11</sup> Further studies will be necessary to resolve the remaining discrepancy.

A more detailed future investigation of MM stabilities should also focus on the influence of the flanking base pairs. This, however, will require a significantly larger database of MM hybridization signals.

## 6.9 Single Base Bulge Defects

Single base insertions and deletions, owing to a surplus unpaired base in one of the two strands, result in bulged duplexes, which like MM duplexes have a reduced binding affinity. In duplexes with single base insertion probes the bulged base is located on the surface-bound probe strand, whereas in duplexes with single base deletion probes the bulged base is located on the target strand.

The positional dependence of the insertion intensity profiles (Figure 6.15A) is very similar to the mismatch intensity profile in Figure 6.4, though the individual insertion profiles (for example the profile of C-insertions - green circles in Figure 6.15) show large deviations from the (moving average) mean profile.

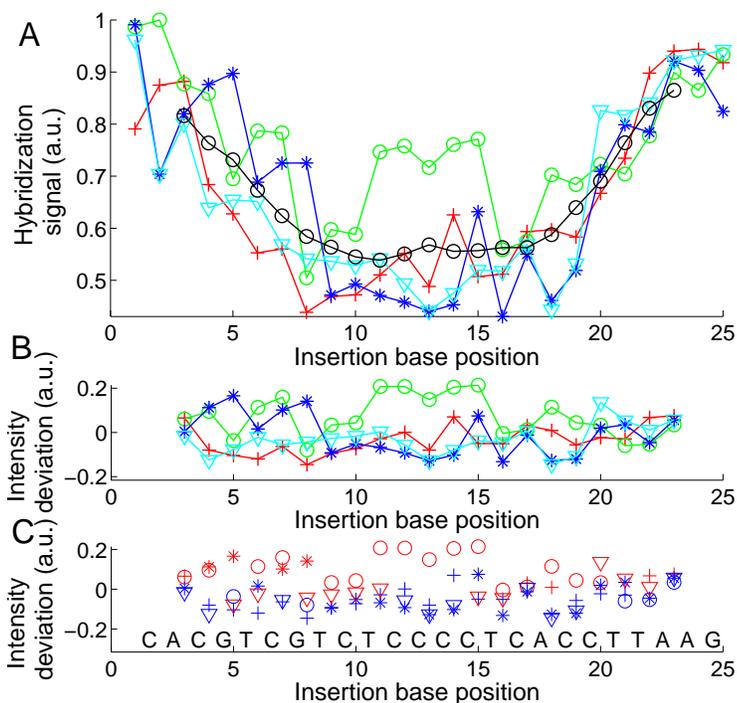
Hybridization signals can be significantly increased over two or more consecutive defect positions. In particular, base insertions next to identical bases (*Group II bulges* [Zhu99]) result in systematically increased binding affinities - in comparison to insertions of non-identical bases (*Group I bulges*). In the notation of Zhu *et al.* [Zhu99] bulged bases without an identical neighboring base (Fig. 2.17A) are defined as *Group I bulges*, whereas bulges with at least one identical neighboring base (Fig. 2.17B) are referred to as *Group II bulges*. Increased stability of duplexes with *Group II bulges* in solution-phase experiments has been described by Ke *et al.* [Ke95]. Fig. 6.15C demonstrates the systematically increased binding affinity of *Group II bulges* in DNA microarray hybridization.

### 6.9.1 Statistical Analysis

The observed stabilization of *Group II bulges* (in comparison to *Group I bulges*) in our microarray experiments is surprisingly large (see discussion below): *Group II bulges* located near the center of 16mer probes often show hybridization signals with a similar intensity as the corresponding PM probe, whereas *Group I bulges* at the same defect position have a significantly smaller binding affinity, with a similar level as single base MMs at the corre-

---

<sup>11</sup>These studies, however, investigated only DNA/RNA hybridization and RNA/RNA hybrids (RNAi: A-form helix between the guide strand and the target mRNA), respectively. No comparison with DNA/DNA hybridization was made.

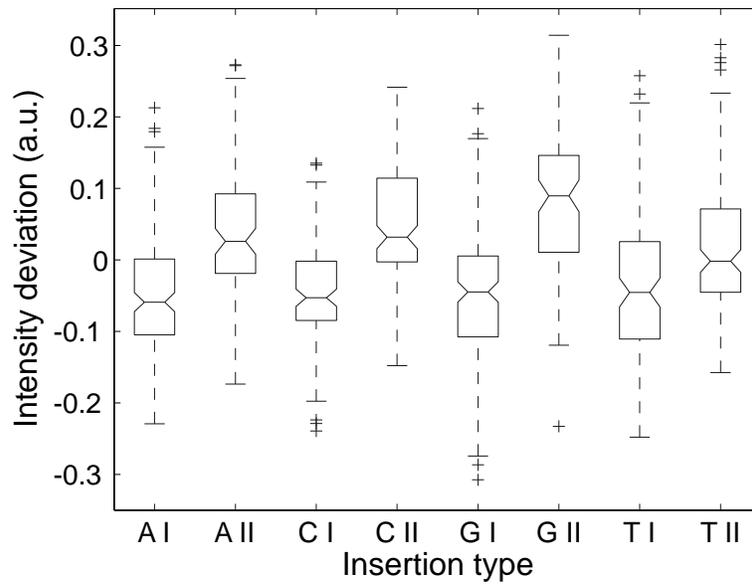


**Figure 6.15:** (A) Single base insertion defect profile (hybridization signal plotted versus the insertion base position) of the probe sequence motif 3'-CACGTCGTCTCCCTCACCTTAAG-5' (complementary to the target *URA*). Symbols correspond to insertion bases (A red crosses; C green circles; G blue stars; T cyan triangles). The mean profile (black line), obtained from the moving average (including all 4 insertion types) over positions  $p - 2$  to  $p + 2$  describes the defect positional influence. (B) and (C) Positional influence is eliminated by subtraction of the mean profile. Elevated intensities are observed for *Group II bulges* - see text - (e.g. C insertions at positions 11 to 15, 6 to 7 and 18 to 20 or G insertions at positions 4 to 5 and 7 to 8). A very distinct increase of the hybridization signal is observed for C insertions into the C-rich subsequence TCCCCT. *Group II bulges* (red markers) have significantly higher intensities compared to *Group I bulges* (blue markers). Further examples are shown in Figs. A.8 - A.10.

sponding defect position (see Fig. 6.19).

A statistical analysis with a dataset (Fig. 6.16) comprising hybridization signal data from 1000 different 20-25mer probes proves the general validity of the above observations: *Group II* hybridization signals are significantly increased with respect to *Group I* hybridization signals. The median normalized hybridization signals of *Group I* insertions do not significantly vary with the type of the bulged base. The largest difference  $\delta I_{bulge}$  between *Group I* and *Group II* hybridization signals is observed for G-insertions.  $\delta I_{bulge}$  is smallest for T-insertions.

A similar experiment performed with 16mer probes (results of the statistical analysis shown in Fig. 6.20) shows some differences with respect to Fig. 6.16: In the 16mer experiment



**Figure 6.16:** Box-whisker plots show the hybridization signal deviations (from the mean profile) for the different insertion base types, which are differentiated according to affiliation to bulge *Group I/II*. The statistical analysis includes about 1000 normalized hybridization signals from 12 different 20 to 25mer probe sequence motifs.

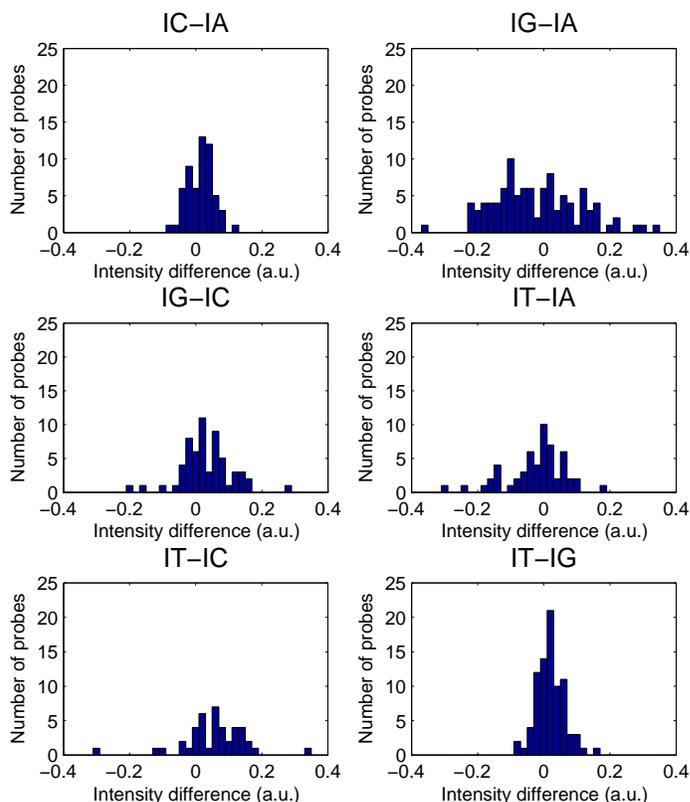
(Fig. 6.20) hybridization signals of *Group I* adenine-insertions are significantly reduced compared to other *Group I* insertion types. The largest difference  $\delta I_{bulge}$  between *Group I* and *Group II* hybridization signals is observed for A- and G-insertions ( $\delta I_{bulge}$  up to 35% of the PM hybridization signal), whereas a significantly smaller  $\delta I_{bulge}$  is observed for C-insertions ( $\delta I_{bulge} \simeq 8\%$  of the PM value) and T-insertions ( $\delta I_{bulge} \simeq 3\%$  of the PM value). The larger variation between different insertion types in Fig. 6.20 may be explained by the larger relative impact of the defect (owing to the shorter duplex lengths).

*Group II bulges* originating from single base deletions in the microarray probe sequences display increased binding affinities as well (Fig. 6.19, orange dashed line). However, the  $\delta I_{bulge}$  for *Group II bulges* originating from single base deletions - unexpectedly - is distinctly smaller than for single base insertions. The *deletion defect profiles* are largely within the hybridization signal range spanned by the single base mismatch defects.

Systematically increased hybridization signals (with respect to the averaged hybridization signal level from other defect types at the same position) have also been observed for certain *Group I bulges*. For guanine-insertions next to thymine bases (e.g. in Fig. 6.19 at base position 15) we found significantly increased hybridization signals. It seems that the insertion of G next to a T, similar like the insertion of a T next to another T (*Group II bulge*), results in an increased binding affinity in comparison to other *Group I bulges*.

We further investigated the degree of correlation of binding affinities between probes with different insertion bases *X* and *Y* (Fig. 6.17). A distinct correlation appears between the

hybridization signals of probes with T- and G-insertions, and also, though less distinct, between A- and C-insertions. In contrast to that, our results show an anti-correlation between G- and A-insertions.



**Figure 6.17:** Histograms of hybridization signal differences  $I_X - I_Y$  ( $X$  and  $Y$  denote the different insertion bases in otherwise identical probe sequences) reveal correlations between the hybridization signals of different insertion types. To exclude the impact of systematically increased intensities of *Group II* insertions only *Group I* insertions are regarded here. Between T- and G-insertions (and between C- and A-insertions) a correlation, as indicated by a narrow distribution with a pronounced peak near zero, is observed. The broad distribution of hybridization signals differences between G and A insertions doesn't show a distinct peak, indicating that there is no correlation but rather an anti-correlation for insertions of A and G.

## 6.9.2 Discussion

We observe significantly increased hybridization signals of single-base insertion defects in which the insertion base is placed next to a like-base. The increased stability of *Group II* bulges in comparison with *Group I* bulges has been investigated previously, however, in solution rather than on microarrays, by [Ke95; Zhu99; Zno02]. According to Ke and Wartell

[Ke95] the increased stability of *Group II* bulges originates from positional degeneracy of the extra unpaired base. Additional conformational freedom, entailing higher entropy, results in lowered duplex free energy. According to Zhu *et al.* [Zhu99] position degeneracy accounts for an average stabilization of  $-0.3$  to  $-0.4$  kcal/mol (in agreement with the theoretical estimate [Zhu99] of  $-R \cdot T \cdot \ln 2 = -0.43$  kcal/mol at  $37^\circ\text{C}$ ) for a two position degeneracy.

Znosko *et al.* [Zno02] reported *Group II* duplexes to be on average  $\delta\Delta G^{37} = -0.8$  kcal/mol more stable than *Group I* duplexes. The latter value matches our observation of significantly increased binding affinities of *Group II* bulges better, since the *Group II* hybridization signals observed were often close to the perfect match hybridization signal.

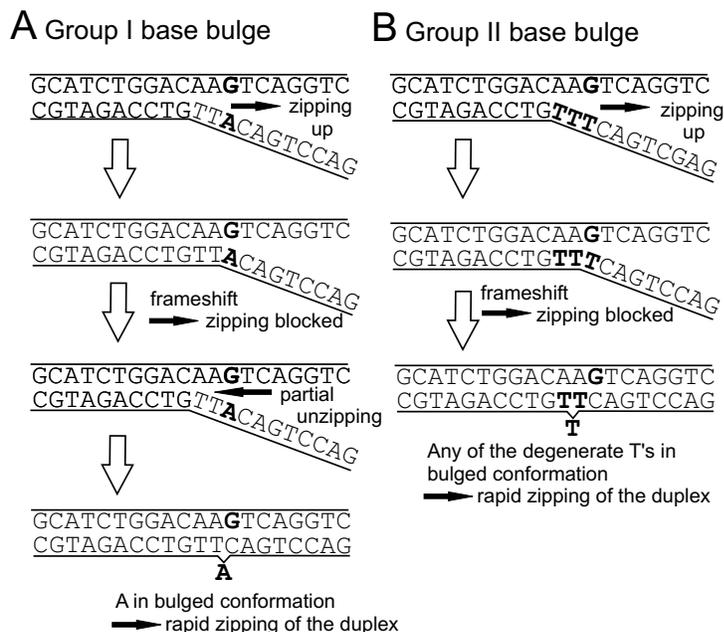
Our investigation shows that (on the microarray) the difference between *Group I* and *Group II* binding affinities  $\delta I_{bulge}$  (inferred from the hybridization signal  $I$ ) is distinctly larger than the defect-type related variation of binding affinities  $\delta I_{MM}$  (see Fig. 6.20).

However, free energy increments between *Group I* and *Group II bulges* previously reported ( $\delta\Delta G_{bulge}^{37} = -0.3$  to  $-0.4$  kcal/mol [Zhu99];  $\delta\Delta G_{bulge}^{37} = -0.8$  kcal/mol [Zno02]) are significantly smaller than the variation of the nearest neighbor duplex free energies within the mismatch defect profiles investigated in our experiments. The MM duplex free energies in a MM defect profile (calculated with MM nearest neighbor parameters of Allawi *et al.* [All97]) vary within a range of  $-\delta\Delta G_{MM}^{37} \simeq 4$  kcal/mol. The standard deviation - with respect to the mean MM free energy - is about 1 kcal/mol.

Thus, there is a discrepancy between our experimental results on DNA microarrays (where  $\delta I_{bulge} > \delta I_{MM}$ ) and the previous estimates of  $\delta\Delta G_{bulge}^{37}$ , since  $\delta\Delta G_{bulge}^{37} < \delta\Delta G_{MM}^{37}$ . One would rather expect  $\delta\Delta G_{bulge}^{37} > \delta\Delta G_{MM}^{37}$ . Therefore, in the context of our experimental results, a value of  $\delta\Delta G_{bulge}^{37} \simeq -0.4$  kcal/mol appears to be too small. This indicates that the model of *Group II bulge* stabilization by entropy increase due to positional entropy may be incomplete.

For explanation of the surprisingly large binding affinity of *Group II* duplexes we postulate the following mechanism (illustrated in Fig. 6.18) based on a molecular zipper model [Gib59; Kit69] of the oligonucleotide duplex:

The surplus (bulged) base acts as a kinetic barrier, interrupting the rapid zipping (consecutive base pairing) of the duplex. The frameshift between the complementary sequences, owing to the unpaired nucleotide prevents hybridization beyond the defect and results in a partially zipped, and correspondingly weakly-bound, duplex. Duplex closure can only progress if the interfering surplus base is giving way (i.e. adopts a favorable looped-out or stacked conformation), thus allowing the subsequent base to form a Watson-Crick base pair with the corresponding complementary base in the target strand. From this point on, the zipping can progress rapidly. Compared to Watson-Crick nearest-neighbor pairs, a

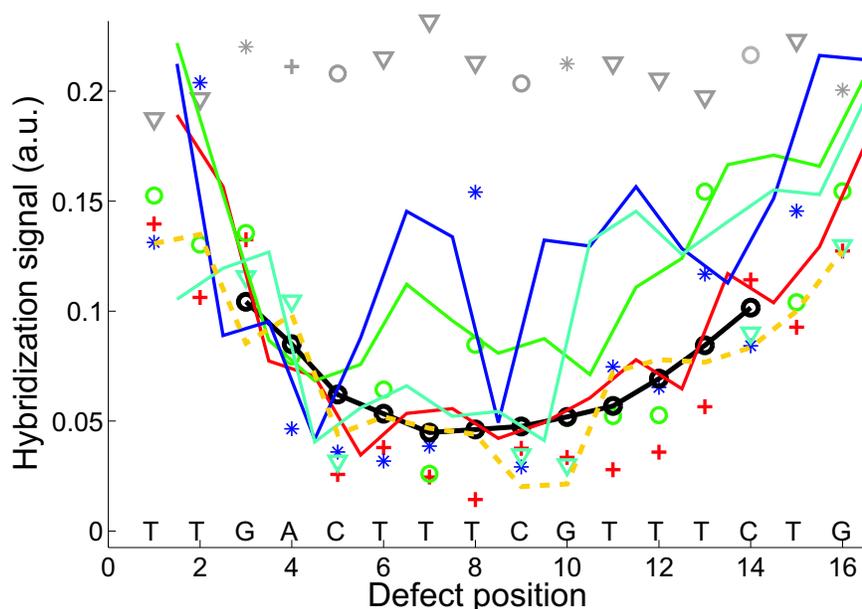


**Figure 6.18:** Proposed mechanism for the increased binding affinity of *Group II base bulges*: A) Destabilizing impact of *Group I base bulges*: The bulge originating from the unpaired base 'A', creates a frameshift of one nucleotide between the complementary probe and target sections, and thus acts like a barrier delaying the formation of a stable duplex. The bulged 'A' needs to adopt a favorable (e.g. looped out) conformation, so that the frameshift is compensated and the zipping of complementary base pairs can continue. Unlike the *Group I base bulge* in A) the *Group II base bulge* in B), originating from the insertion of the surplus base 'T' next to another 'T', is degenerate. The zipping is interrupted at the defect site, which is located at the end of the group of degenerate (identical) bases. As with *Group I* sequences, at the barrier partial unzipping is likely to occur. However, since there is an increased probability that any of the degenerate bases adopts a looped-out or stacked conformation, the formation of a stable duplex is accelerated. Therefore, in duplexes with *Group II* insertions the barrier which is trapping the duplex in a weakly bound partially zipped state can be overcome faster, resulting in increased stability.

bulge defect, similar to a mismatch base pair, decreases the ratio of zipping/unzipping-rates  $k_+/k_-$  of adjacent nearest neighbor pairs. The bulge increases the duplex dissociation rate and thus leads to a reduced duplex binding affinity. For *Group II bulges* the  $k_+/k_-$  ratio is increased with respect to *Group I bulges*: The zipping is delayed at the defect site, which is located at the end of the group of degenerate (identical) bases: As in *Group I* duplexes, partial unzipping will occur at the barrier. However, there is an increased probability that any of the degenerate bases make way (i.e. adopt a bulge conformation) and allow the subsequent base to form a base pair. Since the frameshift is now compensated, the rapid zipping, resulting in a stabilized duplex, can continue.

## 6.10 Comparison of Single Base Mismatches and Single Base Bulges

A direct comparison of *single base mismatch* and *single base bulge* binding affinities (see Figure 6.19) reveals that the *defect positional influence* is largely independent of the defect type. The microarray design employed for this experiment (as shown in Fig. 6.2B) comprises adjacent features for MMs and base bulges to enable a "direct comparison" between the two defect types in the same microarray experiment.



**Figure 6.19:** Position dependent impact of various single base defects on the hybridization affinity for the probe sequence motif 3'-TTGACTTTCGTTTCTG-5' (hybridized with the complementary target sequence *BEI*). The "defect profile" reveals the very similar defect positional influence of single base mismatches, insertions and deletions on duplex binding affinity. Symbols: MM probes with substituent bases A (red crosses), C (green circles), G (blue stars), T (cyan triangles); moving average of all MM intensities (black line); single base insertion probes (solid lines) with insertion bases A (red), C (green), G (blue), T (cyan). Hybridization signals of single base deletions (orange dashed line) are similar to that of MMs at the same position. Perfectly matching (PM) probe replicates (grey symbols), represent a means of quality control, indicating possible gradients on the microarray. Deviations of MM hybridization signals from the mean profile are largely MM-type-specific. Increased hybridization signals of particular insertion probes (in which the extra unpaired base has been inserted next to an identical base  $\rightarrow$  *Group II bulge* [Zhu99]) are due to positional degeneracy of the bulge defects.

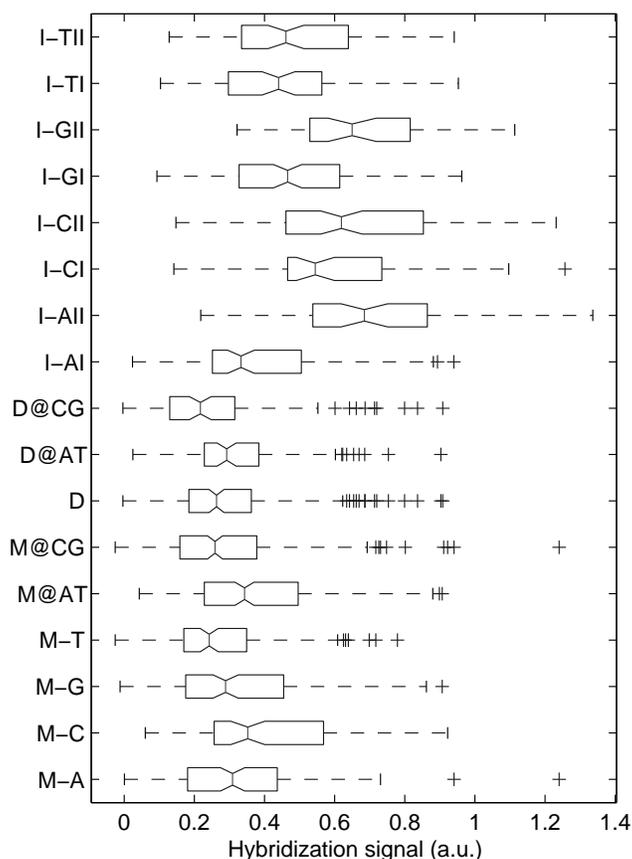
A statistical analysis is shown in Fig. 6.20: Single base insertion probes provide (on aver-

age) significantly larger hybridization signals than MM probes at the corresponding defect position. This may be explained by the reduced number of binding base pairs in the mismatched duplexes (which have one binding base pair less than the PM duplex, whereas a single base insertion leaves the number of binding base pairs unchanged) and by the significantly increased hybridization signals of Group II insertions.

Hybridization signals of MMs 'replacing' C·G base pairs<sup>12</sup> are about 25% smaller (in the median) than those of MMs 'replacing' A·T base pairs. Similarly, single base deletions affecting C·G base pairs result in about 30% smaller hybridization signals than deletions affecting A·T base pairs. This can also directly be observed in the deletion profile in Figure 6.19 (orange dashed line), where the local variations (ups and downs) of the profile curve correlate with deletions affecting either A·T or C·G base pairs. No similar effect is observed for single base insertions because no binding base pair is "destroyed" by the insertion of an extra nucleotide.

---

<sup>12</sup>The perfect matching (PM) duplex has a C·G or G·C base pair at the corresponding position. In the MM duplex the probe is mutated with respect to the PM probe, i.e. the perfect matching base, either C or G, has been substituted by another base which is not complementary to the corresponding base in the target sequence, thus creating a single base MM defect in the duplex (see Fig. 6.1).

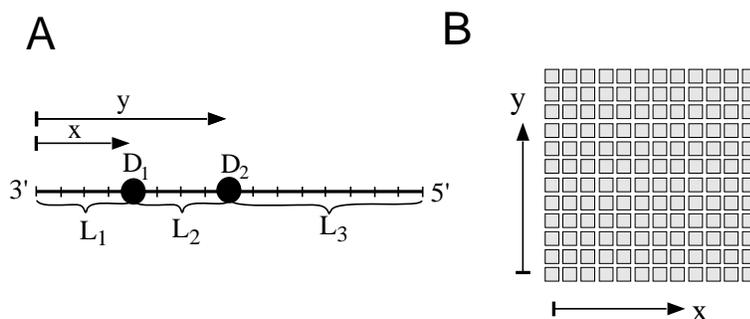


**Figure 6.20:** Comparison of normalized hybridization signals of different point mutation types. To minimize positional influence the statistics include only defect positions 5 to 12, located in the center of the 16mer probes. The 1200 probe sequences were derived from 17 probe sequence motifs. Hybridization signals are normalized with respect to the corresponding perfect match hybridization signal intensities (thus a value of 1 corresponds to the PM hybridization signal intensity). Defect categories: mismatch M-X (X: substituent base); mismatches at A·T and C·G sites M@AT, M@CG; single base deletion D; deletions at A·T and C·G sites D@AT, D@CG; single base insertion I-XI/II (X: insertion base, I/II: *Group I/Group II base bulge*). Hybridization signals from insertion probes (about 50% of the PM hybridization signal for *Group I*; 65% for *Group II* - median values) are significantly higher than that of MM probes (at about 30%). Mismatches at A·T sites result in about 25% larger hybridization signals than MMs at C·G sites. Deletion probes have a median hybridization signal that is slightly lower than the median MM hybridization signal. *Group I base bulges* with the exception of I-AI (33%) have hybridization signals of about 50% of the PM hybridization signal. Hybridization signals of *Group II base bulges* are significantly higher (about 100% for A insertions, and only 5% for T insertions) than that of the corresponding *Group I bulges*.

## 6.11 Binding Affinities of Duplexes Containing Multiple Defects

A significant fraction of the microarray probes generated in the light-directed *in situ* synthesis process contains multiple point-defects (due to stray light, incomplete coupling and incomplete photodeprotection). Hybridization experiments with microarray probes containing multiple (deliberately introduced) defects<sup>13</sup> were performed to investigate to which extend these probes – depending on the number and distribution of defects – contribute to the hybridization signal.

### Influence of the spatial distribution of two defects on oligonucleotide duplex binding affinities



**Figure 6.21:** (A) In the two-defect-experiment defects  $D_1$  and  $D_2$  at varying positions  $x$  and  $y$  divide the duplex into three subsequences of length  $L_i$ . (B) The probe set comprising all configurations of the two defects (in duplicate) is arranged in a compact array. Feature positions (indices  $x$  and  $y$ ) correspond to the defect positions. Compare with experimental results in Figs. 6.22 and Fig. 6.24.

To investigate the influence of the spatial distribution of defects on the microarray, probe-target binding affinities we designed probe sets comprising all two-deletion mutations with respect to the corresponding 20mer probe sequence motifs: Deletions  $D_1$  and  $D_2$  (as shown in Fig. 6.21A) were introduced at positions  $x$  and  $y$ . The positions of the defects were independently varied from base positions 1 to 20, resulting in a  $20 \times 20$  matrix (Figs. 6.21B and 6.22) of 400 probes comprising all two-deletion probes in duplicate (plus 20 single base deletion probes - for  $x$  and  $y$  coinciding).

To extract the influence of the two-defect configuration on the hybridization signal averaging was performed over a set of nine different 20mer motifs. This was necessary to eliminate sequence specific bias (mainly composed of variations owing to increased/reduced

<sup>13</sup> Defects were deliberately introduced into the sequence (with respect to the perfect matching probe motif), in addition to the unavoidable random defects generated *in situ* synthesis process

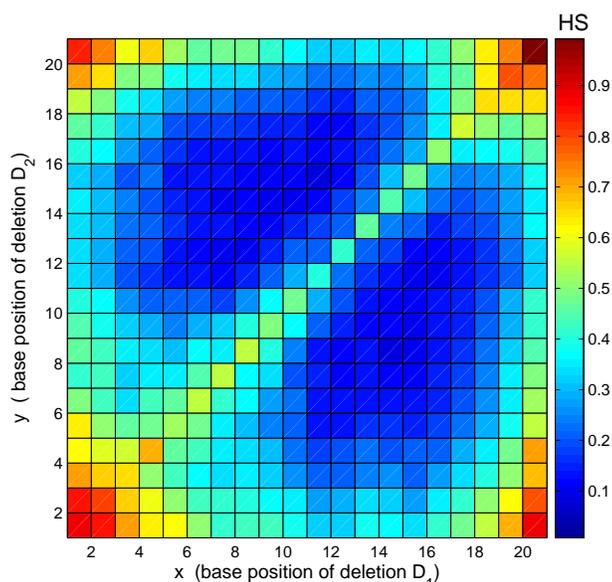
destabilization of defects affecting C·G/A·T base pairs).

Hybridization assays were performed under standard hybridization conditions used in most of the experiments (1 nM target solution in  $5\times$ SSPE, 0.01% Tween-20<sup>TM</sup>, hybridization temperature  $T = 30^{\circ}\text{C}$  to  $40^{\circ}\text{C}$ ).

### Multiple defect experiment

For probes containing more than two defects we applied a statistical approach. Based on a 20mer probe motif (complementary to the target sequence *PET*) we created sets of randomly mutated probes, each containing containing probes with a fixed number (between one and five) of single base deletions at random positions.

#### 6.11.1 Results and Discussion



**Figure 6.22:** Hybridization signals of 20mer probes (normalized with respect to the maximum hybridization signal, which approximately corresponds to the PM hybridization signal) with two single base deletion defects  $D_1$  and  $D_2$  at varying positions  $x$  and  $y$  (compare to Fig. 6.2D). Averaging over data sets obtained from 9 different probe sequence motifs has been performed to eliminate nonpositional contributions (e.g. differences resulting from deletions affecting either A·T or C·G base pairs) from the hybridization signal. The resulting data set shows the influence of the defect distribution on the hybridization signal. Defects at the probe 3'-end (base position 1) affect the hybridization signal slightly less than defects at the 5'-end.

In two-deletion experiments we determined the hybridization signals of 20mer probes with systematically varied configurations of two single-base deletions: The binding affin-

ity is largest when both defects are located close to the same end or separate near both ends (Fig. 6.22). Lowest hybridization intensities are observed for defect configurations dividing the sequence into three roughly equally long subsequences. Closely spaced defects (with a distance of less than 4 bases - located near the diagonal of the plot) result in increased hybridization signals approaching that of single base deletions as the distance between the defects is reduced.

The hybridization experiments with probes containing a varying number of point-defects (base deletions or mismatches) at randomly chosen positions<sup>14</sup> show a broad distribution of binding affinities (see Fig. 6.23B) depending on the number and also on the spatial distribution of the defects.

Hybridization signals of multi-defect probes are described by the empirical relationship

$$f = a \cdot \sum L_i^\nu + b. \quad (6.1)$$

$L_i$  denote the lengths of defect-free subsequences,  $a$  and  $b$  are free parameters. To account for the fact that longer subsequences contribute disproportionately more to the binding affinity than shorter ones, the exponent  $\nu$  is introduced. Effectively  $\nu$  is putting a length dependent weighting factor on the individual lengths  $L_i$ .

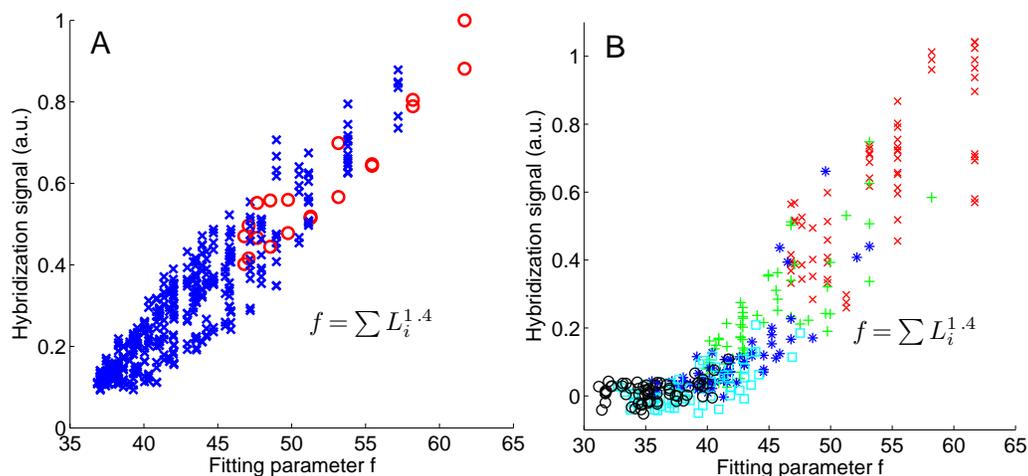
In Fig. 6.23A the hybridization signal intensities of the two-deletion experiment were plotted versus the parameter  $f$  from equation (6.1). As shown in Fig. 6.23B equation (6.1) also predicts hybridization signals for probes with a larger number of deletions.

However, the above considerations were purely empirical. With the zipper-model (see chapter 7 ) we were able to model the experimentally observed binding affinities on a physical basis. The hybridization signal intensities (Fig. 6.24A) are approximately proportional to the Gibbs free energies (Fig. 6.24B) determined from the zipper-model (for explanation see section 7.4).

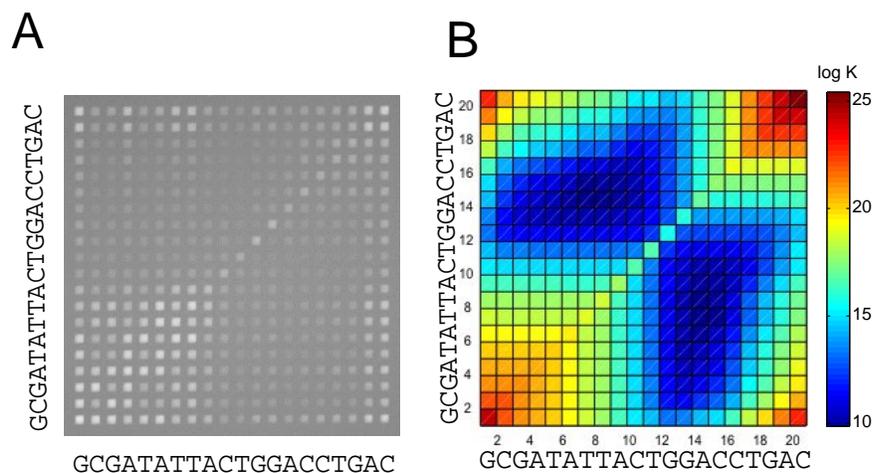
The impact of multiple defects is not additive (as suggested by the nearest neighbor model) but rather depends on the distribution of the defects. Probes with two or more randomly introduced synthesis-defects can have a significant binding affinity if the defects are located close to the duplex ends.

---

<sup>14</sup>The particular probe sequences contain intentionally introduced single base defects. Defects with respect to a common perfect matching probe sequence motif were introduced at randomly chosen base positions.



**Figure 6.23:** Fitting of the hybridization signals of multiple-defect probes. **(A)** Hybridization signals obtained from the two-deletion experiment are plotted versus the fitting parameter  $f = \sum L_i^{1.4}$  ( $L_i$ : length of defect-free subsequences - see Fig. 6.2C) (two deletions: blue crosses; single deletion - for  $x$  and  $y$  coinciding: red circles). **(B)** Similar experiment with a varying number of deletions (at randomly chosen positions) in the 20mer probe sequence motif 3'-TAGTCACGGACACATGATCC-5'. Marker types indicate the number of deletions: 1 red crosses; 2 green crosses; 3 blue stars; 4 cyan squares; 5 black circles). Because only data from a single probe sequence motif was available, non-positional (sequence-related) contributions couldn't be eliminated, thus resulting in increased scattering of the hybridization signal intensities.



**Figure 6.24:** Two-deletion experiment: Systematic variation of the positions of two single base deletions in the probe sequence motif 3'-GCGATATTACTGGACCTGAC-5'. **(A)** Fluorescence micrograph of the hybridization signals. Feature arrangement according to Fig. 6.21. **(B)** Corresponding binding affinities determined with the zipper model (see section 7.3). The color scale is proportional to the logarithm of the binding constant  $K$ . We observe a good agreement with the experimental results in **(A)**.



# Chapter 7

## Modeling the Influence of Point Defects on Oligonucleotide Duplex Stability

### 7.1 The Double-Ended Zipper Model

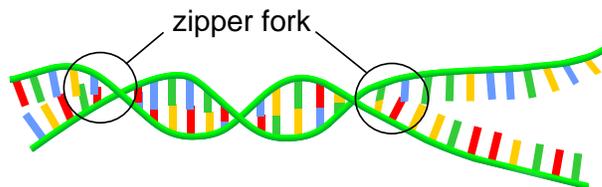
The analysis of the defect profiles in chapter 6 revealed that defect positional influence (DPI) does not just depend on the distance between the defect and the duplex-ends, but also on the nucleotide sequence (see Fig. 6.5). Our results also show that DPI is basically identical for single base mismatches and bulge defects (see Fig. 6.19).

This finding suggests a common mechanism for DPI, that is independent of the defect type. The symmetry of DPI (with respect to the duplex ends) and sequence-specific deviations from the symmetry indicate a zipping-related mechanism.

Rather than to hybridize/denaturate in an all-or-none reaction (as assumed in the simplified two-state model) the oligonucleotide duplex can only sequentially form base pairs or dissociate in a zipper-like fashion.

For our model of oligonucleotide duplex stability we assume that unzipping of the duplex is initiated at the ends only (see Fig. 7.1). Internal denaturation, due to the large bubble initiation barrier (owing to stacking interactions towards both sides of a nucleotide) and due to the relatively short length of the duplexes (throughout this study  $l_D \leq 25$  base pairs), is expected to be negligible [Gib59].

Presuming pure end-domain opening the probability for complete unzipping (resulting in strand dissociation) decreases exponentially with duplex length. However, at sufficiently high temperature (i.e. when  $T\Delta S \simeq \Delta H$ ) denaturation bubbles can more easily open in the interior of the duplex. With increasing duplex length, thus increased melting tempera-



**Figure 7.1:** Zipper-model of the oligonucleotide duplex. The prevailing mode of oligonucleotide duplex hybridization/denaturation is based on a zipper-mechanism. The base pair next to the zipper fork is stabilized by stacking interactions towards one side only, whereas base pairs in the interior of the duplex are stabilized by stacking interactions with the two neighboring base pairs. Due to smaller stacking interactions and due to structural constraints (rigid double helix structure) unzipping occurs mainly at the zipper-forks, whereas internal denaturation - resulting in the opening of a *denaturation bubble* - for short oligonucleotide duplexes is unlikely to occur. Vice versa the sequential closure of the base pairs (zipping) - under suitable hybridization conditions - rapidly propagates via the zipper forks: at the zipper fork the initially far separated bases are brought close together and favorably aligned, thus strongly increasing the probability for Watson-Crick base pair formation.

ture<sup>1</sup>, denaturation by internal bubble formation is eventually dominating over end-domain opening. Therefore long duplexes dissociate mainly via the formation of internal denaturation bubbles. The melting transition of long duplexes is described by the Poland-Scheraga model and the more recent Peyrard-Bishop model (see section 2.3.4).

Provided the individual strands don't form stable secondary structures (e.g. hairpins) and that there are no competing alternative duplex structures other than the linear duplex, the *double-ended zipper model* (Fig. 7.5A) [Gib59; Kit69; Bin06] is appropriate to describe the stability of oligonucleotide duplexes.

In the following we employ the double-ended zipper model to investigate if the experimentally observed defect positional influence could arise from a molecular zipper mechanism. Unlike software for RNA/DNA secondary structure prediction (e.g. MFold or the Vienna Package) the relatively simple model investigated is not meant to determine a potentially complex secondary structure of a duplex. Rather, the application of the zipper model is restricted to short linear duplexes.

On the basis of the zipper model two approaches have been followed:

- The straightforward stochastic simulation (section 7.2) based on the Gillespie algorithm [Gil77] simulates the zipping/unzipping of the individual base pairs. This approach is, however, computationally intensive: the large number of zipping/unzipping steps necessary for a complete duplex dissociation restricts the stochastic simulation to rather

<sup>1</sup> Long duplexes have increased melting temperatures with respect to short oligonucleotide duplexes. However, with increasing duplex length the melting temperature is approaching a saturation value

short-lived duplexes.

- The partition function approach (section 7.3), which describes the equilibrium distribution of partially denatured duplex states, is not subject to the above restriction.

## 7.2 Stochastic Simulation of Oligonucleotide Duplex Stability

### 7.2.1 Implementation of the Stochastic Simulation with the Gillespie Algorithm

In the zipper model the time evolution of the duplex is described by a Markov process: the positions of the zipper forks move stochastically in a (biased) random walk fashion (Fig. 7.2). Unzipping rates are determined by nearest neighbor (NN) interactions, whereas zipping rates of Watson-Crick base pairs are (for simplicity - no data available) assumed to be independent of the type of NN pairs.

In our model, defects are expected to reduce binding interactions at the defect site, and – what is possibly more important – owing to steric hindrance of the mispaired nucleotides, to delay the zipping process: the duplex remains for longer in a weakly bound, partially zipped state in which it is prone to complete dissociation.

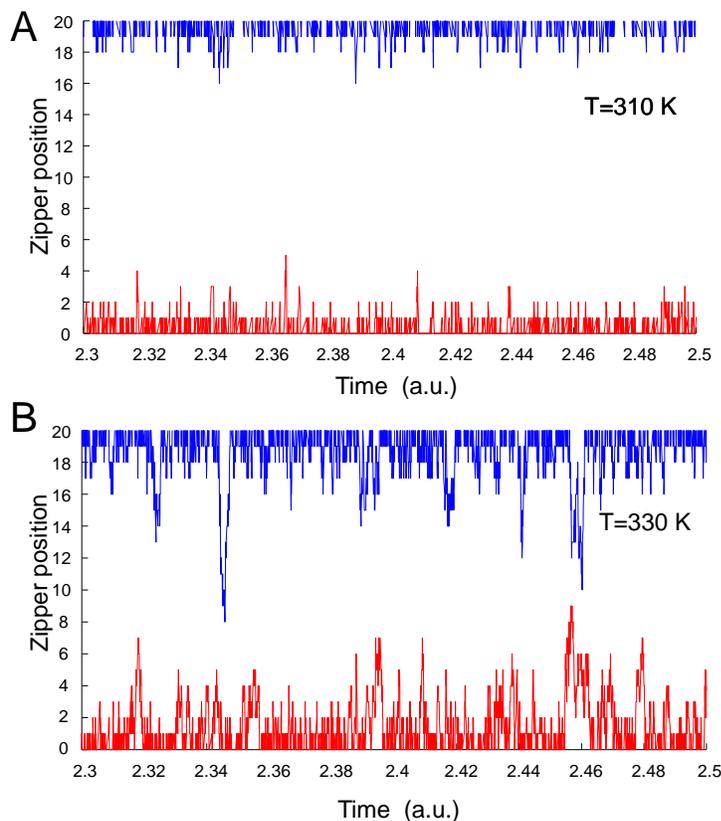
The time evolution of the zipper is simulated with the Gillespie algorithm [Gil77]. In each cycle the stochastic algorithm determines two parameters:

- the time  $\tau$ , how long the zipper (after the preceding step) remains in its current state until the next zipping/unzipping step takes place<sup>2</sup>,
- and which one of the four possible reaction steps  $\mu$  ( $\mu=1$ : zipping at the right end;  $\mu=2$ : unzipping at the right end;  $\mu=3$ : zipping at the left end;  $\mu=4$ : unzipping at the left end) will occur next.

The four possible reaction pathways for zipping/unzipping at the right/left zipper fork are characterized by their rate constants  $k_{r+}$ ,  $k_{r-}$ ,  $k_{l+}$ ,  $k_{l-}$ . Unzipping rates of base pairs (according to the Arrhenius law) are proportional to the Boltzmann factor for base pair dissociation.

$$k_{r/l-} = A \cdot e^{\Delta G^{\circ}/(RT)} \quad (7.1)$$

<sup>2</sup> The Gillespie algorithm doesn't use fixed time steps. Rather it draws a pair of random variables from the reaction probability density function, specifying the the next reaction step and the time it takes until this reaction will occur.



**Figure 7.2:** Simulated time evolution of the end-domain opening for a 20 base pair oligonucleotide duplex (A) at  $T=310^{\circ}\text{C}$  and (B) at  $T=330^{\circ}\text{C}$ . The positions of the left (red) and right (red) zipper fork (corresponding to  $k$  and  $l$  in Fig. 7.5A) follow a biased random walk. The ratio of zipping to unzipping steps is determined by the temperature and by the strength of the nearest neighbor interactions between the individual base pairs. In the Gillespie-based stochastic simulation the time steps follow a poisson-distribution. Strand dissociation is assumed when the zipper forks meet. Strand dissociation is a rather unlikely event at a temperature of 310 K - as can be seen in (A) most of the time only the outermost base pairs are unzipped.

Nearest neighbor (NN) free energies  $\Delta G^{\circ}$  (accounting for hydrogen bonding and stacking interactions) are calculated from Watson-Crick NN thermodynamic parameters [San98]. The preexponential factor  $A$  is assumed (with a large uncertainty<sup>3</sup>) to be at the order of  $10^6 \text{ s}^{-1}$ .

The zipping rate  $k_{r/l+}$ , owing to the short range of the stabilizing interactions, is assumed to be independent of the particular Watson-Crick NN pair. However, defects (single base MMs or single base bulges), due to sterical hindrance, are expected to strongly interfere

<sup>3</sup> The preexponential factor and equally the zipping rate are very difficult to determine experimentally. Reported values of the zipping rate vary between  $10^4$ - $10^5$ /s (fluorescence correlation spectroscopy of the fluctuations of a quenched fluorophore [AB03]) and  $10^7$ - $10^8$ /s (NMR measurement of the feasibility of imino-proton exchange [Gue87]).

with the zipping process.

Equilibrium calculations [Cra71; App65; Zim60] have shown that for formation of helices from short molecules the ratio  $s = k_+/k_-$  between the rates for formation of base pairs  $k_+$  and breakage of base pairs  $k_-$  is required to be in the range of  $9 > s > 4$ .

At each iteration step a pair of random variables  $(\mu, \tau)$  is drawn from the reaction probability density function

$$P(\mu, \tau) = k_\mu \exp\left(-\sum_{\nu=1}^n a_\nu \tau\right). \quad (7.2)$$

The poisson-distributed time step  $\tau$  is determined according to equation 7.3 using a uniform random number  $URN_a$  on the interval  $[0,1]$ .

$$\tau = \ln(1/URN_a)/(k_{r+} + k_{r-} + k_{l+} + k_{l-}) \quad (7.3)$$

The reaction pathway  $\mu$  is chosen with the random number  $URN_b$  according to equation (7.4).

$$m = URN_b \cdot (k_{r+} + k_{r-} + k_{l+} + k_{l-}) \quad (7.4)$$

$$a_1 = 0 + k_{r+}; \quad a_2 = a_1 + k_{r-}$$

$$a_3 = a_2 + k_{l+}; \quad a_4 = a_3 + k_{l-}$$

$$0 \leq m < a_1 \rightarrow \mu = 1$$

$$a_1 \leq m < a_2 \rightarrow \mu = 2$$

$$a_2 \leq m < a_3 \rightarrow \mu = 3$$

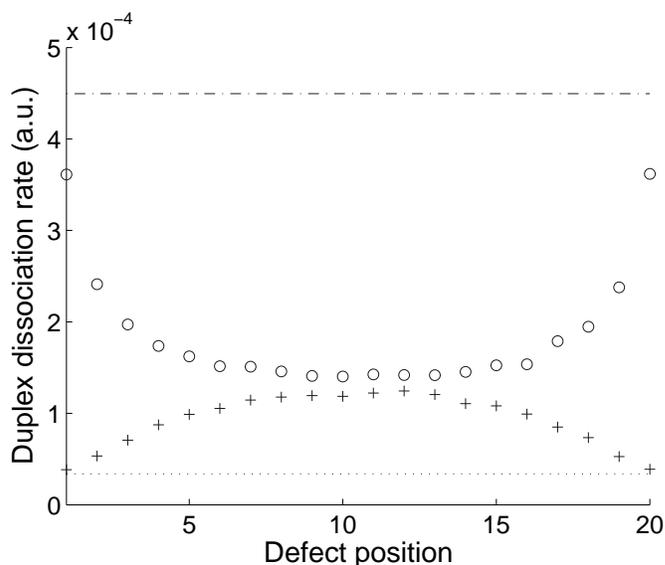
$$a_3 \leq m < a_4 \rightarrow \mu = 4$$

Even though the simulation considers only a single DNA duplex, the evolution of the zipper state over typically  $10^6$  to  $10^7$  iteration steps provides the ensemble average of partially denatured duplex states in equilibrium.

The duplex dissociation rate is determined as the number of complete duplex dissociations per simulation time. The duplex nucleation rate is assumed to be constant. This should be a good approximation, provided the duplexes to be compared have the same length.

## 7.2.2 Simulation Results

To investigate the influence of defect position (DPI) on the duplex dissociation rate, simulations were performed with the probe sequence motifs oligo-(C · G)<sub>19</sub>(A · T) and oligo-(A · T)<sub>19</sub>(C · G). Defects (i.e. inhomogeneities: weakly/strongly binding base pairs A · T and C · G within the strongly/weakly bound oligo-C · G/oligo-A · T sequence) were introduced at



**Figure 7.3:** Influence of defect position on the duplex dissociation rate. A stochastic simulation (based on the Gillespie Algorithm) was employed to determine the duplex dissociation rates for the homopolymer sequences  $(C \cdot G)_{19}(A \cdot T)$  (crosses) and  $(A \cdot T)_{19}(C \cdot G)$  (circles).  $A \cdot T$  and  $C \cdot G$  base pairs, respectively, were employed as single base defects within a homopolymer sequence. Here the defects were not considered as structural defects but rather as more/less stable Watson-Crick base pairs within a homopolymer duplex sequence. The position of the defects was systematically varied to investigate the positional influence on duplex stability. Dotted and dash-dotted lines depict the dissociation rates for defect-free duplexes  $(C \cdot G)_{20}$  and  $(A \cdot T)_{20}$ , respectively. Since the simulation temperatures for the two series were different ( $T=373$  K for  $(C \cdot G)_{19}(A \cdot T)$  and  $T=338$  K for  $(A \cdot T)_{19}(C \cdot G)$ ) the absolute dissociation rates are not comparable. To keep the simulation time reasonably short (overnight calculation) the simulation temperatures had to be chosen unrealistically high (for comparison - melting temperatures predicted by the DINAMelt-server for a target concentration of 1 nM and 1 M  $[Na^+]$ :  $T_m = 353.5$  K and 317 K, respectively).

systematically varied positions. Sequences have been chosen to demonstrate the influence of defect position. Even though these homopolymer-sequences are not really appropriate for the double-ended zipper-model (because alternative strand alignment is ignored), the examples demonstrate (Fig. 7.3) that defect position has a distinct influence on duplex dissociation rates. The significant DPI on the duplex dissociation rate (varying over an order of magnitude) is contrasting results from the two-state nearest neighbor model, which does not describe a positional influence of defects.

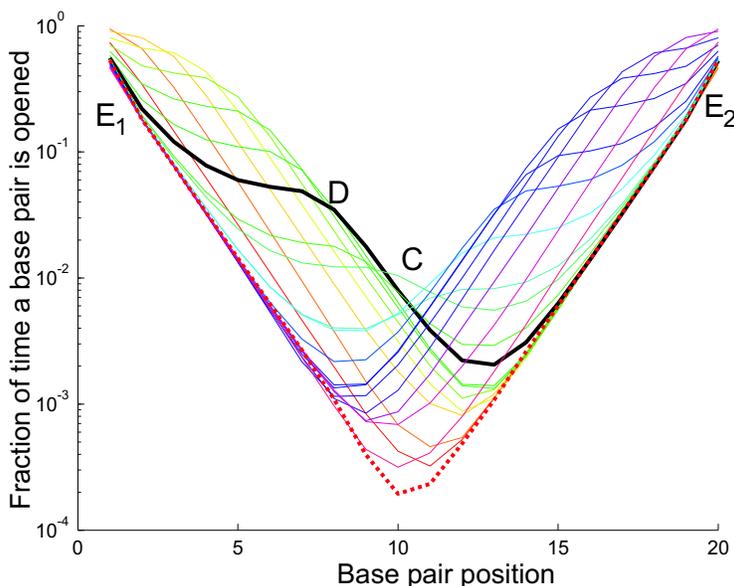
Since the duplex nucleation rate  $k_{nuc}$  is hardly affected by single base defects, the variation of the duplex dissociation rate  $k_{diss}$  should be reflected in the hybridization signal intensity  $I_{hyb}$ . Assuming that the surface density  $\sigma_D$  of the hybridized duplexes is far from satura-

tion, the hybridization signal is expected to be approximately inverse proportional to  $k_{diss}$ .

$$I_{hyb} \sim \sigma_D \sim k_{nuc}/k_{diss} \quad (7.5)$$

However, such an inverse proportionality between the experimentally observed hybridization signal and predicted  $k_{diss}$  couldn't be confirmed: Further implications (see section 7.4) give rise to a more complicated relation between the microarray hybridization signal and the duplex binding constant.

The semi-logarithmic plot of base pair dissociation probabilities (Fig. 7.4) demonstrates



**Figure 7.4:** Base pair opening probabilities for the duplex  $(C \cdot G)_{19}(A \cdot T)$  as a function of base pair position. The individual curves show the base pair opening probabilities for different positions of the defect (A·T base pair). The color spectrum from red to violet corresponds to defects at base pair positions 1 to 20. The black curve has been highlighted for further explanations in the text. The base pair opening probabilities exponentially decay towards the center of the duplex (as demonstrated by the dotted red curve which corresponds to the defect-free duplex  $(C \cdot G)_{20}$ ). Defects, however, result in deviations from the exponential decay and thus lead to increased opening probabilities. To investigate base pair dissociation in the center of the duplex (small opening probabilities) the simulation temperature had to be chosen very high (373 K), since at lower simulation temperatures infrequent dissociation of center base pairs doesn't allow statistical analysis. For the defect-free duplex with a minimum base pair opening probability of about  $2 \cdot 10^{-4}$  (in the center) and a presumed zipping rate of  $10^6 \text{ s}^{-1}$  we estimate a duplex half-life on the order of 1 s.

that for the defect-free PM duplex the base pair opening probabilities decrease exponentially towards the center of the duplex. Duplexes with point defects display a similar expo-

nential decrease. However, the defects shift the opening probability towards a higher level. In the middle of the duplex the base pair opening probability is increased by a factor 2 to 40 with respect to the PM duplex - depending on the position of the defect.

Defects, characterized by a reduced zipping rate  $k_+$  and/or an increased unzipping rate  $k_-$ , represent a barrier for the fast sequential zipping of the duplex. The ratio  $k_-/k_+$  at the defect site determines how long it takes to overcome the barrier.

As shown in Fig. 7.4 the increased opening probability at the defect site affects the binding of other base pairs: in the range  $\overline{DC}$  between the defect D and the center C of the duplex the opening probability decreases exponentially (with the same exponent as for the corresponding PM duplex) but at a higher level compared to the defect-free duplex. In the range  $\overline{DE}_1$  between the defect and the proximate duplex end the opening probabilities are also increased: Increased opening probability of the defect base pair implies increased opening probabilities of the base pairs between the defect and the proximate duplex end. However, the relative impact of the defect on opening probabilities is decreasing towards the duplex ends.

The position of the defect determines how much the opening probability in the range  $\overline{DC}$  is increased, thus how much the duplex dissociation rate is increased in comparison to the defect-free duplex.

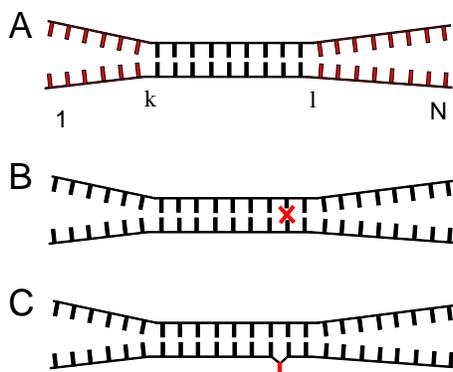
In the range  $\overline{CE}_2$  located in the other half of duplex (opposite the defect) base pair opening probabilities are largely unaffected and show an exponential decrease towards the middle of the duplex.

The stochastic simulation is computationally intensive. Simulations are therefore restricted to hybridization conditions where the duplexes have a short life time. Partially denatured states which occur very rarely (e.g. once per  $10^8$  zipping steps) are likely to be missed by the stochastic approach. To increase the probability for complete duplex denaturation the simulation temperatures had to be chosen unrealistically high (the chosen simulation temperatures are typically above the duplex melting temperatures). To circumvent this limitation a partition function approach (section 7.3) has been implemented.

With the partition function approach the calculation of the equilibrium distribution of duplex states requires only fractions of a second rather than overnight computation. Duplex stabilities at low temperatures ( $\rightarrow$ very small dissociation rates) can be investigated without restrictions. Because of this drastic improvement all further investigations are performed with the partition function approach.

However, an important advantage of the stochastic approach should be mentioned: The partition function approach is limited to the description of a duplex ensemble in thermodynamic equilibrium, whereas the stochastic simulation can be employed to investigate the time evolution of nonequilibrium processes.

