DNA Molecular Recognition Specificity: Pairwise and in Competition

Dissertation

zur Erlangung des Grades
der Doktorin der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von

Mina Mohammadi-Kambs

Saarbrücken

2019

Tag des Kolloquiums:	09.06.2020
Dekan:	Univ.-Prof. Dr. Guido Kickelbick
Berichterstatter:	Univ.-Prof. Dr. Albrecht Ott Univ.-Prof. Dr. Gregor Jung Prof. Dr. Enrico Carlon
Vorsitz:	Univ.-Prof. Dr. Ralf Seemann
Akad. Mitarbeiter:	Dr. Mohammad Reza Shaebani

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.

— John R. R. Tolkien,
"The Hobbit or There and Back Again"

Abstract

Despite its importance in biological systems, the molecular recognition of DNA hybridization within complex, competitive environments is poorly understood. The present thesis investigates DNA hybridization in thermal equilibrium for DNA strands bound to the surface of a microarray as well as in solution in presence of one or more competitors. For the latter we employ fluorescence anisotropy and fluorescence correlation spectroscopy to determine binding affinities of two DNA strands in a pairwise manner and in presence of a single competitor. Our results reveal that there must be a non-trivial interaction between the competing strands that extends beyond simple double helix formation. This is a signature of cooperative behavior, which can lead to more complex binding phenomena than previously thought. Moreover, we find surprising differences between the results of both techniques, which we attribute to differing sensitivities to distinct microstates of double helix formation. The second part of this work is performed with surface-bound DNA and devoted to experimentally determine a sufficient number of differing bases between two sequences to avoid cross-hybridization. We construct a set of 23 non-interacting sequences with a length of 7 bases. We conclude that for systems of increasing complexity a high level of discrimination between many competitors is essential for accurate recognition.

Zusammenfassung

Trotz der Relevanz für biologische Systeme sind die Mechanismen molekularer Erkennung bei der Hybridisierung von DNA in komplexen Umgebungen kaum verstanden. Die vorliegende Arbeit untersucht DNA Hybridisierung im thermischen Gleichgewicht mit DNA-Strängen sowohl an die Oberfläche eines Microarrays gebunden als auch in Lösung in Gegenwart von Konkurrenten. Für letztere verwenden wir Fluoreszenzanisotropie sowie -korrelationsspektroskopie, um Bindungsaffinitäten zweier DNA-Stränge paarweise und in Anwesenheit einzelner Konkurrenten zu bestimmen. Unsere Ergebnisse zeigen, dass es nicht triviale Wechselwirkungen zwischen den beteiligten Strängen geben muss, die über die einfache Bildung einer Doppelhelix hinausgehen. Diese Beobachtung deutet auf kooperatives Verhalten hin und zeigt, dass DNA-Hybridisierung komplexer abläuft als bisher angenommen. Außerdem finden wir eine unerwartete Diskrepanz beider Methoden, die auf unterschiedliche Sensitivitäten für bestimmte Mikrozustände der gebundenen DNA zurückgeht. Im zweiten Teil der Arbeit widmen wir uns Experimenten mit oberflächengebundener DNA. Wir bestimmen eine ausreichende Anzahl sich unterscheidender Basenpaare zweier Stränge, um nicht spezifische Hybridisierung zu vermeiden, und zeigen, dass sich damit ein Satz aus 23 nicht interagierenden Strängen á 7 Basen konstruieren lässt. Wir schließen, dass für zunehmend komplexe Systeme ein hoher Diskriminierungsgrad zwischen vielen Konkurrenten unabdingbar für präzise Erkennung ist.

Contents

Chapter 1

Introduction

Molecular recognition, specific non-covalent binding between two or more molecules is one of the fundamental mechanisms in biological processes. The main elements of life such as replication, metabolism or information processing occur as a result of this type of interaction among biological molecules. Molecular recognition must occur with high accuracy in cell. Non-covalent binding refers to hydrophobic forces [1,2], hydrogen bonding [3], van-der-Waals forces, $\pi - \pi$ interactions [3,4], or electrostatic effects [5]. Molecular recognition plays a crucial role between two proteins, an antigen and its corresponding antibody, or between two DNA strands, among others. In 1984 Emil fisher proposed a model, called *lock and key* (LK), to describe the binding between an enzyme and a substrate [6]. According to this analogy, the lock is a molecular receptor like an enzyme and the key is the substrate. Only the correctly sized and formed key or a substrate can fit into a key hole of the lock, in this case the active site of an enzyme. Although the LK analogy offers an intuitive picture of molecular binding, it fails to describe the complexity of the molecular recognition in biological systems to the core. To account for a number of properties that appear in realistic systems, some extensions to the LK analogy have been proposed. For instance, Daniel Koshland [7] postulated an *induced fit* model, which suggests that if the active site of an enzyme does not initially perfectly match to the substrate, the substrate induces conformational changes in enzyme during interactions, leading to a more accurate binding.

Despite these modifications, the lock and key model remains with some shortcomings. A limitation of the lock and key model emerges, if one considers the case of numerous similar keys, which almost fit to a specific lock, but only a few of them are able to open it. This situation entails two complications undermining molecular recognition. On one hand, the almost fitting keys might occupy the lock and thereby keep the correct one from opening it. On the other hand, some keys might be similar enough to open the lock although its owner was never intended to have access. This situation is comparable to a crowded biological environments including many similar binding partners.

Binding between not fully complementary molecules still occurs, although with lower affinity. These non-specific bindings introduce errors to the system and make the binding mechanism difficult to understand. In the cell recognition requires accurate means of discriminating a specific molecule among many similar competitors at high density.

DNA hybridization, a process of binding two single DNA strands forming a double helix, is a highly sequence specific molecular recognition mechanism which plays an important role in biology, clinical or biotechnological applications, as for instance the physical realization of nanodevices and nanocircuits [8–11], DNA hybridization catalysis [12–14] or DNA-based computing [15–20]. Although many works have been dedicated to describe DNA double helix formation, DNA hybridization in complex situations remains poorly understood. Many techniques have been employed to understand the thermodynamics of DNA hybridization such as gel electrophoresis [21–23], isothermal urea titration [24], microcalorimetry [25–27], or UV absorption [28, 29]. Hybridization experiments are often accompanied by theoretical models relying on the nearest neighbor model which is based on experimentally obtained binding energies in thermal equilibrium [30–34]. In many cases the predictions are very accurate in spite of unavoidable approximations. The nearest neighbor model has been implemented in various software packages such as Mfold [35] and Nupack [36], making their predictions available to a wide range of researchers.

In more general scenarios beyond pairwise binding, it might be expected that investigating the uncertainty of binding within a competitive and crowded environment yields valuable insights towards understanding how cells deal with this molecular ambiguity. An in vitro example of such an environment is DNA microarrays with wide range of applications in biological systems. The major source of false recognition on DNA microarrays arises from single or few mismatched bases between two strands leading to unspecific cross-hybridizations [37–39]. Many researches are focused on the importance of cross-hybridization for quantitative analysis of DNA microarray data in particular for the single nucleotide polymorphisms detection [40–42] or the precise determination of gene expression level [43, 44]. Previous theoretical studies proposed strategies to create a set of surface-bound sequences without any cross-hybridization [45–50].

To investigate the physical binding properties of molecules especially in a solution mixture of many competing binding partners fluorescence techniques can be applied. One possible method is fluorescence anisotropy (FA) which is used in many clinical domains for understanding the mechanisms of drug action [51–54] or detection of chemical contaminations in food [55]. The technique can also be employed in biology to study molecular binding and determining the binding affinity of two interacting biological molecules, for instance DNA strands and/or proteins [56–61]. A FA measurement is based on a change in the rotational mobility of a fluorescent molecule as a result of formation of molecular complexes. Fluorescence correlation spectroscopy

(FCS) is another method which relies on a slow-down of a diffusion caused by a binding of a light molecule to its complementary partner, independent of its thermal rotation [62, 63].

The main focus of this work is to investigate DNA hybridization in particular in competitive environments. On that account we study two major systems: first, hybridization of DNA strands bound to microarray surfaces, which are exposed to a highly crowded solution of various free strands. We propose a strategy, borrowed from concepts known to coding and graph theory, to construct sets of non-interacting sequences. Experimentally, we prove that the sequences of these sets are indeed mutually independent by showing that no cross-hybridization is detectable despite the presence of competitors.

Besides experiments with surface bound strands, we also perform binding assays in bulk. We apply both fluorescence anisotropy and fluorescence correlation spectroscopy to obtain complementary information on the binding of short DNA sequences (12-16 bases) to their complement in a pairwise manner as well as in presence of a single competitor. We show that the binding affinity obtained from FCS for a case with a single mismatch at the center of a strand varies by more than two orders of magnitude from FA value as well as NUPACK prediction.

To explain our findings for competition cases we derive models that describe the simultaneous or exclusive binding of both competitors, which, however, neglect any direct interaction. In certain situations we find a clear disagreement between the models and our findings which suggests non-negligible interaction between two competing targets that not only changes the underlying binding constants, but even to a different degree for each. Our results suggest that cooperative processes in competitive biological environments play a major role in the accuracy of the molecular recognition.

The content of the present thesis is divided into the following chapters: in Chap. 2, we present some selected fundamentals about the biological and chemical properties of DNA and introduce the principles of experimental techniques applied throughout this thesis including light directed in situ synthesis of DNA strands, fluorescence anisotropy and fluorescence correlation spectroscopy which help understanding the results of this work. The design and performance of our employed experimental setups are discussed in Chap. 3. Both Chap. 4 and Chap. 5 are devoted to the obtained experimental results: in Chap. 4 we focus on understanding DNA hybridization in solution, especially in presence of a single competing partner. Chap. 5, on the other hand, investigates and discusses strategies to avoid cross-hybridization on a set of surface-bound DNA strands and proves the independency of such a set. Eventually, we summarize all results of the present thesis in Chap. 6.

Chapter 2

Fundamental

2.1 DNA Structure

The structure of deoxyribonucleic acid or DNA was first discovered by James D. Watson and Francis Crick [64, 65]. A DNA is a complex molecule carrying the genetic information and instruction for functioning and constructing living organism. It also serves as a hereditary material in organisms of all types. The genetic information in DNA is stored on its four chemical bases: adenine (A), thymine (T), cytosine (C) and guanine (G) as illustrated in Fig. 2.1. DNA is a chain-like macromolecule composed of a string of nitrogen-containing units, called nucleotides made of aromatic nucleobases which are either pyrimidine (C and T) or purine derivative (A and G), pentose sugar 2-deoxyribose and phosphate group. The carbon atoms of the deoxyribose sugar are numbered (see Fig. 2.2). The free hydroxyl group at 3' atom form a phosphodiester bond with the phosphate group at the 5' end of an adjacent nucleotide. As a result, DNA strands always have a 5' and 3' end. Fig. 2.2 depicts a structural formula of a single DNA strand 5'-ACGT-3'. The orientation of the strand is determined by the arrangement of the ribose rings.

If two single complementary DNA strands bind to each other, the result is the well-known DNA double helix structure [64]. While a single strand DNA is very flexible and has a very short persistence length of 0.8-3 nm, depending on the ambient salt concentration [66–68], a double helix is much stiffer and has a longer persistence length of 40-50 nm [69]. The binding of two strands happens when the complementary bases of two single strands, A with T and C with G form hydrogen bonds, known as Watson-Crick pairing. Since the guanine and cytosine are linked with three hydrogen bonds and adenine and thymine with only two hydrogen bonds, the binding of guanine to cytosine is more stable than adenine to thymine (see Fig. 2.1). Note that duplex formation occurs in an anti-parallel manner, i.e. the orientation of the sugar molecule is opposite in the two single strands. The duplex stability not only originates from the hydrogen bonds between the nucleotide bases but the base

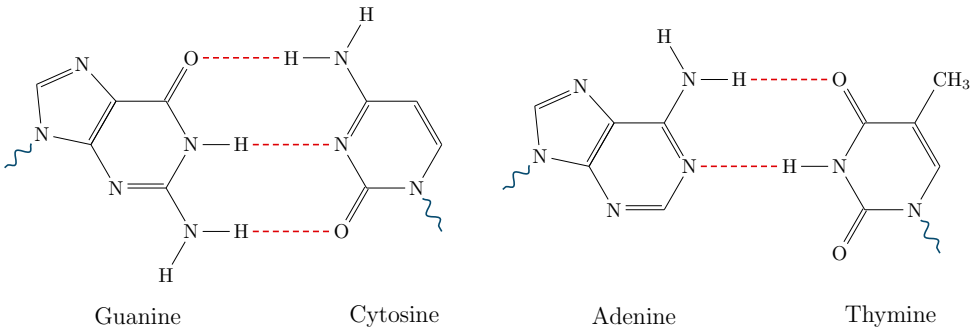

Figure 2.1. Four DNA nucleobases adenine (A), thymine (T), guanine (G), and cytosine (C). The red dashed lines depict the Watson-Crick base pairing between adenine and thymine (A-T) using two hydrogen bonds as well as guanine and cytosine (G-C) with three hydrogen bonds.

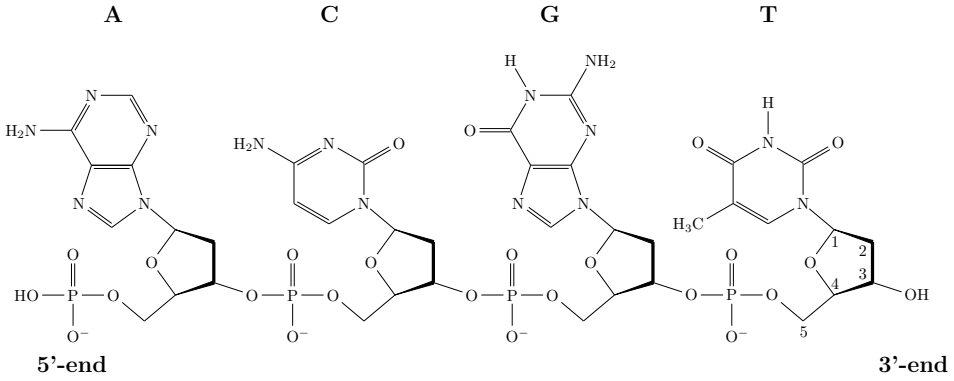

Figure 2.2. Structure of a DNA strand 5'-ACGT-3'. The carbons of a deoxyribose sugar are numbered from 1 to 5. The 3' atom carries a hydroxyl group which is connected via a phosphodiester bond to a phosphate group of an adjacent nucleotide. The orientation of a strand depends on whether the phosphate or the deoxyribose are the terminal molecules.

stacking or $\pi - \pi$ interactions between adjacent nucleobases also has a large contribution for further stabilization [70]. A number of theoretical [71–76] and experimental [77–82] studies address the characterization of the base pairing and base stacking as well as their contribution in duplex stability.

2.2 Nucleic Acid Hybridization

As we mentioned above, the stability of a DNA double helix is not only resulting from hydrogen bonds but it also originates from the attractive so-called base stacking or $\pi - \pi$ interaction between the stacked aromatic bases. The process of binding two single strands and forming a double helix is called

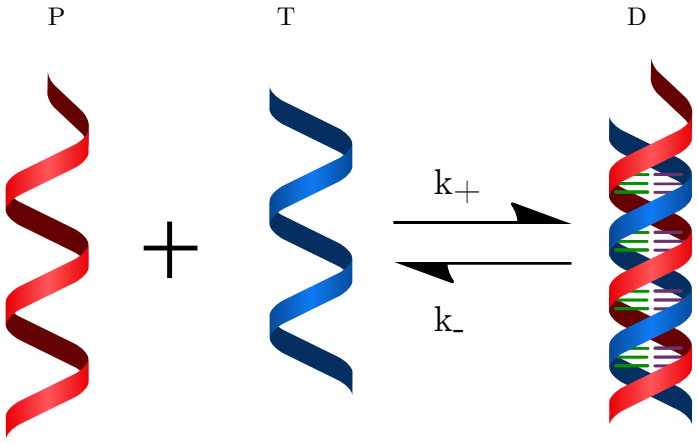

Figure 2.3. DNA hybridization reaction according to two-states model. Two single DNA strands probe P and target T form a double helix D with an association rate k_+. The double helix can unwind to two single strands P and T with a dissociation rate of k_-.

DNA hybridization. However, in thermodynamic equilibrium the helix conformation is balanced by DNA denaturation, a process in which a double helix unwinds and separates into two single strands. The thermodynamics of DNA hybridization, binding of two single strands probe P and target T forming a duplex D, is simply described using a two-state model (refer to Fig. 2.3), neglecting the possibility of any intermediate state. The hybridization reaction in this case is given by

$$P + T \underset{k_-}{\overset{k_+}{\rightleftharpoons}} D \tag{2.1}$$

where k_+ and k_- are the association and dissociation rates. By decreasing the temperature and increasing the ionic concentration of the hybridization buffer the probability of the double helix formation increases. However, the strength of the duplex binding also depends on the amount of cytosine and guanine (GC content) of DNA strands. DNA hybridization is a highly sequence specific process, i.e. even a single mismatch within a DNA duplex can reduce the binding significantly [83–85]. In the following we describe some aspects of DNA hybridization such as kinetics, melting temperature and the double helix stability.

The rate equation for the formation of duplex D is given as

$$\frac{\mathrm{d}c_D}{\mathrm{d}t} = k_+ c_P c_T - k_- c_D. \tag{2.2}$$

In equilibrium $\frac{d(c_D)}{dt} = 0$ and therefore we obtain the binding constant K as

$$K = \frac{k_+}{k_-} = \frac{c_D^{\mathrm{eq}}}{c_P^{\mathrm{eq}} c_T^{\mathrm{eq}}}, \tag{2.3}$$

where $c_P^{\text{eq}} = c_P^{\text{ini}} - c_D$ and $c_T^{\text{eq}} = c_T^{\text{ini}} - c_D$. The superscript 'eq' refers to an equilibrium concentration and the subscript 'ini' to an initial concentration. The Gibbs free energy of duplex formation ΔG_D is related to the equilibrium binding constant K through

$$\Delta G_D = -RT \ln K, \tag{2.4}$$

where $R = 1.987\,\text{cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ is the ideal gas constant and T the hybridization temperature. On the other hand, the Gibbs free energy change ΔG can be obtained by the changes in enthalpy ΔH and entropy ΔS as

$$\Delta G = \Delta H - T\Delta S, \tag{2.5}$$

where ΔH and ΔS can be predicted from the nearest neighbor thermodynamic parameters (see Sec. 2.4.1). Combining Eq. (2.4) and Eq. (2.5) can be written as

$$\frac{1}{T} = \frac{R}{\Delta H} \cdot \ln K + \frac{\Delta S}{\Delta H}. \tag{2.6}$$

The changes in enthalpy ΔH and entropy ΔS can be determined according to Eq. (2.6) by plotting $\ln K$ as a function of $1/T$. Knowing both parameters ΔH and ΔS, it is possible to evaluate ΔG using Eq. (2.5).

2.3 DNA Duplex Melting Temperature

DNA denaturation denotes a process, in which a double stranded DNA separates to two single strands by rising a temperature up to a specific point

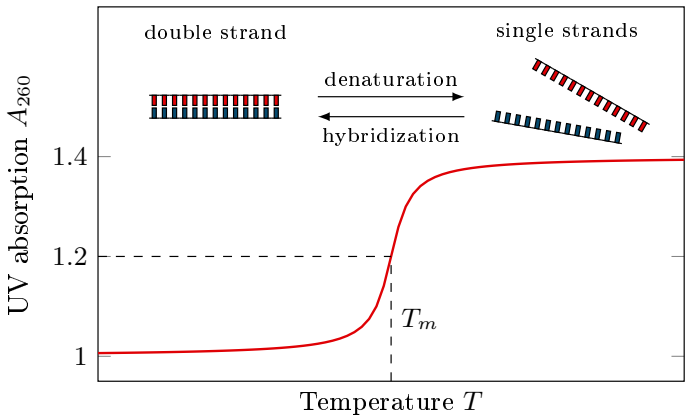

Figure 2.4. DNA melting transition. The UV absorbance of a duplex decreases due to the larger stacking interactions. DNA denaturation occurring by increasing the temperature leads to an increase in UV absorption by 20-40 %. The melting temperature is by definition a midpoint of the melting transition.

or increasing the pH value of the solution. The temperature, in which 50 % of the DNA double strands are single stranded is called melting temperature T_m. DNA strands denaturation provides important information about the stability of a duplex, interaction of single bases and the influence of the salt concentration [64, 86, 87]. A common technique to determine a melting temperature of a DNA duplexes is UV absorption [88]. The UV absorption for a single stranded DNA at around 260 nm (A_{260}) is elevated since the bases are free. For a double helix the UV absorbance reduces. Fig. 2.4 depicts the absorbance change as a function of temperature. The melting transition occurs by increasing the UV absorbance by about 20-40 %.

2.4 Langmuir Model

The equilibrium reaction between target and probe strands with double helix as shown in Eq. (2.1) is described by Langmuir-type adsorption isotherm [89]. If the targets in the experiment are in excess over the probes, the target concentration can be taken as constant $c_T = c_T^{\text{ini}}$. Considering a probe concentration as $c_P = c_P^{\text{ini}} - c_D$, according to Eq. (2.3) we obtain

$$K = \frac{c_D^{\text{eq}}}{c_T^{\text{ini}} \left(c_P^{\text{ini}} - c_D^{\text{eq}} \right)}. \tag{2.7}$$

The fraction of hybridized probe $\Theta = \frac{c_D^{\text{eq}}}{c_P^{\text{ini}}}$ can be written as

$$\Theta = \frac{K \cdot c_T^{\text{ini}}}{1 + K \cdot c_T^{\text{ini}}}. \tag{2.8}$$

Eq. (2.8) is defined as a classical Langmuir isotherm [90]. Inserting $K = \exp\left(-\Delta G_D / RT\right)$ in Eq. (2.8), we obtain a relation between the fraction of hybridized probe Θ and the duplex free energy ΔG_D as following [89, 91, 92]

$$\Theta = \frac{\exp\left(-\Delta G_D / RT\right) \cdot c_T^{\text{ini}}}{1 + \exp\left(-\Delta G_D / RT\right) \cdot c_T^{\text{ini}}}. \tag{2.9}$$

For the case where the concentration of targets and probes are comparable, the target concentration does not remain constant and therefore we substitute the target concentration in Eq. (2.3) by $c_T = c_T^{\text{ini}} - c_D$ instead (compare to Eq. (2.7)). We derive the equilibrium duplex concentration as

$$\begin{aligned} c_D^{\text{eq}} =& \frac{1}{2} \left[c_T^{\text{ini}} + c_P^{\text{ini}} + \frac{1}{K} - \sqrt{\left(c_T^{\text{ini}} + c_P^{\text{ini}} + \frac{1}{K} \right)^2 - 4 \cdot c_T^{\text{ini}} \cdot c_P^{\text{ini}}} \right] \\ =& \mathcal{L}\left(c_T^{\text{ini}}, c_P^{\text{ini}}; K \right), \end{aligned} \tag{2.10}$$

which we define as the extended Langmuir isotherm [93]. For the case where two non-interacting target strands, A and B compete to bind to their common complementary probe and form a double helix D_A and D_B, we can write the differential equation Eq. (2.2) as:

$$\begin{aligned}\frac{\mathrm{d}c_D^A}{\mathrm{d}t}(t) &= k_+^A \cdot c_P(t) \cdot c_A(t) - k_-^A c_D^A(t) \\ \frac{\mathrm{d}c_D^B}{\mathrm{d}t}(t) &= k_+^B \cdot c_P(t) \cdot c_B(t) - k_-^B c_D^B(t),\end{aligned} \tag{2.11}$$

where $c_P(t) = c_P^{\mathrm{ini}} - c_D^A(t) - c_D^B(t)$. Solving the equations at equilibrium, i.e. $\frac{\mathrm{d}c_D^A}{\mathrm{d}t}(t) = 0$ and $\frac{\mathrm{d}c_D^B}{\mathrm{d}t}(t) = 0$ and considering $K = \frac{k_+}{k_-}$ as a duplex binding constant, we obtain the equilibrium concentrations of duplexes D_A^{eq} and D_B^{eq} as:

$$\begin{aligned}D_A^{\mathrm{eq}} &= \frac{c_P^{\mathrm{ini}} \cdot K_A \cdot c_A^{\mathrm{eq}}}{K_A \cdot c_A^{\mathrm{eq}} + K_B \cdot c_B^{\mathrm{eq}} + 1} \\ D_B^{\mathrm{eq}} &= \frac{c_P^{\mathrm{ini}} \cdot K_B \cdot c_B^{\mathrm{eq}}}{K_A \cdot c_A^{\mathrm{eq}} + K_B \cdot c_B^{\mathrm{eq}} + 1},\end{aligned} \tag{2.12}$$

where, c_A^{eq} and c_B^{eq} are the equilibrium concentrations of target A and B. Accordingly, the ratio of the equilibrium concentrations of the formed duplexes D_A^{eq} and D_B^{eq} is described by

$$\frac{D_A^{\mathrm{eq}}}{D_B^{\mathrm{eq}}} = \frac{c_A^{\mathrm{eq}}}{c_B^{\mathrm{eq}}} \cdot \frac{K_A}{K_B}. \tag{2.13}$$

Therefore, the competition situation in two-state model can be solely explained by the ratio of the individual binding affinities of the competing targets K_A/K_B and the equilibrium concentrations $c_A^{\mathrm{eq}}/c_B^{\mathrm{eq}}$. Employing Eq. (2.4) in Eq. (2.13) we find the ratio of the duplexes concentrations as

$$\frac{D_A^{\mathrm{eq}}}{D_B^{\mathrm{eq}}} = \frac{c_A^{\mathrm{eq}}}{c_B^{\mathrm{eq}}} \cdot \exp\frac{(\Delta\Delta G)}{RT}, \tag{2.14}$$

where $\Delta\Delta G = \Delta G_A - \Delta G_B$ is a difference between the binding free energies.

2.4.1 Nearest Neighbor Model

As described above, DNA duplex stability not only depends on hydrogen bonds but also on base stacking interactions between adjacent base pairs, which have a larger contribution than hydrogen bonds [94]. It turns out that the stability of a duplex depends on the identity of neighboring bases [32, 34]. In the following we show the nearest neighbor (NN) model in Gotoh format [95] in which the two contributions, hydrogen bonding and base stacking

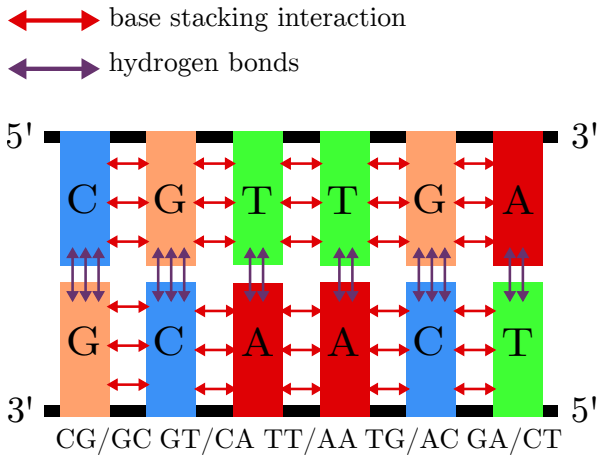

Figure 2.5. The nearest neighbor model for an exemplary duplex formation including hydrogen bonds (violet arrows) between the bases of two strands and base stacking interactions (red arrows) between the adjacent base pairs. While a GC base pair is stabilized by three hydrogen bonds, AT pair by two. In the presented format of NN model, both hydrogen bonds and base stacking interactions contribute to form a NN parameter. The sum of all individual NN parameters results in the stability of the duplex (adapted from [92]).

interactions, are combined into a single nearest neighbor parameter. Fig. 2.5 depicts the NN model for a duplex: The hydrogen bonds between the bases of two strands are shown in violet arrows while the red arrows indicate the base stacking interactions between adjacent base pairs. The corresponding NN pairs are displayed below the double strand. The duplex free energy is calculated by [34]

$$\Delta G_{\text{total}} = \sum_i n_i \Delta G_i + \Delta G_{\text{init}} + \Delta G_{\text{sym}}, \tag{2.15}$$

where n_i in the first term is the occurrence of each nearest neighbor i and ΔG_{i} is the summation over the standard free energy changes of all the possible Watson-Crick NNs. The second term ΔG_{init} is the initiation energy of duplex formation, i.e. the energy of each initial base pair. In order to take into account the differences between duplexes with terminal AT or CG, two initiation parameters are introduced [34,96,97]. The symmetry term ΔG_{sym} is added only for the self-complementary duplexes [98]. For instance, the nearest neighbor interaction for the duplex shown in Fig. 2.5 is provided by [34]:

$$\begin{aligned}\Delta G_{\text{total}} =& \Delta G_{\text{init}} + \Delta G(\text{CG/GC}) + \Delta G(\text{GT/CA}) \\ &+ \Delta G(\text{TT/AA}) + \Delta G(\text{TG/AC}) \\ &+ \Delta G(\text{GA/CT}),\end{aligned} \tag{2.16}$$

Table 2.1. Watson-Crick nearest neighbor thermodynamic parameters in 1 M NaCl as reported in [99]. ΔH, ΔS and ΔG correspond to the changes in enthalpy, entropy and free Gibbs energy, respectively. A nearest neighbor parameter gives the stability of a base pair duplet. For instance, GA/CT refers to a sequence consisting of two bases 5'-GA-3' which is paired with a complementary sequence 3'-CT-5'. For symmetry reason this leads to 10 different NN parameters. While 'init' terms AT and GC account for the initiation free energies, the symmetry term appears in case of self-complementary duplexes. The right column represents the calculated Gibbs free energy parameters using the values from second and third columns (ΔH and ΔS) at 37 °C.

NN-pair	ΔH (kcal/mol)	ΔS (cal/(K · mol))	ΔG_{37° (kcal/mol)
AA/TT	-7.9	-22.2	-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
init AT	2.3	4.1	1.03
init CG	0.1	-2.8	0.98
symmetry	0	-1,4	0.4

where the slashes represent the base stacking between adjacent base pairs, e.g. CG/GC means 5'-CG-3' is base paired with 3'-GC-5'. The ΔG_{sym} equals zero because the duplex is non-self-complementary. The NN free energy parameters for DNA Watson-Crick pairs in 1 M NaCl at 37 °C are presented in Tab. 2.1 according to [99]. Using the values from Tab. 2.1 in Eq. (2.16), we obtain the following free energy for the exemplary duplex shown in Fig. 2.5

$$\begin{aligned}\Delta G_{37^\circ} =& 0.98 + 1.03 - 2.17 - 1.44 - 1.00 - 1.45 - 1.30 \\ =& -5.35\,\text{kcal/mol}.\end{aligned} \tag{2.17}$$

In analogy the changes in enthalpy ΔH and entropy ΔS for duplex formation accordingly can be calculated using NNs parameters from Tab. 2.1 and the Gibbs free energy ΔG at any given temperature T is then obtained by $\Delta G = \Delta H - T\Delta S$.

2.5 DNA Microarray Synthesis

Solid-phase synthesis is a technique in which the polymer-chains are end-tethered to a solid substrate. The synthesis is done step by step using the selective protecting group chemistry as following: The first monomer is tethered to the functionalized substrate and the chemical protecting group prevents the uncontrolled coupling until a chemical deprotection reaction removes the protecting group. Subsequently, the next monomer can bind to the previous one. This method can be used to synthesize nucleic acid sequences.

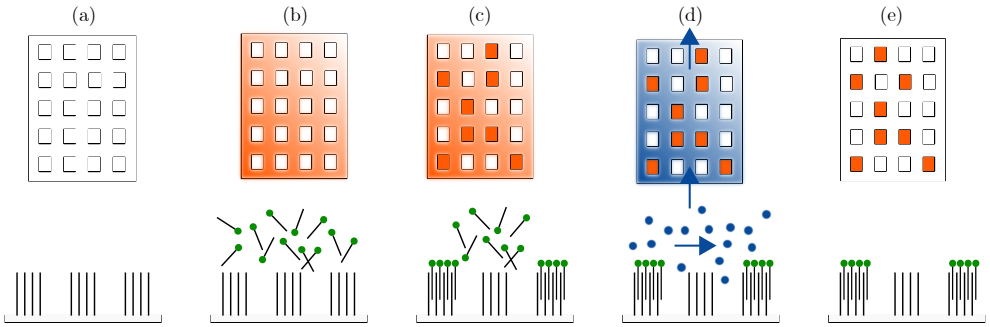

Figure 2.6. Hybridization of DNA strands on a microarray surface. The upper pictures show the microarray surface including the features while the pictures at the bottom line illustrate the surface tethered probe sequences. (a) The position of the immobilized probes on a microarray is defined. (b) The probes are exposed to the hybridization solution including fluorescently labeled target shown by green circles. (c) The fluorescently-labeled targets diffuse freely on the surface and eventually hybridize to their complementary probes at equilibrium. The orange features correspond to the probes which their complementary targets are present and therefore exhibit a measurable fluorescent signal. The white features with no detectable signal belong to the probes that their complementary targets are not present in the solution. (d) The unbound targets are washed a way from the surface and only the hybridized targets remain on the microarray. (e) Since the position and the sequence belong to each feature are known, measuring the fluorescent intensity at the position of each feature gives the information about the binding of a probe and target (adapted from [92]).

DNA microarrays enable the study of the complex mixture of single-stranded DNA sequences in parallel manner. On this account, the well-known oligonucleotide sequences, so-called probes, as illustrated in Fig. 2.6 (a) are end-tethered in a regular arrangement at the defined positions, called the features on the microarray surface. Each feature, includes a large number of the same probes. The probes as are shown in (b) are exposed to the fluorescently labeled DNA strands, the targets, that are complementary to the probes (the green circles). The labeled targets which hybridize to their complementary probes accumulate on the corresponding feature and makes the

fluorescent intensity at this point measurable. These features are shown in orange color while the probes that their complementary targets are not present in a solution are shown with white features. The labeled unbound targets are, however, washed away from the microarray surface (see (c) and (d)). On the other hand, since the position of each feature is defined, the intensity of the hybridization can be assigned to the corresponding probe sequence (e). Microarrays are only semi-quantitative due to the factors affecting the microarray measurements such as, poor predictability of individual probe-target affinities, competitive hybridization between different targets leading to a cross-hybridization or variation in synthesis quality. However, the microarrays became powerful tools for a broad range of applications including gene expression profiling [100–104], genotyping assays [105–111], resequencing [112–114] and pathogen detection [115, 116].

2.5.1 Light Directed In Situ Synthesis

The probes on the microarray surface are made of individual building blocks, called phosphoramidites [117–119] consisting of nucleobases, deoxyribose and a phosphorus group at the 3'-end of the sugar, which can chemically react with the hydroxyl group at the 5'-end of the 2-deoxyribose ring of another

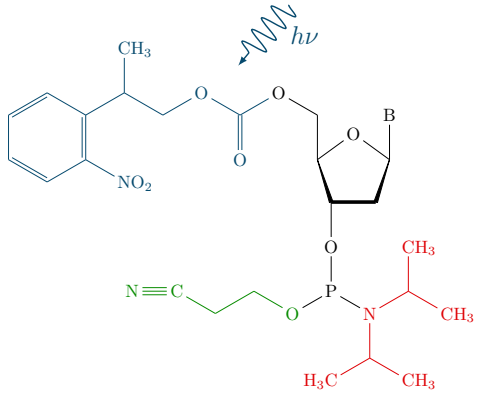

Figure 2.7. The structural formula of phosphoramidite composed of nucleobase, sugar deoxyribose and a phosphate group at the 3'-end of the sugar. To prevent unwanted reaction a phosphorus group is protected by a 2-cyanoethyl group (green molecule) and a diisopropylamino (red molecule). Furthermore, the 5'-hydroxyl of the pentose ring is protected by a NPPOC (2-(2-Nitrophenyl)-propoxycarbonyl) group shown in blue. As this protecting group is photolabile, it can be removed by exposure to UV light. This enables coupling of another building block. The functional group B corresponds to one of the DNA bases, adenine, thymine, cytosine or guanine which are provided by additional protecting groups ib, tac, ipac. The protecting groups are all removed at the last step of the synthesis (adapted from [92]).

phosphoramidite. Fig. 2.7 illustrates the structure of a phosphoramidite. To avoid uncontrolled coupling of phosphoramidite building blocks, a photolabile protecting group for instance NPPOC (2-(2-Nitrophenyl)-propoxycarbonyl) substitutes a 5'-hydroxyl moiety. In the light-directed in situ synthesis under UV radiation the protecting group is cleaved and the 5'-hydroxyl can be coupled with the 3'-phosphate of another phosphoramidite building block leading to a synthesis in the 3' to 5' direction. Further protecting groups are necessary to avoid unwanted reactions: The phosphorus group on the other hand is protected by diisopropylamino (in red) and 2-cyanoethyl groups (in green) while the four bases adenine, thymine, cytosine and guanine (shown as B) are protected by ib, tac, ipac protecting groups. All protecting groups are removed at the final deprotection step of the synthesis. For more details about the synthesis steps see [92, 120]

Fig. 2.8 shows the schematic of the synthesis cycle for building the probes on a microarray surface. The synthesis substrate is initially functionalized by the photo-labile protecting group NPPOC, shown with blue spheres here. A controlled UV exposure removes the protecting groups at desired positions where phosphoramidite building blocks should be attached in the following coupling step. The free hydroxyl moieties are the binding site for the next phosphoramidite building block. The cycle continues by further removal of a new protecting groups and accumulation of the subsequent phosphoramidite at the vacant positions.

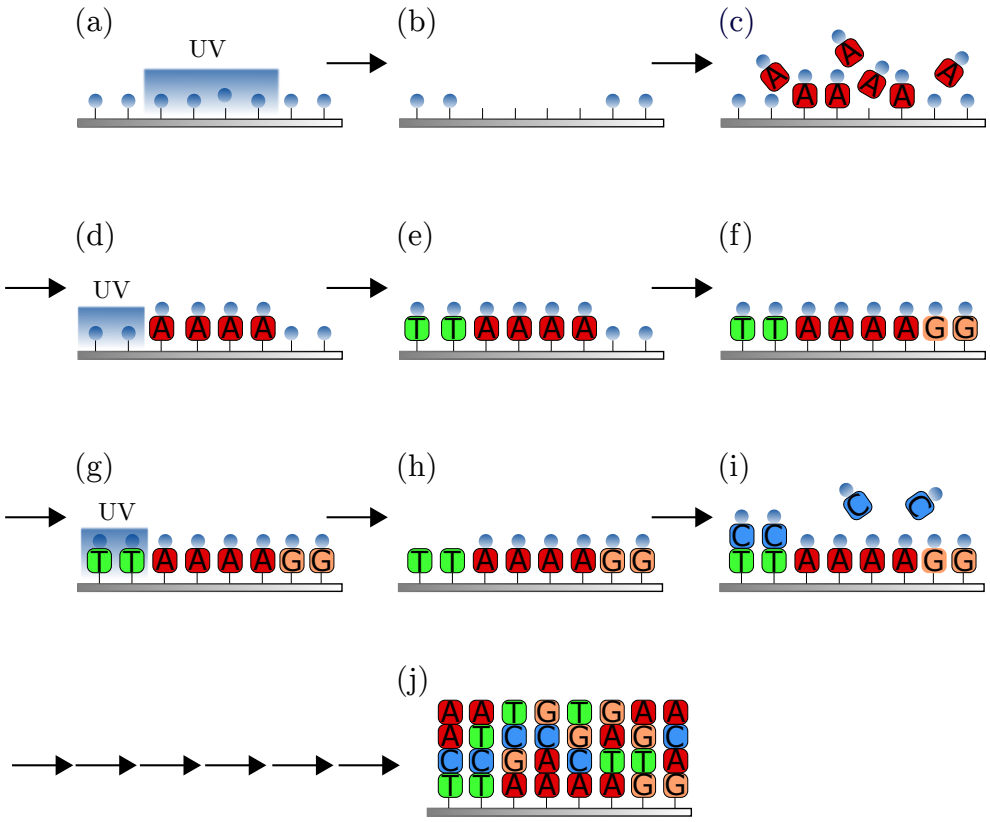

Figure 2.8. Light directed in situ synthesis for DNA microarray. (a) The microarray substrate is initially functionalized with protecting groups NPPOC (blue spheres). (b) The protecting groups are removed at the desired feature areas by the radiation of UV light. (c) The free hydroxyl moieties created by the photo-cleavage of the protecting groups are the binding sites for the subsequent adenosine-phosphoramidite. The protecting groups imported with the building blocks prevent further couplings. (d) Removal of the protecting groups for the probes which their first base is thymine. (e) Coupling of the next thymine-phosphoramidite. (f) The process for coupling new amidite building blocks is repeated until the desired sequences are generated at each feature area on the microarray. (adapted from [92])

2.6 Fluorescence Anisotropy

Fluorescence light emitted during an optical dipole transition of an atom or molecule is polarized. Depending on the specific transition and angle of observation, the light might be linearly, circularly, or in general elliptically polarized. For a situation as depicted in Fig. 2.9: A laser emits light in x-direction, which is linearly polarized along the y-axis. A sample, a single fluorescent molecule, absorbs the light and becomes excited. The subsequently emitted fluorescence light is typically linearly polarized as well. By observing the fluo-

rescence along the z-direction, we find that it exhibits polarization components parallel and perpendicular to the initial excitation polarization, characterized by the respective intensities $I_{\parallel}$ and $I_{\perp}$. These intensities indicate the orientation of the emission dipole within the xz-plane: assume for simplicity that excitation and emission dipole of the fluorophore point in the same direction and the fluorophore cannot rotate. In this case the excitation and emission are parallely polarized and all fluorescence intensity will be found in $I_{\parallel}$. If on the other hand the fluorophore rotates during the time passing between excitation and emission, a part of the intensity will appear in $I_{\perp}$. Fluorescence anisotropy is a technique that exploits this transition of fluorescence intensity from $I_{\parallel}$ to $I_{\perp}$ in order to gain information on the rotation velocity of a molecule and thereby its mass [121].

