Cooperative Biomolecular Binding

High specificity of competitive single stranded DNA hybridization, influence of DNA methylation on the duplex stability, and epigenetic regulation of in vitro gene switches

Dissertation
zur Erlangung des Grades

des Doktors der Naturwissenschaften (Dr. rer. nat.)

der Naturwissenschaftlich - Technischen - Fakultät II
-Physik und Mechatronik-
der Universität des Saarlandes

von
Dipl.-Phys. Marc Schenkelberger

Saarbrücken

2013

Tag des Kolloquium:	13.03.2014
Dekan:	Prof. Dr. Christian Wagner
Mitglieder des Prüfungsausschusses:	Prof. Dr. Albrecht Ott, Prof. Dr. Christian Wagner, Prof Dr. Josef Käs, Prof. Dr. Karsten Kruse, Dr. Jean-Baptiste Fleury

für meine Eltern Wolfgang und Silke

Abstract

Cooperative molecular binding and recognition is essential for the function of the biological organism. Within a molecular environment consisting of many competitors, complex molecules or molecular scaffolds need to find their matching binding partner with ultrahigh specificity. The cooperative interaction between several molecular players often generates biological functions that can no longer be observed if the molecules are separated from each other. The present thesis focuses on the cooperative molecular binding of DNA and proteins. It is shown that when several, single-stranded DNA molecules with slight sequence variations are competing for binding to one complementary strand, the error of recognition is reduced by multiple orders of magnitude in thermodynamic equilibrium. This effect is seemingly dramatic when noting the similar binding affinities of the strands without competition. Furthermore, it is investigated how DNA methylation, one of the most important epigenetic factors, modifies the specificity of DNA hybridization and influences the stability of the double helicoidal DNA molecule. Ultimately, methylated DNA is used as a template for an *in vitro* study of the transcriptional regulation of epigenetic switches. The cooperative interaction between proteins and DNA within a complex molecular environment are mandatory for the efficient switching of genes.

Kurzzusammenfassung

Das kooperative Binden und Erkennen von Biomolekülen ist ein wesentlicher Bestandteil für die Funktionalität des biologischen Organismus. Innerhalb einer molekularen Umgebung von vielen Konkurrenten müssen komplexe Moleküle ihren passenden Bindungspartner mit sehr hoher Spezifizität finden. Die kooperative Wechselwirkung zwischen mehreren molekularen Akteuren erzeugt oft biologische Funktionen, die nicht mehr erkennbar sind, sobald die Moleküle voneinander getrennt werden. Die vorliegende Arbeit konzentriert sich auf das kooperative, molekulare Binden von DNA und Proteinen. Es wird gezeigt, dass wenn mehrere, einzelsträngige DNA Moleküle geringfügig unterschiedlicher Sequenz um das Binden an eine komplementäre Spezies konkurrieren, der Fehler der molekularen Erkennung um mehrere Größenordnungen reduziert wird. Dieser Effekt ist offensichtlich sehr drastisch, wenn man die ähnlichen Bindungsaffinitäten der DNA Stränge ohne Konkurrenz berücksichtigt. Des Weiteren wird untersucht, wie DNA Methylierung, einer der wichtigsten epigenetischen Faktoren, die Spezifizität der DNA Hybridisierung beeinflusst und sich auf die Stabilität der DNA Doppelhelix auswirkt. Letztendlich wird die methylierte DNA als Template für eine *in vitro* Untersuchung der transkriptionellen Regulierung von epigentischen Schaltern verwendet. Es wird gezeigt, dass die kooperative Wechselwirkung zwischen Proteinen und DNA, innerhalb einer komplexen molekularen Umgebung, für das effektive Schalten von Genen zwingend erforderlich ist.

Contents

1. Introduction

The functionality of all biological systems depends on the specific interaction of complex biomolecules or molecular scaffolds. Within a mixture of many similar competitors of comparable binding affinity, a particular target molecule has to find its matching complement with ultrahigh accuracy. The specific binding of transcription factors to DNA, the first steps in virus-cell recognition, or the charging of tRNAs with amino acids are only a few examples for such molecular recognition systems. Although the interactions of one single recognition pair are relatively simple, the multiplicity of cooperating molecules often generates the biological function of the total system in the first place. The properties of a system studied *in vivo* are often no longer observable *in vitro*, when the relevant molecular players are decoupled from the complex molecular environment; protein complexes often gain properties that neither of the subunits possess without the others.

This work deals with the cooperative molecular recognition of DNA. DNA is a double helicoidal bio-polymer, whose base sequence contains the genetic information that is essential for protein manufacturing, biological function, and phenotype of the organism. In this sense, DNA is often called the "blueprint of life". Formation of a DNA double strand is a highly cooperative, poorly understood recognition process that is based on the complementarity of the matching base pairs adenine-thymine (AT) and cytosine-guanine (CG). Besides the hybridization of these complementary bases, there is also the possibility of erroneous base pair formation, as for instance AA or CT mismatches. These so called single nucleotide polymorphisms (SNPs) are the main cause for cancer or autoimmune diseases, since they can lead to erroneous protein synthesis. In order to react to environmental factors, nature has established the possibility to modify gene expression and phenotypes by mechanisms other than changes in the underlying DNA sequence. One of the most important of these so called epigenetic factors is DNA methylation, a chemical, switch-like modification of the DNA strand. The influence of DNA methylation on the cooperative recognition within a complex molecular environment is roughly understood. Ultimately, the double helicoidal DNA strand is the

template for protein expression, one of the most complex and powerful molecular recognition systems. A diversity of molecular players are involved in this highly cooperative mechanism and need to bind specifically to the DNA molecule as a function of gene sequence, methylation state, and other signals.
In the context of this work, the cooperative molecular recognition of DNA is studied by using three different examples:

1. **Competitive DNA hybridization** deals with the molecular competition between two single-stranded DNA species A and B (targets) with slight sequence variations, for duplex formation with one complementary species (probe) in thermodynamic equilibrium. It is generally believed that the accuracy of this molecular recognition solely depends on the difference between the individual binding free energies of the competitors A and B, as long as thermodynamic equilibrium conditions apply. The present work shows that the molecular competition can reduce the recognition error by multiple orders of magnitude. It is demonstrated that thermal fluctuations lead to an antagonistic and cooperative interaction between the competing target species that causes a modification of the binding free energy landscape in favor of one particular DNA strand. This is without any energetic cost, except the introduction of the competitor. For the study of DNA hybridization, fluorescently labeled oligonucleotides are used, which are allowed to bind to complementary DNA. The hybridization is detected via total internal reflection fluorescence (TIRF) or Förster resonance energy transfer (FRET). Base pair fluctuations of formed duplexes are analyzed with fluorescence correlation spectroscopy (FCS). Compared to other methods, like for instance mass adsorption or plasmon resonance, the optical setups are relatively simple and also provide a better signal to noise ratio. DNA hybridization also plays an essential role for biotechnological methods, like for instance *in situ* hybridization, polymerase chain reactions (PCR), and the DNA microarray technology, a poorly understood but frequently used method for the massively parallel analysis of gene expression. In this sense, the results of this work also contribute for optimizing and understanding these biotechnological methods.

2. The second part of this work deals with the question how **DNA methylation** affects the stability of double-stranded DNA and how it improves the recognition specificity of hybridization. Here, the study also includes investigations about the molecular competition between methylated and non-methylated DNA. It is shown that methylation of solely one or two bases within a short DNA strand leads to an

increased affinity of at least one order of magnitude for binding of this strand to its complement. The hybridization of methylated DNA is also detected via TIRF.

3. The last study deals with the **epigenetic regulation of gene expression**. Two transcriptional regulation systems are investigated: the procaryotic epigenetic switch based on the methylation depended DNA binding proteins Lrp and PapI, as well as the eucaryotic methylation binding domain (MBD) of the human repressor MeCP2. The experimental study shows that the cooperative interaction of the molecular players within a complex molecular environment are mandatory for the specific switching of genes. Furthermore, an artificial epigenetic two state switch based on both, procaryotic and eucaryotic DNA methylation, is designed and tested. The epigenetic regulation of gene expression is investigated in a homemade E. coli based *in vitro* cell free expression system. The system contains all the relevant energy sources and molecular machinery for expressing a gene from a DNA template. The study involves cloning of different regulatory DNA operons into circular plasmids, using methods of molecular biology. The methylation dependent DNA binding proteins, directly expressed in the cell free system, can bind to these operons and consequently activate or hinder the measurable expression of a reporter protein. Compared to *in vivo* methods, the cell free expression system enables the study of the transcriptional regulation within a totally endogenous E. coli system, including all the relevant molecular players, but under highly controlled experimental conditions. In this sense, a secondary objective of the experimental study is the characterization of the cell free expression system as a tool for designing synthetic gene circuits.

The present thesis is divided in seven chapters. Chapter 2 gives an overview about the fundamentals of competitive DNA hybridization, transcriptional regulation of gene expression, and epigenetics. It further informs about the state of the art of the current research by appreciating the relevant literature. Chapter 3 introduces all the experimental materials and methods, which are used in the context of this work. The chapters 4, 5, and 6 represent the main part of this thesis. All experimental and numerical findings concerning the competitive DNA hybridization, the influence of DNA methylation on the DNA duplex stability, and the *in vitro* study of epigenetic switches are presented and discussed in detail. The final chapter 7 briefly summarizes the most important results of the present work and further gives an outlook on future studies.

2. Fundamentals

2.1. Structure of the deoxyribonucleic acid

The deoxyribonucleic acid (DNA) carries the genetic information that is essential for the function of almost all biological organisms [1]. In all higher eucaryotic cells, the DNA is located in the nucleus and packed into chromosomes. Each chromosome consists of one double helicoidal DNA molecule. Fig. 2.1 shows a schematic of the structure of DNA. Like all nucleic acids, DNA consists of individual components, the so called nucleotides. These nucleotides consist of the sugar deoxyribose, phosphate, and a heterocyclic nucleobase. There are four kinds of bases: adenine (A), thymine (T), guanine (G), and cytosine (C). The sequence of several nucleotides in a raw constitutes a single-stranded DNA polymer. Here, all deoxyribose and phosphate molecules form the so called DNA backbone. The five carbon atoms of the deoxyribose are usually consecutively numbered. The 3′ atom carries a hydroxyl group that links the deoxyribose to the 5′ atom of the phosphate of the adjacent nucleotide via a phosphodiester bond. Thus, a DNA single strand possesses an orientation, the 5′ end and the 3′ end, depending on either the deoxyribose or the phosphate is the terminal molecule. If two DNA single strands are brought together they can hybridize to one helicoidal double strand. This requires the specific formation of complementary Watson-Crick base pairs (AT and GC) via hydrogen bonds. This base pairing is often called "Lock and Key" binding and therefore defines the sequence specificity of DNA hybridization. A GC base pair is more stable than an AT base pair due to the increased number of hydrogen bonds. Besides the hydrogen bonds, the helicoidal DNA molecule is additionally stabilized by stacking interactions that actually generate the major part of the binding free energy but without any sequence specificity. This attractive interaction results from the overlap of the π electron orbitals of the stacked nucleotides as well as the van der Waals linking of adjacent bases. Duplex formation happens in an anti-parallel manner, what means that the 3′ end of one single strand is on the opposite side of the 5′ end of the second single strand. The pitch of the right handed DNA double helix in the B conformation is 3.4 nm

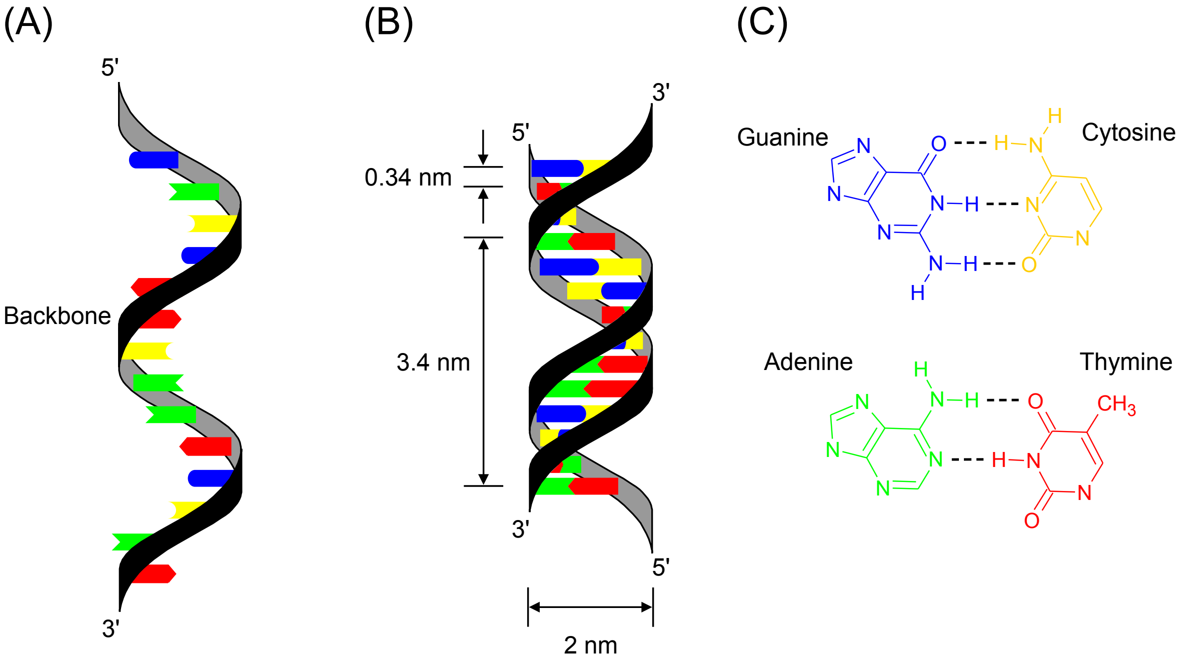


Figure 2.1.: Schematic illustration of single-stranded **(A)** and double-stranded **(B)** DNA, as well as the chemical structure of Watson-Crick base pairing **(C)**. The deoxyribose and phosphate molecules constitute the DNA backbone. A single strand possesses an orientation, the 5′ end and the 3′ end, depending on which molecule, deoxyribose or phosphate, is the terminal one. Formation of a double strand requires the formation of Watson-Crick base pairs AT or CG. The pitch of the right handed DNA double helix in the B conformation is 3.4 nm (about 10 bases) and the diameter is 2 nm.

and the diameter is 2 nm. The process of DNA hybridization highly depends on the ion concentration of the solvent. The DNA backbones of two single-stranded molecules are negatively charged and this leads to a repulsive electrostatic interaction when these molecules come into close distance. The positively charged ions of the solvent act as counterions and screen the repulsive force of the DNA backbones. Compared to a single strand, a double helicoidal molecule possesses less entropy due to the reduced number of degrees of freedom. The persistence length of a single strand is about 1 nm that is approximately the distance between two adjacent bases. Compared to this, the persistence length of a double strand is about one to two orders of magnitude higher [2]. A formed double strand can be separated again into two single strands. This reversible reaction is called DNA denaturation [3].

2.2. Theoretical assessment of DNA hybridization

2.2.1. The Langmuir model

The hybridization of two single-stranded DNA oligonucleotides to one double-stranded molecule is a highly cooperative process [4]. The rate of duplex formation is independent of the length of the single-stranded DNA molecule, which makes the formation of a primary double-stranded region in-between completely single-stranded regions to the limiting process. Subsequent to this initial nucleation, the closure of the following base pairs happens in a highly cooperative manner through molecular zipping [5]. From this it follows that DNA hybridization is typically described with a two-state model that assumes the direct and reversible transition between double strands D and single strands T and P. Assuming an ensemble of solved DNA molecules, the hybridization reaction can be described with the equation

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

where $[T]$ and $[P]$ are the concentrations of single-stranded oligonucleotides T and P, respectively, and $[D]$ is the concentration of formed duplexes D. The unit of the respective DNA concentration is 1 mol/l = 1 M. k_+ is the rate of association (unit $1/(\mathrm{M} \cdot \mathrm{s})$) and k_- (unit 1/s) is the rate of dissociation. Fig. 2.2 shows a schematic of the hybridization reaction with all the relevant molecular players. The donations T and P for the single-stranded species stem from the nomenclature **T**arget and **P**robe defined for hybridization experiments performed on a DNA microarray [5]. In this case, the hybridization takes place on a biochip, since the probes are immobilized to a surface that is in contact with a solution of concentrated target molecules. Probe immobilization does not invalidate the assumptions of the two-state model. Both, bulk hybridization and surface based hybridization are described by Eq. (2.1), if the mass transport of target molecules in bulk to the complementary probes on the surface can be neglected and the rate constants k_+ and k_- are not significantly modified. In this case, the number of formed duplexes on the surface can be expressed in units equivalent to the duplex concentration in bulk. However, in thermodynamic equilibrium, the mass transport does not play a role anyway, since it solely affects the kinetics of the system but not the stationary state. Since most of the experiments in the context of this work are performed under thermodynamic equilibrium conditions, the contribution of mass transport is neglected in the following. Furthermore, it is assumed that all binding

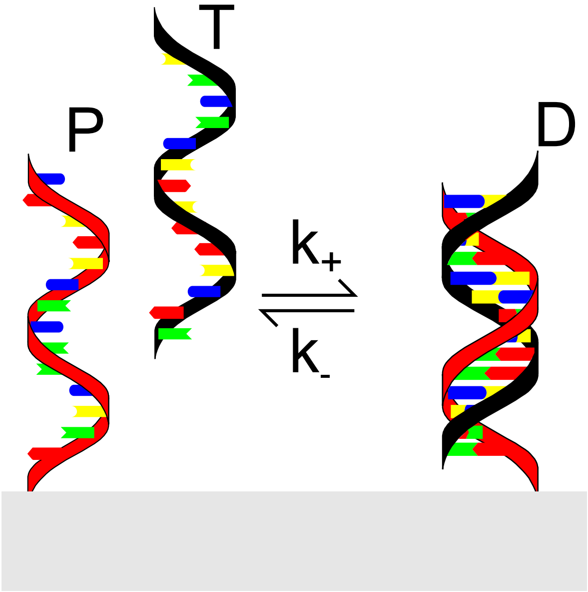


Figure 2.2.: Schematic of the hybridization reaction in the two-state model. Single-stranded oligomers target T and probe P with complementary bases can form a double-stranded duplex D with the association rate k_+. An existing duplex can resolve again into target and probe with the dissociation rate k_-. The nomenclatures "target" and "probe" are used for hybridization experiments performed on a DNA microarray. Here, the probe is immobilized to a surface that is in contact with a solution of concentrated target molecules.

sites are energetically equal and that there is no interaction between adjacent probe molecules. From this it follows that the hybridization kinetics can be described with the differential equation

$$[\dot{D}](t) = k_+[T](t)[P](t) - k_-[D](t) \tag{2.2}$$

In the case of an excess concentration of target molecules over probe molecules ($[T] \gg [P]$), the target concentration can be assumed to remain constant in time

$$[T](t) \approx [T]_0 \tag{2.3}$$

Here, $[T]_0$ is the concentration value of targets at time $t = 0$. However, the concentration of probe molecules $[P](t)$ available for hybridization decreases as a function of time with the number of already formed duplexes $[D]$

$$[P](t) = [P]_0 - [D](t) \tag{2.4}$$

where $[P]_0$ is the concentration value of probes at time $t = 0$. Thus, the solution of Eq. (2.2) is

$$[D](t) = \frac{k_+[P]_0[T]_0}{k_+[T]_0 + k_-} \cdot \left(1 - e^{-(k_+[T]_0+k_-)t}\right) = [D]_{eq}\left(1 - e^{-\gamma t}\right) \tag{2.5}$$

where

$$\gamma = k_+[T]_0 + k_- \tag{2.6}$$

and

$$[D]_{eq} = [P]_0\frac{k_+[T]_0}{\gamma} = [P]_0\frac{k_+[T]_0}{k_+[T]_0 + k_-} = [P]_0\left(1 + \frac{1}{K[T]_0}\right)^{-1} \tag{2.7}$$

K is the equilibrium binding constant of the target molecules (unit 1/M):

$$K = \frac{[D]_{eq}}{[T]_{eq}[P]_{eq}} = \frac{k_+}{k_-} \tag{2.8}$$

where $[D]_{eq}$, $[T]_{eq}$, and $[P]_{eq}$ are the concentrations of duplexes, targets and probes in thermodynamic equilibrium, respectively. Dividing Eq. (2.7) by $[P]_0$ reveals the fraction Θ of occupied probe molecules:

$$\Theta(t) = \frac{[D](t)}{[P]_0} = \Theta_{eq}\left(1 - e^{-\gamma t}\right) \tag{2.9}$$

where

$$\gamma = k_+[T]_0 + k_- \tag{2.10}$$

and

$$\Theta_{eq} = \frac{k_+[T]_0}{\gamma} = \frac{k_+[T]_0}{k_+[T]_0 + k_-} = \left(1 + \frac{1}{K[T]_0}\right)^{-1} \tag{2.11}$$

Eq. (2.11) is defined as the Langmuir isotherm. Fig. 2.3 A shows the curve progression $[D](t)$ as a function of time following Eq. (2.5) for different target concentration $[T]_0$ and for typical values for the association k_+ and dissociation rate k_-. Fig. 2.3 B shows the Langmuir isotherm following Eq. (2.11) as a function of the target concentration $[T]_0$ for different values of the binding constant K.