## 7.3 Partition Function Approach of the Double-Ended Zipper Model

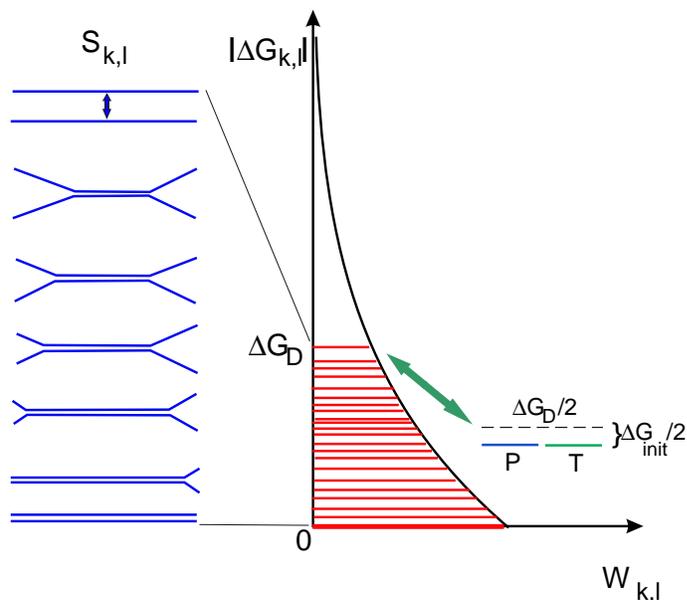


**Figure 7.5:** Double-ended zipper model of the oligonucleotide duplex. (A) Sequential unzipping is initiated at the duplex ends only ( $\Rightarrow$  end-domain opening). Duplexes can only stepwise, in a zipper-like fashion (nucleotide by nucleotide), partially denature or hybridize. The energy level of the partially denatured microstate  $S_{k,l}$  (with respect to the completely hybridized ground state) is determined by summation over the NN free energies of the unzipped NN-pairs (from 1 to  $k$  and from  $l$  to  $N$ ). (B) Single base MMs (non-Watson-Crick base pairing) affect the stabilities of two adjacent NN-pairs. (C) Base insertions and deletions result in bulged duplexes with an unpaired base. The surplus base (depicted in a looped out conformation), similar as a MM defect, results in significant duplex destabilization.

The statistical mechanics of the double-ended zipper model (Fig. 7.5) was first discussed by Gibbs and DiMarzio [Gib59]. Kittel [Kit69] demonstrated that with the double-ended zipper a phase transition (duplex melting) can be described if the degeneracy of states due to rotational freedom of the links between the nucleotides is considered.

In the given context the partition function approach describes the distribution of partially denatured duplex microstates in thermodynamic equilibrium. The microstates are populated according to the Boltzmann-distribution, which determines the statistical weights of individual microstates (Fig. 7.6). The canonical partition function  $Z$  is calculated as the sum of the statistical weights of all microstates of the double-ended zipper. Its value is a measure for the number of thermally populated microstates at a given temperature. The probability  $P_i$  that a system occupies microstate  $i$  (with the energy level  $E_i$ ) is  $\frac{1}{Z}e^{-E_i/kT}$ .

Based on the partition function various thermodynamic parameters (e.g. thermodynamic potentials, heat capacities etc.) of the system can be derived. In the following we investigate if the double-ended zipper model can reproduce our experimental results on mismatched duplex binding affinities.



**Figure 7.6:** In thermodynamic equilibrium the partially denatured duplex states  $S_{k,l}$  (indices  $k$  and  $l$  describe the zipper configuration - see Fig. 7.5) are populated according to a Boltzmann distribution. The lowest free energy level, that of the completely hybridized duplex  $S_{0,N}$  is used as reference and set to  $\Delta G_{0,N} = 0$  kcal/mol. The contribution of the duplex initiation parameter  $\Delta G_{37}^{\circ}{}_{init} = 1.96$  kcal/mol is relatively small (for comparison: the average NN pair free energy  $\Delta G_{37}^{\circ}{}_{NN}$  is about  $-1.4$  kcal/mol) and independent of the duplex sequence and is therefore neglected in all following considerations. The duplex dissociation free energy  $\Delta G_D$  is assumed to be equally distributed between the separated probe P and target T strands.

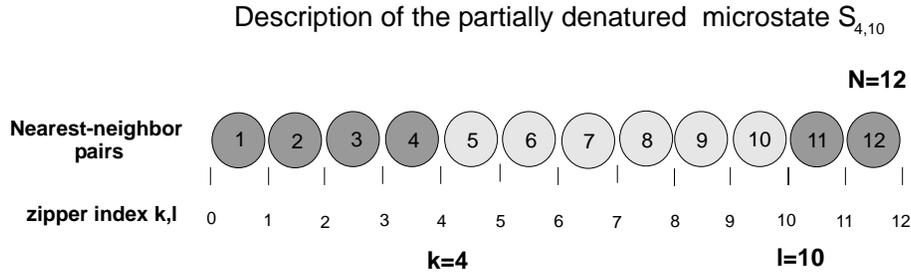
### 7.3.1 Implementation of the Partition Function Approach (PFA)

The canonical partition function  $Z$  of the duplex (equation 7.6) is calculated as the sum of the statistical weights  $w_{k,l}$  of all partially denatured microstates  $S_{k,l}$  of the duplex. Indices  $k$  and  $l$  refer to the positions of the zipper forks as depicted in Figs. 7.5 and 7.7.

$$Z_D = \sum_{k=0}^{N-1} \sum_{l=k+1}^N w_{k,l} = \sum_{k=0}^{N-1} \sum_{l=k+1}^N e^{\Delta G_{k,l}^{\circ}/RT} \quad (7.6)$$

The statistical weight  $w_{k,l}$  (equation 7.7) of the partially denatured state  $S_{k,l}$  is calculated from the sum of nearest neighbor (NN) free energies  $\Delta g_i^{\circ}$  of the unzipped duplex sections (equation 7.8).

$$w_{k,l} = e^{\Delta G_{k,l}^{\circ}/RT} \quad (7.7)$$



**Figure 7.7:** Theoretical description of the duplex microstate  $S_{k,l}$ .

The microstate  $S_{k,l}$  (here, as an example, the microstate  $S_{4,10}$  is shown) is defined by the position of the zipper forks at positions  $k$  and  $l$  (compare with Fig. 7.5). The number of nearest neighbor-pairs  $N=12$  is equal to the number of base pairs less one. Unzipped NN-pairs between the duplex ends and the zipper forks at  $k=4$  and  $l=10$  are featured dark.

$\Delta G_{k,l}^\circ$  is the free energy of the partially denatured state  $S_{k,l}$  relative to the completely hybridized ground state of the duplex.

$$\Delta G_{k,l}^\circ = \sum_{i=1}^k \Delta g_i^\circ + \sum_{i=l+1}^N \Delta g_i^\circ \quad (7.8)$$

$$\Delta G_{0,l}^\circ = \sum_{i=l+1}^N \Delta g_i^\circ \quad \Delta G_{k,N}^\circ = \sum_{i=1}^k \Delta g_i^\circ$$

Index values  $k=0$  and  $l=N$  indicate that the particular duplex state is completely closed at the left, or right end, respectively. The index  $i$  refers to NN-pairs (as shown in Fig. 7.7). NN free energies of Watson-Crick NN-pairs are deduced from unified NN parameters [San04].

$$\Delta g_i^\circ = \Delta h_i^\circ - T \cdot \Delta s_i^\circ \quad (7.9)$$

Assuming that the probe and target strands chosen for this study have no secondary structure and that each strand takes up half of the duplex dissociation energy  $\Delta G_D^\circ$ , we can estimate the partition functions of probes  $Z_P$  and targets  $Z_T$  as

$$Z_P = Z_T = e^{\Delta G_D^\circ / (2RT)} \quad \Delta G_D^\circ = \sum_{i=1}^N \Delta g_i^\circ \quad (7.10)$$

For simplicity duplex initiation free energies have been neglected here.<sup>4</sup>

<sup>4</sup> The duplex initiation free energy (originating from an unfavorable entropy change due to loss of translational degrees of freedom) is small with respect to  $\Delta G_D^\circ$  and independent of the duplex sequence.

Based on the duplex sequence we can now calculate the duplex binding constant

$$K = \frac{Z}{Z_P Z_T} = \frac{Z}{e^{\Delta G_D^\circ/RT}}. \quad (7.11)$$

### 7.3.2 Consideration of Point Defects

Point defects are accounted for with the corresponding defect NN free energy contribution  $\Delta g_{def}^\circ$  at the defect site. The following analytical derivation of the *defect positional influence* (DPI) for homopolymer sequences<sup>5</sup> shows that partition function values (provided as a function of defect position  $x$  - equation 7.16) are increased for defects located near the duplex ends.

#### Analytical derivation of the defect positional influence for a homopolymer sequence

According to equations 7.7 and 7.8 defects affect statistical weights of partially denatured states  $S_{k,l}$  only if the defect at position  $x$  is included in the unzipped sections of the duplex (Fig. 7.8A). Thus, the partition function  $Z(x)$  of a single defect duplex can be separated in two parts (see Fig. 7.8B) :  $Z_A(x)$  comprising microstates whose statistical weights are unaffected by the defect, and  $Z_B(x)$ , comprising those microstates whose statistical weights are affected by the defect (see also Fig. 7.8A). In  $Z_B(x)$  owing to the point defect a NN-pair with the free energy contribution  $\Delta g^\circ$  has been substituted by the defect free energy contribution  $\Delta g_{def}^\circ$ .

We can now factor out the defect type dependent impact of the defect:

$$Z(x) = Z_A(x) + Z_B(x) e^{(\Delta g_{def}^\circ - \Delta g^\circ)/RT} \quad (7.12)$$

Equivalently, using the partition function of the perfect match duplex

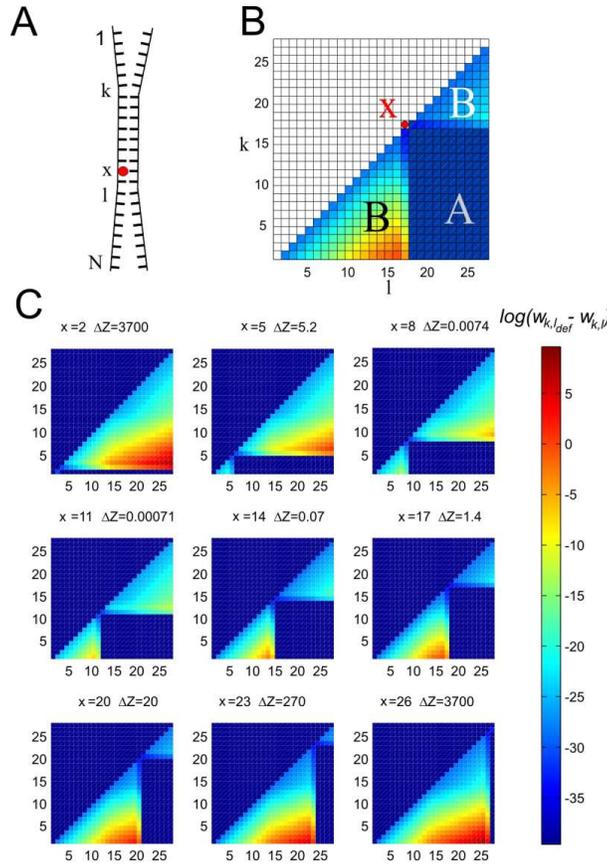
$Z_{PM} = Z_A(x) + Z_B(x)$  we can also write:

$$\begin{aligned} Z(x) &= Z_A(x) + Z_B(x) - Z_B(x) + Z_B(x) e^{(\Delta g_{def}^\circ - \Delta g^\circ)/RT} \\ Z(x) &= Z_{PM} + \Delta Z(x) \end{aligned} \quad (7.13)$$

Thus

$$Z(x) = Z_{PM} + Z_B(x) (e^{(\Delta g_{def}^\circ - \Delta g^\circ)/RT} - 1) \quad (7.14)$$

<sup>5</sup> For the purpose of a simple description NN free energy parameters are assumed to be identical.



**Figure 7.8:** Influence of the defect position on the statistical weights of the partially unzipped microstates of the oligonucleotide duplex  $(C \cdot G)_{12}(A \cdot T)_{12}$ . The difference between the individual weights of the mismatched duplex  $w_{k,l_{def}}$  and the corresponding microstates of the perfect matching (PM) duplex  $w_{k,l}$  (shown in (B) and (C)) strongly depends on the position  $x$  of the defect. In the following we assume a destabilizing defect. (A) Depending on defect position  $x$  and on the positions  $k$  and  $l$  of the two zipper forks, a defect either reduces the free energy (and thus the increases statistical weight  $w_{k,l_{def}}$ ) of the partially hybridized microstate  $S_{k,l}$ , or not: If a defect is located within the hybridized section (between  $k$  and  $l$ ) the statistical weight  $w_{k,l_{def}}$  of the unzipped state  $S_{k,l}$  remains unchanged with respect to the corresponding statistical weight  $w_{k,l}$  of the PM duplex. Whereas if the defect is located within the unzipped sections the statistical weight  $w_{k,l_{def}}$  of the microstate  $S_{k,l}$  is increased with respect to the weight  $w_{k,l}$  of the PM duplex. (B) Impact of a defect at position  $x$  on the statistical weights of the microstates. The individual matrix elements correspond to microstates  $S_{k,l}$  of the zipper. Their color shows the logarithmized deviation from the corresponding PM-duplex matrix element  $\log(w_{k,l_{def}} - w_{k,l_{PM}})$ . In the area A, corresponding to microstates where the defect is embedded in the hybridized duplex section, the difference is zero, as statistical weights are unchanged. In region B microstates are affected by the defect: Here, with respect to the perfect matching reference, free energies  $\Delta G_{k,l}^{\circ}$  are modified by the amount  $\Delta g_{def}^{\circ} - \Delta g_{PM}^{\circ}$ . Destabilization by the defect results in increased statistical weights of partially unzipped microstates. Part (C) shows the matrix depicted in (B) for different defect positions  $x$ . For defects close to the duplex ends the partition function  $Z$  is significantly increased with respect to the PM partition function.

Inserting the expression for  $Z_{PM}$  and summing up the statistical weights

$w_{kl} = e^{\Delta g^\circ(k+N-l)/RT}$  over all partially denatured states contributing to  $Z_B(x)$ , we obtain:

$$Z(x) = \sum_{k=0}^{N-1} \sum_{l=k+1}^N w_{kl} + \left( \sum_{k=0}^{x-1} \sum_{l=k+1}^x w_{kl} + \sum_{k=x}^{N-1} \sum_{l=k+1}^N w_{kl} \right) \times \left( e^{\frac{\Delta g_{def}^\circ - \Delta g^\circ}{RT}} - 1 \right) \quad (7.15)$$

This can be approximated by:

$$Z(x) \simeq Z_{PM} + \left( e^{\frac{(N-x)\Delta g^\circ}{RT}} + e^{\frac{x\Delta g^\circ}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}^\circ}{RT}} - 1 \right) \quad (7.16)$$

Equation 7.16 reflects that defects near the duplex ends increase end-domain opening. The number of thermally populated (partially denatured) duplex states, and thus the partition function  $Z$  is increased.

The defect destabilization  $\delta\Delta g_{def}^\circ = \Delta g_{def}^\circ - \Delta g^\circ$  describes the NN free energy difference between the defective duplex and the perfect matching duplex.  $\delta\Delta g_{def}^\circ$  is equivalent to the two-state nearest neighbor free energy increment between the PM and the MM duplex. For a single base mismatch defect  $\delta\Delta g_{def}^\circ$  is distributed over the two affected nearest neighbor pairs.

In equation 7.16  $\delta\Delta g_{def}^\circ$  has been factored out, revealing a generalized (i.e. defect-type independent) position dependence, which is governed by the distance between the defect and the duplex ends (see Fig. 7.9,  $Z$  is proportional to the binding constant  $K$ ). The stability of the duplex NN pairs  $\Delta g^\circ$  determines the slope  $dZ(x)/dx$  near the duplex ends. The defect destabilization  $\delta\Delta g_{def}^\circ$  determines how much  $Z$  is elevated with respect to the PM partition function  $Z_{PM}$  and thus how far the DPI propagates into the interior of the duplex .

The partition function the perfect matching duplex  $Z_{PM}$  is well approximated<sup>6</sup> by the statistical weight of the ground state with the value  $Z_{PM} \simeq 1$ .

To calculate the binding constant  $K$  of the duplex (according to equation 7.11) we still need the partition functions of the single stranded probe and target molecules: Probe and target strands each take half of the duplex dissociation energy. Since we assume that the single stranded species exist only in an unfolded state the partition functions for probes and targets are approximated by

$$Z_P = Z_T = e^{\Delta G_D^\circ/(2 \cdot RT)} = e^{N\Delta g^\circ/(2 \cdot RT)} e^{\delta\Delta g_{def}^\circ/(2 \cdot RT)}. \quad (7.17)$$

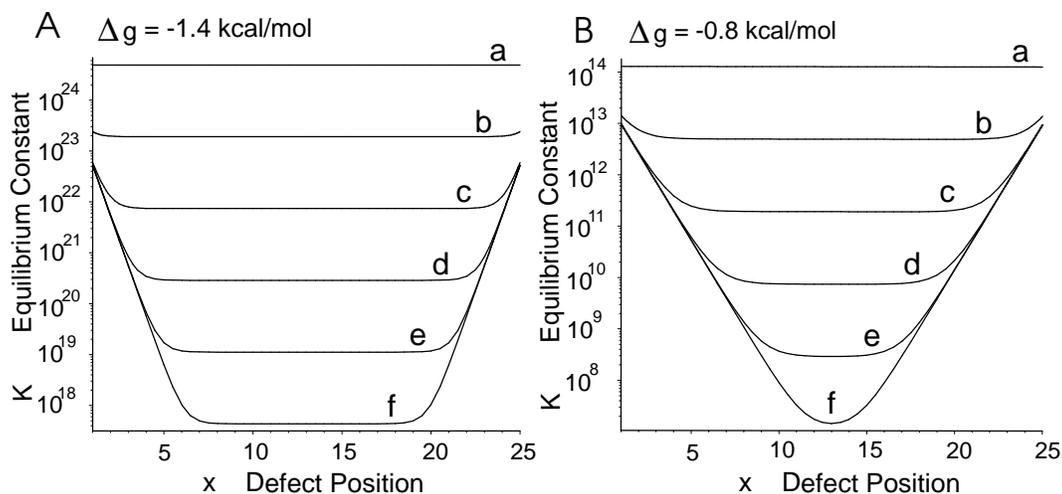
---

<sup>6</sup>  $Z_{PM}$  is about 1.25 for  $\Delta g^\circ = -1.4$  kcal/mol and  $T = 310$  K. With increasing temperature, as  $g_{def}$  approaches a value 0 kcal/mol, partially unzipped duplex states are increasingly populated:  $Z_{PM} \simeq 10$  for  $\Delta g = -0.1$  kcal/mol (at  $T \simeq 90^\circ\text{C}$ )

With  $Z_{PM} \approx 1$  and  $K = Z/(Z_P \cdot Z_T)$  we obtain the duplex binding constant

$$K = \frac{\left( e^{\frac{x\Delta g^\circ}{RT}} + e^{\frac{(N-x)\Delta g^\circ}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}^\circ}{RT}} - 1 \right) + 1}{\left( e^{\frac{N\Delta g^\circ}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}^\circ}{RT}} \right)}. \quad (7.18)$$

Fig. 7.9 illustrates the impact of the defect position  $x$  and defect destabilization  $\delta\Delta g_{def}^\circ$  on the duplex binding constant  $K$  (as described by equation 7.18) for two duplexes with different stabilities.



**Figure 7.9:** Positional influence of single base MM defects on the duplex binding affinity for two different NN pair free energies  $\Delta g^\circ$  at a temperature of 310 K. (A)  $\Delta g^\circ = -1.4$  kcal/mol corresponds to the average NN-pair binding free energy; (B)  $\Delta g^\circ = -0.8$  kcal/mol corresponds to a weakly bound sequence of A·T and T·A base pairs. Curves **a** to **f** correspond to defect destabilization parameters  $\delta\Delta g_{def}^\circ$  ranging from 0 to 5 kcal/mol (incrementally increased by 1 kcal/mol). In this example the defect destabilization  $\delta\Delta g_{def}^\circ$  is quoted per affected NN pair. Since MM defects affect two adjacent NN pairs the total free energy difference between MM and PM duplexes is  $2 \cdot \delta\Delta g_{def}^\circ$ .

## Application of equation 7.18

- for the perfect matching duplex ( $e^{\frac{\delta\Delta g_{def}}{RT}} = 1$ )

$$K = \frac{\left( e^{\frac{x\Delta g}{RT}} + e^{\frac{(N-x)\Delta g}{RT}} \right) (1 - 1) + 1}{\left( e^{\frac{N\Delta g}{RT}} \right) \cdot 1} = \frac{1}{e^{\frac{N\Delta g}{RT}}} \quad (7.19)$$

This is equivalent to the two-state equilibrium constant of the PM duplex.

- for a duplex with a terminal defect ( $x \simeq 0$  or  $x \simeq N$ )

$$K = \frac{(1 + 0) \left( e^{\frac{\delta\Delta g_{def}}{RT}} - 1 \right) + 1}{\left( e^{\frac{N\Delta g}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}}{RT}} \right)} = \frac{e^{\frac{\delta\Delta g_{def}}{RT}}}{\left( e^{\frac{N\Delta g}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}}{RT}} \right)} = \frac{1}{e^{\frac{N\Delta g}{RT}}} \quad (7.20)$$

This is equivalent to the equilibrium constant of the PM duplex.

- for a duplex with a center-defect ( $x \simeq N/2$ )

$$K = \frac{0 \cdot \left( e^{\frac{\delta\Delta g_{def}}{RT}} - 1 \right) + 1}{\left( e^{\frac{N\Delta g}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}}{RT}} \right)} = \frac{1}{\left( e^{\frac{N\Delta g}{RT}} \right) \left( e^{\frac{\delta\Delta g_{def}}{RT}} \right)}. \quad (7.21)$$

This is equivalent to the two-state equilibrium constant of the defective (e.g. mismatched) duplex.

Defects near the duplex ends cause only a small reduction of binding constant  $K$  (with respect to the PM stability), whereas defects in the middle of an oligonucleotide duplex cause a larger destabilization -  $K$  approaches the value of the two-state equilibrium constant.

### 7.3.3 Discussion

In our experiments we observed a distinct influence of the defect position on duplex binding affinities: The experiments show a largely monotonous decrease of hybridization signals over a range of typically 5-8 defect positions for 16mer probes, and up to 14 positions (e.g. in Fig. A.9) for some 25mer sequence motifs, from the duplex ends towards the center of the duplex. This distinct positional influence is confirmed by recent microarray studies [Wic06; Poz06] reporting a similar monotonous decrease of hybridization signals (on average) over 6-9 base positions.

In previous work [All97] NN-pair free energy increments  $\delta\Delta g_{def}^{\circ}$  (at  $T=37^{\circ}\text{C}$ ) for single

base MMs have been reported to be in the range of 1 to 3 kcal/mol per NN-pair. The total destabilization of single base MM duplexes (comprising two affected NN pairs) with respect to the corresponding PM duplexes is on average about 3.5 kcal/mol (at T=310 K).<sup>7</sup> Employing this  $\delta\Delta g_{def}^\circ$  value in equation 7.18, DPI propagation is restricted to the 2 outermost NN-pairs for the oligo-C·G-duplex (in Fig. 7.9A) and to the 5 outermost NN-pairs for the oligo-A·T-duplex (in Fig. 7.9B), respectively.

We find that in equation 7.18 - using "unified" nearest neighbor parameters from [San98] and MM nearest neighbor parameters from [All97] - the influence of the defect position is significantly underestimated. A similar (presumed) underestimation of defect positional influence (DPI) by the commercial DNA melting prediction software OMP (based on a multi-state equilibrium approach) has also been noticed by Wick *et al.* [Wic06].

In order to reproduce the distinct positional influence observed in the experiments we need to assume a significantly increased  $\delta\Delta g_{def}$  or an increased  $\Delta g^\circ$ . A significant increase of  $\Delta g$  can already be achieved by increasing the temperature from 310 K (average  $\Delta g^\circ \simeq -1.4$  kcal/mol) to 330 K (average  $\Delta g \simeq -0.98$  kcal/mol). Thus the temperature dependence of the NN free energy parameters may be the key for understanding the strong positional influence on DNA microarray surfaces. Though speculative, it is also possible that on the microarray surface, for example due to surface effects, the NN interactions are decreased with respect to the NN interactions in bulk solution.

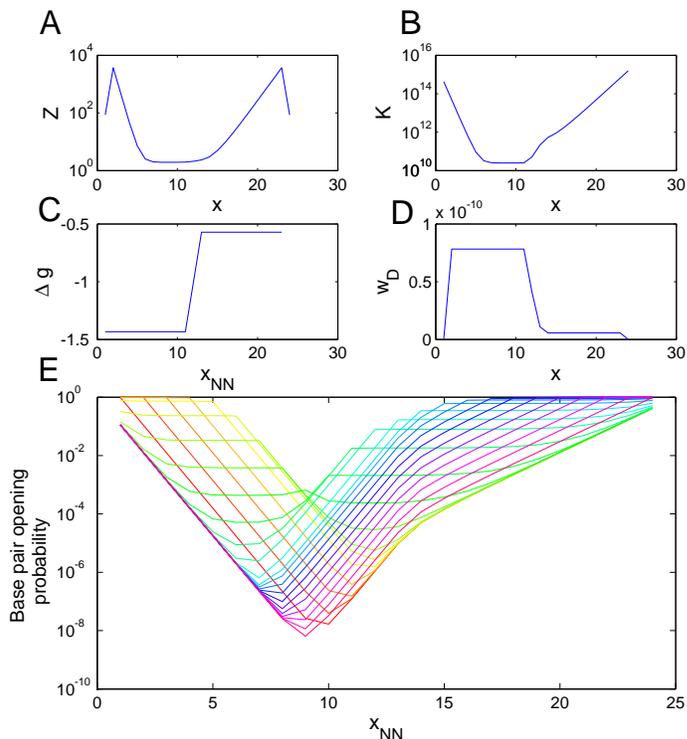
In a partition function based numerical analysis, to model the experimental results, we have chosen  $\Delta g_{def}$  as a free parameter. With a simulation temperature of T=330 K and with a defect NN parameter  $\Delta g_{def,330 K}=2.5$  kcal/mol (corresponding to a  $\delta\Delta g_{def}$  between 2.6 and 4.2 kcal/mol per NN pair) the numerical model shows a distinct positional influence similar to our experimental results. As discussed above, this  $\Delta g_{def}$  value is significantly larger than the mismatch NN-parameters described by Allawi *et al.* [All97].

The numerical analysis in Fig. 7.10 shows that defect positional influence on the duplex binding constant is largely determined by the partition function  $Z$ . In agreement with equation 7.18 the shallower slope on the right side of Fig. 7.10B corresponds to a sequence of weakly bound (A·T) base pairs, whereas the steeper slope on the left originates from a sequence of stable (C·G) base pairs.

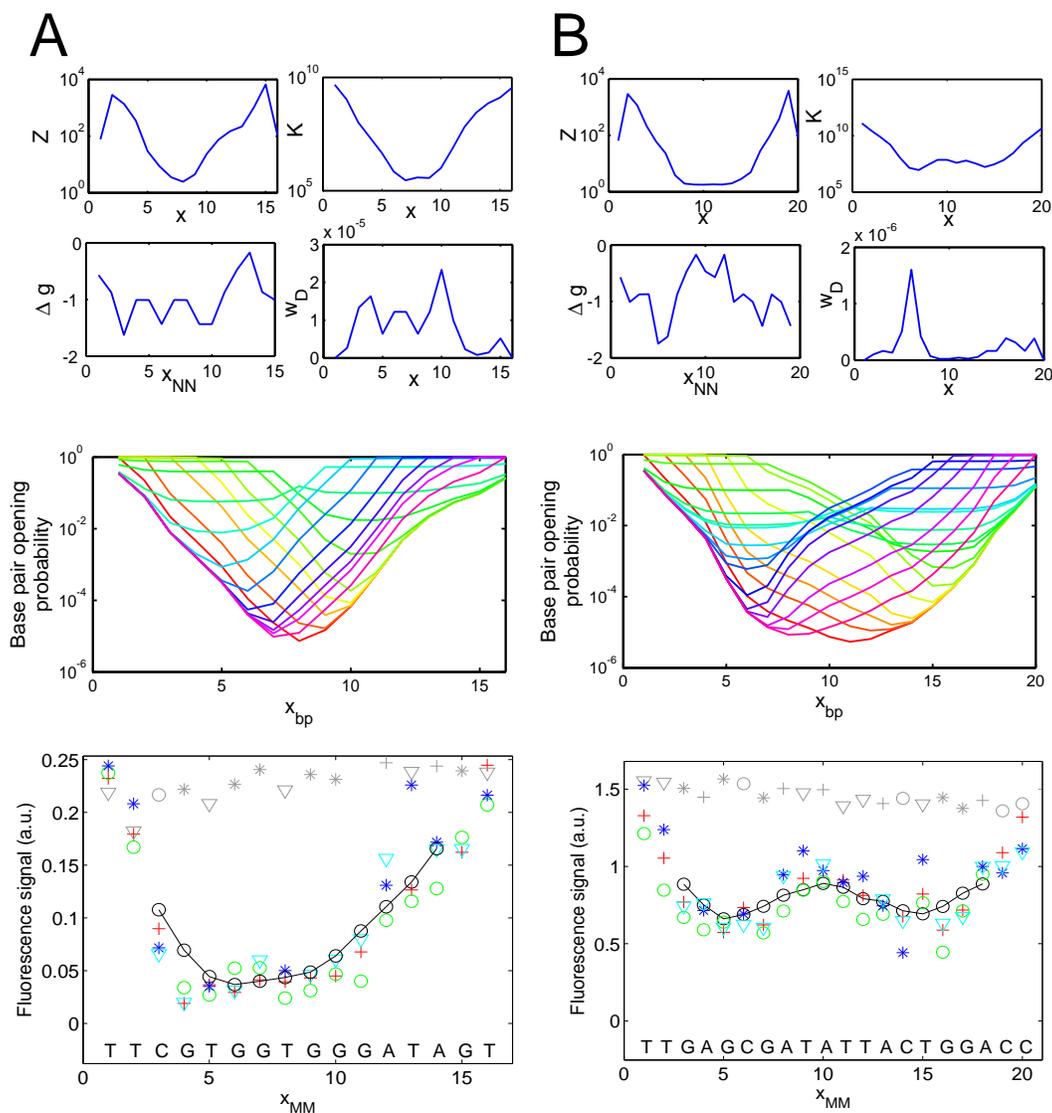
Fig. 7.11 compares results of the numerical analysis with the corresponding experimentally determined mismatch hybridization signals (sub-figures at the bottom). We find a good agreement between the hybridization signal intensity and the logarithm of the binding constant  $K$ .

---

<sup>7</sup> This total  $\delta\Delta g_{def}^\circ$  of about 3.5 kcal/mol, which was calculated on the basis of the two-state nearest neighbor model (using MM nearest neighbor parameters from [All97]), includes the destabilizing impact of the two affected NN pairs. Averaging was performed over various MM types.



**Figure 7.10:** Numerical analysis (partition function approach) to investigate the defect positional influence on oligonucleotide duplex binding affinity. For this example we chose the asymmetric sequence motif  $(C \cdot G)_{12}(A \cdot T)_{12}$ . Simulation temperature  $T=330$  K;  $\Delta g_{def}^{37}=2$  kcal/mol per each of the two affected NN pairs. (A) Partition function of the duplex  $Z(x)$  as a function of defect position  $x$ . The slope  $dZ(x)/dx$  is steeper towards the left end of the duplex (consisting of strongly bound CC/GG nearest neighbor pairs) than towards the right side comprising of the weaker AA/TT pairs. The origin of the position dependence of  $Z(x)$  is depicted in Fig. 7.8. (B) The logarithmic plot of the duplex binding constant  $K(x)$  (equation 7.11) reflects the strong impact of defect position on the hybridization signal observed experimentally. A comparison with experimental data is shown in Fig. 7.11. Duplex stability is least for defects located in the center of the duplex. The position dependence of the binding constant originates largely from the partition function  $Z(x)$ . (C) Nearest neighbor free energies (in kcal/mol) of the individual NN-pairs comprising the duplex. (D) The statistical weight of the completely dissociated state  $w_D$  is reciprocal to the equilibrium constant in the two-state NN model. Variations in curve (D) reflect variations of  $\delta\Delta g_{def}$  and are thus dependent on the defect type. In the profile of binding constants  $K(x)$  (B) defect type dependent influences originating from the statistical weight of the dissociated state (from which the partition functions of the single stranded species are derived - equation 7.17) are significantly smaller than the positional influence introduced by the duplex partition function  $Z(x)$ . The plot of the base pair dissociation probabilities (E), as anticipated, shows an exponential decrease towards the center of the duplex. The exponent, which is determined by the sequence of nearest neighbor parameters, is significantly different in the two sections of the asymmetric sequence. Defect position, encoded by the color spectrum - ranging from red (defect at the left end) to violet (defect at the right end) - the defect is located at the sharp kink - strongly affects partial denaturation of the duplex.



**Figure 7.11:** Comparison of simulation results (top and center) with the experimentally determined hybridization affinities (bottom) for two probe sequence motifs in (A) and (B). The four small subfigures show the partition function  $Z$  and the duplex binding constant  $K$  as a function of the defect position  $x$ , the NN-free energies  $\Delta g^\circ$  of particular NN-pairs as a function of NN-pair position  $x_{NN}$ , and the statistical weight for complete duplex dissociation  $w_D$  as a function of defect position. The middle subfigure shows the base pair opening probabilities (the fraction of strands in which the corresponding base pair is unzipped) as a function of base pair position  $x_{bp}$ . The various curves correspond to different defect positions (red - defect at left end; pink - defect at right duplex end). The bottom subfigure shows the experimentally determined MM defect profile. Legend: A - red crosses, C - green circles, G - blue stars, T - cyan triangles. Moving average of all mismatch types - black line. Grey symbols correspond to PM probes. Irregularities in  $Z(x)$  at the duplex ends are due the fact that only a single NN-pair is affected by a MM-base pair at the duplex end. Simulation parameters:  $T=330$  K,  $\Delta g_{def}^\circ=2$  kcal/mol. It is needs to be emphasized, that  $\log(K(x))$ , rather than  $K(x)$ , resembles the position dependence of the experimentally observed hybridization signal.

However, the logarithmic relationship between  $K$  and the hybridization signal intensity requires an explanation. In the following we investigated how the hybridization signal is linked to the duplex stability.

## 7.4 Relation Between the Hybridization Signal and Duplex Stability

### Microarray hybridization experiment

To investigate how the fluorescence intensity of hybridized targets is related to duplex stability on the microarray surface we performed a hybridization assay: A set of probe sequences with gradually increasing length (e.g. from 12 to 27 nt - see inset in Figure 7.12) - thus with an approximately linearly increasing binding free energy  $\Delta G_D$  - provides the relation between the duplex binding free energy and the microarray hybridization signal.

Experimental results in Fig. 7.12 show a sigmoid relation between the hybridization signal intensity and the probe length. The transition region – in which the hybridization signal intensity is (in a first-order approximation) growing linear with the duplex length – has a width of a least 13 base pairs (corresponding to a binding free energy range  $\delta\Delta G_{D37}^{\circ}$  of approx. 20 kcal/mol). The large deviation from the *Langmuir isotherm*<sup>8</sup> (left curve in Fig. 7.12) is in accordance with previous observations [Car06; Bin06] reporting a strong destabilization of surface tethered duplexes.

### Discussion

The equilibrium between single stranded probes  $P$  and targets  $T$  and hybridized duplexes  $D$  for the hybridization reaction



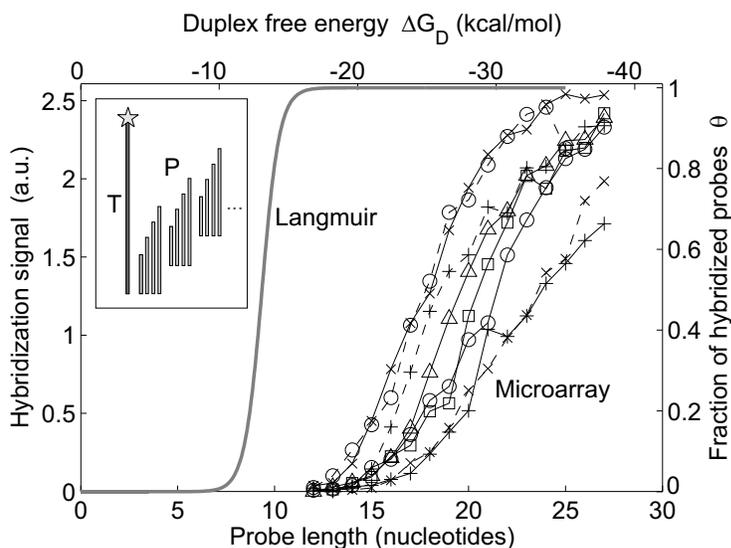
is described by a Langmuir-type adsorption isotherm (equation 7.24). Since (under the experimental conditions employed) targets are in excess, the target concentration  $[T]=[T_0]$  is assumed to be constant. Using the law of mass action 7.23 and considering that the concentration (surface density) of unhybridized probes equals the initial concentration  $P_0$  less the concentration of hybridized probes, i.e.  $[P] = [P_0] - [D]$

$$K = \frac{[D]}{[T][P]} = \frac{[D]}{[T_0][P_0 - D]} \quad (7.23)$$

we derive the Langmuir equation

$$\theta = \frac{[D]}{[P_0]} = \frac{K \cdot [T]}{1 + K \cdot [T]}. \quad (7.24)$$

<sup>8</sup> The *Langmuir isotherm* typically relates the surface coverage  $\theta$  to the concentration (or pressure) of the adsorbed molecule species in the liquid (or gas) above the surface. However, here we employ the Langmuir equation (7.24) to investigate the surface coverage  $\theta$  as a function of the binding free energy.



**Figure 7.12:** Microarray hybridization signal intensity as a function of the oligonucleotide duplex length. The inset (left) shows the scheme of the hybridization experiment: Within subsets of probes  $P$  the probe length is gradually increased from 12 to 27 nt. From each subset we obtain the hybridization signal intensity  $I_{hyb}$  as a function of the probe length  $l_p$ . In the following we will refer to  $I_{hyb}(l_p)$  as "transfer function". All probes were hybridized with the common target sequence  $URA$  (1 nM in  $5\times SSPE$ , for 20 minutes at  $45^\circ C$ ). Additionally, to obtain slightly different subsets of probe sequences, we shifted the probe sequences gradually (in steps of two nucleotides) along the target sequence  $T$ : different binding affinities of the individual subsets (targeting different subsequences of the target  $T$ ) result in significantly different transfer functions. Assuming that the duplex stability increases approximately linear with the duplex length, a length increment of one nucleotide on the lower scale (duplex length) corresponds to a duplex binding free energy increment of  $-1.4$  kcal/mol (average NN free energy parameter  $\Delta g_{37}^\circ$ ) on the upper scale. The experimental results show that the transfer function has a sigmoid shape: The hybridization signal intensity approaches a value of 0 for probes with a length of  $\leq 12$  nt. Within the transition region the hybridization signal (in a first order approximation) increases linear with the probe length. The width of transition region is at least 13 base pairs (corresponding to a  $\delta\Delta G_{D37}^\circ$  range of approx. 20 kcal/mol) - for some transfer functions the transition region extends over an even wider range. For comparison with the experimentally determined transfer function  $I_{hyb}(l_p)$ , a theoretical transfer function  $\theta(\Delta G_D)$  (equation 7.25) - derived from the Langmuir equation is shown (assumed NN free energy:  $-1.4$  kcal/mol,  $T=310$  K, target concentration: 1 nM). In comparison to  $I_{hyb}(l_p)$ ,  $\theta(\Delta G_D)$  has a narrow transition region ( $\delta\Delta G_D^\circ \approx 3$  kcal/mol). The midpoint of the transition is located at a probe length of 9 nucleotides (in the experiment the midpoint is located at  $l_p \geq 17$  nucleotides).

Since the hybridization signal intensity  $I_{hyb}$  is proportional to the fraction of hybridized probes  $\theta = [D]/[P_0]$ , we refer to  $\theta$  as the hybridization signal. In the following we assume that the target concentration is in large excess (under the experimental conditions employed), thus the free target concentration can be considered constant, i.e.  $[T] = [T_0]$ . Inserting the two-state equilibrium constant we obtain a sigmoidal relation (see Fig. 7.12) between the hybridization signal and the duplex binding free energy  $\Delta G_D^\circ$ .

$$\theta = \frac{e^{-\Delta G_D^\circ/RT} \cdot [T_0]}{1 + e^{-\Delta G_D^\circ/RT} \cdot [T_0]} \quad (7.25)$$

This sigmoidal relation between the hybridization signal and duplex free energy has been reported previously [Yil04] for solution-phase hybridization (in the context of fluorescence in situ hybridization (FISH)).

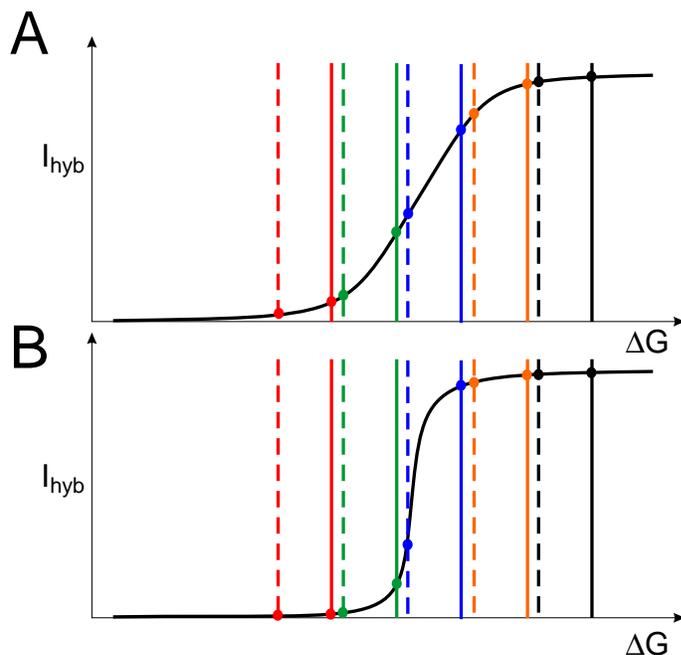
The relatively narrow transition region described by the *Langmuir isotherm* (equation 7.25) (width of the transition region:  $\delta\Delta G \approx 3$  kcal/mol) can not reproduce the experimentally observed DPI of the hybridization signal, since the duplex binding free energy range of the individual defect profiles is expected to be of about the same size or larger than the transition region: Defect profiles (hybridization signal versus defect position and defect type - see Fig. 6.1) cover a broad range of hybridization signal intensities: for 20mer duplexes a hybridization signal range between PM (100%) and center MM (20%) equals the hybridization signal difference between 20mer PM duplexes and 15mer PM duplexes (Fig. 7.12). This range (approx. 7 kcal/mol) is significantly broader than the linear transition range of the *Langmuir isotherm* (approx. 3 kcal/mol) in Fig. 7.12.

The proportionality between free energy increments and hybridization signal increments - within the linear transition range - confirms the previous observation  $\delta I_{hyb} \sim \delta(\log(K))$  (see Fig. 7.11). To gradually differentiate between different binding affinities - like in the defect profiles (e.g. Fig. 6.4) or in Fig. 7.12 a relatively broad transition region is required. The broadened transition (its presumed origin - from heterogeneity of binding affinities - is discussed in the following section), with respect to the idealized adsorption characteristics described by the Langmuir equation, is an important characteristics of DNA microarray hybridization.

A linear relation between free energy increments and hybridization signal increments is only possible within the approx. linear transition range. The relatively narrow transition region described by the *Langmuir isotherm* (equation 7.25) (width of the transition region:  $\delta\Delta G \approx 6$  kcal/mol - approximately linear over ca. 3 kcal/mol) can not reproduce the experimentally observed DPI of the hybridization signal, since the duplex binding free energy range of the individual defect profiles is expected to be of about the same size or larger than

the transition region.<sup>9</sup>

A broadened transfer function  $I_{hyb}(\Delta G)$  is expected to be beneficial, for example, in



**Figure 7.13:** Beneficial effect of a broadened transfer function: The transfer function  $I_{hyb}(\Delta G)$  in (A) - resembling the transfer function observed in the experiment in Fig. 7.12 - is significantly broader than the transfer function in (B) which is derived from the Langmuir equation. The vertical lines correspond to binding free energies within the probe set of an SNP microarray (solid lines correspond to the PM duplexes, dashed lines correspond to the MM duplexes). In (B) good discrimination between the PM and the corresponding MM duplex is only achieved for duplexes with binding free energies within the narrow transition range (blue and green lines), whereas in (A) a good discrimination is achieved over a significantly larger binding free energy range.

genotyping assays: In Fig. 7.13B a good discrimination between PM (solid vertical lines) and MMs (dashed vertical lines) is achieved only over a narrow range of binding free energies. In an SNP microarray assay it would be necessary that the binding affinities are located in (or near to) the narrow transition range. A narrow transition as described by the Langmuir equation would impose a severe constraint on the choice of microarray probe sequences. The broad transition function in Fig. 7.13A enables MM/PM discrimination within a broader transition range - thus the requirement for microarray probe sets compris-

<sup>9</sup> Defect profiles (hybridization signal versus defect position and defect type - see Fig. 6.1) cover a broad range of hybridization signal intensities: for 20mer duplexes a hybridization signal range between PM (100%) and center MM (20%) equals the hybridization signal difference between 20mer PM duplexes and 15mer PM duplexes (Fig. 7.12). This range ( $5 \cdot 1.4 = 7$  kcal/mol) is significantly broader than the linear transition range of the *Langmuir isotherm* (approx. 3 kcal/mol) in Fig. 7.12.

ing almost identical binding free energies is relaxed.

### 7.4.1 Heterogeneity of Binding Affinities

Various reasons have been proposed to explain the deviation of the microarray hybridization binding isotherm from the *Langmuir isotherm*: these include competitive hybridization effects [Hal04], electrostatic repulsion [Vai02], entropic blockage [Hal05] and variability of the probe sequences [Bin06].

Electrostatic repulsion, entropic blockage and variability of probe sequences (owing to sequence defects, e.g. point mutations and strand truncations [Job02] introduced in the *in situ* synthesis process of the DNA microarray) result in heterogeneity of binding affinities. Heterogeneity of binding affinities has also been observed for immobilized antibody probes: Vijayendran *et al.* [Vij01] investigated the heterogeneity of surface-immobilized antibody-receptor binding affinities. They observed that a more uniform alignment of the surface-bound antibodies improves heterogeneity of binding affinities with respect to randomly immobilized antibodies. The chemical microenvironment (distance to adjacent probes etc.) may also play a role.

A single uniform binding constant results in a Langmuir-type hybridization isotherm (equation 7.24), whereas the presumed heterogeneous distribution of binding affinities on the microarray surface results in a broadened effective isotherm.

The *Sips isotherm* (arising from a gaussian distribution of binding affinities)

$$\theta = \frac{[D]}{[P_0]} = \frac{(K \cdot [T_0])^\alpha}{1 + (K \cdot [T_0])^\alpha}. \quad (7.26)$$

has been reported to provide a better description of surface hybridization than the *Langmuir isotherm* [Pet02; Gla06; Bin06]. In particular, it describes the broadened transition region observed experimentally (Fig. 7.12). The Sips exponent  $\alpha \leq 1$  is a measure for the heterogeneity of binding affinities. For a value of  $\alpha = 1$ , which corresponds to a uniform binding affinity, equation 7.26 is identical with the *Langmuir isotherm* (equation 7.25).

### 7.4.2 Impact of Random Defects Introduced in the *in situ* Synthesis Process

The photolithographic *in situ* synthesis process used for the fabrication of DNA chips generates a variety of defects (mainly single base mismatches, base bulges and strand truncations) which affect the binding affinity of the individual probes. Owing to these synthesis-related defects microarray features comprise heterogeneous distributions of binding affini-

ties rather than a uniform binding affinity. Assuming a stepwise synthesis error rate of 10%, most of the 25mer probes (more than 90%) contain at least one synthesis defect [Nai06b]. Since the number of defects per probe molecule follows a binomial distribution the majority of the probe strands comprises between one and three defects.

## Numerical Simulation

For a theoretical investigation of the impact of synthesis errors on the hybridization signal, we have created distributions of probe sequences which are equivalent to the heterogeneous probe composition of a microarray feature. To simplify matters only MM defects were assumed. The number of defects per duplex is binomial distributed. Binding constants  $K_i$  of the individual (randomly mutated) sequences were calculated with the partition function approach (PFA) described in section 7.3.2. The contribution of each individual probe to the total hybridization signal  $\theta_{total}$  is determined by the Langmuir isotherm according to equations 7.27 and 7.28. Probes hybridize with different binding constants  $K_i$  to the same target T.

$$\theta_i = \frac{K_i \cdot [T_0]}{1 + K_i \cdot [T_0]} \quad (7.27)$$

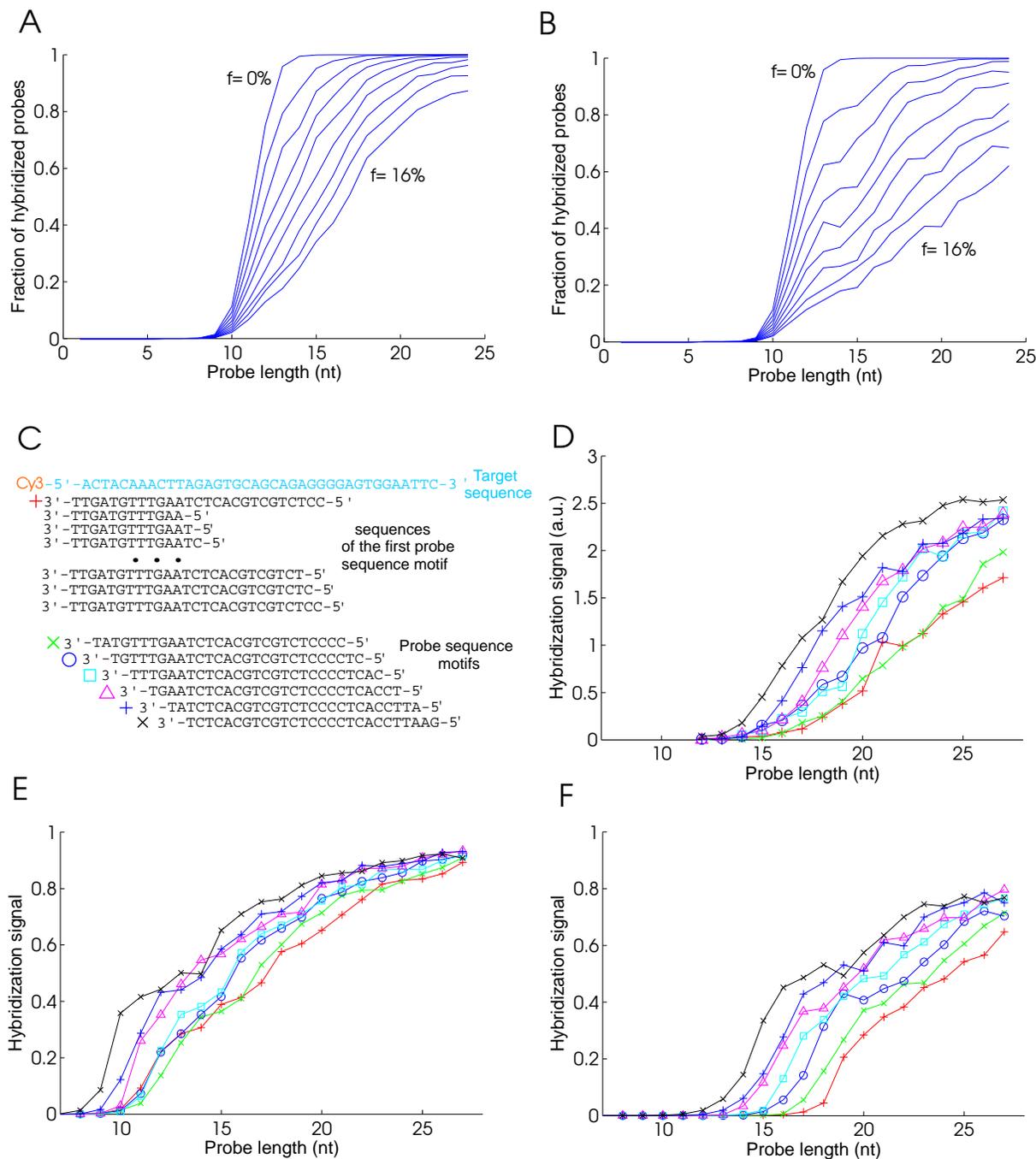
We obtain the total hybridization signal by summing up the over the distribution of probes, thereby accounting for the molar fraction  $x_i$  of the individual probe species.<sup>10</sup>

$$\theta_{total} = \sum_i^N x_i \theta_i = \sum_i^N x_i \frac{K_i \cdot [T_0]}{1 + K_i \cdot [T_0]} \quad (7.28)$$

### Fig. 7.14 (see next page) Heterogeneity of binding affinities

Numerical simulation on the influence of synthesis errors on the hybridization signal: Using the partition function approach we have calculated hybridization signals for distributions of probe sequences containing various synthesis defects - similar to those expected in a single microarray feature. In (A) and (B) the probe length  $l_P$  (assumed to be roughly proportional to  $\Delta G_D^\circ$ ) is varied between 6 and 25 base pairs. The relation  $\theta(l_P)$  (which basically describes the transfer function  $\theta(\Delta G_D^\circ)$ ) was determined according to equation 7.28. Synthesis error rates  $f$  (fraction of errors per synthesis step) have been varied between 0% and 16% in steps of 2%. The simulation code assumes single base MM defects with a  $\Delta g_{def}^\circ = -0.5$  kcal/mol in (A), and  $\Delta g_{def}^\circ = +2$  kcal/mol in (B). The relatively weak defects in (A) like the strong defects in (B) result in a significant broadening of the transfer function in respect to the narrow transition range for  $f=0\%$  (which is corresponding to the Langmuir isotherm). Parts (C) to (F) compare experimental results (shown in D - identical with Fig. 7.12) to the simulation results. (C) For a series of probe sequence motifs (shifted in steps of 2 base positions

<sup>10</sup> A similar approach has been proposed by Vijayendran *et al.* [Vij01].



**Figure 7.14:** Numerical simulation on the influence of synthesis errors on the hybridization signal. A detailed description is provided in the text.

along the complementary target sequence) we have generated several sets of probe sequences (sequence motifs shown) with incrementally increasing length. Experimentally obtained hybridization signals of the corresponding series are shown in (D). Part (E) shows the corresponding simulation results, taking into account synthesis errors (parameters:  $T=310$  K, target concentration=1 nM, error rate  $f=10\%$ ,  $\Delta G_D^0=2$  kcal/mol). Significantly improved agreement with the experimental results is achieved in (F) using an increased temperature  $T=333$  K and a reduced error rate  $f=6\%$ . Like in the experiment the minimum probe length required for hybridization is about 12 to 16 nt.

The heterogeneous distribution of binding efficiencies originating from synthesis defects results in a "stretched isotherm" similar to a Sips isotherm (see Fig. 7.14). This is due to the fact that probes containing synthesis defects do significantly contribute to the hybridization signal, though with a reduced binding efficiency. Therefore (with respect to defect-free probes) the midpoint of the transition region is shifted towards higher duplex binding free energies (i.e. longer probes).

The resulting effective isotherm, with its stretched (in a first order approximation) linear transition region can explain the experimentally observed hybridization characteristics on microarray surfaces as shown in Fig. 7.12 and for example in Fig. A.9.

It should be mentioned that the heterogeneity of binding affinities of surface-immobilized probes is not restricted to DNA microarrays fabricated by *in situ* synthesis: Heterogeneity of binding affinities has also been described for surface-immobilized proteins (antibody assays) [Vij01]. Surface immobilization leads to heterogeneity of the micro-environment of individual probes (e.g. distance to the next probe molecule, orientation of the molecule), and thus results in a distribution of binding affinities.

### **Influence of Electrostatic Blocking and Competitive Effects**

Electrostatic repulsion between free targets in solution and surface bound probes/targets strands has been proposed as an a major reason for the deviation from Langmuir-type behavior [Hal04; Bin06]. The increase of the surface charge during the hybridization process, due to binding of negatively charged target strands increases the negative surface potential. This goes along with a decrease of the apparent binding constant and results in a Sips-like adsorption isotherm [Bin06].

In our study the increase of surface charge owing to the hybridization of negatively charged target strands is expected to be small (estimated 5-20 %) compared to the initial charge arising from probe strands alone. This is for two reasons: Due to synthesis defects a large fraction of the probe strands doesn't significantly contribute to hybridization (thus, the fraction of hybridized probes is typically far from 100%), and, owing to the short length of the targets employed, the amount of charge per hybridized target molecule is rather small.

In the experiments performed in this study electrostatic blocking introduces an approximately constant free energy penalty, and hence does not provide an explanation for the observed Sips-like isotherm.

However, this may be different if relatively long targets strands (50-500 base pairs long) carrying a correspondingly larger charge, are involved, as for example in gene expression assays.

Competitive bulk hybridization (competition between surface and bulk hybridization) has also been proposed to give rise to Sips-like isotherms [Hal04]. In our study, since we focused on experiments with a single target sequence, we can exclude such an influence from competitive effects.

## 7.5 Approximation of the Partition Function Approach (PFA) with a Position Dependent Nearest Neighbor (PDNN) Model

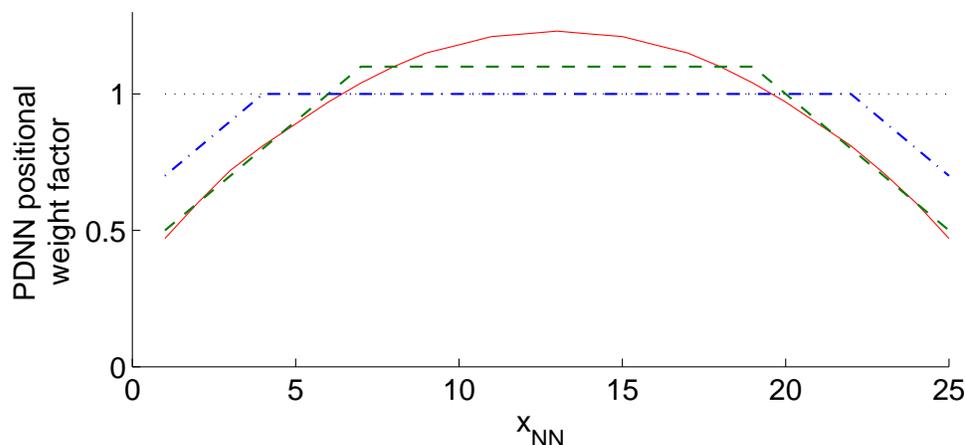
So far we considered the impact of structural defects (like single base MMs) on duplex stability. However, in the framework of our model we can regard any NN pair as a "defect" and investigate its position-dependent contribution to duplex stability. In the empirical PDNN model [Zha03] (see section 2.3.2) like in the zipper model nearest neighbor pairs close to the duplex ends contribute less to duplex stability than those in the interior of the duplex.