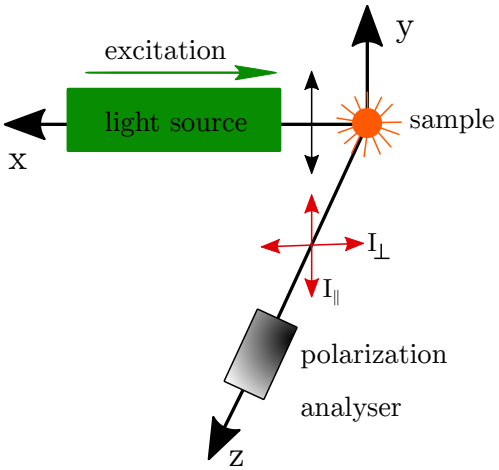

Figure 2.9. Schematic drawing of the anisotropy measurement setup. The laser which emits light at x-direction and linearly polarized along the y-axis excites a fluorescent molecule located in the origin of the axis. The emission is accordingly linearly polarized and detected along the z-direction. It displays two polarization components, parallel $I_{\parallel}$ and perpendicular $I_{\perp}$ to the excitation polarization.

In the previous consideration we assume a single fluorescent molecule as a sample, while in reality the sample contains a large number of molecules, which are interacting with their environment. As a consequence, the detected intensities $I_{\parallel}$ and $I_{\perp}$ do not picture the emission pattern of a single dipole anymore but rather exhibit the polarization of a statistical mixture. Fig. 2.10 illustrates the situation in more details. The top row represents a sample solution containing fluorescent molecules, whose dipole moments are randomly oriented as indicated by arrows. The bottom row shows distribution functions that qualitatively describe the respective orientations. (a) shows the molecules prior to the excitation. The random orientation corresponds to a uniform, i.e. constant, distribution. In (b) the molecules are exposed to

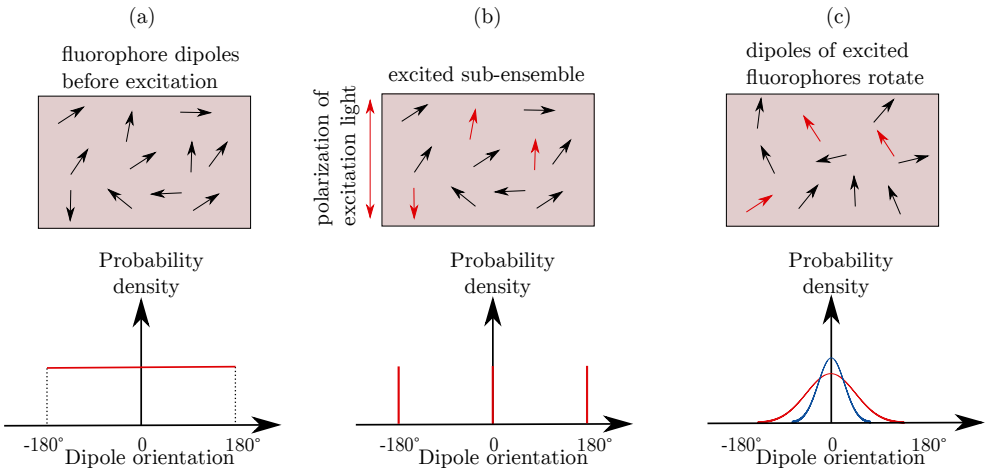

Figure 2.10. The top row: (a) The sample contains fluorescent molecules, which their dipoles are randomly oriented as indicated by arrows. (b) Exciting the sample with a linearly polarized light, only the fluorophores with a dipole moment parallel to the polarization of the excitation light are stimulated (red arrows). (c) Due to the rotational diffusion of the molecules, the dipoles of the excited fluorophores rotate and accordingly the distribution of excited molecules broadens. The respective orientations of the molecules are described by distribution functions as illustrated in a bottom row: (a) the constant distribution for randomly oriented molecules, (b) a sharp delta-like distribution of excited molecules and (c) a broaden distribution caused by rotational diffusion.

the linearly polarized light, those with dipole moments oriented in parallel to the excitation polarization are stimulated (red arrows). This step, often referred to as photoselection, leads to a sharp delta-like distribution of excited molecules. (c) The fluorophores start to randomly collide with other molecules from their environment causing them to rotate at different velocities and in different directions, which is known as rotational diffusion [121]. The distribution of excited molecules broadens due to the rotational diffusion. The extent of the polarization of the sample is described as an anisotropy and it depends on how far the molecule rotates during the lifetime of its excited state. The anisotropy r is defined as a ratio of the polarized component to the total intensity according to

$$r = \frac{I_{\parallel} - gI_{\perp}}{I_{\parallel} + 2gI_{\perp}}, \tag{2.18}$$

where the grating factor g equals the ratio of both detector sensitivities for the horizontally and vertically polarized components of the emission, and therefore nothing but a setup related parameter (cf Sec. 3.2.3 for more information). It

can be shown that for $g = 1$ the anisotropy can assume values between -0.2 and 0.4, where $r = 0$ corresponds to $I_{\parallel} = I_{\perp}$ and therefore to entirely depolarized light [121]. According to our discussion, it is clear that the anisotropy must depend on how far a fluorophore rotates in average before the fluorescence light is emitted. The relation between the characteristic rotational correlation time ϕ and the life time τ to the measured anisotropy r is given by the Perrin equation according to

$$r = \frac{r_0}{1 + \tau/\phi}, \tag{2.19}$$

where r_0 is the intrinsic anisotropy in the absence of any depolarizing processes such as energy transfer or rotational diffusion [121]. It is clear that the anisotropy for the fast decay (small τ) or slow rotation (large ϕ) is maximized [121]. It is often interesting to observe the variation of anisotropy due to the changes in the molecular mass and accordingly the rotation of the molecule. For a good sensitivity of anisotropy r, the lifetime of the fluorophore must be in the order of the rotational correlation time of a molecule. For instance the rotational diffusion time for a 12-mer DNA strand is around 6 ns [122], which is indeed comparable to the lifetime of most available fluorophores which are in the range of 0.1-10 ns [122].

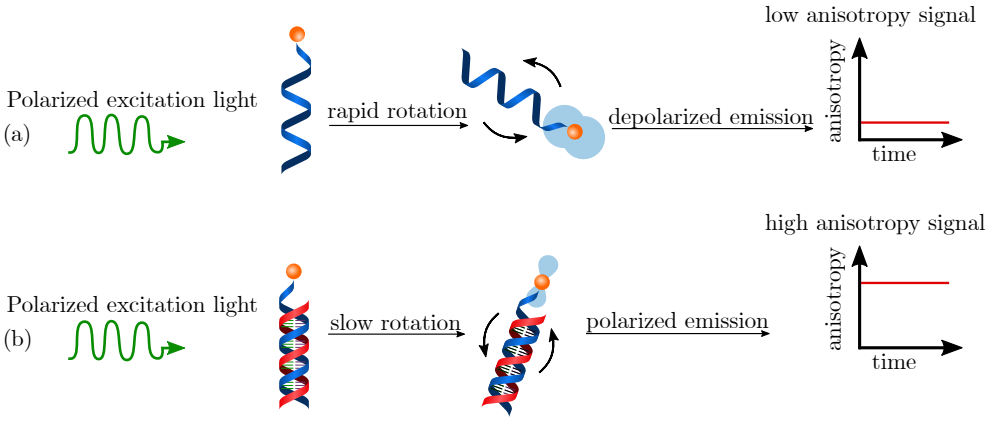

Figure 2.11. (a) A fluorophore which is attached to the DNA strand in both cases is excited by a linearly polarized light. In (a) a single labeled DNA strand rotates very fast since it is very light. This leads to a depolarization of the emission light and accordingly low anisotropy signal. In (b), however, the fluorophore is attached to a double stranded DNA which is heavier and therefore the whole molecule here rotates slower, leading to a polarization of the emission and a measurable anisotropy signal.

If two molecules bind to each other, their rotational diffusion due to the increase in a size of a complex decreases. This leads to an increase in the anisotropy signal. This can be applied to study molecular reaction for instance for DNA hybridization as illustrated in Fig. 2.11. In (a) a fluorescent

molecule which is attached to a single DNA strand is excited by a linearly polarized light. Since the overall molecule is comparatively light, it exhibits a rapid rotational diffusion. As a consequence, the anisotropy of the emission is depolarized and therefore we measure a low anisotropy signal. In (b), however, the hybridized DNA double helix has a larger mass and accordingly it rotates slowly. The emitted fluorescence in this case retains more of the original polarization resulting in a higher anisotropy signal. The samples in our experiments do not contain only single or double stranded DNA, but mixtures of both. The overall anisotropy of a mixture of N fluorescent species is a weighted average written as

$$r = \sum_{i=1}^{N} q_i r_i, \tag{2.20}$$

where i indicates a different emitting species which emits fluorescence while q_i and r_i are the molar fraction and the anisotropy of each components, respectively. In Sec. 4.3 we will show how we use this principle to determine the binding affinity of two DNA strands by measuring the anisotropy as a function of the unlabeled DNA concentration.

2.7 Fluorescence Correlation Spectroscopy

The fluorescence emission of a given fluorescent molecule (fluorophore) $I(t)$ does not certainly remain constant as a function of time t, but it fluctuates over time. The fluctuations may partly be random originating from background fluorescence or noise associated with the detector, but they can also reflect important information on an emission dynamics of a system. These fluctuations stem from the variation in the emission properties of the fluorescent species due to the molecular dynamics or interactions [62,123–126] or the changes in the number of fluorescent emitters in the observation volume [127]. The latter is used in fluorescence correlation spectroscopy (FCS) to investigate the diffusion of a fluorescent particle in a liquid solution.

The basic idea of FCS is based on collecting fluorescence light only from a small detection volume, which ideally contains only a single species in average. Whenever a fluorescent particle diffuses in or out of the detection volume, the fluorescence intensity significantly changes. The temporal variation in intensity is measured in terms of the autocorrelation of the emission signal and the time constants of the fluctuation can be directly extracted from it. On that account the intensity is written as $I(t) = \langle I(t) \rangle + \delta I(t)$, where $\langle \bullet \rangle$ denotes a temporal averaging and $\delta I(t)$ are the fluctuations of the signal around $\langle I(t) \rangle$. The normalized autocorrelation $\tilde{G}(\tau)$ as a function of a timelag

τ can then be written as

$$\tilde{G}(\tau) = \frac{\langle I(t) \cdot I(t+\tau) \rangle}{\langle I(t) \rangle^2} = G(\tau) + 1 \quad \text{with} \tag{2.21}$$

$$G(\tau) = \frac{\langle \delta I(t) \cdot \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2}, \tag{2.22}$$

where we used $\langle \delta I(t) \rangle = 0$. Although other possible processes such as chemical kinetics or blinking contribute to the overall dynamics, we are here only interested in the diffusion of the fluorophores described by an autocorrelation function $G_D(\tau)$. In order to obtain an appropriate expression for it one has to consider the exact shape of the detection volume. Most practical implementations of FCS are performed using a confocal microscope: An objective with a high numerical aperture is used to focus the excitation laser light into the solution containing fluorescent sample and collect the subsequently emitted fluorescence emission. The collimated fluorescence is then focused via a tube lens onto a small pinhole, which blocks all out-of-focus light. Only emission light which was emitted from the focal region of the objective is efficiently transmitted. This region is referred to as confocal volume and its shape can be described by a Gaussian function according to

$$f(x,y,z) \propto \exp\left[-2\left(x^2 + y^2 + \frac{z^2}{\kappa^2}\right)/w_{xy}^2\right], \tag{2.23}$$

where z denotes the propagation direction and x and y span the transversal plane. The parameter w_{xy} is the $1/e^2$ radius of the beam waist in the transversal plane. With the beam waist in axial direction w_z, the axial to lateral aspect ratio is given by $\kappa = w_z/w_{xy}$, leading to the mathematical expression of the confocal volume

$$V_c = \pi^{3/2} w_z w_{xy}^2 = \pi^{3/2} \kappa w_{xy}^3. \tag{2.24}$$

Typically it is $\kappa > 1$, which corresponds to a prolate ellipsoidal shape of the confocal volume as shown in Fig. 2.12. Note that in the FCS method Sec. 3.3 we use the transversal diameter $d_c = 2w_{xy}$ rather than the radius. Assuming the diffusion of the particles is caused by Brownian motion [128], it is possible to show that the desired autocorrelation function can be described by [62,129]

$$G_D(\tau) = \left[N\left(1 + \frac{\tau}{\tau_D}\right) \cdot \left(1 + \frac{\tau}{\kappa^2 \tau_D}\right)^{1/2}\right]^{-1}, \tag{2.25}$$

where $N = c \cdot V_c$ is the number of fluorescent molecules in the confocal volume for a given concentration c and the characteristic diffusion time τ_D that a molecule resides within the detection volume in average. As can be seen the height of the correlation in the limit of $\tau \to 0$ is inversely proportional to the mean number of the molecules in the observation volume, i.e.

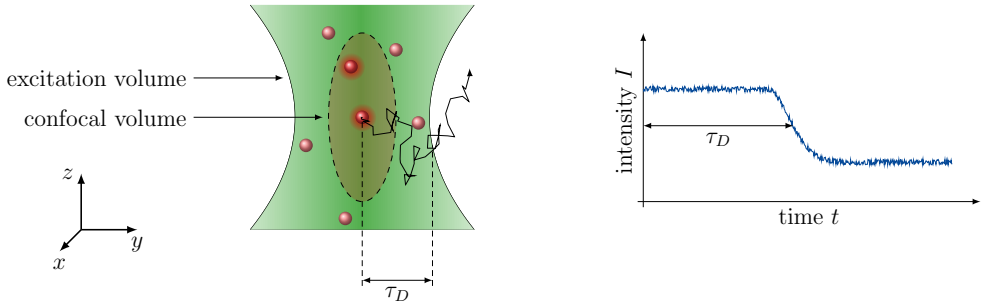

Figure 2.12. Fluorescent molecules shown with red spheres diffuse in and out of focal volume. The fluctuation in fluorescence intensity is defined by the characteristic diffusion time τ_D which is a time that a particle spends inside a focal volume.

$G_D(\tau \to 0) = 1/N$, which confirms our initial assertion that the concentration and the confocal volume should be chosen small. For more details about the experimental setup for FCS method refer to Sec. 3.3.

Besides many applications in biochemical research and biophysics such as studying molecular diffusion or chemical kinetics in equilibrium [130,131], FCS can be used to deduce the binding constant between two interacting molecules in a solution as well [57, 59, 62, 132]. The underlying idea is as following: heavy molecules diffuse slower compared to light molecules and accordingly, it is possible to distinguish between the diffusion times τ_{D_1} of a single free molecule, which is light, and τ_{D_2} of the same molecule bound in a bigger, heavier complex with one or more other particles. In the most simple case of a two-component binding assay, one molecule is fluorescently labeled and the other one remains unlabeled. To determine the binding constant, the titration of the unlabeled molecule in presence of a labeled one is performed. The diffusion times of a free labeled molecule τ_{D_1} as well as the molecular complex in saturation, with an excess of a unlabeled partner τ_{D_2} can be determined by two dimensional single-component diffusion model as

$$G_i(\tau) = 1 + \frac{1}{N_i} \frac{1}{1 + \tau/\tau_{D_i}} \quad \text{with } i = 1, 2. \tag{2.26}$$

For a non-saturation case with a mixture of labeled molecule being in both bound and unbound states, the overall correlation function is given by [62]

$$G(\tau) = 1 + \frac{1}{N} \left(\frac{1-q}{1 + \tau/\tau_{D_1}} + \frac{q}{1 + \tau/\tau_{D_2}} \right), \tag{2.27}$$

where q is the fraction of the molecules which are in a bound conformation. The binding constant is then determined by the concentration dependency of q [62, 133, 134].

Chapter 3

Material and Experimental Methods

In this chapter we introduce three different techniques: light-directed in situ synthesis of DNA microarray, fluorescence anisotropy (FA) and fluorescence correlation spectroscopy (FCS). They were applied in the scope of this work. As the light-directed in situ synthesis setup was built and operated by our collaborators at Vienna university we here only review its major parts and their interplay. In case of the FA and FCS setup we additionally focus on the considered design aspects.

3.1 Light Directed In Situ Synthesis of DNA Microarray

3.1.1 DNA Synthesizer

DNA microarrays allow the simultaneous detection of many, various DNA sequences. They consist of single-stranded DNA oligonucleotides immobilized on a surface which we refer as *probes*. The so-called *features* are small areas containing a large number of surface-attached probes of one kind. These probes are exposed to a bulk mixture of fluorescently labeled target sequences, fully or partially complementary to the surface bound probes. As all probes and their respective positions on the microarray are known, it is possible to obtain information on the simultaneous hybridization of numerous target sequences to the probes. In particular, we employ DNA microarrays to study the competition of various targets binding to the same probe.

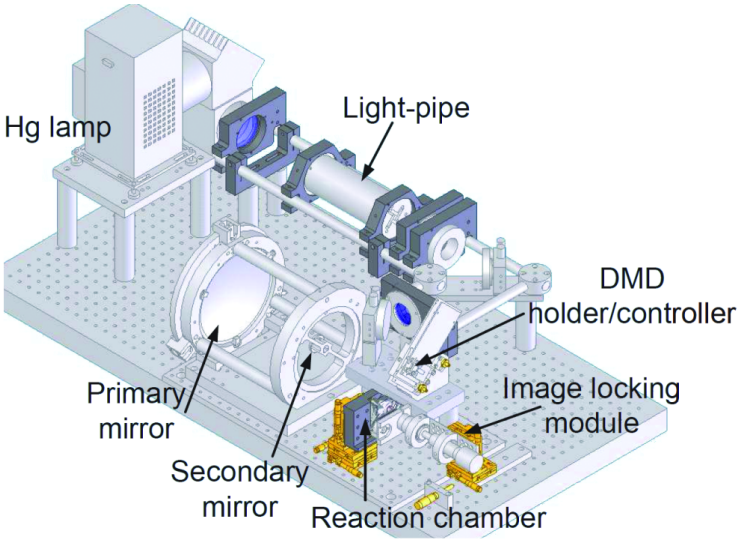

Figure 3.1. Schematic of the optical system of the maskless array synthesizer. The emission of a mercury lamp (Hg lamp) is imaged to a digital micromirror device (DMD) after passing from a homogenizing light pipe and two lenses. The reflected light from the On position mirrors are reflected to the synthesis surface using primary and secondary mirrors. The mirrors are controlled by an external PC (taken with permission from [136]).

The principle of the microarray synthesizer is based on a maskless microprojection photolithography system (MPLS) [92, 136, 137]. The microarray synthesizer that is used in this study consists of two main parts being a nucleic acid synthesizer (Expedite 8909) which delivers solvents and reagents to the reaction chamber as well as an optical system that is illustrated in Fig. 3.1. The optical assembly functions similar to a photolithographic system. A high pressure mercury short-arc lamp (#6286, Newport) is filtered with two 350-450 nm dichroic mirrors (#66218, Newport). The resulting UV light, after passing through a homogenizing light pipe [138], two lenses and a shutter is imaged onto a spatial light modulator, more specifically a digital micromirror device (DMD, 0.7 XGA, Texas Instruments). The DMD contains $1024 \times 768 = 786432$ micromirrors that are individually controlled and positioned in on- or off-state. The DMD is aligned such that light reflected from micromirrors being in an on-state is directed towards the synthesis plane. Light reflected from micromirros in the off-state, on the other hand, is rejected from the optical path. The reflected light from DMD is projected to the synthesis substrate with an Offner relay optical system which provides a 1:1 imaging of the light pattern onto the synthesis surface [139–141]. The external computer that controls the display of the virtual masks on the DMD is synchronized to the DNA synthesizer and therefore to its fluidic system that

provides all chemicals and reagents required for oligonucleotide synthesis. Additionally, it receives a command from the oligonucleotide synthesizer to start a synthesis.

In our experiments we used Schott Nexterion glass slides, which were functionalized with N-(3-triethoxysilylpropyl)-4-hydroxybutyramide (SIT8189.5, Gelest) prior to the synthesis [142]. For more detail about the surface preparation refer to [143]. For the first set of experiments (Sec. 5.4) we used a single mirror design while the features size for the second set of experiments (Sec. 5.6) is 2×2 mirrors.

The probes on the microarray are made of individual building blocks called phosphoramidites. The phosphoreamidites contain nucleobases, deoxyribose, and a phosphorus group at the 3'-end of the sugar, which can react with the nucleophilic hydroxyl-group at the 5'-carbon of the 2-deoxyribose ring of another phosphoramidite (refer to Chap. 2). In this study the phosphoramidite with photolabile 2-(2-nitrophenyl)-propoxycarbonyl (NPPOC) protection group is used. Under UV absorption the photoreactive protection group NPPOC is selectively cleaved to expose the 5' terminal hydroxyl for the next coupling cycle [92,136,144]. The microarray synthesis used in the presented work have a stepwise coupling efficiency of $\geq 99\,\%$ [136].

3.1.2 Hybridization Solution

To prepare the hybridization solution we used 5×SSPE (saline sodium phosphate-EDTA) buffer for all presented experiments. The buffer contains 0.75 M NaCl, 50 mM NaH_2PO_4 and 5 mM Ethylenediaminetetraaceticacid (EDTA) solved in deionized water. We add 0.01% (v/v) Tween-20 to prevent the unspecific adsorption of DNA targets on the microarray surface. The pH value is adjusted to 7.5 using NaOH. The single DNA nucleotides are HPLC-purified and ordered in a lyophilized from Metabion. The samples are dissolved with an amount of deionized water yielding a stock solution of 100 nM concentration. The DNA samples are aliquoted to several smaller concentrations using 5×SSPE buffer and stored in $-22\,^{\circ}\mathrm{C}$. For a short time usage of the aliquots they are kept at $4\,^{\circ}\mathrm{C}$ instead to avoid continues freezing and thawing. In case the nucleotides are modified with fluorophores, they must be protected from any unnecessary light exposure and accordingly kept in dark.

3.1.3 Microarray Scanning and Image Analysis

In order to obtain the hybridization intensities, the fluorescently labeled targets are first optically excited. The image of the hybridized microarray is then obtained by detecting the subsequently emitted fluorescence. In this work, we use two approaches corresponding to two sets of experiments we performed.

(i) The microarrays from the first set of experiments (see Sec. 5.4) were analyzed using a fluorescence microscope (IX81, Zeiss) accompanied by

an electron multiplying EM-CCD camera (C9100-02, Hamamatsu) controlled by the image acquisition software Simple PCI. To extract the average intensities of each feature from the obtained pictures, a home-built java program called ScanArray is used. The detailed information about the software is available at [92].

(ii) For the second set of experiments (compare Sec. 5.6) we need to remove the hybridization solution from the surface before scanning the array. On that account we first wash the microarray surface using a non-stringent washing buffer with high salt concentration (6×SSPE) for one minute to stabilize any possible duplexes. Afterwards we use a final washing buffer (0.1×saline-sodium citrate) for a few seconds. This step removes remaining salt from the previous washing step and thereby prevents salt from crystallizing on the microarray during subsequent drying. To scan the microarrays we employ a Genepix 4400A Molecular device scanner with a resolution of 2.5 μm [144,145]. For quantitative image analysis of fluorescence intensities on a microarray we use a NimbleScan (Roche-NimbleGen) [146], which assigns the obtained intensities to each feature based on the corresponding microarray design [146].

3.2 Fluorescence Anisotropy

As discussed in Chap. 2 fluorescence anisotropy (FA) measurements are based on the rotational diffusion of the fluorescent molecules. Therefore, FA is a suitable method to detect changes in the rotational movement of molecules due to a change of their mass. In the scope of this work, we employ FA to study DNA hybridization under different conditions.

The home-built setup is illustrated in Fig. 3.2. We use a cw frequency doubled Nd:YAG laser (Compass 215M-50, Coherent) at an output wavelength of 532 nm to excite the fluorophore attached to the DNA strands. The excitation laser power varies, however, depending on the fluorophore and its concentration. As the laser can only be operated at a fixed optical power we attenuate it by an appropriate set of neutral density filters (ND filters, Thorlabs) to the desired value. For a better illumination of the sample we expand the laser beam by a factor of 5 using a Galilean telescope, combination of two lenses (ACN127-020-A with f=-20 mm and AC254-100-A-ML with f=100 mm, Thorlabs). Subsequent to the telescope, the excitation light is forwarded to a polarizing beam splitter (PBS A, item CCM1-PBS25-532/M, Thorlabs). While its horizontally polarized element is blocked, the vertically polarized element of a laser light enters a cuvette (105.250-QS, Hellma) containing the sample. Here, the fluorescently labeled DNA strands are excited and the subsequently emitted fluorescence is collected by a plano-convex lens (LA1540-A, Thorlabs) perpendicular to the propagation direction of the excitation light. The residual of the excitation laser light is blocked with a notch

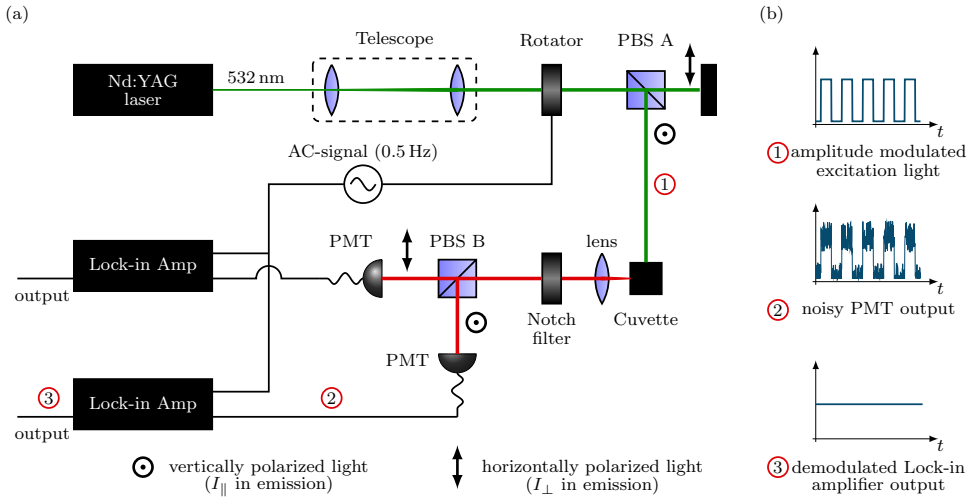

Figure 3.2. Illustration of the home-built experimental setup used to measure fluorescence anisotropy. The excitation light at 532 nm (green) is expanded by a telescope. Its vertically polarized component is separated at a polarizing beam splitter (PBS A) and used to excite a fluorescently labeled sample. The re-emitted fluorescence light (red) is collected by a lens and split at PBS B into components parallelly and perpendicularly polarized with respect to the excitation light. Residual excitation is removed by a notch filter. Two photomultiplier tubes (PMTs) are used to measure the fluorescence intensities in both output channels of PBS B. Lock-in amplifiers in connection with a polarization rotator (driven at 0.5 Hz) and PBS A improve the signal-to-noise ratio. (b) (1) Excitation light is modulated by the rotator and PBS A. (2) The output signal of the PMTs exhibits the same modulation, but superimposed with noise. (3) Eventually, the lock-in amplifiers demodulate the detected signal, yielding a clean DC output. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

filter (NF533-17, Thorlabs). The parallel ($I_\parallel$) and perpendicular components ($I_\perp$) of the emitted light are then separated by a polarizing beam splitter (PBS B, CCM1-PBS21/M, Thorlabs) into two channels. Each component is detected with an individual photomultiplier (PMT, H9305-04, Hamamatsu).

The electrical output of the PMTs is quite noisy, leading to large error bars of the obtained anisotropies. To improve the signal-to-noise ratio we used a DSP lock-in amplifier (7265 Signal Recovery, Advanced Measurement Technology (AMETEK)). To modulate the amplitude of the excitation light an achromatic switchable polarization rotator (ARCoptix) is inserted prior to PBS A. The rotator is driven at a frequency of 0.5 Hz leading to the modulated pattern of excitation light and PMT output as shown in Fig. 3.2 (b). The latter is fed into the lock-in amplifier with the cut-off frequency set to

50 mHz. The demodulated lock-in output is now free of all high frequency noise contributions. In the following we explain how the lock-in amplifier in combination with the rotator works.

3.2.1 Amplitude Modulation

To achieve the required amplitude modulation of the excitation light, we employ the aforementioned sequence of optical polarization rotator and PBS. The rotator switches the polarization of the excitation light between horizontal and vertical at a frequency of 0.5 Hz. As only the vertical component of the light is reflected by PBS, the polarization modulation translates to the amplitude modulation at its output. For a maximal contrast of the modulation pattern the principle axis of a rotator and PBS are aligned with respect to each other. In order to rotate the incident light by 90° the principle axis of the rotator must be either parallel or perpendicular to the polarization of the incoming light. We use a half-wave plate prior to the rotator at the excitation path to insure that the laser light is only vertically polarized. Fig. 3.3 shows the modulated optical output of the PBS (red) along with the electrical output of its driver (black) for different frequencies. For a fast operation of the lock-in amplifier high modulation frequencies are required. It can be seen, however, that the rotator overshoots at its falling edge, which becomes dominant towards high frequencies (compare (a) through (d)). Therefore, for the presented experiments we employed a repetition time of $T = 2\,\mathrm{s}$, corresponding to the frequency of 0.5 Hz as shown in (a).

3.2.2 Lock-in Amplification Technique

Lock-in amplifier can be used to detect a weak DC signal that is mostly obscured by a large background noise. The lock-in amplifier takes the input signal $V_{\mathrm{in}} = V_i \sin(\omega_i \cdot t + \theta_i)$ at a frequency of ω_i and multiplies it with a reference signal $V_{\mathrm{ref}} = V_r \sin(\omega_r \cdot t + \theta_r)$ at ω_r. The resulting mixed signal V_{mix} can be written as

$$\begin{aligned} V_{\mathrm{mix}} =& V_i V_r \sin((\omega_i - \omega_r) \cdot t + \theta_i) \cdot \sin((\omega_i - \omega_r) \cdot t + \theta_r) \\ =& \frac{1}{2} V_i V_r \cos((\omega_i - \omega_r) \cdot t + \theta_i - \theta_r) \\ &+ \frac{1}{2} V_i V_r \sin((\omega_i + \omega_r) \cdot t + \theta_i + \theta_r), \end{aligned} \tag{3.1}$$

where we used the trigonometric equation to separate the expression into the difference frequency and sum frequency terms. In a second step the Lock-in amplifier integrates V_{mix} over a given period. If this time is long compared to the difference and sum frequency terms in Eq. (3.1) and $\omega_i \neq \omega_r$ they will be averaged to zero. If it is, however, $\omega_i = \omega_r$, we obtain the averaged output

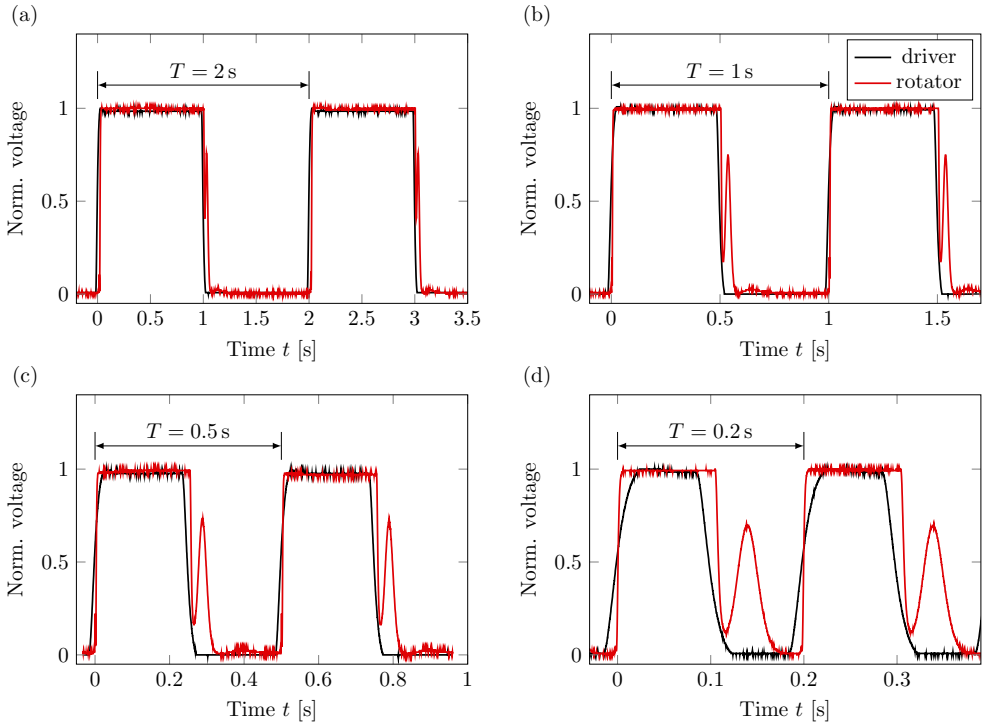

Figure 3.3. The electrical output pulse of the driver (black) and the laser pulse after the PBS (red) at different driven frequencies of 1 Hz, 2 Hz and 5 Hz at (b), (c) and (d) lead to a disturbing overshoot at the rotator signal. However, by setting the frequency to 0.5 Hz as shown in (a) we mostly avoid the overshoot.

$\bar{V}_{\mathrm{mix}}$ as

$$\bar{V}_{\mathrm{mix}} = \frac{1}{2} V_i V_r \cos(\theta_i - \theta_r). \tag{3.2}$$

This effect can be used in our experiment as a narrow filter, which mostly removes the noise caused by the detectors. We modulate a DC signal at a frequency of 0.5 Hz by using the combination of the rotator and a PBS (refer to Sec. 3.2.1). As we use the electric output of the polarization rotator driver as an external reference signal source for the lock-in amplifier, the demodulation and integration described above suppresses all noise frequency components other than 0.5 Hz while recording the desired DC signal. The integration time of the lock-in amplifier is set to 20 s.

3.2.3 Grating Factor Measurement

As illustrated in Fig. 3.2, the anisotropy r is obtained by first separating the emitted fluorescence at a PBS into two components parallelly and perpendicularly polarized with respect to the excitation polarization and subsequently

measuring the corresponding intensities $I_\parallel$ and $I_\perp$. The anisotropy can be computed following the relation

$$r = \frac{I_\parallel - gI_\perp}{I_\parallel + 2gI_\perp}. \tag{3.3}$$

Both detection channels might exhibit differing sensitivities η either due to different transmission losses or detector efficiencies in the individual channels. These effects are accounted for in Eq. (3.3) by the grating factor $g = \eta_\parallel/\eta_\perp$. In the scope of our experiments, we determine g in the following manner: first, a half-wave plate (HWP) is inserted in the emission beam path prior to the PBS. Rotating the HWP by an angle β redistributes the polarization components of the fluorescence between both detectors. In the following we derive the dependency of the detected signals on β and g.

Assume that $I_\parallel$ and $I_\perp$ are the polarization components of the fluorescence prior to the HWP. With its fast axis set to an angle of β with respect to the excitation polarization, the emitted fluorescence is rotated by 2β. Accordingly, at the output of the PBS we find the new intensities $I'_\parallel$ and $I'_\perp$ given by

$$I'_\parallel = I_\parallel \cos^2(2\beta) + I_\perp \sin^2(2\beta) \quad \text{and} \tag{3.4}$$

$$I'_\perp = I_\parallel \sin^2(2\beta) + I_\perp \cos^2(2\beta). \tag{3.5}$$

For what follows, we define the polarization of the emitted light by $P = \left(I_\parallel - I_\perp\right)/\left(I_\parallel + I_\perp\right)$ [121] and the overall intensity by $I_0 = I_\parallel + I_\perp$. Using the trigonometric identities $\cos^2(2\beta) = (1 + \cos(4\beta))/2$ and $\sin^2(2\beta) = (1 - \cos(4\beta))/2$, Eq. (3.4) and Eq. (3.5) can be rewritten as

$$I'_\parallel = \frac{I_0}{2}\{1 + P\cos(4\beta)\} \quad \text{and} \tag{3.6}$$

$$I'_\perp = \frac{I_0}{2}\{1 - P\cos(4\beta)\}. \tag{3.7}$$

Note that the polarization P, similar to the anisotropy r, indicates the rotational mobility of the fluorophores: assuming that the dyes do not rotate at all, we find $P = 1$, which corresponds to $I_\parallel = I_0$ and $I_\perp = 0$, i.e. to emission from dipoles ideally aligned with the excitation light polarization. In case of fast rotation, however, the emission is entirely isotropic ($I_\parallel = I_\perp = I_0/2$) resulting in $P = 0$. As described in Sec. 3.2, we measure the fluorescence intensities using photomultiplier-tubes, whose output is subsequently amplified by Lock-in amplifiers. The obtained output voltages $U_\parallel$ and $U_\perp$ are proportional to the input intensities, which can be written as

$$U_\parallel = \eta_\parallel I'_\parallel = \frac{\eta_\parallel I_0}{2}\{1 + P\cos(4\beta)\} \quad \text{and} \tag{3.8}$$

$$U_\perp = \eta_\perp I'_\perp = \frac{\eta_\perp I_0}{2}\{1 - P\cos(4\beta)\}, \tag{3.9}$$

where $\eta_\parallel$ and $\eta_\perp$ correspond to the desired detection sensitivities. We see that the voltages of both detection channels oscillate with opposite phase around average values given by $\langle U_{\parallel,\perp}\rangle = g_{\parallel,\perp} I_0/2$. Moreover, the amplitude of these oscillations is determined by the polarization P. Eventually, by introducing an angle-offset β_0, we obtain the fitting functions

$$U_\parallel = \langle U_\parallel\rangle \{1 + P\cos(4\beta - 4\beta_0)\} \quad \text{and} \tag{3.10}$$

$$U_\perp = \langle U_\perp\rangle \{1 - P\cos(4\beta - 4\beta_0)\}\,. \tag{3.11}$$

The desired grating factor g can now be simply expressed by

$$g = \frac{\eta_\parallel}{\eta_\perp} = \frac{\langle U_\parallel\rangle}{\langle U_\perp\rangle}. \tag{3.12}$$

The measurements of both voltages $U_\parallel$ and $U_\perp$ as a function of β are shown in Fig. 3.4 for single stranded DNA labeled with (a) Atto 532 and (b) Cy3. The red and blue lines correspond to fits based on Eq. (3.10) and Eq. (3.11). As both parameters P and β_0 must be identical for both channels, they were only fitted for $U_\parallel$, but fixed for $U_\perp$. The averages intensities $\langle U_\parallel\rangle$ and $\langle U_\perp\rangle$ obtained from the fit are indicated by the red and blue dashed lines in Fig. 3.4, respectively. Eventually, their ratio reveals $g_{\mathrm{Atto\,532}} = 0.54$ and $g_{\mathrm{Cy3}} = 0.57$ according to Eq. (3.12).

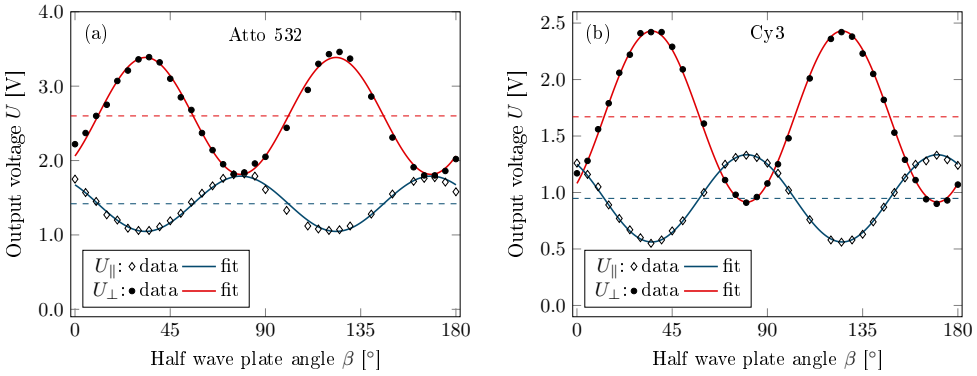

Figure 3.4. Measured voltages of both detectors $U_\parallel$ (circles) and $U_\perp$ (diamonds) as a function of half-wave plate angle β. The blue and red curves correspond to fits based on Eq. (3.10) and Eq. (3.11). The measurements were performed for single stranded DNA samples labeled with (a) Atto 532 and (b) Cy3. The dashed lines correspond to mean voltages $\langle U_\parallel\rangle$ and $\langle U_\perp\rangle$ obtained from the fitting procedure. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

3.2.4 Anisotropy Error Estimation

As we discussed above we employed a lock-in amplifier to reduce the noise of the PMT signals. The integration time of the lock-in amplifier was set

to 20 s, which implies that drifts, oscillations, and noise sources with characteristic time constants above 20 s still appear in our measured time trace, which is displayed in Fig. 3.5 (a) for an exemplary anisotropy measurement with a duration of 10 min. As can be seen, the time trace shows clear oscillations, indicating correlated noise. The standard deviation quantifies errors for uncorrelated noise sources and is therefore unsuitable for our measurements. Instead, we divide the overall measurements into adjacent segments of length τ_{corr}, where τ_{corr} is the characteristic time over which it sustains the correlation. We can now assume that the averages of all individual compartments are independent and their standard deviation is a meaningful measure of the overall error. To obtain τ_{corr}, we evaluate the autocorrelation of the time trace, as shown in Fig. 3.5 (b). Fitting the central peak with a Gaussian function (red curve), we find $\tau_{\mathrm{corr}} = 50\,\mathrm{s}$ (half width at 5% of a maximum), which is in a reasonable relation with the integration time of the lock-in amplifier. In the Chap. 4 all anisotropy results are displayed along with the twofold standard deviation, i.e. a 95% confidence interval, as described here. For this particular time trace shown in Fig. 3.5 (a), we obtain an average anisotropy of $r = 0.0734$ and an absolute error of $\Delta r = 0.0004$, which corresponds to a relative error of 6%, as opposed to a relative error of $> 50\%$ we typically obtained without using a lock-in amplifier.