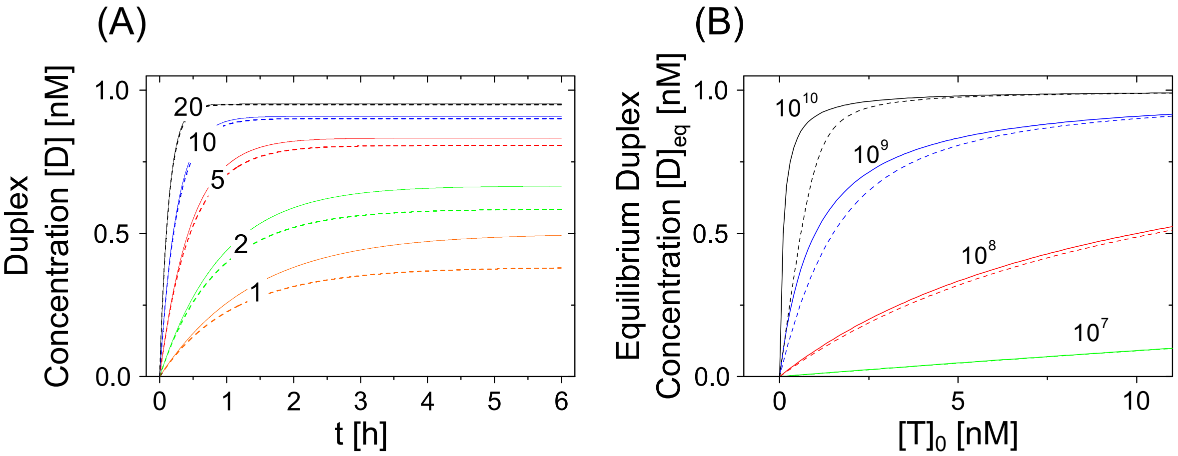


Figure 2.3.: Comparison between the Langmuir model (solid lines) and the extended Langmuir model (dashed lines). For all graphs, the initial concentration of free probe molecules $[P]_0$ is 1 nM. **(A)**, kinetics of the hybridization reaction as described by Eqs. (2.5) and (2.14) for different target concentrations $[T]_0$ (values in nM) and typical values of the association rate ($k_+ = 10^5$ M^{-1} s^{-1}) and the dissociation rate ($k_- = 10^{-4}$ s^{-1}). **(B)**, curve progression of the equilibrium duplex concentration $[D]_{eq}$ as a function of the target concentration $[T]_0$ as described by Eqs. (2.11) and (2.16), for different binding constants K (values in M^{-1}).

2.2.2. The extended Langmuir model

If the concentrations $[T]_0$ and $[P]_0$ in a hybridization experiment are comparable, the target concentration can no longer be considered to remain constant in time. Rather, in analogy to Eq. (2.4), the initial target concentration $[T]_0$ has to be corrected by the concentration of targets that have already hybridized to a probe molecule [6]

$$[T](t) = [T]_0 - [D](t) \tag{2.12}$$

With this, the differential equation (2.2) is modified as follows:

$$[\dot{D}](t) = k_+ \left([T]_0 - [D](t)\right)\left([P]_0 - [D](t)\right) - k_-[D](t) \tag{2.13}$$

with the analytical solution

$$[D](t) = \frac{a}{2} \cdot \frac{1 - ce^{ak_+t}}{1 + ce^{ak_+t}} + \frac{b}{2} \tag{2.14}$$

where

$$a = \sqrt{b^2 - 4[P]_0[T]_0},\ b = [P]_0 + [T]_0 + K^{-1},\ \text{and}\ c = \frac{a+b}{a-b} \tag{2.15}$$

The equilibrium state is described by

$$\begin{aligned} [D]_{eq} &= \frac{1}{2}(b-a) \\ &= \frac{1}{2}\left\{[T]_0 + [P]_0 + K^{-1} - \sqrt{([T]_0 + [P]_0 + K^{-1})^2 - 4[T]_0[P]_0}\right\} \end{aligned} \tag{2.16}$$

Eq. (2.16) is defined as the extended Langmuir isotherm. The curve progression $[D](t)$ as a function of time (Eq. (2.14)) as well as the extended Langmuir isotherm as a function of the target concentration $[T]_0$ (Eq. (2.16)) are also shown in Fig. 2.3 in order to directly compare them with the not-extended model. Fig. 2.3 shows that the difference between the two models can be neglected for $[T]_0 \gg [P]_0$. In this case, the assumption of a constant target concentration $[T]_0$ is valid (see Eq. (2.3)). However, in the case of comparable target and probe concentrations ($[T]_0 \approx [P]_0$) the extended model should be used. This particularly applies for high values of the binding constant K.

2.2.3. Competitive hybridization

In the following, the extended Langmuir model introduced in the previous section is expanded to the case of a molecular competition between two target species T^A and T^B with different binding affinities for one probe species P [7, 8]. This situation constitutes the simplest case of a multiplex hybridization [9]. Assuming a two-state description for this competitive situation, the reaction equations are

$$[T^A] + [P] \underset{k_-^A}{\overset{k_+^A}{\rightleftharpoons}} [D^A] \tag{2.17}$$

and

$$[T^B] + [P] \underset{k_-^B}{\overset{k_+^B}{\rightleftharpoons}} [D^B] \tag{2.18}$$

where $[T^A]$ and $[T^B]$ are the concentrations of the competing target species A and B, $[D^A]$ and $[D^B]$ are the concentrations of duplexes consisting of probe and target molecule A or B, and k_+^A, k_-^A, k_+^B and k_-^B are the respective association and dissociation rate constants. The target species A and B are both complementary to the probe molecule, however species B contains non-matching base pairs that decrease the binding

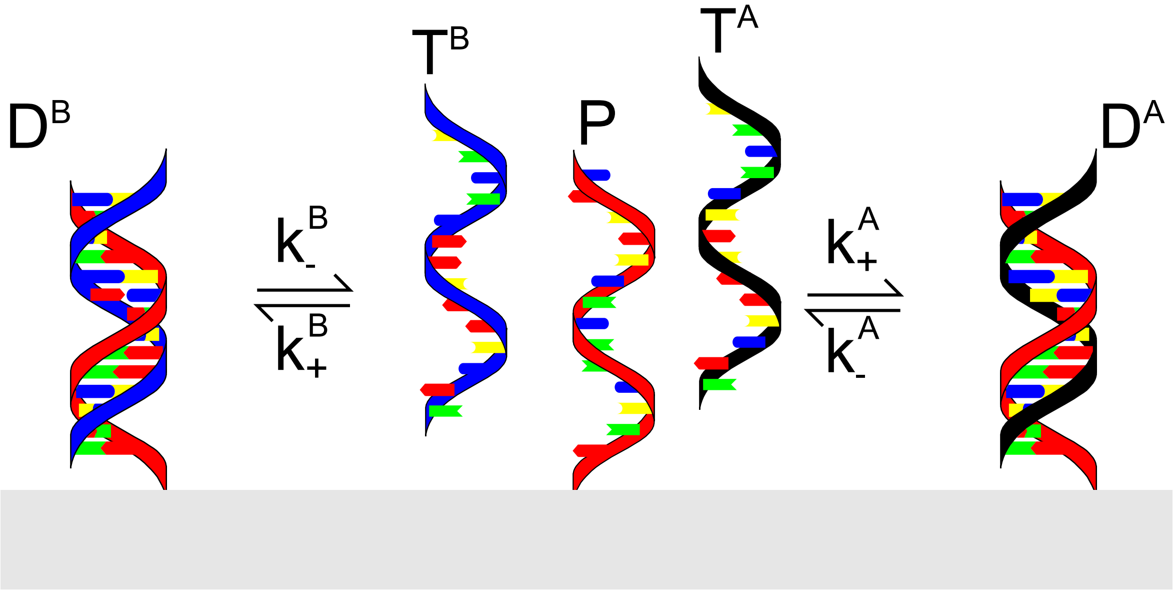


Figure 2.4.: Schematic of the competitive hybridization reaction in the two-state model. A probe molecule P can either be occupied by target T^A or T^B resulting in formation of duplex D^A or D^B. This happens with the association rate k_+^A and k_+^B, respectively. An existing duplex denatures into target and probe with the dissociation rate k_-^A and k_-^B, respectively.

affinity K^B compared to the value of K^A. There is no interaction between the competing target molecules except for the steric hindrance. This means for instance that a probe molecule that is occupied by A cannot be occupied by B. Furthermore, there is no interaction between neighboring binding sites, if the probe molecules are immobilized to a surface. Fig. 2.4 shows a schematic of the competitive hybridization reaction involving all molecular players. In the following, the molecular competition between target species A and B is defined as the competitive system "A vs. B", where A is the "High Affinity Target" (HAT) and B the "Low Affinity Target" (LAT). In analogy to Eq. (2.13), the kinetics of the duplex concentrations, $[D^A]$ and $[D^B]$, can be expressed with the following differential equations

$$[\dot{D}^A](t) = k_+^A \left([T^A]_0 - [D^A](t)\right)\left([P]_0 - [D^A](t) - [D^B](t)\right) - k_-^A[D^A](t) \tag{2.19}$$

and

$$[\dot{D}^B](t) = k_+^B \left([T^B]_0 - [D^B](t)\right)\left([P]_0 - [D^A](t) - [D^B](t)\right) - k_-^B[D^B](t) \tag{2.20}$$

The numerical solution of this coupled system of differential equations is presented in Fig. 2.5. Here, a situation is considered that is typical for the experiments performed in the context of this work. The target concentration $[T^B]_0$ can exceed the target concentration $[T^A]_0$ by several orders of magnitude, in order to investigate the influence of this excess concentration on the equilibrium duplex concentration $[D^A]_{eq}$. Fig. 2.5 illustrates the effect caused by the competitive situation. Due to the excess concentration of $[T^B]_0$ over $[T^A]_0$, the initial slope of $[D^B](t)$ is increased compared to the respective one of $[D^A](t)$. With the initiating increase of $[D^A](t)$, target molecules of species A occupy the probes, which can no longer be occupied by molecules of species B. Consequently, $[D^B](t)$ decreases. For $[\dot{D}^A](t = \infty) = [\dot{D}^B](t = \infty) = 0$ and $K = k_+/k_-$ one reveals the equilibrium duplex concentrations

$$[D^A]_{eq} = \frac{[P]_0 K^A [T^A]_{eq}}{K^A[T^A]_{eq} + K^B[T^B]_{eq} + 1} \tag{2.21}$$

and

$$[D^B]_{eq} = \frac{[P]_0 K^B [T^B]_{eq}}{K^A[T^A]_{eq} + K^B[T^B]_{eq} + 1} \tag{2.22}$$

$[T^A]_{eq}$ and $[T^B]_{eq}$ are the target concentrations of species A and B in thermodynamic equilibrium. Please note that for very low concentrations $[T^B]_{eq} \approx 0$, Eq. (2.21) is identical to Eq. (2.7). The expressions (2.21) and (2.22) are the extended Langmuir isotherms in competition. Dividing Eq. (2.21) by Eq. (2.22) gives the ratio of duplex concentrations of the competing species A and B in thermodynamic equilibrium:

$$\frac{[D^A]_{eq}}{[D^B]_{eq}} = \frac{[T^A]_{eq}}{[T^B]_{eq}} \frac{K^A}{K^B} \tag{2.23}$$

This means that in the two-state model, the competitive situation is fully described by the individual binding affinities of the competing target species and their respective concentrations. As shown in Fig. 2.5, $[D^A]_{eq}$ decreases with increasing concentrations $[T^B]_{eq}$ like expected from Eq. (2.19). One has to mention that even for low target concentrations $[T^A]_0 + [T^B]_0 \leq [P]_0$, $[D^B](t)$ decreases with initiating hybridization of target molecules of species A. This becomes clear by considering for instance the curve for $[D^A](t)$ and $[D^B](t)$ in Fig. 2.5 for a concentration of $[T^B]_0 = 10$ nM. Thus, this reduction of $[D^B](t)$ is not due to the limited number of free probe molecules, but due to the molecular competition itself.

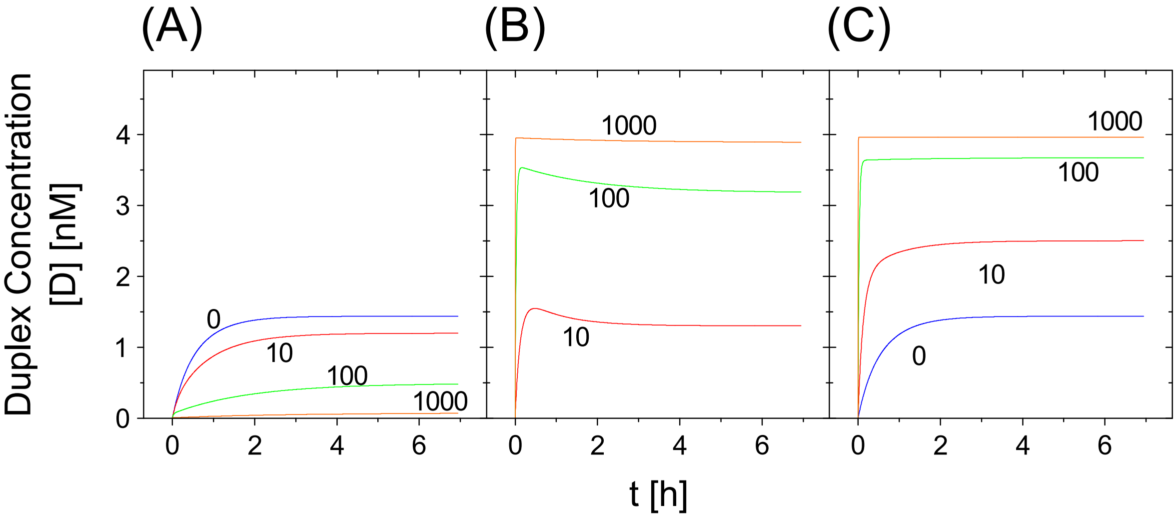


Figure 2.5.: Numerical solution of the coupled differential equations (2.19) and (2.20) for $[P]_0 = 4$ nM, $k_+^A = k_+^B = 10^5$ M^{-1}s^{-1}, $k_-^A = 10^{-4}$ s^{-1}, and $k_-^B = 10^{-3}$ s^{-1}. For all calculations, the concentration $[T^A]_0$ of target species A is 2 nM, while $[T^B]_0$ is varied from 0 nM to 1000 nM. The respective values of $[T^B]_0$ are given in the graph. The graphs show the duplex concentration $[D^A]$ **(A)**, $[D^B]$ **(B)** and the sum $[D^A]+[D^B]$ **(C)** as a function of time.

2.2.4. Correlation between binding constant and free energy

As mentioned above, in the two-state model, the equilibrium binding constant K of a particular target molecule is defined as the ratio of reaction products and educts following

$$K = \frac{[D]_{eq}}{[T]_{eq}[P]_{eq}} \tag{2.24}$$

The binding constant is proportional to the Boltzmann factor

$$K \propto e^{\Delta G/RT} \tag{2.25}$$

where ΔG is the thermodynamic potential of the Gibbs free energy, $R = 1.987$ cal $\cdot$ mol^{-1}K^{-1} is the molar gas constant, and T is temperature. ΔG can be expressed by the changes of enthalpy ΔH and entropy ΔS upon the denaturation of a DNA double strand:

$$\Delta G = \Delta H - T\Delta S \tag{2.26}$$

Combining the Eqs. (2.25) and (2.26) gives

$$\ln(K) = \frac{\Delta H}{R} \cdot \frac{1}{T} - \frac{\Delta S}{R} \tag{2.27}$$

In the two state model, plotting the logarithm of the binding constant K as a function of the inverse temperature $1/T$ (Arrhenius plot) reveals a linear dependence. Here, intercept and slope reveal the change in enthalpy ΔH and entropy ΔS, respectively. This enables conclusions on the duplex stability of the double-stranded DNA molecule at a specified temperature and also the comparison of entropic degrees of freedom of different duplexes. For the case of the competitive hybridization between target species A and B (see section 2.2.3), the combined Eqs. (2.23) and (2.25) link the ratio of duplex concentration to the difference in binding free energy $\Delta\Delta G = \Delta G^A - \Delta G^B$ of the competing target species A and B:

$$\frac{[D^A]_{eq}}{[D^B]_{eq}} = \frac{[T^A]_{eq}}{[T^B]_{eq}} \frac{K^A}{K^B} = \frac{[T^A]_{eq}}{[T^B]_{eq}} \cdot e^{\Delta\Delta G/RT} \tag{2.28}$$

This shows that, besides the concentrations of competing target molecules, solely the difference in the binding free energy $\Delta\Delta G$ defines the competitive situation.

2.2.5. The melting temperature of double-stranded DNA

After transposing Eq. (2.11), one can express the binding affinity K as a function of the fraction Θ_{eq} of occupied probes in thermodynamic equilibrium

$$K = \frac{\Theta_{eq}}{(1 - \Theta_{eq})[T]_0} \tag{2.29}$$

The experimental value of the melting temperature T_m is defined as the temperature where the fraction of occupied probes equals 50%

$$T_m = T\left(\Theta_{eq} = 0.5\right) \tag{2.30}$$

Thus, and by considering Eq. (2.29), one reveals the binding constant K at the melting temperature T_m:

$$K\left(T_m\right) = \frac{1}{[T]_0} \tag{2.31}$$

Adapting Eq. (2.31) into Eq. (2.27) for $T = T_m$ then reveals

$$T_m = \frac{\Delta H}{\Delta S - R\ln([T]_0)} \tag{2.32}$$

Like explained in section 2.2.1, the previous derivation of the melting temperature T_m is only valid in the case of an excess concentration of targets over probes (not extended Langmuir model, $[T]_0 \gg [P]_0$). In the case of target concentrations that are comparable with the probe concentration, $[T]_0$ in Eq. (2.32) has to be replaced by the factor $[S]_1 - [S]_2/2$, where $[S]_1$ and $[S]_2$ are the concentrations of the higher and lower concentrated single strand 1 or 2, respectively [10]. Eq. (2.32) enables the calculation of the melting temperature T_m of a specified target, concentrated at $[T]_0$. The thermodynamic parameters ΔH and ΔS can be calculated with the usually employed nearest neighbor model of DNA hybridization [5, 11].

2.3. The central dogma of molecular biology

The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information [12]. During the transcription, the genetic information, which is encoded by the sequence of the double-stranded DNA molecule, is genewise transcribed into messenger ribonucleic acid (mRNA). During the translation the information carried by the mRNA is then translated by ribosomes into a polypeptide chain that folds into a protein. These molecular processes are usually summarized under the term gene- or protein expression. The dogma is visualized in Fig. 2.6. In the following, the central contents of gene expression are explained in detail, without the claim to be exhaustive. The main focus lies on the protein expression in bacteria, since this is the relevant organism for the context of this work.

2.3.1. Genomic DNA

A gene is a molecular unit of heredity of a living organism and constitutes the blue print for the expression of specified proteins and functional RNA chains [1]. Thus, genes hold the information to build and maintain a biological cell and also to pass this information to offspring. DNA that contains the genetic information is named genomic DNA. The function of other parts of the genome, which is often named “junk DNA”, is not well understood [5]. In procaryotic organisms, as for instance bacteria, the genome is mainly

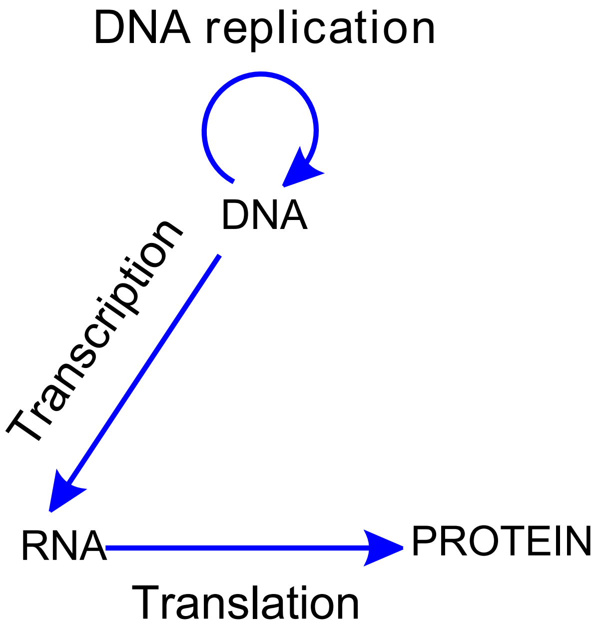


Figure 2.6.: Central dogma of molecular biology [12]. Only the naturally relevant pathways of biological information are shown. The genetic information hold by the DNA is transcribed to mRNA via the transcription process. mRNA is translated into a polypeptide chain via the translation process. The peptide folds into a functional protein. The DNA template is replicated during cell division.

organized in circular DNA plasmids. In eucaryotic organisms, the genome is packed into chromosomes that contain the DNA in a highly compact but well ordered form.