In the following we show that duplex free energy values determined with the PFA can be approximated by a position dependent nearest neighbor (PDNN) model [Zha03; Car06; Hel06] in which the duplex binding free energy  $\Delta G_D$  is calculated as a position-dependent weighted sum of nearest neighbor free energies (equation 7.29).

$$\Delta G_D = \sum_{i=1}^N w_i \Delta G_{NN_i} \quad (7.29)$$

To investigate which position dependence  $w(i)$  provides the best approximation for the PFA we performed a theoretical analysis with a set of one thousand 25mer random sequences (chosen for a similar nucleotide content): Duplex binding free energies were calculated with the partition function approach ( $\Delta G_{PFA}$ ), with the PDNN model ( $\Delta G_{PDNN}$ ), and with the two-state nearest neighbor model ( $\Delta G_{TSNN}$ ). The TSNN model can be regarded as special case ( $w_i = \text{constant}$ ) of the PDNN model.

We determined Pearson's correlation coefficients  $r$  between the corresponding distributions

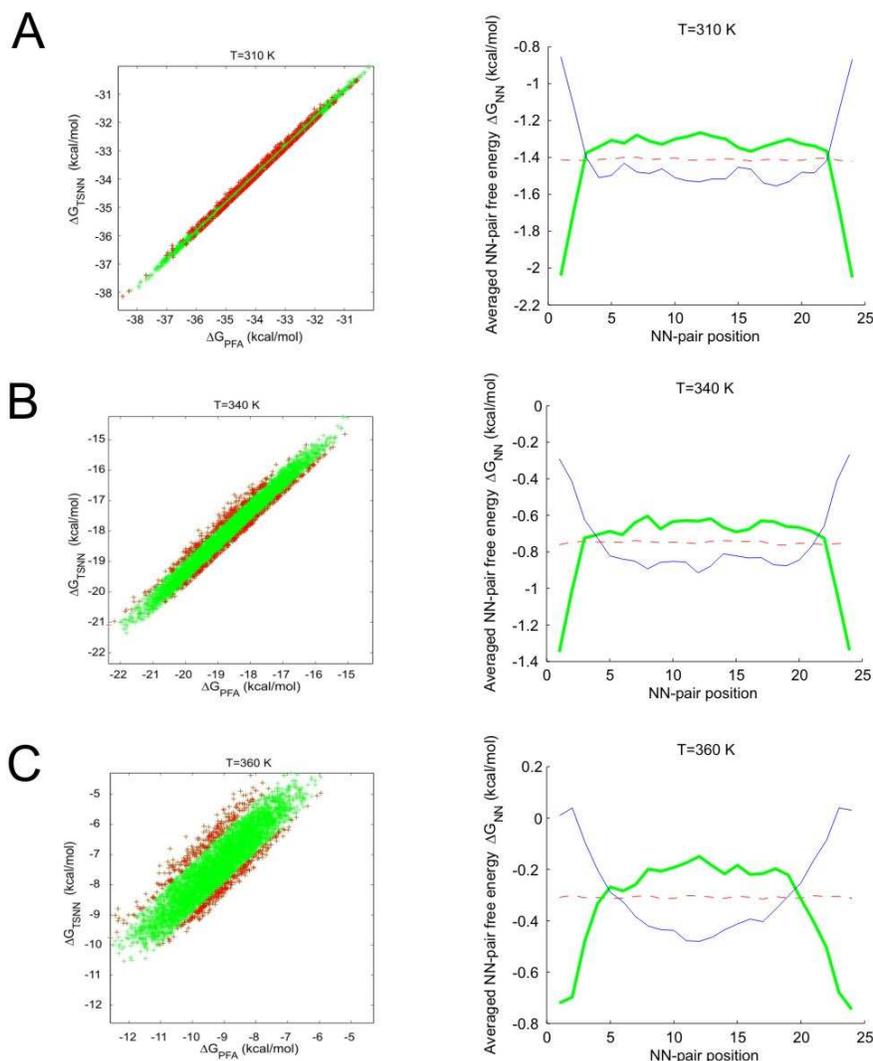


**Figure 7.15:** Approximation of the partition function approach (PFA) by a position dependent nearest neighbor model (PDNN). For a set of one thousand 25mer random probe sequences (chosen for a similar nucleobase content) we compared perfect-match duplex free energies  $\Delta G$  for the PFA, PDNN and two-state nearest neighbor (TSNN) model. Using Pearson's correlation coefficient  $r_P$  for the correlation between  $\Delta G_{\text{PDNN}}$  and  $\Delta G_{\text{PFA}}$  and  $r_T$  for the correlation between the TSNN free energy  $\Delta G_{\text{TSNN}}$  and  $\Delta G_{\text{PFA}}$ , we found that the best fit PDNN weight function  $w(x_{NN})$  strongly depends on temperature.

At  $T=360$  K a parabolic weight function (*solid red line*) reproduces the PFA results significantly better ( $r_P = 0.979$ ) than the two-state nearest neighbor model ( $r_T = 0.892$ ). Approximating the parabolic function by a composed function (*dashed green line*) of decreasing ramps towards the edges and constant weights in the center, we obtain the same Pearson coefficient. However, since the temperature of 360 K is significantly above the melting temperature of the 25mer duplexes (which is approx. 340 K), this result should not be mistaken as an analytical proof of the PDNN model. At  $T=340$  K the best fit ( $r_P = 0.996$  versus  $r_T = 0.981$ ) is achieved with a similar composed function with reduced weight parameters (ramps) only at the three outermost base positions (*blue dashed-dot line*). Towards lower temperatures the PFA converges towards the TSNN model. At 340 K the weight parameters for 360 K provide a relatively poor fit ( $r_P = 0.946$ ). At 310 K the PFA results match that of the TSNN ( $w(x_{NN}) = 1$ ) almost perfectly ( $r_T = 0.999$ ).

of  $\Delta G_{\text{PFA}}$ ,  $\Delta G_{\text{PDNN}}$  and  $\Delta G_{\text{TSNN}}$ . At 360 K a parabolic weight function (Fig. 7.15) provides a significantly better correlation with the PFA ( $r_P = 0.979$ ) than with the two-state nearest neighbor model ( $r_T = 0.892$ ). The parabolic weight function can be well approximated by a ramp function. At 340 K the best correlation is achieved with a ramp function in which only the three outermost NN pairs have a significantly reduced contribution to duplex binding free energy.

The temperature dependence of the correlation between the PFA and the two-state nearest neighbor model is demonstrated in Fig. 7.16. At  $T=310$  K, owing to reduced end-fraying



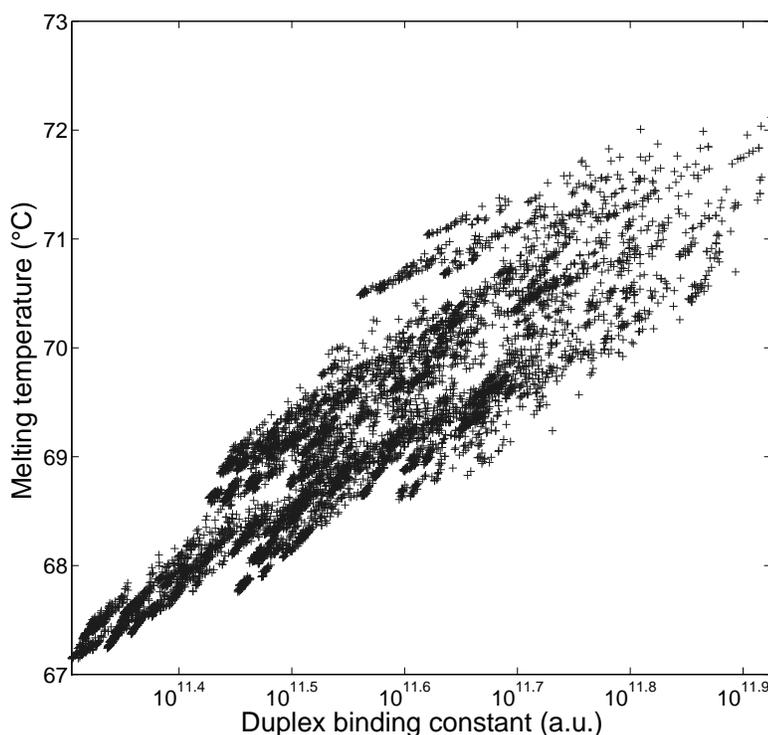
**Figure 7.16:** Theoretical investigation of deviations between the two-state nearest neighbor (TSNN) model and the partition function approach (PFA). To investigate for which sequences the difference between TSNN free energies and PFA free energies is largest, we have created a large set of 5000 random 25mer sequences with a similar nucleobase composition. Plots of TSNN free energies versus PFA free energies (*left*) show a very good correlation at a temperature of 310 K. At higher temperatures (340 K and 360 K) we find significant deviations between the two models. We have selected the 5% of sequences with largest residuals (red markers) and determined the position-dependent distribution of NN free energies (shown *right*) by averaging ( $\rightarrow$ averaged NN-pair free energy versus NN-pair position). The Gibbs free energies in A, B and C refer to  $T=310$  K, 340 K and 360 K, respectively). At 310 K the sequences with the most stable  $\Delta G_{\text{PFA}}$  have their weak NN-pairs at the outermost two base positions (thin blue line) and therefore the more strongly binding NN-pairs in the interior. Vice versa sequences with the weakest  $\Delta G_{\text{PFA}}$  (bold green line) have strong NN-pairs located at the outermost positions. The mean NN free energy (average over all sequences) is indicated by the dashed red line. At 340 K for the most stable sequences (according to PFA) the weakest NN-pairs are concentrated at the 6 outermost base positions (at each duplex end). At 360 K (which is above the melting temperature of the duplexes) the NN-pair stabilities exhibit a parabolic position dependence.

at lower temperatures, the PFA-results converge with that of the two-state nearest neighbor model (Fig. 7.16A).

### Distribution of NN-pair stabilities

To investigate the influence of the positional distribution of stronger/weaker NN-pairs we have created a set of 7500 duplexes, each assembled from the same set of 24 NN-pairs (randomly arranged). Owing to the identical NN-pair content the TSNN binding free energy of the randomly arranged duplex sequences is identical.

On the basis of the partition function approach we determined the binding affinities of the individual duplexes. Good agreement with the UNAFold melting temperatures is shown in Fig. 7.17.



**Figure 7.17:** The comparison between duplex melting temperatures calculated with UNAFold and the corresponding equilibrium constants calculated with our partition function approach (at  $T=340$  K), shows a good agreement between the two models. Equilibrium constants are plotted on a logarithmic scale. Owing to identical NN pairs (randomly arranged) in the chosen set of 7500 25mer sequences the two-state binding constants are identical. The stripe patterns (presumably) originate from sequence similarities (sequences are not completely random - owing to the constraint of identical NN-pairs when the sequences were generated).

The analysis of the positional distribution of NN-pairs in the weakest/strongest 5% of the duplexes is shown in Fig. 7.18:

- in the group of the most stable duplexes the strong NN-pairs are located in the center of the duplexes, whereas
- in the least stable duplexes the strong NN-pairs are located close to the duplex ends.

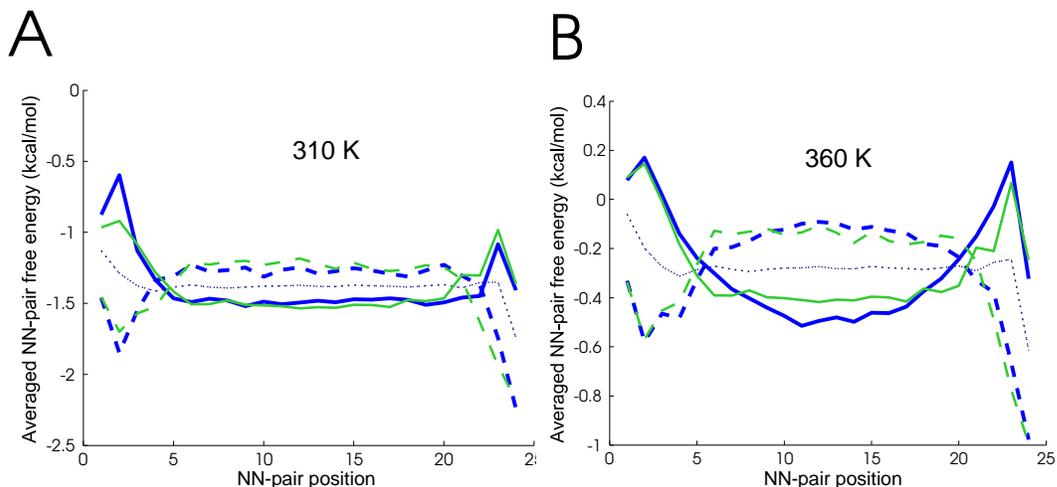
At increased temperatures (360 K in Fig. 7.18B) the positional distribution of NN stabilities is more pronounced (extending to the center of the duplex) than at 310 K where only the outermost base pairs. A similar result has been obtained with the partition function based UNAFold software<sup>11</sup> (DINAMelt web server [Mar05]). Asymmetries at the duplex ends originate from the constraint of identical NN-pairs ( $\Rightarrow$  sequences are not completely random but rather have identical NN-pairs at the duplex ends). A similar result (without artifacts) is shown in Fig. 7.16 (right column). Here the duplexes are composed of sequences with a similar base composition, rather than of identical NN-pairs.

## Discussion

Our theoretical analysis demonstrates that end-fraying is the reason for the reduced stabilizing contribution of base pairs which are located near the duplex ends. This has been previously suggested by Zhang *et al.* [Zha03] for an explanation of the position dependent nearest neighbor model. We have shown that the PDNN model can be derived from the double-ended zipper model.

However, we found that only the outermost base pairs are subject to significant end-fraying. Our statistical analysis (Fig. 7.16B - left) demonstrates that (near the melting temperature) in 25mer duplexes the 4-6 outermost nearest neighbor pairs (at each side, i.e. 8-12 of 24 NN pairs) have a reduced position dependent contribution to duplex stability. This is not exactly the distinct parabola-like position dependence reported by [Zha03], however, even with a positional influence restricted to the duplex ends the distribution of the various NN pairs within the duplex (NN pair free energy versus NN pair position) has a significant influence on duplex stability: The analysis of sequences composed of identical NN pairs ( $\rightarrow$  identical duplex free energies according to the two-state nearest neighbor model) demonstrated that sequences with strong NN-pairs in the center are significantly more stable than sequences with their strong NN-pairs near the fraying ends.

<sup>11</sup>To determine 7500 melting temperatures UNAFold was run via a MatLab script with the `perl` command: `result=perl(['hybrid2.pl'], '-tmin=40', '-A0=0.000000001', '-B0=0.000000001', '-NA=DNA', '-exclude=A', '-exclude=B', '-exclude=AA', '-exclude=BB', 'probeFileName', 'tarFileName')`. Melting temperatures were read from text files generated by UNAFold.



**Figure 7.18:** We performed a theoretical investigation on duplex stability with a set of 7500 randomly created 25mer sequences with an identical NN pair content. For those sequences, owing to identical NN pairs the two-state nearest-neighbor model predicts identical duplex binding free energies. Similar as in Fig. 7.16 we determined those sequences with the largest/smallest  $\Delta G_{\text{PFA}}$  and determined for these subsets the spatial distribution of NN pairs (NN pair stability versus NN pair position). NN free energies are referring to 310 K (A) and 360 K (B), respectively. For the most stable 5% of the duplexes the average NN free energies are plotted as a function of NN pair position (*solid blue curve*), the NN pair distribution of the least stable 5% is shown by *dashed blue curve*. The *dotted line* shows the average nearest neighbor free energy distribution over all 7500 sequences. In parallel we employed the partition function based software UNAFold [Mar05] to determine the melting temperatures of the duplexes. We selected the 5% of the sequences with the highest/lowest melting temperatures and established the corresponding NN pair distribution (highest melting temperatures: *solid green curve*; lowest melting temperatures: *dashed green curve*). We found a good agreement between the results of the PFA and UNAFold. At a temperature of 360 K the PFA shows a significantly stronger position dependence. The asymmetric bias at the duplex ends originates from the fact that the constraint of an identical NN pair content provides sequences with identical NN pairs at the duplex ends.

# Chapter 8

## Microarray Experiments

### 8.1 Influence of Synthesis Defects on Microarray Hybridization Characteristics

#### 8.1.1 Theoretical Considerations

The fidelity of light-directed *in situ* synthesis is affected by point defects introduced in the synthesized probe sequences [Gar02; Job02; Kim03; Nai06b]:

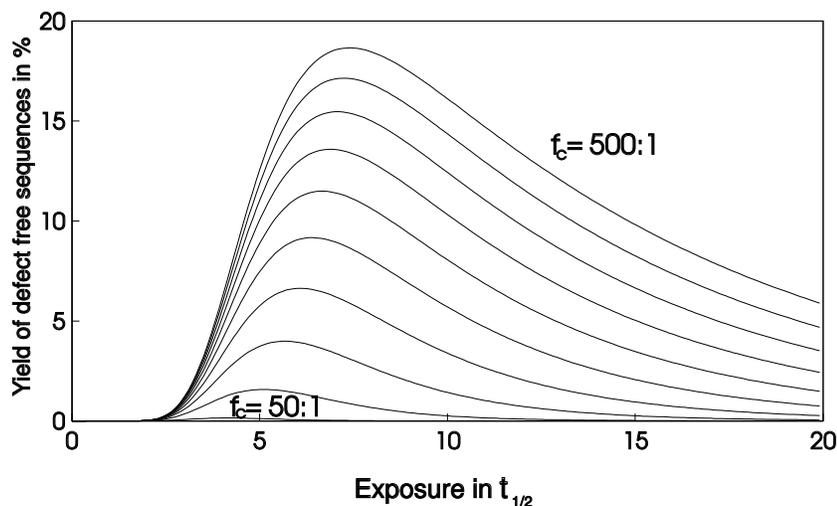
- Limited phosphoramidite coupling efficiency → single base deletions
- Incomplete photo-deprotection (see Fig. 8.3) → single base deletions
- Stray light results in erroneous deprotection → single base insertions and single base mismatches
- Strand breakage → truncated strands

The length of the probe molecules synthesized on the array is typically 15 to 30 nucleotides. Due to random errors, caused by stray light, incomplete photo-deprotection or incomplete coupling of the monomers, the yield  $Y$  of correctly synthesized  $n$ -mer oligonucleotides is limited to

$$\begin{aligned} Y &= E_s^{3 \cdot n} \cdot E_d^n \cdot E_c^n \\ &= (e^{-t/(\tau \cdot f_c)})^{3 \cdot n} \cdot (1 - e^{-t/\tau})^n \cdot E_c^n \end{aligned} \quad (8.1)$$

$E_d$  and  $E_c$  denote the stepwise efficiencies for deprotection and coupling reactions. Coupling efficiencies of NPPOC amidites were reported to be 96-99% (depending on the particular phosphoramidite reagent) [Nuw02].  $E_s$  accounts for erroneous deprotection by

stray light.<sup>1</sup> In the synthesis process a microarray feature is deprotected (on average) only in one of four deprotection steps. Thus, stray light from neighboring features affects the synthesis of a probe sequence three times per coupling reaction. Compared with the time constant for photo-deprotection  $\tau$  the time constant for stray light deprotection, owing to the smaller stray light intensity, is increased by the contrast factor  $f_c$ .  $E_s$  depends not only on the optical performance of the photolithography setup but also on the geometry of the mask patterns (see Sec. 3.2.7). With an averaged estimated stray light intensity of 0.5% of the full exposure intensity (i.e. the local contrast  $f_c$  is 200:1) and the time course of photo-deprotection as shown in Fig. 8.3 the fraction of erroneously deprotected molecules per exposure step is not just 0.5%, but rather 2 to 3%. Therefore  $E_s$  has an average value of about 97 to 98%. With these efficiencies ( $E_s = 0.98$ ,  $E_c = 0.98$  and  $E_d = 0.97$  - estimated from Fig. 8.3) the yield  $Y$  of correctly synthesized sequences on a 25mer microarray is about 6% (compare with Fig. 8.1). For a larger stray light intensity of about 2% of the exposure intensity  $\rightarrow E_s = 0.92$  the yield would be significantly reduced to 0.05%. Therefore a high local contrast ratio over the small distance separating neighboring features is crucial for successful light-directed fabrication of DNA microarrays.



**Figure 8.1:** Yield of defect free 25mer probes sequences according to equation 8.1. The coupling efficiency was set  $E_c=0.97$ . The contrast factor  $f_c$  (ratio between full exposure intensity and stray light intensity) is varied between 50:1 and 500:1 (in increments of 50). Depending on the contrast the optimum yield has its maximum between 5 and 7  $\tau_{1/2}$  (the time at which 50% the protection groups are removed).

A fraction of 6% of error free probe sequences may seem like very little. However, as will be shown later (see section 7.4.2), the fraction of error free probes is not all-important for

<sup>1</sup>  $E_s$  is the fraction of probes which is not affected by stray light. Stray light is caused by optical flare, aberrations, and diffraction at the mirror edges. Most stray light originates from the exposure of neighboring microarray features (local contrast).

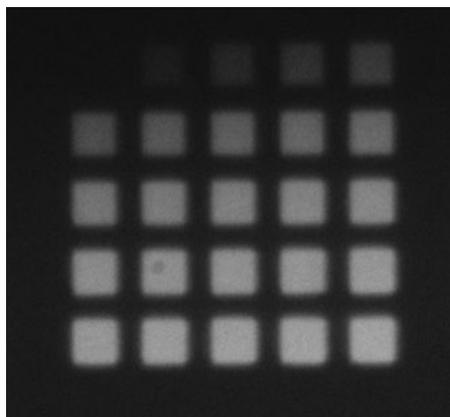
the function of the microarray, since also probes with point mutations participate in the hybridization process. We have demonstrated that despite synthesis defects hybridization on our microarrays is highly specific.

Fig. 8.1 demonstrates that, owing to stray light (even at a relatively high contrast ratio of 500:1), an increase of the exposure beyond 5 to 7 half-exposure times  $\tau_{1/2}$  is not beneficial for the synthesis yield.

A detailed experimental investigation on synthesis-related defects in light-directed *in situ* synthesis and measures to improve the quality of synthesized microarray probes has been published by Richmond *et al.* [Ric04].

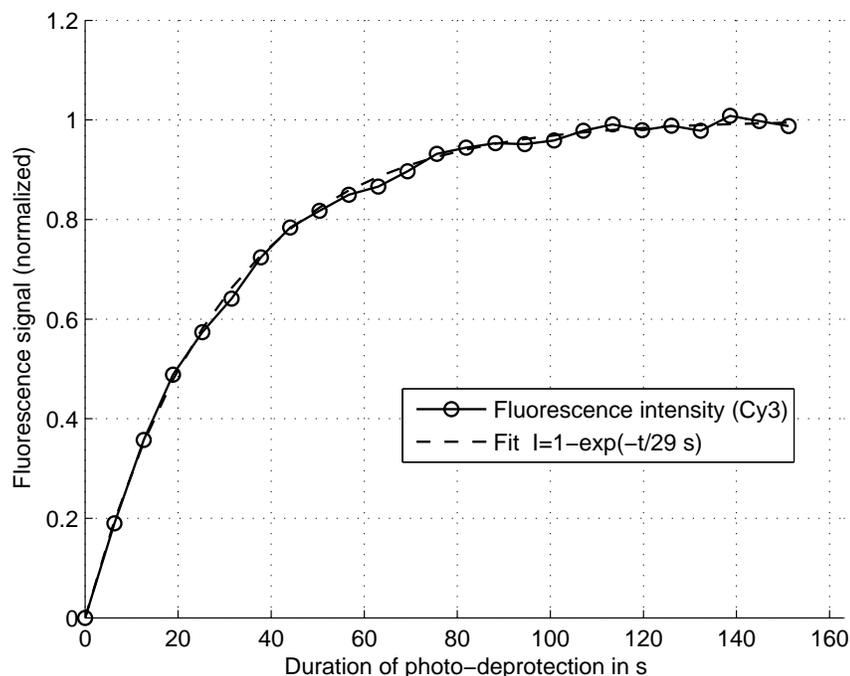
### 8.1.2 Evaluation of the Synthesis Yield - Progress of the Photo-Deprotection

The progress of the photo-deprotection of the NPPOC groups was investigated in an exposure-variation experiment, similar as that described by Luebke *et al.* [Lue02]. Like in the normal synthesis process the initial layer of NPPOC-T-amidite (covering the surface completely) has been deprotected by UV exposure. The deprotection time of the particular features has been varied in increments of 6.3 s. Subsequently Cy3-amidite (Amersham



**Figure 8.2:** Variation of amount of coupled Cy3-phosphoramidite as a function of the UV exposure dose. The fluorescence micrograph shows the fluorescence intensity of Cy3-amidite which has been coupled onto a single layer of deprotected T-amidites on the microarray surface. The length of the UV exposure (photo-deprotection) has been varied in increments of 6.3 s (between 0 and 151.2 s).

Biosciences) was coupled to the deprotected binding sites. Cy3-fluorescence (Fig. 8.2) enables quantification of the deprotected binding sites as shown in Fig. 8.3. The progress of the photo-deprotection is well-fitted by a saturation curve with a half-life of  $\tau_{1/2} = 20.1$  s. With an exposure time of about 100 s (as employed in the microarray synthesis process)



**Figure 8.3:** Degree of photo-deprotection as a function of the deprotection time. The fluorescence intensity of the coupled Cy3-amidite shows the progress of photo-deprotection.

the degree of photo-deprotection is 97%. This value is similar to the estimated coupling efficiencies of NPPOC-amidites of 96-99% [Nuw02] (depending on the individual phosphoramidite reagent). An improvement of the deprotection efficiency, to reduce the rate of synthesis errors [Ric04] could be achieved by increasing

- the exposure time
- the UV light intensity
- the sensitivity of the photo-deprotection reaction [Wol04].

However, one has to be aware that a higher deprotection efficiency also results in an increased stray light dose. To increase the deprotection efficiency from 97 to 98% the exposure needs to be prolonged significantly as complete deprotection is approached asymptotically, whereas the fraction of molecules erroneously deprotected by stray light increases at much faster rate [Gar02]. Thus, as shown in Fig. 8.1, there is an optimum exposure time at which the number of synthesis errors has a minimum.

To ensure a high quality of the synthesized probes, rather than to reduce the deprotection time, a high contrast between the features under exposure and neighboring features, to be protected from stray light, is necessary (see section 3.2.7).

### 8.1.3 Impact of Synthesis Defects on Microarray Synthesis Fidelity

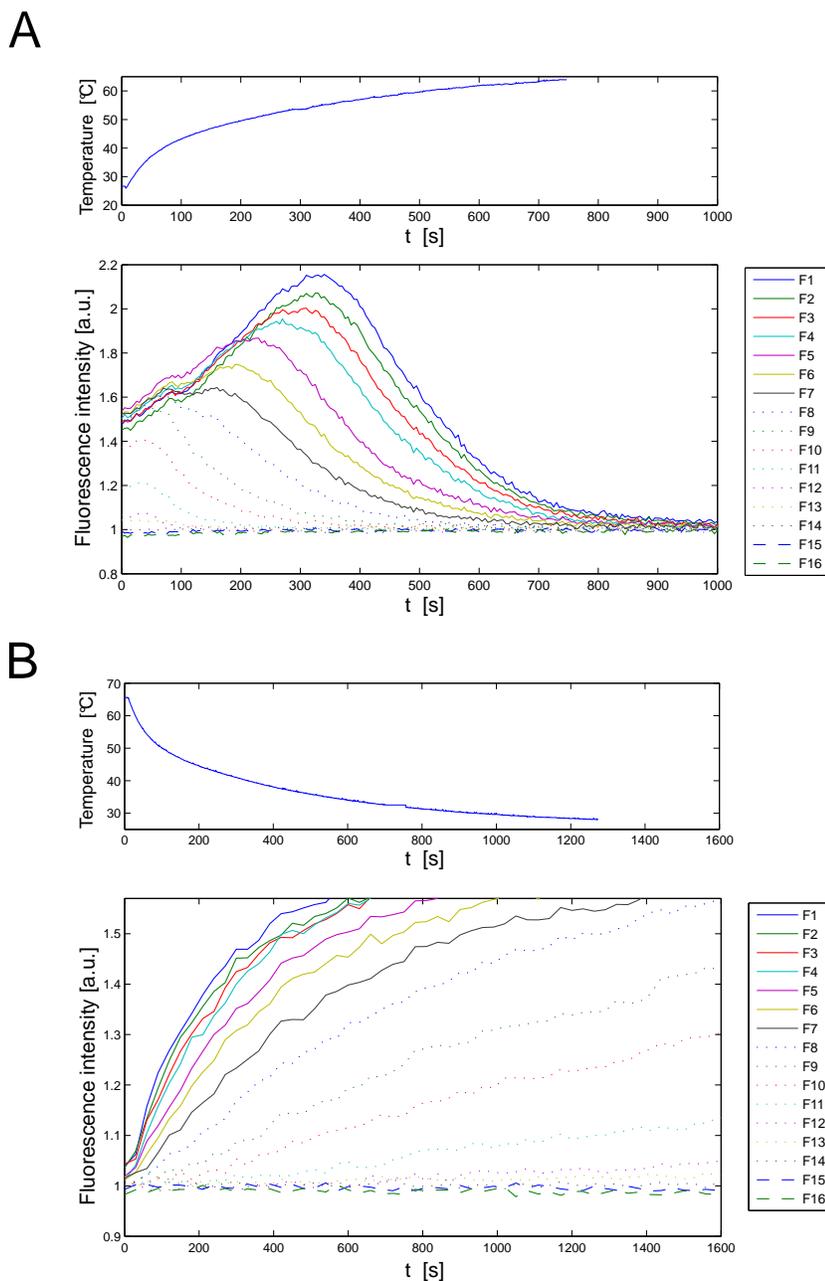
The impact of synthesis defects on the hybridization characteristics of the microarray probes is difficult to estimate, because

- we don't know how many defects are generated in the synthesis process. Moreover, the density of defects varies locally with the imaging quality, and
- the varying number of defects per probe, as well of the varying position of the defects, give rise to a heterogeneity of binding affinities. It is not known how multiple defects affect oligonucleotide duplex binding affinities.

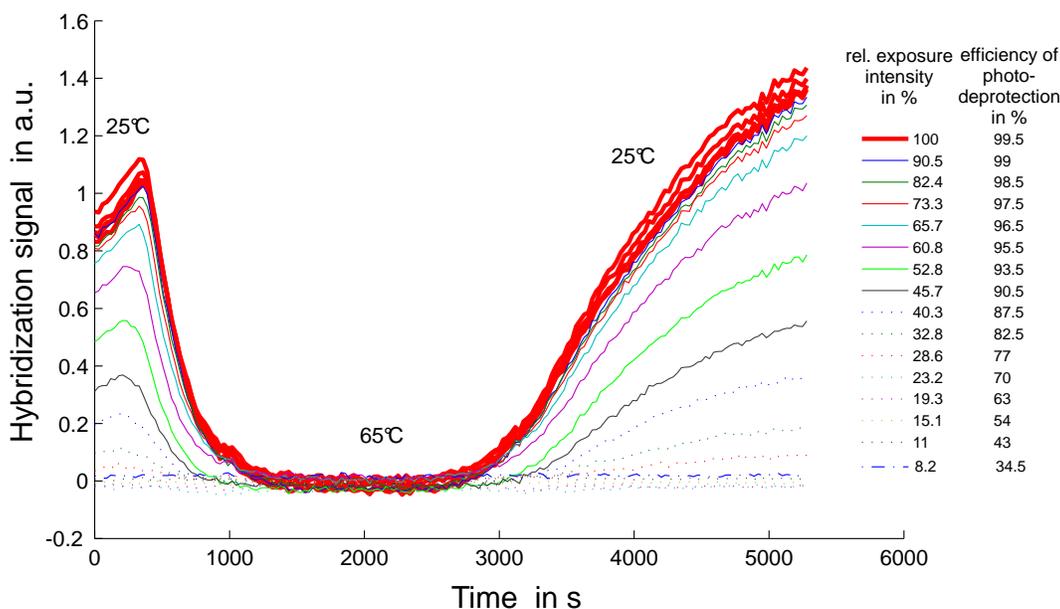
We therefore investigated the impact of synthesis defects with an experiment in which the amount single base deletions (resulting from incomplete photo-deprotection) was incrementally varied.

**Experiment: How do synthesis-related defects affect DNA microarray hybridization characteristics?**

16 microarray features comprising 20mer probes with an identical sequence were deprotected with different UV intensities (controlled by different photolithography mask brightness), so that the photo-deprotection of the NPPOC groups was to a greater or lesser extent completed. The result of the experiment (Figure 8.4) demonstrates that the hybridization and melting behavior, owing to the varying number of single base deletions, strongly depends on the completeness of the photo-deprotection. Deletion defects originating from incomplete photo-deprotection reactions reduce the hybridization signal and result in a significantly reduced melting temperature. A similar experiment was conducted to investigate if the exposure currently employed is sufficient, or if a longer exposure can significantly improve the probe quality. Fig. 8.5 shows the hybridization signals of 16 features (with identical probe sequences but different exposure) as a function of time. The exposure time of the individual features was determined from image grayscale values using the projectors gamma function (see section B.2.1). The total exposure time was 150 s. Thus, the exposure dose value commonly employed in the synthesis process (100 s with a feature brightness of 100%) corresponds to a relative exposure intensity of about 67%. The stability of these probes synthesized with an exposure dose corresponding to 100 s (cyan solid curve) is only slightly smaller than that of probes synthesized with the full exposure of 150 s. An exposure time of 100 s (commonly employed in the synthesis process) is a reasonable compromise between a short exposure time (determining the length of the synthesis process) and the quality of the synthesized probe sequences.



**Figure 8.4:** Influence of deletion defects on the melting **A** and hybridization behavior **B** of 20mer probes. The relative brightness of the features in the JPEG images employed as "virtual photolithography masks" (thus determining the exposure dose of the individual features) was varied in increments of about 3% from 100% (feature F1) to 54% (feature F16). The length of the exposure was 100 s. The upper plots in **A** and **B** show the temperature of the hybridization solution. Diagram **A** demonstrates that the melting temperature of the probes depends strongly on the amount of deletion defects contained in the sequences. Feature F10 begins to melt at about 35 °C, whereas feature F1, containing the least synthesis defects, starts melting only at 55°C. Vice versa, when the temperature of the hybridization solution is reduced (in **B**) the hybridization signal of higher quality probes increases significantly faster and stronger than that of probes containing a larger number of deletion errors. Temperature influence on the Cy3 fluorescence intensity (see section 8.2) is not accounted for.



**Figure 8.5:** Impact of deletion errors on the hybridization/melting behavior of the 20mer probe-sequence 3'-TTGAGCGATATTACTGGACC-5'. The relative exposure intensity, and the corresponding efficiency of photo-deprotection (determined from Fig. 8.3) are shown at right. The temperature is linearly increased from 25 °C to 65°C (between 0 and 2000 s). This is followed by a linear decrease from 65°C to 25°C (between 2000 and 4000 s). Subsequently the temperature is held constant at 25°C.

### Microarray feature size reduction

The synthesis of microarrays with small and densely arranged features, owing to the reduced local contrast ratio, suffers more from stray light than the synthesis of microarrays with larger structures. Stray light induced photo-deprotection of probes in neighboring features gives rise to random MM and insertion defects. Additionally, the quantitative analysis of microarray hybridization signal intensities, owing to the reduced optical contrast, the increased relative impact of small scale inhomogeneities (microarray substrate, particles, etc.), and the requirement of an exact placement of the readout grid, becomes increasingly difficult with decreasing feature size. These limitations require a minimum feature size of about 3x3 pixels (DMD mirrors), plus a separation gap of 2 pixels. The requirement of 25 pixels per feature enables a maximum number of about 30000 features. For a less quantitative analysis (e.g. to investigate gene expression) a feature size of 1 pixel (possibly requiring an increased exposure time and improved image contrast) with a feature separation gap of 1 pixel would enable a total feature number of about 200000 features on a chip area of 10 mm<sup>2</sup>. Quantitativeness in gene expression assays is significantly affected by other factors (e.g. target secondary structure), thus the loss of quantitativeness resulting from feature size reduction presumably wouldn't account significantly.

With 200000 probes, assuming a number of 10 different probe sequences per gene (for the purpose of redundancy - poor predictability of probe-target binding affinity [Poz06]), almost whole-genome expression assays comprising about 20000 genes might be feasible. However, to make use of such a high feature density, more effort needs to be put in the automatization of microarray readout and image analysis.

## 8.2 Temperature Dependence of the Cy3-Fluorescence Intensity

For quantitative analysis of hybridization signal intensities one has to consider the significant temperature dependence of fluorescent markers [Kot00b; Kot00a; Liu05]. The fluorescence intensity versus temperature characteristics depends strongly on the particular fluorescent label employed. For example, the temperature influence on the fluorescence intensity of *Texas Red* is almost negligible, whereas the *Cyanine-3* fluorescence is strongly temperature dependent: in solution between 7.5°C and 80°C an approx. linear decrease of the fluorescence intensity from 100% to about 20% was observed [Liu05]).

Since the experimental conditions described in [Liu05] were differing from experimental conditions employed in our study, the Cy3-fluorescence-temperature dependence was investigated under the same conditions as in our microarray hybridization experiments, i.e. with Cy3-fluorophores attached to surface tethered DNA strands.

In particular we investigated if the fluorescence-temperature dependence is the same for fluorophores in the bulk solution and for fluorophores bound to the microarray surface. An interesting question is whether differences in the lengths of the probe strands - which can be regarded as spacers between the surface and the fluorophores - result in differences of the temperature response.

### 8.2.1 Experiment

The microarray used in this experiment comprises probe sequences of different lengths between 1 and 20 nt. In the final coupling step of the DNA chip synthesis a fluorescent Cy3-marker (Cy3<sup>TM</sup> phosphoramidite, Amersham Pharmacia Biotech.) was coupled to the 5'-end of the surface tethered probe sequences.

The fluorescence of the Cy3-labeled probes decreases with the length of the probe strands. This is most likely owing to the decreasing yield (synthesis defects) of probe strands available for the final Cy3 coupling step.

In the first part of the experiment the fluorescence signal of surface-tethered Cy3-end-

labeled probes was monitored in  $5\times$ SSPE buffer (identical to the hybridization buffer, however, not containing any fluorescent targets strands). The temperature was varied slowly between  $30^{\circ}\text{C}$  and  $70^{\circ}\text{C}$  (see Fig. 8.6A).

In the second part of the experiment we added a high concentration (100 nM) of Cy3-labeled "target" oligonucleotides to the buffer solution. The sequence of the "targets" was chosen to prevent hybridization with the surface-tethered probes. In this way we could simultaneously observe the fluorescence of Cy3-labeled "targets" in bulk solution and surface-tethered Cy3-fluorescence signals (the latter in superposition with the bulk solution fluorescence).

## 8.2.2 Results and Discussion

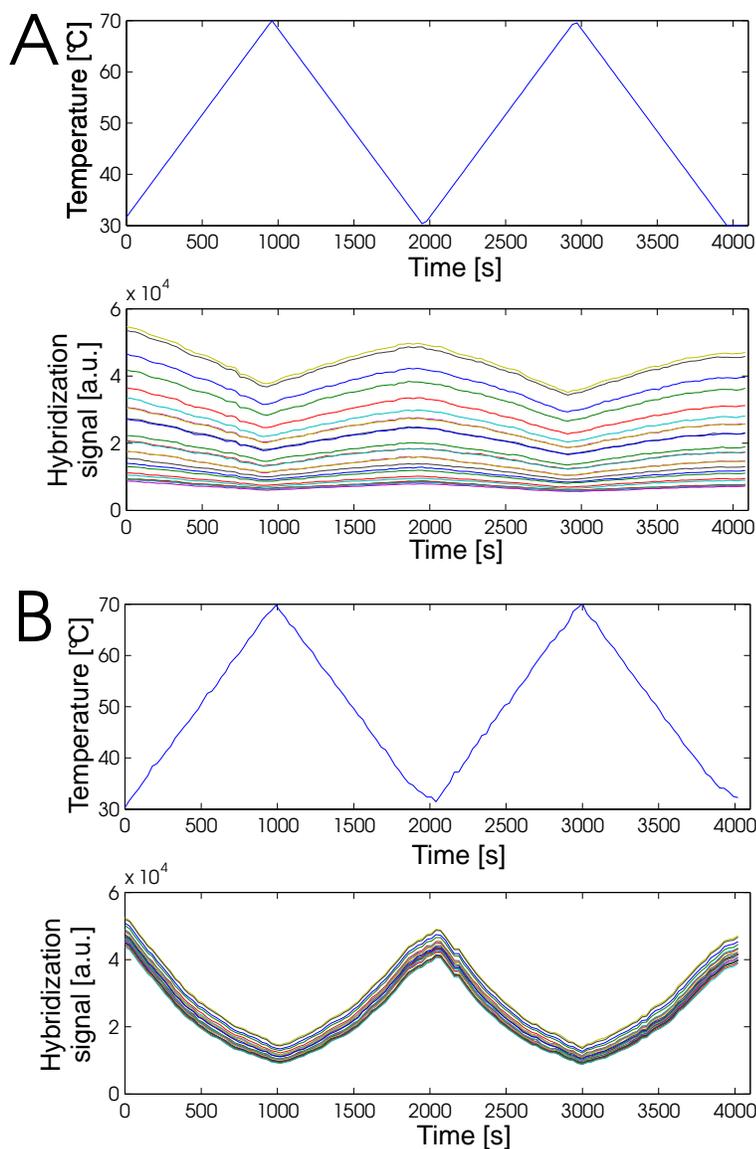
Fig. 8.6A shows the temperature and the fluorescence intensity of the surface bound fluorophores. The temperature ramp was repeated to ensure that intensity changes are reversible and not due to photobleaching or detachment of the probes. The observed small decrease of the fluorescence intensity could be owing to photobleaching.

A significantly stronger temperature influence is observed for Cy3-fluorescence in solution (Fig. 8.6B). The plot of fluorescence intensity versus temperature reveals that the fluorescence intensity on the surface decreases linearly with temperature (Fig. 8.7A), whereas in solution the fluorescence decreases exponentially with temperature (Fig. 8.7B).

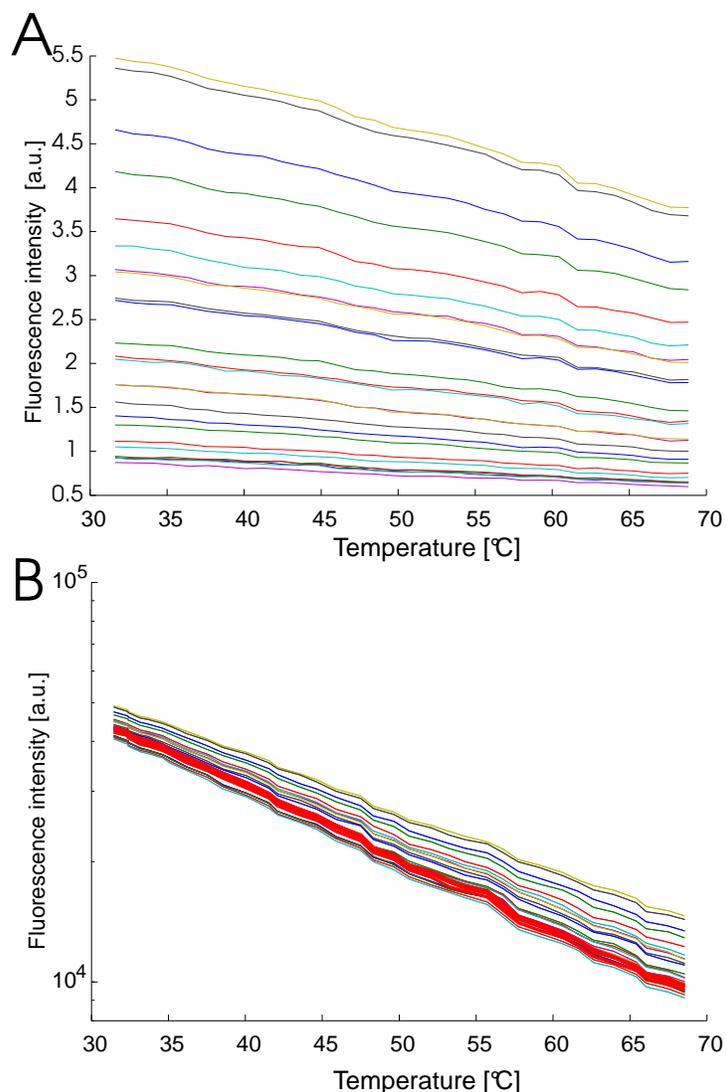
Between  $31.5^{\circ}\text{C}$  and  $68.5^{\circ}\text{C}$  the fluorescence intensity of the surface-tethered Cy3-markers decreases to 67% of the original value, whereas the fluorescence intensity in solution is reduced to 23%. The intensity decrease observed in bulk solution is similar to the intensity decrease described in [Liu05] (decrease to about 35% for the corresponding temperature range).

An explanation for the large differences between the surface and solution fluorescence signal could be that the chemical environment of surface-bound fluorophores is different from that of fluorophores contained in the bulk solution.

The linear decrease of the Cy3-fluorescence in Fig. 8.7A (identical for all features) is independent of fluorophore density. Moreover, the length of the probe strands (i.e. the tether length between the surface and the fluorophore), which was varied between 0.34 and 6.8 nm, didn't have an influence on the fluorescence intensity-temperature characteristics.



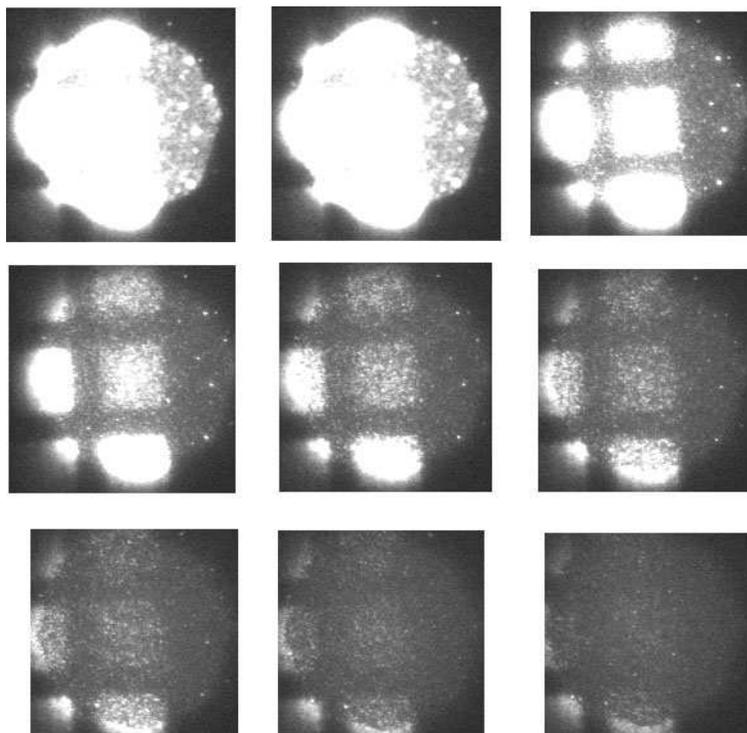
**Figure 8.6:** Temperature influence on the Cy3-fluorescence intensity (A) on the DNA chip surface (surface tethered fluorophores) and (B) in bulk solution. During the experiment the temperature (upper images) was varied between 30°C and 70°C. The individual curves (in the lower images) correspond to fluorescence intensities of individual microarray features with a varying density of Cy3-fluorophores (attached at the 5'-end of DNA probes of varying length). In (B) a high concentration (100 nM) of Cy3-labeled oligonucleotides (sequence chosen not to hybridize with the surface-bound probes) was added: Now the fluorescence of the surface-bound fluorophores is overlaid by the much larger fluorescence of Cy3-labeled oligos in the solution. The fluorescence intensity in the bulk solution is significantly more affected by temperature changes than that of surface bound fluorophores. The temperature ramp was repeated to ensure that a reversible temperature-related effect is observed.



**Figure 8.7:** Temperature dependence of the fluorescence signal of (A) surface-tethered Cy3-labeled probes and (B) a superposition of the same surface tethered Cy3-labeled probes and Cy3-labeled DNA in solution (B). The various fluorescence signals shown in (A) and (B) originate from different microarray features with probe lengths ranging from 1 to 20 bases. In the final step of the DNA chip synthesis the fluorescent Cy3-amidite was coupled to the 5'-end of the surface tethered probe sequences. Between 31.5°C and 68.5°C the hybridization signal of the surface-tethered Cy3-fluorophores decreases linearly to about 67% of the initial intensity. The five smallest signals correspond to background intensities originating from nonspecific binding of the Cy3-amidite to the microarray surface. Plot (B) shows hybridization signals of the the same features, however, superposed by a large fluorescence intensity from Cy3-fluorophores dissolved in hybridization buffer. Unlike plot (A) plot (B) has a semilogarithmic scale. Five of the signals (bold red lines) are genuine solution signals (with no overlying surface signal - apart from a negligible fraction of nonspecifically bound Cy3-molecules). Some surface signals seem to be weaker than the solutions signals. This is owing to an uncorrected gradient in the fluorescence excitation. In solution we observed a significantly stronger decrease of the hybridization signal to about 23% of the original intensity at 31.5°C. Here the temperature dependence is exponential rather than linear.

### 8.3 Single Molecule Imaging on DNA Microarrays - Photobleaching and Photoblinking

Image acquisition with the fluorescence microscope (Olympus IX81, PLAPO 60×1.4NA oil objective) and an electron multiplying EM-CCD camera (Hamamatsu EM-CCD C9100-02) demonstrated our setups capability for single molecule detection.

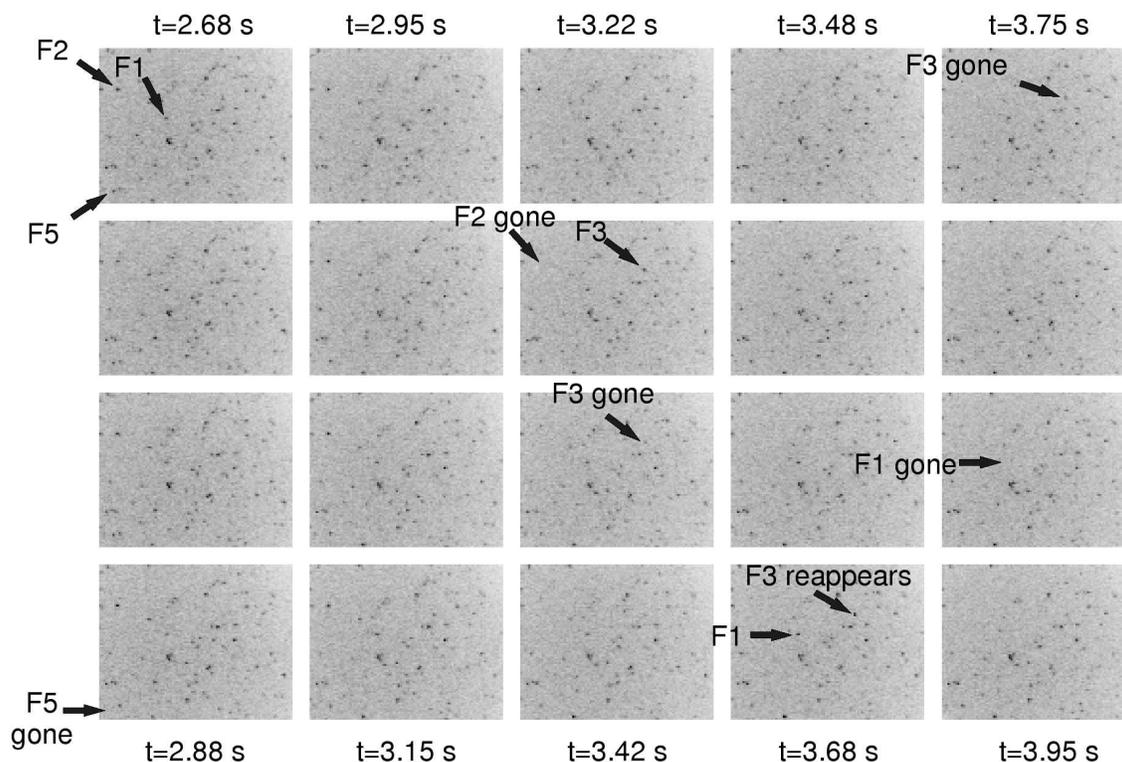


**Figure 8.8:** Photobleaching and photoblinking in a time lapse series of fluorescence micrographs of a hybridized DNA microarray. Square areas are individual microarray features with a size of 17.5  $\mu\text{m}$ . Duration from first to last image: ca. 17 s. With increasing photobleaching individual Cy3 fluorophores become visible. Apart from irreversible photobleaching we observe blinking of individual fluorophores on a one-second-time scale. EM-CCD exposure parameters: exposure time 0.033 s, gain 120, 4x binning.

The series of fluorescence micrographs in Figs. 8.8 and 8.9 shows that photobleaching is overlaid by photoblinking. The observed photoblinking is characterized by long-lasting off-periods (time scale of seconds) of individual the Cy3-fluorophores. The analysis of the time evolution of individual fluorophores in Fig. 8.10 shows that the blinking characteristics of individual fluorophores can be very different: For example, the fluorophore F3 has a significantly higher blinking rate than the fluorophores F2 or F5.

Except for the described experiments, no extensive analysis of the blinking characteristics has been undertaken since single molecule blinking (unlike single molecule imaging) is

not directly in the scope of this work.



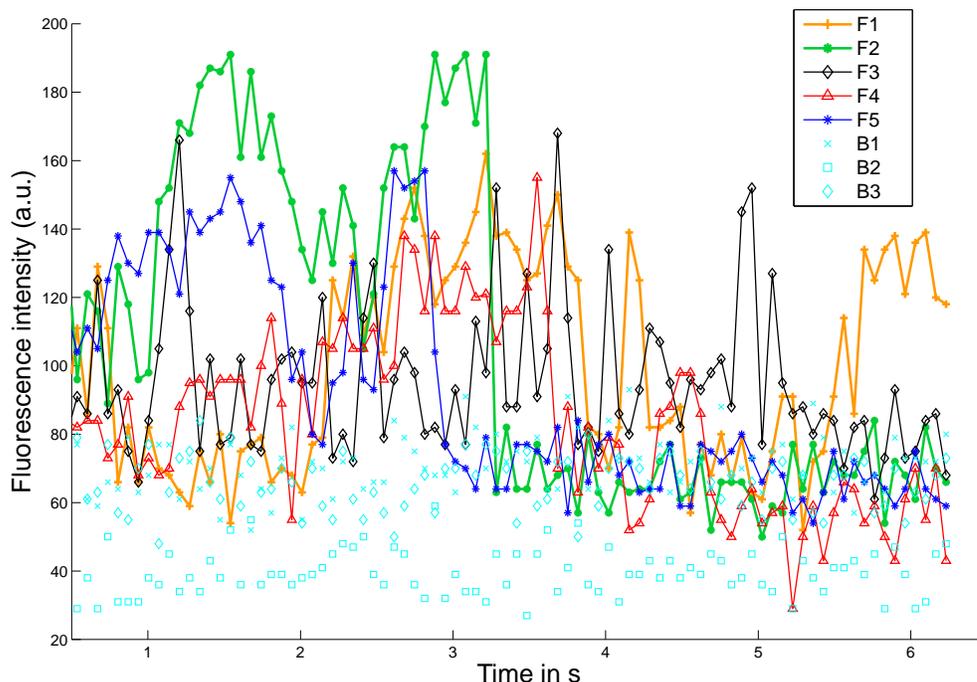
**Figure 8.9:** Photoblinking of individual fluorophores. The Cy3-end-labeled oligonucleotide targets were hybridized on the surface of the microarray. Under the applied observation conditions (high salt buffer, room temperature, no mobile target strands in solution) the duplexes can be regarded as stable. The series of fluorescence micrographs (for contrast enhancement intensity levels were squared and inverted) corresponds to Fig. 8.10 between  $t=2.68$  s and  $t=3.95$  s (time lapse 0.0667s). Image size (ca.  $30\mu\text{m}$ ). For the fluorophores F1, F2, F3 and F5 the time evolution of is shown in Fig. 8.10. Exposure parameters: exposure time 0.033 s, gain 120, 4x binning.

### Long time scale fluorescence intermittency in previous work

The effect of long timescale photoblinking has been described previously, e.g. in [Sab05] and [Sch07]. The underlying physical mechanism for fluorophore blinking is poorly understood [Sab05].

Triplet-blinking occurs on the micro- and millisecond scale. Whereas the long-lasting dark states observed in the experiments have a duration of up to several seconds. Rotation of the molecular dipole may be ruled out by polarization resolved detection [Sch07].

Schuster *et al.* [Sch05] propose the following explanation for long living dark states: Photoinduced charge ejection results in the formation of a radical pair consisting of a dark state fluorophore (cation) and an electron which is trapped in the polymer matrix. This is supported by the observation of a power law distribution of the dark state life-time, which



**Figure 8.10:** Time evolution of the fluorescence of individual fluorophores. The fluorophores F1, F2, F3 and F5 are shown in Fig. 8.9. The corresponding series of fluorescence intensities from randomly chosen background points B1, B2 and B3 indicate the level of noise. Fluorophores F2 and F5 seem to undergo irreversible photobleaching (at  $t \simeq 2.9$  s and  $t \simeq 3.3$  s, respectively), whereas F1 and F3 and F4 recover repeatedly. The blinking rate of F3 is significantly larger than the blinking rate of the other fluorophores considered in this diagram.

suggests that not intrinsic electronic properties of the dye molecule but rather an influence of the local environment is responsible for the observed fluorescence intermittency. The dark state life-time depends the dielectric properties of the surrounding polymer matrix. From the trapped dark state either recombination of the radical pair or, alternatively, formation of a stable non-fluorescent photoproduct (photobleaching) can occur.

In principle the detection of single target molecules is possible with our experimental setup. However, photobleaching is expected to be a serious problem. Photobleaching could be prevented by labeling with antibody-conjugated quantum dots.

## 8.4 Duplex Melting Characteristics on DNA Microarrays

Nucleic acid hybridization, due to base stacking, results in a decreased UV absorbance (hypochromicity) with respect to denatured single strands. Measurement of the UV absorbance – providing the fraction of stacked bases (which is equal to the fraction of hy-

bridized duplexes if a two-state transition is assumed) – is commonly employed for the investigation of the nucleic acid denaturation transition. Unlike typical UV melting curves obtained from absorption spectroscopy measurements, our hybridization experiments on 16 to 25mer microarrays (fabricated by light-directed *in situ* synthesis) don't show a distinct melting transition, but rather a continuous decrease of the hybridization signal with increasing temperature. We have performed experiments to investigate the hybridization/denaturation behavior of oligonucleotide duplexes on DNA microarrays.