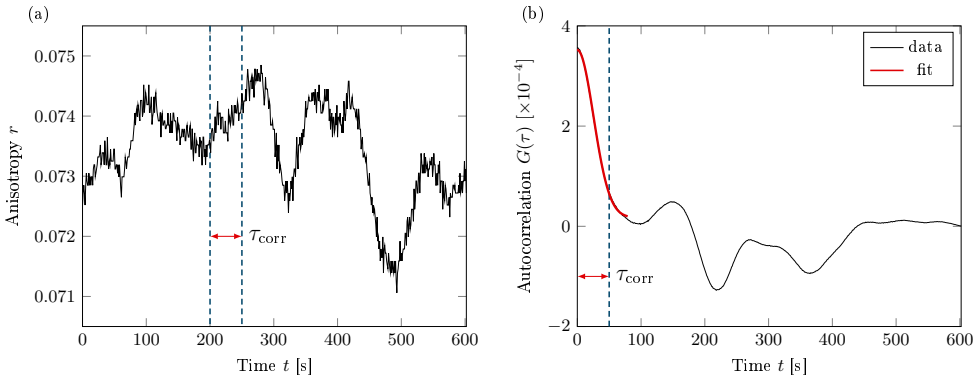

Figure 3.5. (a) The anisotropy as a function of time for a measurement of a duration of 10 min as an example. (b) The corresponding autocorrelation function of the measured anisotropy. Fitting the central peak with a Gaussian function results in $\tau_{\mathrm{corr}} = 50\,\mathrm{s}$. Negative correlation occurs due to the limited time interval.

Note that in the FA experiments the absolute values of anisotropy for the same strand(s) can vary. This depends on the lock-in settings, optical alignment and other technical factors. For measurements with small anisotropy changes the relative errors are apparently larger which may influence the measurement outcome. However, our results are based on relative changes in anisotropy and not the absolute values.

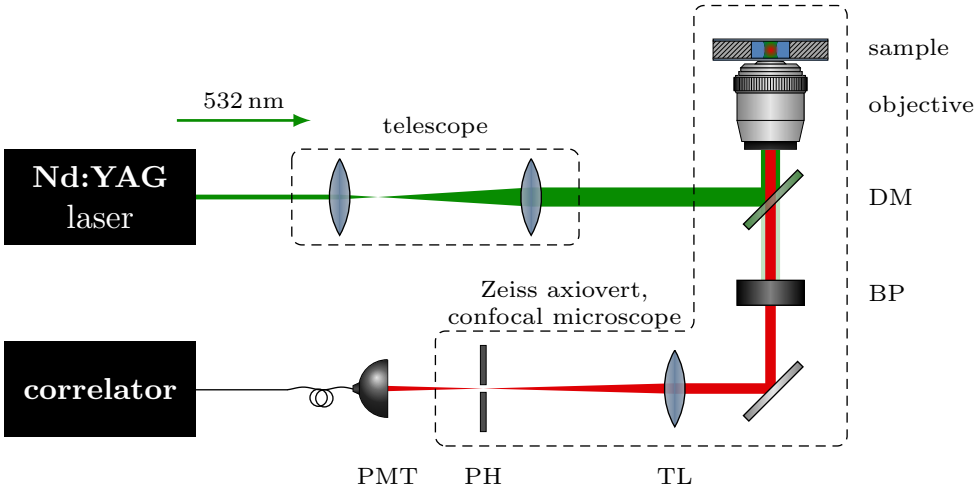

Figure 3.6. The previous FCS setup consist of a commercial inverted microscope. The green laser light (532 nm) after expansion by a telescope is reflected by a dichroic mirror (DM) to a water immersion objective (green path). The laser light is focused to the sample through the objective and excites the fluorescently labeled sample. The emitted light is then transmitted through the DM and is focused through a tube lens (TL) to a $\varnothing$ 50 μm pinhole (PH). A photomultiplier tube (PMT) detects the emitted photons and a correlation hardware records the autocorrelation function (red path).

3.3 Fluorescence Correlation Spectroscopy

Fluorescence correlation spectroscopy (FCS) as we described in Chap. 2 is a technique that is used to investigate the diffusion and the internal kinetics of the fluorescently labeled molecules. Fig. 3.6 shows the schematic of our former setup. Prior to the experiments performed for the presented work, a FCS setup based on a commercial inverted microscope (Axiovert 135, Zeiss) was established within our working group, which was used in the scope of [147]. The setup is schematically depicted in Fig. 3.6 and works as follows: a frequency doubled ND:YAG laser with output wavelength of 532 nm (green beam path) is used as an excitation source. Prior to the microscope, the laser light is expanded with a telescope to illuminate the entire aperture of the microscope objective (water immersion, C-Apochromat 40X, NA=1.2, Zeiss). The excitation light enters the microscope via its rear side and is directed towards the objective by a dichroic mirror (DM). The dichroic mirror is chosen such that it reflects the excitation, but transmits the fluorescence light. The laser light is focused to the sample by the objective, where it excites the fluorophore molecules. The subsequently emitted fluorescence light is collected and collimated by the objective and directed towards the detection channel (red beam path). In order to block residual excitation laser light a band pass filter (BP,

HQ585/40M, Chroma) is inserted to the emission path. Afterwards, the fluorescence is focused by a tube lens (TL, $f = 165\,\mathrm{mm}$, complementary to an objective) onto a pinhole (PH, P50C, $\varnothing$ 50 μm, Thorlabs). Thereby, the overall microscope is operated as a confocal microscope. The pinhole defines the size of the confocal volume. Eventually, we use a photon counting head (H10682, Hamamatsu, photomultiplier tube PMT) to detect the emission. A photon counting hardware correlator (Flex99R-480, Correlator) with a time resolution of 480 ns that is connected to a PC evaluates and records the autocorrelation function. Although this setup was fully functional, it suffered from a number of shortcomings, which restricted its versatility:

(i) The single photon detectors showed afterpulsing, i.e. they generate a second pulse shortly after the detection of a photon, although no second photon was present. This leads to a number of false coincidences at a time scale, where molecular dynamics of interest can occur.

(ii) Due to a small detection efficiency of the PMT ($\eta = 9\,\%$ at 600 nm) the obtained count rates were low. The integration time of correlation measurements scales inverse quadratically with the count rates. Therefore, a series of measurements with a good statistics required a long measurement time.

Fig. 3.7 shows the new setup. The main idea was to construct an entirely homebuilt confocal microscope. This has the advantage that all components are customized and interchangeable and the beam path is accessible and adjustable at any position within the setup in contrast to entirely encapsulated commercial microscope. We used the same green laser for the excitation. However, the light is not directly forwarded to the microscope but first coupled to a single mode fiber (SM fiber, P3-460B-FC-2, Thorlabs) for which we assume a mode field diameter of MFD=3.8 μm[1]. This fiber acts as a spatial mode filter leading to a clean Gaussian beam profile at its output. Moreover, the diameter of the output beam can be chosen easily via the focal length of the collimation lens making the telescope of the previous setup obsolete. For this purpose we employed an aspheric lens (AL, C560TME-A, Thorlabs) with a focal length of f_{col1} =13.86 mm leading to a beam diameter d_{la} of [148].

$$d_{\mathrm{la}} = \frac{4\lambda f_{\mathrm{col1}}}{\pi \mathrm{MFD}} = 2.47\,\mathrm{mm}, \tag{3.13}$$

where $\lambda = 532\,\mathrm{nm}$ is a laser wavelength. Subsequent to the fiber a laser line filter (LLF, FL532-10, Thorlabs) is used to suppress Raman light created by the excitation light inside a fiber. The collimated green light is reflected by a plate beamsplitter (plate BS, BSS10R, Thorlabs, reflectivity 30 %) placed at

[1] Thorlabs reports the MFD of 2.8-4.1 μm at a wavelength 488 nm. Here we assume the middle of this range being 3.5 μm. This corresponds to approximately $(\frac{532}{488})\times$ 3.5 μm=3.8 μm at 532 nm.

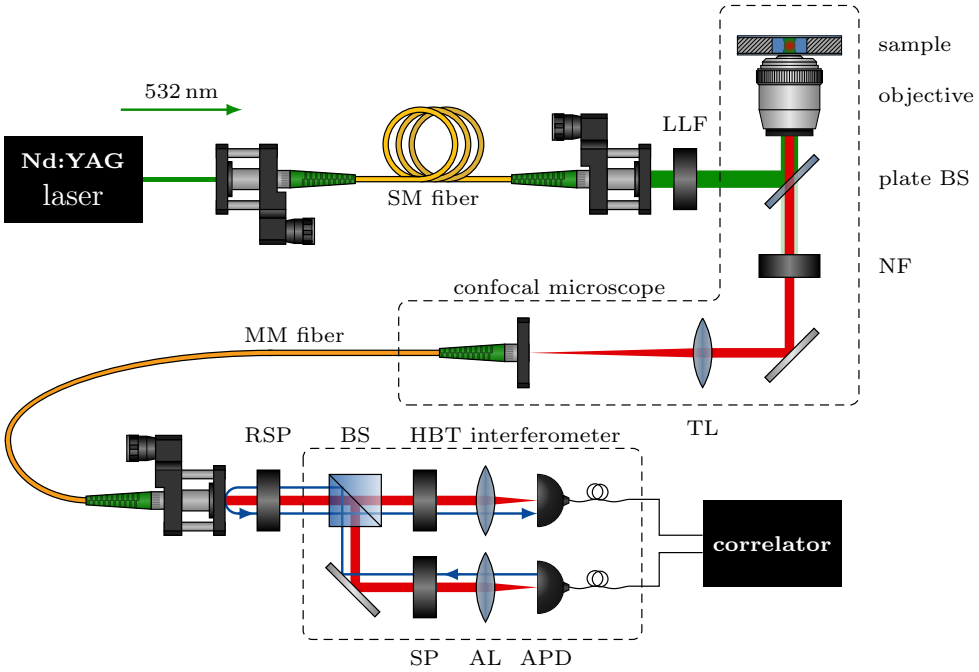

Figure 3.7. The green laser light after coupling in and out of a single mode fiber (SM) and passing through a laser line filter (LLF) is reflected by a plate beam splitter (BS) to the objective. The excitation light is focused through the water immersion objective to the fluorescently labeled sample. The fluorescence emission is collected by the objective and is transmitted through the plate BS. A notch filter (NF) removes the excitation light in the emission path. The fluorescence light is imaged by a tube lens (TL) to the core of the multimode fiber (MM). A shortpass filter (RSP) is used to block the Raman light coming from the water and a sample. The emission is then separated into two channels using a beam splitter (BS) and is focused into two detectors (APD) using two aspheric lenses (AL). Two shortpass filters (SP) are employed in front of each APD to block the afterglow. Eventually a correlator cross-correlates the outputs of the detectors.

45 ° with respect to the incident beam. Here, the achromatic BS was preferred to a DM to sustain flexibility in choosing the excitation wavelength and employed fluorophore. The excitation light passes through the water immersion objective (C-Apochromat, magnification $M = 40$, Zeiss) and enters a chamber including a fluorescently labeled sample. As the diameter of the collimated excitation light d_{la} is much smaller than the aperture of the objective (in the order of 1 cm), we are allowed to use Gaussian beam optics to determine the diameter of the excitation volume d_{ex} in the sample [149]. We consider the lens system of collimation lens and objective as a microscope, imaging the fiber core into the sample. Accordingly, we find a magnification of $f_{\mathrm{obj}}/f_{\mathrm{col1}}$

with f_{obj} =4.13 mm the focal length of the objective and thereby

$$d_{\mathrm{ex}} = \mathrm{MFD} \times \frac{f_{\mathrm{obj}}}{f_{\mathrm{col1}}} = 1.13\,\mu\mathrm{m}. \tag{3.14}$$

The laser light excites the fluorescently labeled molecules which in turn emit fluorescence. The emission is collected by the objective and 70 % of it is transmitted through a plate BS. A notch filter (NF, NF533-17, Thorlabs) is used to block the residual laser light in the emission path. We use a tube lens (TL, 165mm-10x Axio Imager, Zeiss) to image the fluorescence light into the d_{core} =50 μm core of the multimode fiber (MM fiber, M42L05, Thorlabs). The core of the fiber plays a role of a PH and accordingly defines the confocal volume. In particular it determines the resolution of the microscope in axial direction. Due to the limited size of all optical elements a single point in the sample is imaged to an Airy disc with a diameter of $\approx 1.22 \cdot M\lambda/NA$ = 21.6 μm, where NA = 1.2 is the numerical aperture of the objective. As the Airy disc is less than half as wide as the pinhole, the confocal microscope is operated in the geometrical optics limit and the diameter of the confocal volume in axial direction is given by [149]:

$$d_{ax} \approx 1.699 \times \frac{\sqrt{2} \cdot n \cdot d_{\mathrm{core}}}{M \cdot NA} = 3.34\,\mu\mathrm{m}, \tag{3.15}$$

where $n = 1.33$ is the refractive index of water as an immersion medium. In lateral direction we find the diameter of the confocal volume d_c given by

$$d_c = \frac{d_{\mathrm{core}}}{M} = 1.25\,\mu\mathrm{m}. \tag{3.16}$$

Since $d_{\mathrm{ex}} < d_c$, we employ d_{ex} in order to calculate the confocal volume which is given by [141]

$$V_c = \frac{1}{8}\pi^{\frac{3}{2}} d_{\mathrm{ex}}^2 \cdot d_{ax} = 2.97\,\mathrm{fL}. \tag{3.17}$$

To obtain the desired autocorrelation G(τ), we detect the fluorescence photons in a so-called Hanbury-Brown-Twiss interferometer (HBT) [150]. On that account, we first collimate the fluorescence light exiting the MM fiber using a collimation lens with f_{col2} =30 mm (#49-662, Edmund optics). The light then is split at a BS (BS016, Thorlabs) and focused using two lenses (#49-659, Edmund optics) to the active area of single photon counting silicon avalanche photodiodes (APD, PDM series, Picoquant) that detect the emission in each channel. Cross-correlating the output of each APD yields G(τ). This approach has mainly two advantages:

(i) The afterpulses of both detectors are not correlated. Accordingly, no coincidences due to after pulsing will appear in G(τ) when using two detectors.

(ii) The employed APDs feature a high quantum efficiency (η_{APD} =45 % at 600 nm) compared to the PMTs (η_{PMT} =9 % at 600 nm), speeding up the measurements by a factor of 25. However, to achieve the full quantum efficiency, it is necessary to illuminate the entire active area of the APD without cutting any light. Here we choose to focus the light with aspheric lenses (AL, #49-659, Edmund optics) with f_{APD} =20 mm, leading to the beam waist of

$$d_{\mathrm{APD}} = d_{\mathrm{core}} \times \frac{f_{\mathrm{APD}}}{f_{\mathrm{col}}} \approx 33\,\mu\mathrm{m}, \tag{3.18}$$

which equals around 70% of the APD's active area of $\varnothing$=50 μm. To verify the improved detection efficiency we feed the same signal to the PMT and to the HBT interferometer and obtain the count rates of $c_{\mathrm{PMT}} = 50\,\mathrm{kHz}$ and $c_{\mathrm{APD}} = 124\,\mathrm{kHz}$ for a single APD. Considering the maximum afterpulse probability of 3% (factor of 1.03 in equation below) for both detectors and the HBT transmission of the 93% (factor of 0.93) between MM fiber and APD, we obtain the effective count rates (c'_{PMT}, c'_{APD}) as following:

$$c'_{\mathrm{PMT}} = \frac{50\,\mathrm{kHz}}{1.03} = 49\,\mathrm{kHz}. \tag{3.19}$$

$$c'_{\mathrm{APD}} = \frac{124\,\mathrm{kHz}}{0.5 \times 0.93 \times 1.03} = 259\,\mathrm{kHz}, \tag{3.20}$$

where factor of 0.5 considers that the photons are distributed to two detectors at BS inside the HBT interferometer. The ratio of the count rates is $c = \frac{259}{49} = 5.2$, which is exactly the ratio of the detection efficiencies reported in the data sheets of both detectors (η_{PMT} =9 %, η_{APD} =45 % at 600 nm). Single photon counting APDs often show a so-called afterglow [151], i.e the emission of a light pulse during its avalanche process. These photons have a broad spectral range between 700 nm to 1000 nm with a peak at 950 nm. In our HBT interferometer these photons might travel the same path as the fluorescence, be reflected at the facet of the MM fiber and eventually be detected by the other APD, causing cross-talk (compare blue beam path in Fig. 3.7). We suppressed the cross-talk by using two shortpass filters (SP, #64-606, Edmund optics) with a filter edge at 725 nm, one in front of each APD. To measure the correlation we connect the APDs to the two channels field-programmable gate array (FPGA) evaluation kit (Spartan SP-605, Xilinx), which is operated using the software presented in [152]. The time resolution of the device is 4 ns which allows us to access the shorter times compared to the Flex99R-480.

The interaction of an excitation laser light and the water on the objective and inside the sample leads to Raman scattered photons. These photons, if detected, cause a strong background, which effectively decreases the correlation height. Therefore, it is necessary to spectrally remove Raman photons. To identify a suitable filter, we measured the Raman peak of the distilled water for different excitation wavelengths (440, 450, 470, 480) nm using a spectrofluorometer (FP-6500, JASCO). Fig. 3.8 (a) shows the recorded spectra,

which exhibits the Raman peaks for different detection wavelengths λ_{ex} while Fig. 3.8 (b) shows their central Raman wavelength λ_{Raman} as a function of the excitation wavelength λ_{ex}. The central energy of the Raman peak E_{Raman} is always shifted by a constant amount c, the wavenumber of the water, with respect to the excitation energy E_{ex} according to $E_{Raman} = E_{ex} - c$. As the wavelength is inversely proportional to the energy, we obtain

$$\lambda_{Raman} = \left(\frac{1}{\lambda_{ex}} - c\right)^{-1}. \tag{3.21}$$

We fit the data in Fig. 3.8 (b) using Eq. (3.21), where c is a fitting parameter. We find $c = 3350\,\mathrm{cm}^{-1}$ which is in a perfect agreement with literature [153]. Using the value of c, we find a Raman peak of $\lambda_{Raman} = 666\,\mathrm{nm}$ for the green laser excitation $\lambda_{ex} = 532\,\mathrm{nm}$. To block the Raman emission we use a Raman shortpass filter (RSP, FESH0600, Thorlabs) with a cut-off wavelength at $\lambda = 600\,\mathrm{nm}$.

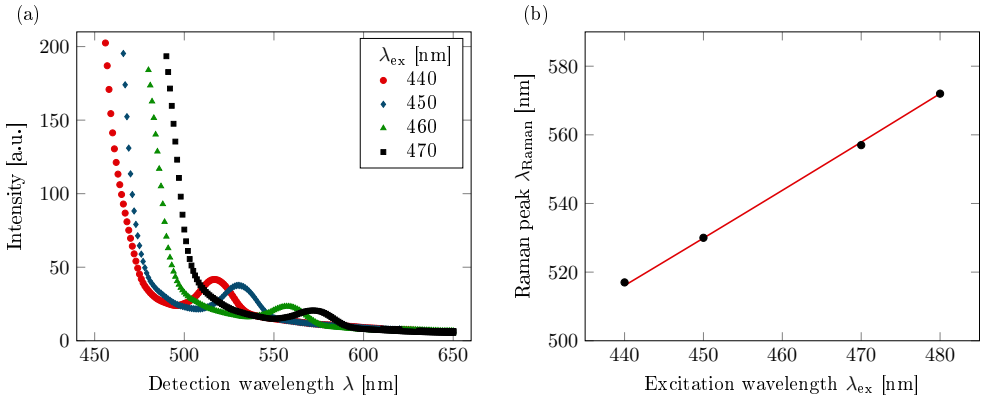

Figure 3.8. (a) The Raman emission for different excitation wavelength λ_{ex}. (b) The Raman peak λ_{Raman} as a function of excitation wavelength λ_{ex}. The fit gives a wave number c of water.

In order to verify the operation of our setup we determined the diffusion constant of the Rhodamin 6G and compared it with the reported value from the literature. Fig. 3.9 shows the measured autocorrelation $G(\tau)$ of the diluted Rhodamin 6g sample. The red curve shows a fit using Eq. (3.22)

$$G(\tau) = a + \frac{1}{N} \cdot \left(\frac{1}{1 + \tau/\tau_{diff}}\right), \tag{3.22}$$

where a, N and τ_{diff} are treated as free parameters. From the fit we obtain $\tau_{diff} = 0.32\,\mathrm{ms}$. The diffusion time can be related to the diffusion coefficient D_{coe} via $D_{coe} = d_{ex}^2/16\tau_{diff}$ [141][2]. Using d_{ex} from Eq. (3.14) and $\tau_{diff} =$

[2] In this equation we should use the diameter of the confocal volume d_c but since $d_{ex} < d_c$, we used d_{ex} instead

0.32 ms we find the diffusion coefficient $D_{\text{coe}} = 2.5 \times 10^{-6}\,\text{cm}^2/\text{s}$ in excellent agreement with the reported value $D_{\text{coe}} = (2.8 \pm 0.3) \times 10^{-6}\,\text{cm}^2/\text{s}$ [133, 154].

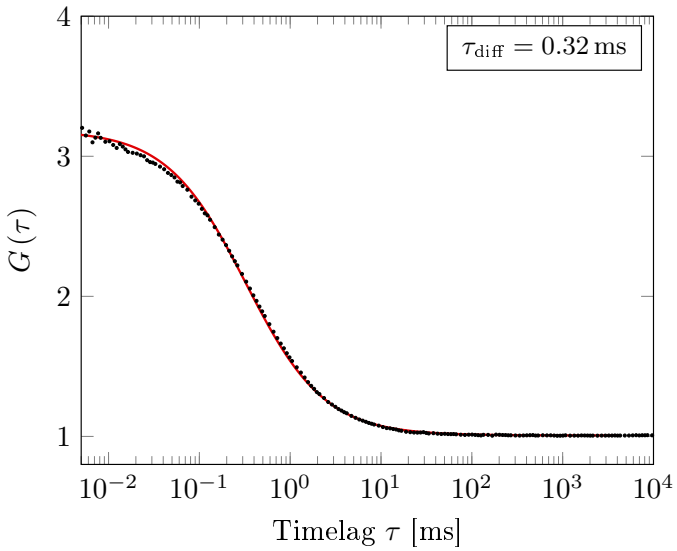

Figure 3.9. Autocorrelation $G(\tau)$ for the Rhodamin 6G. By fitting the data we determine the diffusion time τ_{diff}.

We planned to use our FCS setup to perform supplementary experiments with two color cross-correlation using a green laser to excite both fluorophores. This enables to study the binding and unbinding dynamics of two DNA strands, tagging each with a different fluorophore and cross-correlating their corresponding emission. Such experiments require a pair of two fluorophores that fulfill a number of conditions:

(i) Both can be efficiently excited with green laser

(ii) Their emission spectra need to be well separated, so that optical cross-talk between the detection channels can be supressed using appropriate filters.

(iii) The fluorophores need to be attachable to DNA strands

One of the best available combinations of fluorophores we are aware of that meeting all conditions is Yakima yellow and Atto 647N, whose ideal emission spectra are illustrated in Fig. 3.10 (a) [155]. The employed filters are additionally shown. At the Yakima channel we placed a shortpass filter (SP600, FESH0600, Thorlabs) with the edge at $\lambda = 600\,\text{nm}$ separating it from the Raman light and the emission of Atto 647N. For the Atto 647N channel we used a longpass filter (LP665, et665LP, Chroma) with the edge at $\lambda = 665\,\text{nm}$ instead. Moreover, at this channel we employed a shortpass filter (SP725,

#64-606, Edmund optics) against the afterglow. The integrated areas under the red and blue shaded areas are proportional to the detected fluorescence intensities within the filter windows. We observe that the ideal spectra are well separated. Fig. 3.10 (b) shows the measured spectra of both fluorophores taken with a commercial grating spectrometer (SP2500A, Princeton instruments). It can be seen that both spectra are significantly shifted and broadend compared to (a). In particular, the Yakima fluorescence leaks into Atto 647N channel (gray area) contributing around 7 % to its overall signal. The leakage causes a false cross-correlation that leads to a wrong interpretation of the measurements. As no better fluorophore combination could be found, this approach to two color correlation was not further pursued in scope of the present work.

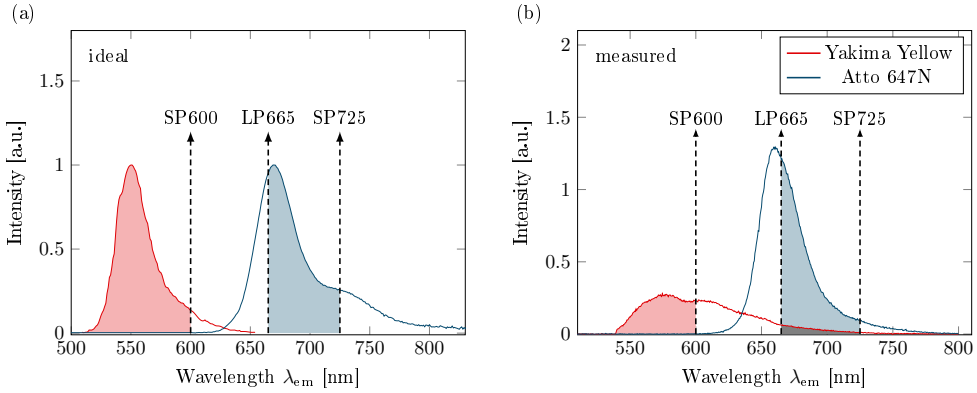

Figure 3.10. (a) The reported emission spectra of Yakima yellow (red) and Atto 647N (blue). (b) Experimentally determined spectrum of Yakima yellow (red) and Atto 647N (blue). The dashed lines in (a) and (b) show the cut-off wavelength of the applied filters. The red and blue regions show the detected emissions of Yakima in channel 1 and channel 2, respectively. The Yakima emission in (b) is shifted compared to (a). The gray region in (b) is the leakage of Yakima into Atto 647N channel.

Chapter 4

DNA Hybridization in Bulk

In this chapter we investigate the binding of DNA strands in bulk experiments. We first determine the binding affinity of two single DNA strands in a pairwise manner using fluorescence anisotropy and fluorescence correlation spectroscopy. It turns out that the results from both techniques disagree for some of the obtained binding constants. This suggests that the definition of a bound state depends on the given experimental situation. Moreover, we study competitive situations where two target strands compete to bind to a common complementary probe. In most cases the results can not be explained based on the determined individual binding constants of the competitors. We conclude that both binding partners interact, leading to cooperative behavior.

4.1 Three Strands Binding Model

To interpret the binding of two DNA strands we use the extended Langmuir isotherm as explained in Sec. 2.4. In the present section we derive model functions that will be used to interpret our measurements for three strands. On that account we establish rate equations for the participating species that are solved at equilibrium, comparable to our approach presented in Sec. 2.4. They yield equilibrium concentrations c_i^{eq} of species i as a function of the underlying binding constants. Employing c_i^{eq} in Eq. (2.20) leads to an anisotropy expression that we use to fit the experimental data.

All competition experiments including three strands that are considered in this work have in common a labeled target L and an unlabeled target U competing to bind to the same probe P. While the initial concentrations of

L and P, c_L^{ini} and c_P^{ini}, are fixed, the concentration of U, c_U^{ini} is varied. Both targets L and U are designed such that they do not hybridize to each other. In the following we develop models describing two situations: **(i)** both targets L and U do not overlap on the probe, thus they can simultaneously bind to P and form a triple P-L-U (Fig. 4.1 (a)). **(ii)** Since the targets overlap on the probe they do not bind to P at the same time. They form the duplexes P-L and P-U with the probe (Fig. 4.1 (b)). We refer to these two situations as simultaneous and exclusive binding, respectively.

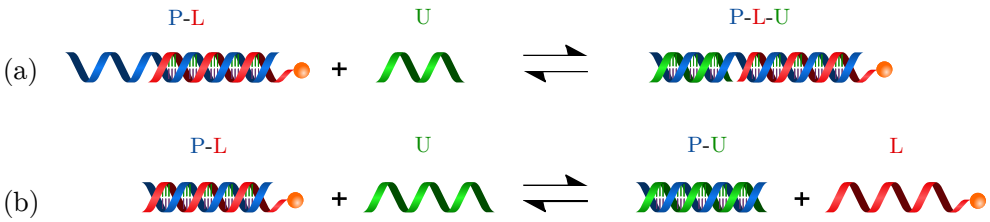

Figure 4.1. (a) Simultaneous binding: the unlabeled green target (U) and the red labeled target (L) do not overlap on the blue probe (P). Therefore, they bind simultaneously to the probe and form a triplet P-L-U. (b) Exclusive binding: the red labeled and green unlabeled targets overlap on the probe. The unlabeled target removes the labeled target and forms a duplex P-U with the probe.

4.1.1 Simultaneous Binding Model

Considering the two target strands L and U do not overlap on the probe P, we assume that the hybridization of one of them to the probe is not influenced by the presence of the other. Introducing the target U to a solution containing P and P-L leads to the formation of duplex P-U and triplet P-L-U. We here show that in this case the change in the anisotropy still follows the extended Langmuir isotherm as explained in Sec. 2.4. The relevant reactions are given by

$$P + U \underset{k_-^U}{\overset{k_+^U}{\rightleftharpoons}} PU \quad \text{and} \tag{4.1}$$

$$PL + U \underset{k_-^U}{\overset{k_+^U}{\rightleftharpoons}} PLU. \tag{4.2}$$

Since we assume that the binding of U to P-L is not influenced by the presence of L, the rate constants k_+ and k_- for both reactions in Eq. (4.1) and (4.2) are equal. The sum of the two corresponding rate equations at equilibrium can be written as

$$0 = k_+^U c_U^{\text{eq}} \left(c_P^{\text{eq}} + c_{PL}^{\text{eq}}\right) - k_-^U \left(c_{PU}^{\text{eq}} + c_{PLU}^{\text{eq}}\right). \tag{4.3}$$

Considering $K_U = k_+^U/k_-^U$ and writing c_U^{eq} and c_P^{eq} as a function of initial concentrations c_U^{ini} and c_P^{ini} we obtain

$$c_{PU}^{\text{eq}} + c_{PLU}^{\text{eq}} = K_U \left(c_U^{\text{ini}} - c_{PU}^{\text{eq}} - c_{PLU}^{\text{eq}}\right)\left(c_P^{\text{ini}} - c_{PU}^{\text{eq}} - c_{PLU}^{\text{eq}}\right). \quad (4.4)$$

Comparing Eq. (4.4) with Eq. (2.10) we find that $c_{PU}^{\text{eq}} + c_{PLU}^{\text{eq}}$, the solution of Eq. (4.4), follows the Langmuir isotherm according to

$$c_{PU}^{\text{eq}} + c_{PLU}^{\text{eq}} = \mathcal{L}\left(c_U^{\text{ini}}, c_P^{\text{ini}}; K_U\right). \quad (4.5)$$

We proceed by introducing the molar fractions q_P^0 and q_{PL}^0 of P and PL, respectively, in case the target strand U is not present, i.e. $c_U^{\text{ini}} = 0$. Since we assume L and U independently bind to the probe, it follows that U hybridizes with probability q_P^0 to P forming P-U and with q_{PL}^0 to P-L forming P-L-U. This can be written as

$$\frac{q_P^0}{q_{PL}^0} = \frac{c_{PU}^{\text{eq}}}{c_{PLU}^{\text{eq}}}. \quad (4.6)$$

Substituting c_{PU}^{eq} from Eq. (4.6) in Eq. (4.5) and using $q_P^0 + q_{PL}^0 = 1$, we can express c_{PLU}^{eq} by

$$c_{PLU}^{\text{eq}} = q_{PL}^0 \mathcal{L}\left(c_U^{\text{ini}}, c_P^{\text{ini}}; K_U\right). \quad (4.7)$$

Writing the total anisotropy r and inserting $q_L + q_{PL} + q_{PLU} = 1$ leads to

$$\begin{aligned} r &= r_L q_L + r_{PL} q_{PL} + r_{PLU} q_{PLU} \\ &= r_L \left(1 - q_{PL} - q_{PLU}\right) + r_{PL} q_{PL} + r_{PLU} q_{PLU} \\ &= r_L + \left(r_{PL} - r_L\right) q_{PL} + \left(r_{PLU} - r_L\right) q_{PLU}. \end{aligned} \quad (4.8)$$

Since any change in c_U^{ini} does not affect the fraction of bound strand L, we find $q_{PL}^0 = q_{PL} + q_{PLU}$. Using this expression in Eq. (4.8) yields

$$r = r_L + \left(r_{PL} - r_L\right) q_{PL}^0 + \left(r_{PLU} - r_{PL}\right) q_{PLU}. \quad (4.9)$$

Applying $q_{PLU} = c_{PLU}^{\text{eq}}/c_L^{\text{ini}}$ and Eq. (4.7) for c_{PLU}^{eq} and introducing the abbreviations $a = r_L + \left(r_{PL} - r_L\right) q_{PL}^0$ and $b = \left(r_{PLU} - r_{PL}\right) q_{PL}^0 / c_L^{\text{ini}}$ leads to the final expression for the anisotropy r

$$r = a + b \cdot \mathcal{L}\left(c_U^{\text{ini}}, c_P^{\text{ini}}; K_U\right). \quad (4.10)$$

Comparing Eq. (4.10) to the extended Langmuir expression for the hybridization of two strands Eq. (2.10), we find that here only a and b are defined differently. We use Eq. (4.10) to fit the anisotropy data in simultaneous binding experiment. While a, b and K_U are fitting parameters, all initial concentrations c_U^{ini} as well as c_P^{ini} are known.

4.1.2 Exclusive Binding Model

In contrast to simultaneous binding situation, here the labeled strand L and the second unlabeled target U compete to bind to the unlabeled probe P. As before, the concentration of the unlabeled strand is varied, while the other two concentrations are kept constant. The two reactions that can occur in exclusive binding of L and U to the probe P are

$$\begin{aligned} P + L &\underset{k_-^L}{\overset{k_+^L}{\rightleftharpoons}} PL \quad \text{and} \\ P + U &\underset{k_-^U}{\overset{k_+^U}{\rightleftharpoons}} PU, \end{aligned} \tag{4.11}$$

which corresponds to the rate equations:

$$\frac{\mathrm{d}c_{PL}(t)}{\mathrm{d}t} = k_+^L c_L(t)\, c_P(t) - k_-^L c_{PL}(t) \quad \text{and} \tag{4.12}$$

$$\frac{\mathrm{d}c_{PU}(t)}{\mathrm{d}t} = k_+^U c_U(t)\, c_P(t) - k_-^U c_{PU}(t). \tag{4.13}$$

Writing the equations in equilibrium, i.e. considering $\dot{c}(t) = 0$, and substituting $c_L^{\mathrm{eq}} = \left(c_L^{\mathrm{ini}} - c_{PL}^{\mathrm{eq}}\right)$, $c_U^{\mathrm{eq}} = \left(c_U^{\mathrm{ini}} - c_{PU}^{\mathrm{eq}}\right)$ as well as $c_P^{\mathrm{eq}} = \left(c_P^{\mathrm{ini}} - c_{PL}^{eq} - c_{PU}^{\mathrm{eq}}\right)$ leads to two coupled nonlinear equations given by

$$K_L\left(c_L^{\mathrm{ini}} - c_{PL}^{\mathrm{eq}}\right)\left(c_P^{\mathrm{ini}} - c_{PL}^{\mathrm{eq}} - c_{PU}^{\mathrm{eq}}\right) - c_{PL}^{\mathrm{eq}} = 0 \quad \text{and} \tag{4.14}$$

$$K_U\left(c_U^{\mathrm{ini}} - c_{PU}^{\mathrm{eq}}\right)\left(c_P^{\mathrm{ini}} - c_{PL}^{\mathrm{eq}} - c_{PU}^{\mathrm{eq}}\right) - c_{PU}^{\mathrm{eq}} = 0. \tag{4.15}$$

The binding affinities are defined by $K^L = k_+^L/k_-^L$ and $K^U = k_+^L/k_-^U$. In the scope of our work these equations were solved numerically yielding c_{PL}^{eq} and c_{PU}^{eq} as a function of c_P^{ini}, c_L^{ini}, and c_U^{ini} as well as K_L and K_U. Note that the solution can also be analytically found by converting the system into a single cubic function, whose roots can be found using Cardano's formula [156]. However, as it is long and offers no valuable insight we omit to give it here. We now rewrite the anisotropy equation Eq. (2.20) for these reactions in the following form:

$$\begin{aligned} r(c_U^{\mathrm{ini}}) &= q_L r_L + q_{PL} r_{PL} \\ &= \frac{c_L^{\mathrm{ini}} - c_{PL}^{\mathrm{eq}}}{c_L^{\mathrm{ini}}} r_L + \frac{c_{PL}^{\mathrm{eq}}}{c_L^{\mathrm{ini}}} r_{PL}. \end{aligned} \tag{4.16}$$

Here, r_L and r_{PL} are the anisotropies of the labeled single strands and the duplexes P-L, respectively. We use Eq. (4.16) to fit the anisotropy data in the exclusive binding experiments. While all initial concentrations are known, K_L and K_U as well as r_L and r_{PL} are free parameters.

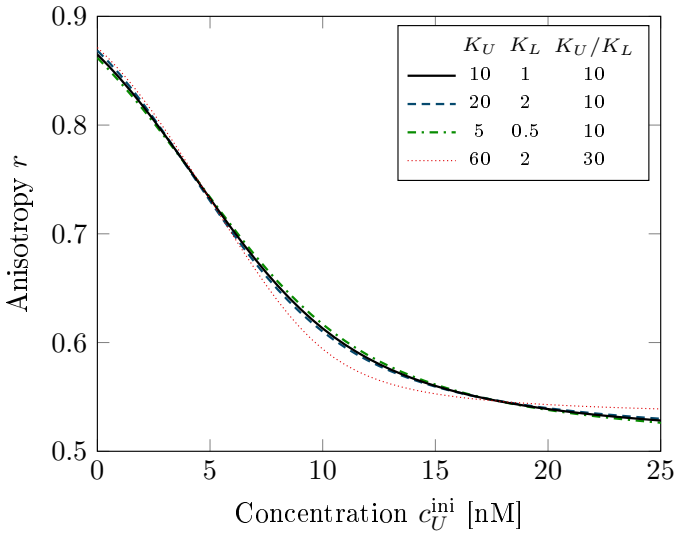

Figure 4.2. Fluorescence anisotropy r as a function of unlabeled strand concentration c_U^{ini} using the exclusive binding model. The black solid reference curve represents the theoretical curve for $K_U = 10 \times 10^9\,\text{M}^{-1}$, $K_L = 1 \times 10^9\,\text{M}^{-1}$, $r_L = 0.5$ and $r_{PL} = 1$. The blue dashed, green dash-dotted, and red dotted curves correspond to the different combinations of binding constants (c.f. inset), that are fitted to the reference curve (r_L and r_{PL} are free parameters). The curves with the ratio $\kappa = K_U/K_L = 10$ are hardly distinguishable from the reference curve but the red dotted curve with $\kappa = 30$ differs. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

We find that fitting K_L and K_U simultaneously leads to a large error and therefore we cannot obtain the absolute values of both K_L and K_U using Eq. (4.16). It turns out that the qualitative shape of the model function is mostly determined by the ratio of the binding constants κ. To clarify this we plot the anisotropy r as a function of c_U^{ini} using Eq. (4.16) for the exemplary binding constants $K_U = 10 \times 10^9\,\text{M}^{-1}$ and $K_L = 1 \times 10^9\,\text{M}^{-1}$ (i.e. $\kappa = 10$) and the anisotropies $r_L = 0.5$ and $r_{PL} = 1$. The black solid curve ('reference' curve) in Fig. 4.2 corresponds to the theoretically generated curve. For what follows, we treat the reference curve as a hypothetical set of experimental data, which we attempt to fit using the exclusive binding model. For the fit, we fix all binding constants and only treat the anisotropies as free parameters. The blue dashed and green dash-dotted curve correspond to combinations of binding constants that yields $\kappa = 10$ as well. It can be seen that the fits convincingly describe the reference curve. The red dotted curve, on the other hand exhibits a larger $\kappa = 30$ and the corresponding fit clearly differs from the reference. Therefore, in the following to fit the data with exclusive binding model, we take the value of one of the binding constant from the individual

binding experiments as a fixed input parameter and only fit the ratio κ. The anisotropies of the single strand (r_L) and molecular complex (r_{PL}) are free parameters. In the following For all the experiments performed with FA, we show the average anisotropies accompanied with the corresponding 95% error bars as we described in Sec. 3.2.4.

4.2 DNA Strands Under Consideration

Table 4.1. Set 1 includes probe P, strand 1 (S_1), strand 2 (S_2). Strands S_1 and S_2 bind to distinct segments of P. The modification strands S_{2-2b}, S_{2-3b} and S_{2-4b} of strand S_2 have additional 2, 3 and 4 overlapping bases, respectively, which overlap with S_1 (marked in red).

Name	Sequence
P	5'-CGTACAAGCTACTGACCTACTTA-3'
S_1	5'-TAGCTTGTACG-3'
S_2	5'-TAAGTAGGTCAG-3'
S_{2-2b}	5'-TAAGTAGGTCAGTA-3'
S_{2-3b}	5'-TAAGTAGGTCAGTAG-3'
S_{2-4b}	5'-TAAGTAGGTCAGTAGC-3'

Table 4.2. Set 2 consists of a probe P, perfectly matching target PM. Sequences MME and MMM are modifications of PM with a single mismatch placed towards the end or in the center of the strand, respectively (marked in red).