2.3.2. Transcription

The first step of gene expression is the transcription in which a gene on the DNA is copied into mRNA by the enzyme RNA polymerase. Fig. 2.7 schematically illustrates the transcription. The process begins with the specific binding of the RNA polymerase to the promoter region of the DNA. The promoter is located a few base pairs upstream (reading direction 5′ → 3′) from the transcription start site, and can be a few 100 base pairs long. In bacteria the promoter contains two short sequence elements approximately 10 and 35 base pairs upstream from the gene. The -10 element has the consensus sequence ATAAT and the -35 element has the consensus sequence TTGACA. Both elements are important recognition sites for the specific binding of so called sigma factors (σ). The σ factor is a subunit of the RNA polymerase that is essential for the

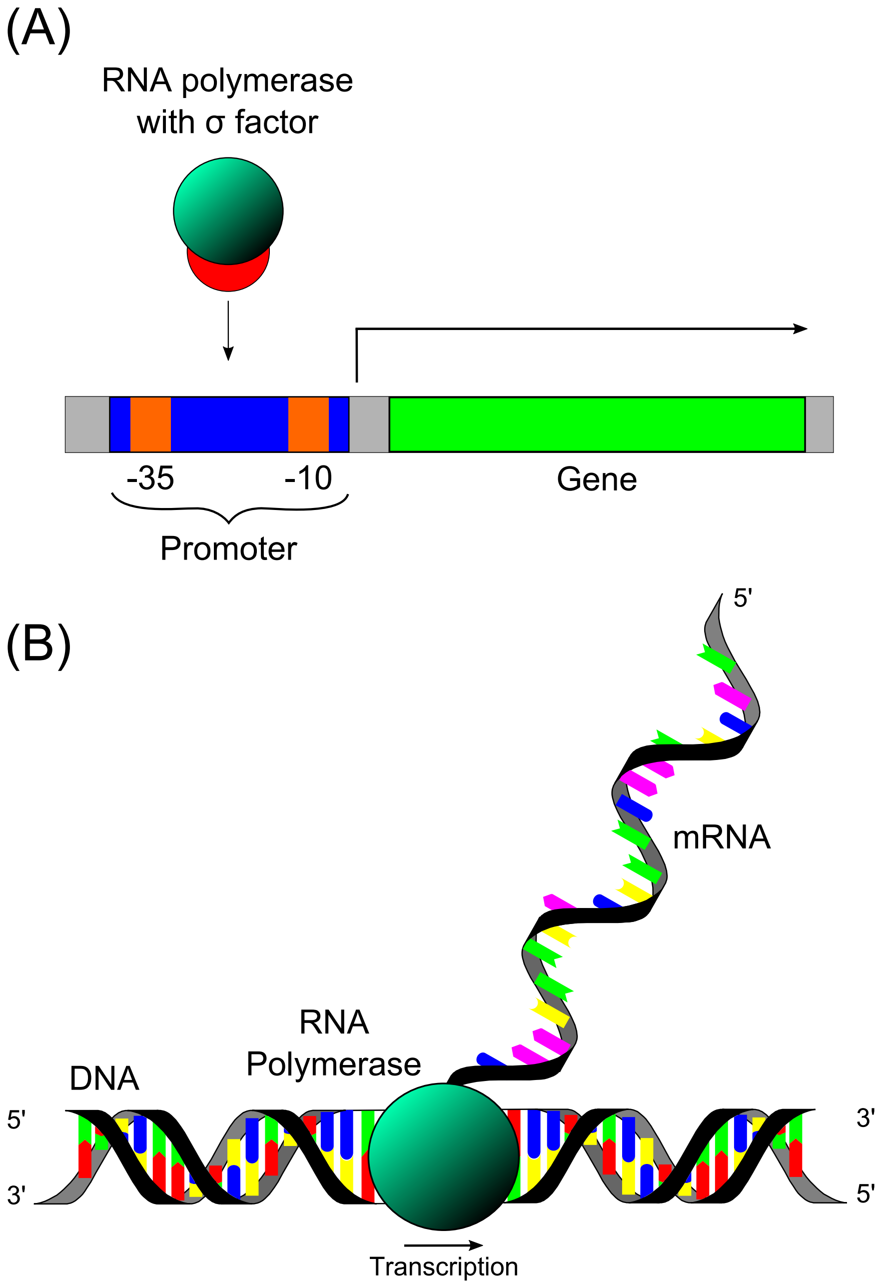


Figure 2.7.: Schematic illustration of the transcription. **(A)**, initiation. The RNA polymerase with its σ subunit specifically binds to the promoter region, located upstream of the gene that is to transcribe. The promoter consists of the -35 and -10 consensus sequences for binding of the transcription factor. **(B)**, elongation. After binding of the RNA polymerase, the sigma factor is released and the enzyme elongates the DNA strand. It uses base pair complementarity in order to transcribe the genetic information hold by the DNA into mRNA.

transcription initiation. The housekeeping σ factor of the bacteria E. coli is σ^{70}, indicating a protein of a molecular weight of 70 kD. The promoter region is also essential for gene regulation on the transcriptional level. Other proteins can also specifically bind to the promoter and thus either enable transcription initiation or prevent it. Such proteins are therefore often named activator or repressor, respectively. Gene regulation on the transcriptional level is explained in more detail in section 2.4. After transcription initiation, the σ factor is released and the RNA polymerase traverses the DNA strand. During this elongation step, the enzyme uses base pair complementarity with the DNA template to transcribe the genetic information hold by the DNA into mRNA. The structure of RNA is similar to that of DNA, however there are differences in chemical stability and biological function. The main difference is the substitution of the DNA-typical base thymine by the RNA-typical base uracil (U). Furthermore, the nucleotides are composed of the sugar ribose instead of deoxyribose (one less oxygen atom). Thus, except of the mentioned differences, transcription produces an exact RNA copy of the coding DNA template. After the gene is completely transcribed, the transcription is terminated either by the formation of a GC rich hairpin loop in the synthesized mRNA strand (intrinsic termination) or by destabilizing the interaction between DNA template and mRNA due to binding of the so called ρ proteins (rho-termination).

2.3.3. Translation

Translation is the process of protein manufacturing that takes place in cellular ribosomes [1]. Fig. 2.8 schematically illustrates the translation. After transcription, the genetic information hold by the mRNA is transcribed into a polypeptide chain. A triplet of three adjacent mRNA bases, the so called codon, codes for one of the 20 naturally relevant amino acids. Thus, the chain is basically synthesized as several amino acids in a row, which are incorporated into the peptide chain by transfer ribonucleic acid (tRNA). Unlike the mRNA, the tRNA is not a template of genetic information, but folds itself into a complex “nanomachine” which is charged with a specified amino acid. The anticodon, an exposed base triplet of the tRNA, can specifically bind to a complementary codon sequence on the mRNA [5]. Upon binding, the tRNA incorporates its charged amino acid into the peptide chain. Afterwards, the discharged tRNA is released by the ribosome which moves to the next codon. In procaryotic organisms the translation is initiated by a special codon, the so called start codon AUG. The end of the translation is reached, when one of the three stop codons UAG, UAA, or AGA appears on the mRNA. These codons are recognition sites not for tRNA but for release factor proteins that prompt

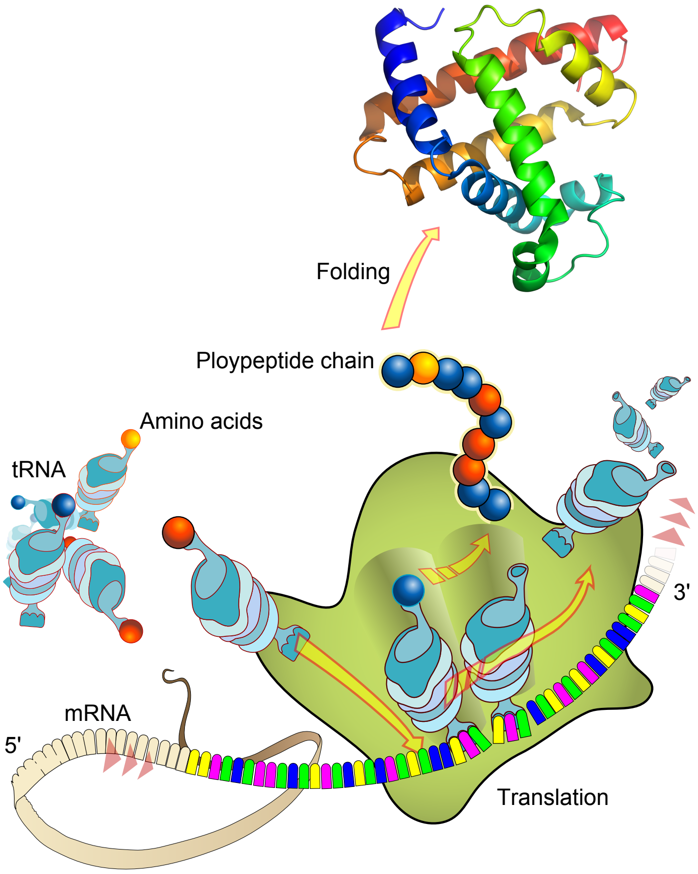


Figure 2.8.: Schematic illustration of the translation. The genetic information hold by the mRNA is translated into a polypeptide chain consisting of amino acids. The incorporation of the amino acid is performed by tRNA. Packets of three mRNA bases (codons) code for one amino acid. After the formation of the peptide chain, it folds into a functional protein. The schematic of the ribosome is adapted from *www.wikipedia.org*.

the disassembly of the entire ribosome-mRNA complex. The already read out mRNA is degraded by nucleases.

2.4. Epigenetic regulation of gene expression

The terminus epigenetics is nowadays commonly used to summarize the study of changes in gene expression and phenotypes caused by mechanisms other than changes in the underlying DNA sequence. However, a commonly accepted and broad definition of the term is not yet available. Epigenetics involve modifications of the genome of a biological organism caused by external factors as for instance environmental chemicals, development, drugs, aging, and diet, however without the occurrence of mutations of the genetic code of the organism. Epigenetic factors are in general hereditary and can cause diseases as for instance cancer, diabetes, autoimmune diseases and mental disorders. Besides histone modification, which is not considered in the context of this work, one of the most important epigenetic factors is DNA methylation. DNA methylation is a biochemical reaction performed by the enzyme methyltransferase (DNA MTase) that adds a methyl group (CH_3) to the nucleotides cytosine or adenine. This is illustrated in Fig. 2.9. In mammalians, DNA methylation typically occurs at so called CpG sites, a cytosine-phosphate-guanine site that is the recognition sequence for CpG methyltransferases. The enzyme symmetrically methylates the top and bottom strand within the recognition site. Methylation of either the top or the bottom strand is called hemi-methylation. Between 60% and 90% of all CpG sites are methylated in mammals [13]. CpG methylation alters the affinity of several regulatory proteins for binding to operator DNA. Thus, it plays a central role for the regulation of gene expression on the transcriptional level with an inverse correlation between methylated CpG sites and transcriptional activity. Besides methylcytosine, the nitrogen base 5-hydroxymethylcytosine (hmC) was found to be abundant in human and mouse brain [14]. hmC emerges from the enzyme driven oxidation of methylcytosine [15]. It can be found in every mammalian cell, however with strongly different levels depending on the cell type. The exact function of hmC is still not fully elucidated, but there are hints that it may regulate gene expression on the transcriptional level. In bacteria, cytosine or adenine methylation acts as a primitive immune system. The genome is usually fully methylated at specific sites and can thus be distinguished from foreign DNA, which is attacked by methylation sensitive restriction enzymes. In E. coli, methylation of the DNA sequence motif 5'-GATC-3' by the enzyme deoxyadenosine methylase (Dam) plays an important role in the timing of initiation of DNA replication, as well as in coordinating cellular events, DNA mismatch repair, and gene regulation [16]. Dam is an enzyme of 32 kD that adds a methyl group to the sixth nitrogen atom of the nucleotide adenine, within its recognition sequence (see Fig. 2.9). In the following, two transcriptional regulation systems based on Dam and CpG methylation are explained in detail. Both systems are studied in the context of this work.

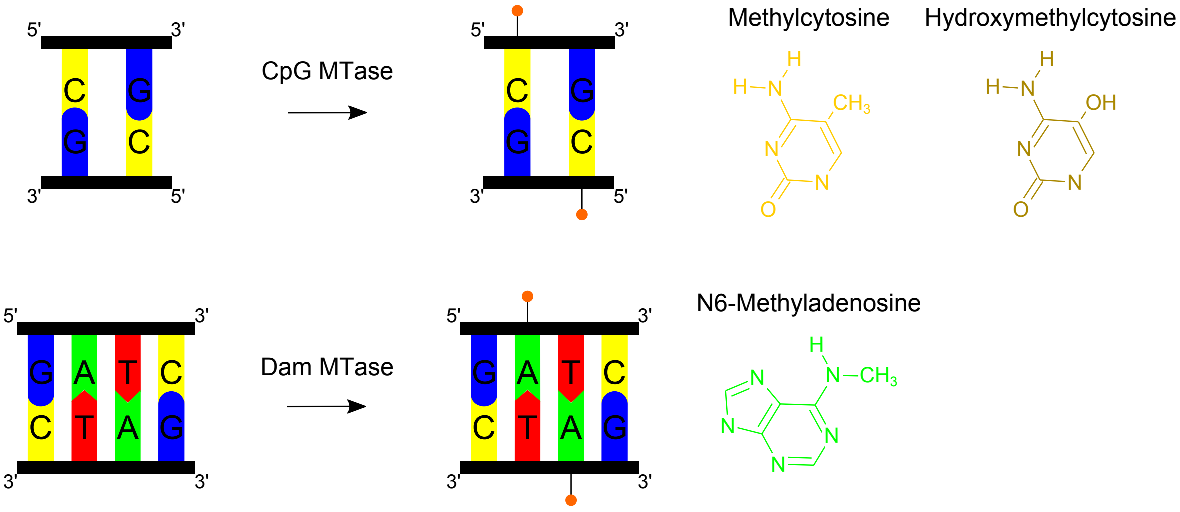


Figure 2.9.: Recognition sequences for CpG and Dam methyltransferases and chemical structure of methylcytosine, hydroxymethylcytosine, and N6-methyladenosin. The orange pin indicates the methyl group in the schematic drawings.

2.4.1. Pyelonephritis-associated pili expression regulation in E. coli

Uropathogenic E. coli (UPEC) bacterias are the main cause of urinary tract infection (UTI) in humans [18]. UTI is among the most common infection diseases in the developed world, with approximately 150 million cases globally per year [19]. The UPEC bacterias attach themselves to epithelial cells of their host bodies by expressing pyelonephritis-associated pili (Pap). Pili are thin, hair-like projections on the surface of the bacteria that can bind to mannose residues on the host cell surface. The expression of Pap is epigenetically regulated depending on the environment of the bacteria. The relevant promoter region is schematically illustrated in Fig. 2.10. The pap operon contains six binding sites for the leucine responsive protein (Lrp, *genbank* ID 949051, [20]). Lrp is a global regulatory protein of 18.8 kD that controls a large number of genes in E. coli. The sequences of the six Lrp binding sites are listed in Table 2.1. The operon 2, proximal to the transcription start of the pap gene, as well as the operon 5 distal to the gene, are each overlapping with a GATC site that is the recognition sequence for the DNA MTase Dam. In phase-OFF cells, Lrp binds cooperatively to binding sites 1, 2, and 3 and thus inhibits the transcription of the pap gene by the RNA polymerase. Additionally, the binding of Lrp protects the GATC^{prox} site located in the

Table 2.1.: Sequences of the six Lrp binding sites of the pap promoter region according to [17]. Only the top strand sequences of the operons are shown.

Lrp binding site	Sequence (5′ → 3′)
1	TTA-GTG-TTT-TGT-TCT-AGT
2	AAC-GAT-CTT-TTA-ACC-CAC
3	TTA-GTT-TTT-TAT-GTT-GTA
4	TAC-GGA-CTT-TCT-TTT-CGC
5	GAC-GAT-CTT-TTA-TGC-TGT
6	CAT-GAT-GTT-TTT-ATC-TGA

operon 2 from methylation by Dam, while GATC^{dist} in the operon 5 is methylated. A roughly understood mechanism that is called “mutual exclusion” prevents simultaneous occupation of all six operons by Lrp [21]. Since methylation decreases the affinity of Lrp for binding to DNA, Dam methylation of GATC^{dist} locks the phase-OFF state. Besides other molecular players and DNA replication [17], switching from phase OFF to ON requires the presence of the local co-regulator protein PapI, coded upstream of the pap promoter region. PapI (*genbank* ID 1039535, [20]) is an unspecific DNA binding protein of 8.7 kD that strongly interacts with Lrp-DNA complexes. It increases the affinity of Lrp for specific binding to operons 4, 5, and 6 [22]. In this case, the pap promoter remains accessible for the annealing of the RNA polymerase and thus the pap gene can be expressed. Additionally, the phase-ON state is stabilized by Dam methylation of GATC^{prox}, which reduces affinity of the PapI mediated binding of Lrp to the proximal operons. Fig. 2.10 also shows the crystal structure of the proteins Lrp and PapI.

2.4.2. The human repressor MeCP2 and its binding domain

Methyl-CpG-binding domain proteins constitute a family of DNA binding proteins that play an essential role in mammalian development. Human proteins MeCP2, MBD1, MBD2, and MBD4 share a structurally similar methylation binding domain (MBD) of about 90 amino acids in length that can specifically bind to methylated CpG sites. In the context of this work, the methylation binding domain (amino acid 78-162, 10 kD) of the transcriptional repressor MeCP2 (*genbank* ID 4204, [20]) is studied with respect to regulating epigenetic gene switches as a function of the methylation state of the regulatory DNA. MeCP2 mediates gene silencing through binding to methylated DNA. The specificity of MeCP2 for methylated DNA

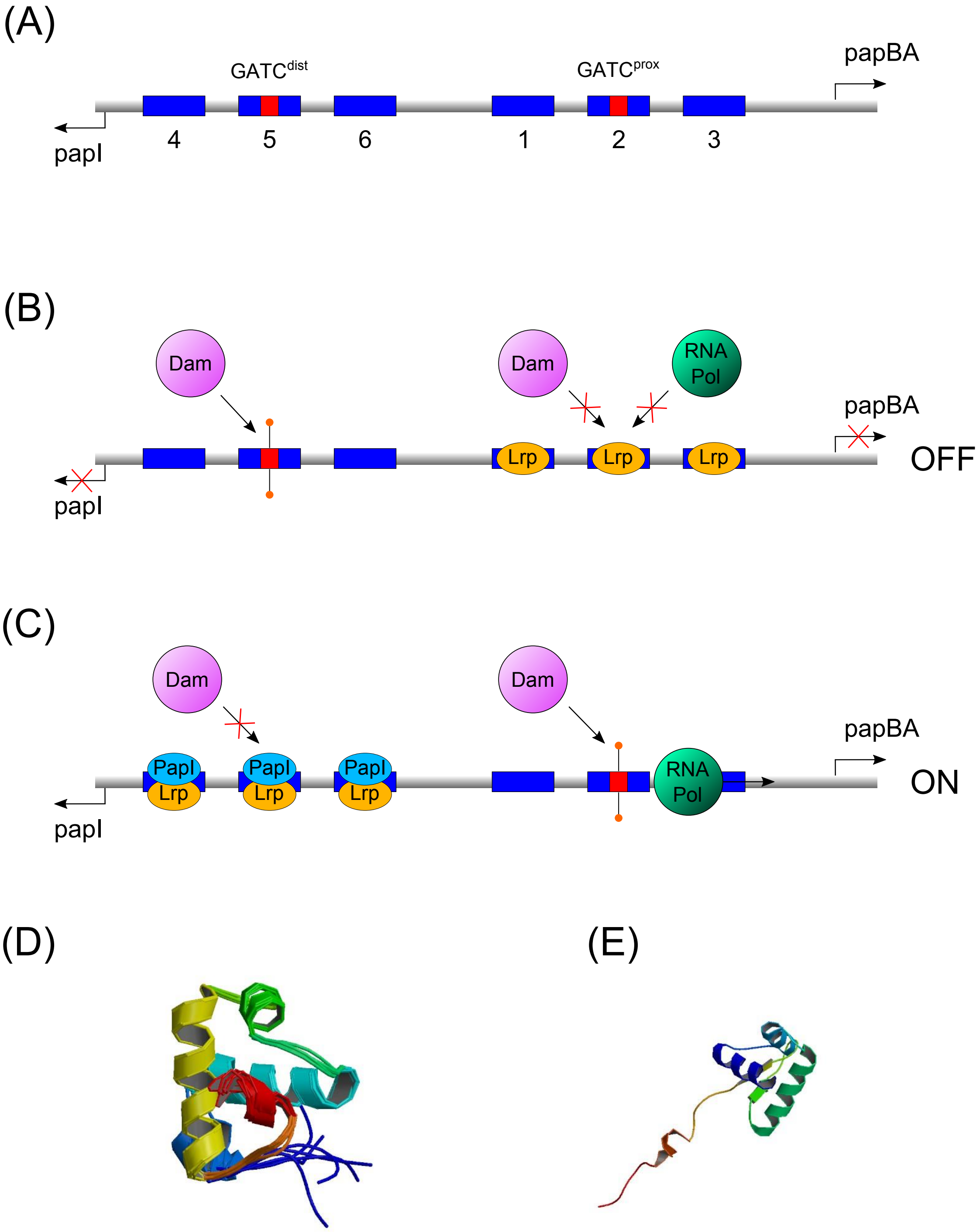
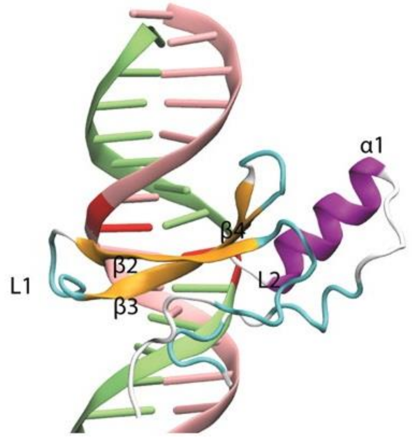


Figure 2.10.: Simplified schematic illustration of the pap phase variation. **(A)**, organization of the pap operon regulating the expression of papBA. The six binding sites for Lrp are shown in blue. The sites 1-3 constitute the proximal promoter region, while the sites 4-6 constitute the distal promoter region. Site 2 and site 5 contain the recognition sequence GATC for the methyltransferase Dam. The gene coding for the local co-repres-

sor PapI is located upstream of the regulatory DNA. **(B)**, OFF state. Lrp binds cooperatively to the proximal promoter sites and blocks the annealing of the RNA polymerase. papBA is not expressed. Lrp prevents site $GATC^{prox}$ located in site 2 from being methylated by Dam, while $GATC^{dist}$ located in site 5 is methylated (orange pin). **(C)**, ON state. A complex consisting of PapI and Lrp binds cooperatively to the distal promoter region and prevents $GATC^{dist}$ from being methylated. The proximal promoter is free for the annealing of the RNA polymerase as well as the methylation of $GATC^{prox}$. **(D), (E)**, crystal structure of Lrp (D) and its co-repressor PapI (E). Both structures are adapted from the *protein data base* [23] (structure ID *2HTJ* and *2L4A*, respectively).

is due to contacts between a hydrophobic patch on the MBD that interacts directly with the DNA methyl groups. hMeCP2 binds to methylated CpG sites in any sequence context and inhibits transcription [24]. MeCP2 possesses highest affinity for binding to the promoter III of the mouse-brain-derived neurodrophic factor (BDNF) gene, which contains a central CpG pair and an AT run [25–27]. The crystal structure of the MBD of MeCP2 bound to its recognition sequence is shown in Fig. 2.11.