### 8.4.1 Experimental Procedures

For the experiments we have fabricated a DNA microarray comprising probes of different lengths ranging from 1 to 23 nucleotides. All probes were complementary to the target sequence *COM*. The target concentration used in the hybridization/melting assay was 1 nM (hybridization buffer: 5×SSPE with 0.1% SDS).

Prior to the *melting experiment* (Fig. 8.11) the targets were allowed to hybridize on the microarray for 30 minutes at a temperature of 30°C. Then the temperature was increased in steps of 5°C to 70°C. A duration of ca. 2000 s for each step ensured that hybridization equilibrium (or close approach to equilibrium) could be achieved.

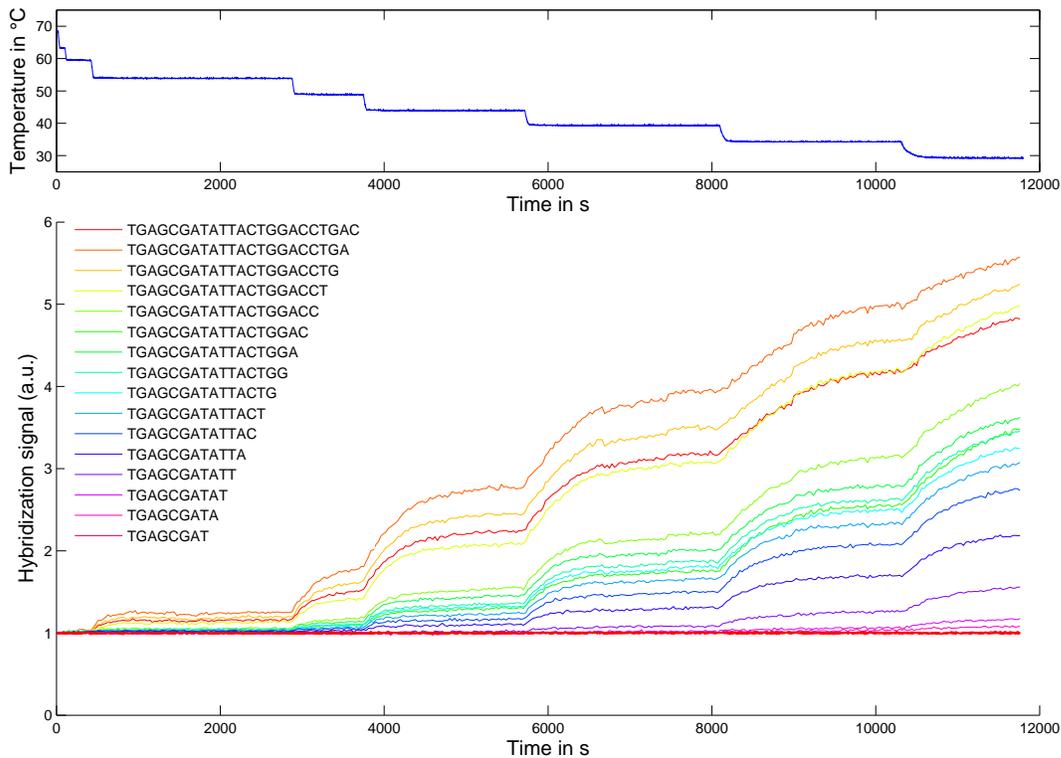
An analog *hybridization experiment* (Fig. 8.12) was performed, however with the temperature program running from 70°C (no initial hybridization) to 30°C, with decrements of 5°C. Time steps, as above, were approximately 2000 s long.

Hybridization signals were acquired in real-time, while the microarray was immersed in the hybridization buffer (with 1 nM Cy3-end-labeled target oligonucleotide), on the fluorescence microscope setup. Fluorescence micrographs were recorded with *SimplePCI* at an interval of 60 s. The temperature of the hybridization chamber was controlled by a software based PID-controller (see section B.11).

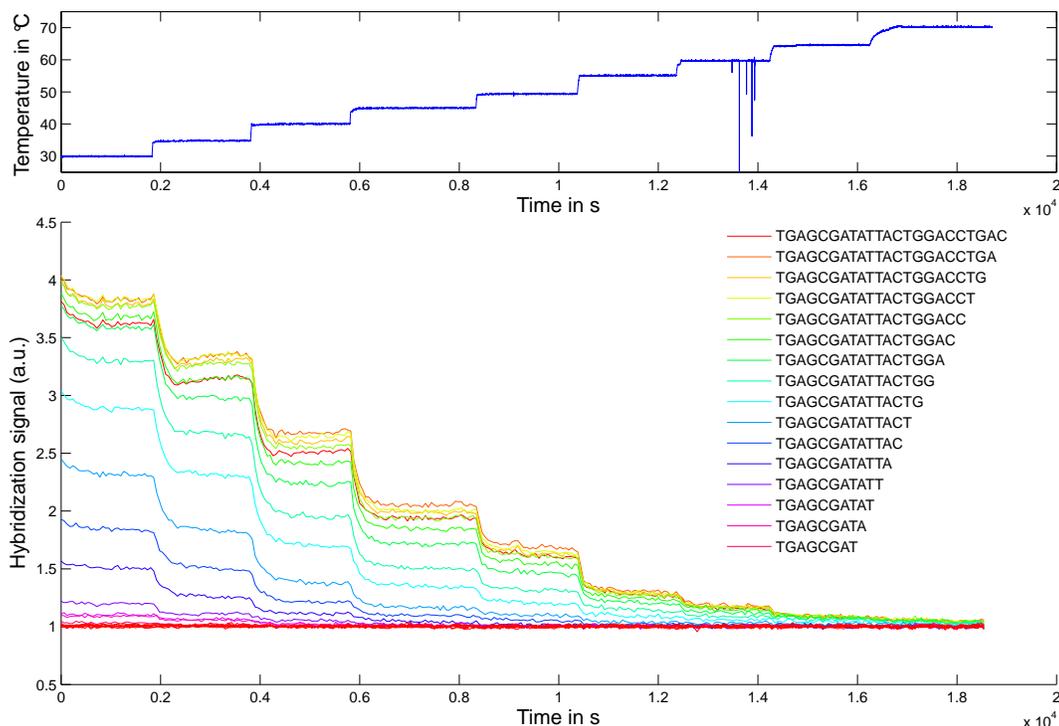
Analysis of the fluorescence micrographs was performed with the *ScanRA* program (see section B.10), which includes a batch processing function for the analysis of image series. Lateral shifting of the feature block, owing to thermal expansion of the hybridization chamber was corrected with the drift-correction function. Analysis of the intensity data was performed with MatLab (Mathworks Inc.).

### 8.4.2 Results and Discussion

Melting curve analysis is commonly performed with UV absorbance measurements, or by measurement of the fluorescence signal of an intercalated dye (e.g. SYBR Green I). In comparison to these measurements performed in solution (see for example [Owc04]), the



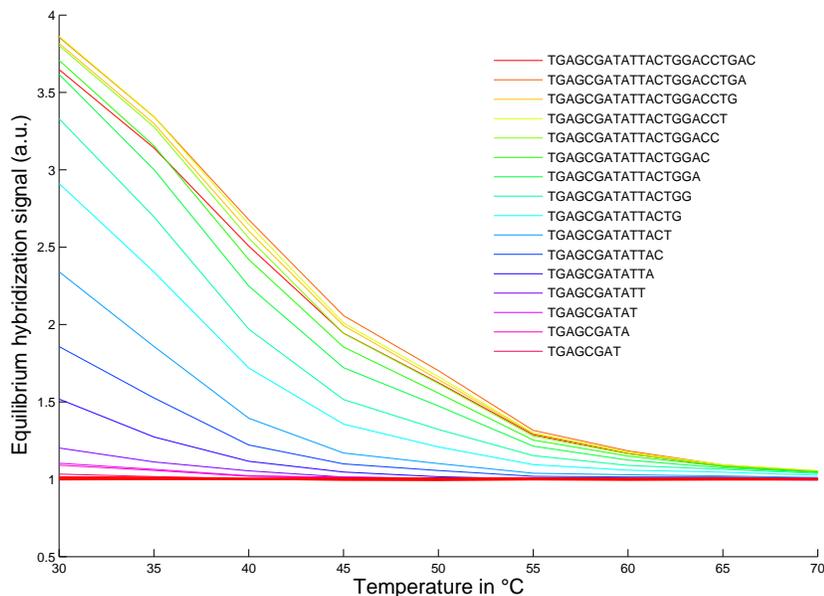
**Figure 8.11:** Stepwise hybridization. The temperature (top) is decreased in increments of 5°C from 70°C to 30°C. The legend shows the corresponding probe sequences. A significant hybridization signal is only observed for temperatures below 60°C. At 30°C hybridization of 9mer probes is observed. Hybridization equilibrium is reached faster at higher temperatures (in about 300 s at 55°C, about 2000 s at 45°C). At lower temperatures the longer probes require more time to reach equilibrium than shorter probes: For example at 35°C the 22mer probes don't reach equilibrium after 2000 s, whereas the 12mer probes reach equilibrium after about 1500 s.



**Figure 8.12:** Stepwise melting. Comparison of hybridization signals from probes of varying length - 1 to 22 nt. The temperature is increased in increments of 5°C from 30°C to 70°C. Probes shorter than 9 nt don't show a significant hybridization signal (red base line). The hybridization signal of the 22mer probe is still noticeable at a temperature of 70°C, whereas hybridization signals of 9mer probes were only observed up to a temperature of 35°C. After a temperature increase (jump) the equilibrium is reached much faster (requiring between 100 and 500 s) than after a corresponding temperature decrease (in Fig. 8.12).

melting transition observed on our microarray surfaces is significantly broadened. As can be seen in Fig. 8.13, the hybridization signal of the probes is gradually decreasing between 30°C and 70°C. The melting curves of the longest probes (17-22 nt) have an inflection point between 30 and 40°C. Melting curves of shorter probes ( $l < 17$  nt) don't have an inflection point in the temperature range investigated.

In a comparable experiment [Wic06] (DNA chips fabricated by light-directed *in situ* synthesis with the Xeotron platform), the melting curves (microarray hybridization signal versus temperature) of 18mer duplexes have an inflection point at about 40°C. The saturation of the hybridization signal (below 30°C) is more pronounced than in our experiment. This could indicate an increased number of synthesis errors in our probes with respect to [Wic06], it may, however, also depend on the stability of the particular sequences chosen for the experiments. Unpublished melting curves presented in product specs. of another commercial microarray platform (based on light-directed *in situ* synthesis) show melting characteristics which are very similar to our results in Fig. 8.13.



**Figure 8.13:** Equilibrium melting curve - equilibrium hybridization signal plotted versus temperature - extracted from Fig. 8.12. We observe no saturation at the low temperature end (even though for longer probes there appears to be an inflection point at 40 °C) - thus, there is no distinct melting transition range.

The significantly broadened melting transition may be owing to heterogeneity of binding affinities originating from synthesis defects. These result in a decreased binding affinity. Thus, the experimental melting temperature is expected to be lower than the theoretical predicted melting temperature of the duplex: melting temperature calculation with the DINAMelt Server (with parameters: 1 nM DNA, 1 M NaCl) provided  $T_m = 63.7^\circ\text{C}$  for the 22mer probe and  $T_m = 22.9^\circ\text{C}$  for the 10mer probe.

In the melting experiment (Fig. 8.12) at temperatures between 60 and 70°C a small hybridization signal is observed<sup>2</sup>, whereas in the hybridization experiment a significant hybridization signal is only observed at temperatures  $\leq 55^\circ\text{C}$ .

Hybridization equilibrium is achieved faster ( $\Delta t = 100$  to 500 s) in the stepwise melting experiment (Fig. 8.12) than in the stepwise hybridization experiment (for the longest sequences  $\Delta t = 100$  s at 55°C and  $\Delta t > 2000$  s at 40°C). The increasingly slow hybridization kinetics is probably owing to a decreasing duplex formation rate  $k_{nuc} \sim [T_0] \cdot [P - D]$  as the fraction of hybridized probes approaches saturation.

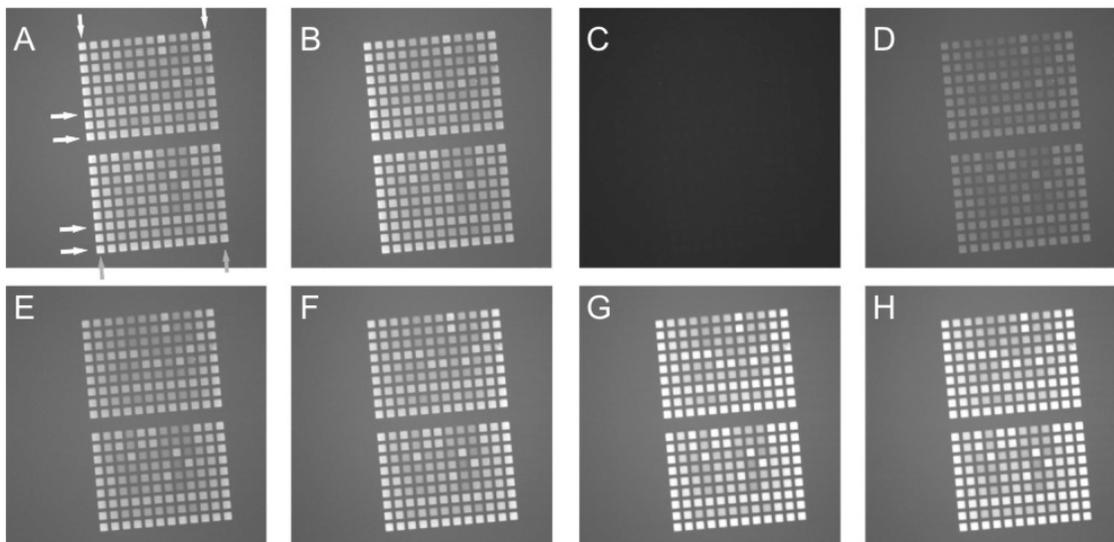
One might also assume that at lower temperatures competitive nonspecific binding may further slow down the hybridization kinetics. However, for the experimental conditions

<sup>2</sup> this could be owing to a small amount of irreversibly adsorbed targets (often remaining from a strong hybridization signal)

employed this can be ruled out since no significant hybridization signal was observed from probes unspecific to the target sequence.

## 8.5 Target Transport Related Effects

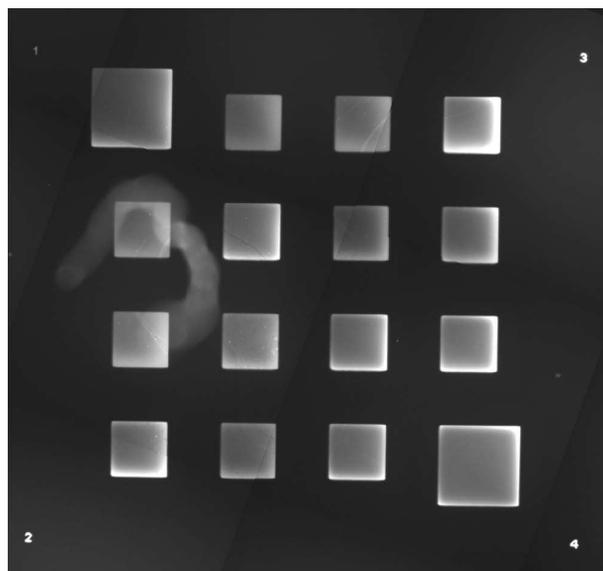
### 8.5.1 Experimental Observations



**Figure 8.14:** Inhomogeneous hybridization due to (assumedly convective) fluid transfer in the hybridization chamber. The image series of fluorescence micrographs A-H shows the time evolution of the hybridization signal. Arrows in image A point to rows and columns of control features with identical PM probes. These control features would ideally have an identical hybridization signal intensity. We find, however, that the leftmost column has a significantly increased intensity with respect to the other control features. In image A (identical to image B) the chip has been hybridized at room temperature. (C) The microarray has been heated up to about 60°C. Due to dissociation of the duplexes the hybridization signal has almost vanished (this is also partly owing to the temperature dependence of the Cy3-fluorescence). D-E Cooling down to room temperature, enables re-hybridization of the targets. In D all control features have the same hybridization signal intensity, whereas in E we find that the lowermost row (at the edge of the feature block) has a significantly higher intensity than the next but one row which is comprising the same control features. In images (F-H) we observe a slightly increased hybridization signal of the control features at the lower/left edges of the feature blocks. This may be indicating a (supposedly convective) target transfer from the lower right towards the upper right. However, inhomogeneities are significantly reduced with respect to image A.

Within the feature blocks shown in Fig. 8.14 the identical PM control features placed next to the cognate (mismatched) probe sequences display varying hybridization signal intensities, and thus indicate a spatially inhomogeneous hybridization. Control features located at

the edges of features blocks can be significantly brighter than identical control features in the interior of the feature block (as shown for instance in Fig. 8.14A). Similar hybridization signal gradients can also be observed in single large features like shown in Fig. 8.15. The distribution of hybridization signal intensities in Figs. 8.14 and 8.15 showing bright intensities on one side of the features/feature block and reduced intensities at the opposite side, indicates that a directed transport of targets (e.g. due to convection in the hybridization chamber<sup>3</sup>) is responsible for the uneven hybridization of the target molecules.



**Figure 8.15:** Target depletion effects are visible in the fluorescence micrograph of a microarray comprising relatively large features (ca. 150  $\mu\text{m}$  in size). Most prominent is the bright hybridization signal mainly at the lower and right edges of the features. At the left edges the intensities are similar to the intensities in the center of the features. Target transport (from the lower right to the upper left) is probably due to laminar flow in the capillary gap between the microarray surface and the cover glass. Different feature intensities result from single base MMs. Features on the diagonal from upper left to lower right are PM features. The hybridization of the 16mer duplexes was performed at room temperature.

## 8.5.2 Discussion

Experimental observations suggest that target molecules are transported not only by diffusion, but also by convection (within the  $\sim 1\text{mm}$  high hybridization chamber), or by a laminar flow in the thin capillary gap between the microarray and a cover glass (as in Fig. 8.15). In case of fast hybridization kinetics probes located near the edge of a feature are more likely to capture targets than probes in the center of a feature: Owing to the local

<sup>3</sup> Convective transport in the hybridization chamber is inferred from the observation of directional movement of fluorescent particulates.

depletion of the hybridization solution [Mic07] fewer targets are left hybridize to probes in the interior of the feature. This is giving rise to a bright rim. Directional transport of the target molecules may result in an uneven distribution of the hybridization signal like shown in Fig. 8.15.

The circular features on printed microarrays<sup>4</sup> often have very bright hybridization signals at the rim compared to reduced hybridization signals in the center. Doughnut-shaped patterns have been attributed to spotting or drying artifacts. However, recently Pappaert *et al.* reported that the ring-pattern may occur when diffusion-limited conditions are present during the hybridization process [Pap06]. Their findings are in good agreement with our observations on *in situ* synthesized microarrays (which are certainly not affected by spotting or drying artifacts). Our experiments confirm the observations of [Pap06]: "the faster the binding kinetics of the target probe complex, the higher the risk for a more dense coverage of the edges of the microarray spot." In the quantitative hybridization experiments throughout this study, uneven hybridization through diffusion limitation effects was prevented by using *probe-target*-complexes with not too high binding affinities (e.g. 16mers rather than 25mers, or probe sequence motifs with suitable CG-contents). Alternatively, suitable binding affinities can be adjusted by variation of the hybridization temperatures.

## 8.6 Influence of Target Secondary Structure on the Duplex Binding Affinity

The selection of appropriate probe sequences is an important issue in the design of DNA microarrays. The sensitivity of the probes within probe-sets (i.e. probes specific to the same gene) on commercial DNA chip platforms can vary over two orders of magnitude [Zha03; Bin04].

"Whereas the specificity of binding is predictable, the efficiency of short oligonucleotide probes binding to long nucleic acid targets has not been predictable. Most short (~20mer) oligonucleotide probes selected for a given target will not bind efficiently to the full length transcript. As a result, many potential antisense agents lack efficacy, and many oligonucleotide probes afford poor signal to noise ratio in hybridization-based measurements." [Lue03]

Possible reasons for the large variation are the target secondary structure [Mir99] (intramolecular base pairing preventing hybridization with the microarray bound probes) and intramolecular folding of the probes [Lue03].

---

<sup>4</sup> The circular shape is a result of the droplet deposition process.

To investigate the reasons for this large variation of binding affinities we performed a simplified hybridization assay, using only one single cRNA transcript as a target sequence. In the tiling array experiment (similar experiments were reported in [Mir99; Lue03; Mat03]) the oligonucleotide probes form a tiling path complementary to the comparatively long cRNA target sequence (see Fig. 8.16A).

Additionally, we performed a similar series of tiling array experiments with a 74mer DNA oligonucleotide target which forms a relatively simple, well-known stem-loop secondary structure.

### 8.6.1 Preparation of the cRNA Target Sequences

The plasmid *pEGFP-Tub* vector (encoding a fusion protein consisting of enhanced green fluorescent protein *EGFP* and human  $\alpha$ -*tubulin*, Clontech Laboratories Inc.) was cloned in *E. Coli* and isolated via miniprep.

#### Amplification and labeling of the transcript by *in vitro* transcription

*In vitro* transcription (IVT) was employed for amplification and labeling of the target sequence. T7 RNA-polymerase (T7 refers to the T7 bacteriophage from which the particular RNAP originates) creates RNA transcripts from a DNA template by polymerization of ribonucleotides. Fluorescent labeling of the RNA transcripts is performed by random incorporation of Cy3-UTP nucleotides. T7 RNA-polymerase is promoter specific. The T7-promoter-sequence was included into the *pEGFP-Tub* template-strand by a PCR step (polymerase chain reaction) prior to the IVT. The upstream PCR-primer comprises the T7-recognition-sequence (5'-TAATACGACTCACTATAG-3', appended to the primer 5'-end) and a complementary section for hybridization with the template. The length of the PCR product is controlled by the selection of the downstream PCR-primer. Two separate PCR reactions with different downstream primers were performed to produce different length PCR-products.

## Standard PCR-protocol

### Buffers and reagents:

- DNA 1  $\mu$ l (undiluted, 1:10, 1:100, 1:1000)
- Upstream primer 2.5  $\mu$ l (10 pmol/ $\mu$ l) T7-F1:  
5'-TAA TAC GAC TCA CTA TAG GGG AAC CGT CAG ATC CGC-3'
- Downstream primer 2.5  $\mu$ l (10 pmol/ $\mu$ l)  
primer sequence P1 for long (ca. 800 nt) cRNA target T1:  
5'-TGG AGA TGC ACT CAC GCA CTC G -3'  
and alternatively  
primer sequence P4 for short (ca. 270 nt) cRNA target T2:  
5'-TGA AGC ACT GCA CGC CGT AGG TC-3'
- 10 $\times$  Taq PCR-Buffer 5  $\mu$ l (containing MgCl<sub>2</sub> 1-10 mM)
- dNTPs (10 mM each nucleotide) 1  $\mu$ l
- Taq-Polymerase (5 U/ $\mu$ l) 0.2-0.5  $\mu$ l
- complete to a final volume of 50  $\mu$ l with nuclease-free water

### PCR amplification cycle:

- I. denaturation 5 min at 95°C
- II. 30 cycles
  - denaturation 30 s 95°C
  - primer annealing 30 s at 64°C
  - elongation 90 s at 72°C
- III. elongation 5 min at 72°C
- IV. PCR completed - storage at 4°C

Gel electrophoresis (1% agarose gel) confirms that the PCR products (depending on the downstream primer chosen) have lengths of ca. 300 nt and ca. 800 nt, respectively. Purification of the PCR-products with QIAquick™ kit.

Quantity of PCR-product from one reaction (UV-absorbance measurement): 0.6  $\mu$ g DNA.

***In vitro* transcription (IVT) protocol**

Labeling-IVT was performed according to the protocol provided in the product specification for FluoroLink™ Cy3™-UTP (Amersham Biosciences).

**Reagents:**

- *in vitro* transcription kit: Ambion T7 MegaScript™
- fluorescent marker: FluoroLink™ Cy3™-UTP (Amersham Biosciences, PA53026)

**Preparation of the IVT:**

- lyophilisation of the DNA template
- dilute DNA template in 7.5 µl nuclease-free water (amount appropriate for IVT reaction)
- thaw reagents from IVT kit (except the enzyme mix) on ice

**IVT labeling protocol:**

Add reagents, in the following order to a nuclease free 1.5 ml conical microfuge tube:

- 10×reaction buffer 2 µl
- A/C/G stock solution 6 µl
- U stock solution 2 µl
- Cy3™-UTP solution (5 mM) 0.5 µl
- enzyme mix (containing T7 RNA-polymerase) 2 µl

**IVT reaction:**

- add DNA template 7.5 µl
- incubate at 37°C for 2-4 hours, in the dark
- remove the template DNA by addition of 1µl DNase I (RNase-free). Mix gently and incubate for 15 minutes at 37°C.

Purification of the cRNA with Qiagen RNeasy™mini protocol for RNA cleanup. Elution with 40 µl nuclease-free water. Storage of the labeled cRNA at -80°C.

### 8.6.2 Design of the Tiling Array Experiment

The cRNA targets (T1 and T2 - see Fig. 8.16A) comprise the *pEGFP-Tub* plasmid sections 578 to 832 (T1) and 578 to 1367 (T2) - numbers refer to base positions specified in the ClonTech pEGFP-Tub Vector Information.

The tiling array comprises sets of 25mer tiling probes complementary to the *pEGFP*-sections 555-1370 (tiling interval: 1 base), see Fig. 8.16B, and 1370-2670 (tiling interval: 2 bases). Only probes targeting the section 555-1370 are expected to specifically hybridize with the target T2. For the probes 1370-2670 no specific targets are available in the ex-

periment. Probes complementary to this section will serve as indicators for nonspecific hybridization.

The microarray further contains a set of "reverse oriented probes". Probe sequences in this probe set are the same as above, however the strand orientation is reversed (example shown in Fig. 8.16E). Furthermore the microarray contains set of "complement probes" (derived from the target specific probe set, by substitution of bases by their complementary bases, e.g. 5'-TATAAGC-3' rather than 5'-ATATTCG-3'). Reversed probes and complement probes are not expected to show specific hybridization signals, rather these probes are employed to reveal non-specific cross-hybridization.

### 8.6.3 Microarray Hybridization - Experimental Procedures

#### 1×MES-hybridization buffer [Nuw02]

##### Reagents:

- MES (2-(N-morpholino) ethanesulfonic acid) 50 mM (zwitterionic buffering agent)
- NaCl 0.5 M
- EDTA 10 mM (chelating agent, employed to bind  $Mg^{2+}$  ions, thus to inhibit nuclease activity)
- Tween-20 0.005% (v/v) (nonionic surfactant, employed to block nonspecific surface adsorption)

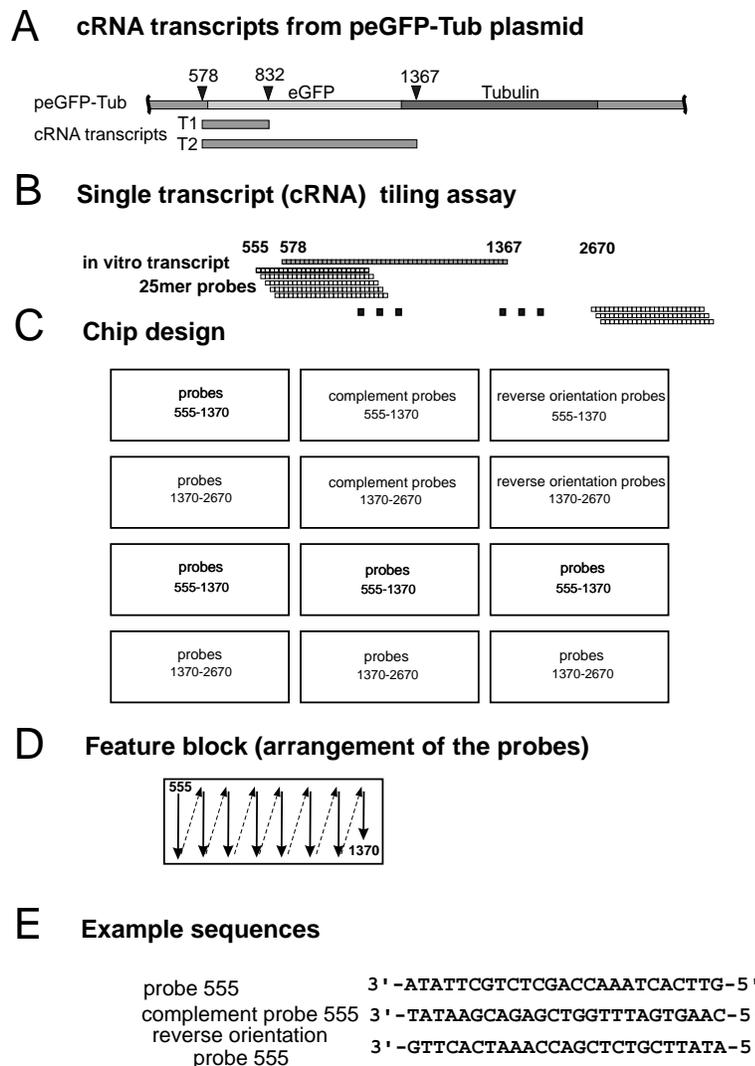
Hybridization buffers were produced with RNase-free MilliQ-water. RNase activity of the MilliQ-water and of the prepared hybridization buffers was tested negative with the Ambion RNaseAlert™ test kit.

Storage of the MES-buffer at 2°C to 8°C. Shield from light. Discard solution if yellow.

**Caution:** With the MES-buffer we experienced irreversible adsorption of fluorescently labeled targets if the temperatures were increased above 55°C (as in melting experiments). This problem (restricting reusability of the DNA microarrays) seems to be related to the use of the MES-buffer. Therefore, in later experiments the MES-buffer has been replaced by 5×SSPE (with 0.01% Tween-20) buffer.

#### Hybridization procedure

Microarray hybridization with the cRNA target (either T1 or T2) was performed in 1×MES-buffer at temperatures of 25°C and 37°C, respectively. The cRNA eluate (obtained from the QIAquick™ purification after IVT) is diluted 1:500 in the MES-buffer. Addition of a fluorescently labeled control-target (DNA oligonucleotide, ca. 100 pM) which specific to control features is useful to create a regular grid for orientation. To prevent evaporation, the



**Figure 8.16:** Tiling array experiment. (A) The fluorescently labeled cRNA transcripts T1 and T2 (employed as hybridization targets) have been produced from the *pEGFP-Tub* plasmid by PCR and in vitro transcription (IVT). T2 corresponds to the section 578-1367 of the *pEGFP-Tub* plasmid (and corresponds to the complete *EGFP*-coding sequence). T1 corresponds to the shorter section 578-832. (B) The tiling array includes a set of 25mer probes which forms a tiling path complementary to the cRNA target sequence. Probes are shifted along the complementary target sequence in increments of 1 (or 2) bases. Further probes that go beyond the actual target sequence range (of probes T1 and T2) (corresponding to *pEGFP*-positions 1370-2670) have also been included. These probes will be used for detection of non-specific cross-hybridization. (C) Arrangement of the feature blocks on the microarray - compare with the fluorescence micrograph of the hybridized microarray in Fig. 8.17. (D) Arrangement of the probes within the feature blocks. Apart from four replicate feature blocks the microarray contains one feature block of probes with reversed strand orientation - explained in (E) - and one with complemented probe sequences in which the bases have been substituted with their complementary bases.

hybridization solution was applied in thin film between the DNA chip and a cover glass, either for 10 minutes (at 25°C) or 2 h (at 37°C).

### Microarray washing procedure

In this experiment (unlike in the other experiments) the strong fluorescence of the cRNA hybridization solution doesn't allow real-time monitoring of the hybridization signal. To remove unbound targets and salt residue the microarrays have to undergo washing.

Unbound targets are washed off (under non-stringent conditions) with 20×SSPE-buffer (presumably 5×SSPE will equally work). Washing in an ethanol/water solution removes salt residue and prevents duplex dissociation.

Washing protocol:

- washing in 20× SSPE, 0.005% Tween-20 (with agitation) for 1 minute
- washing in 1:1 ethanol/water (with agitation) for 30 s
- washing in ethanol (analytical grade) for 30 s
- drying under a stream of nitrogen

Washing is performed at room temperature. Use of pure ethanol in the final washing step prevents droplet formation in the subsequent drying step. The ethanol dries very fast in a thin film and therefore (unlike water) doesn't leave concentrated spots of residues.

The washing procedure doesn't seem to affect the relative hybridization signal values. The hybridization signal after the washing procedure is identical to that observed while the microarray is still in the hybridization buffer.

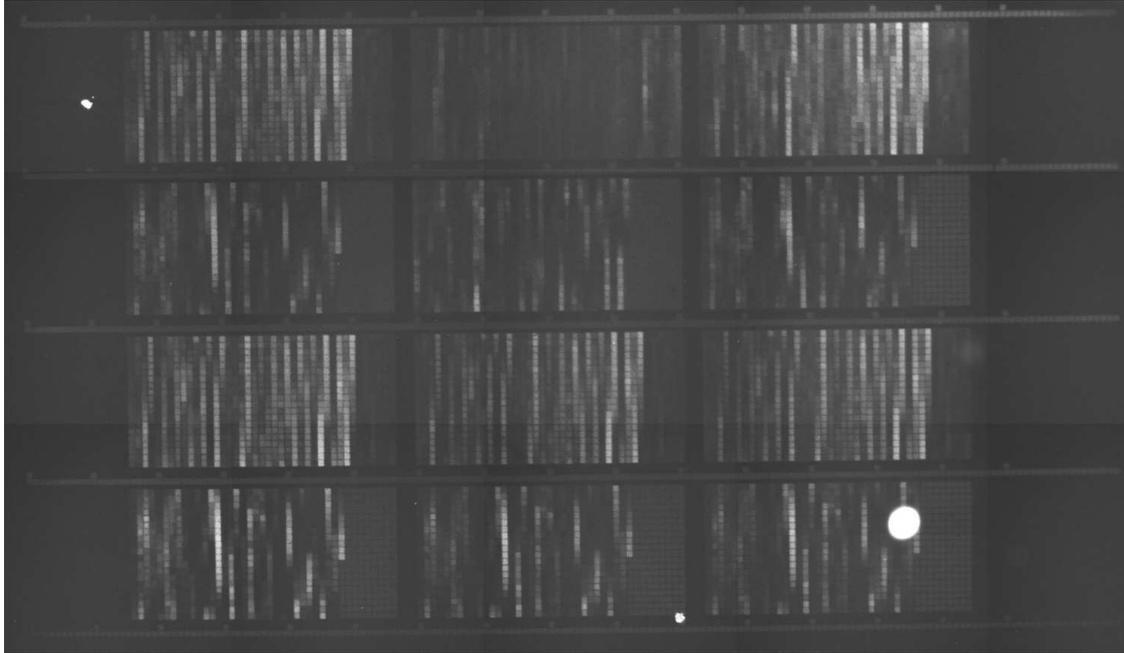
### Microarray stripping procedure

For reuse of DNA microarrays the targets from the previous hybridization experiment have to be removed. For removal of DNA targets heating in 5×SSPE buffer to a temperature of 60-70°C is applied. Often however, the fluorescence signal is not completely removed - in particular if the previous hybridization signal was very strong. It is unclear whether this irreversible hybridization signal is due to covalent binding of the targets.

In case the hybridization is performed with RNA targets there is a more efficient stripping method, making use of the fact that RNA degrades under alkaline conditions, whereas the microarray bound DNA-probes remain intact: Stripping in 100 mM NaOH solution at 45°C (5 minutes) removed fluorescence of cRNA targets completely, whereas traces of the fluorescently labeled control-target (DNA) were still visible. The hybridization capability of the microarray seems unaffected, however, the stripping procedure described above has not been tested extensively.

A similar (somewhat more gentle) method (RNA-degradation with 8 mM NaOH, 250 mM EDTA at 60°C to 62°C, 8 minutes) has recently been described by Hu *et al.* [Hu05].

#### 8.6.4 Results

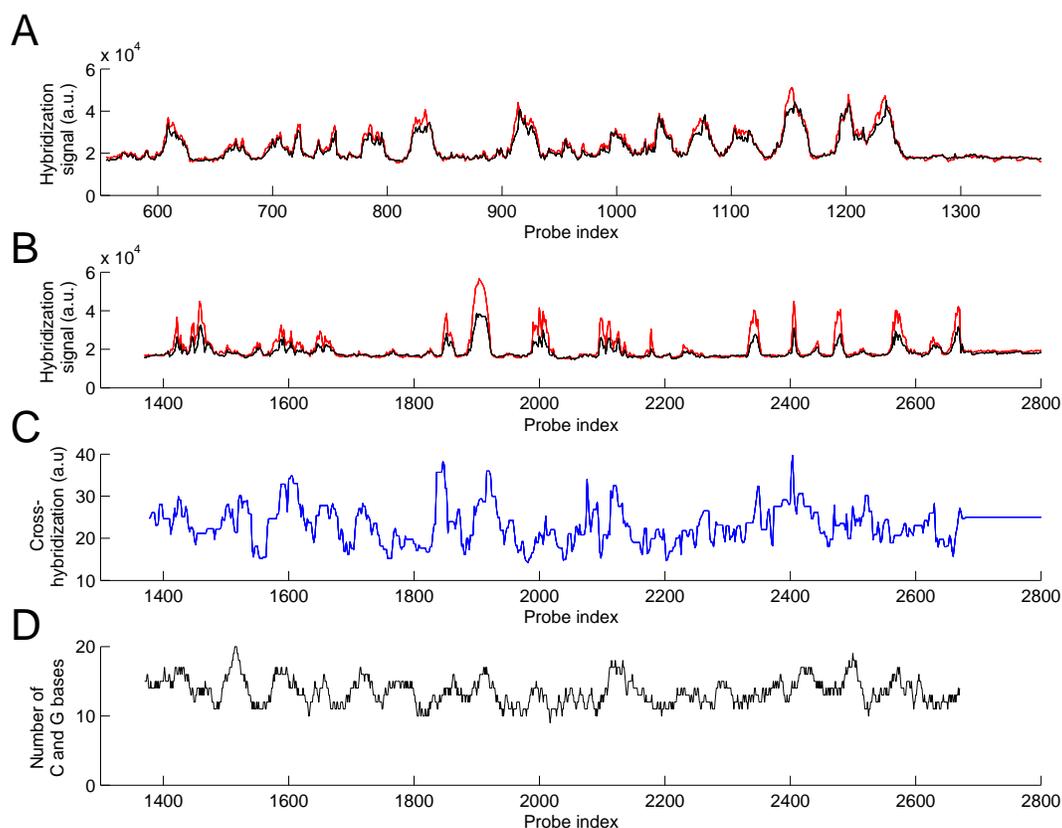


**Figure 8.17:** Image montage of fluorescence micrographs showing the complete DNA microarray (ca. 10000 features) after hybridization with cRNA target T2 (10 minutes at 25°C) and washing. Microarray features - size about 18  $\mu\text{m}$ , corresponding to  $5 \times 5$  DMD pixels - are separated by a 1 pixel gap. The chip design is described in Fig. 8.16C. Control features (horizontal lines between the features blocks and features at the right end of the individual feature blocks - have been hybridized with the fluorescently labeled control target (a short DNA oligonucleotide), and therefore show a weak hybridization signal. Unexpectedly (at 25°C) the reverse oriented probes (feature block top-right) displayed the brightest hybridization signals and the same intensity pattern as the regularly oriented probes.

To investigate the variation of the probe-target binding affinity along the cRNA target sequence we plotted the hybridization signals versus the probe index (number of the base in the *pEGFP-Tub* vector sequence which is pairing with the 3'-terminal base of the corresponding probe - see Fig. 8.16B).

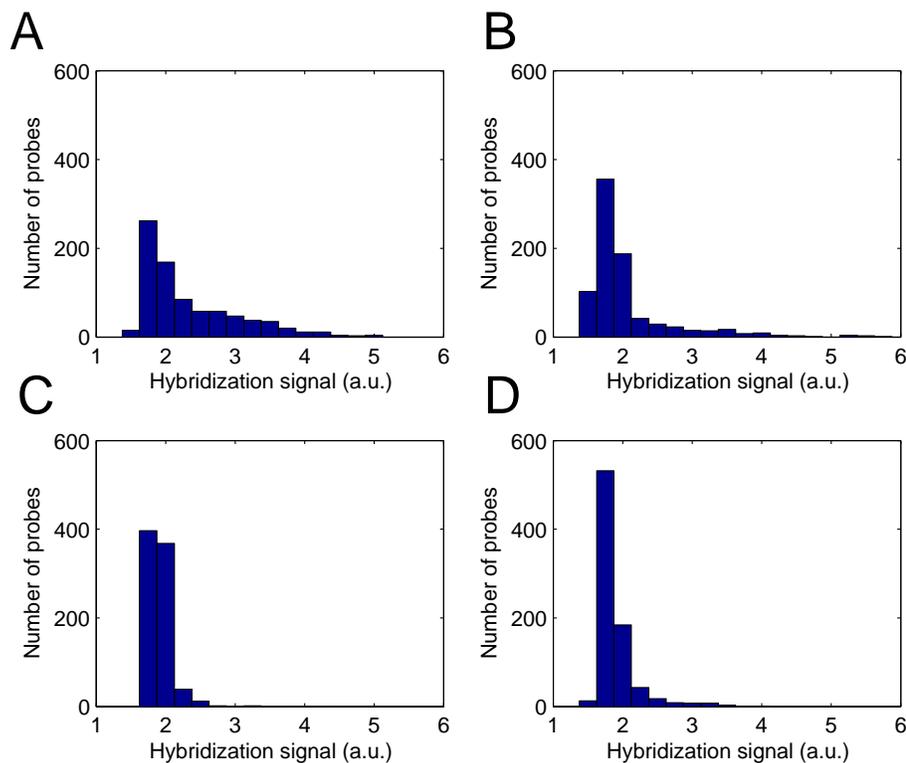
As shown in Figs. 8.17 and 8.18A probes are not equally sensitive for the cRNA target. At 25°C about one third of the target-specific probes has only a small (often negligible) sensitivity for the target sequence. Sections with sensitive probes (typically about 20 bases wide) are interrupted by sections with insensitive probes.

Figure 8.18B demonstrates significant cross-hybridization at 25°C. The probes targeting



**Figure 8.18:** Hybridization with the cRNA target T2 at 25°C. (A) and (B) show a significant variation of the hybridization signals. Hybridization signals from two replicate feature blocks (black and red curves correspond to identical probes at different positions on the same microarray - shown in Fig. 8.17) demonstrate the reproducibility of the analysis. Only probes with indices between 565 and 1355 are expected to specifically hybridize with T2. However, as shown in (B), we observe also strong hybridization signals from probes beyond base position 1355 (for which no specific target sequence is provided). (C) Attempt of a theoretical prediction of the cross-hybridization binding affinities between the probes 1370 to 2670 and the target T2: The affinity-measure employs a weighting factor proportional to the square root of the length of contiguous complementary subsequences and applies a 1.5 times higher weight on C-G base pairs than on A-T base pairs. We observe some correlation with the measured hybridization signals in (B). A significantly better correlation couldn't be expected as target secondary structure is not considered in the prediction. (D) Another affinity-measure, simply based on the number of C and G bases contained in the individual 25mer probe sequences, doesn't correlate with the hybridization signal in (B).

the section 1370 to 2670 of *EGFP-Tub* plasmid are not specific to the cRNA target, and thus are not expected to show a hybridization signal.

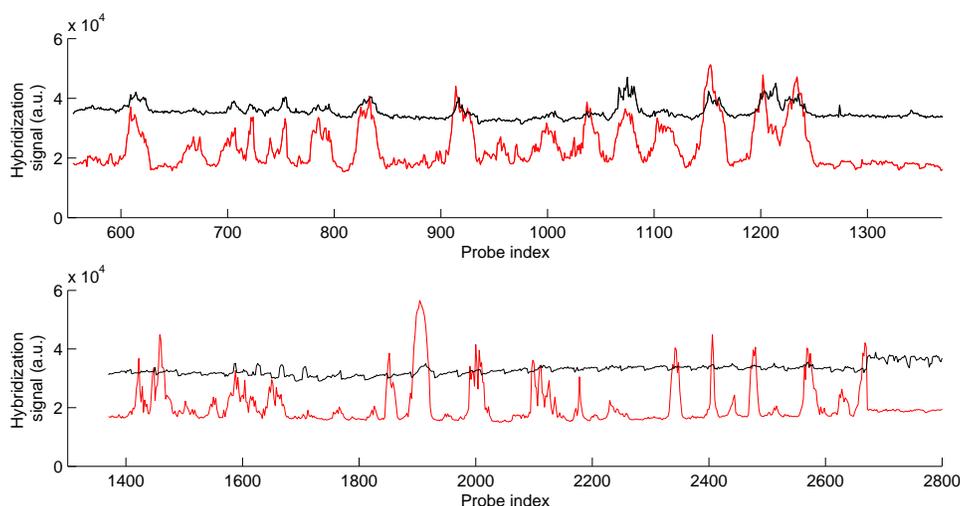


**Figure 8.19:** Histogram plots of the distribution of hybridization signal intensities (target T2) at 25°C. (A) target-specific probes (555-1370), (B) nonspecific probes (1370-2670), (C) nonspecific complement probes (555-1370), (D) nonspecific complement probes (555-1370). The largest fraction of sensitive probes is contained in the target-specific probe set (A). Compared with the nonspecific complement probes (C) and (D) the nonspecific probe set (B) includes a significantly increased fraction of bright probes.

The histogram plots in Fig. 8.19 demonstrate that in non-target-specific probe sets (as shown in Fig. 8.19 B to D) the fraction of probes with large hybridization signals is significantly reduced compared with the target-specific probes in Fig. 8.19A.

At an increased hybridization temperature ( $T = 37^\circ\text{C}$ ) the non-specific hybridization is negligible (Fig. 8.20), however, the hybridization signal of target-specific probes is significantly decreased as well. Only 20-30% of the target-specific probes provide a significant hybridization signal. This confirms results of Luebke *et al.* [Lue03] showing that only 10% of the probes (specific for the cRNA target coding GFP) afforded a signal  $>5\%$  of the highest signal.

Unexpectedly, we found that the reversed orientation probes (Fig. 8.21) at a hybridization temperature of 25°C show the same hybridization pattern - even at slightly increased intensities - like the genuine target-specific probes. However, different from the target-specific



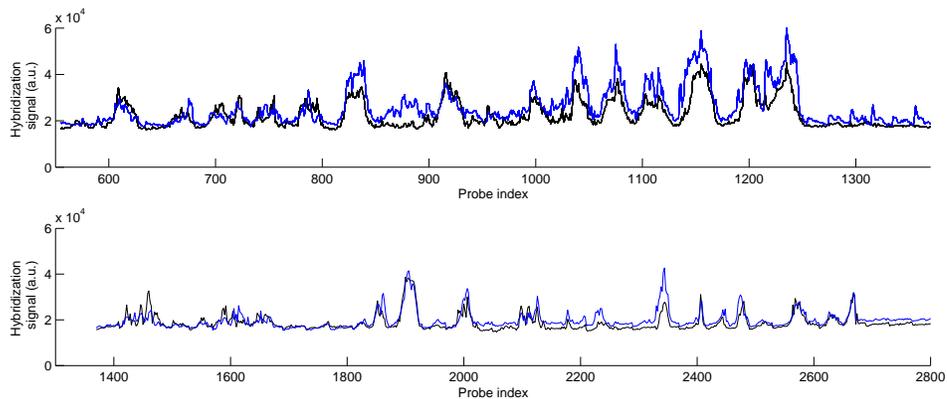
**Figure 8.20:** Hybridization signal of the cRNA target T2 on target-specific probes (upper graph) and nonspecific probes (lower graph) at 37°C (black curve - the corresponding hybridization signal for 25°C is shown in red). The target-specific hybridization is significantly reduced at 37°C. Nonspecific hybridization has almost completely disappeared at 37°C. The ripple on curve is owing to imaging artifacts.

probes, the hybridization signal of the reversed probes has almost completely disappeared at more stringent hybridization conditions at a temperature of 37°C. The presumed, largely specific binding of two parallel-oriented strands raises the interesting question of the underlying duplex structure and binding mechanisms and should therefore be subject to further investigation.

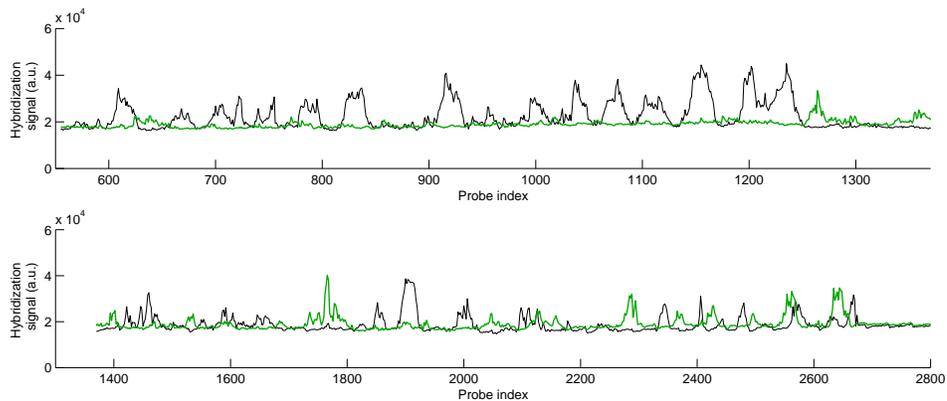
The hybridization signal of the complement probes (derived from the set of target-specific probe sequences by substituting bases with their complementary bases) is not correlated with the hybridization signal of target-specific probes (Fig. 8.22). Several distinct peaks, owing to nonspecific cross-hybridization at 25°C, have disappeared at an increased hybridization temperature of 37°C.

Hybridization of the short cRNA target T1 (at 25°C) afforded the anticipated result (see Fig. 8.23): Hybridization is largely limited to probes covering section 600-720. No significant hybridization signal is observed in the section 720-1370 which is included in the longer target T2, not however in T1.

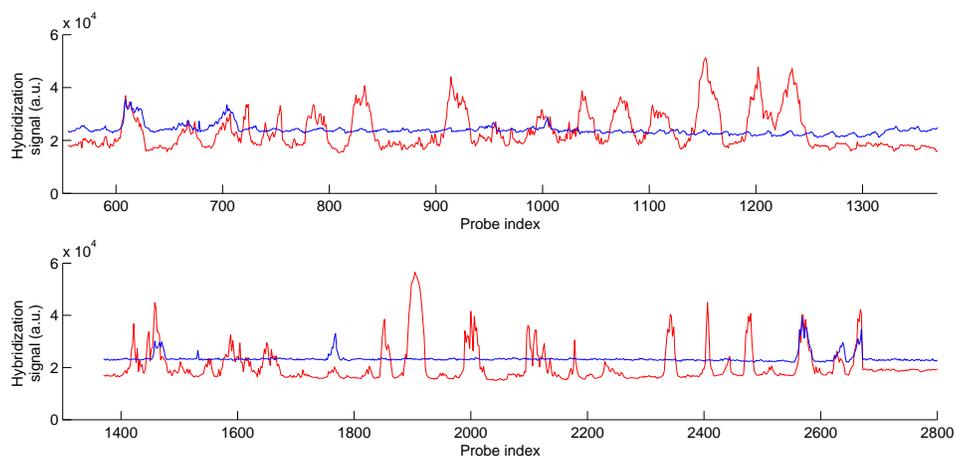
The experiments have been repeated using the same chip design, however with an improved control of the hybridization conditions: Hybridization was performed in a temperature-controlled hybridization chamber enabling temperature control, real-time monitoring and buffer exchange. Hybridization with cRNA target T2 in 5×SSPE buffer (with 0.01% Tween-20) at a temperature of 25°C (and 37°C, respectively, overnight) followed by washing with pure hybridization buffer (inside the hybridization chamber), provided



**Figure 8.21:** Hybridization signal of the cRNA target T2 on reverse orientation probes at a hybridization temperature of 25°C. Reverse orientation probes (blue curve) and the corresponding correctly oriented probes (black curve), rather unexpectedly, agree very well.



**Figure 8.22:** Hybridization signal of the cRNA target T2 on complement probes at a hybridization temperature of 25°C. The hybridization signals of complement probes (green curve) and the corresponding EGFP-Tub specific probes (black curve) are not correlated.



**Figure 8.23:** Hybridization of the short cRNA target T1 (section 578-832) at 25°C. Hybridization signals from T1 (blue curves) shows peaks mainly below probe index 720. These are congruent with peaks from target T2 (red curves). In the section 1370-2670 several distinct peaks, congruent with peaks from target T2 (red curves), originate from nonspecific binding of the probes with target T1. In comparison with the longer target T2 fewer nonspecific peaks (lower graph) are observed with the shorter T1.

basically the same hybridization signals (see Fig. 8.30) as described above.

Interestingly, a gradual temperature increase from 37°C to 65°C results in a gradual decrease of the hybridization signals of all probes (partly owing to the temperature-dependence of the Cy3-fluorescence) rather than in an improved discrimination between specific and nonspecifically bound targets. In contrast to that, we observed a significantly improved discrimination at a hybridization temperature of 37°C compared to a weaker discrimination at hybridization temperature of 25°C. This suggests that the discrimination between specific and nonspecific hybridization should be controlled by the stringency of the hybridization conditions, rather than by stringent treatment after hybridization.

### 8.6.5 Consideration of the Target Accessibility

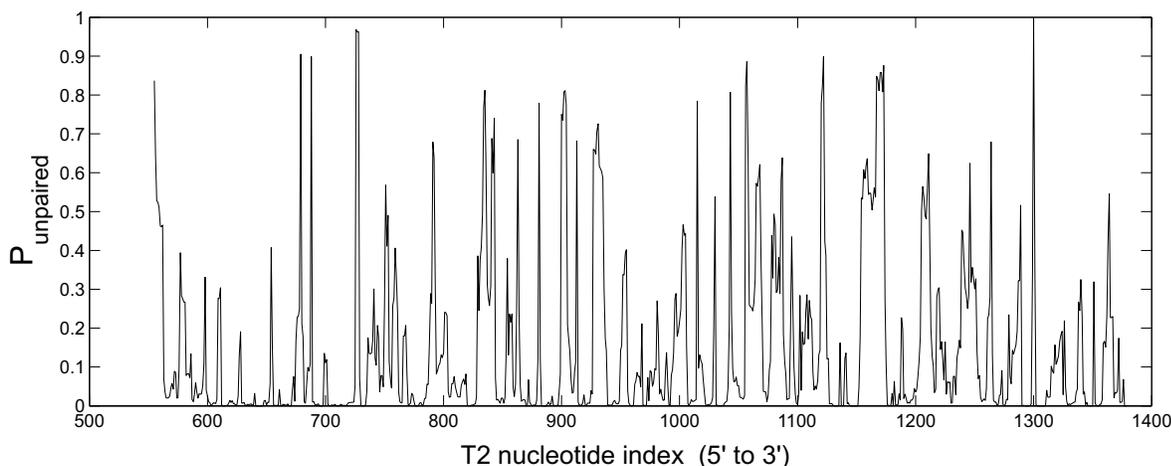
The stable secondary structures of long nucleic acid targets interfere with microarray hybridization. As shown in Fig. B.17 large parts of the target sequence T2 (owing to intramolecular base pairing) fold into stable hairpin structures. Ratushna *et al.* [Rat05] suggested that target secondary structure should be considered in microarray design and in the interpretation of microarray results.

The limited accessibility of the target sequence for nucleic acid probes also affects the efficiency of antisense oligonucleotides [Vic00] and RNA interference (RNAi) [KK03; Wes07]. Software tools for prediction of antisense and RNAi efficiencies consider tar-

get accessibility.

In the Sfold software [Din01; Din03; Din04] (web server at <http://sfold.wadsworth.org>) target accessibility is determined from the Boltzmann ensemble of RNA target secondary structures. Sfold determines the antisense oligonucleotide binding energy as a weighted sum of RNA/DNA *nearest-neighbor* interactions [Sug95]. The weighting factor for each *nearest-neighbor* stacking interaction is calculated from the intramolecular base pairing probability of the corresponding target dinucleotide [Din04].

We compared experimental hybridization efficiencies (from the tiling array experiment in section 8.6.4) to base pairing probabilities (Fig. 8.24) and binding energies (Fig. 8.25) calculated with Sfold.

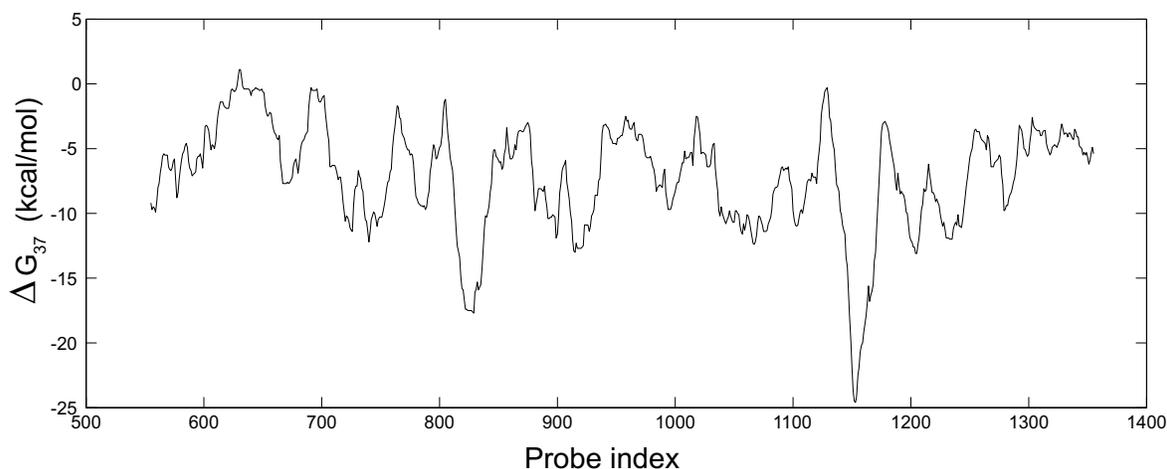


**Figure 8.24:** Prediction of the target accessibility with Sfold: Probability that nucleotides  $i$ ,  $i+1$ ,  $i+2$ , and  $i+3$  are all unpaired. Single stranded regions are more likely to be accessible for hybridization with microarray probes. Increased probability for single stranded section at base positions 785 to 800, 825 to 850, 900 to 930, 1050 to 1090, 1150 to 1170, and 1200 to 1250 is correlated to increased hybridization signals in the tiling array experiment (see Fig. 8.20 top).