Name	Sequence
P	5'-TTACGATCTGATCCTT-3'
PM	5'-AAGGATCAGATCGTAA-3'
MME	5'-AAGGATCAGATCGCAA-3'
MMM	5'-AAGGATCACATCGTAA-3'

In order to study the binding of the DNA strands in bulk we consider two sets of sequences, set 1 and set 2, that are listed in Tab. 4.1 and Tab. 4.2. We did not find any stable secondary structures by checking the designed sequences with the NUPACK web server. Set 1 includes the probe P and the targets S_1 and S_2 that bind to the segment of the probe. Targets S_{2-2b}, S_{2-3b} and S_{2-4b}, on the other hand, are the modification of a S_2 with 2, 3 and 4 additional bases which overlap with S_1 on the probe. In set 2, the targets PM, MME and MMM have the same length as a probe P. While PM perfectly matches to a probe, MME and MMM have a single mismatch towards the end or in the center of the strand, respectively. At the end of Sec. 4.4.3 we show

the results after modifying the sequences in set 2. We perform all experiments below the corresponding melting temperatures. More specifically, we chose room temperature 23 °C (RT) for set 1 and 44 °C for set 2.

Some remarks on our notation: we indicate the hybridization between strands by use of a hyphen, e.g. P-S_1 signifies a binding of the strand 1 to the probe. Moreover, all concentrations are written as c_x, where x indicates the sequence name (e.g. c_{PM} refers to the concentration of PM). All sequences in set 1 are denoted by c_x for the concentration of S_x (e.g. c_1 is the concentration of S_1).

4.3 DNA Strands Binding Assay on Set 1

In this section we investigate the sequences from set 1 as given in Tab. 4.1, where target strands S_1, S_2, S_{2-3b} and S_{2-4b} are shorter than probe P, but fully complementary to it (see Sec. 4.2). First we present our results from anisotropy for the pairwise binding of a target to the probe and later we demonstrate the situation where apart from the probe, two targets are at the same time present in the solution.

4.3.1 Pairwise Binding

To determine the binding constant for individual binding of a target to the probe, all targets here are labeled with Atto 532, which is attached to the 5' end of S_2, S_{2-3b} and S_{2-4b} but to the 3'end of S_1. To determine the binding affinity of the pairwise binding of two DNA strands with fluorescence anisotropy the concentrations of the labeled strand, here the targets, are kept constant while the concentration of the unlabeled probe is changed. The initial concentrations of strands S_1 and S_2 are $c_1^{\text{ini}} = c_2^{\text{ini}} = 10\,\text{nM}$ and for strands S_{2-3b} and S_{2-4b} it is $c_{2-3b}^{\text{ini}} = c_{2-4b}^{\text{ini}} = 5\,\text{nM}$. The parallel and perpendicular components of the emission light $I_\parallel$ and $I_\perp$ for different concentrations of the probe c_P^{ini} are determined. The anisotropy is then calculated using Eq. (2.18). With increasing probe concentration the probe-target duplex concentration increases, which causes a rise in the measured anisotropy until saturation is reached. Fig. 4.3 shows the anisotropy as a function of c_P for P-S_1 (data: circles), P-S_2 (data: diamonds), P-S_{2-3b} (data: triangles) and P-S_{2-4b} (data: squares) hybridizations. In order to obtain the binding constants for these reactions the anisotropy data are fitted using Langmuir isotherm Eq. (2.10). The binding constant K and the anisotropies of the labeled single strand r_L and double helix r_{PL} are free parameters. The fit leads to the individual binding affinities $K_1 = (8.1 \pm 0.1) \times 10^8\,\text{M}^{-1}$ for P-S_1 (red solid curve), $K_2 = (4.8 \pm 1.8) \times 10^9\,\text{M}^{-1}$ for P-S_2 (blue dashed curve), $K_{2-3b} = (1.1 \pm 0.3) \times 10^{10}\,\text{M}^{-1}$ for P-S_{2-3b} (green dotted curve) and $K_{2-4b} = (1.9 \pm 1.5) \times 10^{10}\,\text{M}^{-1}$ for P-S_{2-4b} (violet dash-dotted curve). Since

S_{2-3b} and S_{2-4b} are longer than S_1 and S_2, the corresponding binding constants turn out to be larger. The relative error originates from the different distributions of free to bound strands as it is predicted by the Langmuir model Eq. (2.10). For larger values of binding constants K the number of free strands diminishes leading to a decreased sensitivity and increased relative error. All obtained values are summarized in Tab. 4.3.

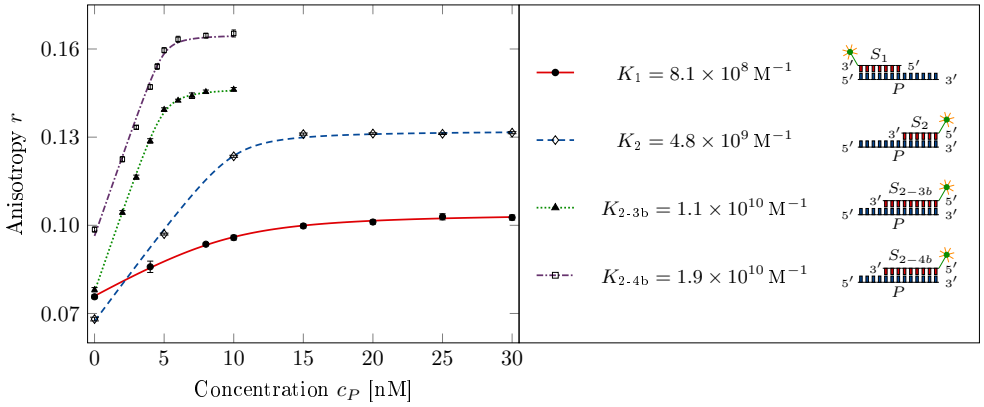

Figure 4.3. Fluorescence anisotropy as a function of probe concentration c_p for P-S_1 (data: circles, fit: red solid curve), P-S_2 (data: diamonds, fit: blue dashed curve), P-S_{2-3b} (data: triangles, fit: green dotted curve) and P-S_{2-4b} (data: squares, fit: violet dash-dotted curve) hybridizations. The data are fitted using the Langmuir isotherm. The dotted and dash-dotted curves are shifted vertically by 0.02 and 0.04, respectively for better legibility. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

Table 4.3. The individual binding constants as obtained by FA for P-S_1, P-S_2, P-S_{2-3b} and P-S_{2-4b} at 23 °C. Units are $10^9\,\mathrm{M}^{-1}$.

Sequence	K $[10^9\,\mathrm{M}^{-1}]$
P-S_1	0.81 ± 0.1
P-S_2	4.8 ± 1.8
P-S_{2-3b}	11 ± 3
P-S_{2-4b}	19 ± 15

4.3.2 Binding of Two Partially Overlapping Strands to a Common Probe

In this part we perform binding experiments with probe and target S_1 as well as the second target either S_2, S_{2-3b}, or S_{2-4b}. While S_{2-3b} and S_{2-4b} share an overlapping tail with the length of 3 and 4 with S_1 on the probe, S_2 and S_1 do

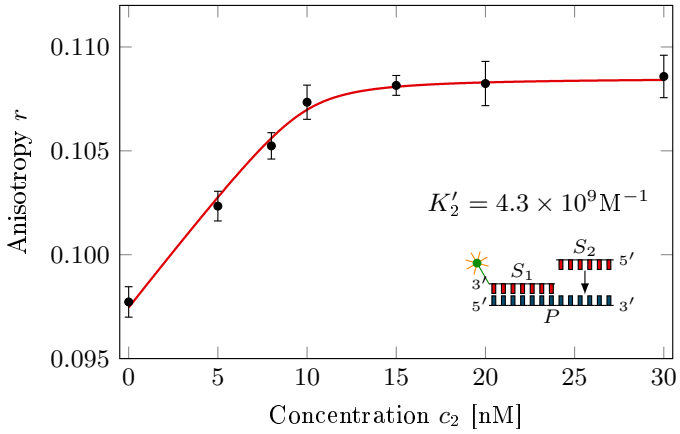

Figure 4.4. Fluorescence anisotropy as a function of c_2 for P-S_1-S_2 binding (data: circles). The rise in an anisotropy shows that target strands S_1 and S_2 simultaneously bind to the probe P, slowing down the rotation of the attached fluorophore. The red solid curve presents the fit against the data using the Langmuir model, for which it was assumed that the presence of S_1 does not influence the binding of S_2 to the probe. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

not overlap. In these experiments P and targets S_2, S_{2-3b} and S_{2-4b} are unlabeled while S_1 is labeled with Atto 532 at 3' end. The initial concentrations of P and S_1 are kept constant at $c_P^{\mathrm{ini}} = c_1^{\mathrm{ini}} = 10\,\mathrm{nM}$. The concentrations of the second targets c_2^{ini}, c_{2-3b}^{ini} and c_{2-4b}^{ini} are varied. Fig. 4.4 shows the change in anisotropy as a function of c_2. The anisotropy increases until saturation is reached. The rise in the anisotropy suggests that S_2 attaches to the existing complex P-S_1, forming a triplet P-S_1-S_2, which slows down the rotation of the attached fluorophore. Note that indeed only the triplet conformation is able to cause the rise in an anisotropy, as strand S_1 carries the fluorescent label and formation of a S_1-S_2 duplex can be ruled out. Since S_1 and S_2 do not overlap on the probe, we assume the binding of S_2 to the P is not influenced by the presence of S_1. Therefore we use simultaneous binding model Eq. (4.10) to describe the experimental data. For the corresponding fit, we keep the binding affinity between P and S_2 (K_2')[1] as well as the anisotropies a and b as free parameters (see Eq. (4.10)). The fit is independent of K_1. The result is plotted as a red curve in Fig. 4.4. We obtain $K_2' = (4.3 \pm 3) \times 10^9\,\mathrm{M}^{-1}$, which compares well to the corresponding value determined by individual binding $K_2 = 4.8 \times 10^9\,\mathrm{M}^{-1}$ (cf. Tab. 4.3). As it is clear in Fig. 4.4, the fit excellently agrees with the data, revealing that the presence of a competitor does not affect K_2. We therefore conclude that any interaction between S_1 and S_2 is

[1] Note that we assign a prime to the binding constant K_2 to distinguish it from the one determined for the individual binding in the previous section

negligible.

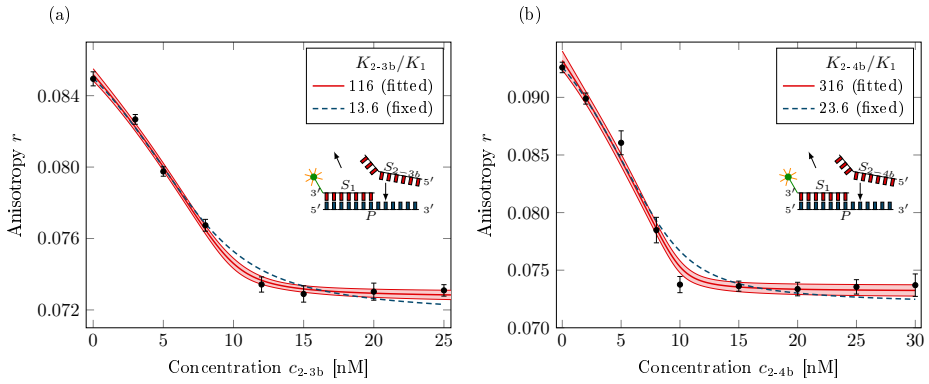

Figure 4.5. Fluorescence anisotropy decays as a function of c_{2-3b} and c_{2-4b} for reactions including P, S_1 and S_{2-3b} (a) and P, S_1 and S_{2-4b} (b) (data: circles). The results are fitted using the exclusive binding model. The blue dashed curves correspond to a fit, where the previously determined binding constants K_1 and K_{2-3b} in (a) and K_1 and K_{2-4b} in (b) are fixed. The red solid curves represent a fit using the ratio of the binding constants κ as a free parameter, keeping only the larger binding constants K_{2-3b} in (a), K_{2-4b} in (b) fixed. The shaded regions correspond to the 95 % confidence band of the fit. The fitted ratios κ^{fit} (red curve) remarkably deviate from the fixed ratios κ^{fix} (blue dashed durve). In all fits the anisotropies of the single strands and molecular complexes are free parameters. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

Next, we perform the same experiment, but with targets S_{2-3b} and S_{2-4b} instead of S_2. The results shown in Fig. 4.5 (a) and (b), respectively, reveal that the anisotropy now drops. The drop implies that the second targets S_{2-3b} and S_{2-4b} compete with S_1 and remove it from the probe due to the overlapping strand. The drop in (b) is faster than in (a) reflecting the larger binding affinity of S_{2-4b} to the probe compared to S_{2-3b}. Assuming there is no interaction between the targets on the probe, we first fit the data using the exclusive binding model Eq. (4.16), taking both binding constants K_1 and K_{2-3b} in (a), K_1 and K_{2-4b} in (b) from the individual binding measurements (Tab. 4.3). The anisotropies of the single strands and the molecular complexes are free parameters (cf. Sec. 4.1.2). The blue dashed curves represent the corresponding fits. While the fit describes the data well for the initial drop at small concentrations $< 10\,\text{nM}$, the agreement for the saturation above 10 nM is not convincing. According to our discussion given at the end of Sec. 4.1, the qualitative shape of the curve in this range is mostly determined by the ratio κ of the binding constants. In the following we treat $\kappa_{2-3b} = K_{2-3b}/K_1$ in (a) and $\kappa_{2-4b} = K_{2-4b}/K_1$ in (b) as additional fitting parameters, keeping only K_{2-3b} and K_{2-4b} from Tab. 4.3 constant. This pro-

cedure assumes that in fact the interaction among the competing targets may change the binding constant K_1 of the weaker competitor - an effect which is known as cooperative binding [157, 158]. The red solid curves represent the corresponding fits which clearly show a better agreement with the data than the blue dashed curves. The improvement can be quantified with the mean square error, which is now reduced by a factor of 3 in (a) and 2 in (b) compared to the blue dashed curve. Moreover, most data are contained within the 95% confidence bands of the fits (red shaded region), suggesting that cooperativity might indeed be a suitable assumption describing the observed drop. We obtain $\kappa^{\text{fit}}_{2-3b} = 116^{+40}_{-29}$ and $\kappa^{\text{fit}}_{2-4b} = 316^{+160}_{-102}$. This means that $\kappa^{\text{fit}}_{2-3b}$ and $\kappa^{\text{fit}}_{2-4b}$ increase around an order of magnitude compared to the fixed ratios $\kappa^{\text{fix}}_{2-3b} = 13.6 \pm 5.4$ and $\kappa^{\text{fix}}_{2-4b} = 23.6 \pm 21.6$, respectively. The error of κ^{fix} is obtained using the linear error propagation of the individual binding constants. Note that for large κ the model becomes increasingly nonlinear in κ, leading to the asymmetric error. The obtained κ^{fit} leads to $K_1 = 0.95 \times 10^8\,\text{M}^{-1}$ in (a) and $K_1 = 0.6 \times 10^8\,\text{M}^{-1}$ in (b). Comparing these values with the value determined by individual measurement $K_1 = 8.1 \times 10^8\,\text{M}^{-1}$ shows that the presence of the second target weakens the binding of S_1 to the probe. Note that since the model is mostly sensitive to the ratio of the binding constants it is not unambiguously possible to individually determine the values of both binding constants in presence of the competitor. In the other words, specifying how much each binding constant changes due to the competition is not possible. However, it is clear that the ratio of the binding constants in competition situation drastically changes compared to the ratio from the individual (pairwise) binding.

Fig. 4.6 shows the same data with the only difference that now the smaller binding constant K_1 is fixed in the respective fits. We obtain $\kappa^{\text{fit}}_{2-3b} = 85^{+28}_{-18}$ in (a) and $\kappa^{\text{fit}}_{2-4b} = 7140^{+16783}_{-4250}$ in (b). Although these fits seem to be in excellent agreement with the data, they lead to the meaninglessly huge binding constants $K_{2-3b} = 6.9 \times 10^{10}\,\text{M}^{-1}$ and $K_{2-4b} = 5.8 \times 10^{12}\,\text{M}^{-1}$. Since we expect the binding constants in presence of a competitor to decrease, this scenario is unlikely to happen. As mentioned before, the sensitivity of the exclusive binding model decreases with increasing κ. More specifically, only the curvature in the region of saturation becomes sharper for large kappa, while the rest of the curve remains unaffected. The resulting κ in (b) is therefore mostly determined by the single data point at 10 nM. In particular the uncertainty of this data point has a severe impact on κ, which probably explains the very large value. Accordingly, it should be taken with a grain of salt.

Eventually, we study the binding of S_1 and S_{2-2b} to the probe, which turns out to be an intermediate situation between simultaneous and exclusive binding. Fig. 4.7 shows the anisotropy as a function of c_{2-2b} keeping the concentrations of the labeled strand S_1 and the unlabeled probe constant at

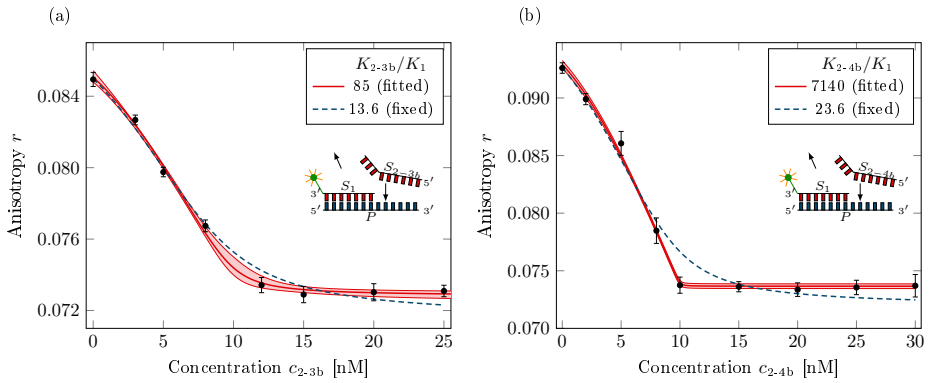

Figure 4.6. Fluorescence anisotropy as a function of c_{2-3b} (a) and c_{2-4b} (b). The data are fitted similar to Fig. 4.5 but here the red solid curves in both cases correspond to the fit, where the smaller binding constant K_1 is fixed instead. The fits, however, lead to an increase in K_{2-3b} (a) and K_{2-4b} (b), which is physically unlikely to happen. Additionally, the extreme large value of κ^{fit} in (b) stems from the model being sensitive to the uncertainty of the data point at 10 nM. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

$c_1^{\text{ini}} = c_P^{\text{ini}} = 10\,\text{nM}$. As it is shown the anisotropy data are scattered around a constant value (gray dashed line). This is interesting, as it is not predicted by any of the two models we introduced earlier. To investigate this in more detail, we consider limiting cases for both simultaneous and exclusive binding models, which are shown as blue and red curves, respectively. To apply both models, we take the predetermined K_1. We moreover require K_{2-2b}, which we estimate via $K_{2-2b} = K_2$ from the individual binding experiments. This value should be a lower bound for K_{2-2b}, as the two additional bases are expected to increase it. Higher values for K_{2-2b} would lead to a curve within the blue shaded region for simultaneous binding and within the red shaded region for exclusive binding. Furthermore, we have seen that cooperativity might reduce K_1. A smaller K_1, however, results in curves within the shaded regions just as well. Since the measured data are between both bounds we suspect that the real kinetics may well include triplet formation (simultaneous binding of both strands to the probe) as well as exclusive binding of S_1 and S_{2-2b} with both effects compensating each other leading to no net change in anisotropy. However, to confirm this conclusion beyond a doubt, it is necessary to perform additional experiments with different strands to detect the same behavior.

To test whether the entropic effects [159] are responsible for the removal of the competing target observed in Fig. 4.5, we modified strand S_1 and S_2 as well as the probe P as summarized in Tab. 4.4: all strands were prolonged by three extra 'T' bases (marked in red), while the probe has an additional 'CTA' bases (marked in green color) in such that both targets have the same

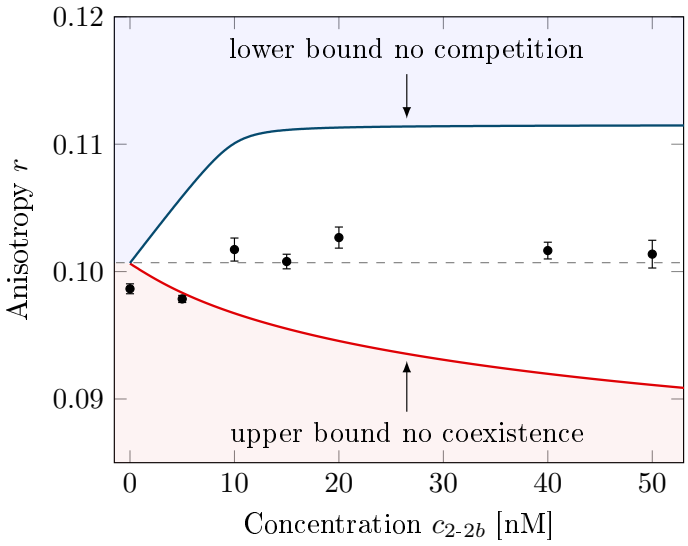

Figure 4.7. Fluorescence anisotropy as a function of c_{2-2b} for the hybridization of S_1 and S_{2-2b} to the probe. The scattered circles are the experimental anisotropy data. The blue curve is a lower bound assuming two strands simultaneously bind to the probe. The red curve is an upper bound if the targets exclusively bind to the probe. The shaded areas indicate where the boundary curves would be for larger K_{2-2b} or smaller K_1.

individual binding constants as shown in Fig. 4.3 but they overlap on a probe via a non-specific overhang 'TTT'. Fig. 4.8 illustrates that the anisotropy as a function of c_1' (a) and c_2' (b) does not decay as it was presented for the specific overhang (see Fig. 4.5). On the other hand, the absolute change of the anisotropy is untypically small, which might give further indication that the present binding situation is too complex to be described by simultaneous or exclusive binding model. We conclude that the entropic repulsion of three non-specific bases overhang is too weak to cause exclusive binding as observed with specifically binding overhangs.

Table 4.4. Set 3: modified sequences from set 1 by adding extra 'TTT' bases to all (marked in red) and a 'CTA' bases to the probe (marked in green).

Name	Sequence
S_1'	5'-TTTTAGCTTGTACG-3'
S_{2-3b}'	5'-TAAGTAGGTCAGTAGTTT-3'
P'	5'-CGTACAAGCTATTTCTACTGACCTACTTA-3'

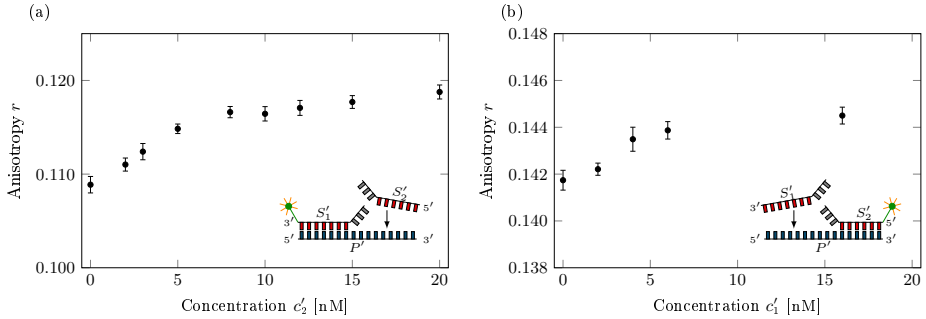

Figure 4.8. Fluorescence anisotropy as a function of c_1' and c_2' for target strands S_1' and S_{2-3b}' binding to part of the probe. Two targets share an unspecific overlap of 'TTT'. The anisotropy in contrast to Fig. 4.5 does not drop, confirming that the non-specific overhang is too weak to create an exclusive binding for each target strands. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

4.4 DNA Strand Binding Assays on Set 2

In this section we study the hybridization of the strands from set 2 as given in Tab. 4.2. All target strands PM, MME and MMM have the same length as the probe P (see Sec. 4.2). In particular, the binding of MMM to the probe is weak because the mismatch at the middle of MMM destabilizes a helix. As a consequence the anisotropy change in the studied range of MMM concentration is quite small. To obtain a detectable change, we prolong MMM strand by adding 100 thymine bases to its 3' end. To make our results comparable, we similarly prolonged MME and PM. In the following we determine the binding constants for the individual binding P-PM, P-MMME and P-MMM with two techniques, fluorescence anisotropy (FA) and fluorescence correlation spectroscopy (FCS). The binding constants are named as K^A and K^{CS}, where A and CS refer to the fluorescence anisotropy and correlation spectroscopy, respectively. The last section addresses the competitive binding situation where two targets are simultaneously present in solution.

4.4.1 Pairwise Binding Using Fluorescence Anisotropy

For this experiment the probe is labeled with Cy3 at 3' end. All targets PM, MMM, MME remain unlabeled. The initial concentration of the probe is $c_P^{\text{ini}} = 10\,\text{nM}$ and the target concentrations c_{PM}, c_{MME} and c_{MMM} are varied. The result is displayed in Fig. 4.9. The red solid, blue dashed, and green dotted curves correspond to fits that yield $K^A_{\text{PM}} = (3.0 \pm 0.85) \times 10^9\,\text{M}^{-1}$, $K^A_{\text{MME}} = (4.3 \pm 0.7) \times 10^8\,\text{M}^{-1}$ and $K^A_{\text{MMM}} = (1.9 \pm 2) \times 10^6\,\text{M}^{-1}$, respectively. As expected the mismatch in the center of MMM destabilizes the duplex more than the mismatch at the end of MME. Therefore, the binding

constant for P-MMM is smaller than for P-MME [89, 92]. As K^A_{MMM} is very small, the corresponding curve does not reach saturation within the studied concentration range, which in turn leads to the relatively large error of the binding constant.

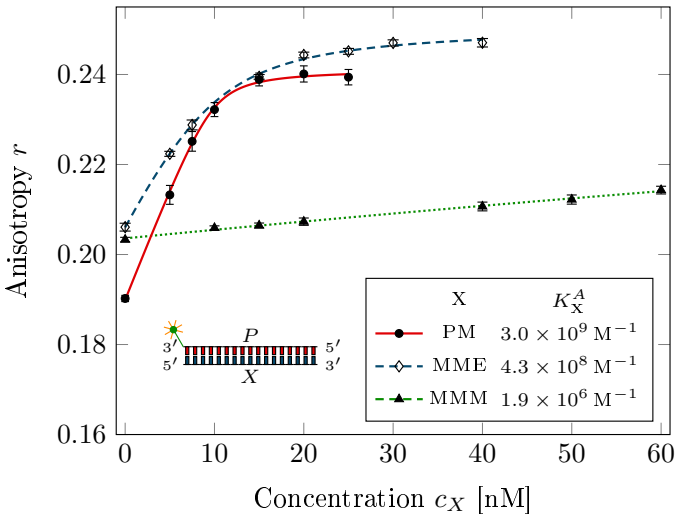

Figure 4.9. Fluorescence anisotropy as a function of target concentrations c_{PM}, c_{MME} and c_{MMM} for P-PM (data: circles), P-MME (data: diamonds) and P-MMM (data: triangles). The red solid, blue dashed and green dotted curves represent the fits using the Langmuir model. Since P-MMM binding is relatively weak, the anisotropy here increases slowly in the investigated concentration range. However, for P-PM and P-MME the anisotropy rises until saturation is reached. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

4.4.2 Pairwise Binding Using Fluorescence Correlation Spectroscopy

As we discussed in Sec. 2.7 the binding constant can be additionally determined using fluorescence correlation spectroscopy. Fig. 4.10 illustrates the measured fluorescence autocorrelation functions for the single strand (circles, $q = 0$), bound double helix (diamonds, $q = 1$) in saturation as well as an exemplary intermediate measurement corresponding to a mixture of both conformations (triangles) for (a) P-PM, (b) P-MME and (c) P-MMM hybridizations. The initial concentration of the Cy3 labeled probe is 5 nM. We first fit the data for the single and double strands using Eq. (2.26). The corresponding fits are shown as red solid and blue dashed curves and lead to the characteristic diffusion times τ_{D_1} and τ_{D_2}, respectively. As can be clearly seen the diffusion times for unbound and bound states are perfectly discrimi-

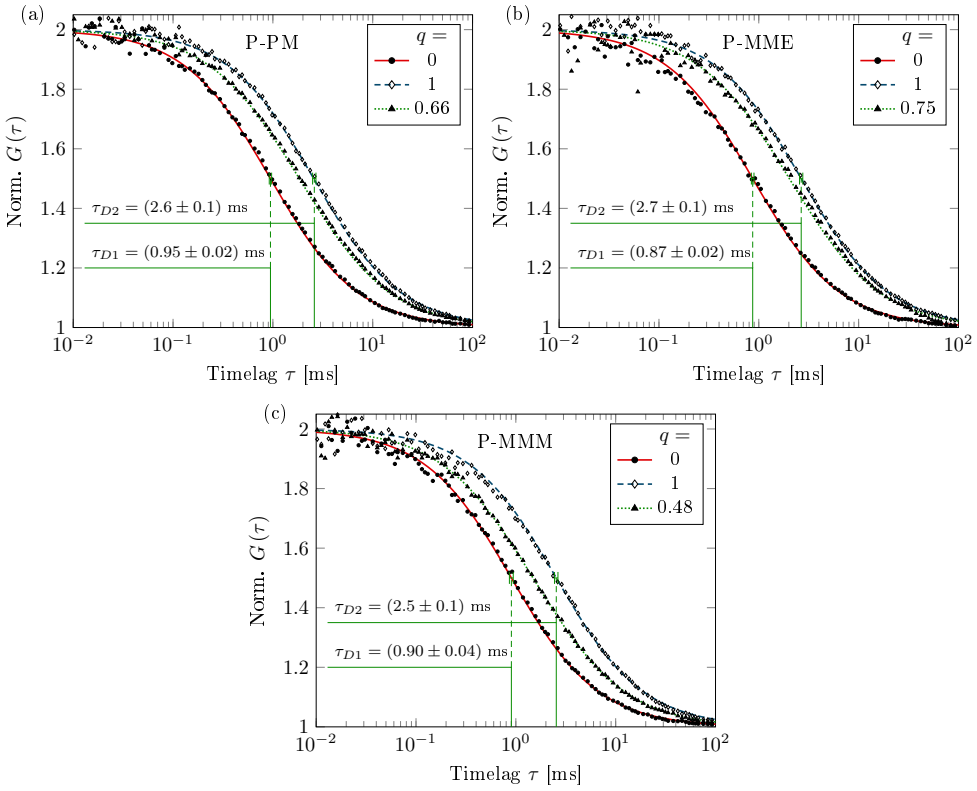

Figure 4.10. Measured fluorescence autocorrelation functions for single labeled stranded P with $q = 0$ (data: circles), and double stranded with $q = 1$ (data: diamonds) for (a) P-PM, (b) P-MME and (c) P-MMM hybridizations. The red solid and blue dashed curves represent the fits using Eq. (2.26). The triangles show the autocorrelation for the mixture of both bound and unbound conformations. The green dotted curves are the corresponding fits using Eq. (2.27). We obtain the molar fractions q from the fits. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

nated. To determine the molar fraction q of the double strands in the mixture of both states, for different target concentrations, we fit the corresponding data to Eq. (2.27), taking the pre-determined τ_{D_1} and τ_{D_2}. The green dotted curves represent the fits for P-PM at $c_{\mathrm{PM}} = 4\,\mathrm{nM}$ as well as P-MME and P-MMM at $c_{\mathrm{MME}} = c_{\mathrm{MMM}} = 5\,\mathrm{nM}$. We obtain $q = 0.66$, $q = 0.75$ and $q = 0.48$, respectively. The correlation measurements for other target concentrations are give in the Appx. A. Fig. 4.11 shows all obtained molar fractions q of the duplex conformation as a function of target concentrations for P-PM, P-MME and P-MMM. Fitting the data using the Langmuir model yields $K^{CS}_{\mathrm{PM}} = (4.6 \pm 2.4) \times 10^9\,\mathrm{M}^{-1}$ (red solid curve), $K^{CS}_{\mathrm{MME}} = (1.5 \pm 0.5) \times 10^9\,\mathrm{M}^{-1}$ (blue dashed curve) and $K^{CS}_{\mathrm{MMM}} = (3.6 \pm 0.8) \times 10^8\,\mathrm{M}^{-1}$ (green dotted curve).

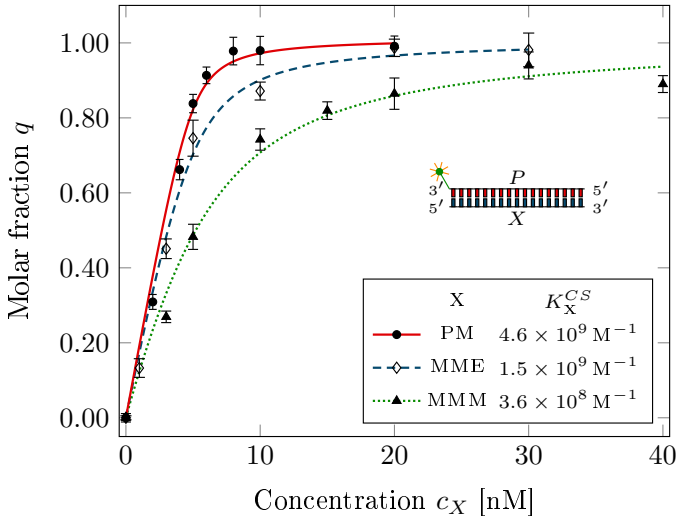

Figure 4.11. The fraction of double strands (q) for P-PM (data: circles), P-MME (data: diamonds) and P-MMM (data: triangles) as a function of c_{PM}, c_{MME} and c_{MMM}. The red solid, blue dashed and green dotted curves represent the fits using the Langmuir model. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

Table 4.5. Comparision of the experimental binding constants as obtained by FCS and FA with the theoretical predictions by NUPACK for P-PM, P-MME and P-MMM. Units are $10^9\,\mathrm{M}^{-1}$.

	bindings constant obtained from		
Sequence	FCS	FA	NUPACK
PM	4.6 ± 2.4	3.0 ± 0.85	3.1
MME	1.5 ± 0.5	0.43 ± 0.065	0.35
MMM	0.36 ± 0.08	0.0019 ± 0.002	0.0023

Tab. 4.5 presents the binding constants obtained from FA and FCS and compares them to predictions obtained from NUPACK. We find that the values from NUPACK match the experimental values from FA while there is a clear difference between them and FCS values, in particular for MMM. Discrepancies between experimental and theoretical values are well known and have been discussed for instance in [23, 24]. In contrast, we here observe a disagreement not only between theoretical and experimental values, but even between two distinct experimental techniques. In particular, we emphasize that the difference significantly exceeds the corresponding confidence intervals. Therefore, we suggest that the discrepancy mainly stems from differences in the underlying principles of both methods. In the following we will elaborate on this point. FA is based on the rotation of the attached fluorophore. A stiff double

helix with a high persistence length [160] entails a slow rotational motion and thereby causes a large anisotropy signal (Fig. 4.12 (a)) compared to the mobile labeled single strand (Fig. 4.12 (b)), While we can assume that the P-PM duplex stiffly binds, the central mismatch of MMM destabilizes the corresponding duplex, diminishes the persistence length. In particular, we find contributions of complexes, in which the labeled end is unbound as illustrated in Fig 3.13 (c). The rotational mobility of a fluorophore in this situation increases compared to a stiff duplex leading only to a small increase in anisotropy. The Nupack algorithm based on the nearest neighbor model [23, 35, 36, 161, 162] relies only on base stacking interactions which can only be present in helix configurations. Therefore, the obtained values from FA agree well with those from Nupack. In contrast to FA, the FCS technique is not sensitive to the difference between partially and stiffly bound microstates as both diffuse as one complex in a similar fashion. Accordingly, partially bound conformations appear as fully bound and FCS leads to the larger binding constant for P-MMM compared to the value obtained by FA. Considering P-MME, however, fluorophore and mismatch are positioned at opposite ends of the double helix (Fig. 4.12 (d)). Therefore, the probability that the mismatch destabilizes the helix at the fluorophore side is comparatively smaller and the rotation of the fluorophore is governed by the mobility of the double helix. Therefore, FA and FCS yield similar values in this case.

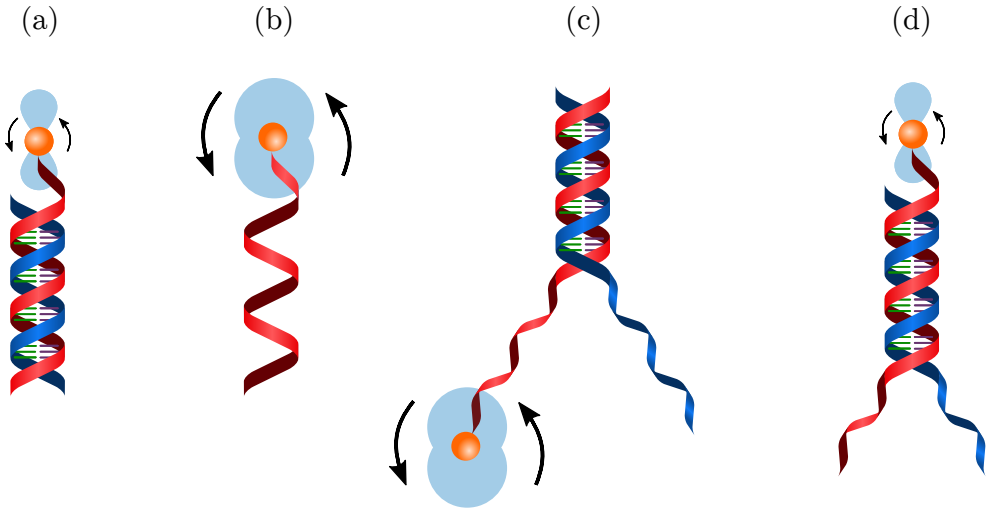

Figure 4.12. (a) The molar mass of a stiff DNA double helix significantly slows down the rotation of the attached fluorophore. (b) For a single strand the attached fluorophore rotates fast because of the small molar mass. (c) The mismatch at the middle of the strand (MMM) might open the helix at the fluorophore side allowing it for a fast rotation. (d) The fluorophore is attached to the opposite end of the strand (MME). Therefore, the rotation of the fluorophore is not much influenced by the opening of the helix.

Our results suggest that other microstates beside the stiff double helix exist and they have different impacts on FCS and FA. To gain further insight on the impact of these so-called intermadiate states [163, 164], we estimate their contribution to the overall binding energy in the following. The interaction between the unbound segments of these strands are caused by "London"- or "dispersion"-forces. The corresponding van-der-Waals energy w between two parallel chain molecules at distance d is given by [165]

$$w(d) = -\frac{3\pi CL}{8\sigma^2 d^5}. \tag{4.17}$$

Here, L is a length of the strand, σ a monomer size and C is the van-der-Waals coefficient. Neglecting the zero-frequency contribution, C can be expressed as [165]

$$C = \frac{\sqrt{3}h\nu_e}{4} \frac{\left(n_1^2 - n_2^2\right)^2}{\left(n_1^2 + 2n_2^2\right)^{3/2}} \times r^6, \tag{4.18}$$

where $h\nu_e$ is the ionization energy and r is a radius of a DNA base, n_1 and n_2 are the refractive index of the molecules and the medium, respectively. We use Eq. (4.17) and (4.18) to calculate the energy for two parallel 16 bp DNA strands ($L = 16\sigma$) with the base radius of $r = 0.5\,\text{nm}$ and assuming that the bases are touching each other, which corresponds to $d = 1\,\text{nm}$. Using refractive index $n_1 = 1.6$ [166] and $n_2 = 1.33$, taking the ionization energy at $\lambda = 350\,\text{nm}$ and $\sigma = 0.5\,\text{nm}$ [167], we obtain the enthalpy of attraction of roughly $w = -1.1\,\text{kcal/mol}$. Neglecting the loss in entropy, we have $\Delta H = \Delta G_w = -1.1\,\text{kcal/mol}$.

On the other hand, using the determined individual binding constants for P-MMM we obtain the Gibbs free energy of $\Delta G^A = -11.63\,\text{kcal/mol}$ and $\Delta G^{CS} = -14.94\,\text{kcal/mol}$. This corresponds to the energy difference of $\Delta\Delta G = -3.3\,\text{kcal/mol}$ for the P-MMM binding in FCS and FA, which is in fair agreement with the estimated van-der-Waals energy of $\Delta G_w = -1.1\,\text{kcal/mol}$. Since we neglected the π-π interaction in our calculation, the unspecific van-der-Waals energy presents only the lower bound for molecular attraction beyond double helix formation [4].