5'--T-C-T-G-G-A-A-C_{CH3}-G-G-A-A-T-T-C-T-T-C--3'

Figure 2.11.: Crystal structure and recognition sequence of the methylation binding domain (MBD) of the human repressor MeCP2 (*protein data base*, structure ID *3C2I*, [23]). The methylcytosine is indicated in red. Only the top strand sequences of the operon is shown. The important motifs and secondary structures are labeled: β-strands (β), α-helix (α), and loops (L)). The picture is adapted from [28].

3. Materials and Methods

3.1. DNA hybridization

3.1.1. Oligonucleotides and buffer solutions

All oligonucleotides used in this work are obtained commercially, (*Metabion*, Germany and *Exiqon*, Denmark) HPLC purified. The concentrations of oligonucleotides are analyzed by UV absorption at 260 nm. No stable secondary structures are found with the dinamelt web server for nucleic acid melting predictions [29]. DNA oligonucleotides can be ordered with modifications. The probe molecules are modified with a C6 amino linker at the 5′ end in order to bind them covalently to the dendrimer coated surfaces (see sections 3.1.3 and 3.1.4). In the case of hybridization experiments in bulk, the probes lack the C6 amino linker and are instead labeled with the fluorescent dye cyanine 3 (Cy-3) or cyanine 5 (Cy-5) at the 3′ end. The optical properties of these dyes are discussed in section 3.1.2. The target molecules are modified with the fluorescent dyes Cy-3 or Cy-5 either externally at the 5′ end (for surface based measurements as well as experiments in bulk; see sections 3.1.5 and 3.1.6, respectively) or internally within the DNA sequence (for fluorescence correlation spectroscopy, see section 3.1.7). In order to study the influence of methylation on the DNA hybridization (chapter 5) methylcytosine (C_{CH_3}) or hydroxymethylcytosine (C_{OCH_3}) can also be included into the DNA sequence. In order to additionally stabilize the oligonucleotide, DNA bases can be replaced by locked nucleic acids (LNA, *Exiqon*, Denmark). Exiqon does not provide cyanine modifications and uses its own fluorescent dyes called “563” and “667” (see appendix section A.8 for a comparison between the dyes provided by Exiqon and the cyanine dyes in terms of stabilizing double-stranded DNA). Appendix Table A.1 gives an overview of all the oligonucleotides used in this work for the study of the DNA hybridization. Some of these molecules were already investigated in [30]. For all hybridization experiments, the standard reaction buffer is 3xSSC (0.57 M monovalent ions), if not stated otherwise.

The buffer is prepared with deionized water and pH is adjusted to 7.5. After delivery, lyophilized DNA single strands are solved in deionized water at a concentration of 100 µM, aliquoted by 5 µl, and stored at −22 °C. In order to prepare a 250 nM target stock solution, one aliquot is thawed and appropriately diluted in 3xSSC buffer. The stock can be stored for a time period of at least four weeks at 4 °C.

3.1.2. Properties of the fluorescent dyes Cy-3 and Cy-5

The DNA molecules are modified with the often employed fluorescent dyes cyanine 3 (Cy-3) and cyanine 5 (Cy-5). Fig. 3.1 shows the excitation and emission spectra of the dyes. Cy-3 has an absorption maximum at 550 nm and an emission peak at 570 nm. Cy-5 has an absorption peak at 649 nm and an emission maximum at 670 nm. As shown in earlier measurements, the fluorescence efficiency of both dyes is a linear function of temperature [5]. Thus, in experiments with varying temperature, it is necessary to eliminate the contribution of the temperature dependent efficiency of the dye (see section 3.1.5). Because of the overlap between the emission spectra of Cy-3 and the excitation spectra of Cy-5 (striped area in Fig. 3.1), both dyes can be used in measurements based on Förster resonance energy transfer (FRET, see section 3.1.6 and 3.1.7) as donor and acceptor. FRET relies on the radiationless energy transfer from an excited Cy-3 molecule to a Cy-5 molecule in close proximity of both fluorescent dyes, which can be detected by the fluorescence emission of Cy-5. Here, the FRET efficiency E is a function of the distance r between donor and acceptor following

$$E(r) = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{3.1}$$

where R_0 is the Förster radius. For details about the principles of fluorescence spectroscopy based on FRET see [31] and [32].

3.1.3. Dendrimer coating of glass surfaces

The procedure of coating glass surfaces with phosphorous dendrimers is adapted from [33] with remote differences [5, 30]. Standard microscopy cover slips with a diameter of 20 mm and a thickness of 0.17 mm are used as substrates for probe DNA immobilization, because they are more robust than rectangular ones and can also be mounted

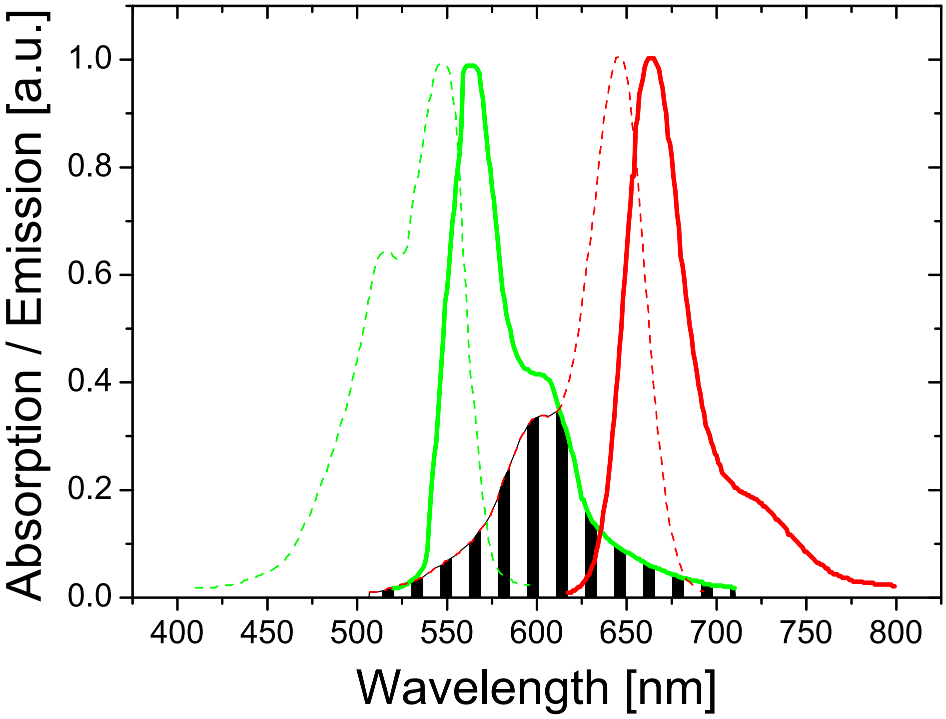


Figure 3.1.: Absorption (dashed line) and emission spectra (solid line) of the fluorescent dyes cyanine 3 (Cy-3, green) and cyanine 5 (Cy-5, red). Due to the overlap between the emission spectra of Cy-3 and the excitation spectra of Cy-5 (striped area), both dyes constitute a donor-acceptor pair and can be used in measurements based on Förster resonance energy transfer (FRET).

into the measurement chamber (Fig. 3.4) very easily. The coverslips are cleaned for 30 minutes in a ultrasonic bath consisting of a detergent solution (5% Deconex), rinsed with deionized water and dried under a stream of nitrogen. In order to remove organic decontaminants and also activate the surface for the subsequent silanization step, coverslips are treated for 30 minutes in a plasma cleaner (air plasma). After that, the slides are silanized with a 10% (v/v) solution of 3-aminopropyltriethoxysilane (APTES, *Sigma-Aldrich*, Switzerland) in a 95:5 (v/v) ethanol/water solution. Silanization is performed at room temperature for 12-15 hours under gentle agitation. After that, the slides are rinsed twice with analytical grade ethanol and once with deionized water and then dried under a stream of nitrogen. Subsequently, the slides are baked at 120 °C for 3 hours. Before the coupling of dendrimers, the substrates are activated in an aqueous solution of a 8% (w/v) potassium hydroxide for 5 minutes, followed by rinsing the slides three times with deionized water and drying them under a stream of nitrogen. Aldehyde-functionalized phosphorus dendrimers (cyclotriphosphazene-PMMH-96 or 24) are dissolved in dichloroethane (0.1% (w/v)). Dendrimers are allowed to couple to the aminosilan for 7 hours under gentle agitation. The dendrimer functionalized surfaces are rinsed with dichloroethane and ethanol (twice in each case) and dried under a stream of nitrogen. The substrates can be stored at 4 °C up to six months under argon atmosphere.

3.1.4. DNA grafting to dendrimer coated surfaces

The procedure of probe immobilization is adapted from [33] with some differences [30]. The probe DNA molecules containing the C6 amino linker are concentrated at 2 µM in a volume of 20 µl of a sodium phosphate buffer (300 mM, pH 9.0). The prepared probe solution is pipetted onto the center of a dendrimer coated glass surface by using a positioning device. The probe molecules are allowed to couple to the dendrimers for 1 hour. In order to avoid the evaporation of the probe solution spot, the immobilization is performed inside a petri dish under an atmosphere of increased air humidity. Fig. 3.2 shows a picture of a dendrimer coated glass surface during immobilization of the probe molecules. After the immobilization, the aldehyde moieties of the dendrimers are reduced to hydroxyl groups. The reduction is performed in an aqueous solution of 0.35% (w/v) sodium borohydride with the addition of 25% (v/v) ethanol for 15 minutes under gentle agitation. After rinsing the slides with deionized water and drying them under a stream of nitrogen, they are ready to use.

3.1.5. TIRF setup for DNA hybridization measurements on surfaces

Total internal reflection fluorescence (TIRF) relies on the evanescent field penetrating into a medium of lower refractive index at the point of total reflection. Here, the intensity I_{ev} of the evanescent wave decreases in an exponential fashion, following

$$I_{ev}(z) = A \cdot e^{-z/d_p} \tag{3.2}$$

where z is the coordinate of the penetration direction and d_p is the penetration depth that typically equals about 100 nm. The appendix section A.2 shows the complete derivation of the evanescent field. The optical setup is shown in Fig. 3.3. A glass

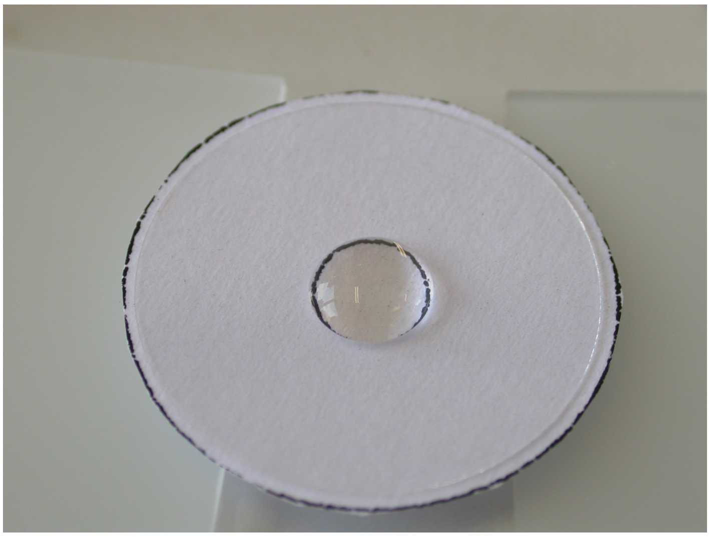


Figure 3.2.: Picture of a dendrimer surface during the 60 minutes immobilization of the probe molecules. The droplet in the center of the surface is the immobilization solution consisting of the probe molecules concentrated at 2 μM. The picture is adapted from [30].

surface with grafted DNA probes (see section 3.1.4) is fixed on a hybridization chamber (Fig. 3.4), which is mounted on the xy-stage of an inverted microscope (Axiovert 135, *Zeiss*, Germany). The excitation beam reaches the cover glass through a dove prism via a layer of immersion oil. To enable simultaneous detection of both fluorescent species during competitive experiments, two lasers of different wavelengths (frequency doubling Nd:YAG laser at 532 nm and helium neon laser at 633 nm) excite the fluorescent dyes Cy-3 and Cy-5, respectively. Before entering the prism, laser beams are expanded and focused into the hybridization chamber. In order to minimize photo bleaching of the dyes during a hybridization experiment, low excitation power is applied. The fluorescence emission is collected by the microscope objective (10x, $NA = 0.25$, *Olympus*, Germany). A beam splitter directs the emission light into two detection systems each consisting of an emission filter (Cy-3 channel: HQ585/40, Cy-5 channel: HQ680/30, *AHF Analysentechnik*, Germany) and a photomultiplier (H9305-04, *Hamamatsu*, Japan). Lock-in amplified signals (7265 Signal Recovery, *Ametec*, USA) are recorded by a PC equipped with a measurement interface. A Pt-100 thermocouple senses the temperature of the hybridization solution and a PID closed loop circuit controls the output to the heating resistors. To avoid a temperature drop towards the observation window, an electrical heating current is applied through the indium tin oxide (ITO) coating of the observation

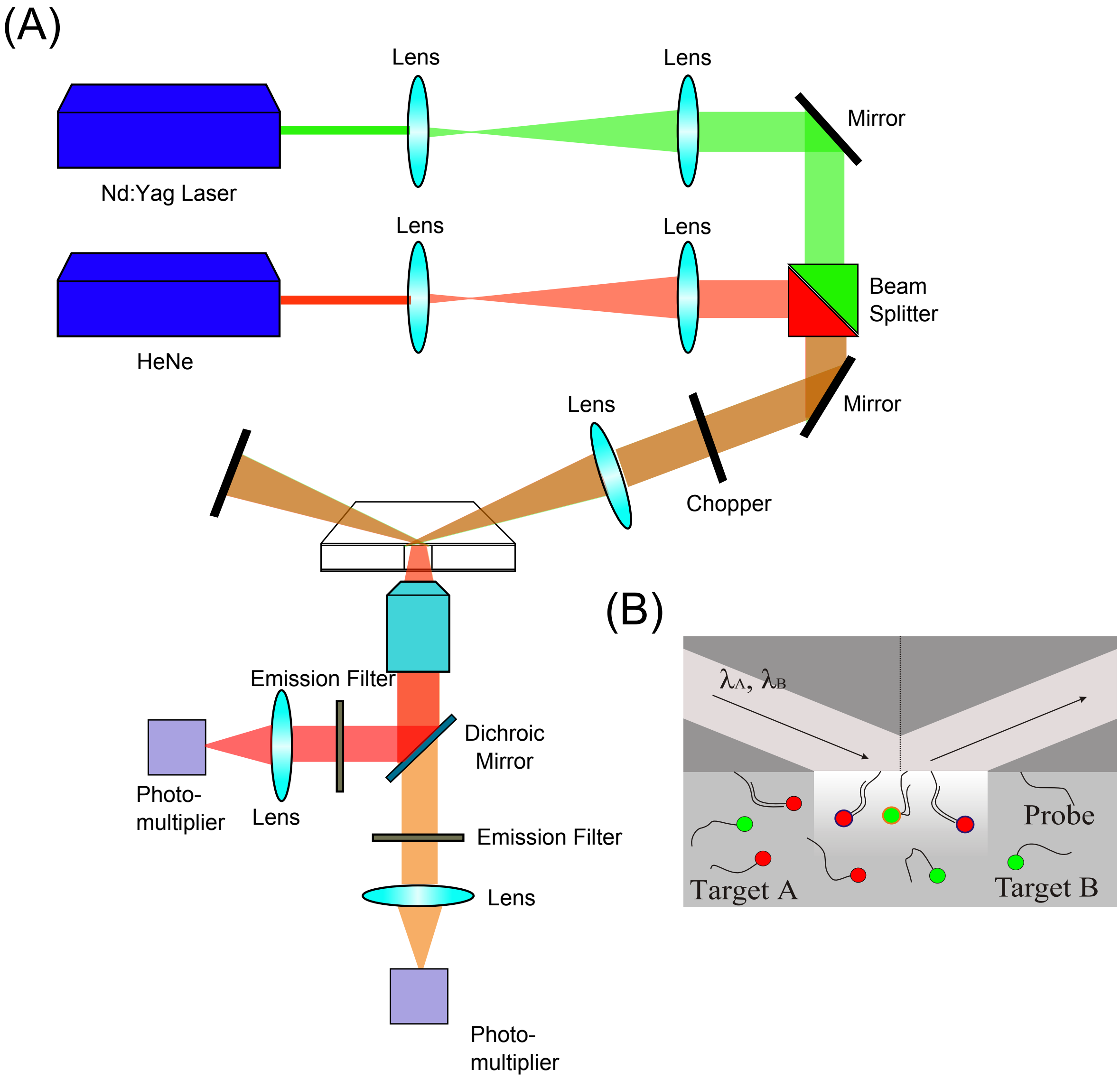


Figure 3.3.: TIRF setup for DNA hybridization measurements on probe grafted surfaces. **(A)**, optical pathway of the laser beams with wavelengths $\lambda_A = 532$ nm and $\lambda_B = 633$ nm. All details are provided in the text. **(B)**, schematic for illustrating the TIRF method during a competitive hybridization experiment between target A and B. The targets are differently labeled with the fluorescent dyes Cy-3 (green) and Cy-5 (red). The evanescent field of the two laser beams with wavelengths λ_A and λ_B penetrates into the measurement chamber at the point of total reflection by typically 100 nm. Only hybridized target molecules are excited.