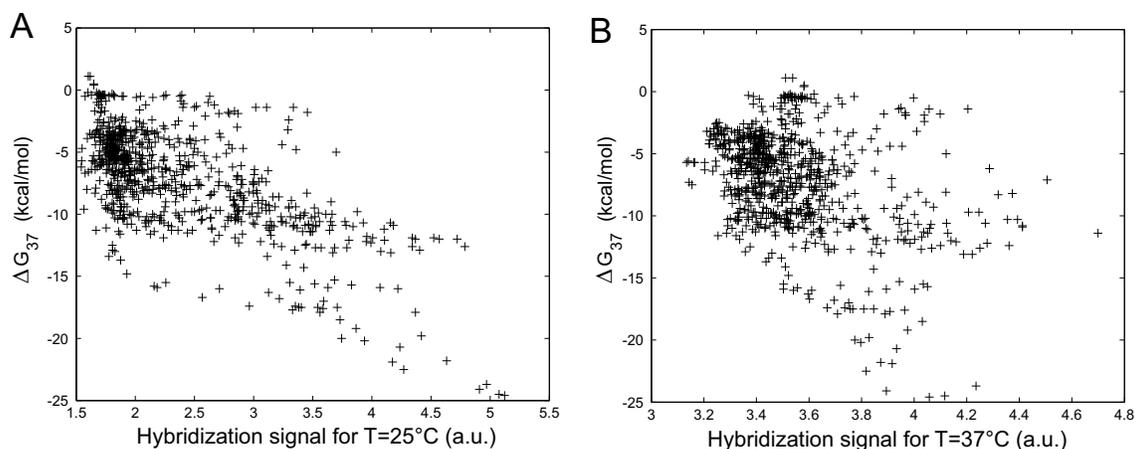
Fig. 8.24 shows the probability that the individual nucleotides along the cRNA sequence T2<sup>5</sup> are unpaired and thus accessible for hybridization. The sequence comprises stable intramolecular base-paired segments which are interrupted by short (mostly) unpaired segments. Unpaired segments frequently correspond to segments with large binding affinities in the tiling array experiment in Fig. 8.20 (top).

Segments with large (negative) binding energies in Fig. 8.25 correspond well to segments of large hybridization signals in the tiling array experiment in Fig. 8.20. The scatter-plots in Fig. 8.26 show the correlation between microarray hybridization efficiencies and Sfold binding energies.

<sup>5</sup> Apart from the end regions the sequence investigated is identical to the cRNA target T2 employed in the tiling array experiment (see section 8.6.1)



**Figure 8.25:** Prediction of antisense binding energies  $\Delta G_{37}$  with Sfold (parameters: 25mer DNA antisense probes,  $T=37^{\circ}\text{C}$ , 1 M [NaCl]). We observe a good correlation between peaks in the hybridization signal (Fig. 8.20 top) and high probe-target binding affinities (low  $\Delta G_{37}$ ).

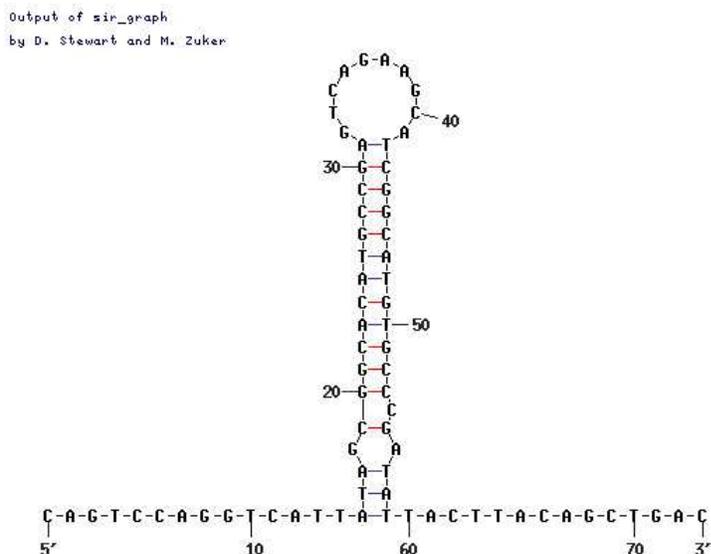


**Figure 8.26:** Plot of the Sfold antisense binding free energies  $\Delta G_{37}$  versus the hybridization signals intensities from the tiling array experiments (as in Fig. 8.20 top) at hybridization temperatures of (A)  $25^{\circ}\text{C}$  and (B)  $37^{\circ}\text{C}$ . Larger (i.e. more negative) Sfold binding free energies correspond to increased hybridization signals. At less stringent hybridization conditions, for  $T=25^{\circ}\text{C}$  in (A), we observe a better correlation than at  $37^{\circ}\text{C}$ . The branch-structures are associated associated to different peaks in Fig. 8.25 and thus to different accessible target regions. The slopes of the individual branches differ significantly.

We have demonstrated that the target accessibility governs microarray hybridization efficiency. The comparison between our experimental results and predicted binding affinities (binding efficiency predicted by the Sfold web server) shows that software tools for the prediction of antisense oligonucleotide efficiency – on the basis of target accessibility – can be employed for the design of efficient (i.e. sensitive) microarray probe sequences.

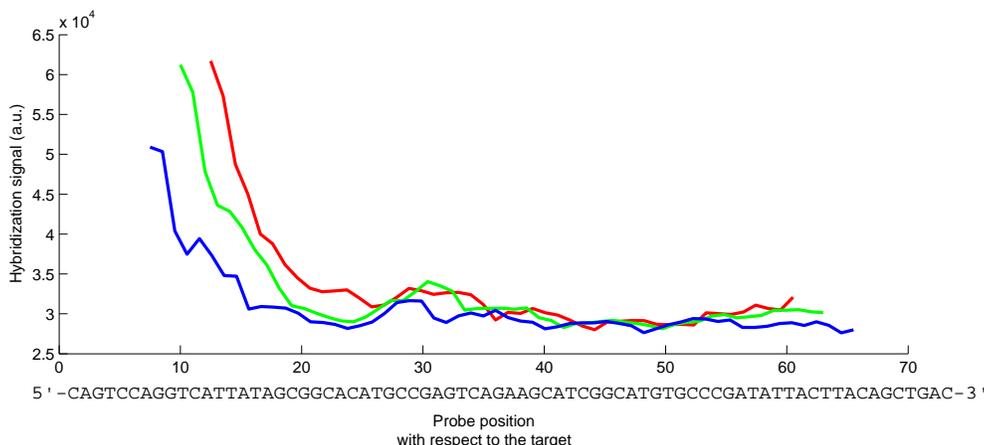
### 8.6.6 Investigation of the Influence of a well-known Target Secondary Structure on the Hybridization Efficiency

To further investigate the influence of target intramolecular base pairing we used a relatively simple DNA oligonucleotide target ('stem-loop-target' SLT) which forms the stable stem-loop secondary structure shown in Fig. 8.27.

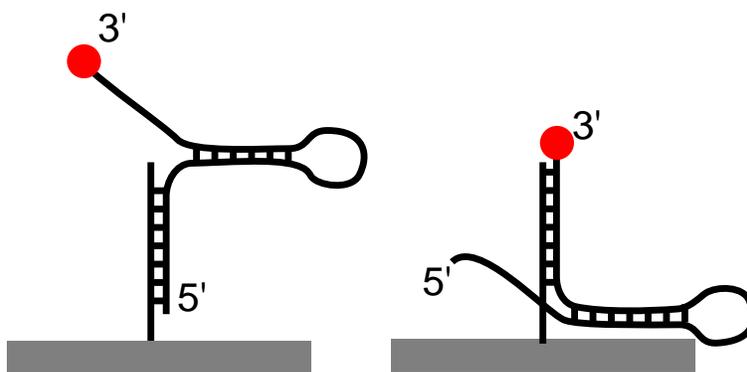


**Figure 8.27:** The 73mer DNA oligonucleotide target sequence SLT was designed to form a relatively simple stem-loop structure. The stem is formed between bases 19-30/34-55. Hybridization with microarray-bound probes is expected to be possible in single stranded regions at the duplex ends (between bases 1-18 and 56-73) and in the single stranded loop region (31-42). A fluorescent marker (Cy3) is attached at the 3'-end. The secondary structure was predicted by the DINAMelt server [Mar05].

We hybridized SLT at a concentration of 1 nM on a tiling array similar as in the above experiments. Hybridization was performed in  $5\times$ SSPE buffer (containing 0.1% SDS) for 90 minutes at a temperature of  $40^{\circ}\text{C}$ . A relatively strong hybridization signal is observed for probes binding to the single stranded region at the target 5'-end (see Fig. 8.28). Probes specific to the stem region - ranging from base positions 19 to 30 - (in particular the shorter 15 and 20mer probes) have significantly reduced hybridization signals (near



**Figure 8.28:** Hybridization signals from tiling array specific to the stem-loop target SLT. The hybridization signal is plotted versus the probe sequence position (center of the probe sequence) with respect to the target sequence (shown below the horizontal axis). Different probe lengths 25mers (red), 20mers (green) and 15mers (blue), have different sensitivities but show the same characteristics. The background intensity level is at 2.8 a.u..



**Figure 8.29:** The large difference between binding affinities at the target 5'-end and at the target 3'-end shown in Fig. 8.28 can be explained by steric hindrance of the rigid double-helical stem structure: Probe targeting the 5'-end of the target sequence (left): The rigid stem-structure can relatively freely move above the probes. Probe targeting the 3'-end of the target sequence (right): The rigid stem-structure is forced onto the microarray surface, leading to increased steric hindrance with respect to the surface and neighboring probes and hybridized duplexes. The position of the fluorophore (red dot) is not expected to play a role.

the background level). Increased hybridization signals are observed for probes binding to the single stranded loop region (31-42). However probes (with their centers) between positions 26 and 35 are significantly more sensitive than those between (35 and 40). In the stem region between positions 42 and 55 the hybridization signal is at the background level. At the single stranded 3'-terminal region a small binding affinity is observed, which is much smaller (i.e. at a level of about 10%) compared to the binding affinity observed at the 5'-end. A possible explanation is that sterical hindrance of the rigid double-stranded stem-structure<sup>6</sup> prevents hybridization at the targets 3'-end (Fig. 8.29).

### 8.6.7 Nonspecific Hybridization – Variation of the Wash Stringency

To investigate the effect of washing on binding specificity we performed a series of washing steps with washing buffers of varying ionic strength.

The experimental design is basically the same as above (section 8.6.4). However, the experiment was performed on another microarray (though with the same chip design). The cRNA targets (T2) have been stored for about one year at -80°C.

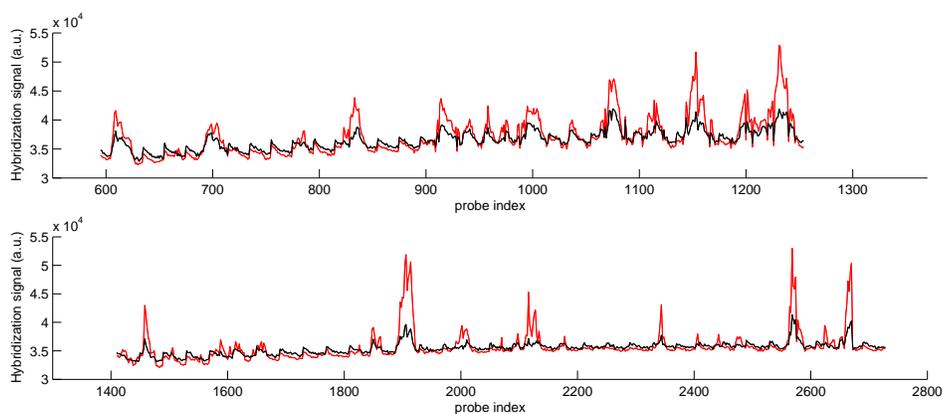
Following overnight hybridization at 23°C (hybridization buffer: 5×SSPE, 0.01% Tween 20) the microarray was washed for 15 minutes (in the flow-through hybridization chamber) with 5×SSPE, 5×SSPE (again), 2×SSPE and 0.5×SSPE. After each wash, the hybridization signal was acquired.

Fig. 8.30 shows the hybridization signal after the first (nonstringent) wash (with 5×SSPE - red curve) and after the last wash (black curve) with the stringent 0.5×SSPE buffer. The hybridization signal is very similar to that in Fig. 8.18, thus demonstrating the good reproducibility of the experiment. Washing under stringent conditions does not increase the discrimination between specifically and nonspecifically bound targets significantly. This is in agreement with Pozhitkov *et al.* [Poz07] who recently reported that they didn't find a statistical difference between the dissociation kinetics of PM and double-mismatch duplexes. A key finding of their study is that nonspecific duplexes do not always dissociate before specific ones.

In Fig. 8.18 we observed significantly improved discrimination for the hybridization performed at 37°C with respect to the hybridization performed at 25°C. Thus, to prevent nonspecific cross hybridization, it appears more effective to apply stringent hybridization conditions rather than stringent washing conditions. However, more experiments need to be done to confirm that washing with increased stringency does not reduce nonspecific

---

<sup>6</sup> Compared to single stranded sections the double stranded stem-region has a strongly increased rigidity (persistence length  $l_p \simeq 45\text{-}50$  nm [Hag88]).



**Figure 8.30:** Influence of the washing stringency on specifically and nonspecifically bound duplexes. Like in Fig. 8.18 the upper graph corresponds to target-specific probes, whereas the lower graph corresponds to nonspecific probes. Washing was performed at 23°C with 5×SSPE (15 min.) (red curve), followed by another wash in 5×SSPE (15 min.), 2×SSPE (15 min.) and 0.5×SSPE (15 min.). Hybridization signals were acquired after every washing step. The hybridization signal acquired after the final washing step is shown in black. The saw tooth patterns at intervals of 20 and 40 steps, respectively, are artifacts owing to brightness gradients in the fluorescence excitation.

cross hybridization.



# Chapter 9

## Summary/Zusammenfassung

### 9.1 Summary

The present thesis describes the development and application of a *DNA microarray synthesizer*. Moreover, in an extensive study based on microarray hybridization experiments, we investigated the impact of single base defects (*single base mismatches* and *single base bulges*) on the binding affinity of oligonucleotide duplexes. We were particularly interested in the influence of defect type and defect position.

The thesis comprises the following topics:

- development of an automated DNA microarray synthesis system on the basis of light-directed *in situ* synthesis.
- quantitative analysis of microarray hybridization signals based on fluorescence microscopy and the use of a sensitive EM-CCD camera
- characterization of the DNA microarrays that have been fabricated with the microarray synthesizer
- detailed investigation of the influence of single base defects (*single base mismatches* and *single base bulges*) on duplex binding affinity
- modeling of oligonucleotide duplex stability and the impact of single base defects on the basis of the double-ended zipper model

In the first half of the thesis, the emphasis is on development, optimization and operation of the *DNA microarray synthesizer*. The automated synthesis-apparatus, which is based on a light-directed (i.e. photolithographically controlled) *in situ* synthesis process [Fod91; SG99], is employed in the fabrication of high-density oligonucleotide microarrays. It provides the basis for the microarray hybridization experiments performed in the second half of the thesis.

Light-directed *in situ* synthesis enables superior flexibility in the choice of microarray probe sequences: unlike with the widely used contact-printing techniques, presynthesized oligonucleotides are not required. With the setup presented in this work, parallel synthesis of 25000 probes-sequences is performed on the surface of a dendrimer-functionalized microarray substrate. Phosphorus (cyclotriphosphazene-PMMH) dendrimer substrates – previously used in the immobilization of presynthesized oligonucleotide probes [LB03] – were employed for the first time, and with great success, as a substrate for light-directed *in situ* synthesis.

In the course of the photolithographically controlled synthesis process photolabile chemical protection groups<sup>1</sup> (5'-nitrophenylpropyloxycarbonyl [Has97]) are cleaved by UV exposure ( $\lambda \simeq 370$  nm), thus to enable coupling of the subsequently provided phosphoramidite building blocks. The coupling reaction extends the growing probe sequences by one nucleotide, which – to prevent uncontrolled coupling – introduces a new 5'-NPPOC protection group.

Spatially controlled photo-deprotection is achieved by the projection of "dynamic photomasks": thereby the image of a micromirror-array-type spatial light modulator<sup>2</sup> (Digital Micromirror Device, DMD<sup>TM</sup>, Texas Instruments Inc.) is projected onto the surface of the microarray substrate. The sequence information of the microarray probe sequences is encoded in the set of photomasks. Photo-deprotection is restricted to those microarray features comprising sequences to which the subsequently provided nucleotide building block (alternately A, C, G or T) is to be attached in the following coupling reaction.

A microscope-projection-photolithography setup has been developed for projection of the photomask patterns onto the substrate. In the setup the spatial light modulator is mounted in the intermediate image plane of a Zeiss Axiovert 135 inverted microscope. The microscope-based design provides superior stability with respect to thermal expansion. Moreover, image reduction reduces the requirements on UV lamp power. A commercially available 5×0.25 NA Fluar microscope objective (Carl Zeiss) enables high transmission and satisfying imaging quality in the near UV around 370 nm. The imaging quality was evaluated and optimized by exposure of a UV-sensitive film. For this purpose a novel UV-sensitive coating, based on the photochromic dye spiropyran has been developed. The technical problem of finding the exact focus for the UV photomask image has been solved with a workaround: fine adjustment of the distance between the DMD and the tube lens enables focusing of the UV photomask image by means of visual focusing on a test pattern (under

---

<sup>1</sup> The 5'-NPPOC protection groups [Has97] are attached to the 5'-end of the 3'-tethered probe sequences.

<sup>2</sup> The DMD<sup>TM</sup> was obtained from a commercial video projector.

green illumination) – despite large uncorrected chromatic aberration.

The fluidics system of the microarray-synthesizer was developed on the basis of a commercial DNA-Synthesizer. Solenoid valves are controlled via a microcontroller-based valve driver, which is connected to the control PC. The microarray synthesizer control-software *DNASyn* (written in Java™) coordinates the projection of the photomasks with the control of the fluidics system and thus enables a fully automated synthesis process.

Microarray synthesis is performed under anhydrous conditions in a hermetically sealed synthesis cell. The design of the synthesis cell was technically demanding since a number of technical requirements had to be considered: sufficient chemical resistance (use of aggressive synthesis reagents, e.g. tetrahydrofurane and acetonitrile), high quality image projection (with UV light) into the interior of the synthesis cell, prevention of argon gas bubbles in the synthesis area, and avoidance of dead volume (to enable fast and complete reagent exchange). The optical flow cell design presented in this work reliably meets the above requirements.

The highly flexible DNA microarray synthesizer – which has been developed from widely available components – has demonstrated reliable operation and is now routinely employed for lab-scale fabrication of tailormade DNA microarrays.

After the fabrication process, microarray hybridization is performed in a temperature-controlled optical flow cell. In most of the hybridization experiments we used fluorescently labeled DNA and RNA oligonucleotides as target sequences<sup>3</sup>. The fluorescent signal of hybridized target molecules was detected with a fluorescence microscope in conjunction with a sensitive EM-CCD camera. Unlike commercial microarray-scanners the microscope-based configuration can be employed in realtime-monitoring of the hybridization process. Readout of the hybridization signal intensity of the individual microarray features is performed with an image processing software which has been developed in the framework of the present thesis.

We investigated the impact of *in situ* synthesis-related point defects (single base mismatches, deletions and insertions) on DNA microarray hybridization: our experimental results show that with an increasing number of deliberately induced deletion-defects the binding affinity (in the median) is reduced. The randomly introduced point defects<sup>4</sup> give rise to a heterogeneous distribution of reduced binding affinities. Moreover, the duplex melting temperature (in the median) is reduced. Similar experiments in section 6.11 demonstrate

---

<sup>3</sup> In the hybridization experiments fluorescently labeled target sequences are applied to the microarray surface in solution. On the surface the diffusing targets can get captured by their complementary probe counterparts. Surface-tethered probes bind (hybridize) to complementary target sequences via base-pairing interactions.

<sup>4</sup> Various amounts of random deletions were created by variation of the photodeprotection time.

that - depending on the distribution of the defects - even probes with multiple defects can significantly contribute to the microarray hybridization signal.

We suspected that observed characteristics of microarray hybridization – the reduced melting temperature (with respect to solution-phase hybridization), and the broad melting transition (we observed almost gradual melting between 30 and 70°C) – originate from random synthesis defects and the associated heterogeneity of binding affinities. This connection has been confirmed by numerical simulations in section 6.11.

Mismatch discrimination<sup>5</sup> is fundamental to number of important genomics applications, such as single nucleotide polymorphism (SNP) analysis and resequencing<sup>6</sup> of gene sequences. In solution-phase-hybridization the influence of single base mismatches has been studied extensively by means of UV-absorption measurements of the duplex melting transition [All97; Sug00; San04].

Recent studies focusing on MM discrimination on DNA microarrays [Wic06; Poz06] show differences with respect to the previous UV absorption studies in solution: in accordance with our results in [Nai06b], [Wic06] and [Poz06] report a dominating influence of defect position. However, between the studies of Wick *et al.* [Wic06] (employing DNA/DNA hybridization) and Pozhitkov *et al.* [Poz06] (RNA/DNA hybridization) there is little agreement in respect to the relative impact of different mismatch base pairs.

In the present work we performed a detailed investigation on the influence of single base defects (*single base mismatches* and *single base bulges*) on oligonucleotide duplex binding affinities (chapter 6). Microarray hybridization experiments were performed with extensive sets of microarray probes containing – deliberately introduced – single base defects with respect to a perfect matching 'probe sequence motif'.<sup>7</sup> Within the probe sets defect type and defect position were varied systematically.<sup>8</sup> Thus, the microarray hybridization signals (fluorescence of hybridized target molecules) provide comprehensive information on duplex binding affinities in relation to defect type, defect position and the composition of the 'probe sequence motif'.

---

<sup>5</sup> Mismatch discrimination: binding affinities of mismatched (MM) duplexes are reduced with respect to the binding affinities of the corresponding perfect matching (PM) duplexes. The discrimination between MM- and corresponding PM-hybridization signals is employed for the detection of point-mutations.

<sup>6</sup> Resequencing is employed to search for single base mutations with respect to a well-known references sequence.

<sup>7</sup> The 'probe sequence motif', from which the individual "point-mutated" probe sequences were derived, matches perfectly with the synthetic oligonucleotide target sequence provided in the hybridization solution. A variety of different 'probe sequence motifs' and corresponding oligonucleotide targets has been employed throughout this study.

<sup>8</sup> Microarray probe sequences were designed to have single base "defects" – i.e mismatches, deletions and insertions, with respect to the perfectly matching 'probe sequence motif'.

An important result of our microarray hybridization experiments is that the mismatch discrimination is governed mainly by the position of the defect (section 6.5): mismatches in the center of the duplex show a more distinct discrimination than comparable defects which are located closer to the duplex ends [Nai06b]. A similar strong influence of the position of MM defects has also been reported in [Wic06] and [Poz06]. In studies on solution-phase hybridization (except for [Kie99] and [Dor03]) an influence of defect position has not been reported. The well-established two-state nearest-neighbor model of oligonucleotide duplex stability – commonly employed to predict duplex stabilities for solution-phase hybridization [All97; San04] – neglects a positional influence, except for terminal base pairs [San04].

We discovered that the influence of the defect position is not restricted to *single base mismatches* but can also be observed for *single base bulge* defects: for the individual 'probe sequence motifs' the position dependence of *single bulge* defects is (aside from defect type-related variations) identical with that of *single base mismatch* defects. Moreover, our results show that the positional influence in the individual 'probe sequence motifs' is not merely determined by the defect-to-end distance but also by the probe sequence.<sup>9</sup> This is presumably related to the extend of end-domain opening (partial denaturation), which is determined by the stability of the nearest-neighbor pairs composing the duplex section between the defect and the proximate duplex end. Our observations suggest that the position-dependent influence of single base defects (independent of the defect type) can be described on the basis of a molecular zipper model.

Systematic variation of the mismatch base pairs (within the probe sets of our microarray experiments) enabled a comprehensive statistical analysis of MM binding affinities (section 6.6). We observed that (in case of DNA/DNA hybridization) the mismatch discrimination is largest for those MM types where the MM base pair is compared to a C-G base pair in the corresponding PM duplex. From our database of about 1000 mismatch binding affinities we determined defect type related contributions to MM discrimination: the established ranking order of MM stabilities<sup>10</sup> is in good agreement with [Wic06]. Moreover, as also reported in [Wic06], the mismatch discrimination of individual MM-types is related to predicted<sup>11</sup> free energy increments  $\delta\Delta G$ . Our analysis shows that MM type dependent contributions to MM discrimination  $\delta I$  – the position-dependent influence has been eliminated – are proportional to  $\delta\Delta G$ . However, the observed large discriminations for the MM

<sup>9</sup> In the present context the 'sequence' is to be regarded as a sequence of weaker and stronger nearest-neighbor pairs.

<sup>10</sup> Since the binding affinities have been normalized with respect to the PM binding affinities, they represent a measure for the mismatch discrimination.

<sup>11</sup> The differences between MM- and PM-duplex Gibbs free energies  $\delta\Delta G$  – a measure for MM discrimination – were predicted from the two-state nearest-neighbor model, thereby using mismatch nearest neighbor free energy parameters from [All97].

types G·G, A·A and T·G significantly deviate from an otherwise monotonic (and approximately linear) relation.

Pozhitkov *et al.* [Poz06] established a ranking order of MM stabilities for RNA/DNA duplexes, which indicates that purine-purine mismatches (i.e. A·G, G·A, A·A and G·G) are among the least stable MM types. However, the correlation with the DNA/DNA stability order established in the present work is relatively small. Suspected differences between DNA/DNA and DNA/RNA mismatch discrimination motivated further hybridization experiments, in which a 'direct comparison' between DNA/DNA and corresponding RNA/DNA binding affinities<sup>12</sup> was performed (section 6.8). We discovered systematic differences with respect to MM discrimination in DNA/DNA and RNA/DNA hybridization: purine-purine mismatches are more discriminative (with respect to other MM types) in case of RNA/DNA hybridization than in case of DNA/DNA hybridization. However, the differences with respect to DNA/DNA hybridization are only moderate – the stability order reported in [Poz06], with purine-purine MMs as the most destabilizing MMs, could not be reproduced in our experiments. Further experiments to resolve the discrepancy, possibly including RNA/RNA hybridization, could provide valuable insights.

We suggest that different helix structures – B-form helix in case of DNA/DNA duplexes and A-form helix in case of RNA/DNA duplexes – are responsible for the observed differences in MM-discrimination. In A-form helices the steric clash between double-ringed purine bases in purine-purine MM-pairs may contribute stronger to destabilization of the mismatched duplex than in B-form DNA/DNA helices.

*Single base bulge* defects are caused by an unpaired base in one of the nucleic acid strands. In our experiments in section 6.9 single bulge defects were created from the PM probe sequence by insertion or deletion of a single base. Like *single base mismatches*, *single bulge* defects can result in a considerable decrease of the probe-target binding affinity. The identity of the neighboring bases determines how significant the binding affinity is affected by the bulge defect: our experimental results show that the binding affinity is significantly less affected if the unpaired (bulged) base is directly adjacent to an identical base. In the presence of such an identical base the position of the physical bulge is ambiguous (*positional degenerate* [Ke95]). According to Wartell and coworkers, who previously observed the stabilization of such *Group II* bulges<sup>13</sup> in solution-phase hybridization exper-

---

<sup>12</sup> A 'direct comparison' was achieved by subsequent hybridizations with RNA and corresponding DNA oligonucleotide targets. These had equivalent sequences – only thymine is replaced by uracil in RNA.

<sup>13</sup> In the notation of [Zhu99] there are two types of *single base bulges*: In *Group I* bulges there is no identical base next to the bulged base, whereas in *Group II* bulge there is at least one identical neighboring base.

iments [Ke95], the *positional degeneracy* entails an increase in entropy and thus results in increased duplex stability [Zhu99]. The unexpectedly large stabilization of *Group II bulges* observed in our microarray experiments cannot be explained solely by this entropy increase, since the resulting stabilization of *Group II bulges* with respect to *Group I bulges*  $\delta\Delta G_{bulge}^{37} \simeq -0.4$  kcal/mol is too small to explain the large differences in hybridization signal intensities. Our explanatory approach is based on the idea that the bulge defect causes a blockage of the zipper mechanism: the surplus nucleotide provokes a 1-nt frameshift between the single stranded sequences and thus prevents the rapid 'zipping up' of the duplex. In case of a *Group II bulge* - owing to the *positional degeneracy* - the barrier can be overcome significantly faster than in case of a *Group I bulge*. This is due to the increased probability that one of the degenerate nucleotides does adopt a favorable conformation which enables the frameshift to be overcome.

In typical microarray experiments the binding affinity between probe and target sequences is strongly affected by the secondary structure of the long target sequences [Lue03]. We performed a *tiling-array experiment*<sup>14</sup> to investigate the influence of target secondary structure on microarray binding affinities (see section 8.6). The cRNA-targets employed in this experiment (with lengths of about 300 and 800 nucleotides) form stable intramolecular secondary structures, which are expected to interfere with hybridization to complementary microarray probes. In agreement with [Lue03] our experimental results show that only piecewise segments of the targets (about 20 to 30% of the total length of the target sequences) are available for hybridization.

We used Sfold [Din04], a software tool for the prediction of efficient antisense oligonucleotides, to investigate the influence of restricted target accessibility on the hybridization efficiency of the individual microarray probes. We found that the predicted binding affinities (accounting for the Boltzmann ensemble of RNA target secondary structures) are correlated with our experimentally determined hybridization efficiencies. Our results suggest that Sfold is suitable for the prediction of efficient microarray probes. Further hybridization experiments are necessary to corroborate the preliminary results.

In order to investigate the underlying physics behind our experimental results – in particular with respect to the large influence of the defect position – we developed a thermodynamic model (see chapter 7) of the oligonucleotide duplex on the basis of the *double-ended zipper model* [Gib59; Kit69]. The *zipper model*, unlike the widely applied *two-state nearest-neighbor model* accounts for partially denatured duplex conformations. We assume that

---

<sup>14</sup>The tiling-array comprises a set of 25mer probes which form a 'tiling-path' along the target sequence. This type of experiment allows to probe the availability of target segments for hybridization.

duplex denaturation occurs via sequential unzipping from the duplex ends. The formation of internal denaturation bubbles is considered to be negligible.

On the basis of the *nearest-neighbor model* we calculate the statistical weights of all partially denatured duplex conformations to finally obtain the partition sum of the oligonucleotide duplex. Our theoretical results show that in the presence of a single base defect the partition function tends to be the larger if a defect is located closer to the duplex end. This is in agreement with our experimental results, where duplexes with defects near the ends are more stable than comparable duplexes with a defect in the center. The numerical analysis shows that the oligonucleotide sequence – in this context considered as a sequence of more or less strong nearest-neighbor interactions – has a significant influence on the position-dependence of the MM discrimination. This becomes apparent if stronger and weaker nearest-neighbor pairs are unevenly distributed within the sequence.

For a comparison of experimentally determined hybridization signal intensities with theoretically predicted duplex stabilities we performed a microarray hybridization experiment (section 7.4) in which the length of the microarray probes (and thus the Gibbs free energy  $\Delta G$  of the duplexes) has been varied incrementally. We observed a sigmoid relation between the hybridization signal (corresponding to the fraction of hybridized probes  $\theta$ ) and the Gibbs energy of the duplexes: Within a broad transition region the hybridization signal increases approximately linearly with the duplex length. The broad transition is not conform with the Langmuir-equation which describes a rather narrow transition. In a numerical simulation we demonstrated that this discrepancy can be explained with the influence of synthesis-related defects: different from the Langmuir equation, which is based on the assumption of a single binding affinity for all probes, the microarray probes in the experiment – owing to a variable number of point defects (located at random positions) – are subject to a heterogeneous distribution of binding affinities. Simulation results confirm that heterogeneity of binding affinities results in a broadened transition region.

The position-dependent influence of point defects is not restricted to *single base mismatches* and *base bulges*, but can also be assigned to any stronger or weaker *nearest-neighbor* pair. Our theoretical investigations in section 7.5 demonstrate that duplexes comprising of identical sets of NN-pairs – which are therefore thermodynamically equivalent on the basis of the two-state *nearest-neighbor model* – have different stabilities, depending on the arrangement of the *nearest-neighbor* pairs. Duplex stability is largest for those sequences where the most stable NN-pairs are located in the center of the duplex and less stable NN-pairs are located near the ends. At room temperature the *zipper model* provides practically the same duplex stabilities as the *two-state nearest-neighbor model*. However, with increasing temperature, owing to increasing end-domain opening, we observe a sig-

nificant positional influence. Thus, we have demonstrated that the *zipper model* provides a theoretical basis for the *positional-dependent nearest neighbor model* (PDNN) [Zha03], which has been considered so far only on a largely empirical basis.

In conclusion, the present thesis describes the development of an automated system for *in situ* synthesis of DNA microarrays. The system provides great flexibility in the fabrication of DNA microarrays at a moderate cost. Since our microarray synthesizer is not a black box technology – like most commercial microarray platforms – it is an interesting basis for further technical development (possibly on the basis of an "open source" platform).

Moreover, the present thesis contributes to a better understanding of microarray hybridization characteristics, in particular with respect to the detection of point mutations. With regard to the growing importance of DNA microarray technologies a more profound understanding of the underlying physicochemical processes will be needed to fully exploit the potential of DNA microarray analysis.



## 9.2 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde ein System zur photolithographisch kontrollierten *in situ* Synthese von DNA Microarrays entwickelt und zur Anwendung gebracht. In einer umfangreichen Studie auf der Basis von Microarray-Hybridisierungsexperimenten wurde untersucht, wie sich verschiedene Einzelbasendefekte (*single base mismatches* und *single base bulges*) in Abhängigkeit von mehreren Parametern – u.a. Defekt-Typ, Defekt-Position und Basen-Sequenz) – auf die Bindungsaffinität von Microarray-Probes auswirken.

Die Arbeit beinhaltet die folgenden Schwerpunkte:

- Entwicklung eines automatisierten DNA Microarray-Synthesizers auf der Basis der photolithographisch kontrollierten *in situ* Synthese (*light-directed in situ synthesis*).
- Quantitative Analyse der Microarray-Hybridisierung mittels Fluoreszenzmikroskopie und EM-CCD Kamera
- Charakterisierung der mit dem Synthesizer hergestellten DNA-Microarrays
- Detaillierte Untersuchung des Einflusses von Einzelbasendefekten
- Theoretische Untersuchungen: Modellierung der Dynamik von Oligonukleotid-Duplexen und der Auswirkung von Einzelbasendefekten bei der Microarray-Hybridisierung auf der Basis eines *Double-ended Zipper* Modells [Gib59; Kit69; Bin06]

Als Grundlage für experimentelle Untersuchungen mit DNA Microarrays wurde - auf der Basis von vorangegangenen Arbeiten [Fod91; Has97; SG99; Nuw02] - ein automatisiertes Microarray-Synthese-System entwickelt (siehe Kapitel 3). Mit dem Microarray-Synthesizer erfolgt die parallele Synthese von etwa 25000 verschiedenen Oligonukleotid-Probesequenzen (basierend auf der Phosphoramidit-Methode) direkt auf der Oberfläche des mit Dendrimer-Molekülen funktionalisierten Microarraysubstrats. Die *in situ* Synthese erlaubt eine hohe Flexibilität bei der Auswahl der Probe-Sequenzen, denn anders als etwa beim weit verbreiteten Kontaktdruckverfahren (*microarray spotting*: Immobilisierung vorsynthetisierter Probe-Sequenzen) werden bei der *in situ* Synthese keine vorsynthetisierten Oligonukleotide benötigt.

Cyclotriphosphazen-PMMH-Dendrimer Substrate (*phosphorus dendrimers*), die bereits sehr erfolgreich bei der Immobilisierung vorsynthetisierter Probes verwendet worden sind [LB03; Cam06], wurden im Rahmen dieser Arbeit erstmals und mit sehr gutem Erfolg als Substrat für die *in situ* Synthese verwendet (siehe Abschnitt 4.2).

Bei der photolithographisch kontrollierten *in situ* Synthese werden durch Belichtung mit UV-Licht ( $\lambda \simeq 370$  nm) photolabile NPPOC Schutzgruppen (5'-Nitrophenylpropyloxy-

carbonyl [Has97]) abgespalten<sup>15</sup>, um auf den so entschützten Microarray Features die Ankopplung des nachfolgend bereitgestellten Phosphoramidit-Basenbausteins (entsprechend den Nukleotiden A, C, G oder T) zu ermöglichen. Die Steuerung der Entschützung erfolgt photolithographisch durch die Projektion "dynamischer Photomasken". Hierzu wird das Bild eines Mikrospiegelarrays (Digital Micromirror Device, DMD<sup>TM</sup>) auf die Oberfläche des Microarraysubstrats projiziert. Die Sequenzinformation der Probe-Sequenzen ist in einem Satz von Synthesemasken codiert, die jeweils vor der dazugehörigen Kopplungsreaktion auf die Substratoberfläche projiziert werden. Bei der Belichtung mit UV-Licht werden nur diejenigen Feature-Bereiche photochemisch entschützt, welche Probe-Sequenzen enthalten, in die der bei der nachfolgenden Kopplungsreaktion bereitgestellte Nukleotidbaustein, an der entsprechenden Basen-Position, eingebaut werden soll.

Die Abbildung der Synthesemasken auf das Microarraysubstrat wurde auf der Grundlage einer Mikroskop-Projektionsphotolithographie-Anordnung (siehe Abschnitt 3.2) realisiert. Das Mikrospiegelarray wurde hierzu in der Zwischenbildebene des inversen Mikroskops angebracht. Die Projektion der "virtuellen Synthesemasken" auf das Microarraysubstrat erfolgt mittels eines handelsüblichen 5×/0,25 Fluor Mikroskopobjektivs (Carl Zeiss).

Für die Justage der Projektionsoptik im nahen UV-Bereich wurde ein neuartiger UV-sensitiver Film auf der Basis des photochromen Farbstoffs Spiropyran hergestellt. Mit Hilfe des photochromen Materials konnte das Photolithographiesystem in Hinsicht auf die Abbildungsqualität evaluiert und optimiert werden. Darüber hinaus wurde ein Verfahren entwickelt, welches eine einfache visuelle Fokussierung der UV-Optik mit Hilfe eines in sichtbarem Licht auf das Microarray-Substrat projizierten Testmusters erlaubt.

Das Fluidiksystem des Microarray-Synthesizers (Abschnitt 3.3) wurde auf der Basis eines handelsüblichen DNA-Synthesizers entwickelt. Die Ansteuerung der Magnetventile erfolgt durch einen Mikrokontroller-basierten Ventiltreiber, welcher seinerseits von einem PC gesteuert wird. Eine im Rahmen dieser Arbeit in der Programmiersprache Java entwickelte Synthesesteuerungssoftware koordiniert die Projektion der Synthesemasken mit der Steuerung des Fluidiksystems und ermöglicht damit einen vollautomatisierten Ablauf der typischerweise etwa 6-8 Stunden dauernden Microarray-Synthese.

Die Synthese erfolgt unter wasserfreien Bedingungen im Inneren der Synthesekammer. Deren Konstruktion wurde von einer Reihe technischer Anforderungen bestimmt: Verwendung chemisch aggressiver Synthesereagenzien, optische Abbildung ins Innere der Kammer, Vermeidung von Streulicht und Reflexionen, Vermeidung des Auftretens von störenden Argon-Gasbläschen im Synthese-Bereich, sowie die Vermeidung von Totvolumen (um einen schnellen und vollständigen Reagenzienaustausch zu gewährleisten). Mit

---

<sup>15</sup>Die NPPOC-Schutzgruppen befinden sich bei den in dieser Arbeit verwendeten 5'-NPPOC Phosphoramiditen am 5'-Ende der in Synthese befindlichen Probe-Sequenzen.

dem in dieser Arbeit beschriebenen Design werden die oben genannten technischen Anforderungen zuverlässig erfüllt. Das aus handelsüblichen Komponenten entwickelte DNA Microarray-Synthesystem erlaubt eine routinemäßige Herstellung maßgeschneiderter Microarrays im Labormaßstab. Vergleichbare Systeme stehen bislang nur wenigen Forschungseinrichtungen zur Verfügung.

Nach der Herstellung erfolgt die Microarray-Hybridisierung unter kontrollierten Bedingungen in einer als optische Durchflusszelle ausgeführten Hybridisierungskammer. Als *Targets* werden fluoreszenzmarkierte DNA- bzw. RNA-Oligonukleotid-Sequenzen verwendet. Das Fluoreszenzsignal von auf den Microarray-Features hybridisierten *Target*-Molekülen wird mit Hilfe eines Fluoreszenzmikroskops und einer empfindlichen EM-CCD Kamera registriert (siehe Abschnitt 5.1). Anders als bei kommerziellen Microarray-Scannersystemen ist bei dieser Anordnung ein Echtzeit-Monitoring des Hybridisierungsverlaufs möglich. Aus den gewonnenen Bilddaten werden mit einer im Rahmen dieser Arbeit entwickelten Bildverarbeitungssoftware die Hybridisierungssignale der einzelnen Microarray-Features ausgelesen.

Im Rahmen der Charakterisierung der Microarrays wurde in Abschnitt 8.1 untersucht, wie sich die im Verlauf der *in situ*-Synthese generierten Einzelbasendefekte auf die Hybridisierung der DNA Microarrays auswirken: Mit einer zunehmenden Anzahl künstlich induzierter Deletion-Defekte (generiert durch eine Verkürzung der Entschützungsdauer) nimmt die Bindungsaffinität der Probes im Median ab, woraus sich eine Abnahme des Hybridisierungssignals ergibt. Die Dissoziation der Duplexe erfolgt – aufgrund der Heterogenität der Bindungsaffinitäten im Median – bei niedrigeren Schmelztemperaturen. Ein ähnliches Experiment in Abschnitt 6.11 demonstriert, dass Microarray-Probes, die synthesebedingt (evtl. mehrere) Einzelbasendefekte enthalten, ganz erheblich zum Hybridisierungssignal beitragen können.

Die Ursache für die signifikanten Unterschiede zwischen Hybridisierungsexperimenten in Lösung und vergleichbaren Experimenten auf Microarray-Oberflächen, ist vermutlich in der durch zufällige Synthesedefekte hervorgerufenen Heterogenität der Microarray-Bindungsaffinitäten zu finden. Dies konnte durch numerische Simulationen (in Abschnitt 7.4.1) bestätigt werden.

Der auf den unterschiedlichen Bindungsaffinitäten von *perfect match* (PM) und *mismatch* (MM) Duplexen beruhende Nachweis von Einzelbasendefekten bildet die Grundlage wichtiger Anwendungen, etwa der Analyse von SNPs (*single nucleotide polymorphisms*: genetisch bedingte Variationen einzelner Basenpaare) oder der Resequenzierung von Genomabschnitten. Der Einfluss von *single mismatch*-Defekten auf die Stabilität von Oligonukleo-

tid-Duplexen wurde in Lösung (Messung des Duplex-Schmelzübergangs mittels UV-Absorptionsspektroskopie) bereits eingehend untersucht [All97; Sug00; San04].

Neuere Studien hinsichtlich der MM-Diskriminierung<sup>16</sup> in DNA Microarray-Experimenten [Wic06; Poz06] weisen deutliche Diskrepanzen sowohl im Vergleich untereinander, als auch im Vergleich mit entsprechenden Bindungsaffinitäten aus UV-Absorptionsexperimenten [Sug00; Poz06] auf.

Im Rahmen dieser Arbeit wurde daher (in Kapitel 6) detailliert untersucht, wie sich unterschiedliche Einzelbasendefekte (*single base mismatches* und *single base bulges*) auf die Bindungsaffinität zwischen Microarray-*Probes* und synthetischen Oligonukleotid-*Target*-Sequenzen auswirken. In entsprechenden Microarray-Hybridisierungsexperimenten wurden, ausgehend von einem festgehaltenen Sequenzmotiv, systematisch Defekt-Typ und Defekt-Position variiert. Hierzu wurden bei der Herstellung der Microarrays *Probe*-Sequenzen generiert, die hinsichtlich ihrer Komplementarität zu der im jeweiligen Telexperiment verwendeten *Target*-Sequenz gezielt eingebaute Einzelbasendefekte enthalten.

Unsere experimentellen Resultate (Abschnitt 6.5) zeigen, dass bei der Hybridisierung auf dem Microarray die destabilisierende Wirkung von Einzelbasen-Mismatch-Defekten sehr stark von der Position des Defekts abhängig ist. Mismatch-Defekte in der Mitte des Duplexes resultieren in einer deutlich stärkeren Mismatch-Diskriminierung als vergleichbare Defekte deren Position sich näher an den Duplexenden befindet [Nai06b]. Ein ähnlich starker Einfluss der Defektposition wird auch von [Poz06] und [Wic06] berichtet. Im Gegensatz dazu ist hinsichtlich der Hybridisierung frei-beweglicher DNA-Stränge in Lösung mit Ausnahme von [Kie99] und [Dor03] kein entsprechender Einfluss der Defektposition dokumentiert [San04].

Wir konnten erstmals zeigen, dass der Einfluss der Defektposition nicht auf *single base MMs* beschränkt ist sondern auch bei *single bulge*-Defekten zu beobachten ist: Für die einzelnen Sequenzmotive ist die Positionsabhängigkeit von *single base bulge*-Defekten (abgesehen z.B. von Defekt-Typ-abhängigen Variationen) identisch mit der bei den *single base MM*-Defekten beobachteten Positionsabhängigkeit. Unsere Messungen zeigen auch, dass die positionsabhängige Wirkung eines Defekts nicht nur vom Abstand des Defekts zu den Duplexenden, sondern auch von der jeweiligen Duplexsequenz abhängig ist. Dies ist möglicherweise darauf zurückzuführen, dass das Ausmaß der partiellen Denaturierung an den Duplexenden (*end fraying*) von der Stabilität der in den endnahen Sequenzabschnitten enthaltenen *nearest-neighbor*-Paare bestimmt wird. Dies lässt vermuten, dass der positionsabhängige Einfluss von Defekten unabhängig vom Defekt-Typ auf der Grundlage eines

---

<sup>16</sup>Mismatch-Diskriminierung: Diskriminierung zwischen *perfect-match* Duplexen (vollständig komplementären Duplexen) und *mismatch* Duplexen, deren Bindungsaffinität durch eine Einzelbasenfehlpaarung vermindert ist. Die Diskriminierung zwischen dem MM Hybridisierungssignal und dem dazugehörigen PM Hybridisierungssignal wird u. a. zum Nachweis von Punktmutationen verwendet.

Zipper-Models ("molekularer Reißverschluss") beschrieben werden kann.

Die systematische Variation der Defekt-Typen erlaubt eine umfassende Analyse der Bindungsaffinitäten der verschiedenen MM-Basenpaare (Abschnitt 6.6). Die statistische Untersuchung zeigt, dass im Falle von DNA/DNA-Hybridisierung (Verwendung von DNA-Targets) die Mismatch-Diskriminierung bei denjenigen MM-Typen am stärksten ist, bei denen ein C-G-Basenpaar (im entsprechenden PM-Duplex) durch den MM-Defekt beeinträchtigt wird. Die von uns ermittelte Reihe von MM-Bindungsaffinitäten<sup>17</sup> in DNA/DNA-Duplexen stimmt weitgehend mit einer entsprechenden Reihe von Wick *et al.* [Wic06] überein. Bezüglich einer weiteren Studie [Poz06] (basierend auf RNA/DNA-Hybridisierung) ist die Übereinstimmung dagegen vergleichsweise gering.

Dies motivierte weitere Hybridisierungs-Experimente (Abschnitt 6.8), in denen die Bindungsaffinitäten von DNA/DNA- und RNA/DNA-Mismatch-Duplexen direkt miteinander verglichen werden sollten. Hierbei konnten systematische Unterschiede hinsichtlich der MM-Diskriminierung bei DNA/DNA- und RNA/DNA-Hybridisierung beobachtet werden: Purin-Purin Mismatch-Basenpaare (A-G, G-A, A-A und G-G) zeigen bei der RNA/DNA-Hybridisierung eine etwas stärker ausgeprägte MM-Diskriminierung als bei der Hybridisierung von entsprechenden DNA/DNA-Duplexen. Die beobachteten Unterschiede sind vermutlich auf die unterschiedlichen Helix-Strukturen von A-Form RNA/DNA-Duplexen und B-Form DNA/DNA-Duplexen zurückzuführen.

*Single bulge*-Defekte werden durch eine ungepaarte Base in einem der beiden Stränge hervorgerufen. In unseren Experimenten (Abschnitt 6.9) wurden *single bulge*-Defekte (ausgehend von einem zum jeweiligen Target komplementären Probe-Sequenzmotiv) durch Einfügung (*insertion*) bzw. Entfernung (*deletion*) einzelner Nukleotide erzeugt. Ähnlich wie *single MMs* resultieren *single bulge*-Defekte in einer deutlichen Verringerung der Probe-Target-Bindungsaffinität. Die Identität der benachbarten Basen bestimmt, wie stark die Bindungsaffinität von einem *bulge*-Defekt beeinträchtigt wird: Unsere Untersuchungen zeigen, dass die Bindungsaffinität deutlich weniger vermindert wird, wenn sich die ungepaarte Base (innerhalb des Einzelstrangs) in direkter Nachbarschaft zu einer identischen Base befindet. Die Position des *bulges* d.h. der ungepaarten Base innerhalb der Duplexstruktur ist beim Vorhandensein identischer Nachbarbasen nicht eindeutig bestimmt (*positional degeneracy* [Ke95]). Eine Stabilisierung solcher (*Group II*) *bulge* Defekte wurde schon an freien Duplexen in Lösung beobachtet und ist laut Zhu und Wartell [Zhu99] auf die Entropiezunahme infolge der Positionsentartung zurückzuführen. Die in unseren Experimenten beobachtete, unerwartet deutliche Stabilisierung ist jedoch vermutlich nicht

<sup>17</sup>Da diese MM Bindungsaffinitäten auf die jeweiligen PM Bindungsaffinitäten bezogen sind stellen sie ein Maß für die MM-Diskriminierung dar.

allein aufgrund der geringfügigen Stabilisierung infolge dieser Entropiezunahme zu erklären. Unser Erklärungsansatz beruht auf einer durch den *bulge*-Defekt verursachten Blockade des Zipper-Mechanismus: Die durch den *bulge*-Defekt hervorgerufene Verschiebung zwischen den Einzelstrang-Sequenzen (*frameshift*) verhindert ein schnelles Schließen (*zipping up*) des Duplex. Diese Barriere kann beim Vorliegen eines *Group II bulges* – aufgrund der Positionsentartung – schneller übersprungen werden<sup>18</sup> als bei *Group I bulge*-Defekten (bei welchen keine Positionsentartung vorliegt).

Die Bindungsaffinität zwischen *Probe*- und *Target*-Sequenzen wird sehr stark von der Sekundärstruktur der *Target*-Sequenzen beeinflusst [Lue03]. Für ein Experiment zur Untersuchung des Einflusses solcher Sekundärstrukturen (Abschnitt 8.6), wurden fluoreszenzmarkierte cRNA-*Targets* mit einer Länge von 300 bzw. 800 Nukleotiden hergestellt. Bei diesen Längen sind stabile intramolekulare Sekundärstrukturen zu erwarten, die in den dazugehörigen Sequenzabschnitten eine Hybridisierung mit komplementären Microarray-*Probes* verhindern. Tatsächlich konnte in dem *tiling-array*-Experiment<sup>19</sup> nur auf etwa 20 bis 30% der Länge dieser *Target*-Sequenzen eine signifikante Hybridisierung erzielt werden.

Mit Hilfe von Sfold [Din04], einem Software-Tool welches u. a. zum Auffinden effektiver Antisense Oligonukleotide dient, wurde untersucht, wie sich die infolge der Sekundärstruktur verminderte Zugänglichkeit von großen Teilen der Targetsequenz auf die Bindungsaffinität der einzelnen Probesequenzen auswirkt. Unsere Ergebnisse zeigen, dass die mit Hilfe von Sfold auf theoretischer Grundlage (unter Berücksichtigung des Boltzmann-Ensembles von Target-Sekundärstrukturen) ermittelten Bindungsaffinitäten mit unseren experimentell bestimmten Hybridisierungssignalen korreliert sind. Unsere Ergebnisse legen nahe das Sfold auch zum Auffinden effizienter Microarray-Probe-Sequenzen geeignet ist. Weitere Microarray-Hybridisierungsexperimente mit anderen *Target*-Sequenzen sind erforderlich um die im Rahmen der vorliegenden Arbeit gewonnenen Ergebnisse zu untermauern.

Auf der Basis des *double-ended Zipper*-Modells [Gib59; Kit69] wurde ein thermodynamisches Modell des Oligonukleotid-Duplexes entwickelt (Kapitel 7), um die experimentellen Ergebnisse, insbesondere den starken Einfluss der Defektposition, genauer zu untersuchen. Im Gegensatz zum in der Praxis am häufigsten verwendeten *two-state nearest-*

---

<sup>18</sup>Die Stabilisierung von *Group II bulges* beruht der erhöhten Wahrscheinlichkeit, dass eine der identischen Basen eine günstige Konformation einnimmt, bei der ein rasches Fortschreiten des Zipping-Prozesses möglich ist.

<sup>19</sup>Das *tiling-array*-Experiment beinhaltet einen Satz von 25mer *Probe*-Sequenzen die entlang der sehr viel längeren *Target*-Sequenz relativ zueinander versetzt angeordnet sind. Diese Art von Experiment verfolgt den Zweck, die Bindungsaffinität der einzelnen *Target*-Bereiche zu sondieren.

*neighbor* Modell werden beim *Zipper* Modell auch die an den Enden partiell denaturierten Duplexkonformationen berücksichtigt. Ausgehend von den *nearest-neighbor* Wechselwirkungen benachbarter Basenpaare werden für die einzelnen Duplexkonformationen die statistischen Gewichte und daraus schließlich die Zustandssumme berechnet. Die theoretischen Betrachtungen zeigen, dass die Zustandssumme beim Vorliegen von Einzeldefekten umso größer ist, je näher der Defekt bei den Duplexenden liegt. Dies bestätigen die experimentellen Ergebnisse: Oligonukleotid-Duplexe mit endnahen Defekten sind stabiler als entsprechende Duplexe mit in der Mitte liegenden Defekten. Eine numerische Analyse des Defekt-Positionseinflusses auf die Bindungsaffinität zeigt, dass die Oligonukleotidsequenz, in diesem Fall als Abfolge unterschiedlicher starker *nearest-neighbor*-Wechselwirkungen betrachtet, wie bei auch experimentell beobachtet, einen signifikanten Einfluss auf die Positionsabhängigkeit der Bindungsaffinität haben kann. Dies wird vor allem offensichtlich, wenn innerhalb der Duplex-Sequenz stärkere und schwächere NN-Paare ungleichmäßig verteilt sind.

Um die experimentell bestimmten Hybridisierungssignale mit den auf theoretischer Basis ermittelten Duplexstabilitäten vergleichen zu können wurde in einem Microarray-Hybridisierungsexperiment (Abschnitt 7.4) die Länge der Probes – und somit die Gibbs-Energie  $\Delta G$  der DNA-Duplexe – schrittweise variiert. Wir beobachten einen sigmoidalen Zusammenhang  $\theta(\Delta G)$  zwischen dem Anteil hybridisierter Probes und der freien Enthalpie der Duplexe  $\Delta G$ . Über einen relativ weiten Übergangsbereich nimmt das Hybridisierungssignal näherungsweise linear mit der freien Enthalpie der Duplexe zu. Damit weicht das experimentelle Ergebnis deutlich von einem theoretischen Verlauf ab, der durch die Langmuir-Adsorptionsgleichung beschrieben wird - dieser weist einen vergleichsweise schmalen Übergangsbereich auf. Die Diskrepanz konnte anhand einer numerischen Simulation mit dem Einfluss von Synthesedefekten erklärt werden: Die in den Experimenten vorliegende breite Verteilung von Bindungsaffinitäten, die durch eine variable Anzahl von Defekten in der *Probe*-Sequenz hervorgerufen wird (die sich zudem an unterschiedlichen Positionen befinden), resultiert in einem stark verbreiterten Übergangsbereich in  $\theta(\Delta G)$ .

Die untersuchte Positionsabhängigkeit von Defekten kann auch auf die mehr oder weniger starken NN-Wechselwirkungen von Watson-Crick-Basenpaaren übertragen werden. Unsere Untersuchungen in Abschnitt 7.5 zeigen: Duplexe, die aus identischen NN-Paaren zusammengesetzt, und somit auf der Grundlage des *two-state nearest-neighbor* Modell thermodynamisch äquivalent sind, weisen im *Zipper*-Modell die größte Stabilität dann auf, wenn die stabilsten NN-Paare in der Mitte des Duplex und die schwächsten NN-Paare entsprechend an den Enden des Duplexes angeordnet sind. Bei Raumtemperatur sind die Ergebnisse des *Zipper*-Modells mit denen des *two-state nearest-neighbor* Modells praktisch

identisch. Erst mit zunehmender Temperatur ist infolge der verstärkten Denaturierung an den Duplexenden die beschriebene Positionsabhängigkeit zu beobachten. Dieses Ergebnis liefert erstmals eine theoretische Grundlage für das bislang nur auf empirischer Basis beschriebene positionsabhängige *nearest-neighbor* Modell (PDNN).

Im Rahmen der vorliegenden Arbeit wurde auf der Basis von handelsüblichen Komponenten ein flexibles System zur *in situ*-Synthese von DNA-Microarrays entwickelt. Aufgrund seiner technischen Möglichkeiten (bei vergleichsweise niedrigen Investitionen), aber auch weil es im Gegensatz zu kommerziellen Microarray-Plattformen keine Black-Box-Technologie darstellt, dürfte das hier im Detail beschriebene System eine interessante Ausgangsbasis für die Entwicklung von Microarray-Synthesizern sein. Eine (evtl. auf einer "Open Source"-Basis betriebene) Weiterentwicklung des Microarray-Synthesystems wäre wünschenswert, damit diese vielversprechende und vielseitig einsetzbare Zukunftstechnologie bald breite Anwendung finden kann.