4.4.3 Binding of Two Fully Overlapping Strands to a Common Probe

We consider PM and MMM in competition to bind to the same complementary probe P. Here the PM strand is labeled by Cy3 at 3' end. To make the changes in anisotropy more visible, we add 100 'T' bases to the 3' end of the probe. While the initial concentrations of the probe and the labeled PM are $c_\text{P} = c_\text{PM} = 10\,\text{nM}$, c_MMM is varied. Fig. 4.13 presents the drop in anisotropy by increasing c_MMM, which indicates that the fraction of P-PM

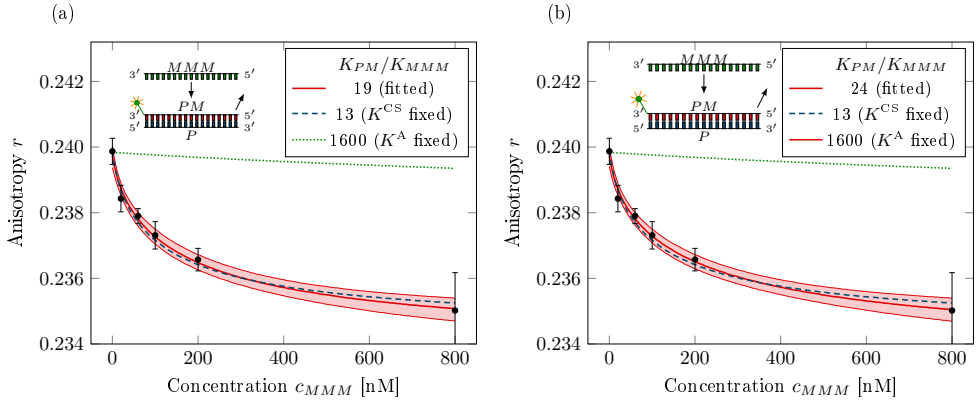

Figure 4.13. The change in anisotropy as a function of c_{MMM} for PM and MMM in competition to bind to the probe (black circles). The green dotted curve corresponds to a plot, using the exclusive binding model and taking the individual binding constants from FA, which does not at all describe the experimental data. The blue dashed curve represents a fit, where the individual binding constants are fixed and taken from FCS. The red solid curves, on the other hand, present a fit with κ as a free parameter and K_{PM} in (a) and K_{MMM} in (b) as fixed parameters, taken from the pairwise experiments. The anisotropies of the single strands and molecular complexes in all fits are free parameters. The red curves perfectly fit the data. The red shaded regions correspond to the 95 % confidence bands of the fits. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

duplexes diminishes while the number of P-MMM duplexes increases. The green dotted curve is a plot using the exclusive binding model and taking the pre-determined individual binding constants K^A_{PM} and K^A_{MMM} from FA (or Nupack) with $\kappa^{A,\mathrm{fix}} = K^A_{\mathrm{PM}}/K^A_{\mathrm{MMM}} = 1600$. It is clear that this curve does not describe the experimental data at all. In another attempt we fit the data employing the exclusive binding model, taking the individual binding constants K^{CS}_{PM} and K^{CS}_{MMM} from FCS with $\kappa^{CS,\mathrm{fix}} = 13 \pm 9.6$. The error of $\kappa^{CS,\mathrm{fix}}$ is propagated from the experimentally pre-determined K^{CS}_{PM} and K^{CS}_{MMM}. The only fitting parameters here are the anisotropies of the single strand and molecular complexes. The fit is displayed as blue dashed curve and it is easy to recognize that it is in a far better agreement with the data than the green curve. It might come as a surprise that the binding affinities obtained from FCS are more suitable to describe the competition experiment performed with FA than the binding affinities from FA itself. We suggest that this is also a signature of FA and FCS being sensitive to partially bound microstates in a different matter. In Fig. 4.14 we illustrate two different microstates: (a) MMM completely removed the PM and (b) MMM only removed the labeled side of PM, but PM is still partially bound. Considering a PM as a labeled strand, in both cases the rotational mobility of the fluorophore is

high, which means that both microstates will contribute to the anisotropy in a similar fashion. This implies that a stiffly bound MMM (a) and a partially bound MMM (b) have identical signatures in this experiment. A conclusion that we were drawing for the individual binding experiments, with MMM being a labeled strand, for FCS, but not FA. This might be rephrased in the following way: a partially bound MMM can be seen as unbound in the individual FA binding experiments presented in Fig. 4.9. However, it counts as a bound microstate in the individual FCS binding experiment Fig. 4.11 as well as in the competition experiment Fig. 4.13. Therefore, the individual binding constants from FCS much better describe the competition data from FA.

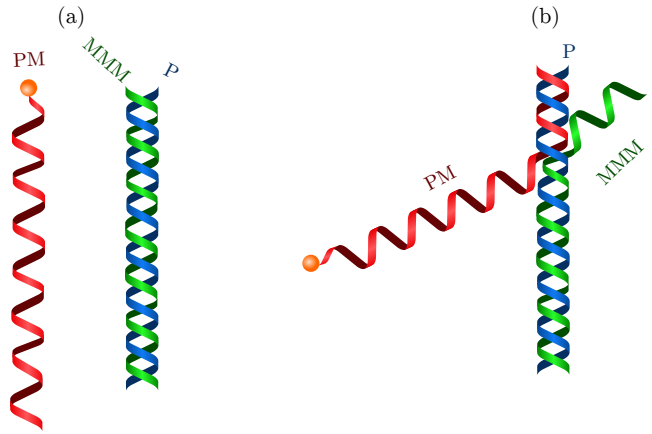

Figure 4.14. (a) The MMM strand completely removes the PM and forms a helix with the probe. (b) The loosely bound MMM weakens the binding of the PM to the probe. The fluorophore at the loose end rotates fast. Therefore, the PM in such a configuration contributes to the anisotropy in the same way as a free single strand. The PM strand in both situations (a) and (b) is counted as unbound.

To test whether there is any interaction among PM and MMM on P, we fit the data using the exclusive binding model where the ratio κ as well as the anisotropies of the single strand and the molecular complexes are free parameters (red solid curve). One of the binding constants, in Fig. 4.13 (a) the larger one K_{PM}^{CS}, is kept constant. This leads to an excellent fit with $\kappa^{\mathrm{fit}} = 19^{+4}_{-3}$. The difference between κ^{fit} and $\kappa^{CS,\mathrm{fix}}$ is relatively small. Moreover, the blue dashed curve is within the confidence band of the red curve, which reveals that the interaction between fully overlapping targets is less pronounced compared to partially overlapping ones (Fig. 4.5). Fig. 4.13 (b) shows the same fit, with the difference that now K_{MMM}^{CS} is now fixed instead. We obtain $\kappa^{\mathrm{fit}} = 24 \pm 5$, leading to $K_{\mathrm{PM}}^{CS} = 8.6 \times 10^9\,\mathrm{M}^{-1}$. This means that K_{PM}^{CS} increases compared to its value from individual experiment (Tab. 4.5) in presence of MMM, which is physically unlikely to happen.

In order to complement our data set, we repeat this competition exper-

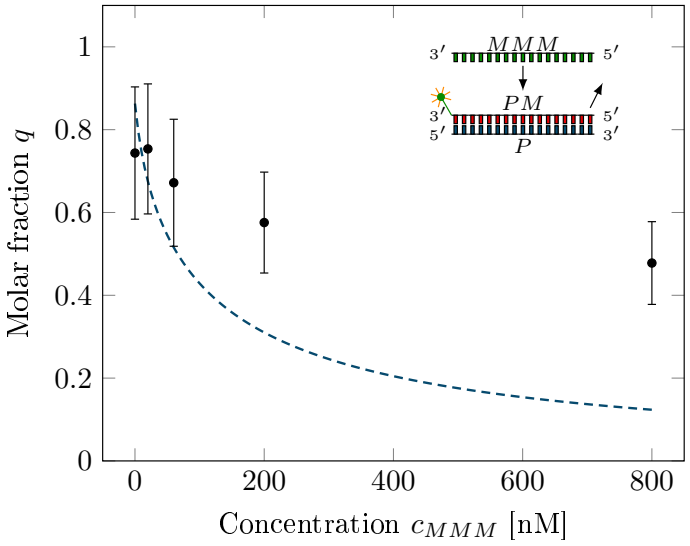

Figure 4.15. The molar fraction of double helix P-PM as a function of c_{MMM} for the same configuration as studied in Fig. 4.13 using FCS. The blue dashed curve corresponds to a fit, taking the binding constants from FCS, does not describe the experimental data. To account for an increase in viscosity the data point at 800 nM was correspondingly corrected. Adapted with permission from [135], IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft.

iment with the same configuration as shown in Fig. 4.13, but monitor the removal of PM with FCS instead of FA. Fig. 4.15 illustrates the molar fraction as a function of c_{MMM}, which we obtained from a set of individual autocorrelation measurements as presented in Appx. A. The blue dashed curve represents the probe occupancy behavior using the pre-determined binding constants from FCS (K^{CS}_{PM} and K^{CS}_{MMM}), which clearly deviates from the measured data points.

Eventually, we attempt to corroborate our claim that the centrally placed mismatch in MMM destabilize the helix, which in turn leads to partially bound microstates with small anisotropies. In order to do that, we repeat the same experiment, but add on each side of MMM and P two additional complementary bases (see Tab. 4.6). Fig. 4.16 (a) shows the corresponding individual binding experiment of MMM and P, which yields to a binding affinity of $K_{\mathrm{MMM}} = (6.2 \pm 1.3) \times 10^8\,\mathrm{M}^{-1}$. This value is significantly larger than $K^{A}_{\mathrm{MMM}} = 1.9 \times 10^6\,\mathrm{M}^{-1}$ (compare Tab. 4.2). As expected, the additional bases diminish the impact of the single mismatch and lead to a larger binding constant.

We repeat the competition experiment for the prolonged probe and MMM while the PM stays the same as previous (see Tab. 4.6). Fig. 4.16 (b) shows the drop in the anisotropy as a function of the prolonged MMM concentration

Table 4.6. Set 4 includes a perfect match PM, prolonged probe P and MMM with 20 bases. The four additional bases are marked in green. The MMM strand possess a mismatch in the center (marked in red).

Name	Sequence
P	5'-TTTTACGATCTGATCCTTTTT-3'
PM	5'-AAGGATCAGATCGTAA-3'
MMM	5'-AAAAGGATCACATCGTAAAA-3'

c_{MMM}. We fit the data using exclusive binding model and keeping both binding constants fixed. For the binding of PM to the prolonged probe we used the value $K_{\mathrm{PM}} = 8.6 \times 10^9\,\mathrm{M}^{-1}$ from the NUPACK prediction and the experimentally pre-determined $K^{\mathrm{FA}}_{\mathrm{MMM}} = 6.2 \times 10^8\,\mathrm{M}^{-1}$. The green dotted curve represents the corresponding fit which is perfectly matching the data. Adding extra bases to the probe and MMM increases the rigidity of the formed helix, which in turn reduces the impact of partially bound microstates. Therefore, the individual binding constants from FA (or NUPACK) are suitable to describe the data from the competition in this experiment while in Fig. 4.13 (green dotted curve) they could not. On the other hand, the red solid curve represents a fit, where the ratio of the binding constants κ is a fitting parameter while the larger binding constant K_{PM} is fixed. We obtain a better fit with $\kappa^{\mathrm{fit}} = (24 \pm 0.7)$, close to $\kappa^{\mathrm{A,fix}}$ from FA, illustrating that the cooperativity in this experiment has a smaller impact.

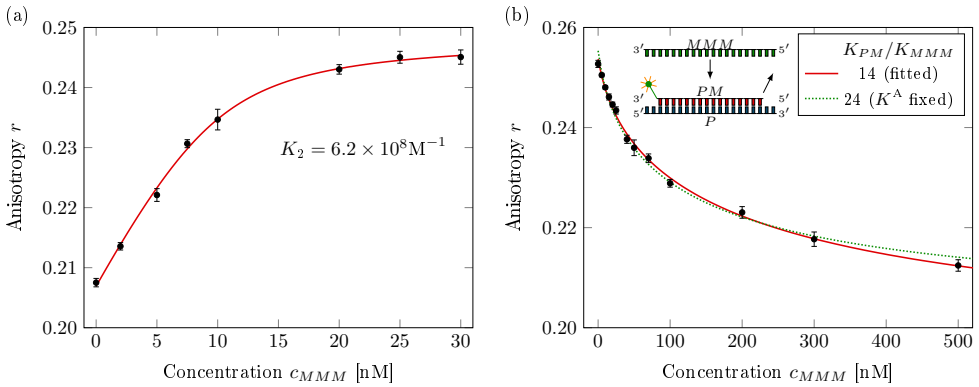

Figure 4.16. (a) The change in anisotropy as a function of c_{MMM} for P-MMM. The modified probe and MMM consist of 20 bases while modified PM has 16 bases. The data are fitted using the Langmuir model. (b) The anisotropy drops as a function of c_{MMM} for PM and MMM in competition to bind to the probe. The data are fitted using the exclusive binding model. The green curve corresponds to a fit, taking the individual binding constants from FA (or NUPACK) as constant parameters while the red curve represents a fit using K_{PM} as fixed and κ as fitting parameters.

4.5 Summary and Discussion

In this study, we compare the binding affinities obtained from pairwise binding of two DNA strands to binding constants as they emerge if two targets compete for their common complementary probe. On that account, we employed the fluorescence anisotropy (FA) technique, which is based on a rotational diffusion of the fluorescently labeled molecule. Hybridization of two complementary DNA strands, forming a double helix, leads to a decrease in the Brownian rotational diffusion of the molecule. On the other hand, if two strands bind only weakly, the rotational mobility of the attached fluorophore in such loosely bound microstates feature a single unbound strand which does not contribute in an anisotropy signal. We show that the binding constants results from FA agrees perfectly to the values predicted by the Nupack software package. The Nupack algorithm based on a nearest neighbor model [23,35,36,161,162] relies on stacked bases. This type of interaction is only present in helicoidal conformation that is detected in FA.

Additionally we used fluorescence correlation spectroscopy (FCS), a technique that is sensitive to the translational diffusion of the labeled molecule: for instance, a formation of a helix slows down the diffusion. In contrast to FA, here the loosely bound states still diffuse together and appear as bound. Since it is not possible to distinguish such states from helix conformations by FCS the binding constants are larger than in FA. For the case of P-MMM, the mismatch at the center of the strand loosens the helix and the loss of helical structure diminishes the persistence length, thereby increasing the rotational mobility of the attached fluorophore. For this case the binding constant obtained by FCS deviates by more than two order of magnitudes from FA (or Nupack). It seems that here the contribution of the loosely bound microstates to the binding is relatively more and accordingly the difference between two techniques is larger. For P-MME, however, a mismatch and fluorophore are positioned at opposite ends of a double helix. Therefore, it is unlikely that the mismatch destabilizes the helix at the fluorophore side and FA and FCS yield similar values. High concentration of ions as in our buffer tends to screen the electrical interactions between the DNA strands in solution, however, non-specific "London-" or "dispersion-" forces are not suppressed in presence of ions. For two parallel single strands we estimate the enthalpy of attraction to be approximately -1.1 kcal/mol, assuming a distance of 1 nm between the centers of their bases. This energy is in vicinity of the effective energy difference between P-MMM binding in FA and FCS. Since we did not take into account the contribution of $\pi - \pi$ interactions [4] in our estimation, this van-der-Waals energy yields a lower bound for non-specific interactions.

For two fully overlapping targets competing for binding to the same probe, we find that our results from FA cannot be explained by their corresponding individual binding constants (compare the dotted green plot in Fig. 4.13 to the experimental data). Binding in competition cannot be inferred from

simple nearest neighbor interactions that lead to Boltzmann statistics. We find, however, that the binding constants from FCS could reasonably describe our observation from FA. This emphasizes that binding constants obtained by FCS correspond to a more meaningful statistical weight. This suggests that other microstates that do not contribute to the double helix play a role in diminishing the statistical weight of nearest-neighbor helicoidal conformations. According to Fig. 4.13 a better fit is accomplished by an increase in $\kappa = K_{PM}/K_{MMM}$ compared to the pairwise assessment from FCS. This leads to a decrease in K_{MMM} in competition compared to its individual value. The result of the competition in FCS, however, reveals a strong deviation from the prediction that relies on the individual affinities from FCS. For not fully overlapping targets, where each of the competing partners has a specific overhang on the probe, we observe that the ratio of the binding affinities in competition varies by one order of magnitude compared to the corresponding value from pairwise assessment.

The deviation from the Boltzmann binding picture only occurs if the binding microstates of both competitors are affected in presence of each other but to a different degree. The degree of cooperativity seems to increase by extending the overlapping length of a competing partner. Such a change cannot emerge within the nearest neighbor model because the stacking microstate energy levels as well as their distribution do not depend on the presence of a competitor. Additionally, our observation cannot be sufficiently explained by simple entropic repulsion since this would reduce the binding constants of both competitors by the same amount. Moreover, for the case with a non-specific foothold, we find that the entropic repulsion of the overhang bases is too weak to cause the exclusive binding of the competitors as observed with specifically binding overhangs.

Our results emphasize that in situations beyond a simple helix formation binding constants may have to be interpreted carefully, considering the applied measurement technique and the type of interaction at the binding site. These situations may involve other molecular conformations besides the double helix. Moreover, we conclude that the competitive situations are more complex to be described by the binding constants deduced from pairwise considerations. This reveals a non-trivial, cooperative interaction between the competitors.

Chapter 5

DNA Hybridization on Microarray Surfaces

Contributions and copyright notice: The experimental results presented in this chapter were obtained within a collaborative project under the joint supervision of Prof. Mark Somoza (University of Vienna) and Prof. Albrecht Ott. All experiments were designed and analysed by Mina Mohammadi-Kambs (M.M.K.). The synthesis of all microarrays was done by Kathrin Hölz (K.H.) from university of Vienna. Hybridizing and scanning the microarrays were done by M.M.K for the experiments presented in Sec. 5.4 and by K.H. for the experiments illustrated in Sec. 5.6. The numerical algorithm presented in Sec. 5.3 was set up with the help of Benjamin Kambs.

Note that the main results presented in this chapter were originally published in [168] (Copyright © 2017 by American Chemical Society). Further permissions related to the material excerpted should be directed to the ACS.

In this chapter we study the hybridization of DNA strands on a microarray in a competitive environment. As explained earlier, the microarray consists of single-stranded DNA, called probes, that are immobilized on the surface (cf. Sec. 2.5). The probes are exposed to a mixture of the labeled targets in solution. Although one expects that perfectly matching probes and targets preferably hybridize to each other, hybridization of non-complementary partners still occurs. These non-specific bindings have lower binding affinities compared to the perfectly matching strands. The cross-hybridization is a direct result of the imperfect recognition among similar probes. Fig. 5.1 illustrates immobilized probes (blue strands) that are exposed to the labeled targets (red strands). Each target is complementary to the probe with the same numbering but it has mismatches (the colored crosses) with the other

probes. For example target 1 is perfectly matching to probe 1 while it has one, two and three mismatches with probes 4, 3 and 2, respectively.

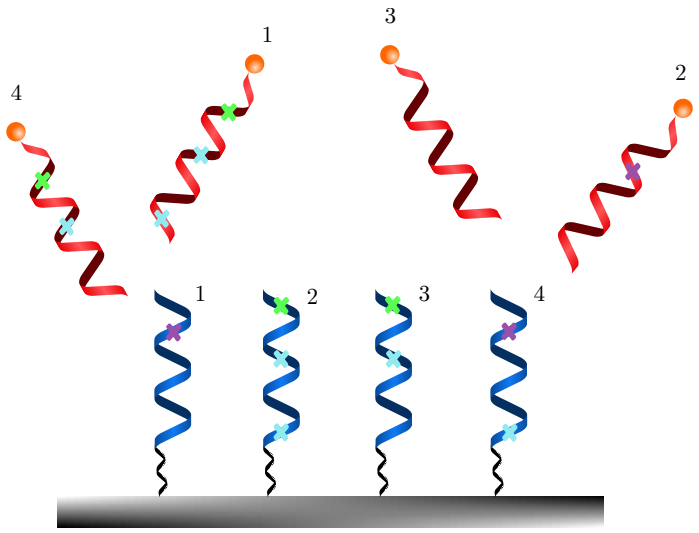

Figure 5.1. The surface bound probes are shown in blue and the labeled target strands in solution are in red. The probes and the targets with the same numbering are perfectly matching while each target strand have mismatches respect to the other probes (colored crosses). For example target 1 is complementary to probe 1 but it has one, two and three mismatches with probes 4, 3 and 2, respectively. Therefore, target 1 may hybridize to other probes apart from its corresponding complementary. The similarity between the probes on the microarray leads to the cross-hybridization.

The complexity of the system increases by adding more strands to the solution that can cross hybridize to several probes on the surface. Such cross-hybridization introduces errors during the transmission of information and makes the results from microarrays experiments difficult to interpret. Previous theoretical works established various strategies to find sets of non-interacting probes [45–50]. For instance, the thermodynamics model was employed to predict the duplex and mismatched binding stabilities [169,170]. The approach is based on discriminating perfectly matching and mismatched hybridizations by ensuring sufficiently large differences between their free energies. For that it is necessary to know the binding free energies of all pairs of sequences, in particular also for non-specific bindings. These energies can be calculated within the nearest neighbor model [33] for single mismatch hybridization. However, to our knowledge similar predictions for an arbitrary number of mismatches do not exist. On the other hand, Pozhitkov *et al.* [169] predicted hybridization intensities for perfect match and single mismatch binding resulting from Gibbs free energy (ΔG) calculations in bulk and compared them with experimental results on the surface. Their evaluation revealed a poor statistical relationship

between predicted and experimental intensities. Although it was shown that in special cases it is possible to bring experimental results to a fair agreement with predictions based on theoretical values of ΔG in bulk [89, 92], the approach is not easily transferable due to a large variety of surface chemistries that can lead to artifacts as well as point defects during the in situ synthesis process. Therefore, estimating the hybridization intensities for microarrays from available thermodynamics parameters in bulk is not straight forward. In this work we follow a simplified approach: we assume that for the sequences of one set with a length L a minimum number of mismatches d_m exists so that no two sequences cross hybridize as long as they have d_m mismatches. In the following we offer a brief introduction to the information theoretical concept of Hamming distance, from which the idea above is borrowed. By implementing methods from graph theory and employing the well-known local search algorithm [171,172], we then construct set of probes, which differ by at least d_m bases. Such sets are refereed to as independent. Experimentally, we determine a necessary d_m for the length of $L = 7$ and eventually demonstrate that the corresponding set is independent i.e. free of cross-hybridization.

In the second part of the chapter we illustrate how by introducing the unlabeled complementary targets of selected probes we can diminish the effect of the cross-hybridization.

5.1 Sequences Without Runs of Guanine

For each length of the sequence L, we can construct 4^L possible sequences with all permutations of four DNA bases (A, T, C, G). Some of the resulting sequences, however, contain undesired conformations which prevent them from hybridizing to their complement. One group is the sequences including runs of at least four guanine which we refer to as 4G sequences. The guanine-rich region of the DNA strands can form a complex structure such as a G-quadraplex [173]. The previous studies [173–175] show that probes containing multiple guanines in a row tend to increase cross-hybridization signals and reduce target-specific hybridization, making the gene expression measurements unreliable. Moreover, they exhibit abnormal binding behavior in microarrays as well as in gene expression assays [173–175]. Therefore, it is better to avoid them in microarray designs. In this work we first find such structures and then disregard them as well as their complement (at least four cytosine (4C) in a row) from the selected probes on our microarray set. While this is done computationally during a set generation presented in Sec. 5.3, it is possible to numerically estimate the number of eliminated sequences. For this reason we construct sequences of length $L = m + k$, where m is the number of guanine (G) and k the number of non-guanine bases ($\overline{G}$=C, T, or A). It is possible to unambiguously identify any given sequence by stating the number of $\overline{G}$ bases preceding each guanine G, which is illustrated for $L = 8$ in Fig. 5.2

(a). Reading a sequence from left to right, we first encounter a G which is categorized as *no* $\overline{G}$ *in front*. The second G is at a third position with one $\overline{G}$ preceding it, i.e. it falls into a category of *one* $\overline{G}$ *in front*. The third G at position four is assigned to the same category as there is no additional $\overline{G}$ in between. Labeling these categories with letters A, B, C, ... corresponding to *no, one,* and *two* $\overline{G}$ *in front*, the illustrated sequence can be written as ABBCD. We realize that for a given k there are at maximum $k+1$ groups and each sequence can be written as $L-k$ of those letters. To avoid sequences with the runs of at least four guanine (4G), a non-guanine base must interrupt a run of 3G. Using our notation, this is equivalent to not having more than three times the same letter in a sequence, e.g. 4×A or 4×B are not allowed (compare Fig. 5.2 (b) and (c)). We can realize such sequences by using a modified model borrowed from combinatorics: Assuming three sets, each containing $L-k$ letters A, B, C, ..., every sequence can be created by randomly picking letters from these sets. As each letter exists only three times, it is not possible to obtain a sequence with runs of G that exceed 3G in a row. As it turns out, the number of possibilities $P(L)$ to select $L-k$ elements out of 3 identical sets including $k+1$ elements is given by quadrinomial coefficient [176, 177]

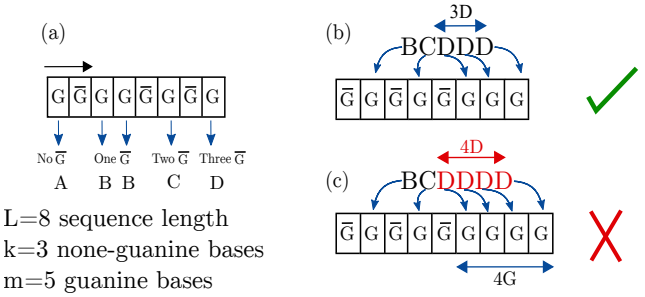

Figure 5.2. (a) The black arrow shows the direction of reading a sequence from left to right. The sequence has a length of $L = 8$ with $m = 5$ guanine base and $k = 3$ non-guanine bases. The bases that are not guanine are shown as $\overline{G}$. Each G base is categorized in different groups A, B, C and D depending on how many $\overline{G}$ base came in front of it. (b) The string written as BCDDD can be transformed to the sequence consists of G and $\overline{G}$ bases. To avoid runs of 4G, each group should appear at maximum 3 times (BCDDD). (c) Group D appears more than 3 times (BCDDDD). This leads to an unwanted run of four guanine.

$$P(L) = \binom{k+1}{L-k}_3 . \tag{5.1}$$

The number of possible sequences without any runs of 4G ($N_{4\overline{G}}$) is the summation of $P(L)$ over all k for a given length L multiplied by 3^k different

possibilities for substituting all $\overline{G}$ bases by C,T or A. Therefore we have:

$$N_{4\overline{G}}(L) = \sum_{k_{\min}}^{L} \binom{k+1}{L-k}_3 \cdot 3^k, \tag{5.2}$$

where $k_{\min} = \left\lceil \frac{L-3}{4} \right\rceil$ is the minimum number of $\overline{G}$ that are necessary to avoid runs of at least four guanine. Accordingly, we can write the number of the sequences including runs of 4G as

$$N_{4G}(L) = 4^L - N_{4\overline{G}} = 4^L - \sum_{k_{\min}}^{L} \binom{k+1}{L-k}_3 \cdot 3^k, \tag{5.3}$$

where 4^L is the number of all possible sequences with the length of L.

To verify Eq. (5.3) we numerically calculated $N_{4G}(L)$ by creating all 4^L sequence for $L \leq 7$ and discarding the ones containing runs of at least four guanine. We find that the numerical values match the analytical results. Fig. 5.3 presents the fraction of $N_{4G}(L)$ compared to total 4^L possible sequences as a function of sequence length L. The inset of the figure shows that for very short length $L \leq 7$ the fraction is below 1.5 % while for longer lengths it increases. For instance for $L > 200$ around more than 50 % of all possible sequences include runs of 4G. Accordingly, we only expect to find a minor effect of discarding 4G sequences while constructing our sets without cross-hybridization (compare Sec. 5.3).

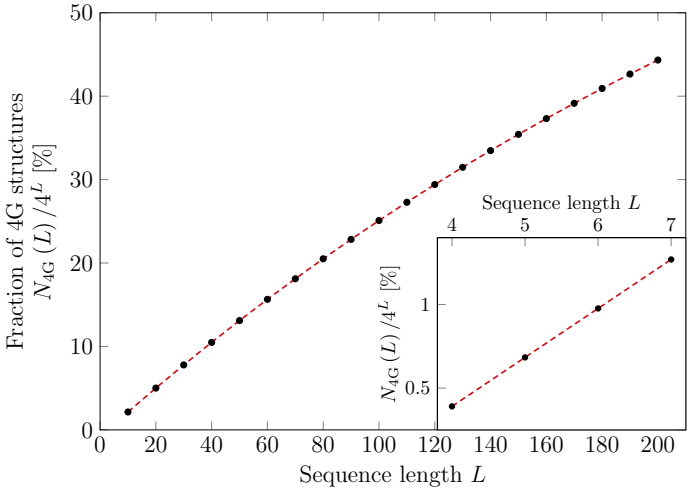

Figure 5.3. Fraction of all 4^L sequences containing runs of at least four guanine ($N_{4G}(L)$) for different sequence length L. The inset shows that the fraction for short lengths $L \leq 7$ stays below 1.5 % while it rises to almost 50 % for sequence lengths of 200. Dashed lines are a guide for the eye. Adapted with permission from [168] published by ACS.

5.2 Coding Theory and Hamming Distance

Coding theory is a branch of mathematics that studies the properties of the codes in various scientific disciplines such as information theory [178, 179]. A code is a set of codewords and each codeword consists of letters that are taken from an alphabet of size n. The length of the codeword is L. In analogy, here a DNA strand with a length L is taken as a codeword and DNA bases are the letters of an alphabet with four elements (A,T,C,G). Accordingly, a code is a set of DNA strands. An important measure in coding theory is Hamming distance d which is the number of positions that two codewords of the same length differ. Accordingly, the number of bases that two DNA strands differ is defined as a Hamming distance. Fig. 5.4 (a) and (b) illustrate each two codewords of length $L = 7$ with a Hamming distance of 3, for a binary alphabet (a) and DNA sequences (b). Here we assume there is a minimum Hamming distance d_m in such a way that as long as $d \geq d_m$ there is no cross-hybridization among the sequences of one set. Sequences with $d < d_m$ on the other hand cross hybridize. Fig. 5.4 (c) represents an exemplary set of four DNA strands with the minimum Hamming distance $d_m = 2$. This means that each pair of sequences in this set have $d \geq 2$.

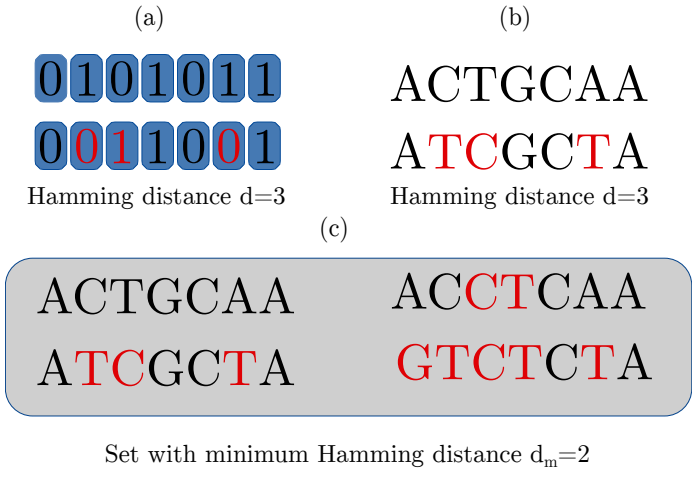

Figure 5.4. (a) Two codewords made of binary numbers 0 and 1. (b) Two DNA sequences consist of four DNA bases A,T,C,G. The Hamming distance in both (a) and (b) is $d = 3$. (c) A set of four sequences with the minimum Hamming distance $d_m = 2$.

For a given sequence of a length L, the number of the sequences with Hamming distance d is simply calculated by

$$H_d(L) = \binom{L}{d} \times 3^d. \tag{5.4}$$

Fig. 5.5 represents $H_d(L)$ for different Hamming distance d for $L = 7$ (circles) and $L = 8$ (triangles). The red solid and blue dashed curves therein are ob-

tained from Eq. (5.4) by substituting factorials by the corresponding gamma functions. The figure depicts an increase in the number of possible sequences by increasing a Hamming distance for $d < L - 2$.

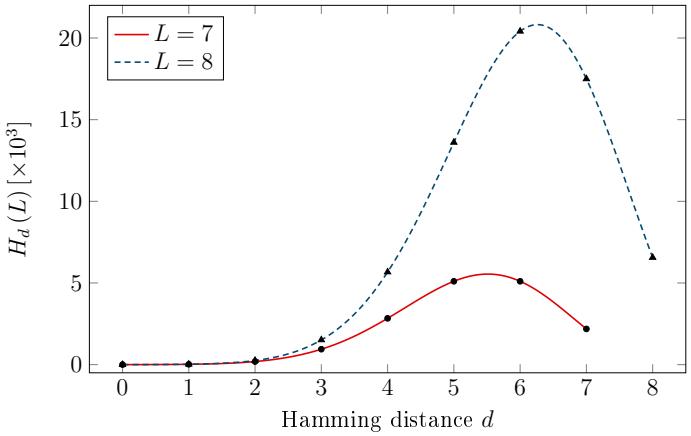

Figure 5.5. Number of possible sequences $H_d(L)$ as a function of Hamming distance d for $L = 7$ (data: circles) and $L = 8$ (data: triangles). The red and blue curves represent the corresponding gamma functions for the factorials in Eq. (5.4).

In order to find the number of sequences $N(L, d_m)$ with a length L and a given minimum Hamming distance d_m that do not cross hybridize, i.e. $d \geq d_m$, we need to sum $H_d(L)$ over all $d \geq d_m$ according to

$$N(L, d_m) = \sum_{d=d_m}^{L} H_d(L) = \sum_{d=d_m}^{L} \binom{L}{d} \times 3^d. \tag{5.5}$$

For example the number of sequences with $L = 7$ and $d_m = 5$ that do not cross hybridize is

$$N(L = 7, d_m = 5) = \binom{7}{5} \times 3^5 + \binom{7}{6} \times 3^6 + \binom{7}{7} \times 3^7 = 12393. \tag{5.6}$$

Fig. 5.6 illustrates the number of sequences with $d \geq d_m$ ($N(L, d_m)$) for $L = 7$ as a function of minimum Hamming distance d_m, while the inset shows $H_d(L)$. The integrated area of the shaded region depicts the number of the sequences as calculated by Eq. (5.6), which do not cross hybridize, assuming a minimum Hamming distance of $d_m = 5$.

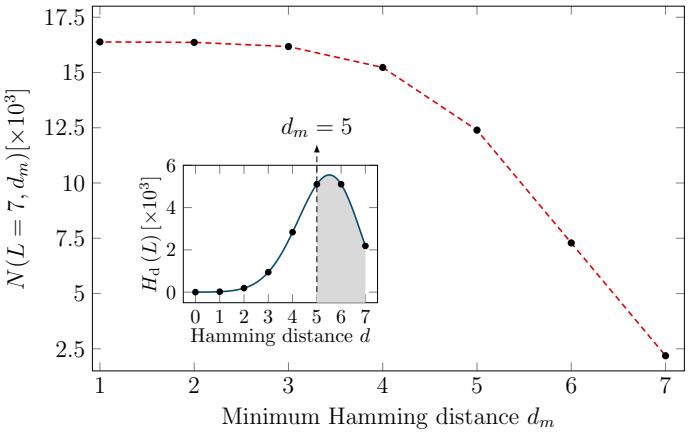

Figure 5.6. The number of sequences $N(L, d_m)$ with $d \geq d_m$ as a function of minimum Hamming distance d_m for a given length of $L = 7$. This number decreases for large d_m. The dashed line is a guide for the eyes. The inset shows $H_d(L)$ for $L = 7$ and the integrated shaded region equals $N(L = 7, d_m = 5)$. Adapted with permission from [168] published by ACS.

5.3 Maximum Independent Set

After establishing analogies between coding theory and a set of non-interacting (independent set) DNA sequences, we transfer results from coding theory to our current problem. Finding the largest set of independent sequences is a NP-hard problem (not solvable in polynomial time) with no exact answer [180,181]. However, there are methods leading to the set with the approximate largest possible size [182,183]. Here, by applying a well-known local search algorithm [50,171] and implementing graph theoretical methods we find a maximum set of independent strands, called MIS [183,184].

A graph as shown in Fig. 5.7 (a) consists of vertices (red circles) and edges (blue lines), which connect pair of vertices. Two vertices are adjacent if they are connected with an edge. In analogy, here we consider DNA sequences as vertices. In Fig. 5.7 (b) we show a small set of five sequences with $L = 3$ and $d_m = 2$. If two strands hybridize to each other we connect them by an edge, i.e. the blue lines. An independent set is a set that no two vertices (sequences) are connected by an edge (sequences in a black dashed box). To find the largest possible set of independent sequences for a given L and d_m we create an adjacency matrix. This square matrix has $4^L \times 4^L$ elements. If sequences i and j cross hybridize, i.e. $d < d_m$, they are adjacent. In this case the elements of the matrix are $A_{ij} = 1$ and otherwise $A_{ij} = 0$. Note that sequences are not self-adjacent, $A_{ii} = 0$. For instance, the adjacency matrix in Fig. 5.7 (c) corresponds to the set of five sequences with $d_m = 2$ depicted in (b).

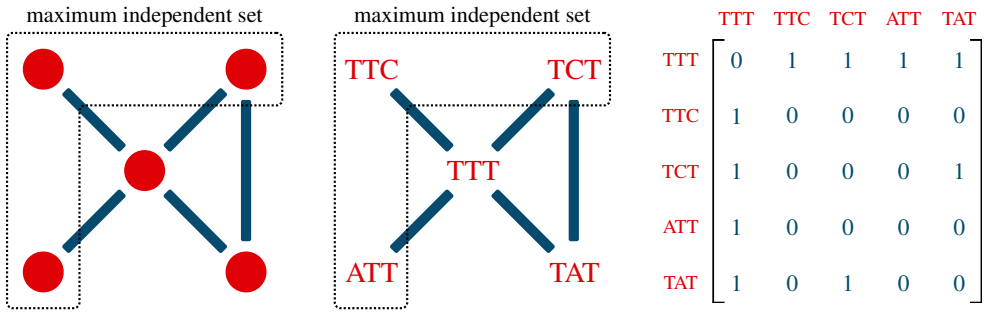

Figure 5.7. (a) A graph made of vertices (red circles) along with the edges (blue lines). (b) The sequences with $d < 2$ cross hybridize and therefore they are connected by edges, while sequences with $d \geq 2$ are independent. The vertices and sequences within the black box in (a) and (b) indicate a maximum independent set. (c) The corresponding adjacency matrix for the sequences that are shown in (b). The elements of an adjacency matrix are $A_{ij} = 1$, if two sequences cross hybridize, and otherwise are $A_{ij} = 0$. Adapted with permission from [168] published by ACS.

We start with a pool of all 4^L available sequences and an initially empty set of independent sequences. The algorithm selects an arbitrary sequence from the pool and transfers it to the set. Using the adjacency matrix that we constructed beforehand we identify all sequences that cross hybridize with the initial sequence and we cross them from the pool. In the next iteration we pick a new sequence from the pool and repeat the procedure until the pool of available sequences is depleted. The selection of a sequence in each iteration is random. Therefore, we may get different sizes for MIS by every time running the algorithm. The algorithm does not try all combinations of sequences thus it does not necessarily find the overall maximum independent set, but instead proposes multiple sets whose sizes are maximized within the respective run of the algorithm. We consider the largest set among them as an approximation of the MIS. As the size of the adjacency matrix increases exponentially with the sequence length (4^{2L} elements for the length of L), the algorithm requires a huge amount of memory for longer lengths. Fig. 5.8 (a) presents set sizes that we found with our algorithm for $4 \leq L \leq 7$ and different minimum Hamming distances $2 \leq d_m \leq 6$. We compared them to lower and upper bounds on the MIS size named Gilbert-Varshamov and Singleton-bounds [178, 185] provided by coding theory. They are given by

$$\underbrace{\frac{4^L}{\sum_{d=0}^{d_m-1}\binom{L}{d}\times 3^d}}_{\text{Gilbert-Varshamov bound}} \leq M(L,d_m) \leq \underbrace{4^{L-d_m+1}}_{\text{Singleton bound}} . \tag{5.7}$$

Accordingly, Fig. 5.8 (b) shows the set sizes for $4 \leq L \leq 7$ and $d_m = 4$ within

the corresponding bounds (shaded region). As can be seen for $L = 4, 5, 6$ our set sizes coincide with the Singleton bound, indicating that we found a MIS. The set sizes are additionally summarized in Tab. 5.1 and compared to values from literature [50] and the singleton bound. As can be seen, the set sizes in the present work are hardly smaller than the literature set sizes.

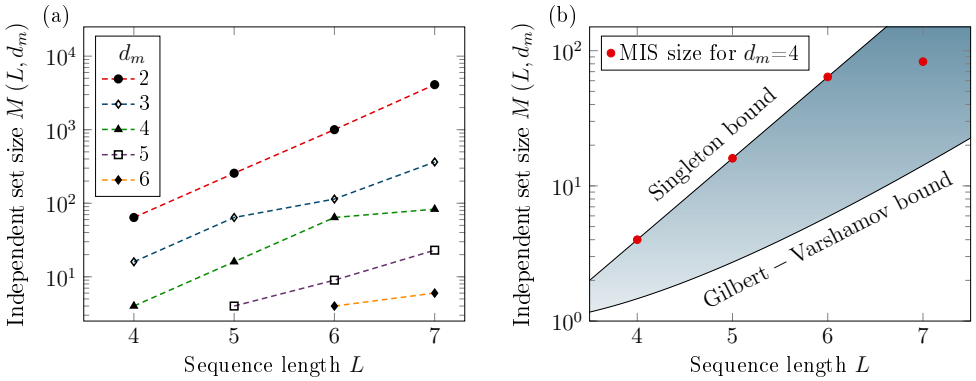

Figure 5.8. (a) Maximum set sizes obtained from our local search algorithm for different lengths and minimum Hamming distances. (b) All maximum set sizes for $4 \leq L \leq 7$ and $d_m = 4$ are within the upper and lower bounds referred to as Singleton and Gilbert-Varshamov bounds, respectively. Adapted with permission from [168] published by ACS.

Fig. 5.9 presents a distribution of the set sizes for $L = 7$ and $d_m = 5$. The blue columns show the set sizes without any 4G and 4C sequences while the red columns correspond to the set sizes including all the sequences. Removing 4G and 4C sequences changes the pool of the available sequences which in turn affects the distribution of the set sizes, i.e. the height of the red and blue histograms. However, in both cases we obtain 23 as a MIS size. As we show in Fig. 5.3 the fraction of 4G sequences for short lengths is very small, thus crossing them does not change the pool and therefore the MIS size significantly. However, for longer lengths we expect that discarding these sequences changes the set sizes.

The presented experiments in the following sections are based on the sets without 4G and 4C sequences. However, if it is experimentally desire to obtain more homogeneous hybridization signals, it is possible to further restrict the pool of sequences with a GC content of 50 % [46, 50]. Tab. 5.2 shows set sizes obtained by our algorithm for these sequences. As can be seen the sizes are smaller compared to Tab. 5.1. However, our set sizes are often larger than the sizes reported in literature [46, 50].

Table 5.1. Comparing the set sizes achieved by our algorithm to the values reported in literature and Singleton bound. The set sizes that are written after the slashes correspond to the set without 4G and 4C sequences.