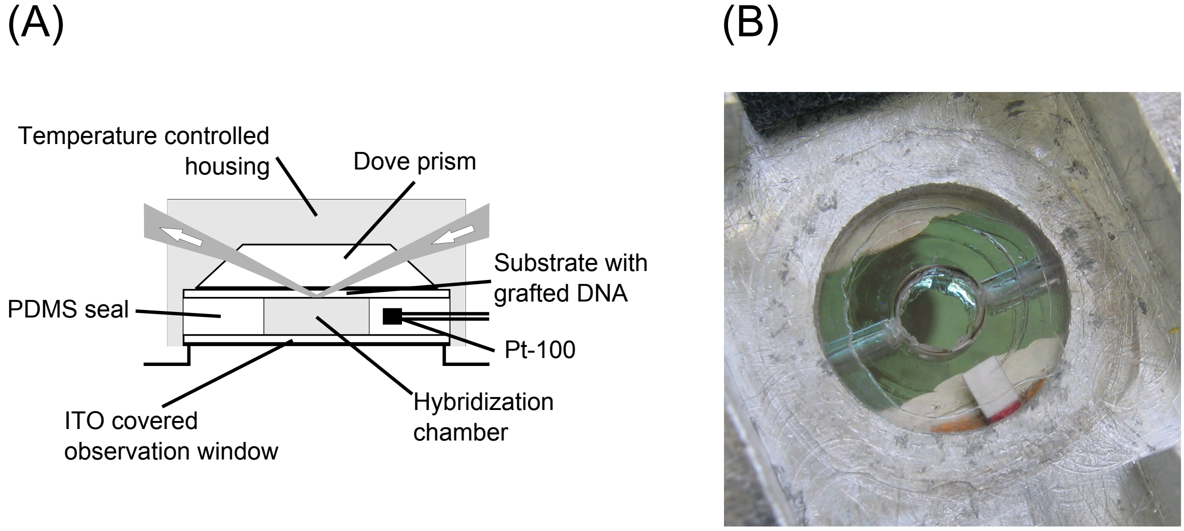


Figure 3.4.: Schematic drawing **(A)** and picture **(B)** of the hybridization chamber. All details are provided in the text. The picture in (B) is adapted from [30].

window. The chamber is made of polydimethylsiloxane (PDMS) and is of circular cross section (viewed from the top) with a diameter of 4 mm and a height of 2.5 mm. Details about the TIRF setup can also be found in [30].

Approach for measuring DNA hybridization kinetics

Before the beginning of an experiment, the target solution is prepared in a standard reaction tube (*Eppendorf*, Germany) by appropriately diluting the 250 nM stock with 3xSSC. A volume of 0.5 ml is adequate to fill the hybridization chamber via openings in the side walls by using a syringe. In the beginning of the experiment, the chamber contains 3xSSC buffer incubated at the desired temperature. The visibility field of the photomultiplier is identical to the visibility field of the microscope ocular and is focused on the center of the hybridization chamber. After switching on the Nd:YAG laser, the evanescent field is scattered by tiny particles in the immersion oil. This scattered light is used to focus the laser beam on the center of the chamber. After starting the measurement software and recording the first data point (offset signal), the buffer is replaced by the prepared target solution. The fluorescence signal is proportional to the amount of hybridized target molecules and therefore defined as the hybridization signal D. It is recorded as a function of time until it remains constant and thermodynamic equilibrium is reached. For quantitative analysis of the DNA hybridization experiments, D is calibrated into units equivalent to the duplex concentration in the hybridization chamber (appendix section A.3). Due to a possible, unwanted energy transfer from Cy-3 to Cy-5 (see section 3.1.2) during competitive hybridization experiments, this calibration is only performed for the Cy-5 channel of the experiment, while the hybridization signal of the Cy-3 channel is not quantitatively analyzed. A cover slip with grafted probe DNA can be used for multiple times. Removal of hybridized targets before the start of a new experiment is performed by filling the chamber with a 10 mM sodium hydroxide solution for 5 seconds at 44 °C (see appendix section A.4).

Approach for measuring denaturation curves

The recorded hybridization signal D consists of two contributions, the amount of hybridized DNA and the temperature dependent efficiency of the dye. Thus, in experiments where D is measured as a function of temperature (denaturation curves), it is necessary to separate the two contributions. Therefore, temperature ramps with the fluorescently labeled variant of the probe molecule immobilized to the surface have been performed in earlier measurements. In this case the number of dyes excited by the evanescent field does not change with temperature and the observed linear dependence of the fluorescent signal on temperature is a property of the dye alone. This contribution determines the linear baseline BL of a denaturation experiment. It is removed following Eq. (3.3) to give the fraction of occupied binding sites Θ:

$$\Theta = 1 - \frac{D}{BL} \tag{3.3}$$

An example for removing the temperature baseline in a denaturation experiment is given in the appendix section A.7.

3.1.6. FRET setup for DNA hybridization measurements in bulk

In order to assess the hybridization of single-stranded DNA molecules in bulk, an experimental setup based on FRET (see section 3.1.2) is used. In this case the probe molecules are not immobilized to a surface but also present in solution. The principle of the bulk measurement is shown in Fig. 3.5. In the case of a competitive hybridization

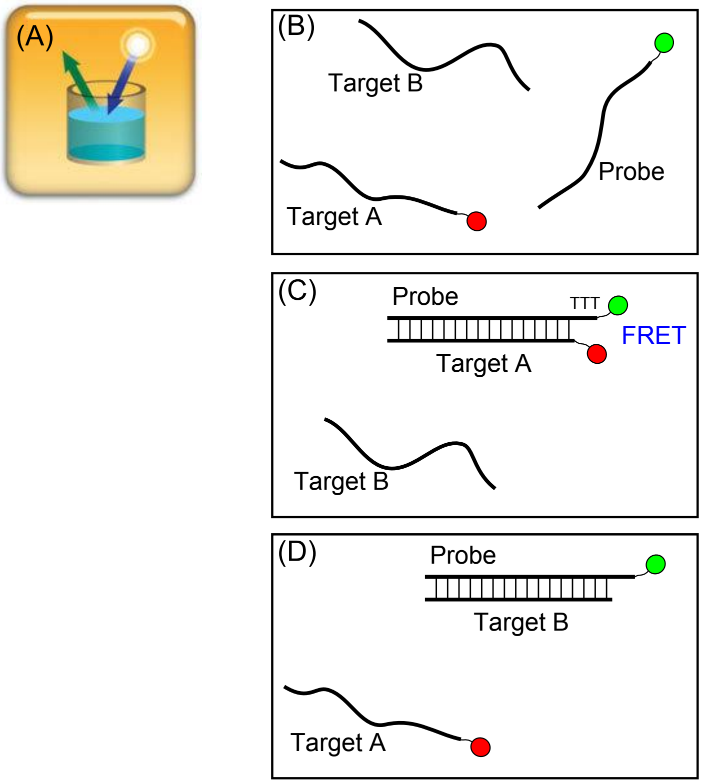


Figure 3.5.: FRET method for competitive DNA hybridization measurements in bulk. **(A)**, schematic of the measurement principle of the plate reader. The Cy-3 molecules in the assay are excited by a xenon flash light with a wavelength of 544 nm and the FRET based emission of either the quenched donor at 580 nm or the acceptor molecule Cy-5 at 670 nm is recorded. **(B)**, target A is tagged with the fluorescent dye Cy-5 at the 5' end of the strand while the probe is tagged with Cy-3 at the 3' end. Target B remains unmodified. **(C)**, after the hybridization of target A to the probe molecule, there is a radiationless energy transfer between Cy-3 and Cy-5. A spacer consisting of 3 thymine bases is used to avoid contact quenching. **(D)**, hybridization of the untagged target B to the probe does not lead to any energy transfer. In order to measure the hybridization of target B in presence of target A, target B is Cy-5 tagged and target A remains unmodified (not shown).

between target A and B, one of the two competing species is tagged with the acceptor Cy-5 at the 5' end of the strand, while the complementary probe species is labeled with the respective donor Cy-3 at the 3' end. A second, competing target molecule B remains unmodified. After duplex formation of the tagged molecules, the Cy-3 emission is quenched due to the proximity to Cy-5 and the resulting energy transfer following Eq. (3.1). The probe molecules used for the bulk measurements are lacking the C6 amino linker at the 5′ end. In order to avoid contact quenching, the distance between the fluorescent dyes Cy-3 and Cy-5 in a DNA duplex is increased by introducing a spacer of three T bases into the probe molecule.

Approach for measuring DNA hybridization kinetics

The three DNA species (target A, target B, and probe) are premixed in a standard reaction tube, which is pre-incubated at the desired temperature. After mixing, 200 µl of the solution is transferred into a non-absorptive 96 well microplate (*Nunc*, Germany). The fluorescent signal is immediately measured by using a plate-reader (*Polar Star Optima*, *BMG Labtech*, Germany), which is also incubated at the desired temperature. 544 nm is used as the excitation wavelength of Cy-3 and either the emission of Cy-3 at 580 nm (donor channel) or of Cy-5 at 670 nm (acceptor channel) is measured.

3.1.7. FCS setup for measuring base pair fluctuation dynamics

Fluorescence correlation spectroscopy (FCS) is a single molecule technique that can address fluctuation dynamics in double-stranded DNA [34]. Fig. 3.6 shows the experimental setup. The probe and target molecules are tagged with the fluorescent dye Cy-3 and Cy-5 and thus constitute a FRET pair (see section 3.1.2). The fluorescent dyes are used as internal modifications within the DNA strand. Due to FRET, the equilibrium base pair fluctuations in the duplex translate into fluctuations in fluorescence intensity at a given wavelength and the correlation function reveals the characteristic relaxation dynamics of the investigated duplex. The excitation light of a frequency doubling Nd:YAG laser (532 nm) is first expanded by a telescope and then coupled into an inverted microscope (Axiovert 135, *Zeiss*, Germany) via the rear site port. The light beam is focused into the sample through the objective lens (C-Apochromat 40X, $NA = 1.2$, water immersion, *Zeiss*, Germany) and the emitted light is collected through the same objective. The excitation light is reflected by a dichroic mirror while the emitted light passes through a Cy-3 filter before it is filtered by a 50 µm diameter pinhole. The emitted light is detected by a photon counter (H8259-01, *Hamamatsu*, Japan) whose signal is fed into a hardware correlator (Flex 99r480, *Correlator.com*). The autocorrelation function as well as the photon flux is recorded as a function of time and displayed on a PC. Low excitation power is used in order to avoid photo bleaching as well as triplet formation of the fluorescent dyes. The contribution of the afterpulsing of the photon

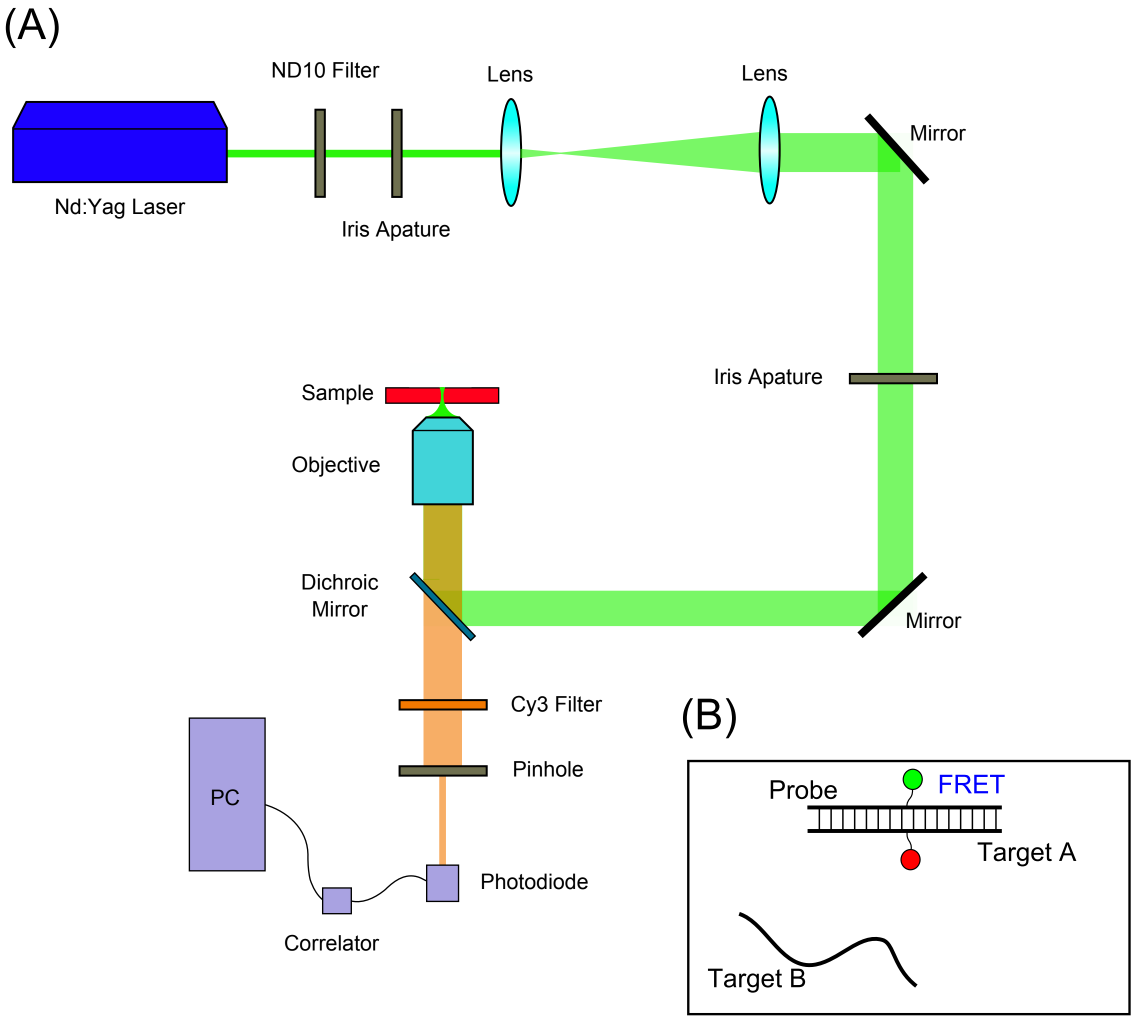


Figure 3.6.: FCS setup for the measurement of base pair fluctuation dynamics in double-stranded DNA. **(A)**, optical pathway. The expanded light from a frequency doubling Nd:YAG laser (532 nm) is focused into a measurement chamber with the objective lens. The chamber contains the interacting DNA molecules at specified concentrations and at controlled temperature. The emitted light of the fluorescent dye Cy-3 passes through a dichroic mirror as well as a Cy-3 filter, before it is filtered through a 50 µm diameter pinhole. The fluorescence fluctuations are read by a photon counter whose signal is fed into a correlator. The signal is transferred to a PC. **(B)**, schematic of labeling the DNA molecules for a FCS measurement. The probe is internally tagged with Cy-3 while the target is modified with Cy-5. Due to FRET, the base pair fluctuations of the duplex translate into fluorescence fluctuations. For measuring the fluctuation dynamics of a hybridized target A in the presence of a competing target B, target B remains unmodified.

counter is at least one order of magnitude faster (approximately 3 µs) than the specific signal of the base pair fluctuations. Thus, afterpulsing does not need to be taken into account. The measurement chamber consists of polydimethylsiloxane (PDMS) and is covered by two glass slides at the top and the bottom. It has the same geometry as the hybridization chamber explained in section 3.1.5. The temperature of the chamber is read by a Pt-100 thermocouple and stabilized by a controller regulating the output to a heating foil.

Approach for measuring base pair fluctuation dynamics

Before the start of the measurement, the DNA molecules are premixed at specified concentrations in a standard reaction tube. A 2-fold excess concentration of targets (20 nM) over probes (10 nM) is used to ensure that every probe molecule is hybridized to a target. The tubes are incubated in boiled water, which is allowed to cool down to room temperature over a time period of 12 hours. This guarantees hybridization of targets and probes. In order to start the experiment, the DNA solution is transferred into the measurement chamber via openings in the side walls by using a syringe. The measured correlation function $G_{meas}(t)$ consists of two contributions, the correlation $G_{diff}(t)$ due to the diffusion of the DNA molecules through the focused laser beam in the measurement chamber, and the wanted correlation $G(t)$ of the base pair fluctuation dynamics. In order to eliminate the contribution of the diffusion, the measured correlation function $G_{meas}(t)$ is divided by G_{diff} measured on double-stranded molecules lacking the quencher Cy-5:

$$G(t) = \frac{G_{meas}(t)}{G_{diff}(t)} \tag{3.4}$$

All correlation functions presented in this work are mean values of at least four individual experiments measured on four identical DNA solutions. An detailed example for measuring the correlation function of the base pair fluctuation dynamics in double-stranded DNA including the elimination of the contribution of the diffusion is given in the appendix section A.9.

3.2. In vitro gene expression

3.2.1. DNA plasmids

Plasmid DNA is the template for most of the *in vitro* expression experiments performed in this work. A plasmid or vector is a closed, circular, double-stranded DNA molecule that is, however, not part of the bacterial chromosome. It contains the relevant gene for the protein to express, a promoter region to regulate the gene expression, and an origin of replication. If the origin of replication is compatible to the bacterial strain, the plasmid can be replicated independently of the chromosomal DNA. It further contains an antibiotics resistance gene that provides a selection advantage for the host bacterium. Fig. 3.7 shows the general structure of a DNA plasmid including the most important regions.

All plasmids used in the context of this work stem from the vector pBEST-Luc (*Promega*, USA). pBEST-Luc contains a ColE1 origin of replication that produces a few hundred plasmid copies per E. coli cell. In order to avoid toxicity due to an over-expression of the recombinant protein during plasmid amplification, the ColE1 origin can be replaced by a p15A origin with about 10-15 copies per E. coli cell. A more general solution for this problem is reported in [35]. Using the strong lambda phage promoter Pr flanked by the operons OR1 and OR2 enables the temperature sensitive lambda repressor Cl857 to stabilize the plasmid during cell growth and thus avoid toxicity. The principle of the plasmid design involves cloning of the relevant genes and regulatory parts into pBEST-Luc plasmids according to procedures of molecular biology (see section 3.2.2). All cell free experiments are based on the expression with the E. coli housekeeping transcription factor σ^{70}. Thus, the relevant regulatory parts are cloned to overlap with the -35 and -10 recognition sequences of σ^{70}. The untranslated region UTR1, containing the T7 g10 leader sequence, enables highly efficient translation initiation and is therefore included behind the promoter region [35]. This is followed by inclusion of the gene coding for the molecular players Lrp, PapI, or the methylation binding domain of MeCP2 (see chapter 2.4) or the reporters enhanced green fluorescent protein (eGFP) truncated and modified in N-terminal [35], or mCherry. The efficiency of gene expression is increased by introducing the strong transcription terminator T500. All plasmids contain a resistance gene for the antibiotic ampicillin. Appendix Table B.1 gives an overview of all the plasmids used in the context of this work. In the following, the procedure of plasmid design and amplification is explained in detail.

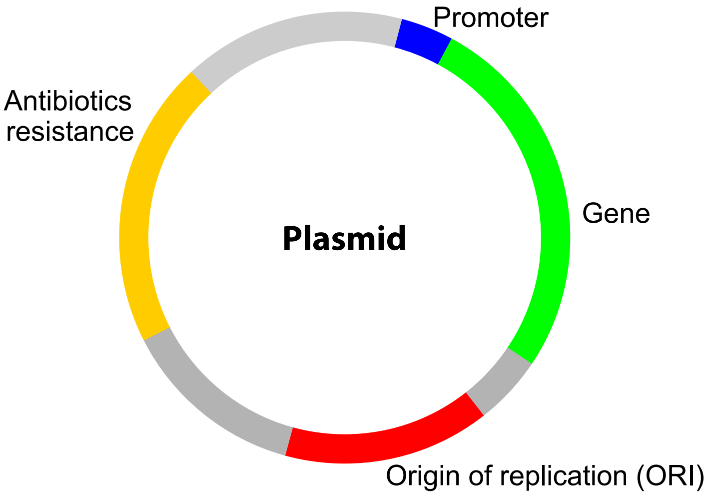


Figure 3.7.: Schematic illustration of a circular DNA vector including the promoter, the respective gene which is to express, the origin of replication (ORI), and a gene for the antibiotics resistance.