In Hinblick auf die zunehmende Bedeutung der DNA-Microarray Technologie ist ein fundiertes Verständnis der zugrunde liegenden physikalisch-chemischen Zusammenhänge erforderlich. Vor allem in Hinblick auf die Untersuchungen zur Detektion von Punktmutationen wurde in der vorliegenden Arbeit dazu beigetragen.





# Bibliography

- [AB03] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky. Bubble dynamics in double-stranded DNA. *Physical Review Letters*, 90(13):138101, April 2003.
- [Alb03] T. J. Albert, J. Norton, M. Ott, T. Richmond, K. Nuwaysir, E. F. Nuwaysir, K. P. Stengele, and R. D. Green. Light-directed 5' - 3' synthesis of complex oligonucleotide microarrays. *Nucleic Acids Research*, 31(7):e35, April 2003.
- [Alk82] D. Alkema, P. A. Hader, R. A. Bell, and T. Neilson. Effects of flanking GC base-pairs on internal watson-crick, GU, and nonbonded base pairs within a short ribonucleic-acid duplex. *Biochemistry*, 21(9):2109–2117, 1982.
- [All97] H. T. Allawi and J. SantaLucia. Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry*, 36(34):10581–10594, August 1997.
- [Amb05] T. Ambjornsson and R. Metzler. Blinking statistics of a molecular beacon triggered by end-denaturation of DNA. *Journal of Physics-Condensed Matter*, 17(49):S4305–S4316, 2005.
- [Amb06] T. Ambjornsson, S. K. Banik, O. Krichevsky, and R. Metzler. Sequence sensitivity of breathing dynamics in heteropolymer DNA. *Physical Review Letters*, 97(12):128105, September 2006.
- [And06] D. Andreatta, S. Sen, J. L. P. Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy, R. S. Coleman, and M. A. Berg. Ultrafast dynamics in DNA: "fraying" at the end of the helix. *Journal of the American Chemical Society*, 128(21):6885–6892, May 2006.
- [App65] J. Applequist and V. Damle. Thermodynamics of helix-coil equilibrium in oligoadenylic acid from hypochromicity studies. *Journal of the American Chemical Society*, 87(7):1450–&, 1965.
- [Bar06] A. Barthel and M. Zacharias. Conformational transitions in rna single uridine and adenosine bulge structures: A molecular dynamics free energy simulation study. *Biophysical Journal*, 90(7):2450–2462, April 2006.
- [Bau03] M. Baum, S. Bielau, N. Rittner, K. Schmid, K. Eggelbusch, M. Dahms, A. Schlauersbach, H. Tahedl, M. Beier, R. Guimil, M. Scheffler, C. Hermann, J. M. Funk, A. Wixmerten, H. Rebscher, M. Honig, C. Andrae, D. Buchner,

- E. Moschel, A. Glathe, E. Jager, M. Thom, A. Greil, F. Bestvater, F. Obermeier, J. Burgmaier, K. Thome, S. Weichert, S. Hein, T. Binnewies, V. Foitzik, M. Muller, C. F. Stahler, and P. F. Stahler. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Research*, 31(23):e151, 2003.
- [Bea81] S. L. Beaucage and M. H. Caruthers. Deoxynucleoside phosphoramidites a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, 22(20):1859–1862, 1981.
- [Bei99] M. Beier and J. D. Hoheisel. Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. *Nucleic Acids Research*, 27:1970–1977, 1999.
- [Ben02] R. Benters, C. M. Niemeyer, D. Drutschmann, D. Blohm, and D. Wöhrle. DNA microarrays with PAMAM dendritic linker systems. *Nucleic Acids Research*, 30(2):e10, January 2002.
- [Bha03] G. Bhanot, Y. Louzoun, J. H. Zhu, and C. DeLisi. The importance of thermodynamic equilibrium for high throughput gene expression arrays. *Biophysical Journal*, 84(1):124–135, January 2003.
- [Bin04] H. Binder, T. Kirsten, M. Loeffler, and P. F. Stadler. Sensitivity of microarray oligonucleotide probes: Variability and effect of base composition. *Journal of Physical Chemistry B*, 108(46):18003–18014, 2004.
- [Bin06] H. Binder. Thermodynamics of competitive surface adsorption on DNA microarrays. *Journal of Physics-Condensed Matter*, 18(18):S491–S523, 2006.
- [Bla96] A. P. Blanchard, R. J. Kaiser, and L. E. Hood. High-density oligonucleotide arrays. *Biosensors & Bioelectronics*, 11(6-7):687–690, 1996.
- [Blo03] R. Blossey and E. Carlon. Reparametrizing the loop entropy weights: Effect on DNA melting curves. *Physical Review E*, 68(6):061911, December 2003.
- [Bre86] 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, June 1986.
- [Cam06] A. M. Caminade, C. Padie, R. Laurent, A. Maraval, and J. P. Majoral. Uses of dendrimers for DNA microarrays. *Sensors*, 6(8):901–914, August 2006.
- [Car06] E. Carlon and T. Heim. Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays. *Physica A-Statistical Mechanics and its Applications*, 362(2):433–449, April 2006.
- [Cha05] C. Y. Chan, C. E. Lawrence, and Y. Ding. Structure clustering features on the sfold web server. *Bioinformatics*, 21(20):3926–3928, October 2005.

- [Che07] W. W. Chen, S. Kiriwara, and Y. Miyamoto. Fabrication of three-dimensional micro photonic crystals of resin-incorporating TiO<sub>2</sub> particles and their terahertz wave properties. *Journal of the American Ceramic Society*, 90(1):92–96, January 2007.
- [Chi05] P. Y. Chiou, A. T. Ohta, and M. C. Wu. Massively parallel manipulation of single cells and microparticles using optical images. *Nature*, 436(7049):370–372, 2005.
- [Cog91] J. A. H. Cognet, J. Gabarroarpa, M. Lebret, G. A. Vandermarel, J. H. Vanboom, and G. V. Fazakerley. Solution conformation of an oligonucleotide containing a GG mismatch determined by nuclear-magnetic-resonance and molecular mechanics. *Nucleic Acids Research*, 19(24):6771–6779, December 1991.
- [Con83] B. J. Conner, A. A. Reyes, C. Morin, K. Itakura, R. L. Teplitz, and R. B. Wallace. Detection of sickle-cell beta-s-globin allele by hybridization with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 80(1):278–282, 1983.
- [Cra71] M. E. Craig, D. M. Crothers, and P. Doty. Relaxation kinetics of dimer formation by self complementary oligonucleotides. *Journal of Molecular Biology*, 62(2):383–&, 1971.
- [Cri70] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–&, 1970.
- [Cro64] D. M. Crothers and B. H. Zimm. Theory of melting transition of synthetic polynucleotides: Evaluation of stacking free energy. *Journal of Molecular Biology*, 9(1):1–&, 1964.
- [Cue04] J. A. Cuesta and A. Sanchez. General non-existence theorem for phase transitions in one-dimensional systems with short range interactions, and physical examples of such transitions. *Journal of Statistical Physics*, 115(3-4):869–893, May 2004.
- [Dan07] D. S. Dandy, P. Wu, and D. W. Grainger. Array feature size influences nucleic acid surface capture in DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4):8223–8228, February 2007.
- [Dau93] T. Dauxois, M. Peyrard, and A. R. Bishop. Dynamics and thermodynamics of a nonlinear model for DNA denaturation. *Physical Review E*, 47(1):684–695, January 1993.
- [Der05] G. Derra, H. Moench, E. Fischer, H. Giese, U. Hechtfisher, G. Hensler, A. Korerber, U. Niemann, F. C. Noertemann, P. Pekarski, J. Pollmann-Retsch, A. Ritz, and U. Weichmann. Uhp lamp systems for projection applications. *Journal of Physics D-Applied Physics*, 38(17):2995–3010, September 2005.
- [Deu04] J. M. Deutsch, S. Liang, and O. Narayan. Modelling of microarray data with zippering. *Preprint q-bio.BM/0406039 v1*, 2004. arXiv:cond-mat/0304567.

- [Din01] Y. Ding and C. E. Lawrence. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Research*, 29(5):1034–1046, March 2001.
- [Din03] Y. Ding and C. E. Lawrence. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research*, 31(24):7280–7301, December 2003.
- [Din04] Y. Ding, C. Y. Chan, and C. E. Lawrence. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Research*, 32:W135–W141, July 2004.
- [Din05] Y. Ding, C. Y. Chan, and C. E. Lawrence. RNA secondary structure prediction by centroids in a boltzmann weighted ensemble. *RNA-A Publication of the RNA Society*, 11(8):1157–1166, August 2005.
- [Dod77] J. B. Dodgson and R. D. Wells. Synthesis and thermal melting behaviour of oligomer-polymer complexes containing defined lengths of mismatched da.dg nucleotides. *Biochemistry*, 16(11):2367–2374, 1977.
- [Dor03] D. R. Dorris, A. Nguyen, L. Gieser, R. Lockner, A. Lublinsky, M. Patterson, E. Touma, T. J. Sendera, R. Elghanian, and A. Mazumder. Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios. *BMC Biotechnology*, 3:6, 2003.
- [Eve07] R. Everaers, S. Kumar, and C. Simm. Unified description of poly- and oligonucleotide DNA melting: Nearest-neighbor, poland-sheraga, and lattice models. *Physical Review E*, 75:041918, 2007.
- [Fin72] T. R. Fink and D. M. Crothers. Free-energy of imperfect nucleic-acid helices 1. bulge defect. *Journal of Molecular Biology*, 66(1):1–&, 1972.
- [Fod91] S. P. A. Fodor, J. L. Read, M. C. Pirrung, A. T. Stryer, L. and Lu, and D. Solas. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251(4995):767–773, 1991.
- [Fre86] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Sciences of the United States of America*, 83(24):9373–9377, December 1986.
- [Gao01] X. L. Gao, E. LeProust, H. Zhang, O. Srivannavit, E. Gulari, P. L. Yu, C. Nishiguchi, Q. Xiang, and X. C. Zhou. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Research*, 29(22):4744–4750, 2001.
- [Gao04] X. L. Gao, E. Gulari, and X. C. Zhou. In situ synthesis of oligonucleotide microarrays. *Biopolymers*, 73(5):579–596, April 2004.

- [Gar02] P. B. Garland and P. J. Serafinowski. Effects of stray light on the fidelity of photodirected oligonucleotide array synthesis. *Nucleic Acids Research*, 30(19):e99, October 2002.
- [Gib59] J. H. Gibbs and E. A. Dimarzio. Statistical mechanics of helix-coil transitions in biological macromolecules. *Journal of Chemical Physics*, 30(1):271–282, 1959.
- [Gil77] D. T. Gillespie. Exact stochastic simulation of coupled chemical-reactions. *Journal Of Physical Chemistry*, 81(25):2340–2361, 1977.
- [Gla06] M. Glazer, J. A. Fidanza, G. H. McGall, M. O. Trulson, J. E. Forman, A. Suseno, and C. W. Frank. Kinetics of oligonucleotide hybridization to photolithographically patterned DNA arrays. *Analytical Biochemistry*, 358(2):225–238, November 2006.
- [Got81] 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.
- [Gue87] M. Gueron, M. Kochoyan, and J. L. Leroy. A single-mode of DNA base-pair opening drives imino proton-exchange. *Nature*, 328(6125):89–92, July 1987.
- [Gut05] Z. Guttenberg, H. Muller, H. Habermuller, A. Geisbauer, J. Pipper, J. Felbel, M. Kielpinski, J. Scriba, and A. Wixforth. Planar chip device for pcr and hybridization with surface acoustic wave pump. *Lab On A Chip*, 5(3):308–317, 2005.
- [Hag88] P. J. Hagerman. Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry*, 17:265–286, 1988.
- [Hal04] A. Halperin, A. Buhot, and E. B. Zhulina. Sensitivity, specificity, and the hybridization isotherms of DNA chips. *Biophysical Journal*, 86(2):718–730, February 2004.
- [Hal05] A. Halperin, A. Buhot, and E. B. Zhulina. Brush effects on DNA chips: thermodynamics, kinetics, and design guidelines. *Biophysical Journal*, 89(2):796–811, August 2005.
- [Has97] A. Hasan, K. P. Stengele, H. Giegrich, P. Cornwell, K. R. Isham, R. A. Sachleben, W. Pfeleiderer, and R. S. Foote. Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates. *Tetrahedron*, 53(12):4247–4264, 1997.
- [Hel03] 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, June 2003.
- [Hel06] G. A. Held, G. Grinstein, and Y. Tu. Relationship between gene expression and observed intensities in DNA microarrays a modeling study. *Nucleic Acids Research*, 34(9):e70, 2006.

- [Hin78] B. Hingerty, R. S. Brown, and A. Jack. Further refinement of structure of yeast transfer-RNA phe. *Journal of Molecular Biology*, 124(3):523–534, 1978.
- [Hol91] S. R. Holbrook, C. J. Cheong, I. Tinoco, and S. H. Kim. Crystal-structure of an RNA double helix incorporating a track of non-watson-crick base-pairs. *Nature*, 353(6344):579–581, October 1991.
- [Hu05] Z. Y. Hu, M. Troester, and C. M. Perou. High reproducibility using sodium hydroxide-tripped long oligonucleotide DNA microarrays. *Biotechniques*, 38(1):121–124, January 2005.
- [Hue04] M. L. Huebschman, J. Hunt, B. Munjuluri, A. Takashima, and H. R. Garner. Design and performance of a variable spectrum synthesizer. *Review of Scientific Instruments*, 75(11):4845–4855, 2004.
- [Hun87] W. N. Hunter, T. Brown, G. Kneale, N. N. Anand, D. Rabinovich, and O. Kennard. The structure of guanosine-thymidine mismatches in b-DNA at 2.5- $\text{\AA}$  resolution. *Journal of Biological Chemistry*, 262(21):9962–9970, July 1987.
- [Iva04] V. Ivanov, Y. Zeng, and G. Zocchi. Statistical mechanics of base stacking and pairing in DNA melting. *Physical Review E*, 70(5):051907, November 2004.
- [Jac61] F. Jacob and J. Monod. Genetic regulatory mechanisms in synthesis of proteins. *Journal Of Molecular Biology*, 3(3):318–&, 1961.
- [Job02] M. Jobs, S. Fredriksson, A. J. Brookes, and U. Landegren. Effect of oligonucleotide truncation on single-nucleotide distinction by solid-phase hybridization. *Analytical Chemistry*, 74(1):199–202, 2002.
- [Ke93] S. H. Ke and R. M. Wartell. Influence of nearest-neighbor sequence on the stability of base-pair mismatches in long DNA - determination by temperature-gradient gel-electrophoresis. *Nucleic Acids Research*, 21(22):5137–5143, 1993.
- [Ke95] S. H. Ke and R. M. Wartell. Influence of neighboring base-pairs on the stability of single-base bulges and base-pairs in a DNA fragment. *Biochemistry*, 34(14):4593–4600, 1995.
- [Kie99] R. Kierzek, M. E. Burkard, and D. H. Turner. Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, 38(43):14214–14223, 1999.
- [Kim03] C. Kim, M. Li, N. Venkataramaia, K. Richmond, J. Kaysen, and F. Cerrina. DNA microarrays: an imaging study. *Journal of Vacuum Science & Technology B*, 21:2946, 2003.
- [Kim04] C. Kim, M. Li, M. Rodesch, A. Lowe, K. Richmond, and F. Cerrina. Biological lithography: Improvements in DNA synthesis methods. *Journal of Vacuum Science & Technology B*, 22(6):3163–3167, 2004.
- [Kit69] C. Kittel. Phase transition of a molecular zipper. *American Journal of Physics*, 37(9):917–&, 1969.

- [KK03] F. R. Kretschmer-Kazemi and G. Sczakiel. The activity of sirna in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Research*, 31(15):4417–4424, August 2003.
- [Koe05] R. T. Koehler and N. Peyret. Thermodynamic properties of dna sequences: characteristic values for the human genome. *Bioinformatics*, 21(16):3333–3339, August 2005.
- [Koh06] M. Kohandel and B. Y. Ha. Thermal denaturation of double-stranded DNA: Effect of base stacking. *Physical Review E*, 73(1):011905, January 2006.
- [Koo01] E. T. Kool. Hydrogen bonding, base stacking, and steric effects in dna replication. *Annual Review of Biophysics and Biomolecular Structure*, 30:1–22, 2001.
- [Kot00a] E. Y. Kotova, E. Y. Kreindlin, V. E. Barsky, and A. D. Mirzabekov. Effects of various fluorochromes and competition between labeled oligonucleotides on their hybridization to oligonucleotides immobilized on biological microchips. *Molecular Biology*, 34(2):207–214, 2000.
- [Kot00b] E. Y. Kotova, E. Y. Kreindlin, V. E. Barsky, and A. D. Mirzabekov. Optical properties of fluorochromes promising for use in biological microchips. *Molecular Biology*, 34(2):266–271, 2000.
- [Kro04] A. H. Krotz, C. Rentel, D. Gorman, P. Olsen, H. J. Gaus, J. V. McArdle, and A. N. Scozzari. Solution stability and degradation pathway of deoxyribonucleoside phosphoramidites in acetonitrile. *Nucleosides nucleotides & nucleic acids*, 23(5):767–775, May 2004.
- [LB03] V. Le Berre, E. Trevisiol, A. Dagkessamanskaia, S. Sokol, A. M. Caminade, J. P. Majoral, B. Meunier, and J. Francois. Dendrimeric coating of glass slides for sensitive DNA microarrays analysis. *Nucleic Acids Research*, 31(16):e88, 2003.
- [Lee04] I. Lee, A. A. Dombkowski, and B. D. Athey. Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. *Nucleic Acids Research*, 32(2):681–690, 2004.
- [Lei92] M. Leijon and A. Graslund. Effects of sequence and length on imino proton-exchange and base pair opening kinetics in DNA oligonucleotide duplexes. *Nucleic Acids Research*, 20(20):5339–5343, October 1992.
- [Lev05] R. Levicky and A. Horgan. Physicochemical perspectives on DNA microarray and biosensor technologies. *Trends in Biotechnology*, 23(3):143–149, March 2005.
- [Liu05] W. T. Liu, J. H. Wu, E. S. Y. Li, and E. S. Selamat. Emission characteristics of fluorescent labels with respect to temperature changes and subsequent effects on DNA microchip studies. *Applied and Environmental Microbiology*, 71(10):6453–6457, October 2005.

- [Lu06] Y. Lu, G. Mapili, G. Suhali, S. C. Chen, and K. Roy. A digital micro-mirror device-based system for the microfabrication of complex, spatially patterned tissue engineering scaffolds. *Journal of Biomedical Materials Research Part A*, 77A(2):396–405, May 2006.
- [Lue02] K. J. Luebke, R. P. Balog, D. Mittelman, and H. R. Garner. Digital optical chemistry: A novel system for the rapid fabrication of custom oligonucleotide arrays. *Microfabricated Sensors, Application of Optical Technology for DNA Analysis*, Richard Kordal, Author Usmani and Wai Tak Law, editors, American Chemical Society Publications, 815:87–106, 2002.
- [Lue03] K. J. Luebke, R. P. Balog, and H. R. Garner. Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Research*, 31(2):750–758, January 2003.
- [Mac98] T. Macke and D. A. Case. *Molecular Modeling of Nucleic Acids*, chapter Modeling unusual nucleic acid structures, pages 379–393. American Chemical Society, 1998.
- [Mar05] N. R. Markham and M. Zuker. Dinamelt web server for nucleic acid melting prediction. *Nucleic Acids Research*, 33:W577–W581, July 2005.
- [Mat03] O. V. Matveeva, S. A. Shabalina, V. A. Nemtsov, A. D. Tsodikov, R. F. Gesteland, and J. F. Atkins. Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Research*, 31(14):4211–4217, July 2003.
- [Mau06] K. Maurer, J. Cooper, Caraballo M., J. Crye, D. Suciu, A. Ghindilis, J. A. Leonetti, W. Wang, F. M. Rossi, A. G. Stover, C. Larson, H. Gao, K. Dill, and A. McShea. Electrochemically generated acid and its containment to 100 micron reaction areas for the production of DNA microarrays. *PLOS One*, 1:e34, 2006.
- [McG97] G. H. McGall, A. D. Barone, M. Diggelmann, S.P. A. Fodor, E. Gentalen, and N. Ngo. The efficiency of light-directed synthesis of DNA arrays on glass substrates. *Journal of the American Chemical Society*, 119(22):5081–5090, 1997.
- [Mer63] R. B. Merrifield. Solid phase peptide synthesis - synthesis of a tetrapeptide. *Journal Of The American Chemical Society*, 85(14):2149–&, 1963.
- [Mic07] W. Michel, T. Mai, T. Naiser, and A. Ott. Optical study of DNA surface hybridization reveals DNA surface density as a key parameter for microarray hybridization kinetics. *Biophysical Journal*, 92(3):999–1004, February 2007.
- [Mir99] K. U. Mir and E. M. Southern. Determining the influence of structure on hybridization using oligonucleotide arrays. *Nature Biotechnology*, 17(8):788–792, August 1999.
- [Mod01] M. A. Model and J. K. Burkhardt. A standard for calibration and shading correction of a fluorescence microscope. *Cytometry*, 44(4):309–316, August 2001.

- [Moo05] M. J. Moorcroft, W. R. A. Meuleman, S. G. Latham, T. J. Nicholls, R. D. Egeland, and E. M. Southern. In situ oligonucleotide synthesis on poly(dimethylsiloxane): a flexible substrate for microarray fabrication. *Nucleic Acids Research*, 33(8):e75, 2005.
- [Nai06a] T. Naiser, O. Ehler, T. Mai, W. Michel, and A. Ott. Hybridization to surface-bound oligonucleotide probes: Influence of point defects. 2006. arXiv: q-bio.QM/0612043.
- [Nai06b] T. Naiser, T. Mai, W. Michel, and A. Ott. Versatile maskless microscope projection photolithography system and its application in light-directed fabrication of DNA microarrays. *Review of Scientific Instruments*, 77(6):063711, 2006.
- [Nel81] J. W. Nelson, F. H. Martin, and I. Tinoco. DNA and RNA oligomer thermodynamics the effect of mismatched bases on double-helix stability. *Biopolymers*, 20(12):2509–2531, 1981.
- [Nuw02] E. F. Nuwaysir, W. Huang, T. J. Albert, J. Singh, K. Nuwaysir, A. Pitas, T. Richmond, T. Gorski, J. P. Berg, J. Ballin, M. McCormick, J. Norton, T. Pollock, T. Sumwalt, L. Butcher, D. Porter, M. Molla, C. Hall, F. Blattner, M. R. Sussman, R. L. Wallace, F. Cerrina, and R. D. Green. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Research*, 12(11):1749–1755, 2002.
- [Owc97] R. Owczarzy, P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, and A. S. Benight. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*, 44(3):217–239, 1997.
- [Owc04] R. Owczarzy, Y. You, B. G. Moreira, J. A. Manthey, L. Y. Huang, M. A. Behlke, and J. A. Walder. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry*, 43(12):3537–3554, March 2004.
- [Owc05] R. Owczarzy. Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophysical Chemistry*, 117(3):207–215, October 2005.
- [Pap06] K. Pappaert, H. Ottevaere, H. Thienpont, P. Van Hummelen, and G. Desmet. Diffusion limitation: a possible source for the occurrence of doughnut patterns on DNA microarrays. *Biotechniques*, 41(5):609–616, November 2006.
- [Pat82] D. J. Patel, S. A. Kozlowski, L. A. Marky, J. A. Rice, C. Broka, J. Dallas, K. Itakura, and K. J. Breslauer. Structure, dynamics, and energetics of deoxyguanosine-thymidine wobble base pair formation in the self-complementary d(cgtgaattcgcg) duplex in solution. *Biochemistry*, 21(3):437–444, 1982.
- [Pea94] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. A. Fodor. Light-generated oligonucleotide arrays for rapid dna-sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11):5022–5026, 1994.

- [Pet02] A. W. Peterson, L. K. Wolf, and R. M. Georgiadis. Hybridization of mismatched or partially matched DNA at surfaces. *Journal of the American Chemical Society*, 124(49):14601–14607, 2002.
- [Pet04] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004.
- [Pey99] N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. SantaLucia. Nearest-neighbor thermodynamics and nmr of DNA sequences with internal AA, CC, GG, and TT mismatches. *Biochemistry*, 38(12):3468–3477, March 1999.
- [Pie00] J. Piehler, A. Brecht, R. Valiokas, B. Liedberg, and G. Gauglitz. A high-density poly(ethylene glycol) polymer brush for immobilization on glass-type surfaces. *Biosensors & Bioelectronics*, 15(9-10):473–481, November 2000.
- [Pol66] D. Poland and H. A. Scheraga. Occurrence of a phase transition in nucleic acid models. *Journal of Chemical Physics*, 45(5):1464–&, 1966.
- [Por77] D. Porschke. Elementary steps of base recognition and helix-coil transitions in nucleic acids. *Mol Biol Biochem Biophys.*, 24:191–218, 1977.
- [Poz02] A. E. Pozhitkov and D. Tautz. An algorithm and program for finding sequence specific oligo-nucleotide probes for species identification. *BMC Bioinformatics*, 3:9, 2002.
- [Poz06] A. Pozhitkov, P. A. Noble, T. Domazet-Loso, A. W. Nolte, R. Sonnenberg, P. Staehler, M. Beier, and D. Tautz. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Research*, 34(9):e66, 2006.
- [Poz07] A. E. Pozhitkov, R. D. Stedtfeld, S. A. Hashsham, and P. A. Noble. Revision of the nonequilibrium thermal dissociation and stringent washing approaches for identification of mixed nucleic acid targets by microarrays. *Nucleic Acids Research*, 35(9):e70, May 2007.
- [Pta02] M. Ptashne and A. Gann. *Genes&Signals*. Cold Spring Harbor Laboratory Press, 2002.
- [Rat05] V. G. Ratushna, J. W. Weller, and C. J. Gibas. Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC Genomics*, 6:31, March 2005.
- [Ric04] K. E. Richmond, M. H. Li, M. J. Rodesch, M. Patel, A. M. Lowe, C. Kim, L. L. Chu, N. Venkataramaian, S. F. Flickinger, J. Kaysen, P. J. Belshaw, M. R. Sussman, and F. Cerrina. Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Research*, 32(17):5011–5018, 2004.

- [RL06] E. Rodriguez-Lebron and H. L. Paulson. Allele-specific RNA interference for neurological disease. *Gene Therapy*, 13(6):576–581, March 2006.
- [Rot06] P. W. K. Rothmund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, March 2006.
- [Sab05] C. R. Sabanayagam, J. S. Eid, and A. Meller. Long time scale blinking kinetics of cyanine fluorophores conjugated to DNA and its effect on forster resonance energy transfer. *Journal of Chemical Physics*, 123(22):224708, December 2005.
- [San98] J. SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95(4):1460–1465, February 1998.
- [San04] J. SantaLucia and D. Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440, 2004.
- [Sch95] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270:467–470, 1995.
- [Sch98] M. Schena, R. A. Heller, T. P. Theriault, K. Konrad, E. Lachenmeier, and R. W. Davis. Microarrays: biotechnology’s discovery platform for functional genomics. *Trends in Biotechnology*, 16(7):301–306, July 1998.
- [Sch99] M. Schena, editor. *DNA Microarrays: a practical approach*. Oxford University Press, 1999.
- [Sch02] M. Schena. *Microarray Analysis*. Wiley, 2002.
- [Sch05] J. Schuster, F. Cichos, and C. von Borczyskowski. Influence of self-trapped states on the fluorescence intermittency of single molecules. *Applied Physics Letters*, 87(5):051915, August 2005.
- [Sch06] D. S. Schwarz, H. L. Ding, L. Kennington, J. T. Moore, J. Schelter, J. Burchard, P. S. Linsley, N. Aronin, Z. S. Xu, and P. D. Zamore. Designing siRNA that distinguish between genes that differ by a single nucleotide. *Plos Genetics*, 2(9):e140, September 2006.
- [Sch07] J. Schuster, J. Brabandt, and C. von Borczyskowski. Discrimination of photoblinking and photobleaching on the single molecule level. *Journal of Luminescence*, 127(1):224–229, November 2007.
- [SG99] S. Singh-Gasson, R. D. Green, Y. J. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnology*, 17(10):974–978, 1999.

- [Sin84] N. D. Sinha, J. Biernat, J. Mcmanus, and H. Koster. Polymer support oligonucleotide synthesis .18. use of beta-cyanoethyl-n,n-dialkylamino-/n-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Research*, 12(11):4539–4557, 1984.
- [Ske93] J. V. Skelly, K. J. Edwards, T. C. Jenkins, and S. Neidle. Crystal-structure of an oligonucleotide duplex containing G.G base-pairs influence of mispairing on DNA backbone conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 90(3):804–808, February 1993.
- [Sou75] E.M. Southern. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol.*, 98:503–517, 1975.
- [Sug86] N. Sugimoto, R. Kierzek, S. M. Freier, and D. H. Turner. Energetics of internal GU mismatches in ribooligonucleotide helices. *Biochemistry*, 25(19):5755–5759, September 1986.
- [Sug95] N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, 34(35):11211–11216, September 1995.
- [Sug00] N. Sugimoto, M. Nakano, and S. Nakano. Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes. *Biochemistry*, 39(37):11270–11281, 2000.
- [Sun00] M. Sundaralingam and Y. Xiong. Crystal structure of domain II of x-laevis somatic 5s RNA in two conformations. *Biophysical Journal*, 78(1):311A–311A, January 2000.
- [Sun05] C. Sun, N. Fang, D. M. Wu, and X. Zhang. Projection micro-stereolithography using digital micro-mirror dynamic mask. *Sensors and Actuators A-Physical*, 121(1):113–120, 2005.
- [Tin73] I. Tinoco, P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. Improved estimation of secondary structure in ribonucleic-acids. *Nature-New Biology*, 246(150):40–41, 1973.
- [Toe03] A. Toegl, R. Kirchner, C. Gauer, and A. Wixforth. Enhancing results of microarray hybridization trough microagitation. *Journal of Biomolecular Techniques*, 14:197–204, 2003.
- [Tur92] D. H. Turner. Bulges in nucleic acids. *Current Opinion in Structural Biology*, 2:334–337, 1992.
- [Ura02] H. Urakawa, P. A. Noble, S. El Fantroussi, J. J. Kelly, and D. A. Stahl. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Applied and Environmental Microbiology*, 68(1):235–244, January 2002.

- [Ura03] H. Urakawa, S. El Fantroussi, H. Smidt, J. C. Smoot, E. H. Tribou, J. J. Kelly, P. A. Noble, and D. A. Stahl. Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Applied and environmental microbiology*, 69(5):2848–2856, 2003.
- [Vai02] A. Vainrub and B. M. Pettitt. Coulomb blockage of hybridization in two-dimensional DNA arrays. *Physical Review E*, 66(4):041905, October 2002.
- [vE06] T. S. van Erp, S. Cuesta-Lopez, and M. Peyrard. Bubbles and denaturation in DNA. *European Physical Journal E*, 20(4):421–434, August 2006.
- [Vic00] T. A. Vickers, J. R. Wyatt, and S. M. Freier. Effects of rna secondary structure on cellular antisense activity. *Nucleic Acids Research*, 28(6):1340–1347, March 2000.
- [Vij01] R. A. Vijayendran and D. E. Leckband. A quantitative assessment of heterogeneity for surface-immobilized proteins. *Analytical Chemistry*, 73(3):471–480, February 2001.
- [Wal79] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura. Hybridization of synthetic oligodeoxyribonucleotides to phi-chi-174 DNA effect of single base pair mismatch. *Nucleic Acids Research*, 6(11):3543–3557, 1979.
- [Wal01] S. Walbert, W. Pfliederer, and U. E. Steiner. Photolabile protecting groups for nucleosides: Mechanistic studies of the 2-(2-nitrophenyl)ethyl group. *Helvetica Chimica Acta*, 84(6):1601–1611, 2001.
- [War85] R. M. Wartell and A. S. Benight. Thermal-denaturation of DNA-molecules: a comparison of theory with experiment. *Physics Reports-Review Section of Physics Letters*, 126(2):67–107, 1985.
- [Wat00] J. H. Watterson, P. A. E. Piunno, C. C. Wust, and U. J. Krull. Effects of oligonucleotide immobilization density on selectivity of quantitative transduction of hybridization of immobilized DNA. *Langmuir*, 16(11):4984–4992, May 2000.
- [Wes07] E. M. Westerhout and B. Berkhout. A systematic analysis of the effect of target rna structure an rna interference. *Nucleic Acids Research*, 35(13):4322–4330, 2007.
- [Wet68] J. G. Wetmur and N. Davidson. Kinetics of renaturation of DNA. *Journal of Molecular Biology*, 31(3):349–&, 1968.
- [Wet91] J. G. Wetmur. DNA probes: Applications of the principles of nucleic-acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology*, 26(3-4):227–259, 1991.
- [Wic06] L. M. Wick, J. M. Rouillard, T. S. Whittam, E. Gulari, J. M. Tiedje, and S. A. Hashsham. On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Research*, 34(3):e26, 2006.

- [Woe06] D. F. Woell. *Neue photolabile Schutzgruppen mit intramolekularer Sensibilisierung - Synthese, photokinetische Charakterisierung und Anwendung für die DNA-Chip-Synthese*. PhD thesis, Universitaet Konstanz, 2006.
- [Wol04] D. Woll, S. Walbert, K. P. Stengele, T. J. Albert, T. Richmond, J. Norton, M. Singer, R. D. Green, W. Pfeleiderer, and U. E. Steiner. Triplet-sensitized photodeprotection of oligonucleotides in solution and on microarray chips. *Helvetica Chimica Acta*, 87(1):28–45, 2004.
- [Won04] C. W. Wong, T. J. Albert, V. B. Vega, J. E. Norton, D. J. Cutler, T. A. Richmond, L. W. Stanton, E. T. Liu, and L. D. Miller. Tracking the evolution of the sars coronavirus using high-throughput, high-density resequencing arrays. *Genome Research*, 14(3):398–405, March 2004.
- [Woo88] S. A. Woodson and D. M. Crothers. Structural model for an oligonucleotide containing a bulged guanosine by NMR and energy minimization. *Biochemistry*, 27(9):3130–3141, May 1988.
- [Wu87] H. N. Wu and O. C. Uhlenbeck. Role of a bulged-a residue in a specific RNA protein-interaction. *Biochemistry*, 26(25):8221–8227, December 1987.
- [Yil04] L. S. Yilmaz and D. R. Noguera. Mechanistic approach to the problem of hybridization efficiency in fluorescent in situ hybridization. *Applied And Environmental Microbiology*, 70(12):7126–7139, December 2004.
- [Yoo01] J. S. Yoo, H. K. Cheong, B. J. Lee, Y. B. Kim, and C. Cheong. Solution structure of the SL1 RNA of the m1 double-stranded RNA virus of saccharomyces cerevisiae. *Biophysical Journal*, 80(4):1957–1966, April 2001.
- [Zen06] Y. Zeng and G. Zocchi. Mismatches and bubbles in DNA. *Biophysical Journal*, 90(12):4522–4529, 2006.
- [Zha03] L. Zhang, M. F. Miles, and K. D. Aldape. A model of molecular interactions on short oligonucleotide microarrays. *Nature Biotechnology*, 21(7):818–821, July 2003.
- [Zha07] L. Zhang, C. L. Wu, R. Carta, and H. T. Zhao. Free energy of DNA duplex formation on short oligonucleotide microarrays. *Nucleic Acids Research*, 35(3):e18, February 2007.
- [Zho06] H. Zhou, Y. Zhang, and Z. Ou-Yang. *Handbook of Theoretical and Computational Nanotechnology*, chapter Chapter 9: Theoretical and Computational Treatments of DNA and RNA Molecules, pages 419–487. American Scientific Publishers, 2006.
- [Zhu99] J. Zhu and R. M. Wartell. The effect of base sequence on the stability of RNA and DNA single base bulges. *Biochemistry*, 38(48):15986–15993, 1999.
- [Zim60] B. H. Zimm. Theory of melting of the helical form in double chains of the DNA type. *Journal of Chemical Physics*, 33(5):1349–1356, 1960.

- [Zno02] B. M. Znosko, S. B. Silvestri, H. Volkman, B. Boswell, and M. J. Serra. Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges. *Biochemistry*, 41(33):10406–10417, 2002.
- [Zoc03] G. Zocchi, A. Omerzu, T. Kuriabova, J. Rudnick, and G. Gruner. Duplex-single strand denaturation transition in DNA oligomers. 2003. arXiv:cond-mat/0304567.

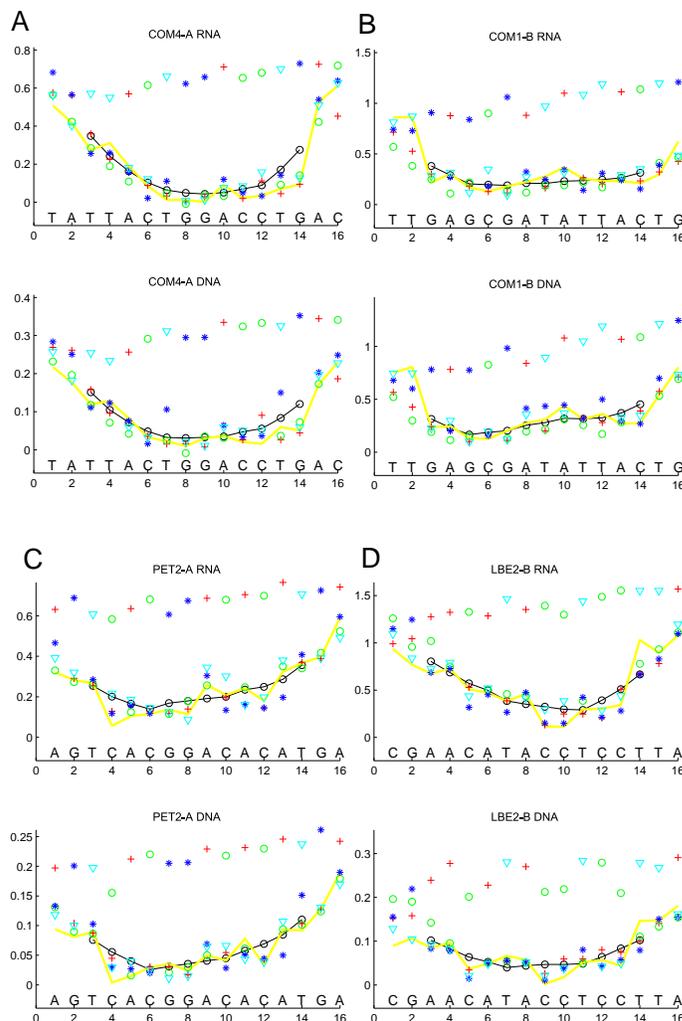


# Appendix A

## Experimental Data

## A.1 Experimental Data

### A.1.1 Comparison Between MMs in RNA/DNA and DNA/DNA Duplexes



**Figure A.1:** Direct comparison of DNA/DNA and RNA/DNA mismatch hybridization signals (see section 6.8). Parts A-D compare defect profiles of different sequence motifs (sequences shown at the bottom of the plots). Hybridizations of RNA targets (top image) and equivalent DNA targets (bottom image) were performed subsequently on the same microarrays. The defect positional influence is very similar for DNA/DNA and RNA/DNA hybridization. However, there are systematic differences between the binding affinities of the various MM types in DNA/DNA and RNA/DNA duplexes. The hybridization signal (in a.u.) is plotted versus defect position. Substitution bases A (red cross), C (green circle), G (blue star) and T (cyan triangle) either result in 3 MM duplexes and one PM duplex at every defect position; Hybridization signals of duplexes with single base deletions (yellow line); moving average MM hybridization signal (black line).

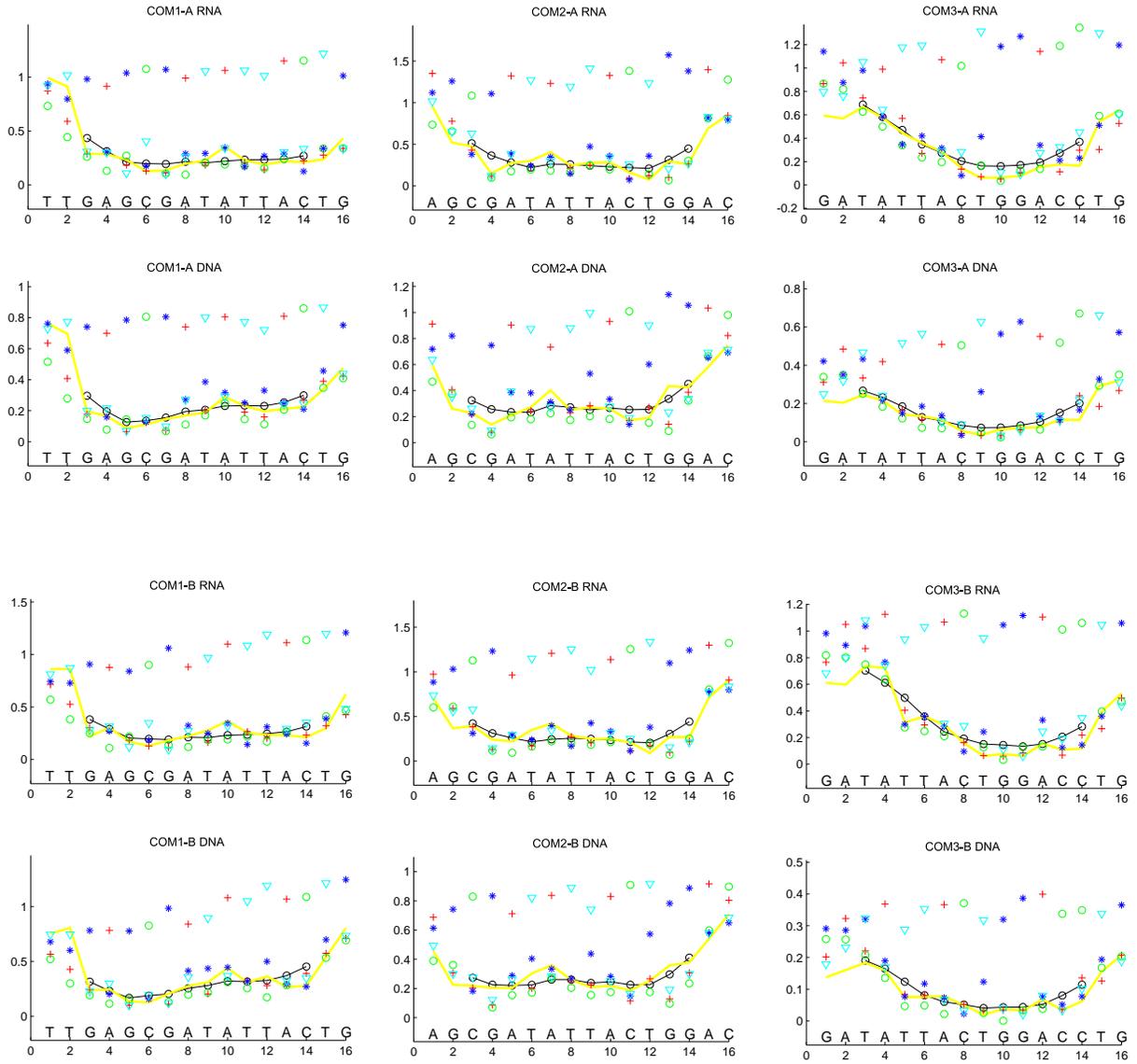


Figure A.2: For details see Fig. A.1.

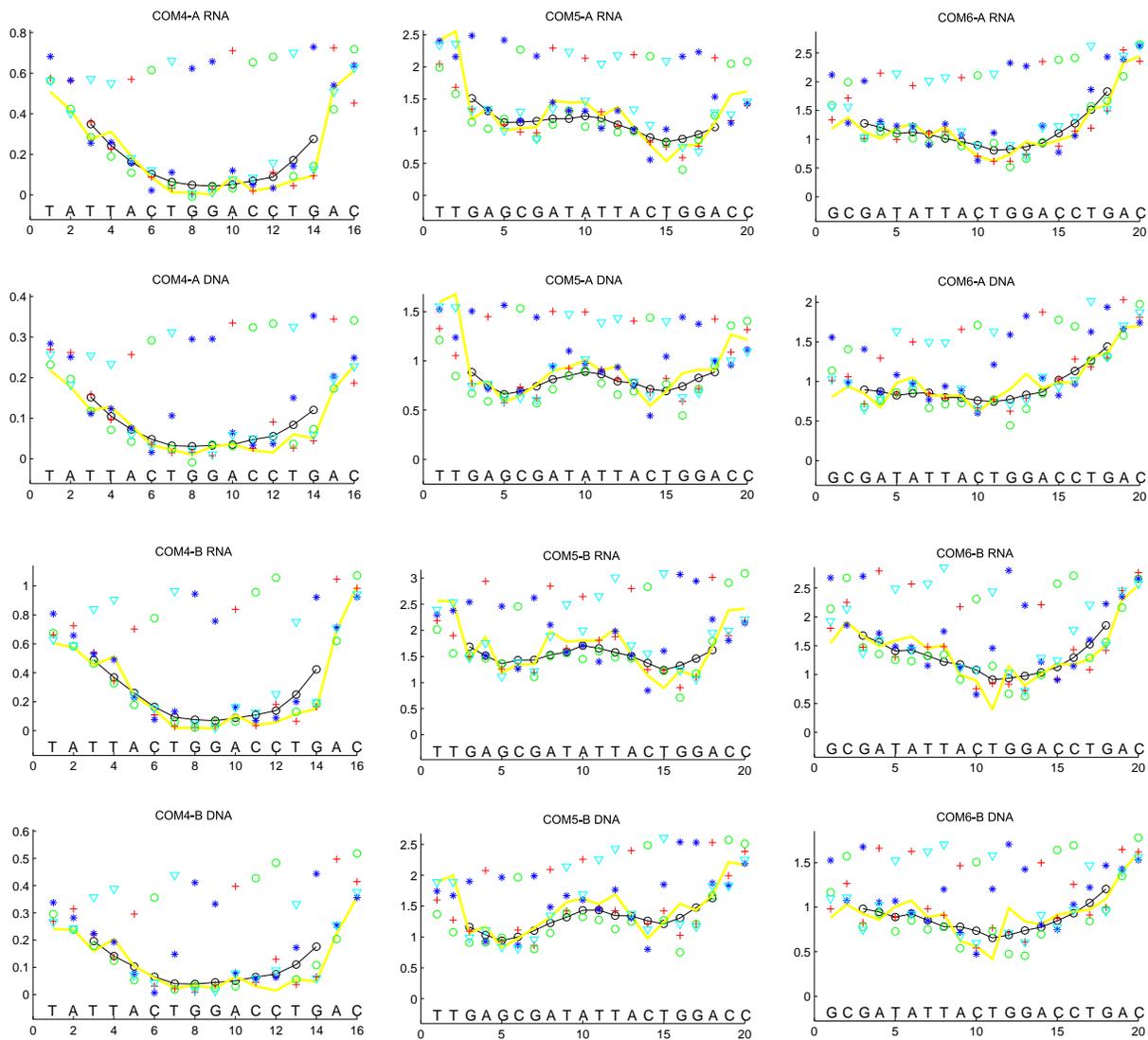


Figure A.3: For details see Fig. A.1.

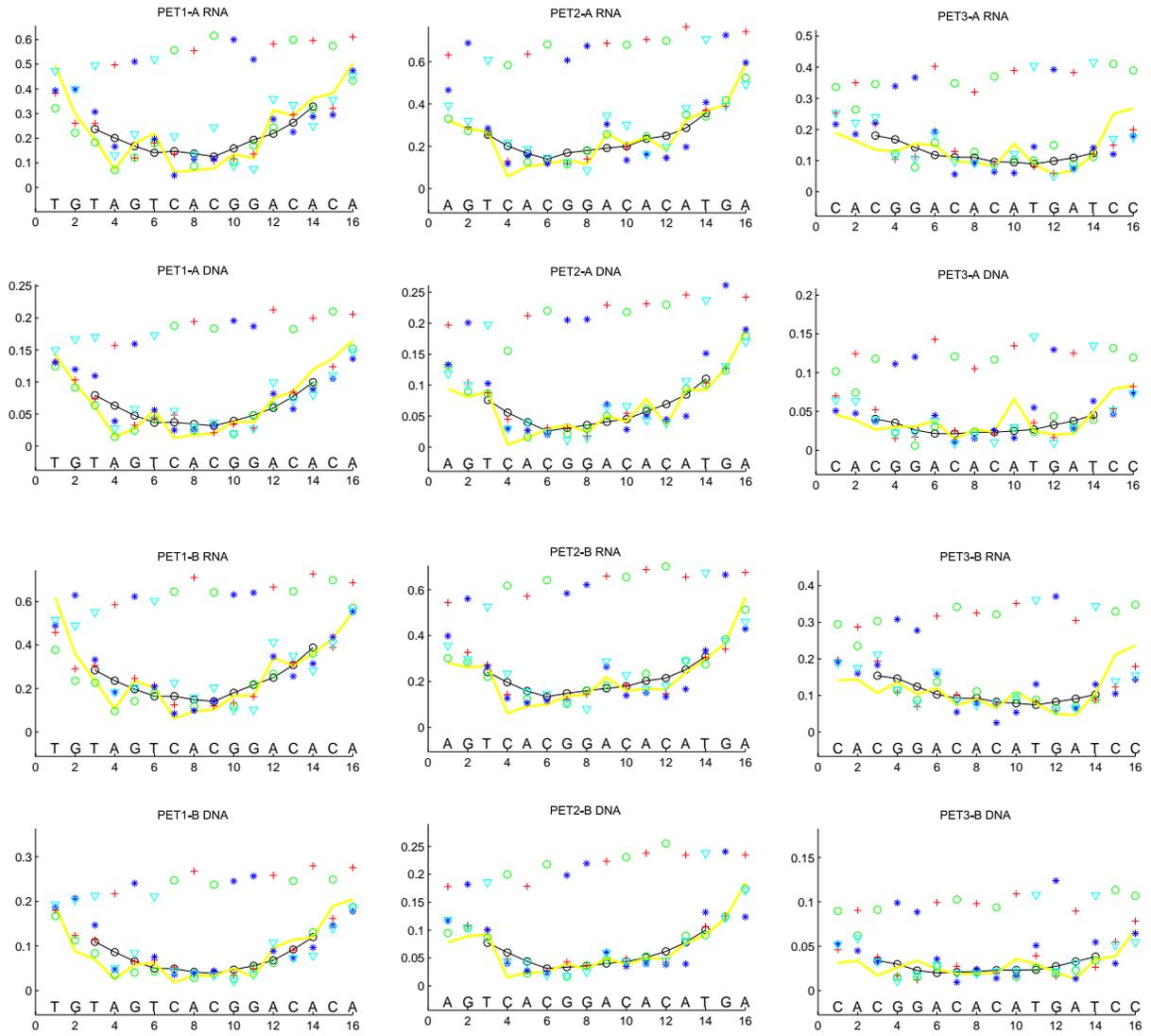


Figure A.4: For details see Fig. A.1.

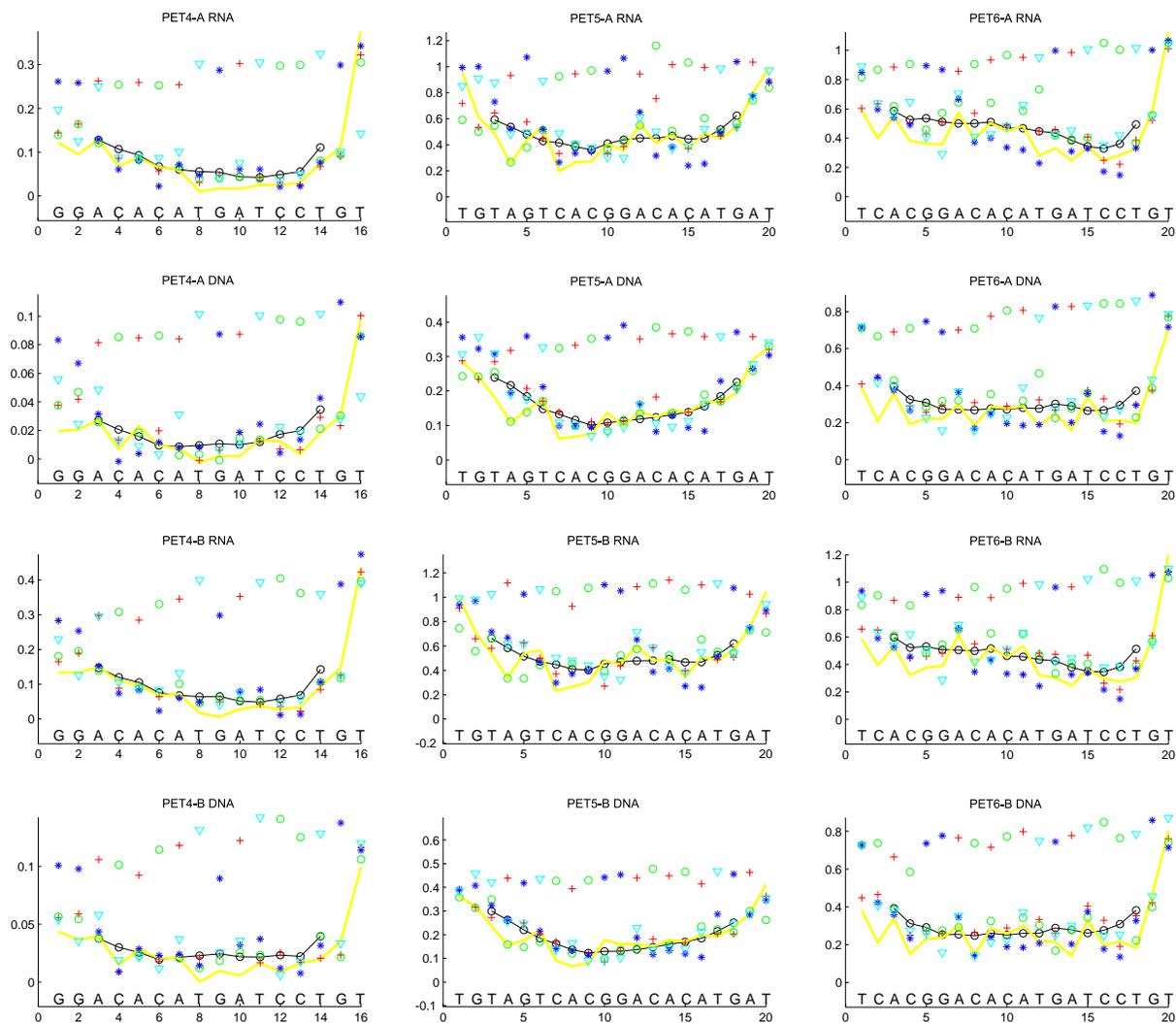


Figure A.5: For details see Fig. A.1.

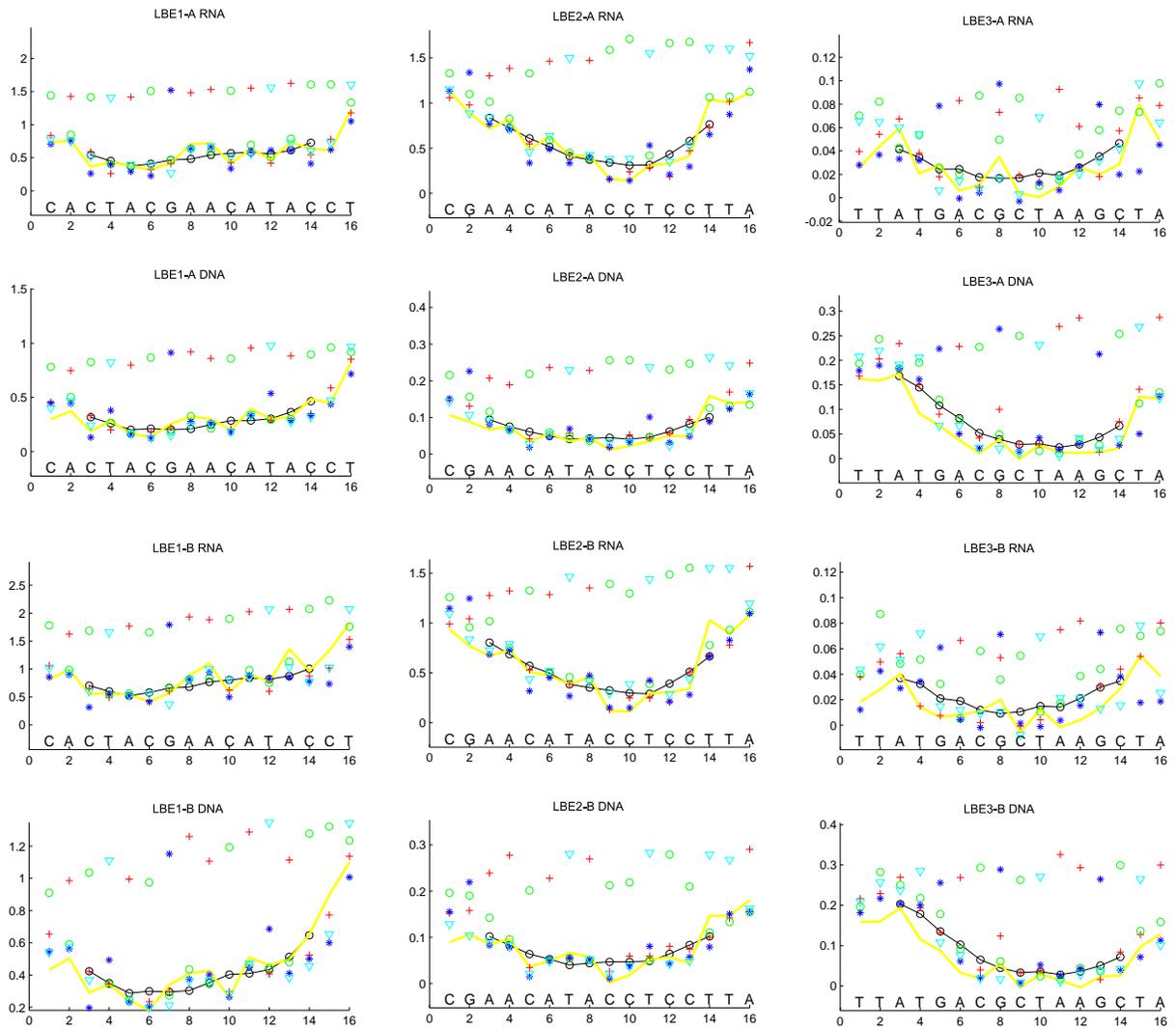


Figure A.6: For details see Fig. A.1.