L	d_m	MIS	MIS [50]	Singleton bound
4	3	16	-	16
4	2	64	-	64
5	4	16	16	16
5	3	64	-	64
5	2	256/252	-	256
6	5	9	9	16
6	4	64	64	64
6	3	114	-	256
6	2	1024/1001	-	1024
7	6	6	8	16
7	5	23	23	64
7	4	83	78	256
7	3	364	-	1024
7	2	4096	-	4096

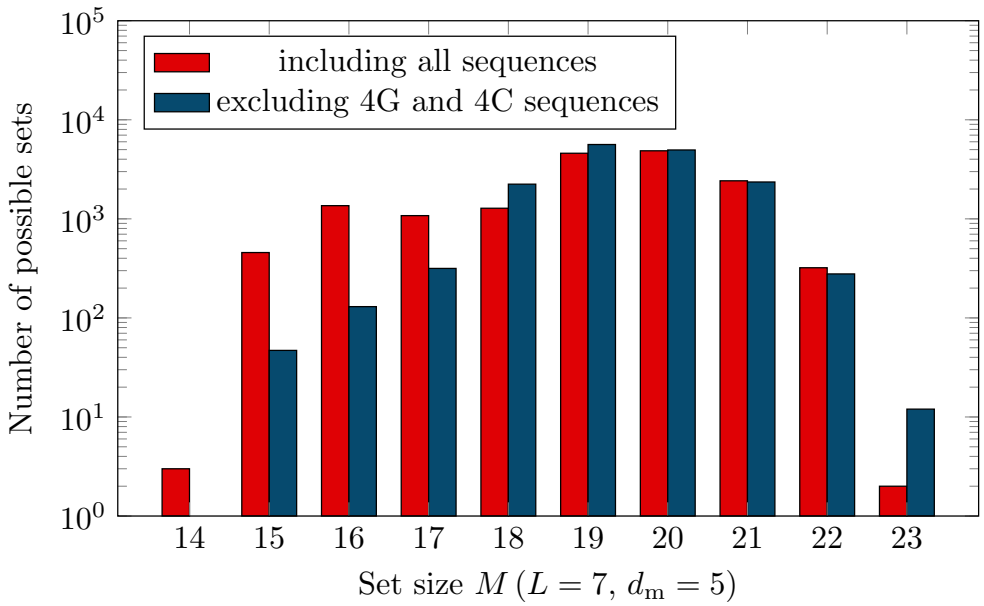

Figure 5.9. The set sizes for $L = 7$ and $d_m = 5$ before (red columns) and after (blue columns) crossing 4G and 4C sequences. The height of the histogram shows the number of sets found for each $M(L, d_m)$. Adapted with permission from [168] published by ACS.

Table 5.2. Comparison of the set sizes achieved by our algorithm to values reported in literature, where the sequences are restricted to a 50 % GC content.

L	d_m	MIS	MIS [46]	MIS [50]
4	3	12	-	-
4	2	48	48	48
5	4	10	-	-
5	3	27	-	-
5	2	156	120	142
6	5	8	-	-
6	4	36	-	-
6	3	82	56	85
6	2	640	-	-
7	6	7	-	-
7	5	21	-	-
7	4	65	-	-
7	3	238	224	230
7	2	2240	-	-

5.4 Determining Minimum Hamming Distance

To find a maximum independent set, first we need to experimentally determine a minimum Hamming distance for a given L. Since $L = 7$ is the longest length that we could study with our algorithm, here we design a microarray including sequences with 7 main bases. As the fluorescence intensity for $L = 7$ is too low, we add four additional bases, two to each terminal of all sequences, i.e. all sequences have a length of 11 bases. The additional bases are chosen to be 'CT' and 'TC' at the 3' and 5' ends, respectively and are always complementary to the corresponding targets. The corresponding sequences are shown in Appx. B. We consider an arbitrary sequence with $L = 7$ as a "PM" target. We immobilize its complementary probe (3'-CTACCGTACTC-5') as well as some of the sequences which possess various number of mismatches in different positions compared to it. The impact of the mismatch position and its distribution along the sequence on the binding affinity and discrimination between two sequences is extensively studied [89, 92]. To investigate the dependency of hybridization probability on the positions of the mismatches we located the mismatches at the ends, in the middle or uniformly distributed them. Fig. 5.10 shows the results from microscope pictures obtained after hybridizing 25 nM PM target on the microarray at 32 °C. The background fluorescence visible in the fluorescence scans originates from unhybridized labeled targets. Each block (a)-(g) represents a set of sequences with one to seven mismatches which refer to as MM1-MM7, respectively. Each feature

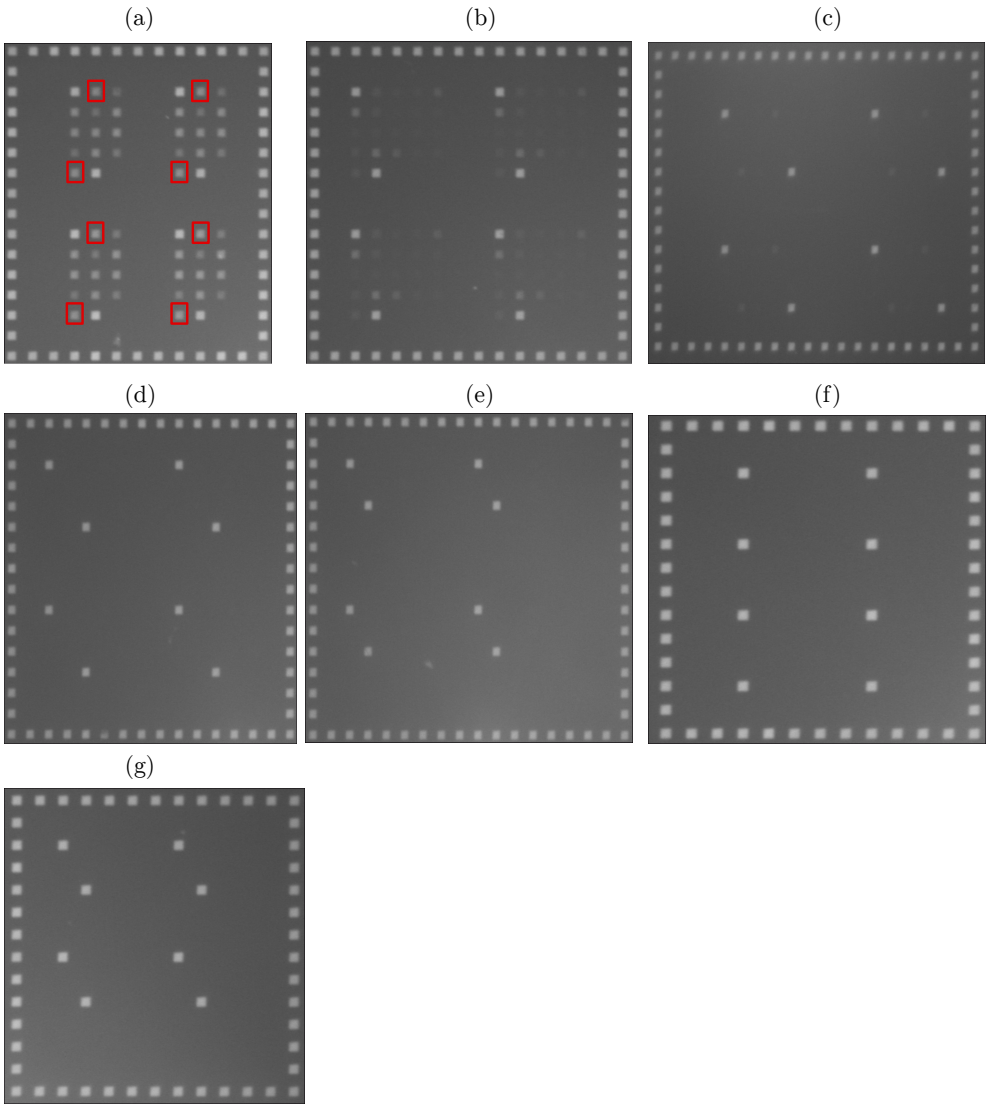

Figure 5.10. The fluorescence intensity of the hybridized PM target on the microarray. Each block (a-g) consists of the sequences with one to seven mismatches. Each sequence is repeated 8 times within a feature block. For instance the red squares show the 8 replicates of sequence S_1. The blocks are surrounded by a frame of PM hybridization.

inside every block corresponds to the fluorescence intensity of the hybridization of PM target to the corresponding sequence. The features on the frames belong to the hybridization of the PM target to its complementary probe. Every sequence appears 8 times within each feature block, e.g. the red squares show the 8 replicates of sequence S_1. Accordingly, we arranged the replicates of other sequences inside each block. Fig. 5.11 shows the computationally reconstructed microarray pictures after background subtraction. For (d) to

(g), corresponding to MM4 to MM7, we detect only hybridization to a PM sequence.

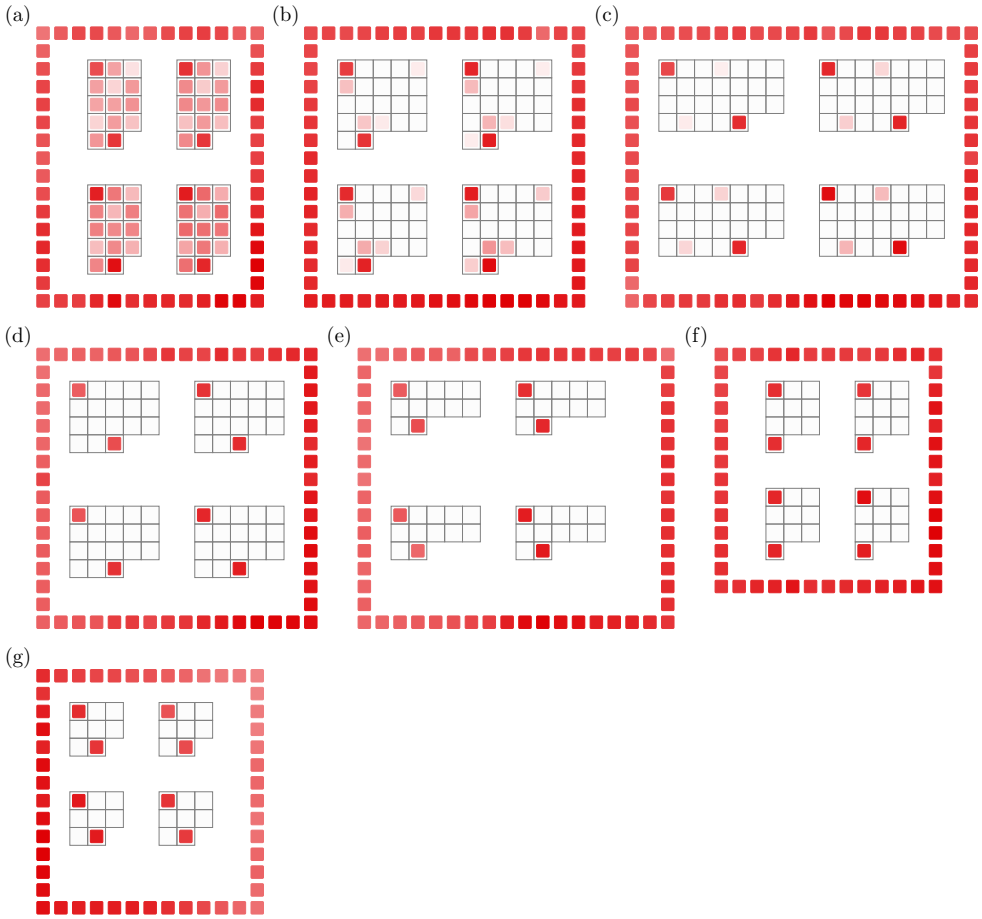

Figure 5.11. The computationally reconstructed feature blocks from Fig. 5.10 after background subtraction. For $MM \geq 4$ (d)-(g) we do not detect any cross-hybridization intensity. Adapted with permission from [168] published by ACS.

The fluorescence intensities on one microarray may vary due to the fluorescent stains, inhomogeneities of the surface or an illumination gradient during synthesis [85]. To have better statistics for each sequence we average the intensity of all 8 replicates and treat the corresponding standard deviation as an estimation for its error. For better comparability we normalize all intensities with respect to the average PM intensity. The stepwise coupling efficiency of the synthesis of our microarrays is $\geq 99\,\%$ [136]. Considering the sequence length of $L = 7$ in our experiments, the yield of probes free of any synthesis defects is approximately $93\,\%$. It is important to note that all sequences are subject to the same synthesis error leading to a homogeneous loss of hybridiza-

tion intensity. Therefore, the relative intensities are not affected by the loss. Fig. 5.12 shows the normalized fluorescence intensity of hybridization for sequences with a single mismatch at different position, i.e. for MM1 as shown in Fig. 5.10 (a). The sequences with a single mismatch at the middle show lower intensities, which is in agreement with earlier investigation [89, 92]. This is explained via a strong destabilization caused by centrally located mismatches, which easily unwind a formed helix.

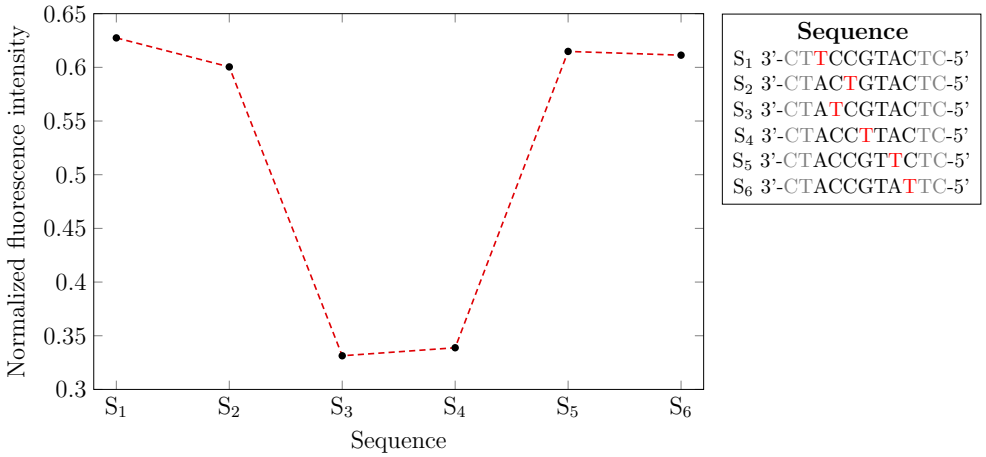

Figure 5.12. The fluorescence intensity of the hybridization for the sequences with a single mismatch located in different position. The sequences with a mismatch at the middle have lower intensity. Adapted with permission from [168] published by ACS.

Table 5.3. Sequences that exhibit maximum fluorescence intensity within each feature block as well as their corresponding normalized intensities $I_{\max}$ and standard deviations σ.

number of mismatches	sequences	$I_{\max} \pm \sigma$
0	3'-CTACCGTACTC-5'	1 ± 0.066
1	3'-CTTCCGTACTC-5'	0.63 ± 0.1
2	3'-CTACCGTCTTC-5'	0.37 ± 0.074
3	3'-CTACCGACTTC-5'	0.27 ± 0.073

To determine the minimum Hamming distance d_m for each mismatches (MM1-MM3) we have considered in particular the sequences which lead to the strongest cross-hybridization intensities ($I_{\max}$). Tab. 5.3 lists the corresponding normalized intensities of the PM along with these sequences for MM1-MM3. Additionally, Fig. 5.13 illustrates the intensities assuming the obtained fluorescence intensities are normally distributed. The peak centers correspond to the respective highest normalized intensity while the standard

deviation determines their width. To discriminate the PM intensity from all other non-specific hybridization, their normal distribution must be very well separated. We see that the intensity distribution of PM overlaps with MM1, while it is separated from MM2 and MM3. Therefore, one might conclude picking two or three mismatches as a minimum hamming distance is sufficient. On the other hand we know that the binding affinity varies depending on the precise sequence and its concentration. Therefore, the intensities of the perfectly matching partners spreads in a wide range, while we here only investigated the intensity of a specific sequence. For instance, we expect that a sequence 3'-CTATATATATC-5' with the length core of 7 bases possessing no G or C bases has a very low binding affinity within the sequences of the same length. We find the free energy of the binding of this sequence to its perfect match using Nupack software as $\Delta G = -9.6\,\mathrm{kcal/mol}$. We then calculated the fluorescence intensity of the hybridization I using Langmuir isotherm [89]

$$I = \frac{\exp\left(-\Delta G_D/RT\right) \cdot c_t^{\mathrm{ini}}}{1 + \exp\left(-\Delta G_D/RT\right) \cdot c_t^{\mathrm{ini}}}, \tag{5.8}$$

where c_t^{ini} is the initial concentration of the target. This intensity is only 16.5 % of the intensity of the PM sequence (3'-CTACCGTACTC-5') that we used on our microarray (Fig. 5.10). This demonstrates that there are some weakly binding sequences with the length 7 whose their perfectly matching hybridization intensities may still be below the 27 % intensity of the sequence with three mismatches (see Tab. 5.3). This clearly shows that choosing even $d_m = 3$ as a minimum Hamming distance is not sufficient to precisely discriminate PM intensity from all mismatch intensities. Therefore, to check the independency of our sets, we consider only $d_m = 4$ and $d_m = 5$ in the following.

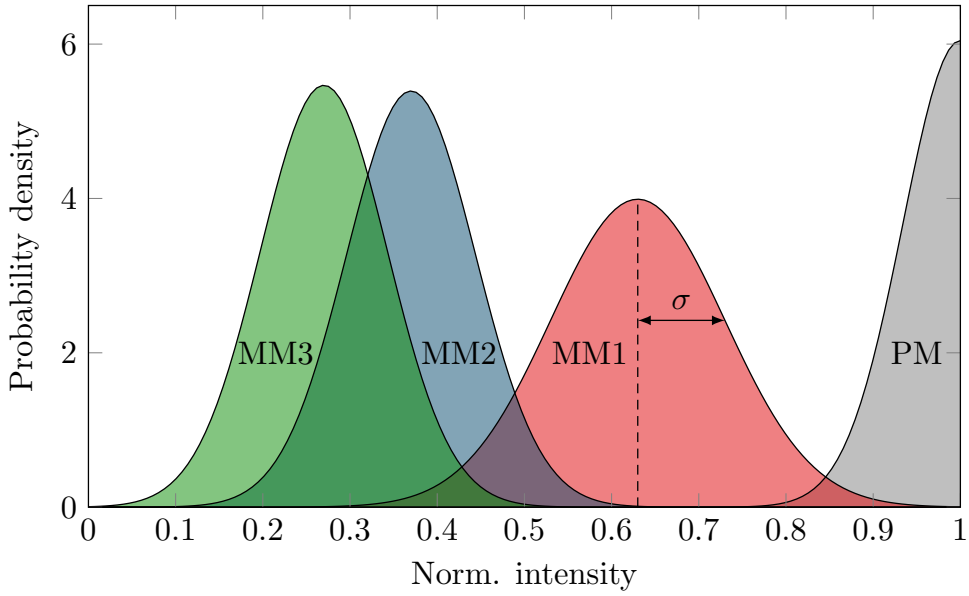

Figure 5.13. Normal distribution of the intensities for PM and MM1-MM3 and their corresponding standard deviations taken from Tab. 5.3. The peak centers correspond to the normalized average intensity. The average normalized intensity of 27 % for MM3 is still too high to precisely discriminate the PM intensity from other cross-hybridized intensities (cf. main text). Adapted with permission from [168] published by ACS.

5.5 Finding an Independent Set

The experiments presented in this section were obtained using microarray, on which we immobilized the 23 and 83 probe sequences from the maximum independent sets for $d_m = 5$ and $d_m = 4$ as obtained from our algorithm. First, we consider the set with $d_m = 5$, which we simultaneously hybridized with the perfectly matching targets of three probes 3'-CTTGTAGGATC-5', 3'-CTTTCCCGCTC-5' and 3'-CTGGCTGCTTC-5'. Fig. 5.14 (a) shows the reconstructed normalized intensities I_{seq} after background subtraction. The green bars indicate hybridization signals at the expected positions corresponding to the perfectly matching probes of the chosen targets. Using the highest intensity as a reference the other two intensities are reaching 24 % and 31 % of this level. The intensity of all the other probes (shown in blue) are scattered with $\sigma = 0.3\,\%$ around the average value of zero, which can be attributed to the background fluorescence noise. These intensities stay well below 2 % within a 5σ confidence interval. Note that negative values correspond to intensities below the average background. To check whether the set with $d_m = 4$ is independent as well, we hybridize only a single complementary target of one of the probe 3'-CTACCTGGCTC-5'. The reconstructed measurement is depicted in Fig. 5.14 (b). As before, the green bars show corresponding PM

hybridization intensity. However, the red bars show that 11 additional probes exhibit significant cross-hybridization signal exceeding the 2 % confidence interval obtained before. This indicates that for $L = 7$ a maximum independent set includes only 23 sequences with $d_m = 5$. Note that the lists of the sequences for both $d_m = 4$ and $d_m = 5$ as well as the original pictures taken with the microscope camera are shown in Appx. B and Appx. C.

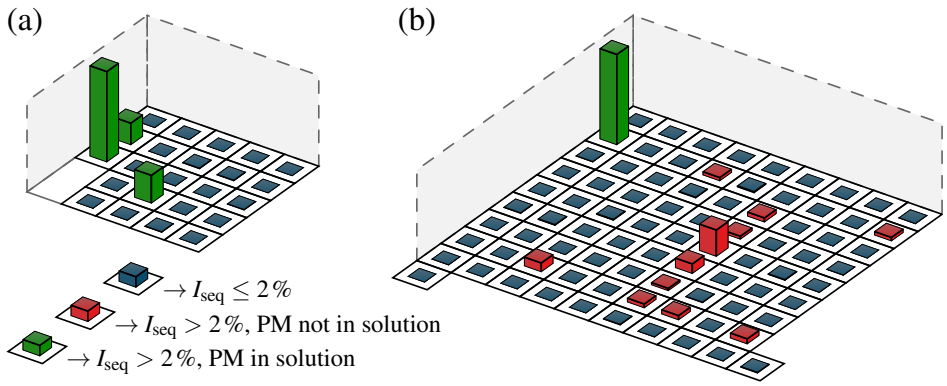

Figure 5.14. Hybridization intensities reconstructed from the raw data. Two sets of sequences with two different minimum Hamming distances. (a) an independent set including 23 sequences with $d_m = 5$ and (b) a set consists of 83 sequences with $d_m = 4$. The green bars are the intensities of the sequences that their PM targets are present in the solution. The blue bars correspond to the intensities with $I \leq 2\,\%$ at the background level while the red bars in (b) are the cross-hybridization intensities with $I > 2\,\%$. From the results we conclude that $d_m = 5$ is the appropriate minimum Hamming distance for $L = 7$. Adapted with permission from [168] published by ACS.

5.6 DNA Hybridization in Presence of Protectors

As we discussed in the previous section, to avoid cross-hybridization in one set we need to determine an appropriate minimum distance between its sequences. The experiments in this section investigate a possible strategy to decrease the necessary Hamming distance, accordingly the cross-hybridization and thereby achieve larger set sizes. Here, we study sets including sequences with $L = 17$ for different Hamming distances. The outline of the idea is as following: besides the surface bound probes and fluorescently labeled targets we additionally add unlabeled sequences to the hybridization solution, called protectors. The protectors are complementary to selected probes. The probes exhibit higher affinities to the protectors than to the mismatched labeled targets. Accordingly, we expect that the protectors win the competition and hence the fluorescence signal from cross-hybridization decreases.

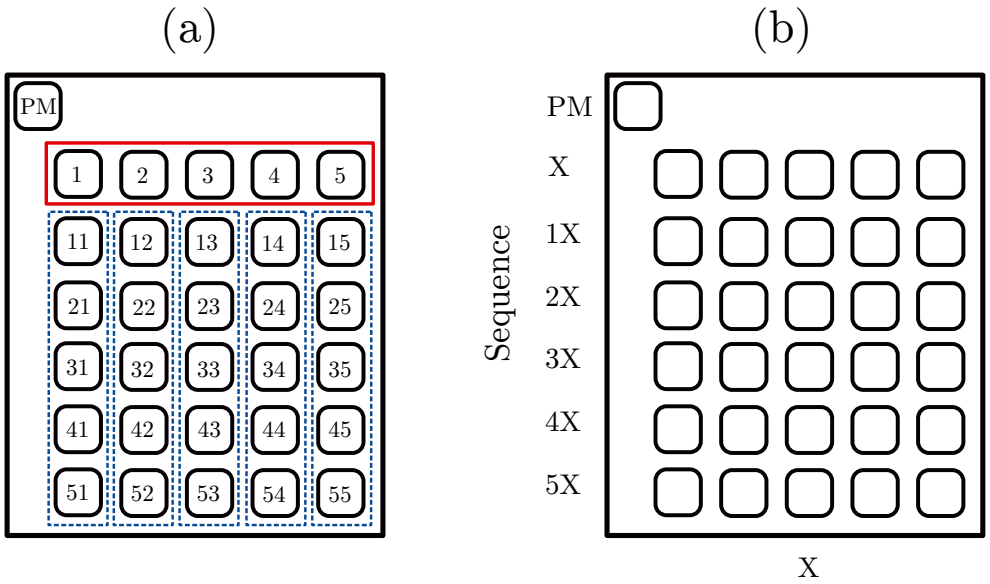

Figure 5.15. Conceptual design of the investigated sets. Each set includes a PM and 30 other sequences called x, $1x$, $2x$, $3x$, $4x$, $5x$, where $x = 1, \ldots, 5$. The first five sequences $x = 1, \ldots, 5$ shown in the red box have difference of d_1 to the PM. For each sequence x there are five sequences shown in the blue box that have Hamming distance of d_1 to sequence x and $d_2 = 2d_1$ to the PM. (a) shows all sequences including their position and names, while (b) illustrates the axes that we use in the following pictures to identify the sequences.

Each set includes the specific sequence 3'-ACTCTATCACGCCTGGT-5' as perfect match (PM) and 30 additional sequences. Fig. 5.15 shows a conceptual design of the investigated sets including the nomenclature chosen to label each sequence. The first five sequences $x = 1, 2, 3, 4, 5$ in the red box (a) have a defined Hamming distance d_1 to the PM sequence. Additionally, for each sequence x there are five further sequences $1x$, $2x$, $3x$, $4x$, $5x$ as shown at the blue dashed boxes, which have a Hamming distance d_1 with respect to sequence x and $d_2 = 2d_1$ to the PM. The Hamming distances between sequences of different blue boxes, on the other hand, are undefined. (b) depicts the axes that we use later in our results to identify the sequences. In particular, we study three sets, set 1 including sequences with $d_1 = 2$ and $d_2 = 4$, set 2 with $d_1 = 3$ and $d_2 = 6$ and set 3 with $d_1 = 4$ and $d_2 = 8$.

For having a better statistics over the intensities at different positions on the chip we have 5 replicates from each sequence. In the actual microarray the sequences and their replicates are not located as illustrated in Fig. 5.16, but randomly distributed. Every chip is divided to four millichips refereed to as MC1, MC2, MC3 and MC4, where each millichip carries the sequences of one set (compare Fig. 5.16). In all millichips we hybridized the Cy3 labeled

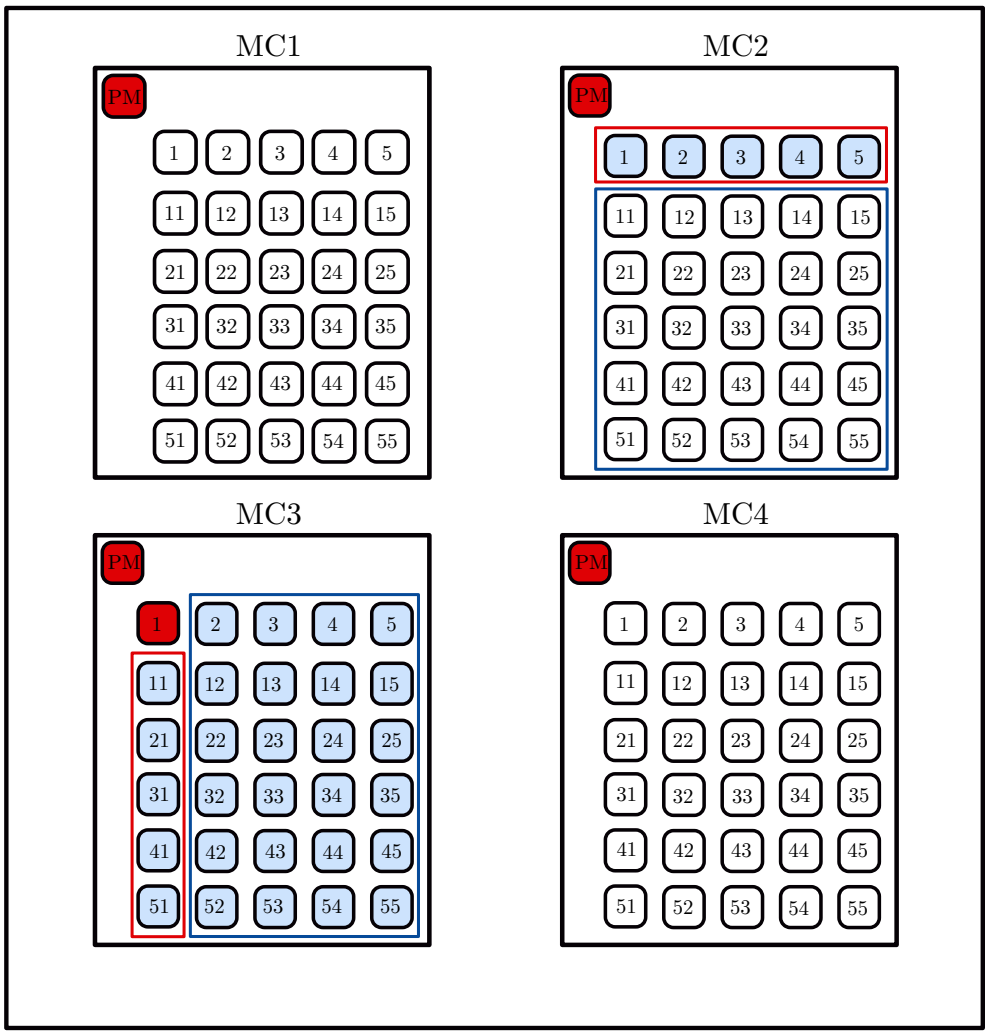

Figure 5.16. For each set we investigate a microarray with four millichips MC1, MC2, MC3 and MC4. All of them consist of the same probes. Each millichip is hybridized individually, independent of others. The hybridization solution for all millichips includes a Cy3 labeled PM target (shown in red). Additionally, in MC2 we added the unlabeled complementary protectors of the first 5 sequences $x = 1, \ldots, 5$ (blue colors). In MC3 we hybridized the Cy3 labeled complementary of sequence 1 besides PM (shown in red) and the unlabeled protectors of the rest (blue colors). The sequences in the red box in MC3 are less protected due to the presence of the labeled target 1 while the sequences in the blue box are less affected by it and more protected. MC4 is used only for some cross-checks.

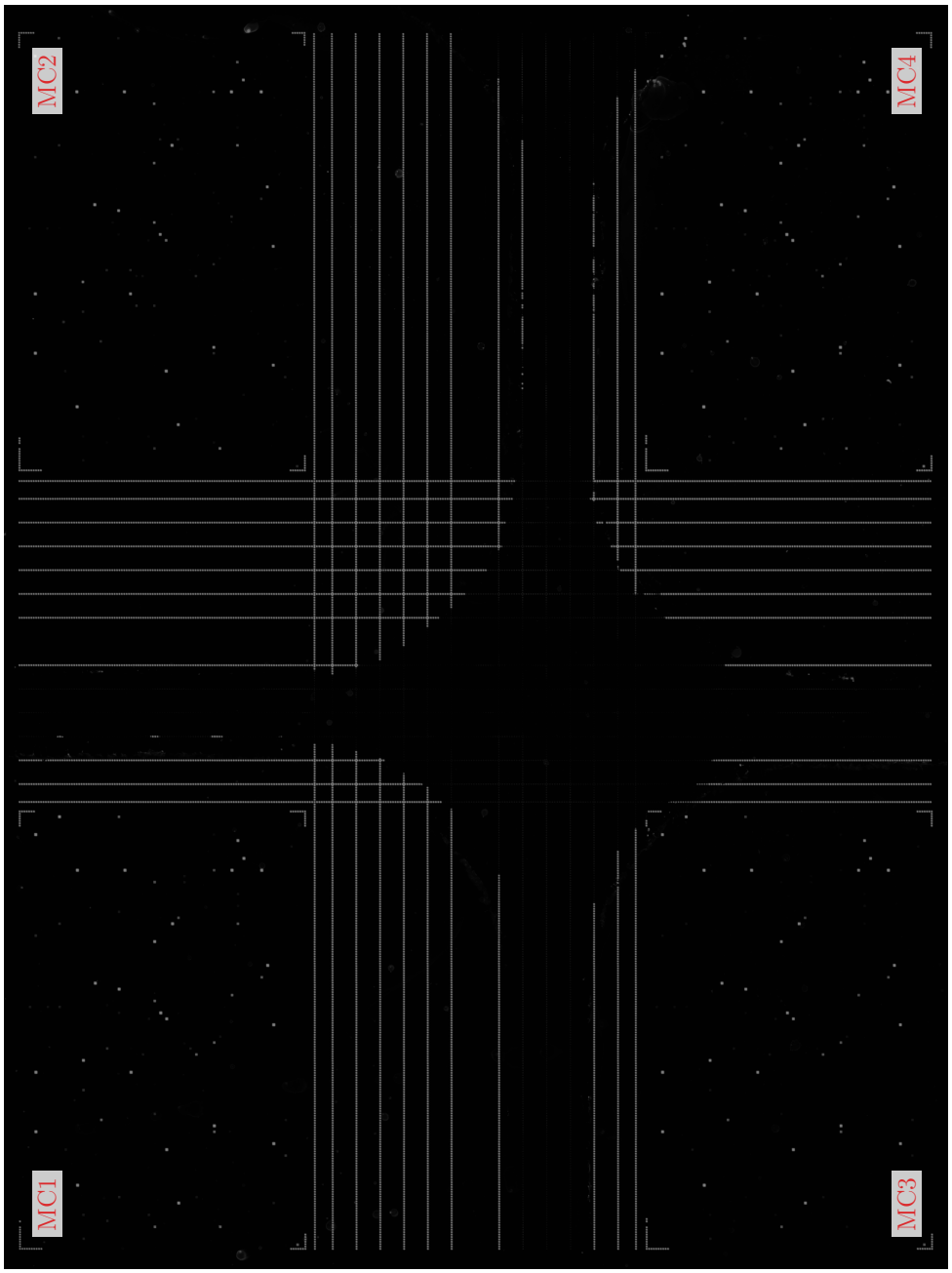

Figure 5.17. Fluorescence image taken with Genepix scanner from set 3 after hybridization. The design and the hybridization solution for all milichips MC1 to MC4 are according to the main text. The four millichips are separated by the PM fiducial. Within each MC the fluorescence intensities of the randomly distributed 5 replicates of all 31 probes can be seen.

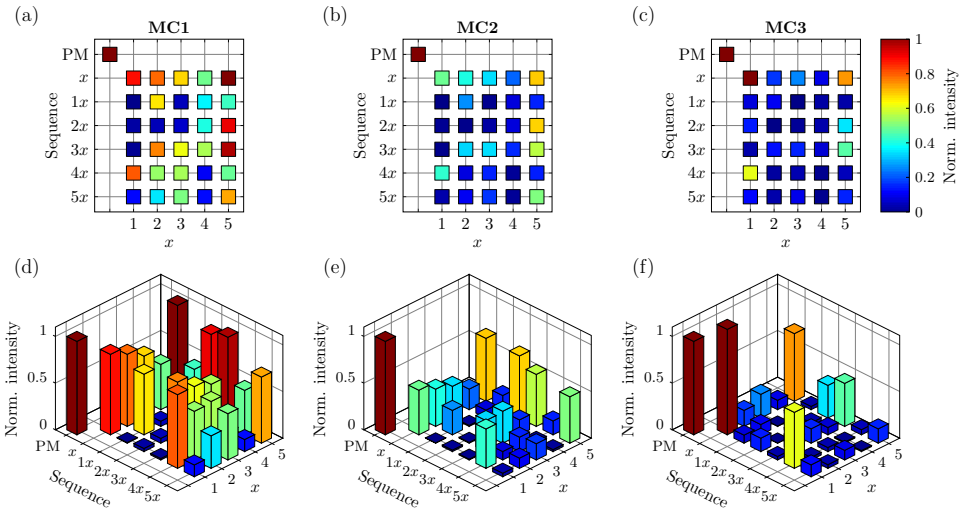

Figure 5.18. The computationally reconstructed chip for set 1 based on the design presented in Fig. 5.15 and fluorescence intensities extracted from the image taken with a scanner. For each sequence the intensity is averaged over 5 replicates and the background is subtracted. The fluorescence intensity of the hybridization is normalized subsequently to the PM intensity. The color of each feature in (a)-(c) as well as the height of the bars in (d)-(f) correspond to the normalized intensities. In all MC we hybridized the Cy3 labeled PM. In MC2 we additionally have the unlabeled protectors of the first five sequences ($x = 1, \ldots, 5$) and in MC3 the Cy3 labeled target 1 and the unlabeled protectors of the rest. The PM cross hybridizes to other surface-bounds probes allover the microarray. Adding the unlabeled protectors in MC2 and MC3 decreases the effect of cross-hybridization.

complementary of PM (shown in red). In MC2 we furthermore added the unlabeled complementaries (the protectors) of the first five sequences (colored with blue). MC3 is hybridized with Cy3 labeled complementaries of both PM and sequence 1, called target 1 (in red), as well as the unlabeled protectors of the remaining sequences (blue color). We used MC4 for various cross-checks which will not be subject of the following discussions. The list of the sequences for all sets are given in the Appx. D. The concentration of each target is chosen to be 5 nM. The labeled targets may bind to the microarray surface at RT unspecifically. This leads to a significant amount of background or false hybridization, which decreases the sensitivity of the measurement and complicates the analysis. We reduce this undesired binding by heating the microarray surface including a hybridization solution to 44 °C, letting it relax back to RT, and eventually hybridizing it for one hour at RT. Fig. 5.17 is a picture taken with a Genepix scanner which shows the result of the hybridization for the set 3.

To analyse the data for each sequence, we averaged the intensity over all 5 replicates, subtracted an average background intensity and normalized each intensity to the PM intensity. Fig. 5.18 (a)-(c) shows the computationally reconstructed results for MC1, MC2 and MC3 for set 1. In (d)-(f) the same results are illustrated with the help of a 3D bar plot. The color of each feature in (a)-(c) and the height of each column in (d)-(f) correspond to normalized intensities of each sequence according to the color bar depicted on the right hand side. For MC1 [(a) and (d)] hybridization of the labeled target PM leads to the cross-hybridization of some other sequences. These non-specific intensities vary depending on the exact sequence and the position of the mismatches. In agreement with our observation in Sec. 5.4 and with [89, 168], our results reveal that the sequences with terminal mismatches show stronger cross-hybridization as we expected. The results shown in Fig. 5.18 depict our first attempt to decrease the cross-hybridization observed for (b) and (e) for MC2 by adding the protectors of sequences $x = 1, \ldots, 5$. It strikes the eye that all cross-hybridization signals are significantly reduced. In particular, we realize that not only the fluorescence intensity of the probes complementary to the protectors are diminished, but also of the remaining probes of the set. The situation for MC3 [(c) and (f)] is more complicated to interpret: as we added the labeled targets of both PM and sequence 1, their corresponding complementary probes show high intensities. The remaining 29 probes are protected and therefore show less cross-hybridization compared to MC1. For the sequences $1x, \ldots, 5x$, however the signal increased compared to MC2, even though their protectors are now present. This is the consequence of these probes having only two mismatches with labeled target 1, which is present in solution, in contrast to four mismatches they have with the labeled target PM.