3.2.2. Methods of molecular cloning

PCR

In order to clone genes into DNA vectors, they have to be amplified from template DNA, as for instance chromosomes or other plasmids. Therefore, polymerase chain reactions (PCR) are performed with a standard thermo cycler (PCR sprint, *Thermo Life Science*, USA) and the Taq PCR Kit (*New England Biolabs*, USA). The principle of PCR is schematically illustrated in Fig. 3.8. During the denaturation step of the PCR, the double-stranded DNA template is denatured at a temperature of 95 °C for 30 seconds. An initial denaturation of five minutes ensures complete melting of all duplexes. After that, temperature is decreased to 56 °C in order to enable hybridization of so called primers. Primers are single-stranded DNA oligonucleotides of about 20 bases in length that are complementary to the 3' ends of each of the sense and anti-sense strand of the gene that is to amplify. Thus, a standard PCR is performed with two primer species, the sense primer and the anti-sense primer, which both have to be designed before. All primers used in this work are obtained commercially (*Metabion*, Germany). After primer annealing, temperature is again increased to 72 °C in order to initiate the elongation step of the PCR. Here, the enzyme DNA polymerase uses base pair complementarity of the template strand to synthesize a new single strand. This is done by incorporating so

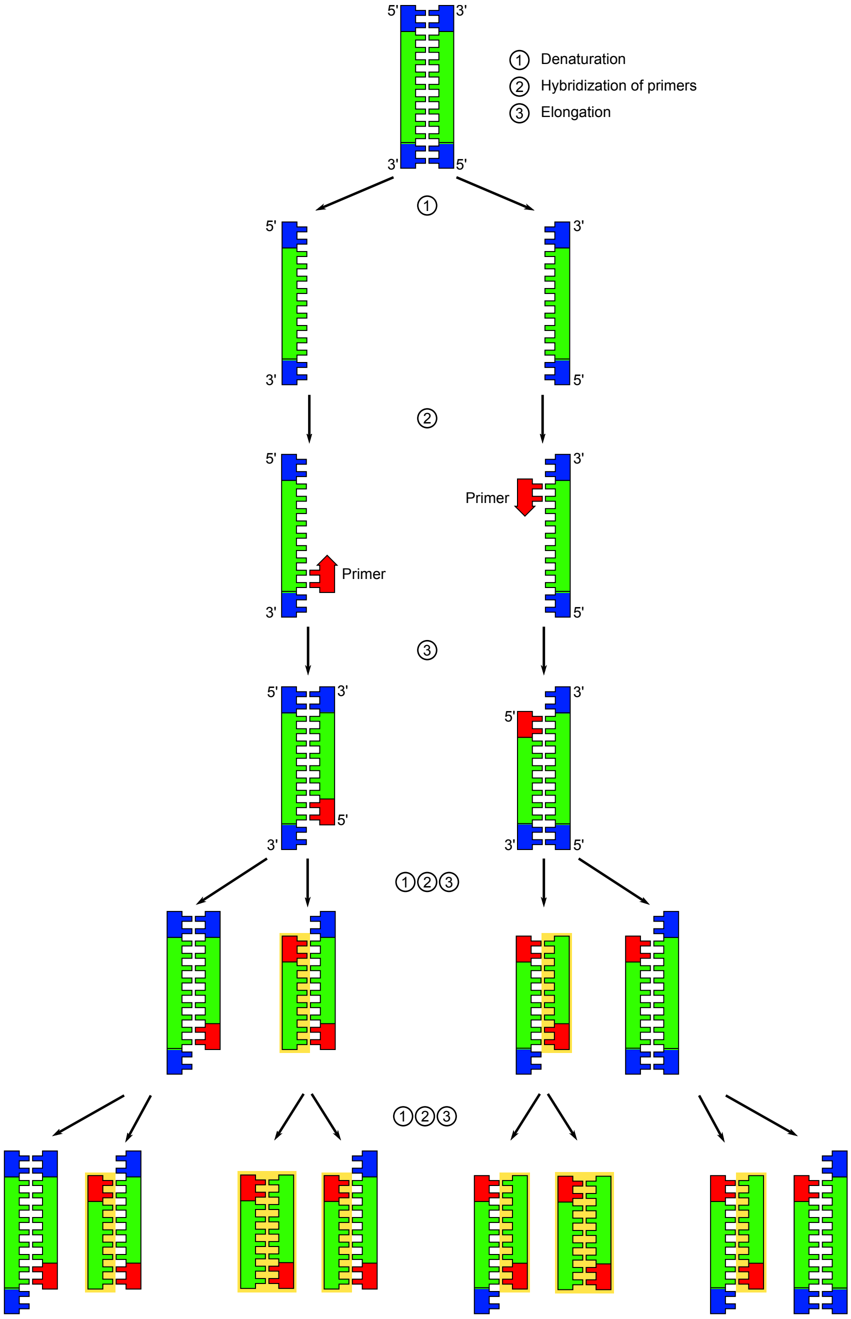


Figure 3.8.: Schematic illustration of the polymerase chain reaction (PCR). After denaturation of the template DNA, single-stranded primers are allowed to hybridize to the 3′ end of the top and bottom strand of the template. The enzyme DNA polymerase then uses base pair complementarity in order to string together dNTPs, thus producing a DNA replicate perfectly complementary to the template. The replicate does also act as a template for the subsequent amplification cycles. The figure is adapted from [36].

called dNTPs (ATP, CTP, GTP, TTP) complementary to the template in 5′ - 3′ direction. The 5′ phosphate group of the dNTP is condensed with the 3′ hydroxyl group at the end of the nascent DNA strand. The duration of the elongation step depends on both, the size of the gene that is to amplify and the polymerization speed of the enzyme. The Taq-polymerase (*New England Biolabs*, USA), used in the context of this work, synthesizes about 1000 bases per minute. Denaturation, primer annealing and elongation constitute 1 cycle of the PCR. In order to reveal a high DNA yield, a typical PCR consists of about 20 cycles. In order to clone the amplified gene into the plasmid later on, the primers also contain the recognition sites for so called restriction endonucleases. The PCR products are purified after PCR amplification with a purification kit (PureLink, *Fisher Scientific*, USA). This is to remove the primers, dNTPs, and reaction buffers used during the PCR. After the purification, the PCR product is solved in deionized water for downstream processing.

Gel electrophoresis

Gel electrophoresis is used for DNA fragment analysis and separation. A mixture of negatively charged DNA fragments of different sizes migrates through the matrix of a 0.8% agarose gel by applying an electric field. Shorter DNA fragments migrate faster than larger ones, so that the molecules can be separated by their size (see Fig. 3.9). The agarose is boiled in a microwave oven and then casted into a gel cassette that contains a template comb to form wells. Agarose is allowed to gelify for one hour at room temperature. It is then loaded into an electrophoresis bath filled with 1xTBE buffer. The DNA samples are mixed with a gel loading dye (GLB) in a standard reaction tube and then pipetted into the wells. Electrophoresis is started by applying a voltage of 140 V to the bath for 30 minutes. After migration, DNA molecules are stained with ethidium bromide in order to make them visible under ultraviolet light. The fluorescent bands of the DNA sample can be compared with molecular weight markers (Marker VII, Marker XVII, *Roche*, Switzerland; 2-log ladder, *New England Biolabs*, USA) containing DNA fragments of defined sizes and quantities. Extraction of separated DNA fragments is performed by cutting out the respective DNA bands under ultraviolet light with a scalpel and subsequent purification with the PureLink gel extraction kit (*Fisher Scientific*, USA). After gel extraction, the DNA fragments can be used in the downstream cloning process.

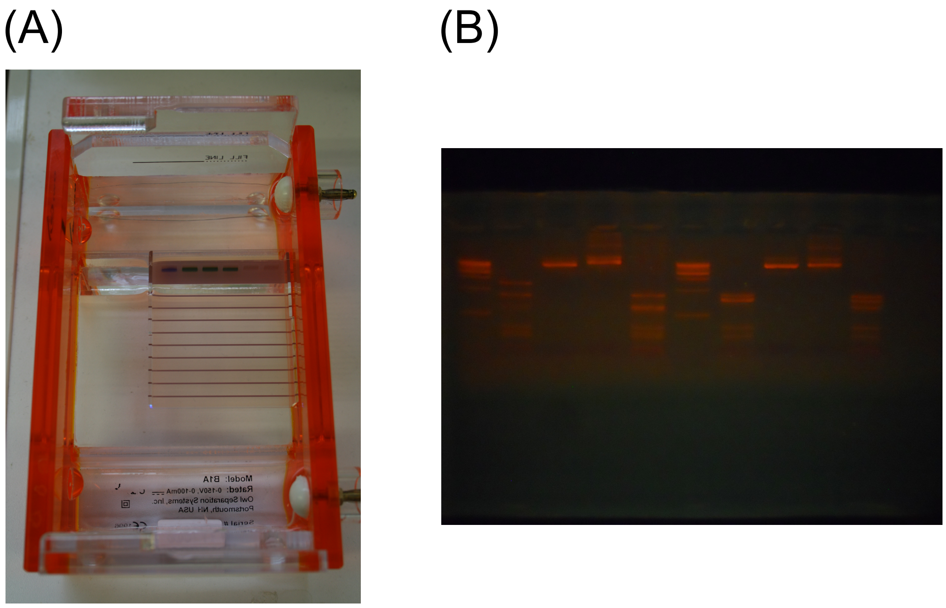


Figure 3.9.: Gel electrophoresis. **(A)**, picture of the electrophoresis chamber with the loaded gel cassette. The DNA solution is pipetted into wells at the top end of the gel loading cassette. Applying an electric field to the chamber causes migration of the DNA fragments through the gel and separation of the molecules as a function of their size. **(B)**, read out of the gel. The DNA fragments are stained with ethidium bromide that makes the DNA visible under UV light.

Restriction digest of DNA

Restriction of double-stranded DNA is an enzymatic reaction that cleaves the DNA molecule at a specific site. Fig. 3.10 schematically illustrates the reaction using the example of a double digestion by the restriction endonucleases NheI and XhoI. The endonuclease specifically binds to its recognition sequence (restriction site) and thus only cleaves the DNA at that specific site. Restriction ensures that all DNA fragments, which contain the recognition sequence, have the sequence located at exactly the same site. There are various enzymes available and all of them are purchased commercially (*New England Biolabs*, USA). Circular DNA plasmids have to be linearized by restriction digest before analyzing the size of the plasmid. This is because plasmids do not migrate linearly during gel electrophoresis due to super coil formation and thus cannot be compared with quantification markers on the gel. Some of the enzymes are blocked or activated by Dam or CpG methylation, which enables evaluation of the methylation state of the DNA by gel electrophoresis. In order to include a PCR amplified gene

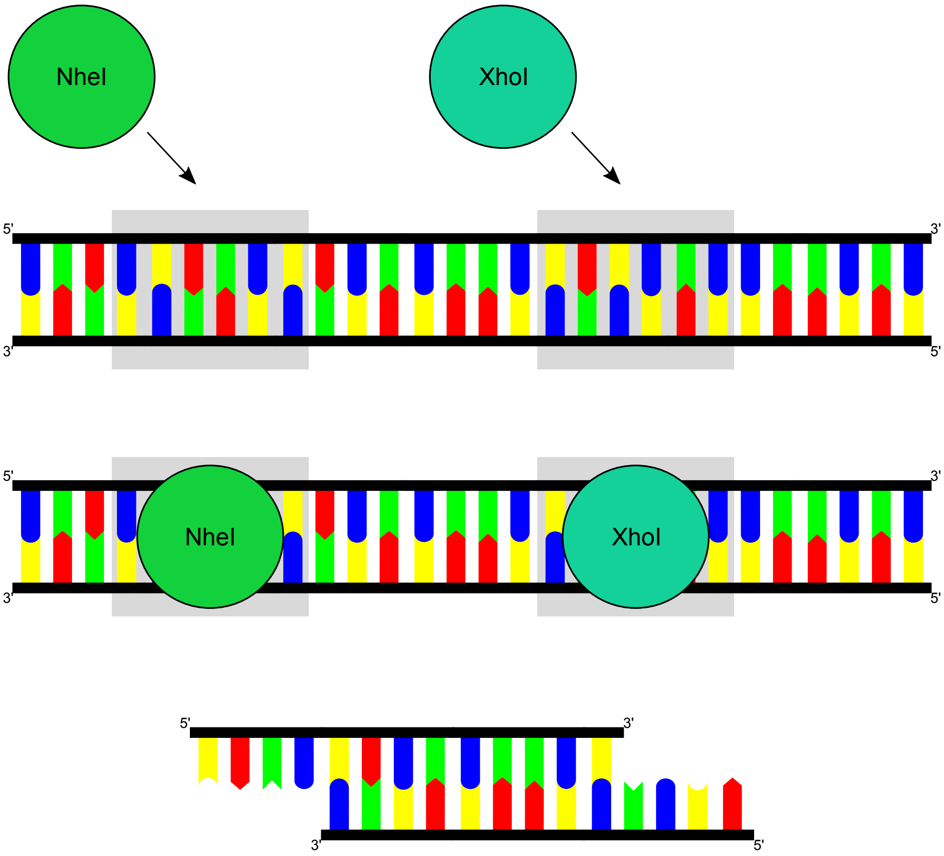


Figure 3.10.: Schematic illustration of a double restriction digest using the example of the endonucleases NheI and XhoI. The enzymes cleave the DNA at specific sites (gray boxes) and the resulting DNA fragments can be separated by gel electrophoresis.

into a plasmid, both, PCR product and plasmid are digested with the same restriction enzymes and the emerging DNA fragments are separated by gel electrophoresis. This produces perfectly matching DNA ends and enables subsequent ligation of the DNA fragments. Restriction of DNA with two different enzymes at the same time is called *double digestion*. *New England Biolabs* provides a double digest finder for their enzymes to determine the compatibility and adequate buffer conditions for the reaction. A typical digestion is performed in a standard reaction tube in a water bath incubated at 37 °C for 1-3 hours. Some special restriction enzymes require different temperatures and the addition of the bovine serum albumin (BSA). One distinguishes between enzymes that produce so called blunt ends and such enzymes that produce sticky ends. In a blunt-ended molecule, both single strands of the DNA duplex terminate in one base pair. A sticky-ended molecule possesses a single-stranded overhang of unpaired bases in the end of the duplex. Restriction enzymes that produce sticky ends are preferred for the downstream cloning process, since they decrease the error rate during ligation.

Ligation

Ligation is an enzymatic reaction that facilitates the joining of two DNA fragments. Here, the enzyme T4 DNA Ligase (*New England Biolabs*, USA) links the 5′ end of one single strand to the 3′ end of the second single strand by formation of a phosphodiester bond. Fig. 3.11 schematically illustrates the ligation of double-stranded DNA. The T4 ligase is most active at 25 °C. However, for a ligation reaction with sticky ends, the activity temperature of the enzyme has to be balanced with the melting temperature of the DNA fragments to be ligated. Thus, a typical ligation reaction is performed in a standard reaction tube in a water bath incubated at 16 °C for 12 hours. Ligation is used to incorporate genes or regulatory parts into an existing DNA plasmid that has been opened by restriction digest before. Typical DNA concentrations for a standard ligation reaction are $0.3 - 0.5$ nM for the plasmid fragment and $3 - 5$ nM for the gene or the regulatory fragment to be included. Only successfully ligated, circular plasmids can be transformed into E. coli cells for the purpose of amplification, since exposed 3′ ends are degraded by the primitive immune system of the bacterium.

Bacterial amplification of plasmid DNA

The designed plasmids with the included genes and regulatory parts are transformed into E. coli chemically competent cells for the purpose of amplification. Fig. 3.12 schematically illustrates the bacterial amplification of plasmid DNA. The typical bacterial strains used in the context of this work are JM109 (*Promega*, USA) for standard amplification, KL740 (*Coli Genetic Stock Center*, USA) expressing the lambda repressor Cl857 for plasmids containing the promoter OR2-OR1-Pr, and Dam-/Dcm- (*New England Biolabs*, USA) for growing DNA plasmids free of Dam methylation. The E. coli cells are mixed with the plasmid DNA in a standard reaction tube and incubated on ice for 10 minutes. Heat-shocking the cells for 20 seconds in a water bath incubated at 37 °C opens pores in the cell membrane and thus enables inclusion of the plasmid into the bacterium. After transformation, cells are allowed to heal in lysogeny broth (LB) medium for 40 minutes under agitation. E. coli cells are then cultured on a LB agar plate containing the antibiotic ampicillin. The standard incubation temperature

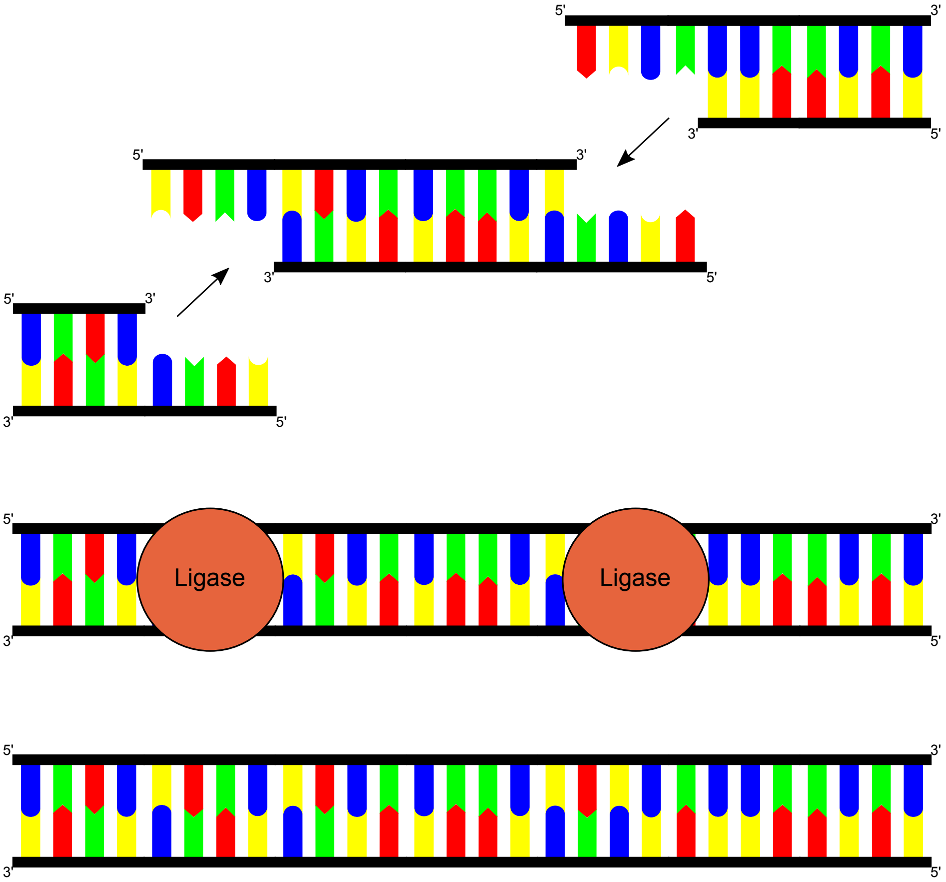


Figure 3.11.: Schematic illustration of a sticky end DNA ligation. After hybridization of the DNA fragments, the ligase enzyme links the 5′ end of one single strand to the 3′ end of the second single strand by formation of a phosphodiester bond.

is 37 °C for JM109 and Dam-/Dcm- competent cells and 29 °C for KL740, since the lambda repressor Cl857 expressed by the KL740 strain is temperature sensitive [35]. After cell growth, the E. coli colonies are transferred from the plate into mini culture tubes containing 5 mL of LB medium with ampicillin. Bacterias are allowed to grow for 8 hours at 37 °C (JM109, Dam-/Dcm-) and for 10 hours at 29 °C (KL740), respectively. Afterwards, the E. coli cell culture is transferred into a standard reaction tube and cells are harvested by centrifugation. The cell pellets are resuspended in a resuspension solution belonging to the GeneElute Mini Prep Kit (*Sigma-Aldrich*, Switzerland), which is used for extracting the plasmids. Bacterial cells are dissolved with a lysis buffer and the plasmid DNA is adsorbed onto a silica column in the presence of high salts. All contaminants including genomic DNA or RNA are removed by two spin wash steps.

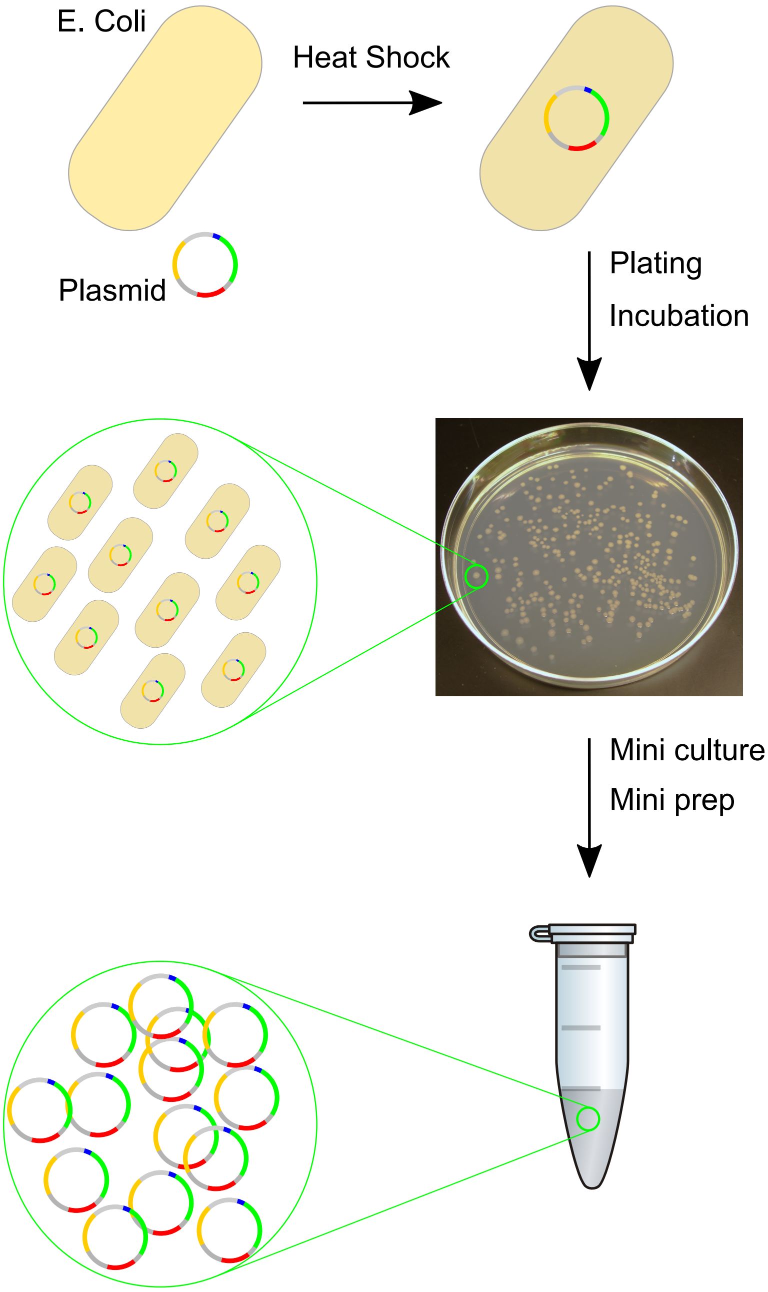


Figure 3.12.: Bacterial amplification of plasmid DNA. The ligated plasmid is transformed into chemically competent E. coli cells via heat shock. After healing, the cells are spread on an agar plate which is incubated for one night. After that, bacterial cells are further amplified by growing them in a nutrition medium (mini culture). Plasmid DNA is extracted from the E. coli cells by performing a mini prep.