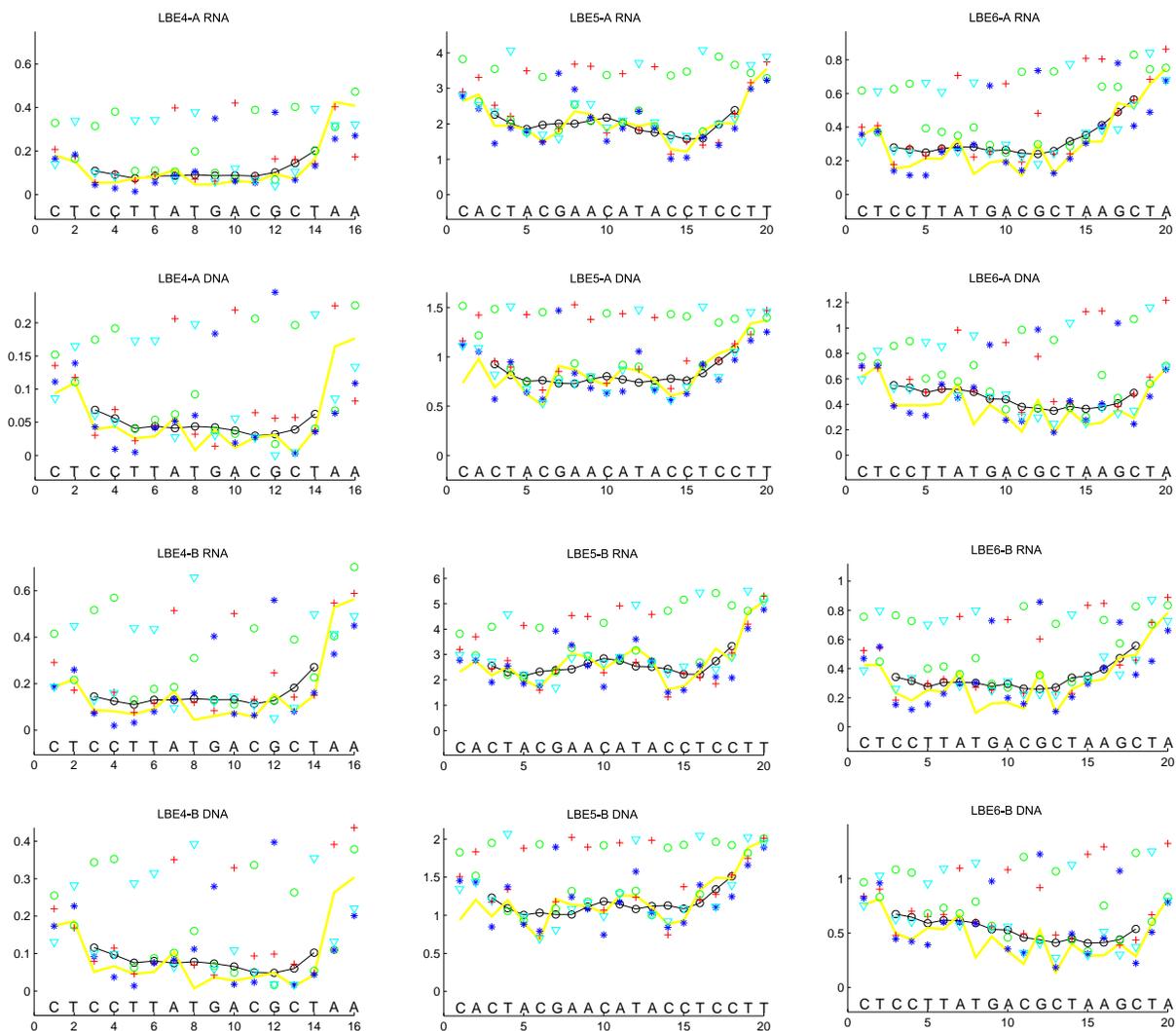


Figure A.7: For details see Fig. A.1.

## A.1.2 Single Base Insertion Defect Profiles

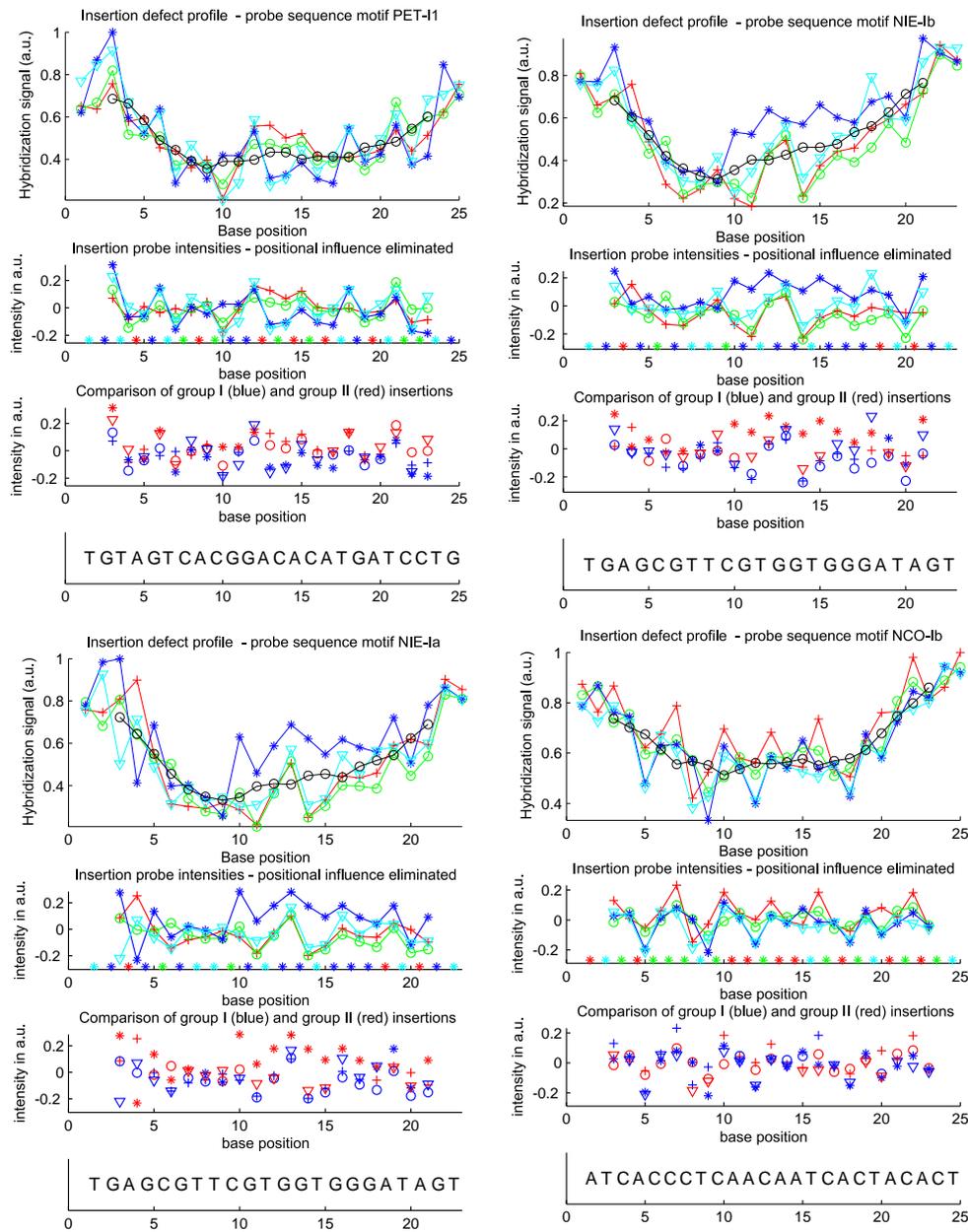


Figure A.8: Single base insertion defect profiles. For details see Fig. 6.15.

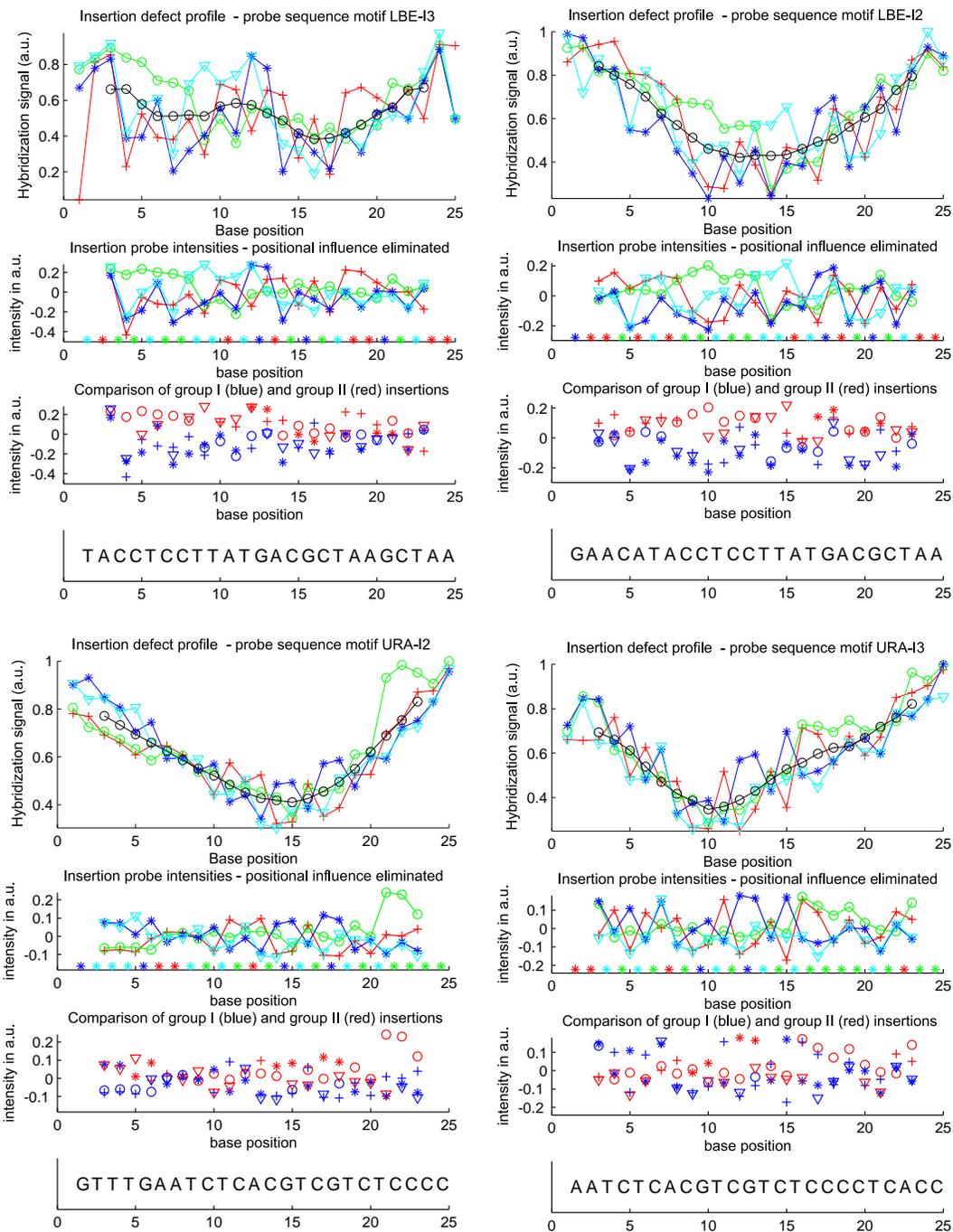


Figure A.9: Single base insertion defect profiles. For details see Fig. 6.15.

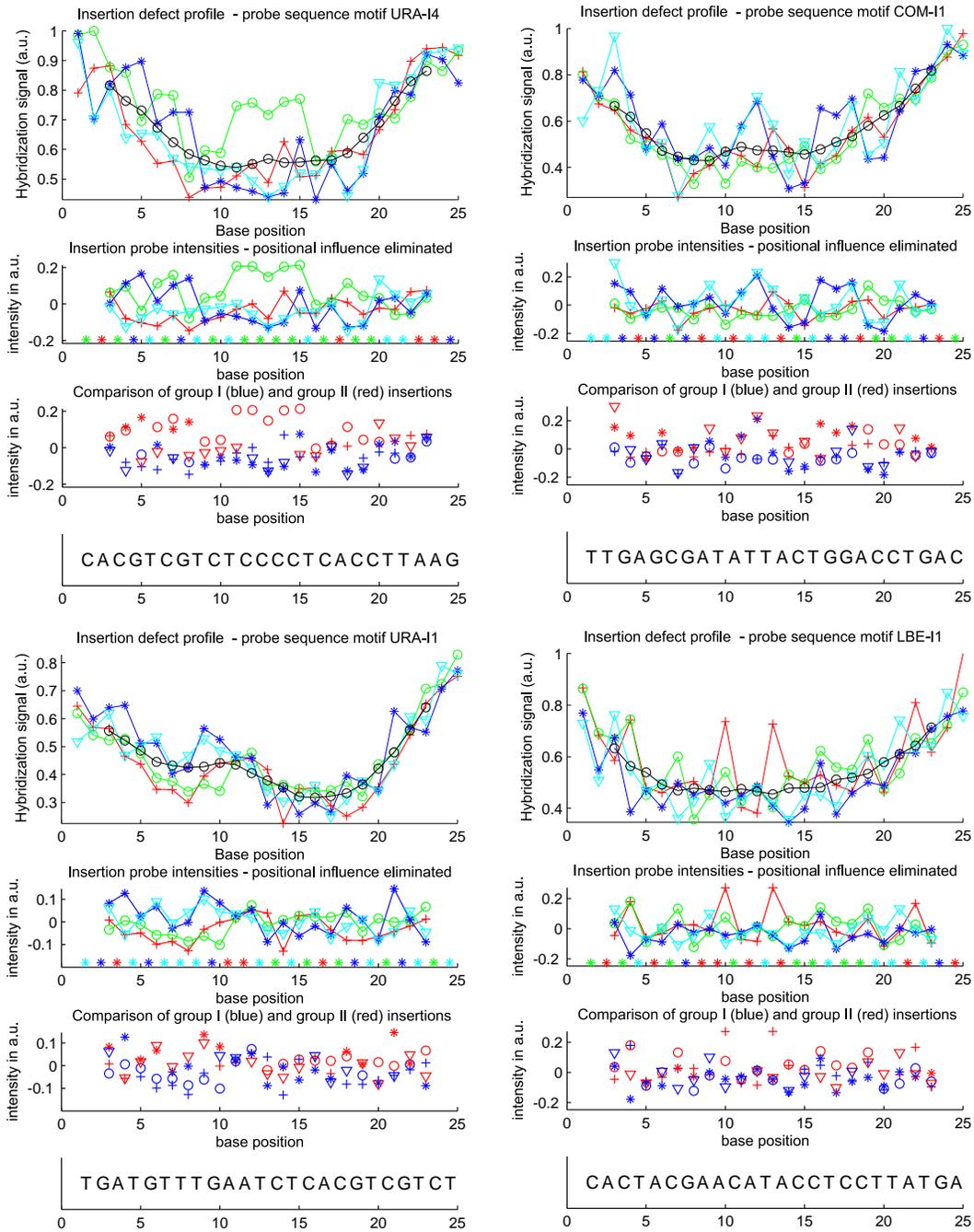
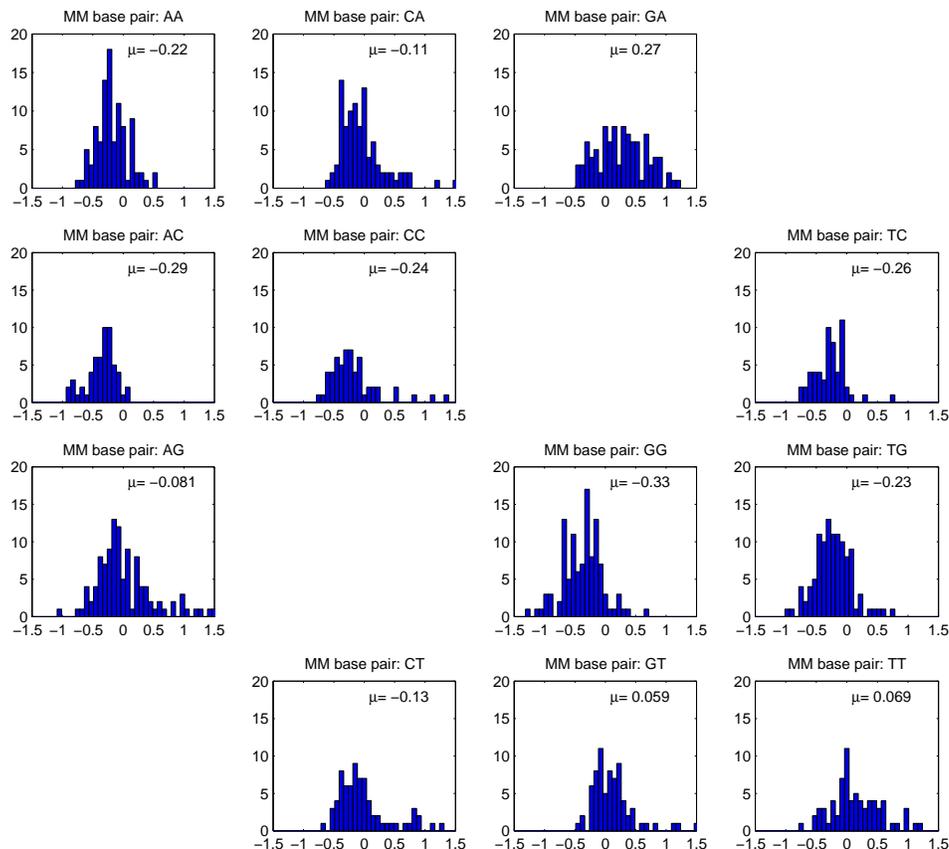


Figure A.10: Single base insertion defect profiles. For details see Fig. 6.15.

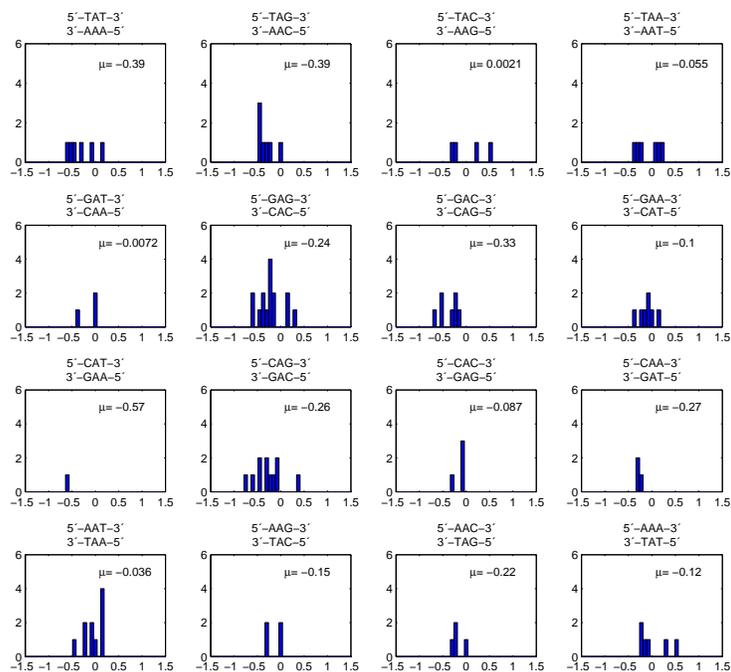
### A.1.3 Single Base Mismatches in DNA/DNA Duplexes - Statistical Analysis to Investigate the Influence of the Flanking Base Pairs



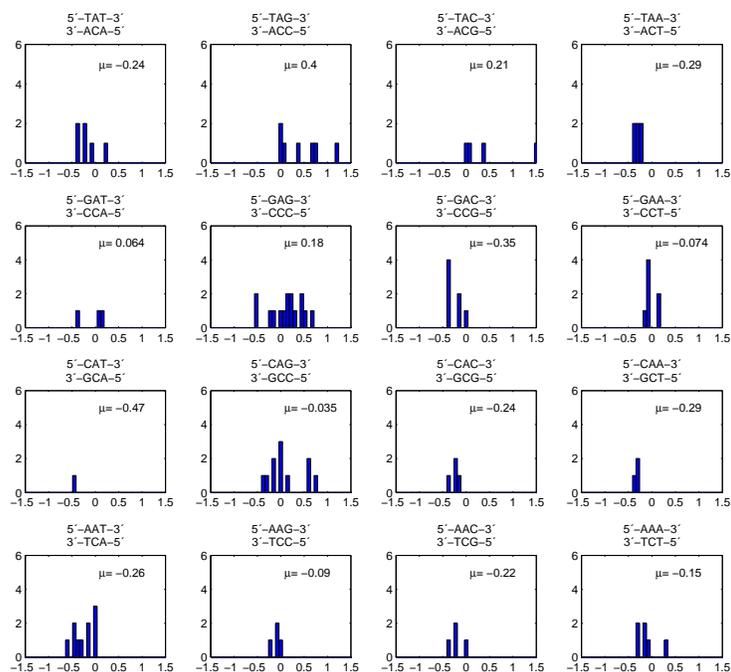
**Figure A.11:** All mismatch base pair types  $X \cdot Y$ . Measured hybridization signal distributions (occurrence versus deviation of the particular hybridization signal from the mean profile) as a function of the MM base pair alone, i.e. independent of the flanking base pairs.  $\mu$  denotes the median value of the distributions. A box-whisker plot of the distributions is shown in Fig. 6.6.

On the following pages (Figs. A.12 - A.22) this data is categorized according to the type of flanking base pairs. Owing to the restricted set of target sequences available for this study the sizes of the data sets measured for the individual defect configurations are very different.  $\mu$  denotes the median values of the distributions.

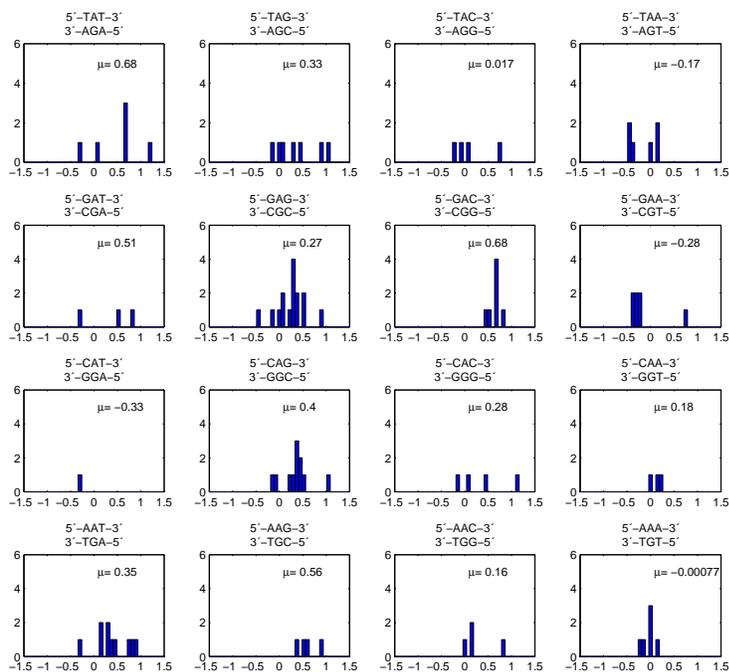
The median values of the nearest neighbor pair dependent subsets are compared in Fig. 6.9.



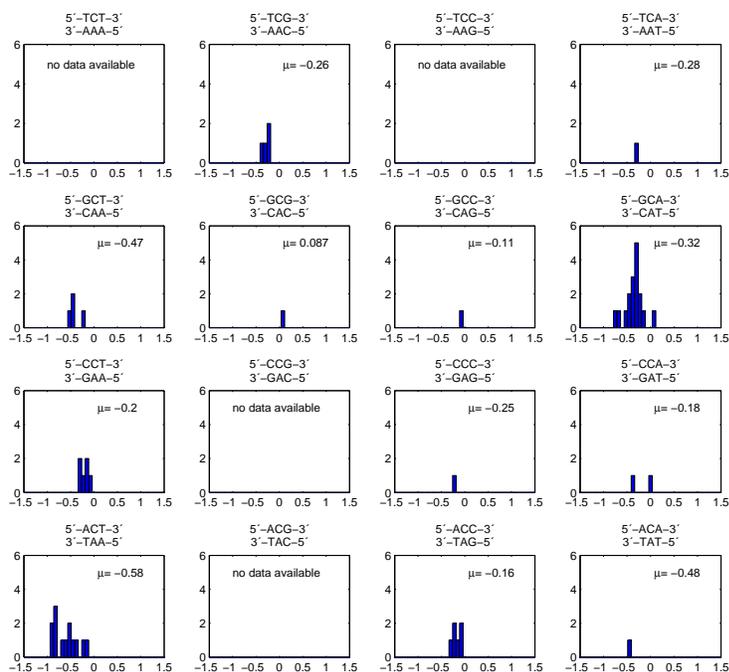
**Figure A.12:** A·A mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



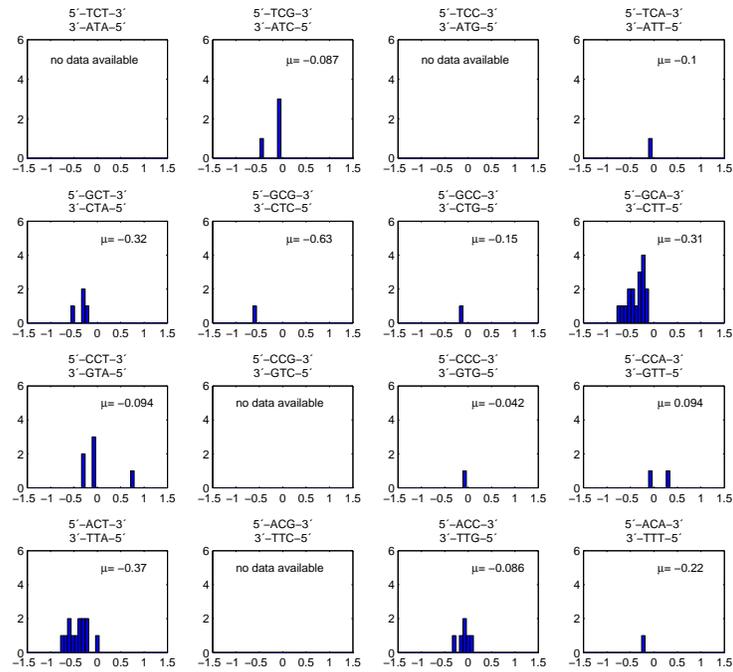
**Figure A.13:** C·A mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



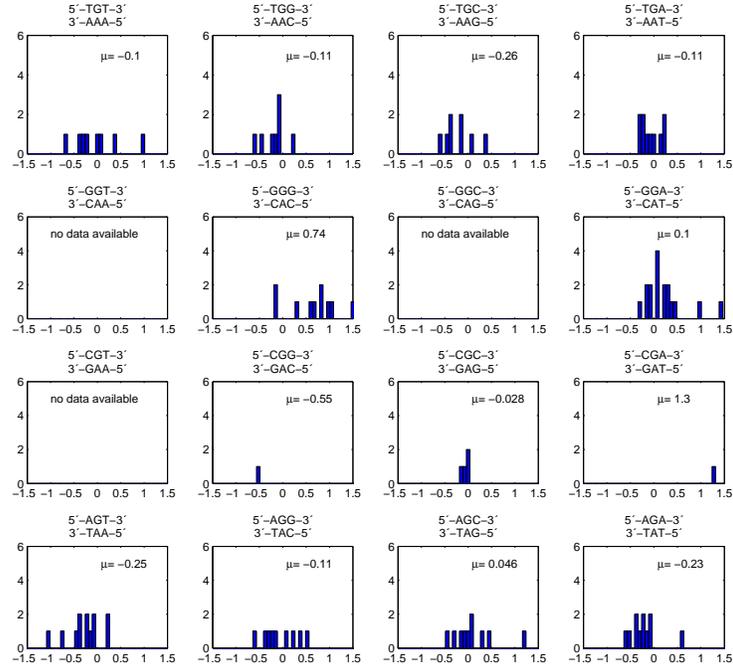
**Figure A.14:** G·A mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



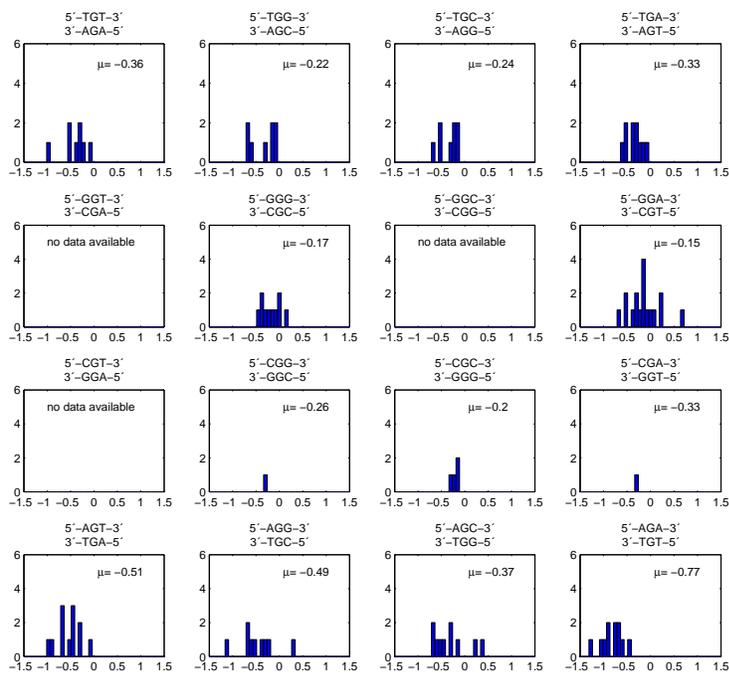
**Figure A.15:** A·C mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



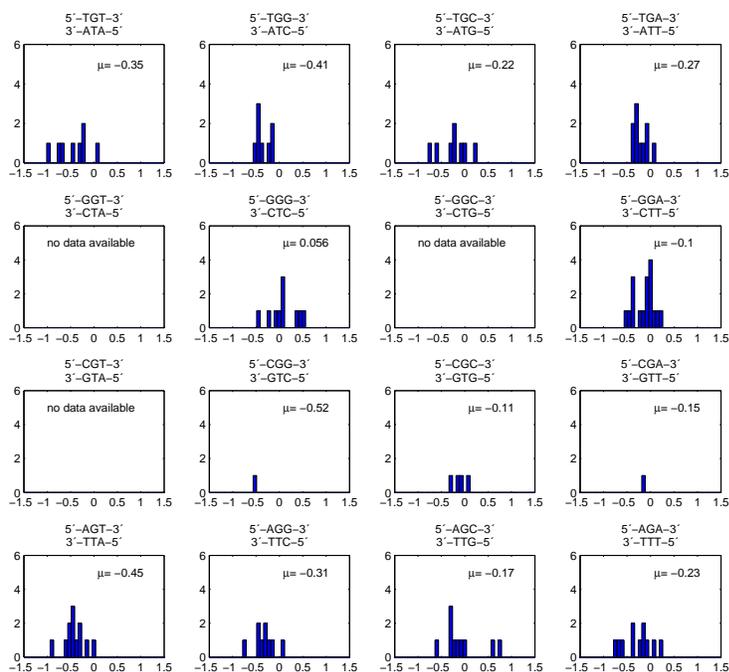
**Figure A.16:** T-C mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



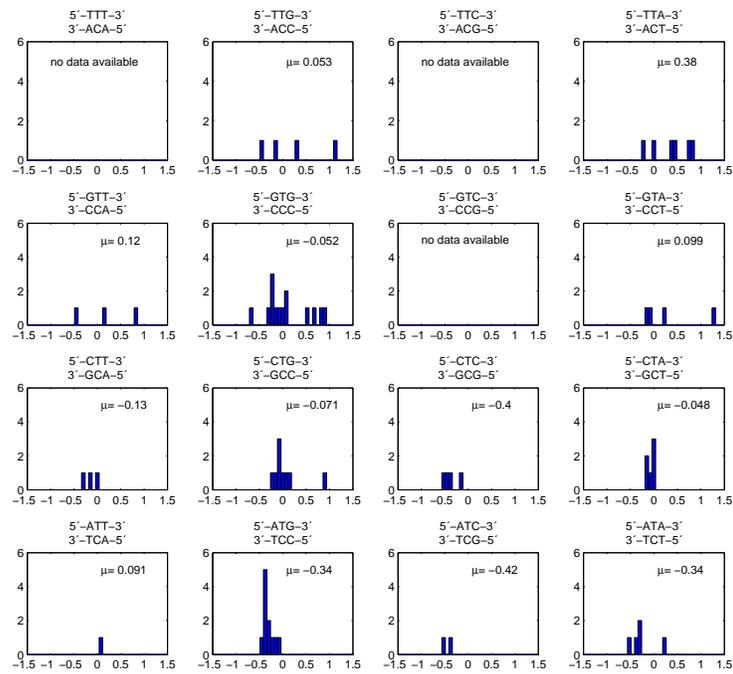
**Figure A.17:** A-G mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



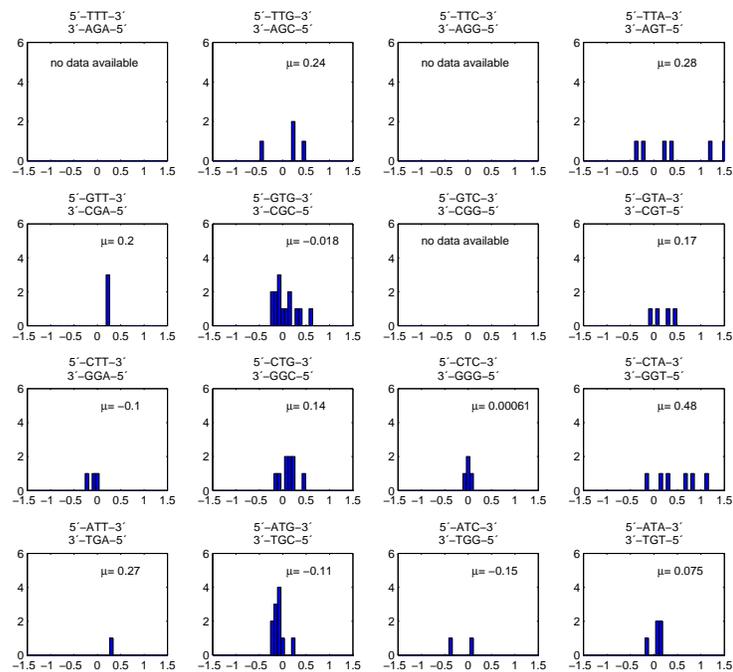
**Figure A.18:** G·G mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



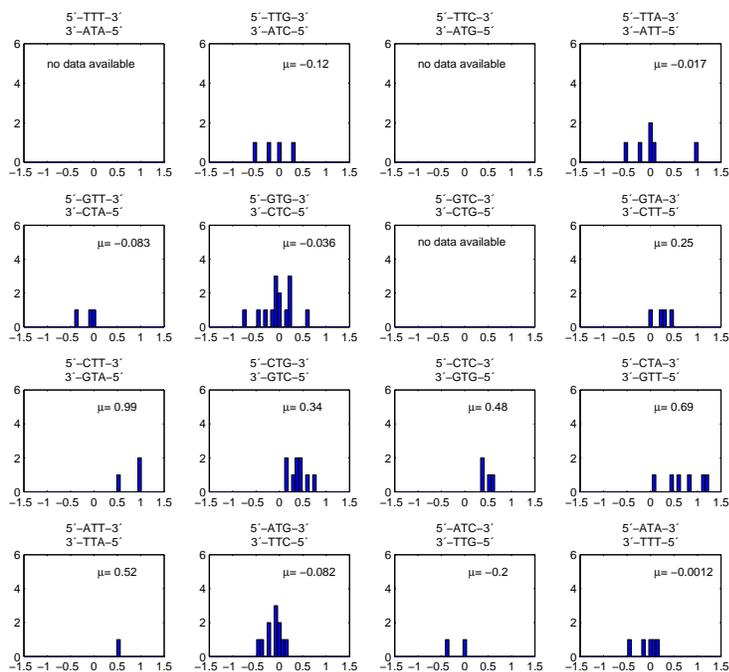
**Figure A.19:** T·G mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



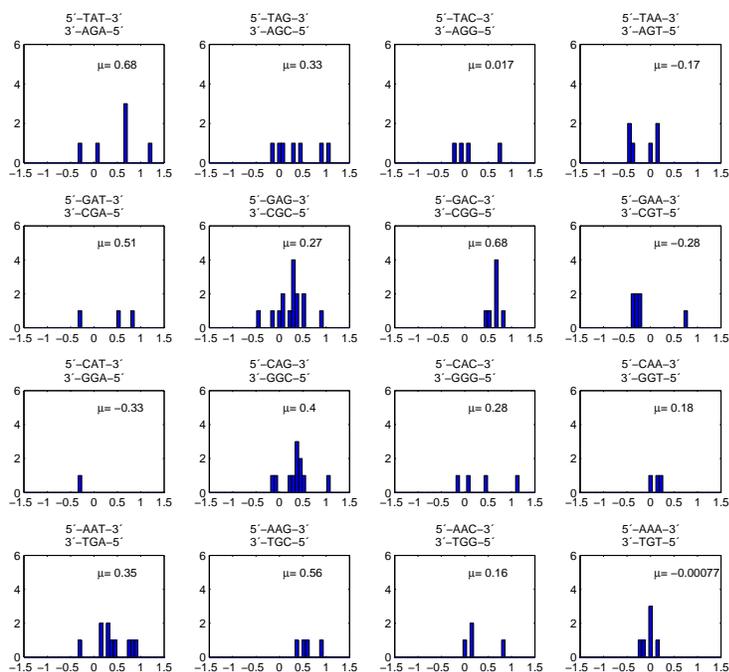
**Figure A.20:** C-T mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.



**Figure A.21:** G-T mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.

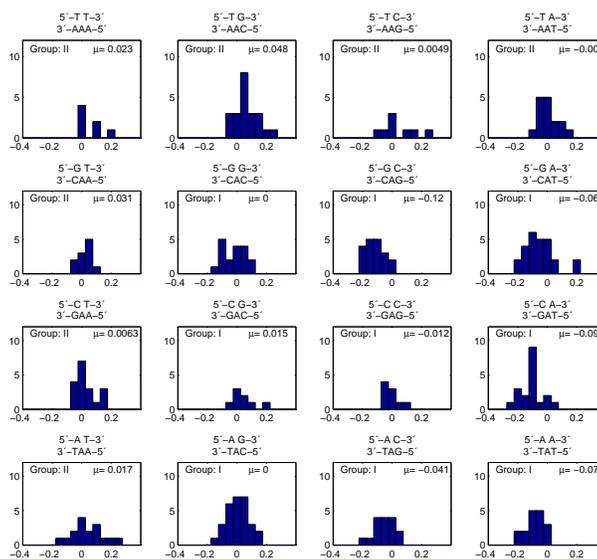


**Figure A.22:** T-T mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.

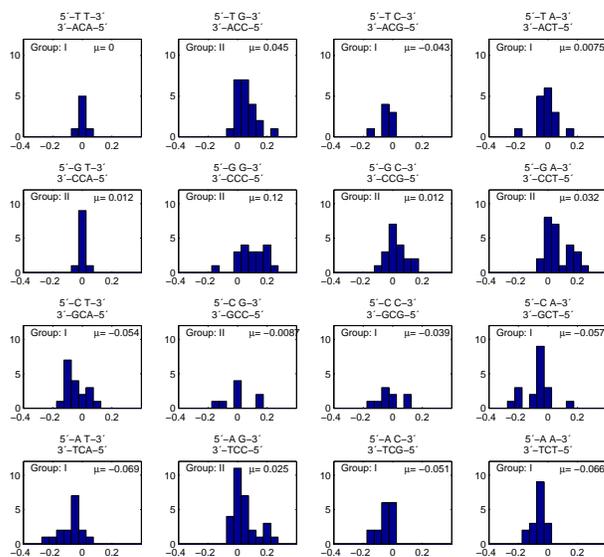


**Figure A.23:** G-A mismatches. Measured hybridization signal distributions categorized according to the flanking base pairs.

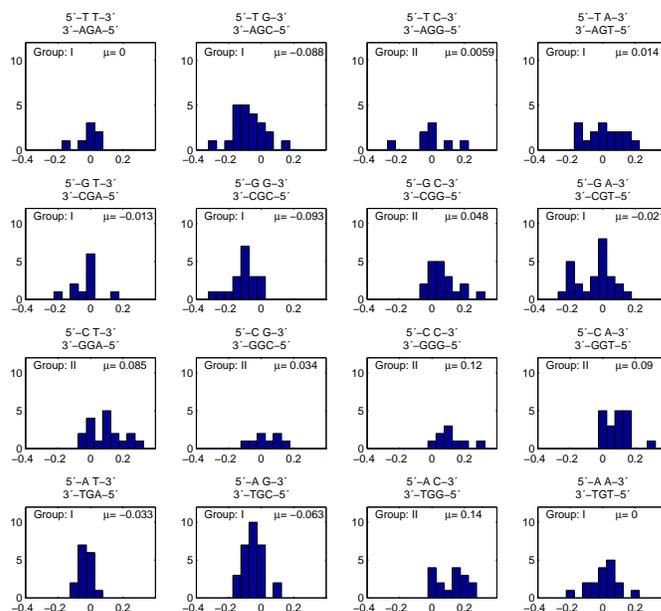
## A.1.4 Single Base Insertions - Statistical Analysis



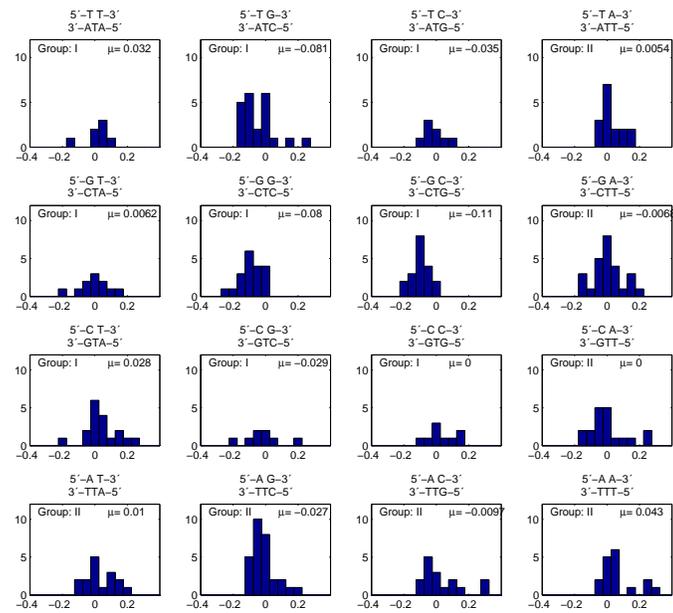
**Figure A.24:** Insertions of adenine bases - influence of the neighboring base pairs. Distribution of hybridization signal intensities (deviation from the mean profile in a.u.).  $\mu$  denotes the median value of the distribution. Group II insertions have at least one identical neighbor base, whereas Group I insertions don't have an identical neighbor. Group II insertions have consistently increased hybridization signals compared to Group I insertions.



**Figure A.25:** Insertions of cytosine bases - influence of the neighboring base pairs. Distribution of hybridization signal intensities (deviation from the mean profile in arbitrary units).



**Figure A.26:** Insertions of guanine bases - influence of the neighboring base pairs. Distribution of hybridization signal intensities (deviation from the mean profile in arbitrary units).



**Figure A.27:** Insertions of thymine bases - influence of the neighboring base pairs. Distribution of hybridization signal intensities (deviation from the mean profile in arbitrary units).

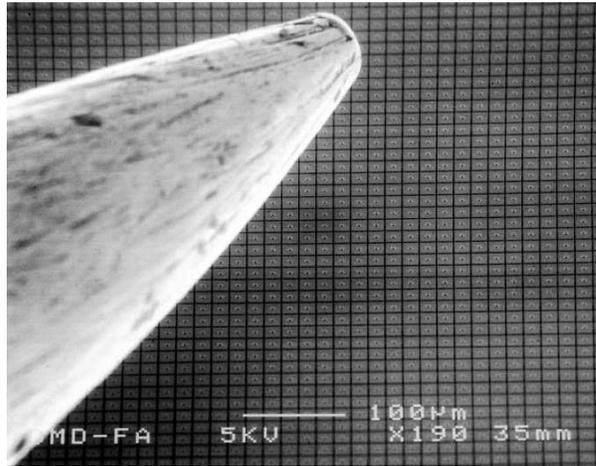


# Appendix B

## Supporting Information

## B.1 The Digital Micromirror Device (DMD™)

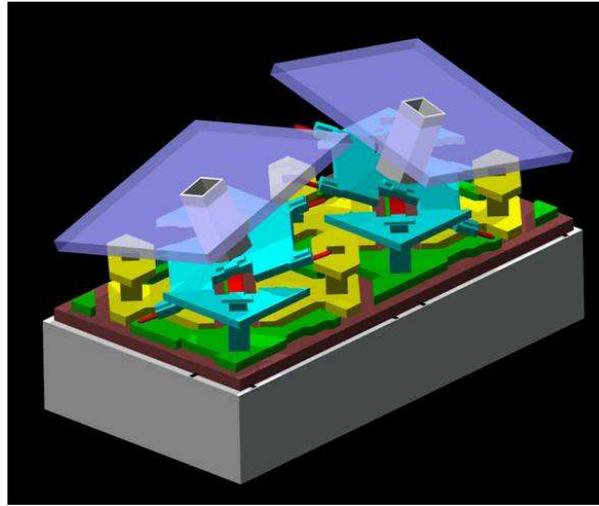
The DMD is an opto-electromechanic spatial light modulator which is commonly employed in video projection systems. The DMD (developed by Texas Instruments Inc.) comprises an array of tiny (16 micron sized) tilting mirrors, each corresponding to a single image pixel. A DMD with XGA resolution comprises  $1024 \times 768 = 786432$  of these individually addressable micromirrors (Fig. B.1).



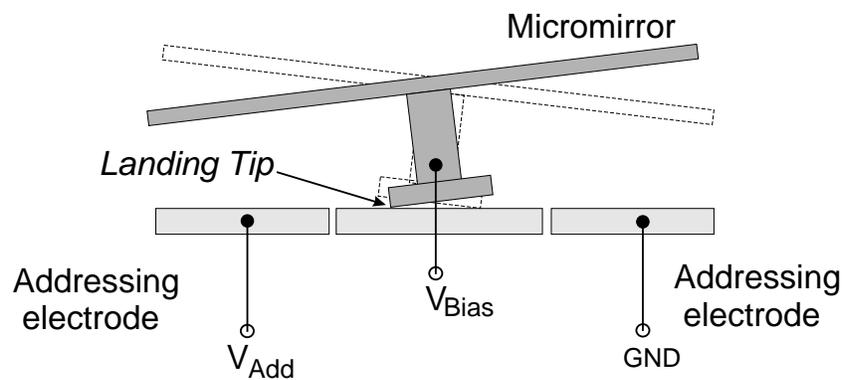
**Figure B.1:** Electron micrograph of a digital micromirror device. A pinhead is shown for size comparison. (Photo: Texas Instruments)

The DMD is a micro-opto-electromechanic system (MOEMS) which is produced with standard semiconductor fabrication techniques (photolithography, etching etc.). Each single mirror, which has a reflective surface of aluminum, is sitting on a torsion bar only a few microns in size (Fig. B.2). By applying a small torque the mirror can be tilted by an angle of  $\pm 10^\circ$  with respect to the DMDs normal axis. The torque is created by electrostatic forces between the mirror and the addressing electrodes beneath (Fig. B.3). The addressing electrodes are connected to SRAM memory cells under each the corresponding micro-mirror. Since the addressing voltage of the SRAM cell (5 V) is not sufficient to reorientate the mirrors, the DMD is operated in a bistable mode, in which a bias voltage of about 26 V is applied to the mirrors. For reorientation of the mirrors (all mirrors simultaneously) the image information is loaded into the array of SRAM cells beneath the mirrors. Then, the temporary removal of the bias voltage allows all mirrors to reorientate into the positions determined by the addressing voltage of the SRAM cells. Reestablishment of the bias voltage latches the mirrors in their new positions.

Depending on the mirrors orientation the light originating from the illumination system is either reflected towards the projection optics (the mirror is ON) or into a light trap (if the mirror is in OFF-position). The technique is therefore called Digital Light Processing



**Figure B.2:** Schematic of the DMD. Each micromirror (coated with reflective aluminum) (purple) is attached to a support post (grey) which is connected with the yoke (cyan) and with the flexible torsion hinge (red). The SRAM cell below each mirror determines the potential of the address electrodes (yellow) and thus (via electrostatic attraction between the mirror and the address electrodes) the orientation of the mirror. Tilt motion is limited to  $\pm 10^\circ$  by the yoke (cyan) touching the landing site. (Image: Texas Instruments)



**Figure B.3:** Working principle of the DMD. Electrostatic forces between the mirror and the addressing electrode result in tilt the mirror around the axis of the torsion hinge. The tilt angle ( $\pm 10^\circ$ ) is limited by the landing tip touching the landing site. See text for further details.

(DLP®). Pixel brightness is determined by pulse-width-modulation (mirrors are switching at a high frequency between ON- and OFF-position). The switching time between ON and OFF state is about 15  $\mu$ s. Compared to liquid crystal (LC)-SLMs, DMDs provide a high contrast ratio combined with small intensity losses (due to polarizers, filters etc.). DMDs are also suitable for UV irradiation (and thus can be used in photolithography applications, e.g. light-directed synthesis of DNA microarrays) whereas liquid crystal polymers quickly degrade under UV exposure.

## B.2 Modification of the DLP Video Projector for use of the Spatial Light Modulator in Photolithography Applications

The DMD system has been obtained from an A+K AstroBeam 540 DLP video projector.<sup>1</sup> The integration the optimized illumination- and projection-optics of the video projector in the UV photolithography setup (interesting due to the high light collection efficiency and illumination uniformity) soon turned out to be inappropriate because of the high UV absorption in the optical train.

To improve the positioning of the DMD the short connector between the DMD board and the DLP electronics main board was replaced by a 40 cm long 148 pin extension cable (Fig. B.4). Removal of the electrical shielding doesn't seem to affect the function of the DMD. The DMD board has been rotated by 45° so that the tilting axis of the micromirrors is oriented in vertical direction. Thus, the vectors of incident light and reflected light are oriented in a horizontal plane. The 120 W UHP lamp (parabolic reflector) of the AstroBeam projector has been transferred into an external housing. Replacement of this lamp by a more powerful 250 W UHP lamp (ellipsoidal reflector) required by-passing the control electronics. Since the video projector mainboard expects a confirmation of the lamp operation via the lamp power supply, the lamp-operation signal needs to be provided manually.

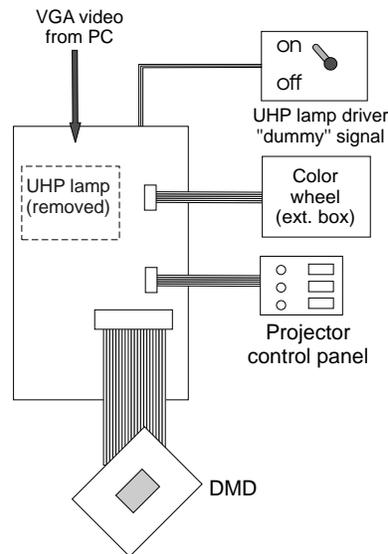
### B.2.1 Gamma-function of the DLP-Projector

The intensity response of display devices (e.g. monitors or video projectors) on image brightness values provided from the computers graphic hardware is determined by the devices *gamma function*.

Personal computer display hardware is currently restricted to 24 bit color depth (8 bits for

---

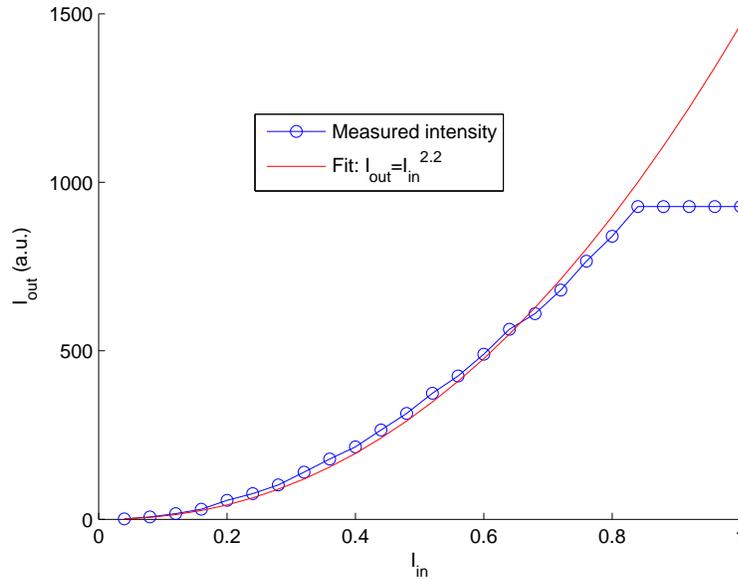
<sup>1</sup> The A+K AstroBeam 540 DLP video projector is very similar to the DAVIS DL X10.



**Figure B.4:** Modifications of the DLP video projector. Since the UHP lamp is removed a "dummy signal" pretending lamp operation, has to be provided to the projector electronics. The dummy signal has to be switched "on" manually ca. 5 s after pressing the power-button (on the control panel). The color wheel - a fast spinning filter wheel - provides a feedback signal to the DLP electronics main board - thus cannot be removed. For safety reasons it has been accommodated in an external box. For better accessibility the DMD board (carrying the Digital Micromirror Device) is separated from the electronics main board (which is fixed on the projector chassis).

each of the 3 color channels). Thus there are in total 256 grey levels available.

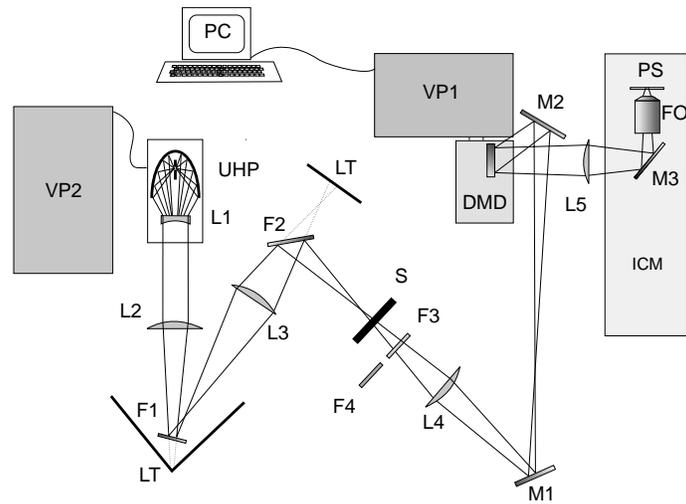
For a variety of experiments it may be interesting vary the exposure intensities by using grey level masks. The DMDs intensity response on the grey level intensity value of a full screen image was measured with a photometer which was located at the focus of the microscope objective (microprojection setup). The *gamma function* of the DMD (light intensity versus image grey level value) is shown in Fig. B.5. The nonlinear response is described by a 2.2 *gamma curve* (power law with an exponent of 2.2), that is typical for many display devices. At large brightness values the intensities are cut off. The cut-off level is determined by the contrast value set in the projectors settings menu.



**Figure B.5:** *Gamma function* of the AstroBeam projector. The intensity response  $I_{out}$  on the image brightness  $I_{in}$  (normalized on a maximum value of 1) follows a power law with an exponent of 2.2. For an image brightness larger than about 80% of the maximum value a cut-off is observed. The position of the cut-off depends on the contrast and brightness values chosen in the AstroBeams "Display Settings Menu".

## B.3 Optics of the Microscope Projection Photo-lithography System

- UHP: Philips UHP-lamp 250W 1.35 TOP 222 H4 elliptical reflector elliptical reflector geometry: major axis  $\sim 80$  mm, minor axis  $\sim 50$  mm)
- L1: plano-concave lens:  $f=50$  mm, diam. 25 mm (silica), placement between UHP lamp window and the outer focal point of the elliptical reflector
- L1-L2: 145 mm
- L2: plano-convex lens:  $f=50$  mm, diam. 50 mm
- L2-F1: 120 mm
- F1: UV cold mirror (UV barrier filter from the Optoma projector lamp module)
- F1-L3: 165 mm
- L3: plano-convex lens (BK7):  $f=100$  mm, diam. 50 mm
- F1-F2: 215 mm
- F2: UV cold mirror (Oriel)
- F2-F3: 165 mm
- F3: UV band pass (bk-370-35-B, Interferenzoptik Elektronik GmbH), diam. 25.4 mm
- F2-L4: 250 mm
- L4: plano-convex lens (BK7):  $f=125$  mm, diam. 50 mm
- L4-M1: 170 mm



**Figure B.6:** Schematic of the microscope projection photolithography system.

- M1: mirror
- M1-M2: 380 mm
- M2: mirror
- M2-DMD: 60 mm
- DMD-L5: ca. 164.5 mm, to be fine-adjusted
- L5: tube lens, Carl Zeiss,  $f=164.5$  mm
- M3: mirror/beam splitter

## B.4 Fabrication of the Synthesis Cell



**Figure B.7:** Punching tool (top) for the fabrication of the PDMS gasket (center). The tool, producing a diamond-shaped cutout (the cell volume) with clean edges, is essential for smooth operation of synthesis apparatus. Wire-cut EDM (electrical discharge machining) has been employed for producing the sharp-edged structure in hardened steel. Dimensions of diamond-shaped cell volume: length 16 mm; width 5 mm. The outer edge of the gasket was cut with another (smaller) version of the punching tool.

Part names are referring to Fig. 3.13.