In order to quantify the change in the intensities after adding the unlabeled complementary protectors, we compute the improvement of MC2 over MC1 η_2^{S1} and MC3 over MC1 η_3^{S1}, where S1 refers to set 1, i.e. $d_1 = 2$ and $d_2 = 4$. We define the improvement as

$$\begin{aligned} \eta_2^{S1} &= 1 - I_2^{S1}/I_1^{S1} \quad \text{and} \\ \eta_3^{S1} &= 1 - I_3^{S1}/I_1^{S1}. \end{aligned} \tag{5.9}$$

Here, I_1^{S1}, I_2^{S1}, and I_3^{S1} denote the relative intensities of set 1 for MC1, MC2 and MC3 as shown in Fig. 5.18. The improvement as defined in Eq. (5.9) yields the drop of intensity compared to MC1. That means an improvement close to 1 indicates a good performance of the respective protector. All improvements obtained for set 1 are shown in Fig. 5.19. As can be seen most improvements exceed 50 % with numerous features even approaching 100 % in particular for MC3. The features illustrated as black or white belong to the sequences that were excluded from the plots and later analysis. Black indicates that the corresponding relative intensity on MC1 is below 5 %. For these sequences the improvement is subject to an large error and therefore un-

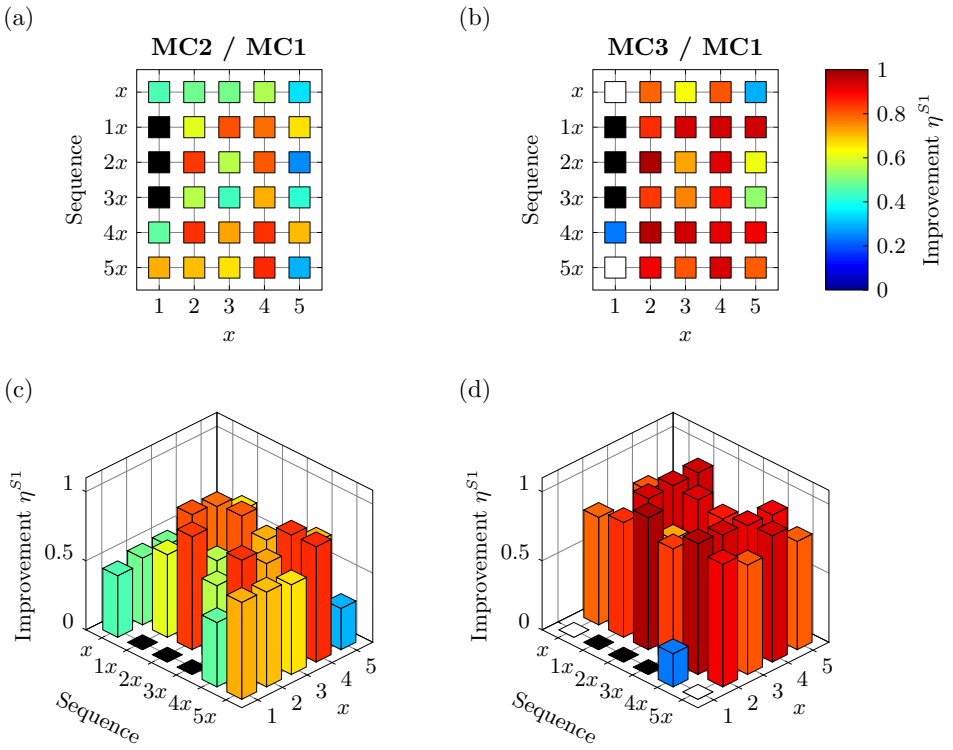

Figure 5.19. Improvements of cross-hybridization intensities (a) and (c) η_2^{S1} for MC2 as well as (b) and (d) η_3^{S1} for MC3 of set 1 according to Eq. (5.9). The color of each feature in (a) and (b) and the height of the bars in (c) and (d) show the drop of measured intensities for MC2 and MC3 with respect to MC1. The black features belong to the sequences, whose intensities in MC1 are lower than 5% of the PM intensity. The white features on the other hand, are the sequences whose intensities in MC3 are elevated compared to MC1, leading to $\eta_3^{S1} = 1 - \frac{I_3^{S1}}{I_1^{S1}} < 0$. The protectors help decreasing the cross-hybridization intensities of most of the sequences.

reliable. White features correspond to negative improvements. These occur only for MC3 and sequences 11, 21, ..., 51 due to the presence of the labeled target 1.

To obtain a trend for the impact of the protectors, we average η_2^{S1} and η_3^{S1} over certain groups of sequences. On that account, we separate MC2 into two groups: (i) the directly protected sequences, whose protectors are in solution (Fig. 5.19 (a) and (c) first row) and (ii) the indirectly protected sequences, whose protectors are not present, but they still profit from the protectors of the first group (row 2 to 6). For the directly protected sequences we find an average improvement of $(46 \pm 7)\,\%$ while the indirectly protected ones are even improved by $(64 \pm 17)\,\%$. It might come as a surprise that the

latter group achieves the better improvement although it is only indirectly protected. One should keep in mind, however, that cross-hybridization within this group stems from binding to the PM, which has four mismatches with respect to them, and is therefore comparatively weak. As protectors have less mismatches compared to the PM, they can remove the PM with ease.

The reduction of the cross-hybridization intensity for MC3 is illustrated in Fig. 5.19 (b) and (d). The additional presence of the labeled target 1 increases the cross-hybridization signals of sequence 51, i.e. $\eta_3^{S1} < 0$. This sequence is accordingly shown in white and also excluded from the averaging. Here for MC3 we average over the improvement of all sequences apart from $x = 1$ (first column) since they are directly influenced by the labeled target 1. We obtain $(80 \pm 16)\,\%$, which is mainly limited due to sequences from the last column (5, 25 and 35). The low improvements for these specific sequences is stemming from the comparatively strong binding affinities resulting from their mismatches being terminally positioned. Generally, the high average improvement clearly demonstrates that cross-hybridization signals can be efficiently minimized with the help of protectors.

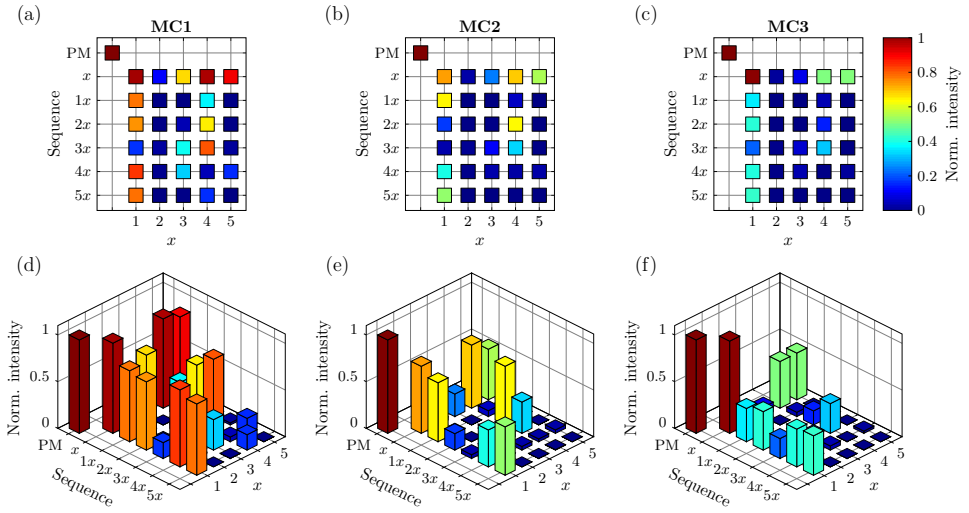

Figure 5.20. The computationally reconstructed chip for set 2 based on the design presented in Fig. 5.15 and the fluorescence intensities extracted from the images taken with the scanner. The PM cross hybridizes to other surface-bounds probes allover the microarray. Adding the unlabeled protectors in MC2 and MC3 decreases the effect of cross-hybridization. However, the cross-hybridization compared to set 1 (Fig. 5.18) is less because of the large Hamming distance ($d_1 = 3$). For more details cf. caption Fig. 5.18 and main text.

In the following we present the hybridization results for set 2 and set 3 similar to what we described above for set 1. All normalized hybridization intensities are shown in Fig. 5.20 and Fig. 5.21 for set 2 and 3, respectively.

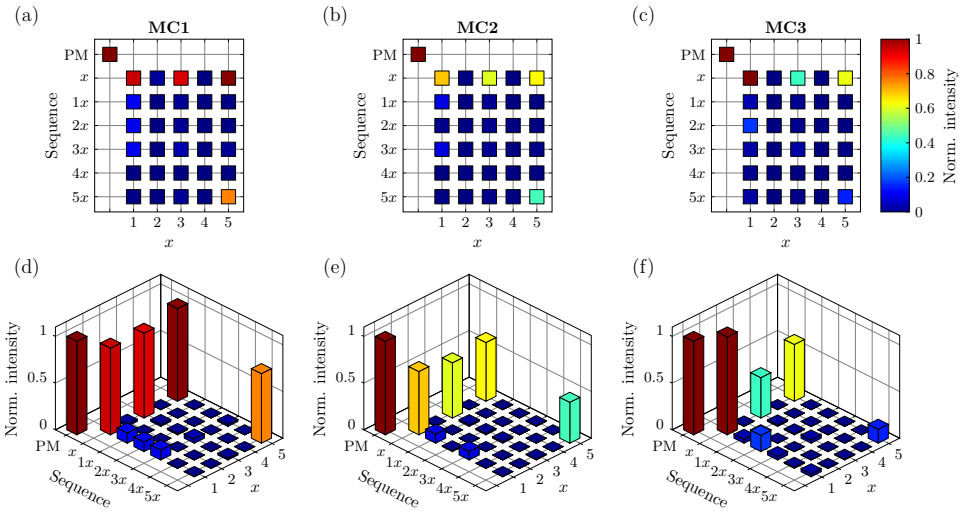

Figure 5.21. The computationally reconstructed chip for set 3 based on the design presented in Fig. 5.15 and the fluorescence intensities extracted from the images taken with the scanner (Fig. 5.17). The PM cross hybridizes to other surface-bounds probes allover the microarray. Adding the unlabeled protectors in MC2 and MC3 decreases the effect of cross-hybridization. However, the cross-hybridization compared to set 1 (Fig. 5.18) is less because of the large Hamming distance ($d_1 = 4$). For more details cf. caption Fig. 5.18 and main text.

For both sets we observe cross-hybridization to the PM in MC1 and MC2 as well as to PM and target 1 in MC3. The amount of cross-hybridization is significantly decreased compared to set 1 since the respective Hamming distances here are larger ($d_1 = 3$ and $d_2 = 6$ for set 2 and $d_1 = 4$ and $d_2 = 8$ for set 3). Also, we see the expected decrease in the cross-hybridization signal in presence of protectors when comparing MC2 to MC1 as well as MC3 to MC1 for both sets. To quantify the effect we determine all improvements η_2^{S2} and η_2^{S3} for MC2 over MC1 as well as η_3^{S2} and η_3^{S3} for MC3 over MC1 with the help of equation Eq. (5.9). The results are illustrated in Fig. 5.22 and Fig. 5.23, respectively. At a first glance, we recognize that we still have a high average level of protection, but it continuously decreases from set 1 to set 3. Moreover, it can be seen that an increasing number of sequences must be excluded from the analysis as their corresponding cross-hybridization intensities were already below our threshold of 5 % for MC1. We furthermore evaluate the average improvements for the directly and indirectly protected sequences of MC2 as well as all sequences of MC3 apart from $x = 1$ in the same fashion as presented for set 1. The results are summarized in Tab. 5.4. Note that the presented errors are not the error of each individual intensity but resulting from the scattering of the intensities around the average value in one group . The averages confirm the trend we suspected from the respective

Table 5.4. Comparison of improvements obtained for MC2 and MC3 for all three sets. The sequences of MC2 are grouped into directly and indirectly protected sequences (cf. main text). All values are given in percent.

	MC2		MC3
Set	directly protected	indirectly protected	improved sequences
S1	46 ± 7	64 ± 17	80 ± 16
S2	46 ± 21	59 ± 29	77 ± 17
S3	33 ± 5	40 ± 37	57 ± 21

plots: from set 1 to set 3 the improvement decreases from 46 % to 33 % for the directly protected sequences, from 64 % to 40 % for the indirectly protected ones, and from 80 % to 57 % for improved sequences (all sequences apart from the first row $x = 1$) on MC3. This observation can be understood keeping in mind that the cross-hybridization for high Hamming distances is anyhow very low. Therefore, there is anyway not much room for the protectors to decrease the cross-hybridization signal.

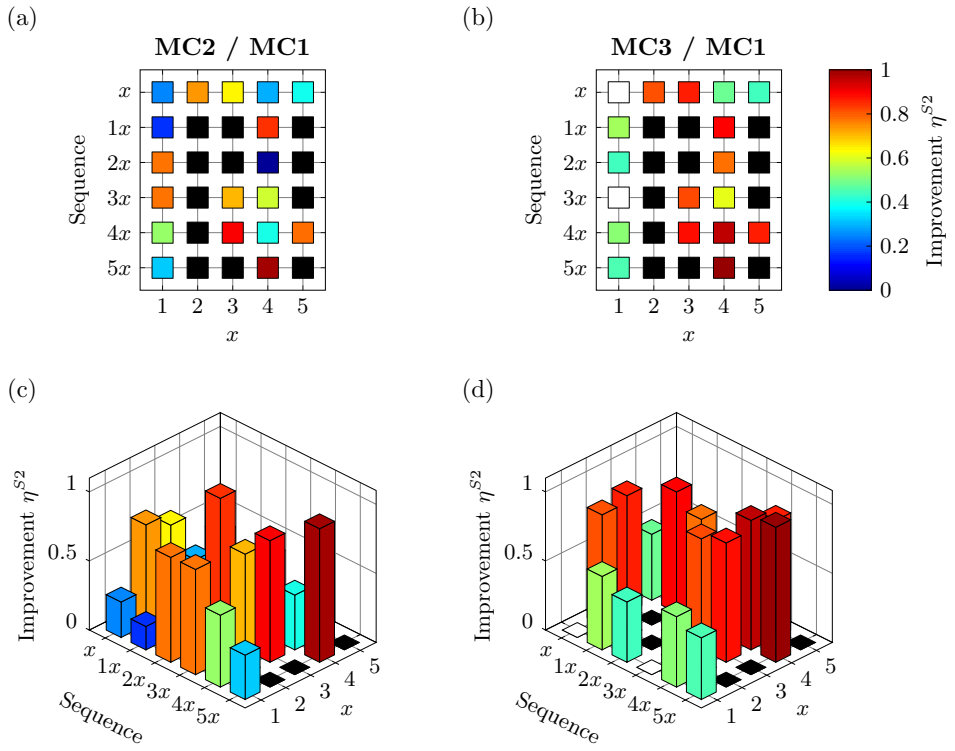

Figure 5.22. Improvements of cross-hybridization intensities (a) and (c) η_2^{S2} for MC2 as well as (b) and (d) η_3^{S2} for MC3 of set 2 according to Eq. (5.9). For details cf. caption Fig. 5.19 and main text.

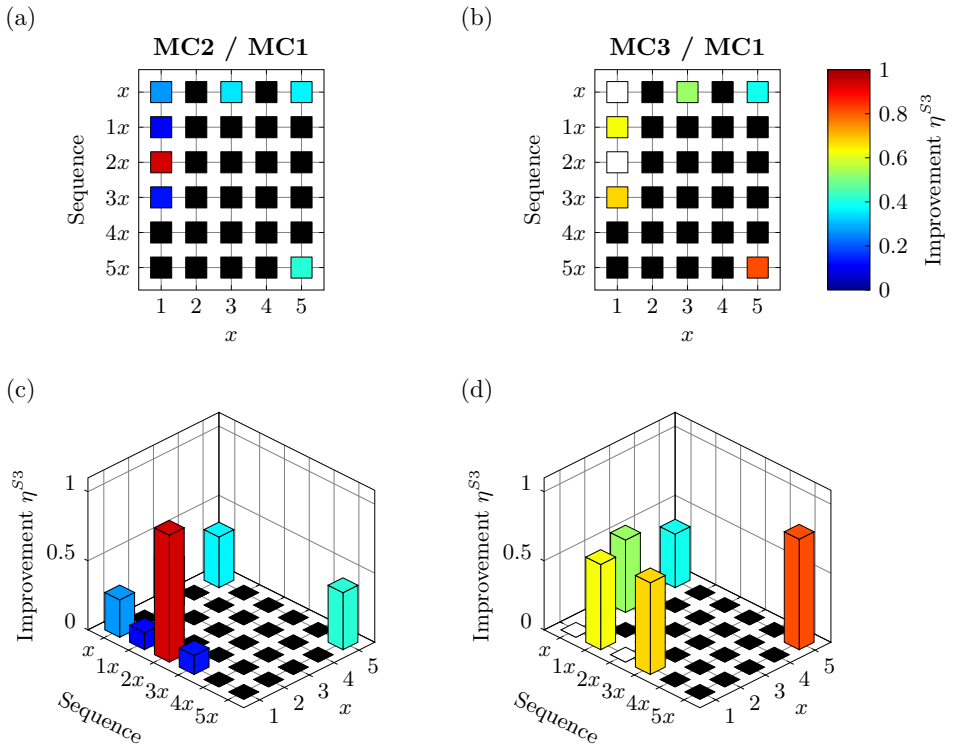

Figure 5.23. Improvements of cross-hybridization intensities (a) and (c) η_2^{S3} for MC2 as well as (b) and (d) η_3^{S3} for MC3 of set 3 according to Eq. (5.9). For details cf. caption Fig. 5.19 and main text.

5.7 Summary and Discussion

In the present chapter we first address the question by how much sequences of a given set must differ so that the hybridization to their fluorescently labeled complementary targets in solution does not lead to any cross-hybridization, i.e. single sequences are well discriminated. For that we performed experiments with DNA sequences that are synthesized on a surface employing light directed in situ synthesis. We first introduced the concept of Hamming distance borrowed from information theory and applied it to DNA sequences. We experimentally determined a necessary minimum Hamming distance for sequences with a length of $L = 7$. By employing a local search algorithm, we constructed independent sets of length $L \leq 7$ for different d_m whose sizes are maximized. We found that for $L = 7$ a minimum Hamming distance of $d_m = 5$ is necessary to avoid cross-hybridization and experimentally verified the independency of such a set. Due to a large d_m the size of an independent set in this case is only 23, which is much smaller than 4^7 possible sequences. The obtained d_m is, however, in a good agreement with the reported discrim-

ination level of $d_m \approx L/2$ in [45, 46]. One should keep in mind that we added four extra bases to each strand to enhance the hybridization intensity. However, since these bases are fixed the effective length of the sequence is still $L = 7$, which in combination with $d_m = 5$ leads to a small set size. For longer length, these additional bases are redundant and therefore, with the same d_m the longer lengths provide larger set sizes.

Although a discrimination of a single-nucleotide in bulk is successfully accomplished [23], the optically directed synthesis technology introduces errors in synthesized sequences [186–190]. However, a minimum Hamming distance d_m can be reduced by rising a temperature [23]. At a given concentration a discrimination increases close to a melting temperature. However, to achieve that it is necessary to select the sequences of a set in such a way that they have similar melting temperatures. Regardless of the help this strategy offers it introduces an extra restriction to the choice of the sequences and does not present a general solution for the problem.

In the second part of the chapter we introduced a strategy to reduce the necessary Hamming distance between the sequences of a set. Here, we focused more on how to avoid cross-hybridization between sequences of a set. We qualitatively show that adding unlabeled complementary strands of all surface bound probes can decrease the cross-hybridization stemming from non-specific binding of a labeled targets. Adding the unlabeled protectors we observed a decrease of up to 80 % in average for the cross-hybridization intensities. For some specific sequences, the cross-hybridization was even almost suppressed. However, the relative effect is stronger for smaller Hamming distance since for larger distances there is not much cross-hybridization that needs to be reduced. Also, we observed that for an increasing number of labeled targets or unlabeled protectors, the complexity of the competition situation rises, making the effect of protectors difficult to quantify. However, the overall outcome clearly illustrates that the employment of protector sequences decreases undesired cross-hybridization and accordingly paves the way for smaller necessary Hamming distances and possibly larger set sizes. In future experiments, protectors can be employed to demonstrate that they help to maximize the size of an independent sets.

Chapter 6

Summary and Conclusion

In the course of present thesis we developed several experimental techniques in order to investigate the physical mechanism of DNA molecular recognition in particular in competitive environments. We focused on establishing a better understanding of the underlying processes in the context of experiments with surface-bound DNA as well as DNA in solution. The first part of this work deals with determining binding affinities of two DNA strands in liquid solution in a pairwise manner as well as in competition. The second part focuses on finding a largest possible set of independent surface-attached probes which do not exhibit any cross-hybridization.

In the bulk experiments first we applied fluorescence anisotropy (FA) technique to determine a pairwise binding affinity of a probe and a target. Fluorescence anisotropy is based on rotational diffusion of a fluorescently labeled molecule. While a labeled single strand rotates fast, hybridization of a single strand to its complementary forming a double helix slows down the rotational diffusion of a helix and increases the corresponding anisotropy. A single mismatch at the middle of a strand destabilizes a formed helix and decreases a persistence length. This leads to an increase in rotational mobility of a molecule and a decrease in anisotropy value. Accordingly, the intermediate conformations where two strands are only weakly bound results in a small anisotropy and these states will still be counted as unbound. Therefore FA is only sensitive to the formation of the helicoidal conformation. We found that the pairwise binding affinities obtained from FA are in a good agreement with the corresponding predicted values from Nupack software package. The Nupack algorithm based on nearest neighbor model considers only the base stacking interactions which are only present in duplex structures.

Moreover, we employed another technique, fluorescence correlation spectroscopy (FCS) based on translational diffusion of a fluorescent molecule. If two DNA strand partially bind to each other, they still diffuse together as a complex. Such microstates appear in a FCS signal while very same conformations as we discussed above do not have a big contribution in FA signal. As

a result of that, FCS leads to larger binding constants than FA. We observed that the binding affinity obtained from FCS for strands with a mismatch at the center is more than two orders of magnitudes larger than the respective value from FA. We suggest that for the mismatched strands, binding microstates of high entropy contribute relatively more to the affinity than the well known nearest neighbor interactions that arise in a stiff double helix.

Considering 3 strands, two of them competing for binding to the same probe, we observed that the individual binding constants obtained from pairwise binding do not describe our results in competition. In this case the ratio of binding affinities changes with respect to pairwise binding, revealing non-negligible and non-trivial interaction. The binding energy of either probe-target pair is not only affected by the presence of the respective competitor, but for each competitor to a different degree, suggesting that DNA binding is cooperative.

In the second part of this study we employed in situ synthesis technique to synthesize short DNA sequences with a core length of $L = 7$ on a microarray surface. To increase the hybridization probability we extended all sequences by adding four fixed extra bases, two to each terminal. We distributed different numbers of mismatches along each sequence at different positions. By hybridizing a perfectly matching target of one of the probe, we obtained various fluorescence intensities indicating cross-hybridization between a labeled target and mismatched probes. Subsequently, we determined a minimum number of mismatches which two sequences of a same length must have in order not to show any cross-hybridization. In analogy to coding theory, we called this distance as a minimum Hamming distance d_m. Accordingly, a set containing sequences with mutual distance of at least d_m should be free of any cross-hybridization. Such a set is referred to as independent. Employing a local search algorithm, we created different sets of sequences for different L and d_m, whose sizes are maximized. These sets are called maximum independent sets MIS. Our set sizes in most of the cases are either equal or exceeding the reported set sizes from literature [46, 50]. We experimentally illustrated that for the length of $L = 7$ a set with $d_m = 5$ does not display any cross-hybridization. We found the size of such an independent set to be only 23, which is surprisingly small compare to the overall 4^7 possible sequences. However, this set size is in a good agreement with a reported discrimination level of $L \approx d/2$ [45, 46], considering the additional four terminal bases. Moreover, we derived an analytical expression for the number of the sequences with runs of at least four guanine. We illustrated that eliminating these sequences for short lengths of DNA does not influence the MIS size, but we expect that it plays a role for longer lengths.

Additionally we introduced an approach to decrease a minimum Hamming distance of a set. We illustrated that adding unlabeled complementary targets of the surface bound probes (protectors) will diminish the cross-hybridization intensity of a labeled target up to 80 % in average. We observed that by in-

creasing the number of labeled targets the complexity of the situation quickly increases, leading to difficulties in quantifying the impact of the protectors. However, our results clearly demonstrated that introducing unlabeled protectors decreases cross-hybridization and accordingly is a compelling option to reduce the minimum necessary Hamming distance, which would result in larger set sizes.

The presented work contributes to the understanding of how molecular recognition works in case of DNA hybridization. In surface experiments we found that we require a large minimum Hamming distance to avoid any cross-hybridization. The bulk experiments on the other hand illustrate that the binding of a target to its complementary probe in presence of a competitor is much more complex than what can be inferred from individual binding mechanisms. We suggest that the cooperative processes increase the accuracy of the recognition in high density environments. We expect the cooperativity to also play a role in other recognition processes inside a cell. Therefore, other biological systems need to be studied along these lines in order to obtain a comprehensive picture of molecular recognition.

Appendices

A Diffusion Measurements for Determining Binding Affinity

Fig. A.1, Fig. A.2 and Fig. A.3 represent the correlation measurements using FCS technique for determining the fraction of occupied probe q for different concentrations of c_{PM}, c_{MMM} and c_{MME} in P-PM, P-MMM and P-MME hybridizations, respectively. Fitting all measurements to two components diffusion model Eq. (2.27), we derive the molar fractions q that we used in Fig. 4.11.

Fig. A.4 illustrates the autocorrelation measurements for finding the molar fraction q in the experiment where PM and MMM strands compete to hybridize to the probe. Here, the PM is a labeled sequence and the concentration of the MMM strand is varied.

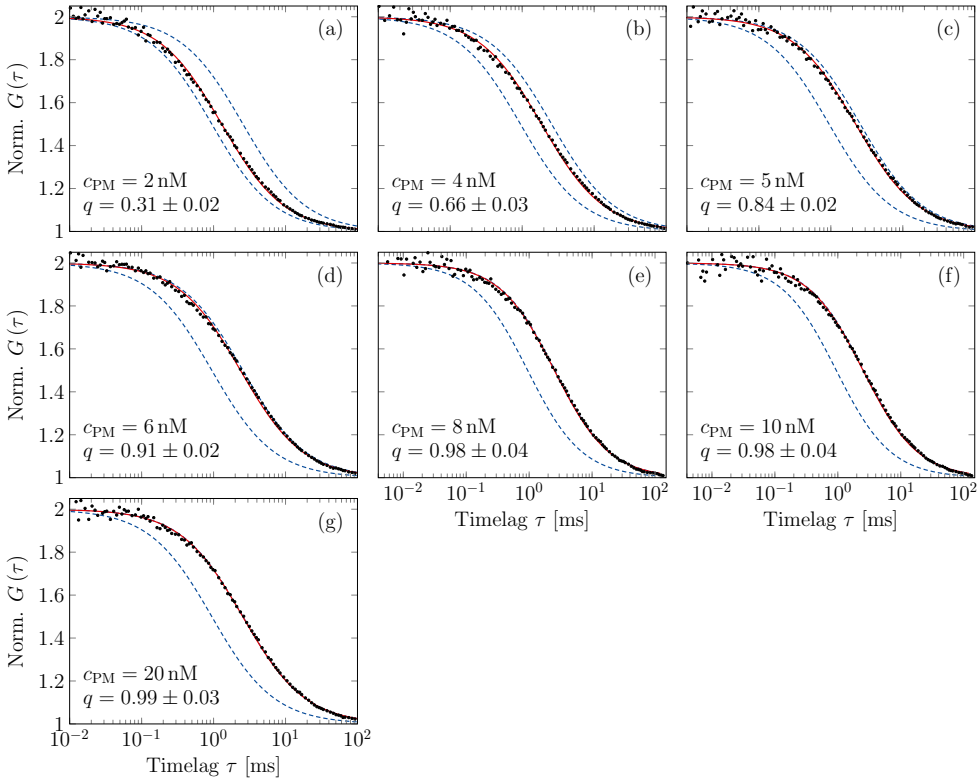

Figure A.1. Measured fluorescence autocorrelation for P-PM hybridization (data:circles) for different concentrations of PM. The blue curves correspond to the autocorrelation functions of a single labeled stranded P with $q = 0$, and double stranded with $q = 1$. The red solid curve represents the fit using Eq. (2.27) leading to a molar fractions q.

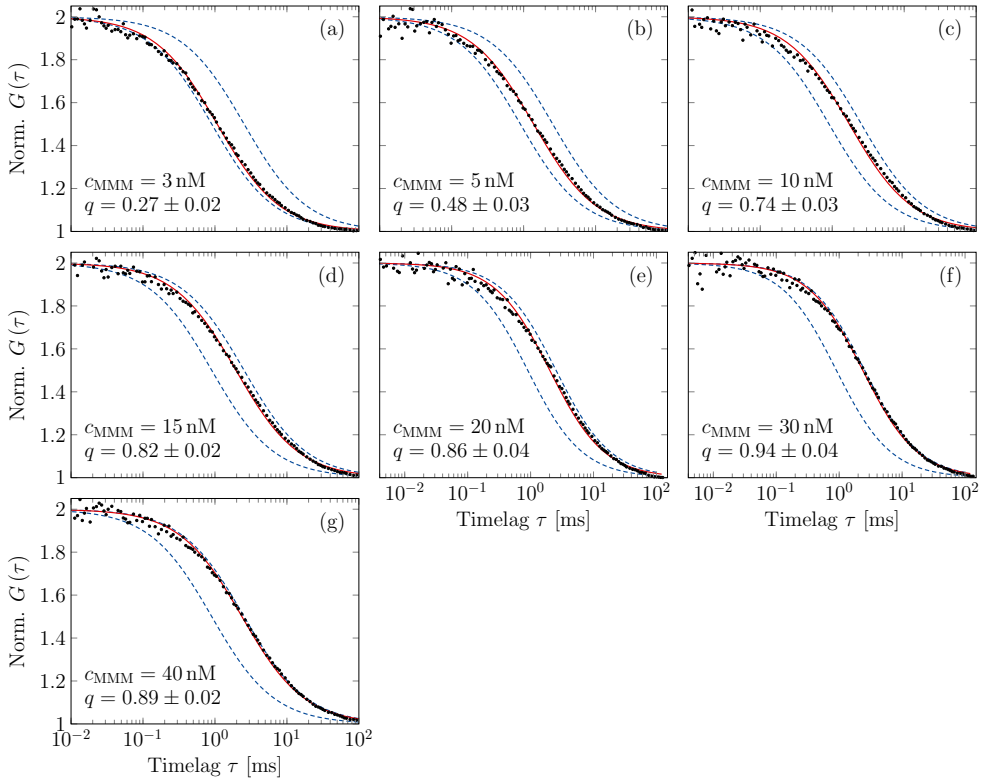

Figure A.2. Measured fluorescence autocorrelation for P-MMM hybridization (data:circles) for different concentrations of MMM. The blue curves correspond to the autocorrelation functions of a single labeled stranded P with $q = 0$, and double stranded with $q = 1$. The red solid curve represents the fit using Eq. (2.27) leading to a molar fractions q.

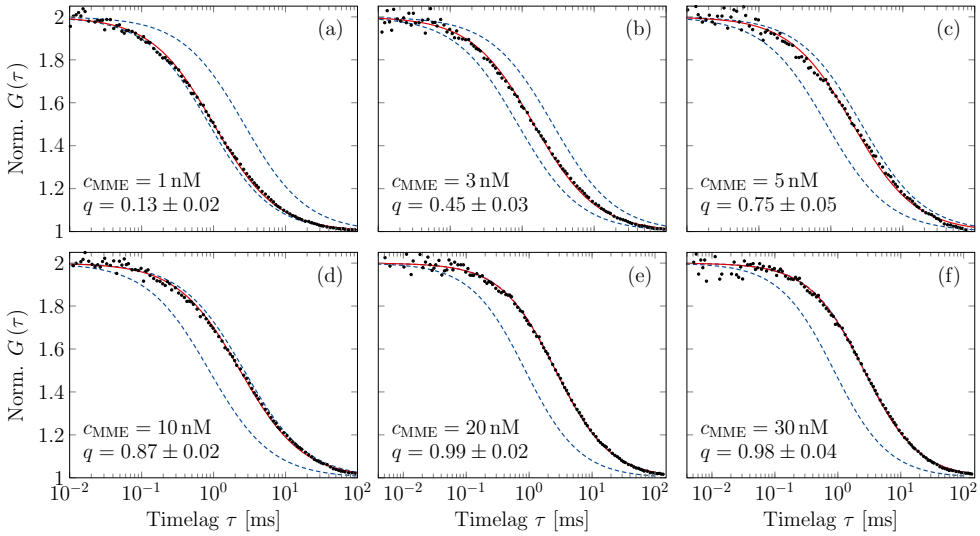

Figure A.3. Measured fluorescence autocorrelation for P-MME hybridization (data:circles) for different concentrations of MME. The blue curves correspond to the autocorrelation functions of a single labeled stranded P with $q = 0$, and double stranded with $q = 1$. The red solid curve represents the fit using Eq. (2.27) leading to a molar fractions q.

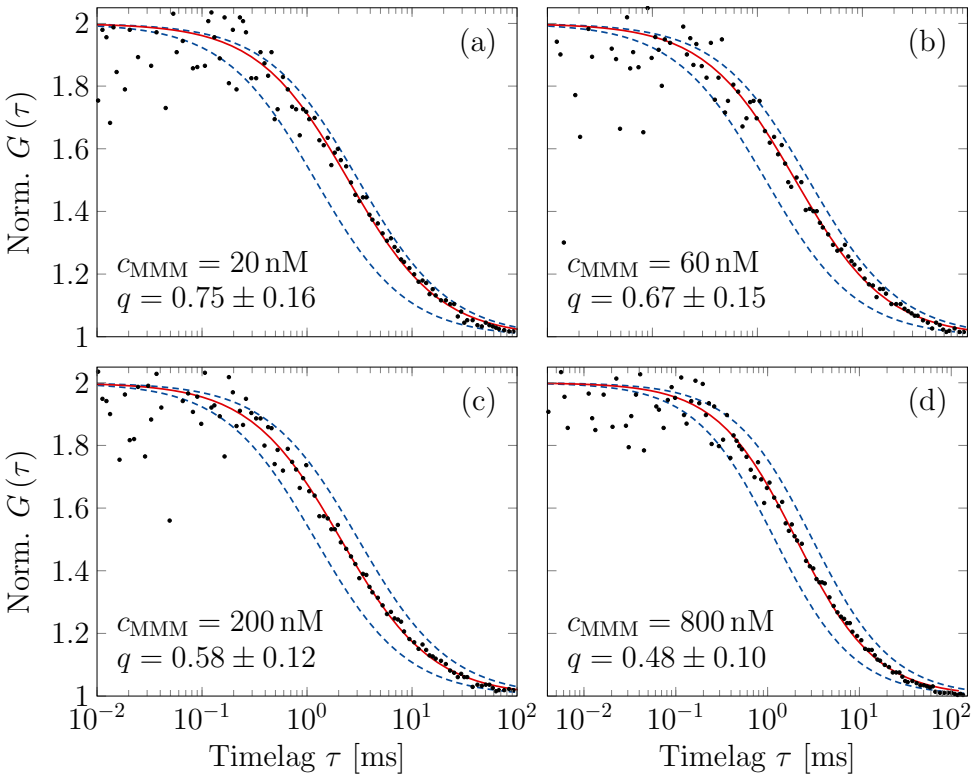

Figure A.4. Measured fluorescence autocorrelation for P-PM-MMM hybridization (data:circles) for different concentrations of MMM. The blue curves correspond to the autocorrelation functions of a single labeled stranded PM with q=0, and double stranded (P-PM) with q=1. The red solid curve represents the fit using Eq. (2.27) leading to a molar fractions q.

B List of Sequences for the First Set of Microarray Experiments

In this section at Tab. B.1 we show the list of sequences we used in Sec. 5.4 to determine the minimum Hamming distance d_m. The first sequence at Tab. B.1 is PM and the other sequences have one to seven mismatches with respect to it, which are marked in red. Tab. B.2 and Tab. B.3 represent the list of sequences with $d_m = 4$ and $d_m = 5$ as mentioned in Sec. 5.5, respectively. All sequences were hybridized to the surface using five thymine (T) bases as a linker.

Table B.1. Sequences used in Sec. 5.4 for determining minimum Hamming distance. The first sequence of a table is PM and the others have different numbers of mismatches compared to it, which are marked in red.

Sequences from 3' to 5'		
CTACCGTACTC	CTCTAGTACTC	CTGCCTACCTC
CTTCCGTACTC	CTACCACTCTC	CTGCTATCCTC
CTATCGTACTC	CTACCGACTTC	CTGCTCTAATC
CTACTGTACTC	CTCCTGAACTC	CTCTGACACTC
CTACCTTACTC	CTCCCATTCTC	CTAGTCACCTC
CTACCGTTCTC	CTCCCGCATTC	CTACGTACTTC
CTACCGTACTC	CTCCCGTCTTC	CTGCTCTCATC
CTCTCGTACTC	CTCCCATATTC	CTCTCGCTGTC
CTACTCTACTC	CTATCATGCTC	CTATGCATTTC
CTACCTCACTC	CTATCATAGTC	CTTGTCATCTC
CTACCGCTCTC	CTACTCAACTC	CTCTTGACATC
CTACCGTCTTC	CTACACTATTC	CTTCACAGTTC
CTCCTGTACTC	CTGTACTACTC	CTTGACCTATC
CTACCTTACTC	CTAGTACACTC	CTCGTCATATC
CTTCCGCACTC	CTACGTACCTC	CTGTTCACGTC
CTTCCGTCCTC	CTACCTGCATC	
CTCCCGTATTC	CTGCTGCAATC	

Table B.2. List of sequences with minimum Hamming distance $d_m = 4$.

Sequences from 3' to 5'	
CTAAAGCAATC	CTCCTTCGTTC
CTAAGGGTGTC	CTCGAAATCTC
CTAATCCTTTC	CTCGACTCGTC
CTACACAACTC	CTCGGCCATTC
CTACCACCATC	CTCGGTTTATC
CTACCTGGCTC	CTCGTGACTTC
CTACGGTCTTC	CTCGTTGAGTC
CTACTTTAATC	CTCTACGTTTC
CTAGAATATTC	CTCTAGCCCTC
CTAGATCGGTC	CTCTCGTGTTC
CTAGCGCTCTC	CTCTCTCAATC
CTAGGAACGTC	CTCTGAGCATC
CTATATACATC	CTCTGGAAGTC
CTATCCGAGTC	CTCTTACTGTC
CTCAACAGATC	CTGAATTGCTC
CTCACAAATTC	CTGACAGCGTC
CTCAGTACCTC	CTGACCCACTC
CTCATAGGCTC	CTGAGAATATC
CTCCCGATATC	CTGATGAGGTC
CTCCGATACTC	CTGCAAGAATC
CTGCCATTTTC	CTTCATCTCTC
CTGCGCGGGTC	CTTCCTTCGTC
CTGCGTAATTC	CTTCGGAGCTC
CTGCTCACATC	CTTCTCGATTC
CTGCTGGTCTC	CTTGAGGGATC
CTGGACCTATC	CTTGCACAGTC
CTGGATGCTTC	CTTGCCGCCTC
CTGGCGTCATC	CTTGCTAGTTC
CTGGGACGCTC	CTTGGAGTTTC
CTGGTCTGTTC	CTTGTCATGTC
CTGTAAAGTTC	CTTGTGTACTC
CTGTAGTTGTC	CTTGTTCCATC
CTGTCTATCTC	CTTTGATGGTC
CTGTGCTAATC	CTTTGCACTTC
CTGTGTCCGTC	CTTTGGCTATC
CTGTTATCCTC	CTTTGTGACTC
CTGTTGCATTC	CTTTTAAAATC
CTGTTTGGATC	CTTTTCCGCTC
CTTAAATCATC	CTTTTGGCGTC
CTTAATAAGTC	CTTTTTTTTTC
CTTACGCCTTC	CTTACTGTATC
	CTTAGCTTCTC

Table B.3. List of sequences with minimum Hamming distance $d_m = 5$.

Sequences from 3' to 5'	
CTGTGACATTC	CTAATTGACTC
CTGTTCACATC	CTACACCCTTC
CTTAAGAGTTC	CTACTGTGGTC
CTTCCAATGTC	CTAGCGAAATC
CTTCGCTAATC	CTAGGATCCTC
CTTGATCAGTC	CTCATACCGTC
CTTGTAGGATC	CTCCATGTATC
CTTTCCCGCTC	CTCGGCATTTC
CTTTGGGCGTC	CTCTAGTACTC
CTTTTTTTTTC	CTGAACTTGTC
CTGCGTAGCT	CTGGCTGCTTC
	CTGGTGCTCTC

C Microarray Raw Data Images

(a) (b)

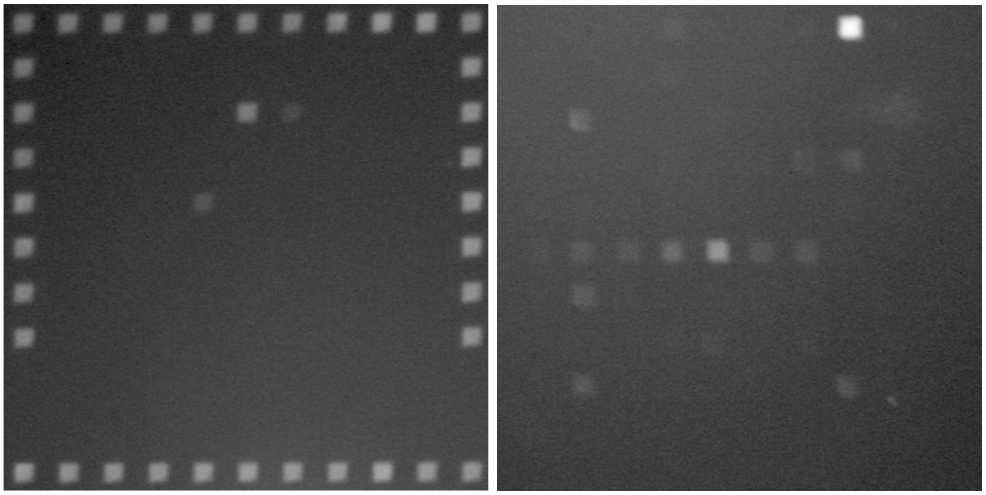

Figure C.5. The images taken with the camera of the microscope, showing the hybridization intensities of two sets of sequences with two different minimum Hamming distances. (a) an independent set including 23 sequences with $d_m = 5$ and (b) a set consists of 83 sequences with $d_m = 4$. For more details refer to Sec. 5.5

D List of Sequences for the Second Set of Microarray Experiments

In this section we present three lists of sequences that we used in Sec. 5.6, set1 with $d_1 = 2$ and $d_2 = 4$, set2 with $d_1 = 3$ and $d_2 = 6$ and set3 with $d_1 = 4$ and $d_2 = 8$.

Table D.4. List of sequences for set 1 with $d_1 = 2$ and $d_2 = 4$.

Name	Sequence from 3' to 5'
PM	ACTCTATCACGCCTGGT
1	ATTCTATCACGCCTAGT
2	ACTGTATCACGCCAGGT
3	ACTCTAGCACGTCTGGT
4	ACTCTATGACGTCTGGT
5	ACTCTATCACGCCTGCA
11	ATTGTATCAGGCCTAGT
21	ATTCTTTCACGGCTAGT
31	ATTCTATCATACCTAGT
41	ATGCTATCACGCCTATT
51	ATTCTGTCTCGCCTAGT
12	AGAGTATCACGCCAGGT
22	ACTGTGTCTCGCCAGGT
32	TCTGTATCACGCCAGGA
42	ACTGTATCACGCCACCT
52	TCTGTATCGCGCCAGGT
13	ACTGTAGCACGTCAGGT
23	ACTCTAGCACGTGAGGT
33	TCTCTAGCACGTCTGGC
43	ACTATGGCACGTCTGGT
53	ACTCTAGCACGTCTGCC
14	TCTCTGTGACGTCTGGT
24	GTTCTATGACGTCTGGT
34	TCTCTATGACGTCTGGC
44	ACTCTATGCTGTCTGGT
54	ACTCTATGACGTACGGT
15	ACTGTATCTCGCCTGCA
25	ACTCTATCACGCCGCCA
35	GTTCTATCACGCCTGCA
45	ACTCTATTTCGCCTGCA
55	TCTCTATCTCGCCTGCA

Table D.5. List of sequences for set 2 with $d_1 = 3$ and $d_2 = 6$.