Finally, the DNA is eluted in deionized water and recovered in a standard reaction tube that can be stored at 4 °C. Picogreen (*Invitrogen*, USA) or QuantiFluor (*Promega*, USA) is used for the determination of the DNA concentration on a plate reader (Polar Star Optima, *BMG Labtech*, Germany). The DNA concentration can be increased by evaporation of the solvent using a heated vacuum centrifuge (Conentrator 5301, *Eppendorf*, Germany). The sequence of all designed DNA plasmids are analyzed by Sanger sequencing performed by an external vendor (*Microsynth*, Switzerland). Sequenced clones are used as templates for further bacterial amplification of plasmid DNA.

In vitro methylation

In order to study the epigenetic regulation of protein expression with the cell free extract, the template DNA has to be methylated after amplification. Using the E. coli strains JM109 or KL740 for amplifying plasmids should produce Dam methylated DNA in the first place. However, it was observed in the context of this work that the global E. coli regulator protein Lrp binds to DNA even during plasmid amplification and thus prevents some Dam sites from being methylated. Additionally, E. coli is incapable of producing CpG methylation *in vivo*. Thus, methylation of the DNA is done *in vitro* using recombined methyltransferases Dam for GATC methylation as well as M.SssI or HhaI for CpG methylation (*New England Biolabs*, USA). The DNA to be methylated is incubated with the methyltransferases at 37 °C for three hours in a water bath and subsequent PCR purification gets rid of all contaminants. Complete methylation of all recognition sites of the methyltransferase is analyzed by restriction digest with methylation dependent restriction endonucleases DpnI (requires Dam methylation of a GATC site), MboI (blocked by Dam methylation of a GATC site), and HhaI (blocked by CpG methylation of a GCGC site) and subsequent gel electrophoresis. Analyzing the methylation state of template DNA is shown in the appendix section B.2.

3.2.3. Preparation of the crude extract

The protocol for cell free extract preparation is completely adapted from [35]. It is a modification of the protocol presented in [37]. The detailed protocol can be found in the appendix section B.5. All the equipment (glassware, reaction tubes, etc.) that comes in contact with organic material is washed thoroughly and sterilized. All buffer solutions are autoclaved at 120 °C for 15 minutes. The crude extract is prepared with an E. coli

BL21 strain (Rosetta2, *Novagen*, Germany) carrying a plasmid with an antibiotics resistance gene for chloramphenicol. The bacterias are grown under agitation at 37 °C in six 2 *l* Erlenmeyer flasks filled with 500 ml 2xYT medium (*Sigma-Aldrich*, Switzerland) containing 34 mg/ml of chloramphenicol. All the following steps are performed on ice or at 4 °C unless stated otherwise. Cell growth is stopped at $OD_{600} = 1.5$ (measured with the photospectrometer Carry, *Varian*, USA) after usually 3 hours and cells are harvested by centrifugation at 4425 x *g* for 12 minutes. The cells are washed twice and resuspended with S30 buffer A, consisting of 50 mM Tris, 60 mM potassium glutamate, 14 mM magnesium glutamate, and 2 mM DTT at a pH value of 7.7. After washing, the cell membranes are broken with a bead beater (mini-bead-beater, *Biospecs*, USA) using 0.1 mm glass beads. Usually, 4.5 g of E. coli cells are mixed with 4.05 ml of S30 buffer A and 23 g of dry beads in a 50 ml falcon. The mixture is homogenized by thoroughly vortexing the viscous solution and then transferred into 1.5 ml screw cap tubes. Bead beating is performed twice at 4600 rpm for 30 seconds. The beads are removed from the extract by centrifuging the mixture through chromatography columns (Micro Bio-Spin, *Biospecs*, USA) at 6805 x *g* for ten minutes. The extract is then centrifuged at 12000 x *g* for 20 minutes followed by incubation of the clear supernatant at 37 °C for 80 minutes. After that, the extract is centrifuged again at 12000 x *g* for 10 minutes and then dialyzed for 3 hours against S30 buffer B (50 mM Tris, 60 mM potassium glutamate, 14 mM magnesium glutamate, 1 mM DTT, pH value 8.2) using 10 kD MWCO dialysis cassettes (Slide-A-Lyzer, *Thermo Life Science*, USA) under permanent agitation. The recovered extract is aliquoted by a volume of 30 µl in standard reaction tubes and shock frozen in liquid nitrogen. The protein concentration of the extract is determined by preparing a Bradford assay that is analyzed with the photospectrometer. The typical protein concentration of the extract is $27 - 30$ mg/ml. The extract can be stored at -80 °C over a time period of at least one year.

3.2.4. Batch mode cell free reaction

The cell free extract used in the context of this work is entirely endogenous. Transcription and translation is preformed by the molecular machinery present in the extract and there is no need to add additional enzymes. As mentioned above, all cell free experiments are based on the expression with the E. coli housekeeping transcription factor σ^{70} and E. coli core mRNA polymerase present in the extract with a concentration of 35 nM and 100 nM, respectively [35]. The principle of a batch mode cell free reaction is to clone the relevant regulatory parts overlapping the -35 and -10 recognition sequence of σ^{70} in such a way that they regulate the expression of a reporter gene (eGFP or mCherry) in a reaction tube (batch mode expression, see Fig. 3.13). The epigenetic regulators Lrp, PapI, and the methylation binding domain of hMeCP2 (amino acid 78-162), studied in the context of the work, are expressed in the same reaction tube and can interact with the regulatory promoters as a function of the methylation state of the DNA. MBD is also obtained commercially (*Biomol*, Germany) in its recombined form. Measuring the fluorescence intensity of the reporter as a function of the reporter plasmid concentration or the regulator plasmid concentration gives information about the methylation dependent interaction between proteins and DNA. A typical cell free reaction consists of 33% (v/v) extract and 66% (v/v) reaction buffer including the DNA templates. The reaction buffer is composed of 50 mM HEPEs (pH value 8), 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/ml tRNA E. coli, 0.26 mM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-posphoglyceric acid (PGA), 1 mM DTT, and 2% (v/v) PEG8000 (all *Sigma-Aldrich*, Switzerland) [35]. Addition of amino acids ($0.5 - 1.5$ mM of each of the 20 amino acids), magnesium glutamate ($0 - 10$ mM), and potassium glutamate ($0 - 120$ mM) are adjusted depending on the extract quality and plasmid concentration. The system is prepared in such a way that the concentrations of these three components can be adjusted independently for any reaction. The recombined lambda phage protein gamS has to be added to the reaction (3.3 μM) when using non-plasmid DNA templates such as PCR products. A typical batch mode reaction is prepared on ice by thoroughly mixing extract, buffer and template DNA at specified concentrations and then split the solution into volumes of 6 μl. In this case, diffusion of oxygen into the extract is not a limiting factor [35]. The reaction tubes are then incubated at 29 °C for at least 12 hours. After expression has finished, 10 μl of the sample is transferred into a 384 multi well plate and the fluorescence intensity of the recombined reporter proteins eGFP or mCherry is read out by using a plate reader (Polar Star Optima, *BMG Labtech*, Germany). The fluorescence signal is calibrated into units equivalent to the concentration of the recombined reporter proteins in the expression reaction. Fig. 3.14 shows the absorption and emission spectra of the two reporter proteins. An excitation wavelength of 485 nm and 595 nm is used for eGFP and mCherry, while the fluorescence emission is collected at 520 nm and 620 nm, respectively. Due to the overlap between the emission spectra of eGFP and the excitation spectra of mCherry (striped area in Fig. 3.1), both dyes constitute a donor-acceptor pair for FRET (compare section section 3.1.6 for the example of Cy-3 and Cy-5) with a Förster radius of 51 Å [38].

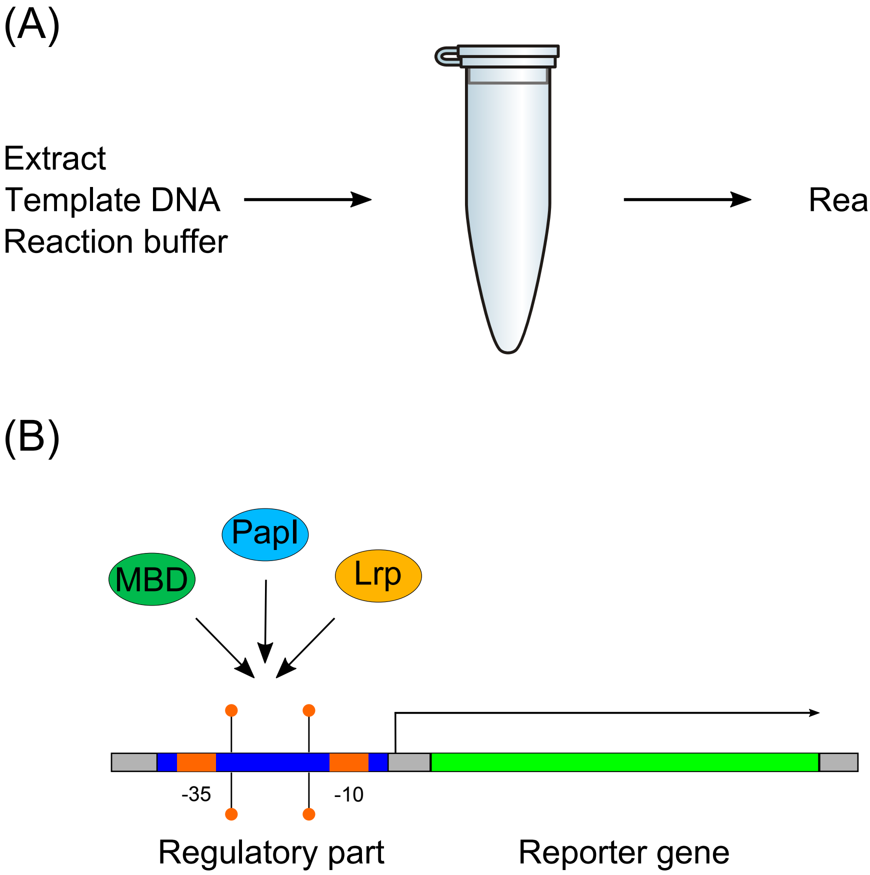


Figure 3.13.: Principle of a batch mode expression experiment. **(A)**, schematic illustration of the principle of a batch mode reaction. Extract and the reaction buffer are mixed together with the template DNA that holds the genetic information of all proteins which are to express. After incubation of the extract, the result of the experiment is read out by measuring the fluorescence intensity of the reporter protein. **(B)**, schematic illustration of the processes on the molecular level. The regulatory proteins (in this work PapI, Lrp, and the MBD of MeCP2) bind to specific recognition sites on the DNA strand, which overlap the -35 -10 consensus sequence of the promoter. This consequently represses the expression of the reporter gene which is observable by an decreased fluorescence intensity.

However, this energy transfer is not wanted during the simultaneous measurement of both fluorescence signals during competitive expression experiments. Thus, a possible modification of the data due to FRET artifacts needs to be taken into consideration for quantitative data analysis of competitive expression experiments. For the direct measurement of the expression kinetics, the plate reader is incubated at 29 °C. In this case, the reaction solution is immediately transferred into the multi well plate after the

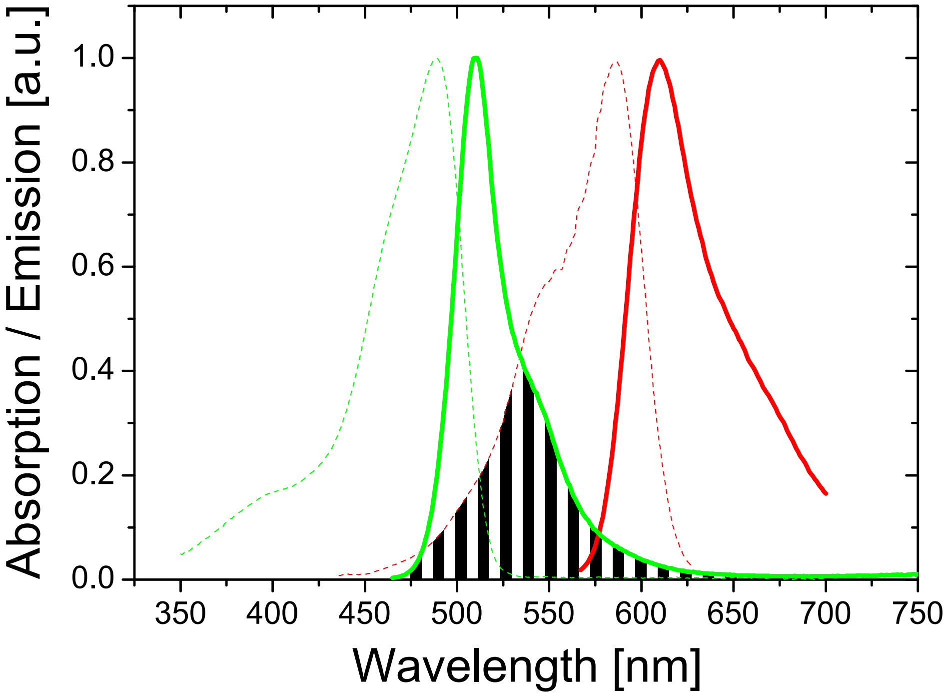


Figure 3.14.: Absorption (dashed line) and emission spectra (solid line) of the reporter proteins eGFP (green) and mCherry (red). Due to the overlap between the emission spectra of eGFP and the excitation spectra of mCherry (striped area), both dyes constitute a donor-acceptor pair for FRET.

preparation and the fluorescence signal of the recombined reporter protein is recorded as a function of time. The expression level of a batch mode reaction is reproducible from batch to batch within a 5% error bar [39].

4. Ultrahigh DNA hybridization specificity due to molecular competition

Competitive DNA hybridization deals with the molecular competition between two single-stranded DNA species A and B with slight sequence variations (targets) for binding to a complementary strand (probe) at a controlled temperature. In earlier measurements [30], a surprising effect was detected during such a molecular competition. In spite of both, comparable binding constants of species A and B and an excess concentration of B over A by several orders of magnitude, species A was able to bind exclusively to the probe molecules in thermodynamic equilibrium, at particular experimental conditions. This result was in contrast to a usually employed thermodynamic description of molecular competition wherein the equilibrium state of the system is fully described by the individual binding affinities of the competing species and their respective concentrations (compare Eq. (2.28)). The following sections deal with the definition (4.1), verification (4.2), characterization (4.3) and explanation (4.4) of the mentioned high specificity of competitive DNA hybridization. A conclusion of all experimental and numerical results is given in sections 4.4.1 and 4.5.

4.1. Occurrence and definition of the high specificity

4.1.1. Binding affinities

In order to describe the competitive DNA hybridization in thermodynamic equilibrium, it is necessary to determine the binding affinities K^A and K^B of the competing target molecules. Therefore, hybridization experiments without competition are performed for each individual target species. Following an increase of the target concentration $[T]_0$, the hybridization signal $D(t)$ is measured until it reaches its equilibrium value D_{eq} and remains constant in time. This procedure is repeated until a final concentration increase does no longer lead to an increased signal, what means that all probe molecules are occupied. Ultimately, the equilibrium values D_{eq} are plotted against the target concentration $[T]_0$ and the fit of the extended Langmuir isotherm

$$[D]_{eq} = \frac{1}{2}\left\{[T]_0 + [P]_0 + K^{-1} - \sqrt{([T]_0 + [P]_0 + K^{-1})^2 - 4[T]_0[P]_0}\right\} \qquad (4.1)$$

(compare Eq. (2.16)) to the data reveals the binding constant K of the investigated target molecule as well as the initial probe concentration $[P]_0$ that is the second fit parameter.

Fig. 4.1 shows the raw data and the extended Langmuir isotherms for the target molecules PM, MM1, MM2, MM3, MM4, PM* and MM* (see appendix Table A.1). The respective binding constants K are combined in Table 4.1. Please note that PM* and MM* hybridize to a different probe strand than PM, MM1, MM2, MM3 and MM4, since the effect of highly specific hybridization needs to be verified for a different sequence motif (see section 4.2.1). The results show that both, the position and the number of non-matching base pairs within short oligonucleotide strands, have significant influence on the binding constant. Here, perfect matches of course possess the highest affinity (PM), followed by mismatches towards the ends (MM2 and MM4) and in the middle of a strand (MM1 and MM3) [5]. The influence of the mismatch position on the occurrence of the high specificity is investigated in detail in section 4.3.2. The target species PM* and MM* posses a higher binding constant compared to the other ones, due to an increased GC content (compare for instance PM*: 53% and PM: 38%). Ultimately, the selection of target species presented in Table 4.1 enables investigation of the high specificity over a broad affinity spectrum.

4.1.2. Highly specific DNA hybridization

The effect of highly specific DNA hybridization appears upon competition between a “High Affinity Target” (HAT) and a “Low Affinity Target” (LAT) for binding to one single probe species [30]. The following experimental approach is used: In the first part of the experiment, only the HAT, concentrated at 5 nM, is introduced into the measurement chamber without any competing species (single hybridization). After that, the hybridization signal of the HAT is measured until it reaches its final value and remains constant in time. In the second part of the experiment, the HAT has to compete

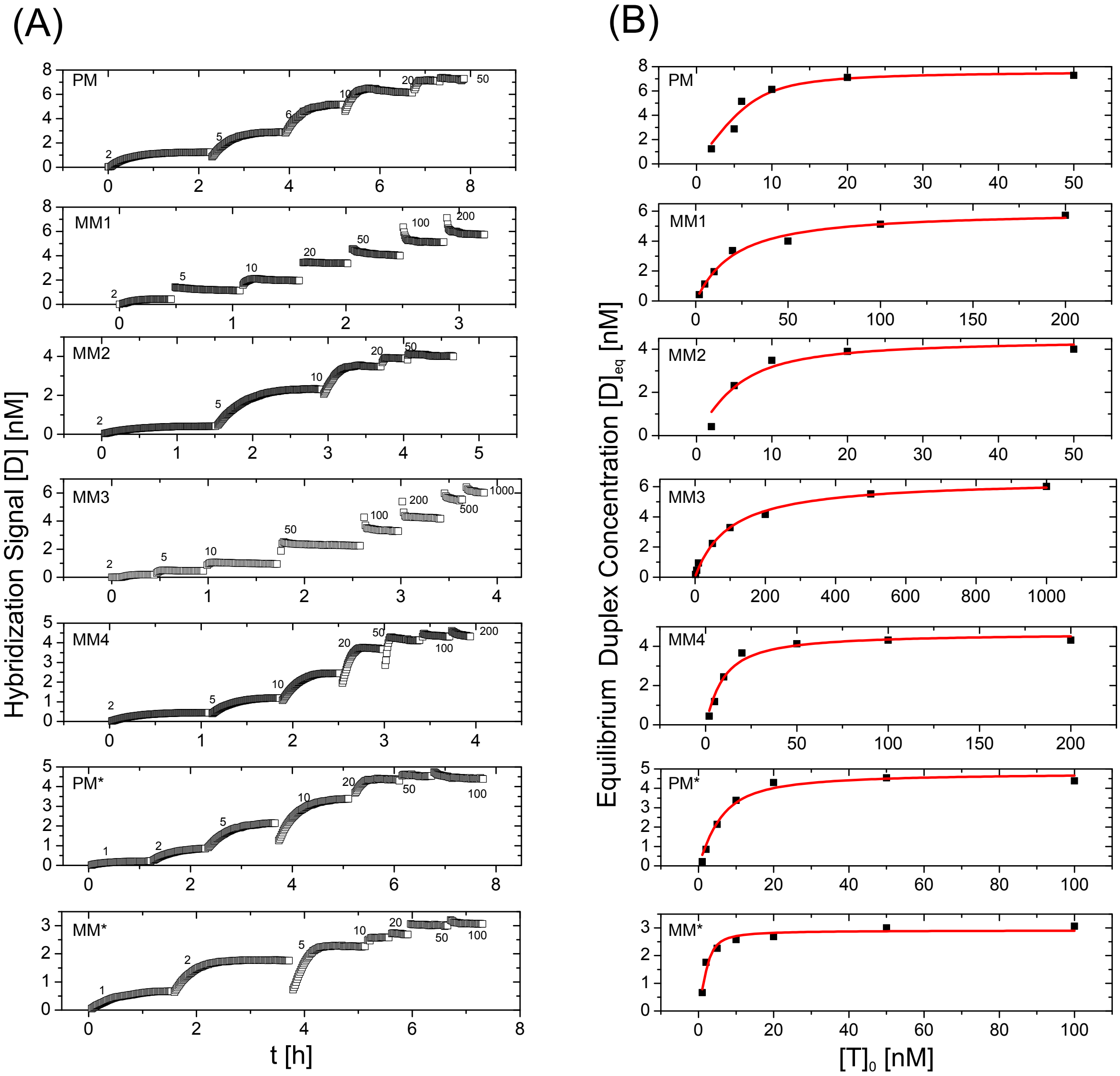


Figure 4.1.: Determination of binding affinities at 44 °C for target molecules PM, MM1, MM2, MM3, MM4 (already presented in [30]) and PM* and MM*. **(A)**, raw data. The graph shows the hybridization signal $[D]$ equivalent to the duplex concentration in the measurement chamber as a function of time. Following an increase of the target concentration $[T]_0$ (respective values in nM), $[D]$ is measured until it reaches its equilibrium value $[D]_{eq}$ and remains constant in time. Procedure is repeated until all probe molecules are occupied. **(B)**, extended Langmuir isotherms. The respective values for $[D]_{eq}$ from (A) are plotted against $[T]_0$. Fitting of Eq. (4.1) to this data (red line) reveals the binding constants K of the investigated target molecule in units 1/M. The values are listed in Table 4.1.