- The top-plate is made from a 10 mm thick plate of transparent Makrolon<sup>®</sup> plastics (polycarbonate). Produce four tapped holes for fastening screws (not too far away from the center of the plate, to enable proper sealing action). Further, two holes for fastening the cell-assembly on the projection lithography setup are required.
- Inlet and outlet tubes are made from syringe needles (0.9×40 mm). By using a drilling machine as a "lathe" the plastic adapter of the syringe needle is reduced to a cylindrical bit as shown in Fig. 3.13.
- Produce holes for inlet/outlet needles. (diam. 1 mm on the upper side of the top plate). At the bottom side of the top-plate the needle (blunt end near the coupling) should protrude 1 mm. The needles are fastened with epoxy glue.
- To obtain a transparent and chemically inert (solvent resistant) surface, a glass microscopy slide is glued onto the lower side of the top-plate. Before gluing (with transparent PDMS silicone rubber), the slide needs to be cut in 3 pieces to produce gaps for the fastening screws. Moreover, two 1 mm diam. holes for the inlet/outlet tubes have to be drilled into the glass slide by using a diamond tool. By gluing the glass slide onto

the top-plate the gaps between the needles and the glass are sealed with PDMS (avoid getting PDMS into the needles!). PDMS (Dow Corning Sylgard<sup>®</sup> 184) was purchased from World Precision Instruments.

- The bottom-plate is made from 5 mm aluminum. The exposure window should not be too large (ideally implemented as a long hole) to achieve proper sealing action by pressing the Chip-substrate/PDMS-gasket against the top-plate.
- Fabrication of the PDMS-gasket: PDMS Sylgard 184 (Dow Corning) is mixed thoroughly (ratio between elastomer base and curing agent: 10:1), degassed and poured into a glass petri dish. Curing for 20 minutes at 80°C. A custom-made punching tool (Fig. B.7) is used to produce the streamlined cutout forming the synthesis volume.
- Connectors: PFA (PTFE) tubes (internal diam. 0.8 mm) fit tightly on the 0.9 mm diam. syringe needles. PTFE tube end fittings (UNF 1/4" 28 G) provide a removable connection with the fluidics system.

## B.5 Technical Notes on Light-directed DNA Chip Synthesis

### B.5.1 Handling of Phosphoramidite Reagents

The coupling efficiency of phosphoramidite reagents is very sensitive to contamination with (even trace amounts of) water. To maintain low moisture conditions the following precautions should be considered:

- Storage under moisture free conditions at  $-20^{\circ}\text{C}$ . Use dry argon atmosphere and desiccant.
- Open storage bottles only in glove box under dry argon atmosphere. Use silica gel beads to maintain a low moisture content in the glove box.
- Use oven-dried glass ware to minimize surface-adsorbed water.
- Dissolve phosphoramidites only immediately before synthesis.
- Use dry MeCN with  $<10$  ppm of water.
- Use molecular sieve bags (in the MeCN storage bottle and in the activator solution) to adsorb water from the solvent.
- Phosphoramidite solutions should be used the same day as prepared.

Solution stability and degradation pathways of deoxyribonucleoside phosphoramidites in MeCN are discussed in [Kro04].

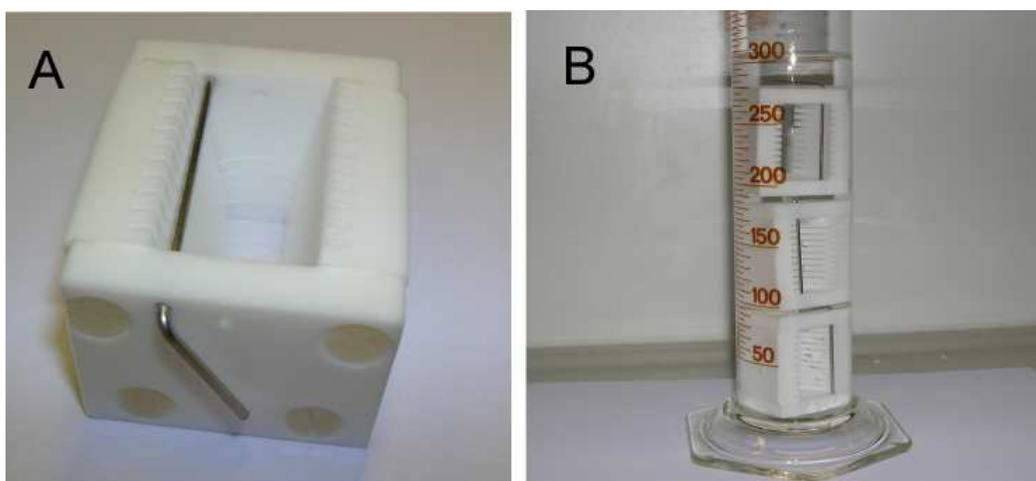
### B.5.2 Additional Notes on the Synthesis

Prior to the first phosphoramidite coupling the substrate is soaked in MeCN for about 2 minutes. The initial coupling is performed for 1 minute and then repeated once. According to Richmond *et al.* [Ric04] an increase of the coupling time (of the first base only) from 20 s to 6 h resulted in an 80% increase in the amount of full-length probes.

Coupling and exposure time, washing steps and image quality are the key parameters for high quality synthesis. According to [Ric04] the number of error-free probe sequences could be increased 100-fold by making several technical improvements on their synthesis apparatus. Improvements include the extension of the coupling time from 20 to 60 s and of the exposure time from 50 to 150 s, additional argon drying steps and modifications on the projection optical system (image-locking).

Upon prolonged exposure the solvents tetrahydrofuran (THF) and pyridine cause significant swelling of the PDMS gasket. Exposure to these solvents (contained in oxidizer and capping reagents) should therefore be minimized.

## B.6 Technical Notes on Microarray Dendrimer Substrate Preparation

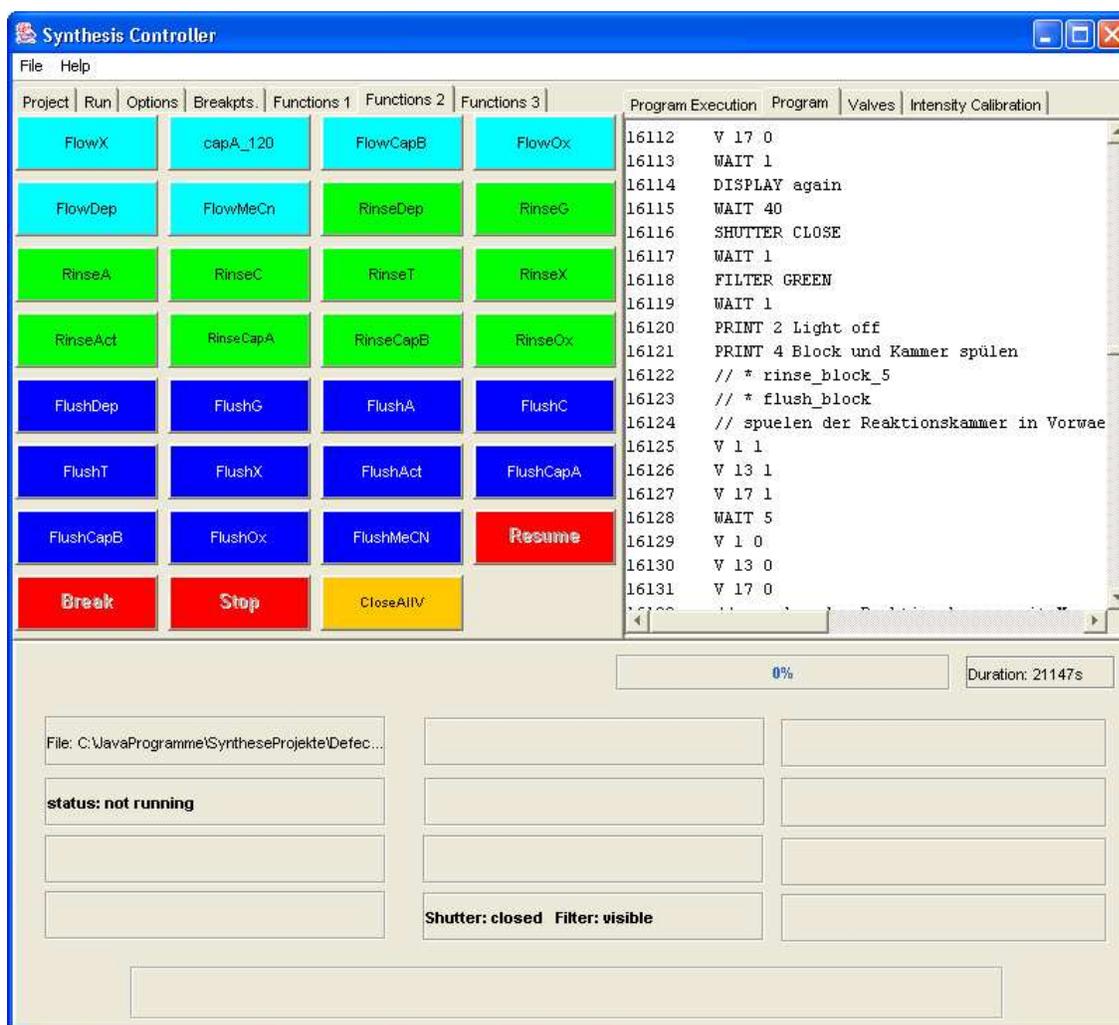


**Figure B.8:** (A) Teflon slide holder for up to 12 round cover glasses. The stainless steel pin secures the glasses. For use with dichloroethane the nylon screws should be replaced by stainless steel screws. (B) Substrate functionalization in a 500 ml graduated cylinder requires about 250 ml reagent solution.

- For dendrimer functionalization of the microarray substrates a compact slide holder for handling of up to 12 cover glasses was developed. Parts of the teflon (PTFE) slide holder are assembled with stainless steel screws and can thus withstand a bath in dichloroethane solution. The holder enables fast and thorough washing and drying of the slides. Use of the holders resulted in significantly increased quality of the substrates and enabled reduction of the reagent consumption.
- To minimize reagent consumption (ethanol analytical grade, dendrimers in dichloroethane) the substrate functionalization is performed in a 500 ml graduated cylinder. Three slide holders (with 36 slides in total) are immersed in about 250 ml of solution.
- Drying of the slides under a nitrogen stream should be performed in such a way that the liquid is blown away from the center of the slides. Drying of droplets on the surface has to be avoided because this can produce irremovable stains.

## B.7 Technical Notes on the Synthesizer Control Software *DNASyn*

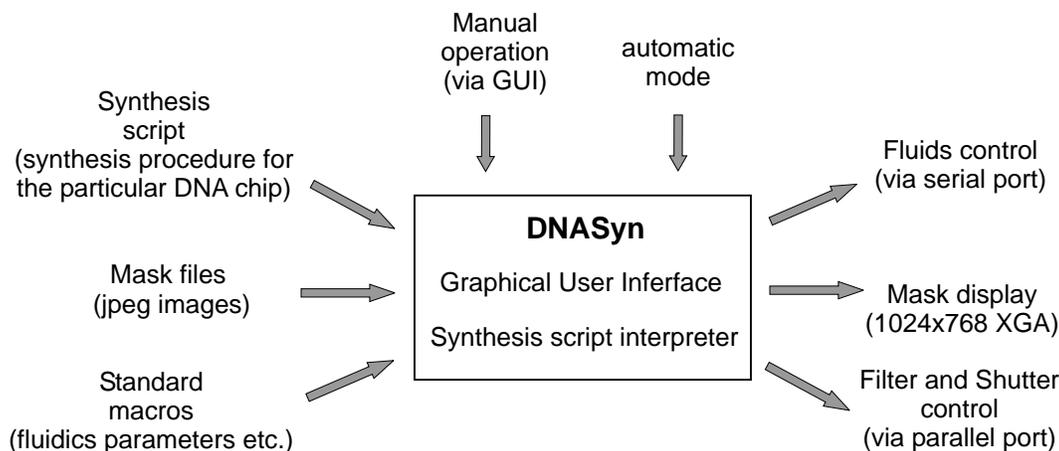
The light-directed fabrication of a DNA microarray has been fully automated. The synthesizer control software *DNASyn* integrates control of the fluidics system with the maskless microphotolithography system (including image display, shutter and filter control).



**Figure B.9:** The graphical user interface of *DNASyn*. The buttons in the left panel enable manual access to user-defined macro functions. The textbox at the right shows the code of the synthesis script loaded.

*DNASyn* was implemented in Java<sup>TM</sup>. It is running with Windows XP Professional (and is also expected to work with Windows 98). The use with Windows XP Home or Windows Vista is not recommended since these operating systems won't allow direct access to the hardware ports via the kernel mode driver *UserPort*.

## B.7.1 Basic Features



**Figure B.10:** Concept of the *DNASyn* microarray synthesis control software.

*DNASyn* includes a flexible macro programming language for the automated control of the synthesis process, and a graphical user interface (GUI) for manual control of various synthesizer functions (see Fig. B.9). The macro language comprises only a small number of basic commands.

### Keywords

START	Begin of the main program
END	End of the main program
MACRO <i>macroname</i> {...}	Macro header
PRINT <i>n note</i>	DNASyn shows text <i>note</i> in output line <i>n</i>
// <i>comment</i>	Comment in the source code
WAIT <i>n</i>	Wait for <i>n</i> seconds
V X Y	Valve operation X: valve number ; Y: 0=close 1=open
DISPLAY <i>imagename.jpg</i>	Virtual mask display
DISPLAY AGAIN	Display the previous image again
SHUTTER <i>ON/OFF</i>	Shutter control
FILTER <i>GREEN/UV</i>	Filter changer control

- Switching of solenoid valves (fluidics operations) is performed with the V X Y command.
- The DISPLAY *imagename.jpg* command loads the JPG image from the synthesis directory and shows it on the DMD. The keyword AGAIN is used to reload the previous image.
- The WAIT *n* command (*n* duration in seconds) is used for time control of the synthesis processes.

- Comments begin with // followed by a space character.

## Macros

Typical routines (e.g. amidite coupling or photo-deprotection) can be combined to macro commands, as shown in the following example.

```
MACRO rinse_20
{
//Rinse synthesis cell with MeCN for 20 s - this is a comment
V 18 1
V 2 1
V13 1
V17 1
WAIT 20
V 2 0
V13 0
V17 0
V18 0
}
```

Macro commands can be called from the main program and from within other macros. Manual control (via button-click in the control panel) is also based on macro commands. Most control panel buttons are assigned a macro function. Macro codes for these functions are listed (and can be modified if necessary) in the file *functions.prg*.

A synthesis program comprises a list of macros (a library of *standard macros* and additional user-defined macros) and the *main program*. Standard macros describe routine synthesis processes. Basically they are not different from user-defined macros, but since they include critical time parameters (duration of fluidics processes, exposure times etc.) and since they may be called from other macros, modifications in standard macros should be considered cautiously. Upon loading a synthesis program (file extension *.prg*) the parser of *DNASyn* initially reads the main program (between the commands *START* and *END*). In the next step macro calls are substituted by the corresponding macro codes. To consider nested macros this is repeated until all macros are resolved. A completely resolved synthesis program for a 25mer array synthesis typically comprises about 40000 commands.

## Frequently used macro functions

flush	flush synthesis cell with argon
flow_X	flow reagent X through the synthesis cell
rinse_X	fill MeCN into the storage bottle for reagent X

rinse_block	rinse valve block with MeCN
flush_block	flush valve block with argon
prime_X	fill the tube between the storage bottle X and the valve block with reagent X
reverse_flush	fast flush of the synthesis cell with argon in reverse direction
deprotect	photodeprotection
couple_X	coupling of the phosphoramidite X
oxidize	oxidization of phosphite bonds

Number-extensions to the functions name (e.g. flush10) specify the duration of the operation (in seconds).

### B.7.2 Communications between the Control PC and the Synthesizer Hardware

For serial communication with the solenoid valve controller the Java Communications API (Sun Microsystems) is employed. The communications parameters have been set to the requirements of the valve controller (see below).

The control of the shutter and filter-changer via the parallel port has been implemented with a Java native code. Direct control of the parallel port requires the java package *parport*. The library *parport.dll* needs to be installed in the directory `Systems32/drivers`. With *parport* the channels of the parallel port can be set and read in a straightforward way. For direct access on the I/O ports (user mode) the driver *UserPort* (written by Tomas Franzon) needs to be installed (for this purpose *Userport.sys* needs to be copied to `System32/drivers`). Possibly the Windows98 compatibility mode needs to be enabled. With the executable *Userport.exe* the access to the parallel port (base address \$387) is set enabled.

### B.7.3 Dual Screen Support

*DNASyn* provides dual screen support to display the control panel and the photolithography mask patterns on different devices - TFT monitor and video projector (DMD), respectively. This requires the use of a dualview graphics card and extension of the Windows desktop onto the second display. The control panel is displayed on the primary screen (TFT-monitor with 1280×1024 pixels). Display of the photolithography masks on the secondary display (video projector) is achieved by opening a window at the corresponding desktop coordinates - no further programming tricks are necessary. The Class `DisplayFrame`, an extension of the Java Class `JWindow` enables display of the masks without

a window frame and without a menu bar. In the initialization method of `DisplayFrame` the command `this.setLocation(1280,0)` opens the `JWindow` on the extended desktop in the area covered by the secondary TFT display.

### B.7.4 Additional Hints

- Mask files (JPEG format required) have to be copied into the same directory (synthesis folder) as the synthesis program.
- Avoid compression of the JPG mask images. Compression losses (gray pixels) result in a grainy structure of the microarray features.
- Display of several images at an interval of less than about 5 s may result in delays.
- Not to interfere with time critical procedures no further programs should be running on the control computer during a synthesis. In particular, Virus scanners should be deactivated. Disconnect the computer from the local area network.
- `DNASyn` should be executed as a jar-file (java archive) from `java.exe` (e.g. with the command `java -jar DNASyn.jar`). Even though this is possible, the fluidics timing behavior can be affected. In the Windows Task Manager an increased priority level should be given to the `java.exe` process. Do not select the highest priority level (real time) as this may affect the operating system stability.

## B.8 Solenoid Valve Driver for Fluidics Control

The valve block of the synthesizer is operated by the microcontroller-based solenoid valve driver *Elub 0670/01* developed by the electronics workshop of UBT. Operating a solenoid valve requires an initial spike voltage of 24 volts. After an adjustable time (100-500 ms) the voltage is reduced to a lower hold voltage of 8 V, to prevent overheating of the coil. Driver operation is controlled via serial communication (RS-232) with the control PC.

#### Serial communications parameters:

9600 baud, 8 data bits, no stop bit, no parity, no handshake

#### Connector (9-pin male) at the controller

Pin 2: TxD (Data out)

Pin 3: RxD (Data in)

Pin 5: GND

#### Connector (9-pin female) at the PC

Pin 2: RxD

Pin 3: TxD  
Pin 5: GND

Valve operation is controlled by transmission of ASCII commands via the PC-serial port RS232:

open valve: **eVenXX** (XX: valve number 1 to 36)

close valve: **aVenXX** (XX: valve number 1 to 36)

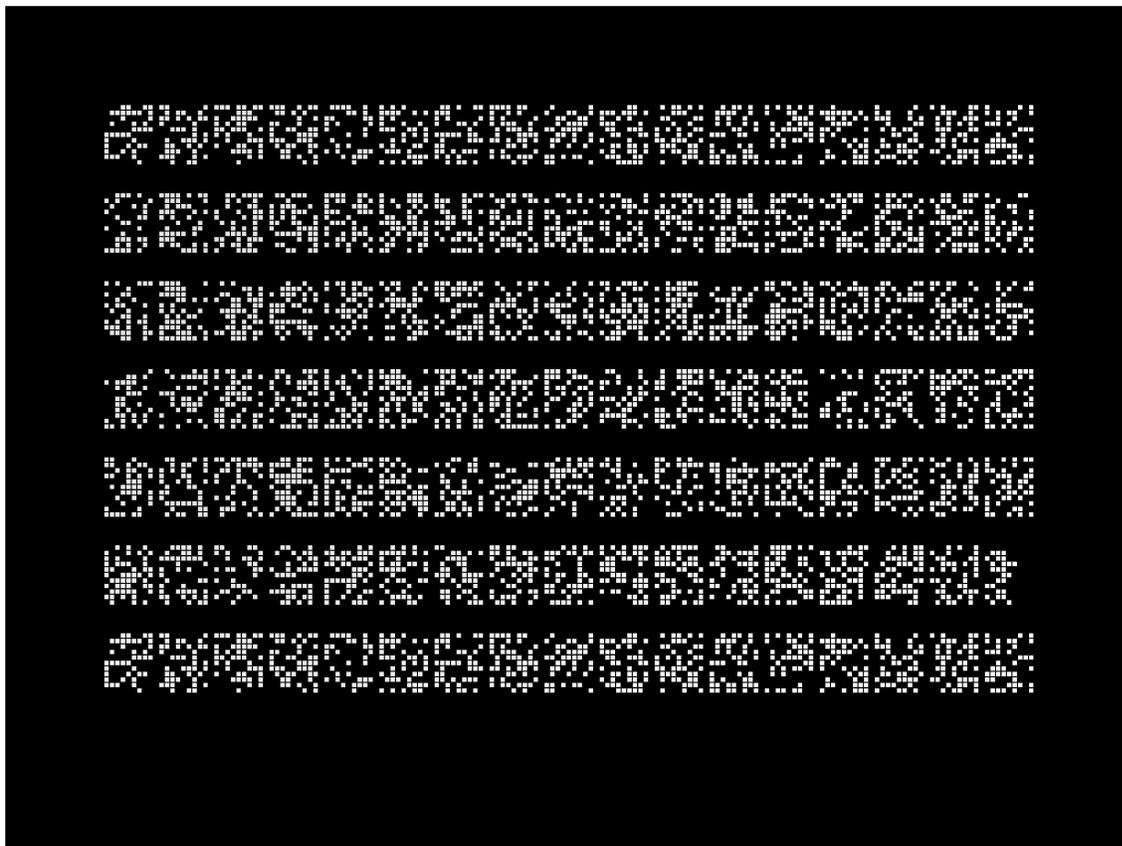
The command **?Ven** is used to request the valve status. The 36-digit answer string sent back to the control PC (e.g. 0010111100...) reports the states (1=open; 0=closed) of all 36 valves that can be addressed (not all in use).

## B.9 DNA Microarray Design

Synthesis masks (shown in Fig. B.11) are generated with the Java programs *MaskDesigner* and *MaskFileGen*. The software concept is described in Fig. B.12. With *MaskDesigner* microarrays can be designed manually via a graphical user interface. Microarray geometry (feature arrangement, feature size and spacing) is created with commands provided in the menu *Edit*. Once the geometry has been defined, the probe sequences (contained in sequences lists, ASCII-format) can be assigned to the feature blocks. Photolithography masks are generated with the command "Generate Masks" in the "Main" menu of *MaskDesigner*. The *MaskDesigner* software can also be employed as a viewer (*View* menu), e.g. to find particular sequences on the microarray or to investigate sequence similarity (e.g. to find the longest common subsequence).

The manual method of chip design is, however, somewhat tedious. To account for the fact that chip designs are often very similar, the Java program *MaskFileGen* has been developed. Based on a short *Chip script file*, in which the geometry of the individual feature blocks and the file path to the corresponding probe sequence lists is specified, the *Chip definition file* is generated. The *Chip definition file* format is equally used by *MaskDesigner* to save and read the data of manually designed chips.

Another, more efficient method is microarray design within MatLab: a MatLab program generates a single large sequence list, which is employed by *MaskFileGen* to generate the *Chip definition file*. The arrangement of the probes in form of feature blocks is considered in the MatLab program: feature blocks (two-dimensional cell arrays containing probe sequences in form of character arrays) are copied straightforward to the appropriate position in the master array (corresponding to the whole-chip-array defined in the *Chip script file*).



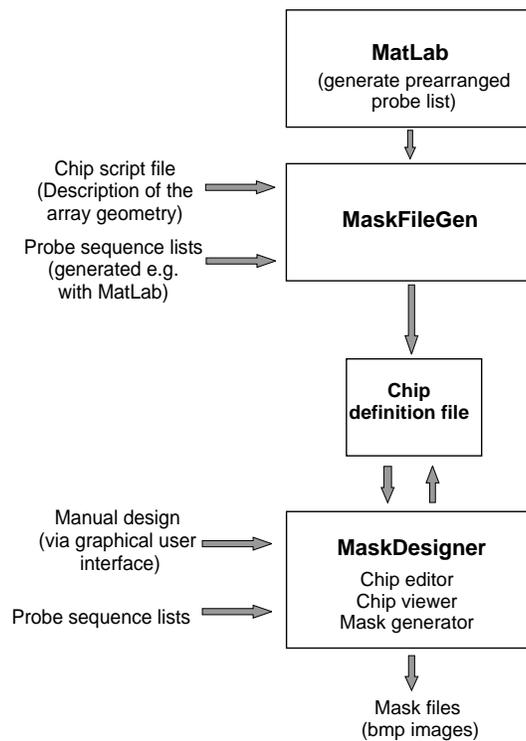
**Figure B.11:** The above example of a synthesis mask pattern (resolution:  $1024 \times 768$  pixels) generated with *MaskDesigner* comprises about 13000 features ( $4 \times 4$  pixels with a one pixel separation gap). About 80 mask patterns are required for the synthesis of a 25mer microarray. Synthesis masks are saved in jpeg format. Minimum compression is applied to prevent pixelation artifacts.

When the master array is complete, a linear list of probe sequences (to be used by *MaskFileGen*) is extracted from the master array. The feature block geometry is encoded in the order of the sequences. The sequence list in conjunction with a standard *Chip script file* is used by *MaskFileGen* to generate the corresponding *Chip definition file*.

For use with the synthesis software *DNASyn* the bmp image format produced by *MaskDesigner* needs to be converted into jpeg format. Conversion of the 80 images is performed with MatLab in a batch process. MatLab further has the advantage that it can generate jpeg images with almost no compression losses. This is important because compression artifacts (pixelation) can affect the quality of the synthesized microarray probes.

### Mask pattern generation with *MaskDesigner*

The principle of mask pattern generation is described in Fig. B.13: Synthesis masks  $M_i$  are affiliated to the corresponding photo-deprotection step  $D_i$ . For features represented in



**Figure B.12:** Synthesis masks are generated with the programs *MaskDesigner* and *MaskFileGen*. Chip designs are saved in the *Chip definition file* format. This file format is also the interface between the script-file based *MaskFileGen* and the graphical user interface based *MaskDesigner*. Arrays of probe sequences can also be created with MatLab: the prearranged probe sequence list (feature block structure encoded in the order of the sequence list) can be used with *MaskFileGen* to produce a *Chip definition file* which can be loaded by *MaskDesigner* to generate a set of synthesis masks.

white color on the synthesis mask  $M_i$  the corresponding microarray features are illuminated in the exposure step  $D_i$ . In the subsequent coupling step  $C_i$  a phosphoramidite building block can attach to deprotected probes. The *coupling sequence* (i.e. sequence of coupling steps  $C_i$  performed in the synthesis process) needs to be specified to generate the set of mask patterns for the corresponding deprotection (exposure) steps  $D_i$ . The default coupling sequence is TACG TACG TACG TACG... .

For a 25mer microarray synthesis in principle  $4 \times 25 = 100$  masks are required. However, with a little optimization typically 80 masks are sufficient. The optimization comprises that a coupling reaction on a particular feature, or more specific, the exposure of the particular feature prior to the next suitable coupling step, is performed as early in the synthesis as possible.



**Figure B.13:** Principle of virtual mask set generation. Different probe sequences located in features 1 and 2 - *probe sequence 1* and *probe sequence 2* - are treated individually. Probe sequences are compared with the *coupling sequence* to establish the *exposure sequence* for each feature. The exposure sequence determines in which mask patterns a feature is undergoing UV exposure (in the corresponding photomask the feature is shown in white). Mask patterns shown at the bottom - for simplicity - comprise two features only.

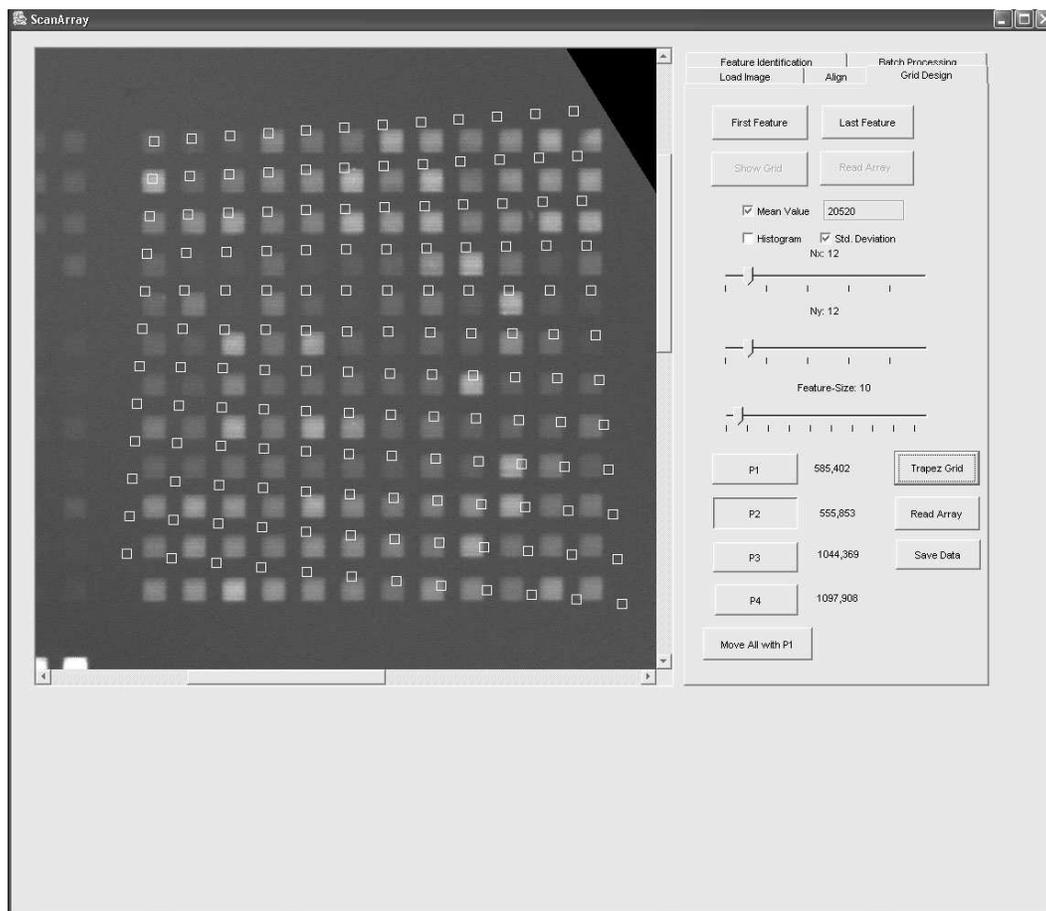
Example: In mask number 4 (employed in the photo-deprotection step prior to coupling of T) the feature corresponding to probe 1 is displayed in white, whereas the feature corresponding to probe 2 is displayed in black. Photo-deprotection of feature 1 in mask 4 (prior to coupling of T) results in coupling of T to the probes in feature 1.

## B.10 Microarray-Analysis with *ScanRA*

*ScanRA* has been developed for quantitative image analysis of microarray hybridization signals. The regular arrangement of the microarray features is employed for placement of a readout grid. Small distortions of the regular feature arrangement (e.g. due to optical distortions) are compensated by using a quadrilateral readout grid (see section 5.2).

### Functions of the software

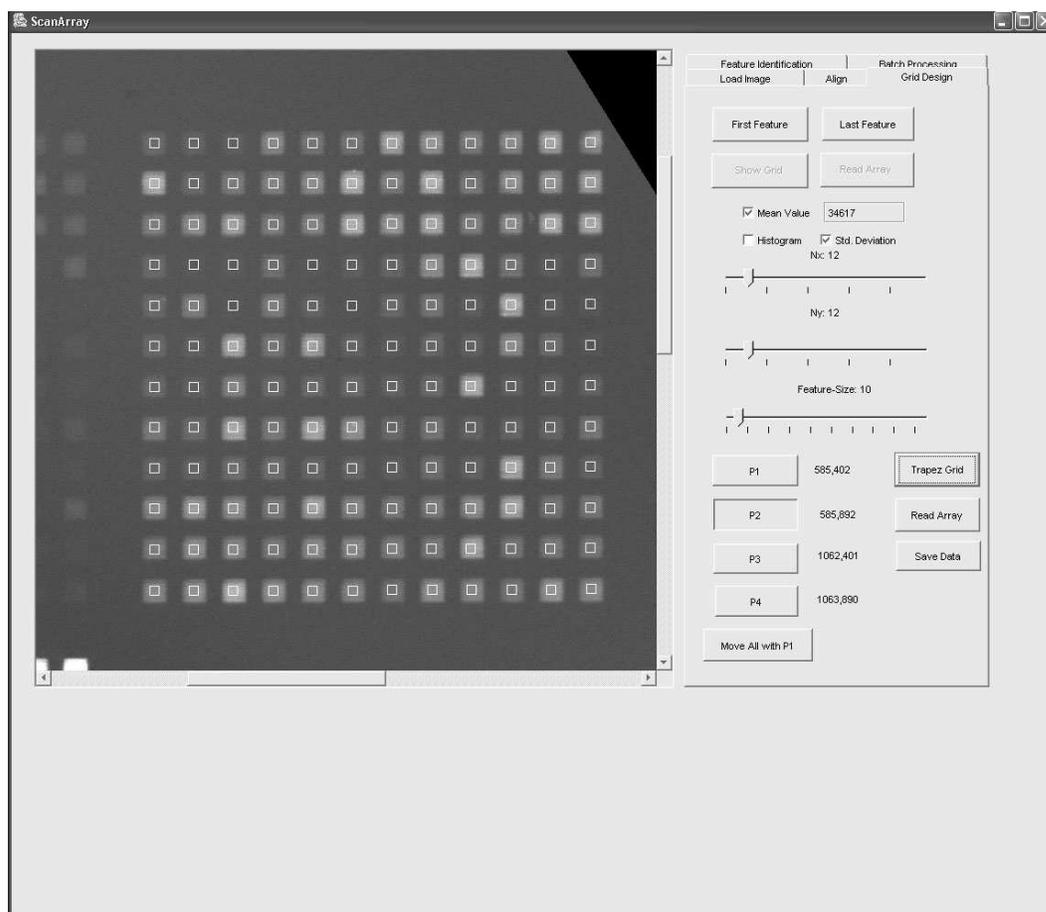
- Loading and processing of 16-bit TIFF images (display is restricted to 8 bit gray scale)
- Image alignment (the microarray needs to be aligned approx. in parallel with the horizontal/vertical axis)
- Placement of the readout grid: the four corner points (set by mouse click into the four corners features of the feature block to be analyzed) define a quadrilateral grid. Further the number of features in x and y direction and the size of the readout frame (chosen to fit into the homogeneous center region of the microarray feature) has to be specified.
- Readout of averaged feature intensities (8 or 16 bit), calculation of the standard deviation of the pixel intensities within the individual feature readout frames (useful for detection of inhomogeneities, particles etc.)
- Batch processing of image series: The readout grid needs to be defined only once. Image drifting (due to thermal expansion of the hybridization chamber) can be com-



**Figure B.14:** Graphical user interface of *ScanRA*. A quadrilateral readout grid (for demonstration the distortion is exaggerated) - to account for small image distortions - is put on the microarray fluorescence micrograph. Feature intensity is integrated over the readout boxes (white boxes).

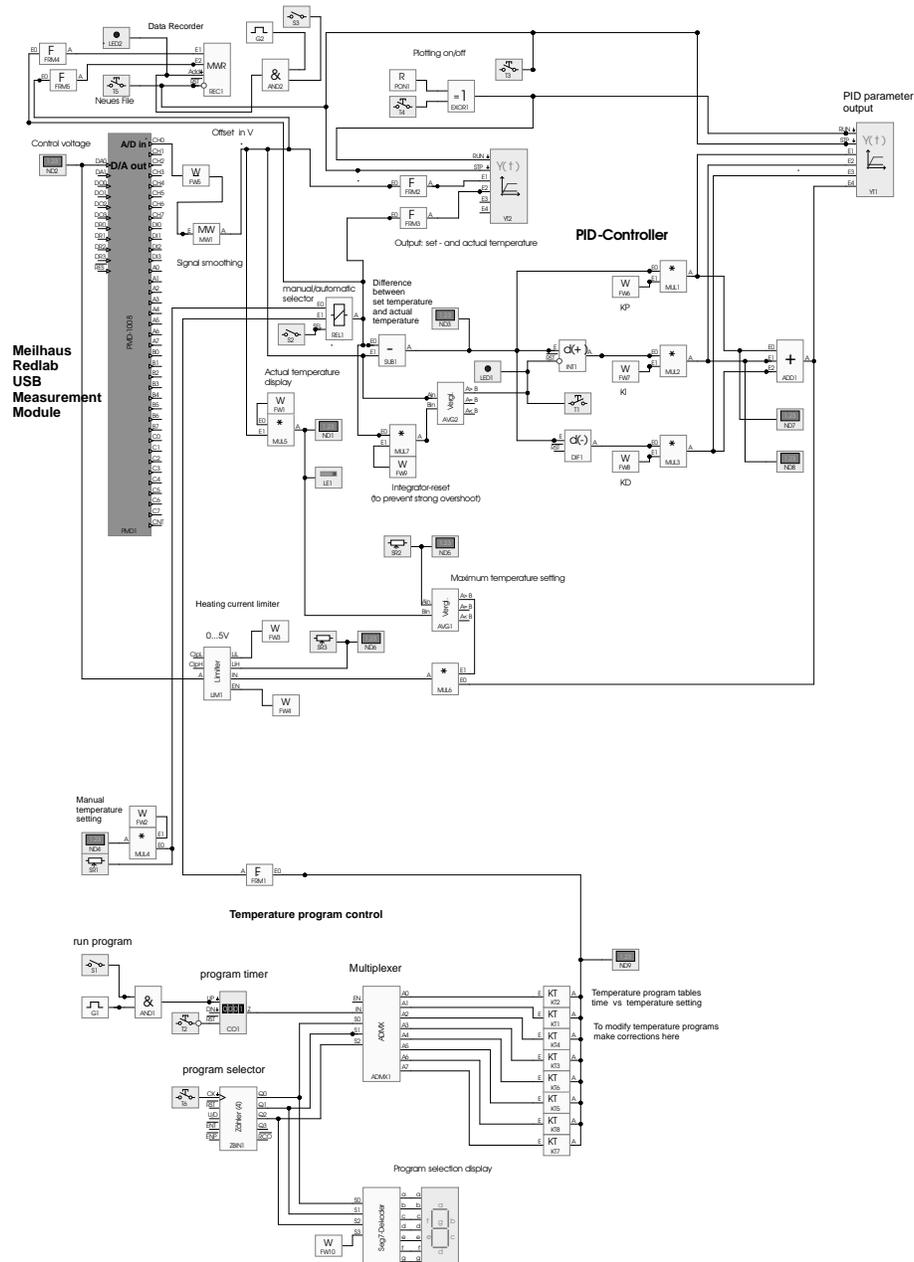
pensated by the drift correction function.

- Feature intensities and standard deviations are saved in Comma Separated Value (CSV) file format.



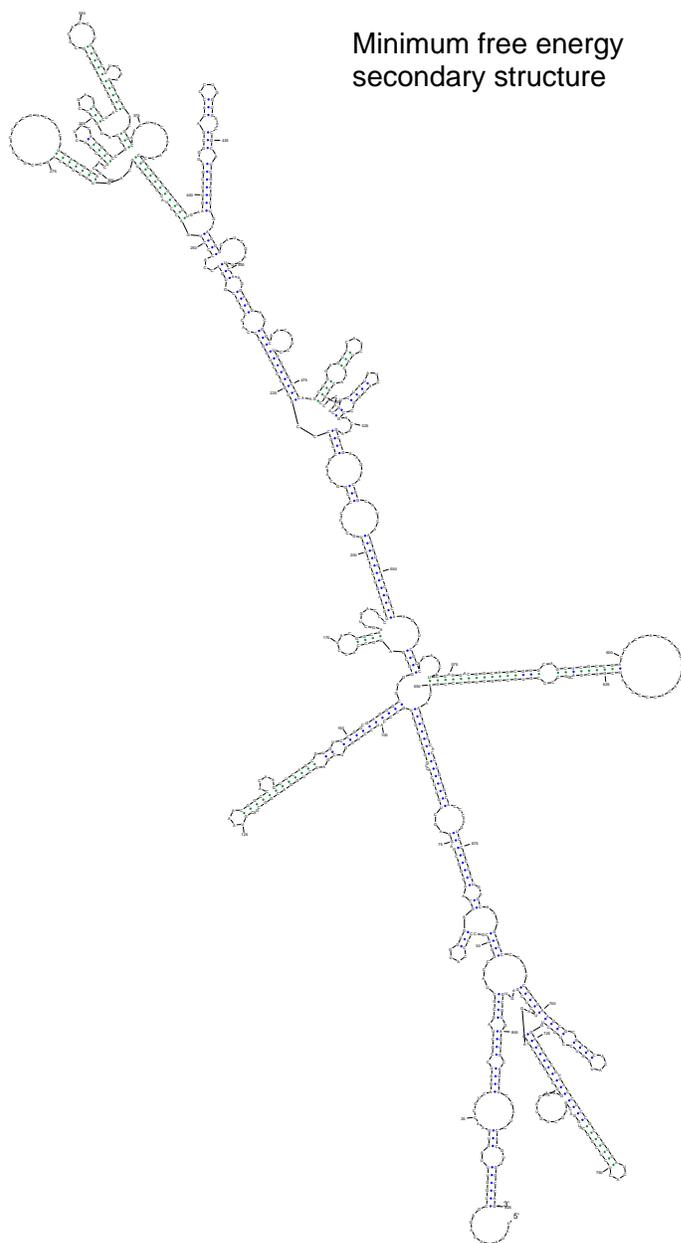
**Figure B.15:** The readout grid is exactly positioned on the microarray features. Averaging over the readout boxes yields the hybridization signals of the individual microarray features.

## B.11 Temperature Control of the Hybridization Chamber



**Figure B.16:** Software-based PID-temperature controller. Implementation with ProfiLab Expert 3.0 (ABACOM GbR). The RedLab measurement module (Meilhaus) is employed for input/output of analog signals. Temperature can be set manual or in a program mode. Programs are entered as tables (ProfiLab-Function "Korrekturabelle") of time versus temperature (recompilation necessary). Between two successive temperature set-points the temperature is varied linearly. The temperature controller application is run on the "microscope control PC" in parallel with the image acquisition-software SimplePCI (Compix Inc.).

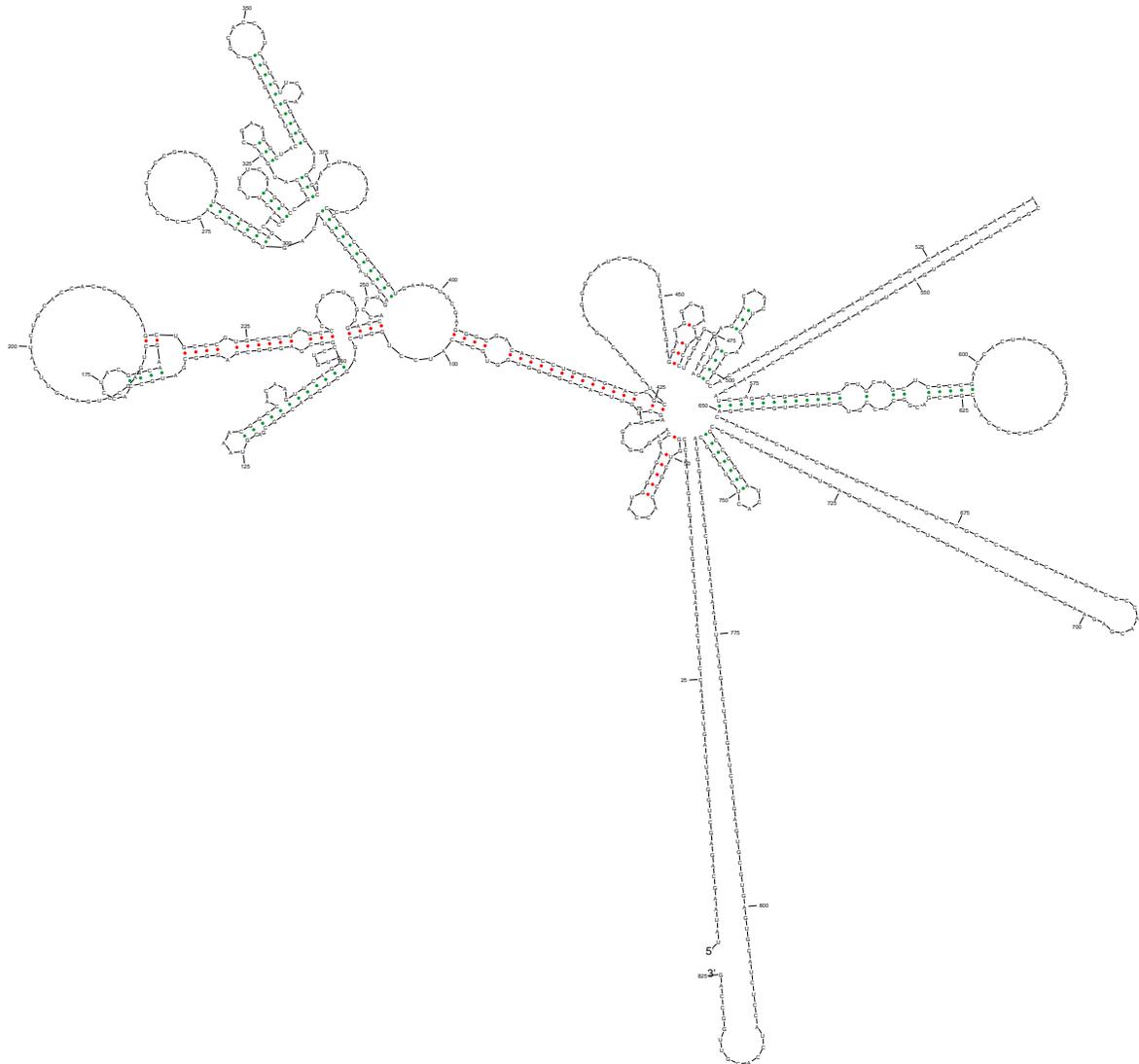
## B.12 cRNA Secondary Structures



$$\Delta G_{37}^{\circ} = -301.8 \text{ kcal/mol}$$

**Figure B.17:** The minimum free energy (MFE) secondary structure of the eGFP cRNA target sequence T2 – see section 8.6.2 – was calculated on the Sfold web server [Din04]. Owing to intrastrand base pairing large parts of the sequence are unavailable for hybridization to DNA microarray probes. The base numbering 1 to 825 corresponds to bases 556 to 1380 of the eGFP-Tub plasmid sequence (see section 8.6.2). Compare with the centroid structure in Fig. B.18. Green dots represent base pairs common in the MFE and centroid structures. Blue dots represent base pairs present only in the MFE structure.

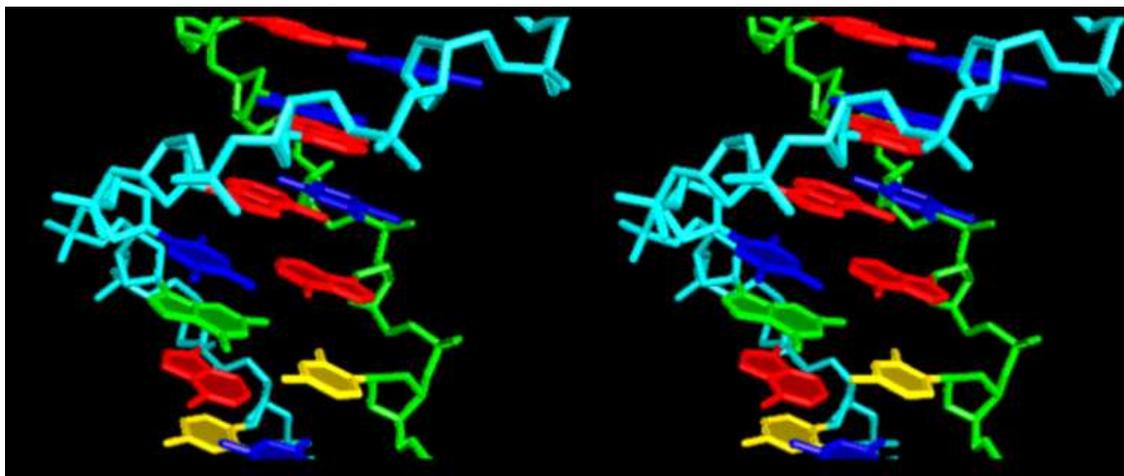
## Ensemble Centroid



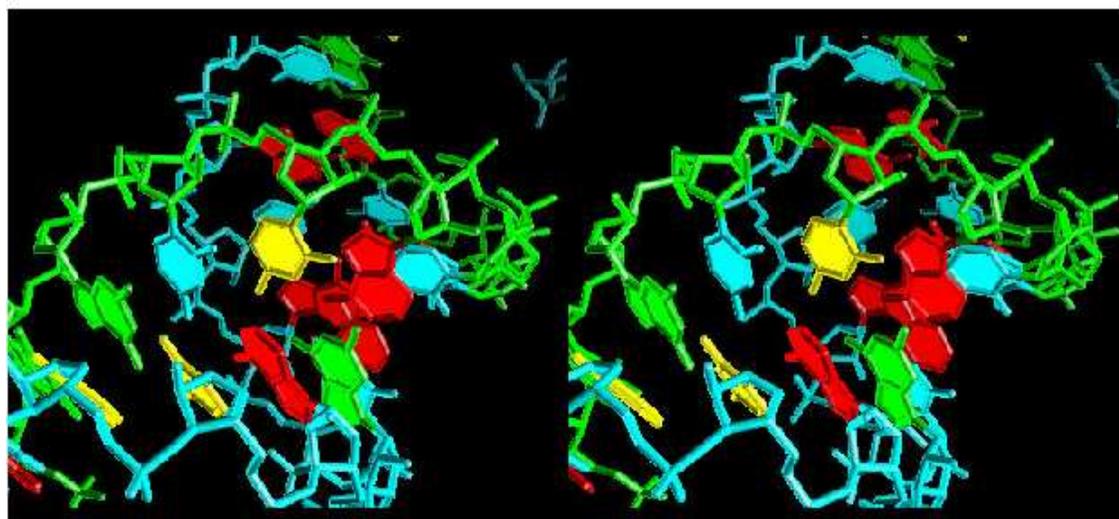
$$\Delta G_{37}^{\circ} = -200.66 \text{ kcal/mol}$$

**Figure B.18:** Centroid secondary structure [Din05] of the eGFP cRNA target sequence T2. The centroid structure was calculated on the Sfold web server [Din04; Cha05] from a Boltzmann-weighted structure ensemble. "The centroid structure can be considered as the single structure that best represents the central tendency of the set" [Cha05]. Compare with the minimum free energy secondary structure in Fig. B.17. Green dots represent base pairs common in the MFE and centroid structures. Red dots represent base pairs present in the centroid structure, not however in the MFE structure.

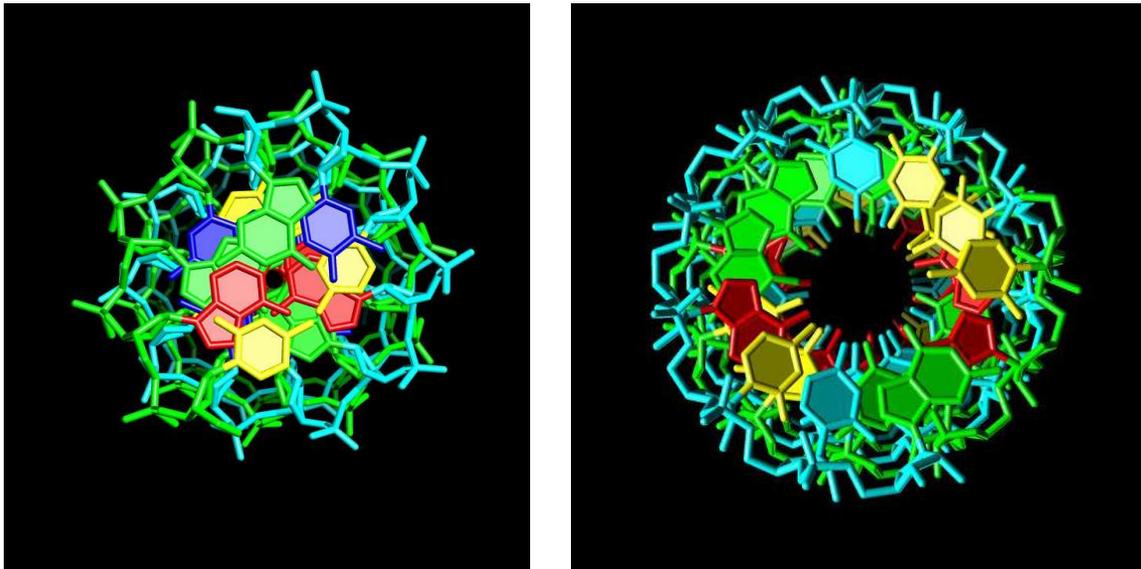
## B.13 3-D Visualization of Nucleic Acid Structures



**Figure B.19:** B-DNA structure - stereo view (use cross-eye-technique for 3D effect). Stereo images of the ideal B-DNA structure were created with UCSF Chimera.



**Figure B.20:** A-RNA structure - stereo view (use cross-eye-technique for 3D effect). Stereo images of the ideal A-RNA structure were created with UCSF Chimera.



**Figure B.21:** Top views of the helix structures - B-DNA (left) and A-RNA (right)  
- demonstrate significant differences in base stacking



## List of Publications

T. Naiser, T. Mai, W. Michel, A. Ott. A versatile maskless microscope projection photolithography system and its application in light-directed fabrication of DNA microarrays. *Review of Scientific Instruments*, 77(6): 063711, 2006.

W. Michel, T. Mai, T. Naiser, A. Ott. Optical study of DNA surface hybridization reveals DNA surface density as a key parameter for microarray hybridization kinetics. *Biophysical Journal*, 92(3):999-1004, 2007.

T. Naiser, O. Ehler, J. Kayser, T. Mai, W. Michel, A. 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* 2008, **8**:48

### Submitted manuscripts

T. Naiser, J. Kayser, T. Mai, W. Michel, A. Ott. DNA hybridization to surface bound probes: point defects in experiment and model.  
Submitted to *Phys. Rev. Lett.*

T. Naiser, J. Kayser, T. Mai, W. Michel A. Ott. Point defects and the stability of surface bound oligonucleotide duplexes - experiments and model.  
Submitted to *Biophysical Journal*



# Danksagung

Als erstes möchte ich mich bei meinem Doktorvater, Herrn Prof. Dr. Albrecht Ott, für die großartige Betreuung meiner Doktorarbeit bedanken. Vielen Dank Albrecht, für die ausgesprochen freundschaftliche Zusammenarbeit mit Dir. Danke für die großen Freiräume die Du mir bei der Ausgestaltung der vorliegenden Arbeit gewährt hast, und auch dafür, dass es praktisch jederzeit möglich war Dich um Rat zu fragen und mit Dir wissenschaftliche Probleme zu erörtern.

Bei meinen Mit-Doktoranden Timo Mai, Pablo Fernandez, Harish Bokkasam, Jérôme Goidin, und ganz besonders bei Wolfgang Michel, mit dem ich über die vergangenen Jahre das Büro geteilt habe, möchte ich mich ebenfalls sehr herzlich für die freundschaftliche Zusammenarbeit und die angenehme und anregende Arbeitsatmosphäre in unserer Arbeitsgruppe bedanken. Dieser Dank gilt ebenso Pramod Pullarkat und Jordi Soriano-Fradera, die mir beide als Postdocs während der vergangenen Jahre stets mit viel Rat und Tat zur Seite standen, und ein wichtiger Quell der Motivation für mich waren – und sicher immer bleiben werden. War 'ne tolle Zeit mit Euch! Zum guten Arbeitsklima in der Arbeitsgruppe haben auch die Diplomanden Oliver Ehler, Jona Kayser, Benjamin Tränkle und Philipp Baaske ihren Beitrag geleistet. Die gesellige Zeit bei unseren EP1-Kaffeepausen – zusammen mit Michael Küken, Ernesto Nicola, Sten Rüdiger, Snigdha Thakur, Cyril Colombo und Roberto Bernal – werd' ich nie vergessen Freunde!

Durch ihre kompetente technische Unterstützung bei molekularbiologischen, softwaretechnischen und elektronischen Problemen haben auch Tobias Mummert, Andrea Hanold, Ralf Pihan und Paul Hurych ganz wesentlich zum Gelingen dieser Arbeit beigetragen. Habt vielen Dank! Auch Jürgen Gmeiner vom Lehrstuhl EP II möchte ich für seine Unterstützung in Fragen der organischen Chemie an dieser Stelle meinen Dank aussprechen. Für eine herausragende technische Unterstützung möchte ich mich auch bei Herrn Krejtschi und seinem Mechanikwerkstatt-Team bedanken.

Besonders herzlicher Dank gilt der guten Seele des Lehrstuhls Margot Lenich für eine tolle Unterstützung in administrativer Hinsicht, aber auch für viele nette Unterhaltungen zwischendurch. Danke Margot!

Für ihren Beitrag zur guten Arbeitsatmosphäre am Lehrstuhl EP I möchte ich mich auch bei Herrn Dr. Uwe Schmelzer, Herrn Prof. Dr. Pascher und seinen Mitarbeitern - meinen Freunden und Kollegen - Jens Fürst, Wolfgang Kellner, Ralf Lang und Andreas Winter sehr herzlich bedanken.

Bei Prof. Dr. Frank Jülicher möchte ich mich dafür bedanken, dass er mir einen mehrmonatigen Gastaufenthalt in seiner Arbeitsgruppe am MPIPKS in Dresden, und somit interessante

Einsichten in the theoretische Biophysik ermöglicht hat. Für die interessanten Diskussionen (die meinen Blick auf die theoretischen Grundlagen der Microarray-Technologie gelenkt haben) und für die freundschaftliche Aufnahme am MPIPKS möchte ich auch Ralf Everaers, Karsten Kruse und Sarah Hallerberg meinen Dank aussprechen.

Großer Dank gilt auch Herrn Prof. Dr. Josef Käs und Herrn Prof. Dr. Thomas Fischer, die bereitwillig die Gutachten für meine Arbeit übernommen haben, wohl ohne zu erahnen worauf sie sich da (hinsichtlich des Umfangs meiner Dissertation) einlassen haben.

Für das Korrekturlesen und Verbesserungsvorschläge möchte ich mich an dieser Stelle auch noch einmal bei Wolfgang, Oliver, Jona und Tobias herzlich bedanken.

Last – not least – gilt mein Dank auch meiner Familie, die mir in den vergangenen Jahren den nötigen Rückhalt gegeben hat, um das vorliegende Monumentalwerk ;-) zu einem erfolgreichen Abschluss zu führen.

*Bayreuth, 16.07.2008*