Name	Sequence from 3' to 5'
PM	ACTCTATCACGCCTGGT
1	TGACTATCACGCCTGGT
2	ACTGTATCTCGCCTCGT
3	TCTCAATGACGCCTGGT
4	ACTCTATCACGCCTTCA
5	CCTCTATGACGCCTGGA
11	TGAGATTCACGCCTGGT
21	TGACTATCACGCCTAAC
31	TGACTATCGCGCATGAT
41	TGACTCTCACGCCTGTA
51	TGAACATCACGCCTGGA
12	TCTGTATGTCGCCTCGA
22	TCTGTATCTCACCTCGA
32	TTGGTATCTCGCCTCGT
42	ACTGTATCTCGCCACAA
52	ACGGTAACTCGACTCGT
13	TCTCACTGACTCCTCGT
23	TCTCAATGACGCCTAAG
33	TCTAATAGACGCCTGGT
43	TTGAAATGACGCCTGGT
53	TCTCAATGACGTCTGAA
14	TCGCTATCATGCCTTCA
24	ACTCTATCACGTGGTCA
34	TTGCTATCACGCCTTCA
44	TCTCTGTCACACCTTCA
54	CCTCTATGACGCTTTCA
15	CCTGTATGTCGCTTGGA
25	CCTGTATGACGTCTGTA
35	CCTTTATGACGCTTGAA
45	CGTCTATGACGCCTATA
55	CCTCAATGACAGCTGGA

Table D.6. List of sequences for set 3 with $d_1 = 4$ and $d_2 = 8$.

Name	Sequence from 3' to 5'
PM	ACTCTATCACGCCTGGT
1	TGCATATCACGCCTGGT
2	ACTCTGTAACGTATGGT
3	ACTCTATCACGCCACTG
4	ACTTTATAAGACCTGGT
5	TGTCTATCACGCCTGAC
11	TGCAACGTACGCCTGGT
21	TGCATATCACGCCACCA
31	TGCAACGTACGCCTGGT
41	TGCATAACTCGCATGGC
51	TGCATATTGCGCCTGAA
12	AGTCTGTATCGTACGGA
22	TGTCTGTAACGTATGAA
32	ACTAAGTAACGTATGTC
42	ATGTAGTAACGTATGGT
52	ACTCTGTAACGTAAACG
13	ACTCTATCAAAGTACTG
23	ACTCTATCATTGTACTG
33	TTGATATCACGCCACTG
43	TCTATACCAAGCCACTG
53	TCTGAATCACGCAACTG
14	GTTTTATAAGACCTGAC
24	TCTTTGTAAGACCAGCT
34	GAATCATAAGACCTGGT
44	ACTTTATAAGAAGGCGT
54	ACCTTCTAAGACGTAGT
15	TGTCTAGTACTTCTGAC
25	TGTCCATTTCGACTGAC
35	TGTCTATCACGGTACAC
45	TGTCGTTCAATCCTGAC
55	TGAGCTTCACGCCTGAC

References

[1] M. Lockett, H. Lange, B. Breiten, A. Heroux, W. Sherman, D. Rappoport, P. O. Yau, P. W. Snyder, and G. M. Whitesides, "The Binding of Benzoarylsulfonamide Ligands to Human Carbonic Anhydrase is Insensitive to Formal Fluorination of the Ligand," Angew. Chem. Int. Ed. **52**, 7714 (2013).

[2] B. Breiten, M. R. Lockett, W. Sherman, S. Fujita, M. Al-Sayah, H. Lange, C. M. Bowers, A. Heroux, G. Krilov, and G. M. Whitesides, "Water Networks Contribute to Enthalpy/Entropy Compensation in Protein–Ligand Binding," J. Am. Chem. Soc. **135**, 15579 (2013).

[3] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Chemical Foundations," in "Molecular Cell Biology," (W. H. Freeman, New York, 2000), chap. 2.

[4] S. Grimme, "Do Special noncovalent π-π stacking interactions really exist?" Angew. Chem. Int. Ed. **47**, 3430 (2008).

[5] I. Cosic, "Macromolecular bioactivity: is it resonant interaction between macromolecules?-theory and applications," TBME **41**, 1101 (1994).

[6] E. Fischer, "Einfluss der Configuration auf die Wirkung der Enzyme," Ber. Dtsch. Chem. Ges. **27**, 3479 (1894).

[7] D. E. Koshland, "Application of a Theory of Enzyme Specificity to Protein Synthesis," Proc. Natl. Acad. Sci. U.S.A **44**, 98 (1958).

[8] R. Penchovsky and R. R. Breaker, "Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes," Nat. Biotechnol. **23**, 1424 (2005).

[9] A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills, M. I. Blakey, and F. C. Simmel, "DNA Fuel for Free-Running Nanomachines," Phys. Rev. Lett. **90**, 118102 (2003).

[10] G. Seelig, D. Soloveichick, D. Y. Zhang, and E. Winfree, "Enzyme-Free Nucleic Acid Logic Circuits," Science **314**, 1585 (2006).

[11] B. M. Frezza, S. L. Cockroft, and M. R. Ghadiri, "Modular Multi-Level Circuits from Immobilized DNA-Based Logic Gates," J. Am. Chem. Soc. **129**, 14875 (2007).

[12] G. Seelig, B. Yurke, and E. Winfree, "DNA Hybridization Catalysts and Catalyst Circuits," in "DNA Computing," , C. Ferretti, G. Mauri, and C. Zandron, eds. (Springer, Berlin, Heidelberg, 2005), p. 329.

[13] J. S. Bois, S. Venkataraman, H. M. T. Choi, A. J. Spakowitz, Z. G. Wang, and N. A. Pierce, "Topological constraints in nucleic acid hybridization kinetics," Nucleic Acids Res. **33**, 4090 (2005).

[14] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce, "Programming biomolecular self-assembly pathways," Nature **451**, 318 (2008).

[15] J. Bath and A. J. Turberfield, "DNA nanomachines," Nat. Nanotechnol. **2**, 275 (2007).

[16] Y. Krishnan and F. C. Simmel, "Nucleic Acid Based Molecular Devices," Angew. Chem. Int. Ed. **50**, 3124 (2011).

[17] D. Y. Zhang and G. Seelig, "Dynamic DNA nanotechnology using strand-displacement reactions," Nat. Chem. **3**, 103 (2011).

[18] Y. J. Chen, B. Groves, R. A. Muscat, and G. Seelig, "DNA nanotechnology from the test tube to the cell," Nat. Nanotech. **10**, 748 (2015).

[19] S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor, and Y. Krishnan, "A DNA nanomachine that maps spatial and temporal pH changes inside living cells," Nat. Nanotech. **4**, 325 (2009).

[20] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," Science **266**, 1021 (1994).

[21] M. A. Heffler, R. D. Walters, and J. F. Kugel, "Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system," Biochem. Mol. Biol. Educ. **40**, 383 (2012).

[22] T. Wöhl, M. Brecht, F. Lottspeich, and H. Ammer, "The use of genomic DNA probes for in-gel hybridization," Nat. Chem. **16**, 739 (1995).

[23] D. Y. Zhang, S. X. Chen, and P. Yin, "Optimizing the specificity of nucleic acid hybridization," Nat. Chem. **4**, 208 (2012).

[24] A. Idili, F. Ricci, and A. Vallée-Bélisle, "Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves," Nucleic Acids Res. **45**, 7571 (2017).

[25] C. Crane-Robinson, A. Dragan, and C. M. Read, "Defining the thermodynamics of protein/DNA complexes and their components using

micro-calorimetry," in "DNA-Protein Interactions: Principles and Protocols, Third Edition," , B. Leblanc and T. Moss, eds. (Humana Press, Totowa, NJ, 2009), p. 625.

[26] J. B. Chaires, "Calorimetry and Thermodynamics in Drug Design," Annu. Rev. Biophys. **37**, 135 (2008).

[27] A. Cooper, "Microcalorimetry of Protein-Protein Interactions," in "Protein Targeting Protocols," , R. A. Clegg, ed. (Humana Press, Totowa, NJ, 1998), p. 11.

[28] J. L. Mergny and L. Lacroix, "Analysis of thermal melting curves," Oligonucleotides **13**, 515 (2003).

[29] L. A. Marky and K. J. Breslauer, "Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves," Biopolymers **26**, 1601 (1987).

[30] H. DeVoe and I. Tinoco, "The stability of helical polynucleotides: Base contributions," J. Mol. Biol. **4**, 500 (1962).

[31] P. N. Borer, B. Dengler, I. Tinoco, and O. C. Uhlenbeck, "Stability of ribonucleic acid double-stranded helices," J. Mol. Biol. **86**, 843 (1974).

[32] K. J. Breslauer, R. Frank, H. Blöcker, and L. A. Marky, "Predicting DNA duplex stability from the base sequence," Proc. Natl. Acad. Sci. U. S. A. **83**, 3746 (1986).

[33] J. SantaLucia and D. Hicks, "The thermodynamics of DNA structural motifs," Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).

[34] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).

[35] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," Nucleic Acids Res. **31**, 3406 (2003).

[36] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce, "NUPACK: Analysis and design of nucleic acid systems," J. Comput. Chem. **32**, 170 (2011).

[37] J. Bishop, A. M. Chagovetz, and S. Blair, "Kinetics of Multiplex Hybridization: Mechanisms and Implications," Biophys. J. **94**, 1726 (2008).

[38] J. Bishop, C. Wilson, A. M. Chagovetz, and S. Blair, "Competitive displacement of DNA during surface hybridization," Biophys. J. **92**, L10 (2007).

[39] A. Harrison, H. Binder, A. Buhot, C. J. Burden, E. Carlon, C. Gibas, L. J. Gamble, A. Halperin, J. Hooyberghs, D. P. Kreil, R. Levicky, P. A. Noble, A. Ott, B. M. Pettitt, D. Tautz, and A. E.

Pozhitkov, "Physico-chemical foundations underpinning microarray and nextgeneration sequencing experiments," Nucleic Acids Res. **41**, 2779 (2013).

[40] A. Halperin, A. Buhot, and E. B. Zhulina, "On the hybridization isotherms of DNA microarrays: the Langmuir model and its extensions," J. Phys. Condens. Matter **18**, S463 (2006).

[41] A. Halperin, A. Buhot, and E. B. Zhulina, "Hybridization isotherms of DNA microarrays and the quantification of mutation studies," Clin. Chem. **50**, 2254 (2004).

[42] A. Halperin, A. Buhot, and E. B. Zhulina, "Sensitivity, specificity, and the hybridization isotherms of DNA chips," Biophys. J. **86**, 718 (2004).

[43] G. A. Held, G. Grinstein, and Y. Tu, "Relationship between gene expression and observed intensities in DNA microarrays," Nucleic Acids Res. **34**, e70 (2006).

[44] G. Bhanot, Y. Louzoun, J. Zhu, and C. DeLisi, "The importance of thermodynamic equilibrium for high throughput gene expression arrays," Biophys. J. **84**, 124 (2003).

[45] A. G. Frutos, Q. Liu, A. J. Thiel, A. M. Sanner, A. E. Condon, L. M. Smith, and R. M. Corn, "Demonstration of a word design strategy for DNA computing on surfaces," Nucleic Acids Res. **25**, 4748 (1997).

[46] M. Li, H. J. Lee, A. E. Condon, and R. M. Corn, "DNA Word Design Strategy for Creating Sets of Non-interacting Oligonucleotides for DNA Microarrays," Langmuir. **18**, 805 (2002).

[47] W. Liu, S. Wang, L. Gao, F. Zahng, and J. Xu, "DNA Sequence Design Based on Template Strategy," J. Chem. Inf. Comput. Sci. **43**, 2014 (2003).

[48] O. D. King, "Bounds for DNA codes with constant GC-content," Electron. J. Comb. **10**, 33 (2003).

[49] R. Montemanni and D. H. Smith, "Construction of Constant GC-Content DNA Codes via a Variable Neighbourhood Search Algorithm," J. Math. Model. Algorithms **7**, 311 (2008).

[50] D. C. Tulpan, H. H. Hoos, and A. E. Condon, "Stochastic Local Search Algorithms for DNA Word Design," in "DNA Computing," , M. Hagiya and A. Ohuchi, eds. (Springer, Berlin, Heidelberg, 2003), p. 229.

[51] H. Zhang, Q. Wu, and M. Y. Berezin, "Fluorescence anisotropy (polarization): from drug screening to precision medicine," Expert Opin. Drug Discov. **10**, 1145 (2015).

[52] A. Bertz, J. E. Ehlers, S. Wöhl-Bruhn, H. Bunjes, K. H. Gericke, and H. Menzel, "Mobility of green fluorescent protein in hydrogel-based drug-delivery systems studied by anisotropy and fluorescence recovery after photobleaching," Macromol. Biosci. **13**, 215 (2012).

[53] J. C. Owicki, "Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer," J. Biomol. Screen. **5**, 297 (2000).

[54] D. S. Smith and S. A. Eremin, "Fluorescence polarization immunoassays and related methods for simple, high-throughput screening of small molecules," Anal. Bioanal. Chem. **391**, 1499 (2008).

[55] H. Zhang, S. Yang, K. D. Ruyck, N. V. Beloglazova, S. A. Eremin, S. D. Saeger, S. Zhang, J. Shen, and Z. Wang, "Fluorescence polarization assays for chemical contaminants in food and environmental analyses," Trends Anal. Chem. **114**, 293 (2019).

[56] H. Huang, H. Wei, M. Zou, X. Xu, B. Xia, F. Liu, and N. Li, "Modulating fluorescence anisotropy of terminally labeled double-stranded DNA via the interaction between dye and nucleotides for rational design of DNA recognition based applications," Anal. Chem. **87**, 2748 (2015).

[57] T. D. Pollard, "A Guide to Simple and Informative Binding Assays," Mol. Biol. Cell **21**, 4061 (2010).

[58] D. M. Jameson and J. A. Ross, "Fluorescence Polarization/Anisotropy in Diagnostics and Imaging," Chem. Rev. **110**, 2685 (2010).

[59] A. M. Rossi and C. W. Taylor, "Analysis of protein-ligand interactions by fluorescence polarization," Nat Protoc. **6**, 365 (2011).

[60] V. J. LiCata and A. J. Wowor, "Applications of fluorescence anisotropy to the study of protein-DNA interactions," Methods Cell Biol. **84**, 243 (2008).

[61] T. Deng, J. Li, J. H. Jiang, G. L. Shen, and R. Q. Yu, "A Sensitive Fluorescence Anisotropy Method for Point Mutation Detection by Using Core–Shell Fluorescent Nanoparticles and High-Fidelity DNA Ligase," Chemistry. **13**, 7725 (2007).

[62] D. Grünwald, M. C. Cardoso, H. Leonhardt, and V. Buschmann, "Diffusion and Binding Properties Investigated by Fluorescence Correlation Spectroscopy (FCS)," Curr. Pharm. Biotechnol. **6**, 381 (2005).

[63] A. Michelman-Ribeiro, D. Mazza, T. Rosales, T. J. Stasevich, H. Boukari, V. Rishi, C. Vinson, J. R. Knutson, and J. G. McNally, "Direct Measurement of Association and Dissociation Rates of DNA Binding in Live Cells by Fluorescence Correlation Spectroscopy," Biophys. J. **97**, 337 (2009).

[64] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," Nature. **171**, 737 (1953).

[65] R. M. Litman, "Rosalind Franklin and DNA," BioScience **26**, 572 (1976).

[66] M. Kohandel and B. Y. Ha, "Thermal denaturation of double-stranded DNA: Effect of base stacking," Phys. Rev. E **73**, 011905 (2006).

[67] B. Tinland, A. Pluen, J. Sturm, and G. Weill, "Persistence Length of Single-Stranded DNA," Macromolecules **30**, 5763 (1997).

[68] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy," Biophys. J. **86**, 2530 (2004).

[69] P. J. Hagerman, "Flexibility of DNA," Ann. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).

[70] E. T. Kool, "Hydrogen bonding, base stacking, and steric effects in DNA replication," Annu. Rev. Biophys. Biomol. Struct. **30**, 1 (2001).

[71] V. I. Danilov and I. S. Tolokh, "Nature of the stacking of nucleic acid bases in water: a Monte Carlo simulation," J. Biomol. Struct. Dyn. **2**, 119 (1984).

[72] P. Cieplak and P. A. Kollman, "Calculation of the free energy of association of nucleic acid bases in vacuo and water solution," J. Am. Chem. Soc. **110**, 3734 (2001).

[73] L. X. Dang and P. A. Kollman, "Molecular dynamics simulations study of the free energy of association of 9-methyladenine and 1-methylthymine bases in water," J. Am. Chem. Soc. **112**, 503 (1990).

[74] C. A. Hunter, "Sequence-dependent DNA structure. The role of base stacking interactions," J. Mol. Biol. **230**, 1025 (1993).

[75] R. Luo, H. S. Gilson, M. J. Potter, and M. K. Gilson, "The physical basis of nucleic acid base stacking in water," Biophys. J. **80**, 140 (2001).

[76] J. Šponer, P. Jurečka, and P. Hobza, "Accurate interaction energies of hydrogen-bonded nucleic acid base pairs," J. Am. Chem. Soc. **126**, 10142 (2004).

[77] M. E. Burkard, R. Kierzek, and D. H. Turner, "Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs," J. Mol. Biol. **290**, 967 (1999).

[78] S. Bommarito, N. Peyret, and J. SantaLucia, "Thermodynamic parameters for DNA sequences with dangling ends. Nucleic Acids," Nucleic Acids Res. **28**, 1929 (2000).

[79] M. Petersheim and D. H. Turner, "Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp," Biochemistry **22**, 256 (1983).

[80] S. M. Freier, D. Alkem, A. Sinclair, T. Neilson, and D. H. Turner, "Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes," Biochemistry **24**, 4533 (1985).

[81] 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," Proc. Natl. Acad. Sci. U.S.A. **83**, 9373 (1986).

[82] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix," Nucleic Acids Res. **34**, 564 (2006).

[83] 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 selfcomplementary d(CGTGAATTCGCG) duplex in solution," Biochemistry **21**, 437 (1982).

[84] 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**, 2509 (1981).

[85] T. Naiser, J. Kayser, T. Mai, W. Michel, and A. Ott, "Stability of a surface-Bound oligonucleotide duplex inferred from molecular dynamics: A study of single nucleotide defects using DNA microarrays," Phys. Rev. Lett. **102**, 218301 (2009).

[86] R. Blossey and E. Carlon, "Reparametrizing the loop entropy weights: Effect on DNA melting curves," Phys. Rev. E. **68**, 061911 (2003).

[87] C. Schildkraut and S. Lifson, "Dependence of the melting temperature of DNA on salt concentration," Biopolymers **3**, 195 (1965).

[88] J. Applequist and V. Damle, "Thermodynamics of helix-coil equilibrium in oligoadenylic acid from hypochromicity studies," J. Am. Chem. Soc. **87**, 1450 (1965).

[89] T. Naiser, J. Kayser, T. Mai, W. Michel, and A. Ott, "Position dependent mismatch discrimination on DNA microarrays – experiments and model," BMC Bioinform. **9**, 509 (2008).

[90] I. Langmuir, "The constitution and fundamentel properties of solids and liquids, part I. solids," J. Am. Chem. Soc. **38**, 2221 (1916).

[91] L. S. Yilmaz and D. R. Noguera, "Mechanistic Approach to the Problem of Hybridization Efficiency in Fluorescent In Situ Hybridization," Appl. Environ. Microbiol. **70**, 7126 (2004).

[92] T. Naiser, "Characterization of Oligonucleotide Microarray Hybridization," Ph.D. thesis, Universität Bayreuth (2008).

[93] 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," Biophys. J. **92**, 999 (2007).

[94] D. M. Crother and B. H. Zimm, "Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy," J. Mol. Biol. **9**, 1 (1964).

[95] 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**, 1033 (1981).

[96] H. T. Allawi and J. J. SantaLucia, "Thermodynamics and NMR of internal G.T mismatches in DNA," Biochemistry **36**, 10581 (1997).

[97] D. M. Gray, "Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA · RNA hybrids and DNA duplexes," Biopolymers **42**, 795 (1997).

[98] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules* (W H Freeman and Co Ltd, Oxford, 1980).

[99] H. T. Hatim and J. SantaLucia, "Thermodynamics and NMR of Internal G · T Mismatches in DNA," Biochemistry **36**, 10581 (1997).

[100] L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, and J. B. Reece, *Campbell Biology* (Pearson, Oakland, 2019), 11th ed.

[101] D. Gresham, M. J. Dunham, and D. Botstein, "Comparing whole genomes using DNA microarrays," Nat. Rev. Genet. **9**, 291 (2008).

[102] C. C. Chou, C. H. Chen, T. T. Lee, and K. Peck, "Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression," Nucleic Acids Res. **32**, e99 (2004).

[103] 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 Res. **12**, 1749 (2002).

[104] P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, r. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisén, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics," Science **353**, 78 (2016).

[105] D. Gresham, D. M. Ruderfer, S. C. Pratt, J. Schacherer, M. J. Dunham, D. Botstein, and L. Kruglyak, "Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray," Science **311**, 1932 (2006).

[106] V. Trevino, F. Falciani, and H. A. Barrera-Saldaña, "DNA microarrays: a powerful genomic tool for biomedical and clinical research," Mol. Med. **13**, 527 (2007).

[107] D. J. Wang, J. B. Fan, C. J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M. S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. J. Hudson, R. Lipshutz, M. Chee, and E. S. Lander, "Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome," Science **280**, 1077 (1998).

[108] K. L. Gunderson, F. J. Steemers, G. Lee, L. G. Mendoza, and M. S. Chee, "A genome-wide scalable SNP genotyping assay using microarray technology," Nat. Genet. **37**, 549 (2005).

[109] H. Matsuzaki, H. Loi, S. Dong, Y. Y. Tsai, J. Fang, J. Law, X. Di, W. M. Liu, G. Yang, G. Liu, J. Huang, G. C. Kennedy, T. B. Ryder, G. A. Marcus, P. S. Walsh, M. D. Shriver, J. M. Puck, K. W. Jones, and R. Mei, "Parallel Genotyping of Over 10000 SNPs Using a One-Primer Assay on a High-Density Oligonucleotide Array," Genome Res. **14**, 414 (2004).

[110] T. Pastinen, M. Raitio, K. Lindroos, P. Tainola, L. Peltonen, and A. C. Syvänen, "A System for Specific, High-throughput Genotyping by Allele-specific Primer Extension on Microarrays," Genome Res. **10**, 1031 (2000).

[111] H. Schulze, T. Barl, H. Vase, S. Baier, P. Thomas, G. Giraud, J. Crain, and T. T. Bachmann, "Enzymatic on-Chip Enhancement for High Resolution Genotyping DNA Microarrays," Anal. Chem. **84**, 5080 (2012).

[112] T. Pastinen, A. Kurg, A. Metspalu, L. Peltonen, and A. C. Syvänen, "Minisequencing: A Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays," Genome Res. **7**, 606 (1997).

[113] I. Pe'er, N. Arbili, and R. Shamir, "A computational method for resequencing long DNA targets by universal oligonucleotide arrays," Proc. Natl. Acad. Sci. **99**, 15492 (2002).

[114] N. Berthet, I. Leclercq, A. Dublineau, S. Shigematsu, A. M. Burguière, C. Filippone, A. Gessain, and J.-C. Manuguerra, "High-density resequencing DNA microarrays in public health emergencies," Nat. Biotechnol. **28**, 25 (2010).

[115] B. E. E. Cleven, M. Palka-Santini, J. Gielen, S. Meembor, M. Krönke, and O. Krut, "Identification and characterization of bacterial pathogens causing bloodstream infections by DNA microarray," J. Clin. Microbiol. **44**, 2389 (2006).

[116] S. F. Al-Khaldi, M. M. Mossoba, M. M. Allard, E. K. Lienau, and E. D. Brown, "Bacterial identification and subtyping using DNA microarray and DNA sequencing," in "Microbial Systems Biology: Methods and Protocols," , A. Navid, ed. (Humana Press, Totowa, NJ, 2012), p. 73.

[117] T. Naiser, T. Mai, W. Michel, and A. Ott, "A versatile maskless microscope projection photolithography system and its application in light-directed fabrication of DNA microarrays," Rev. Sci. Instrum. **77**, 063711 (2006).

[118] A. Hasan, K. P. Stengele, H. Giegrich, P. Cornwell, K. R. Isham, R. A. Sachleben, W. Pfleiderer, and R. S. Foote, "Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates," Tetrahedron **53**, 4247 (1997).

[119] M. Beier and J. D. Hoheisel, "Versatile derivatisation of solid support media for covalent bonding on DNA-microchips," Nucleic Acids Res. **27**, 1970 (1999).

[120] K. Hölz, J. K. Hoi, E. Schaudy, V. Somoza, J. Lietard, and M. M. Somoza, "High-Efficiency Reverse (5'→3') Synthesis of Complex DNA Microarrays," Sci. Rep. **8**, 15099 (2018).

[121] J. R. Lakowicz, "Fluorescence Anisotropy," in "Principles of Fluorescence Spectroscopy," (Springer, 2006), chap. 10, pp. 353–382.

[122] W. Eimer and R. Pecora, "Rotational and translational diffusion of short rodlike molecules in solution: Oligonucleotides," J. Chem. Phys. **94**, 2324 (1991).

[123] S. Deshayes and G. Divita, "Fluorescence technologies for monitoring interactions between biological molecules in vitro," Prog. Mol. Biol. Transl. Sci. **113**, 109 (2013).

[124] S. Weiss, "Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy," Nat. Struct. Biol. **7**, 724 (2000).

[125] J. C. Politz, E. S. Browne, D. E. Wolf, and T. Pederson, "Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells," Proc. Natl. Acad. Sci. **95**, 6043 (1998).

[126] L. Wawrezinieck, H. Rigneault, D. Marguet, and P. F. Lenne, "Fluorescence Correlation Spectroscopy Diffusion Laws to Probe the Submicron Cell Membrane Organization," Biophys. J. **89**, 4029 (2005).

[127] Z. Petrášek and P. Schwille, "Fluctuations as a source of information in fluorescence microscopy," J. Royal. Soc. Interface **6**, S15 (2008).

[128] A. Einstein, "Investigations on the Theory of the Brownian Movement," Ann. Phys. **7**, 281 (1956).

[129] S. T. Hess, S. Huang, A. A. Heikal, and W. W. Webb, "Biological and Chemical Applications of Fluorescence Correlation Spectroscopy: A Review," Biochemistry **41**, 697 (2002).

[130] H. Neuweiler, A. Schulz, M. Böhmer, J. Enderlein, and M. Sauer, "Measurement of Submicrosecond Intramolecular Contact Formation in Peptides at the Single-Molecule Level," J. Am. Chem. Soc. **125**, 5324 (2003).

[131] P. Kask, P. Piksarv, M. Pooga, U. Mets, and E. Lippmaa, "Measurement of Submicrosecond Intramolecular Contact Formation in Peptides at the Single-Molecule Level," Biophys. J. **55**, 213 (1989).

[132] A. Allikalt, S. Kopanchuk, and A. Rinken, "Implementation of fluorescence anisotropy-based assay for the characterization of ligand binding to dopamine D_1 receptors," Eur. J. Pharmacol. **839**, 40 (2018).

[133] D. Magde, E. L. Elson, and W. W. Webb, "Fluorescence correlation spectroscopy. II. An experimental realization," Biopolymers **13**, 29 (1974).

[134] L. Jermutus, R. Kolly, Z. Földes-Papp, J. Hanes, R. Rigler, and A. Plückthun, "Ligand binding of a ribosome-displayed protein detected in solution at the single molecule level by fluorescence correlation spectroscopy," Eur. Biophys. J. **31**, 179 (2002).

[135] M. Mohammadi-Kambs and A. Ott, "DNA oligomer binding in competition exhibits cooperativity," New J. Phys. **21**, 113027 (2019).

[136] C. Agbavwe, C. Kim, D. G. Hona, K. Heinrich, T. Wang, and M. M. Somoza, "Efficiency, Error and Yield in Light-Directed Maskless Synthesis of DNA Microarrays," J. Nanobiotechnology **9**, 57 (2011).

[137] K. Hölz, E. Schaudy, J. Lietard, and M. M. Somoza, "Multi-level patterning nucleic acid photolithography," Nat. Commun. **10**, 3805 (2019).

[138] Y. K. Cheng and J. L. Chern, "Irradiance formations in hollow straight light pipes with square and circular shapes," J. Opt. Soc. Am. **23**, 427 (2006).

[139] A. Offner, "New concepts in projection mask aligners," Opt. Eng. **14**, 142130 (1975).

[140] A. Suzuki, "Complete analysis of a two-mirror unit magnification system. Part 1," Appl. Opt. **22**, 3943 (1983).

[141] S. Rüttinger, V. Buschmann, B. Krämer, R. Erdmann, R. Macdonald, and F. Koberling, "Comparison and accuracy of methods to determine the confocal volume for quantitative fluorescence correlation spectroscopy," J. Microsc. **232**, 343 (2008).

[142] 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," J. Am. Chem. Soc. **22**, 5081 (1997).

[143] M. Sack, N. Kretschy, B. Rohm, V. Somoza, and M. M. Somoza, "Simultaneous Light-Directed Synthesis of Mirror-Image Microarrays in a Photochemical Reaction Cell with Flare Suppression," Anal. Chem. **85**, 8513 (2013).

[144] M. Sack, K. Hölz, A. K. Holik, N. Kretschy, V. Somoza, K. P. Stengele, and M. M. Somoza, "Express photolithographic DNA microarray synthesis with optimized chemistry and high-efficiency photolabile groups," J. Nanobiotechnology **14**, 14 (2016).

[145] Molecular Devices, *Datasheet for GenePix microarray scanner, Model 4400A*.

[146] NimbleGen Systems, Inc., *Datasheet for NimbleScan, version 2.1* (2002-2005).

[147] M. Schenkelberger, "Cooperative biomolecular binding," Ph.D. thesis, Universität des Saarlandes (2013).

[148] M. Born and E. Wolf, "Geometrical Theory of Optical Imaging," in "Principles of optics," (Cambridge University Press, 1999), chap. 4, pp. 353–382.

[149] S. Wilhelm, B. Gröbler, M. Gluch, and H. Heinz, "Confocal laser scanning microscopy – principles," Carl Zeiss Jena. Accessed online on Nov. 19th 2019.

[150] R. Hanbury Brown and R. Q. Twiss, "Correlation between Photons in two Coherent Beams of Light," Nature. **177**, 27 (1956).

[151] F. Zappa, A. L. Lacaita, S. D. Cova, and P. G. Lovati, "Solid-state single-photon detectors," Opt. Eng. **35**, 938 (1996).

[152] S. Kalinin, R. Kühnemuth, H. Vardanyan, and C. A. M. Seidel, "A 4 ns hardware photon correlator based on a general-purpose field-programmable gate array development board implemented in a compact setup for fluorescence correlation spectroscopy," Rev. Sci. Instrum. **83**, 096105 (2012).

[153] J. B. Brubach, A. Mermet, A. Filabozzi, A. Gerschel, and P. Roy, "Signatures of the hydrogen bonding in the infrared bands of water," J. Chem. Phys. **122**, 184509 (2005).

[154] P. O. Gendron, F. Avaltroni, and K. J. Wilkinson, "Diffusion coefficients of several rhodamine derivatives as determined by pulsed field gradient–nuclear magnetic resonance and fluorescence correlation spectroscopy," J. Fluoresc. **18**, 1093 (2008).

[155] Data retrieved from Fluorophores.org, Technische Universität Graz, available at `http://www.fluorophores.tugraz.at` (2019).

[156] M. Hazewinkel, ed., *Encyclopaedia of Mathematics* (Springer, Netherlands, 1990). Section on *Cardano formula*.

[157] C. Bohr, K. Hasselbalch, and A. Krogh, "Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt," Acta Physiol. **16**, 402–412 (1904).

[158] J. Wyman and S. J. Gill, *Binding and linkage: Functional chemistry of biological macromolecules* (University Science Books, Mill Valley, 1990).

[159] A. O. Schmitt and H. Herzel, "Estimating the Entropy of DNA Sequences," J. Theor. Biol. **188**, 369 (1997).

[160] K. A. Schallhorn, K. O. Freedman, J. M. Moore, J. Lin, and P. C. Ke, "Single-molecule DNA flexibility in the presence of base-pair mismatch," Appl. Phys. Lett **87**, 033901 (2005).

[161] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner, "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure," Proc. Natl. Acad. Sci. U.S.A **101**, 7287 (2004).

[162] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure," J. Mol. Biol. **288**, 911 (1999).

[163] D. Xu, K. O. Evans, and T. M. Nordlund, "Melting and premelting transitions of an oligomer measured by DNA base fluorescence and absorption," Biochemistry **33**, 9592 (1994).

[164] H. Ma, C. Wan, A. Wu, and A. H. Zewail, "DNA folding and melting observed in real time redefine the energy landscape," Proc. Natl. Acad. Sci. U. S. A. **104**, 712 (2007).

[165] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press Limited, 2011), 3rd ed.

[166] W. Jung, H. Jun, S. Hong, B. Paulson, Y. S. Nam, and K. Oh, "Cationic lipid binding control in DNA based biopolymer and its impacts on optical and thermo-optic properties of thin solid films," Opt. Mater. Express **7**, 3796 (2017).

[167] Q. Chi, G. Wang, and J. Jiang, "The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory," Physica A **392**, 1072 (2013).

[168] M. Mohammadi-Kambs, K. Hölz, M. M. Somoza, and A. Ott, "Hamming Distance as a Concept in DNA Molecular Recognition," ACS Omega **2**, 1302 (2017).

[169] A. Pozhitkov, P. A. Noble, T. Domazet-LoŠo, 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 Res. **34**, e66 (2006).

[170] M. R. Shortreed, S. B. Chang, D. Hong, M. Phillips, B. Campion, D. C. Tulpan, M. Andronescu, A. Condon, H. H. Hoos, and L. M. Smith, "A thermodynamic approach to designing structure-free combinatorial DNA word sets," Nucleic Acids Res. **33**, 4965 (2005).

[171] H. H. Hoos and T. Stützle, *Stochastic Local Search: Foundations and Applications* (Morgan Kaufmann, San Francisco, 2005).

[172] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," Docklady Akad. Nauk, S.S.S.R. **117**, 739 (1957).

[173] C. J. Lech, B. Heddi, and A. T. Phan, "Guanine base stacking in G-quadruplex nucleic acids," Nucleic Acids Res. **41**, 2034 (2013).

[174] C. Wu, H. Zhao, K. Baggerly, R. Carta, and L. Zhang, "Short oligonucleotide probes containing G-stacks display abnormal binding affinity on Affymetrix microarrays," Bioinformatics **23**, 2566 (2007).

[175] S. Benabou, R. Ferreira, A. Aviñó, C. González, S. Lyonnais, M. Solà, R. Eritja, J. Jaumot, and R. Gargallo, "Solution equilibria of cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops," Biochim. Biophys. Acta **1840**, 41 (2014).

[176] S. B. Provost and W. M. Ratemi, "Polynomial expansions via embedded Pascal's triangles," Acta Comment. Univ. Tartu. Math. **15**, 45 (2011).

[177] C. Smith and V. E. Hoggatt, "A study of the maximal values in Pascal's quadrinomial triangle," Fibonacci Q. **17**, 264 (1979).

[178] S. E. Rouayheb and C. N. Georghiades, "Graph Theoretic Methods in Coding Theory," in "Classical, Semi-classical and Quantum Noise," , L. Cohen, H. V. Poor, and M. O. Scully, eds. (Springer US, Boston, MA, 2012), p. 53.

[179] J. Bierbrauer, *Introduction to Coding Theory* (Chapman and Hall/CRC, New York, 2016), 2nd ed.

[180] M. R. Garey and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness* (Cambridge University Press, 2014).

[181] S. Butenko, P. M. Pardalos, I. Sergienko, V. Shylo, and P. Stetsyuk, "Finding maximum independent sets in graphs arising from coding theory," Proc. ACM Symp. Appl. Comp. p. 542 (2002).

[182] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo, "The Maximum Clique Problem," in "Handbook of Combinatorial Optimization: Supplement Volume A," , D.-Z. Du and P. M. Pardalos, eds. (Springer US, Boston, MA, 1999), p. 1.

[183] J. M. Robson, "Algorithms for Maximum independent Sets." J. Algorithm. **7**, 425 (1986).

[184] R. E. Tarjan and A. E. Trojanowski, "Finding a Maximum Independent Set," SIAM J. Comput. **6**, 537 (1976).

[185] A. Barg, S. Guritman, and J. Simonis, "Strengthening the Gilbert–Varshamov bound," Lin. Algebra Appl. **307**, 119 (2000).

[186] A. Koren, I. Tirosh, and N. Barkai, "Autocorrelation analysis reveals widespread spatial biases in microarray experiments," BMC Genom. **8**, 164 (2007).

[187] S. Ma, I. Saaem, and J. Tian, "Error correction in gene synthesis technology," Trends Biotechnol. **30**, 147 (2012).

[188] H. Koltai and C. Weingarten-Baror, "Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction," Nucleic Acids Res. **36**, 2395 (2008).

[189] P. K. Tan, T. J. Downey, E. L. Spitznagel, P. Xu, D. Fu, D. S. Dimitrov, R. A. Lempicki, B. M. Raaka, and M. C. Cam, "Evaluation of gene expression measurements from commercial microarray platforms," Nucleic Acids Res. **31**, 5676 (2003).

[190] G. L. G. Miklos and R. Maleszka, "Microarray reality checks in the context of a complex disease," Nat. Biotechnol. **22**, 615 (2004).

Publications

- M. Mohammadi-Kambs, K. Hölz, M. M Somoza, and A. Ott, "Hamming Distance as a Concept in DNA Molecular Recognition," ACS Omega **2**, 1302 (2017).
- M. Mohammadi-Kambs and A. Ott, "DNA oligomer binding in competition exhibits cooperativity," New J. Phys. **21**, 113027 (2019).
- M. Schenkelberger, C. Trapp, T. Mai, M. Mohammadi-Kambs, V. Giri, and A. Ott, "Ultrahigh molecular recognition specificity of competing DNA oligonucleotide strands in thermal equilibrium: a cooperative transition to order," (In preparation).

Acknowledgment

I must first thank my supervisor Prof. Dr. Albrecht Ott. He must have had a lot of trust to hire me as an outsider and stay patient with me through the first year of my PhD, the time I mostly needed to learn many things which were out of my expertise. Thank you Albrecht for always having time to answer my questions, discuss with me about our projects and more importantly listen to all my complaints.

I thank Prof. Dr. Gregor Jung from the biophysical chemistry group of Saarland university for accepting to be my second referee.

I am thankful to all my colleagues for keeping a friendly and pleasant atmosphere in our working group. I would like to specially thank Jenny and Varun not only for being great colleagues but also being my dear friends who always had open ears for me.

I have done part of my thesis in collaboration with Prof. Dr. Mark Somoza and Dr. Kathrin Hölz from the organic chemistry department of University of Vienna. I thank both of you for hosting me in your lab and helping me to move forward with my experiments. I am so much glad, Kathrin, that I got to know you through our collaboration and now I have you as a real friend beside me.

I must mention my friends for life, Samira, Sarah and Mozhdeh, each living in a different spot around the world. Thanks for being always so supportive and trying to keep our friendships as close as possible although we were apart in distance.

I thank my best friend in Saarbrücken, Hristina. I very much remember our first piano lesson and I already knew you are always going to be there for me. Thanks my dear for being so kind, understanding and supportive. I promise you to start playing piano very soon.

Ich möchte gerne meiner Yogalehrerin, Ilka, für die großartigen Yogastunden danken. Unsere gemeinsame Zeit hat mir definitiv sehr geholfen mich zu beruhigen. Ich freue mich jede Woche wieder auf unsere Yogastunde am Donnerstag. *Namaste.*

Ich danke meiner zweiten Familie in Deutschland Eckhard, Viola und Christopher. Ich bin sehr glücklich, dass ihr mich mit offenen Armen in eure Familie aufgenommen habt. Nur dank euch ist Benjamin so ein toller Mann geworden wie er heute ist.

I thank my mother, Batool, for instilling the right values in me. You are the strongest, kindest, most self-giving and caring woman I've ever seen. My dear brother, Amir, I am very lucky to have you as the greatest support in my life. Beyond any doubt I could not reach what I have without you always keeping my back. My dear sisters, sister-in-law, brothers-in law, Fariba, Elham, Azar, Mahtab, Shahram and Ali, I want you all to know that I live every day of my life remembering the memories we made through the years I was still living close to you. Thanks for motivating me, giving me self-confidence and involving me in your every day life despite our distance. My cutest nieces, Mana, Ashida and Anita, I am so sorry that I cannot be present for you and watch you growing up but be sure that I love you more than anything.

I dedicate my last words to my dear husband, Benjamin. I am sure I could never go through all the highs and lows of last years without having you and your support with me. I remember everything you have done to cheer me up when I was down. Thanks for all the popcorn, nachos and guacamole you made for me at 11:00 pm after my demand. They indeed helped me to chill out. Thanks for all our non-stop Netflix-watching evenings, spontaneous traveling and treating me almost patiently when I was in my crazy *Let's go shopping* mood. Sorry if I spoiled some of our nice spare time with talking about my work and of course thank you for listening to me and providing me with many helpful solutions. I am looking forward to our life in the new era beginning after our PhD.

Every flight begins with a fall.

— George R. R. Martin,
"A Game of Thrones"