Table 4.1.: Sequences and binding affinities of the target molecules used in surface based hybridization experiments. Non-matching bases are highlighted in red and underlined. The values for K at 44 °C are determined with the extended Langmuir isotherm (compare Fig. 4.1). Note that target molecules PM* and MM* hybridize to a different probe molecule than the other ones (compare section 4.2.1).

Target	Sequence (5′ → 3′)	K [10^7 1/M]
PM	AAG-GAT-CAG-ATC-GTA-A	90 ± 50
MM1	AAG-GAT-CAC-ATC-GTA-A	6 ± 1
MM2	AAG-GAT-CAG-ATC-GCA-A	40 ± 20
MM3	AAG-GAT-CTC-ATC-GTA-A	1 ± 0.1
MM4	AAA-GAT-CAG-ATC-GAA-A	10 ± 3
PM*	GGG-CAG-CAA-TAG-TAC	200 ± 30
MM*	GGG-CAG-CTT-TAG-TAC	200 ± 80

with 1 µM of the LAT for binding to the probes and the hybridization signal of the HAT is measured as a function of time. With the target concentrations $[HAT]$ and $[LAT]$, the probe concentration $[P]$, and the respective binding affinities K^{HAT} and K^{LAT} determined in section 4.1.1, it is possible to compare the measured equilibrium values with the prediction of Eq. (2.21):

$$[D]^{HAT} = \frac{[HAT]_{eq} \cdot K^{HAT} \cdot [P]_0}{K^{HAT} \cdot [HAT]_{eq} + K^{LAT} \cdot [LAT]_{eq} + 1} \tag{4.2}$$

Fig. 4.2 shows the result of the competitive hybridization experiment between PM and MM2 as well as PM and MM1 at 44 °C. Here, the PM constitutes the HAT and the mismatching targets MM2 and MM1 constitute the LAT. In the case of competition between PM and MM2, a 200-fold excess concentration of MM2 targets over PM targets leads to a reduced PM hybridization level by a factor of 94%. This reduction can be described with Eq. (4.2), thereby using the concentrations $[PM]_0$ and $[MM2]_0$ of the competing molecules and the respective binding affinities K^{PM} and K^{MM2}. The system exhibits *standard specificity*. However, in the case of "PM vs. MM1", an equivalent excess of MM1 targets over PM targets does not lead to any measurable decrease of D^{PM}. Eq. (4.2) fails to predict the correct equilibrium value since the PM hybridization in competition is unexpectedly specific. The system shows *high specificity*.

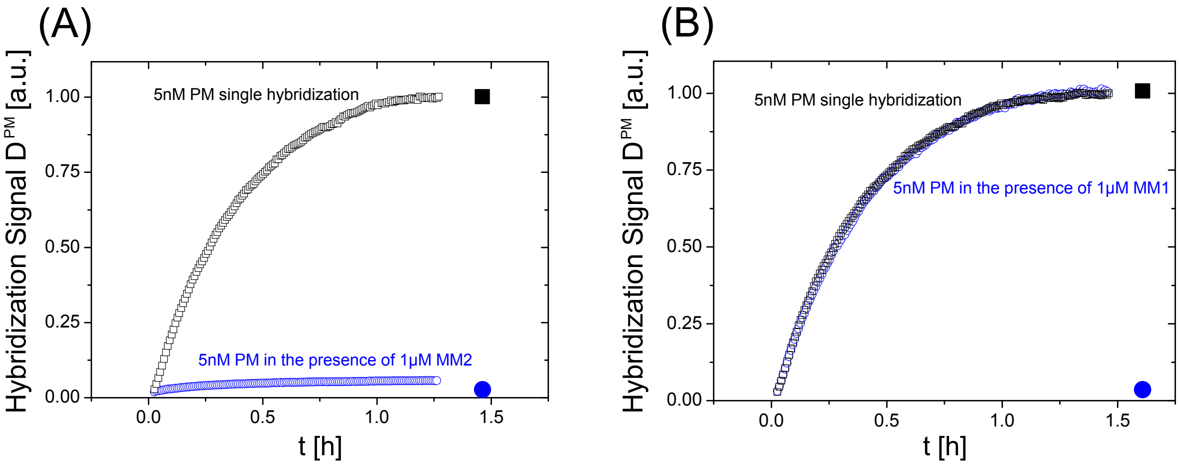


Figure 4.2.: Standard and high specificity during competitive DNA hybridization (also presented in [30]). **(A)**, competition between PM and MM2 at 44 °C. The graph shows the hybridization signal of the PM as a function of time in the case of a single hybridization (no competition, black) and in the presence of 1 µM of MM2. The target concentration of the PM is 5 nM in both measurements. D^{PM} is normalized to the final value of the single hybridization. The presence of MM2 targets diminishes the amount of hybridized PM by a factor of 94%. The filled symbols show the equilibrium duplex concentration, D_{eq}^{PM}, predicted by Eq. (4.2), thereby using the respective target concentrations and binding affinities (compare Table 4.1). The predictions are in good agreement with the experimental values (standard specificity). **(B)**, same experiment as shown in (A) but for competition between PM and MM1. The presence of 1 µM MM1 has no measurable influence on D^{PM} compared to the single hybridization. The predictions of Eq. (4.2) fail to describe the equilibrium value in the case of competition (high specificity).

4.1.3. Competitive systems of standard and high specificity

In the following experiments, the effect of the high specificity is analyzed over a broader affinity spectrum. Therefore, the measurements shown in section 4.1.2 are repeated for different competitive systems "HAT vs. LAT", involving all target molecules PM and MM1-MM4 listed in Table 4.1. Compared to the experiments presented in the previous section 4.1.2, the only difference is that the concentration of the LAT is increased in a stepwise manner from 1 nM to 1 µM. For six competitive systems "HAT vs. LAT", Fig. 4.3 shows the HAT hybridization kinetics during competition with the LAT. The corresponding HAT equilibrium values for a LAT concentration of $[LAT] = 0$ nM and 1 µM are combined in Fig. 4.4. The results reveal two characteristic scenarios:

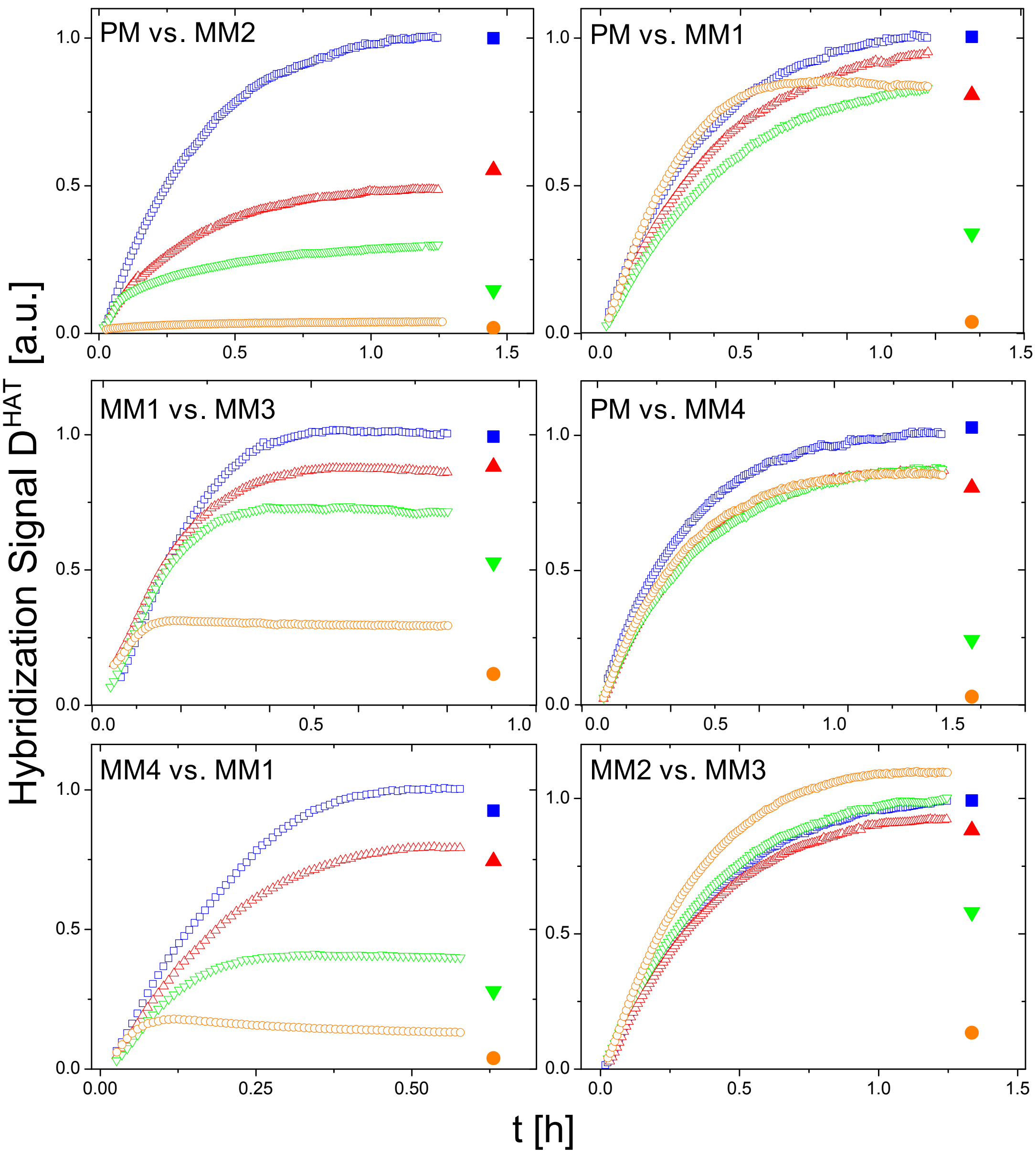


Figure 4.3.: Competitive hybridization experiments at 44 °C among the strands PM and MM1-MM4, listed in Table 4.1 (also presented in [30]). In each experiment, the HAT hybridization signal is measured as a function of time for increasing LAT concentrations ($[LAT]_0 = 0$ nM, blue; 10 nM, red; 100 nM, green and 1 μM, orange). The HAT concentrations are 10 nM (for “MM4 vs. MM1” and “MM1 vs. MM3”) and 5 nM (all others). D^{HAT} is normalized to the measured equilibrium value of the single hybridization ($[LAT]_0 = 0$ nM). The filled symbols in each graph are the predicted equilibrium values of Eq. (4.2) for D^{HAT}_{eq}, thereby using the binding affinities K^{HAT} and K^{LAT} (see Table 4.1).

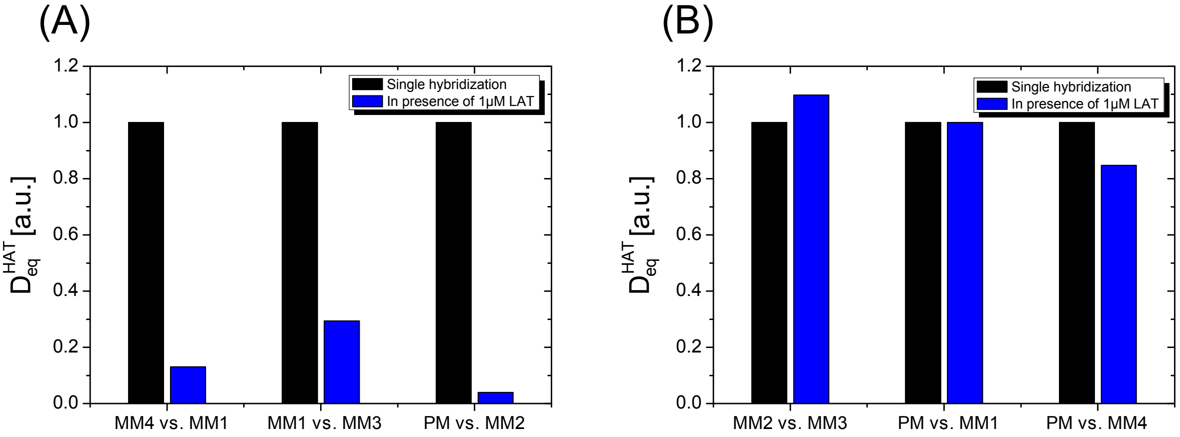


Figure 4.4.: Equilibrium values of all competitive systems "HAT vs. LAT" presented in Fig. 4.3. The graphs show the measured HAT equilibrium value D_{eq}^{HAT} for a LAT concentration of 0 nM (single hybridization, black) and 1 μM (blue). The respective competitive systems "HAT vs. LAT" are named on the x-axis. D_{eq}^{HAT} is normalized to the respective equilibrium value of the single hybridization (black bar). Two characteristic scenarios can be distinguished: **(A)**, standard specificity. D_{eq}^{HAT} decreases by increasing $[LAT]_0$ and Eq. (4.2) correctly predicts the equilibrium values from the individual binding affinities listed in Table 4.1. **(B)**, high specificity. The HAT dominates the reaction regardless of LAT concentration. The binding affinities do not reflect the measured values of D_{eq}^{HAT} and Eq. (4.2) fails.

1. Competitive systems of standard specificity:

   The presence of the LAT reduces the equilibrium duplex concentration D_{eq}^{HAT}. In this case, D_{eq}^{HAT} can be predicted from the target concentrations $[HAT]_0$ and $[LAT]_0$ as well as the respective binding affinities K^{HAT} and K^{LAT}, following Eq. (4.2).

2. Competitive systems of high specificity:

   The HAT hybridizes as if the LAT was not present at all in spite of a LAT excess concentration of at least two orders of magnitude. In this case, the binding affinities K^{HAT} and K^{LAT} do not reflect the measured equilibrium values of D_{eq}^{HAT} and Eq. (4.2) fails.

The results show that the effect of the high specificity occurs over a broad affinity spectrum and is not limited to the competitive system "PM vs. MM1" investigated in section 4.1.2. They further reveal that the competitive measurements divide into classes of high specificity and standard specificity.

4.2. Verification of the high specificity

4.2.1. Sequence motif invariance

The target molecules PM* and MM* presented in Table 4.1 are designed to investigate if the high specificity is also observable with a different sequence motif that is unrelated to the motif used in sections 4.1.2 and 4.1.3. The sequences of PM* and MM* closely follow the molecules "target of interest" and "competitor" presented in [8]. The experiments in Fig. 4.3 are repeated for the case of competition between PM* and MM*, while PM* and MM* constitute the HAT and LAT, respectively. The measured equilibrium values are compared with the predictions of Eq. (4.2), thereby using the binding affinities K^{PM^*} and K^{MM^*} also listed in Table 4.1.

Fig. 4.5 shows the competitive hybridization between PM* and MM* at 44 °C. Also in the case of this sequence motif, the effect of the high specificity is clearly observable. Eq. (4.2) fails to describe the equilibrium values from the binding affinities K^{PM^*} and K^{MM^*}. Thus, the effect of high specificity is not limited to the competitive systems shown in Fig. 4.4 (B).

4.2.2. Reversibility of the PM hybridization

The following experiment is performed to investigate if the hybridization of the PM is reversible. An irreversibility could cause the observed high specificity, since the annealing of the PM would block the probes for subsequent binding of a competing target species. In the beginning of the experiment, the PM, concentrated at 10 nM and labeled with Cy-5, is introduced into the measurement chamber and the hybridization signal D^{PM} is measured until the equilibrium value D^{PM}_{eq} is reached. After that, the target solution in the chamber is replaced by an equivalent solution of Cy-3 labeled PM molecules. The hybridization signal D^{PM} is measured again.

Fig. 4.6 shows that the hybridization of the PM indeed is reversible at 44 °C. The data verifies that it is possible to release the PM-Cy5 targets from the probe molecules by introducing an equivalent solution of PM-Cy3 targets into the measurement chamber. Thus, one can conclude that the high specificity is not caused by an irreversible binding of the HAT.

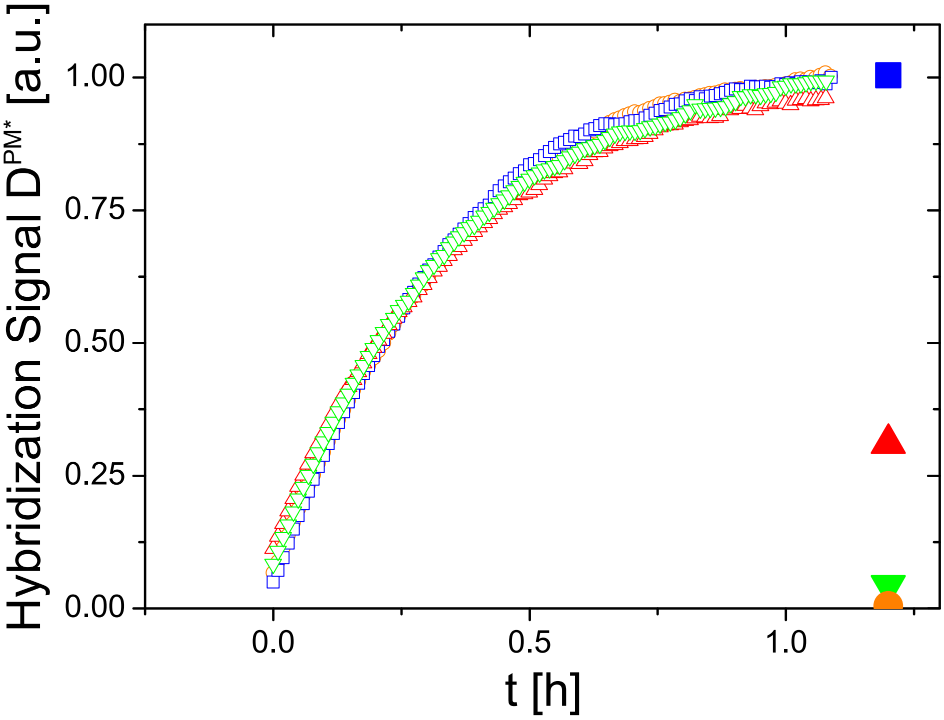


Figure 4.5.: Competitive hybridization between PM* and MM* at 44 °C. The graph shows the PM* hybridization signal as a function of time for increasing MM* concentrations ($[MM^*]_0$ = 0 nM, blue; 10 nM, red; 100 nM, green and 1 μM, orange; $[PM^*]_0 = 5$ nM in each measurement.). D^{PM^*} is normalized to the equilibrium value of the single hybridization (blue curve). The respective filled symbols are the predictions of Eq. (4.2) for the equilibrium values $D_{eq}^{PM^*}$, thereby using the binding affinities K^{PM^*} and K^{MM^*} (see Table 4.1). Also in the case of this sequence motif, the effect of the high specificity is clearly observable and Eq. (4.2) fails.

4.2.3. Influence of ion concentration

In the following, the effect of the high specificity is analyzed over a broader ion concentration range of the hybridization buffer. Since the hybridization of single-stranded DNA highly depends on the ion concentration of the solvent, it is necessary to rule out that the effect of the high specificity is caused by such artifacts. Thus, competitive hybridization experiments between PM and MM1 as well as PM and MM2 are performed with 0.25xSSC buffer (62.5 mM monovalent ions) and with 5xSSPE (0.95 M monovalent ions). In the latter case, the experiment is performed with a DNA-microarray setup (see [36, 40] for details).

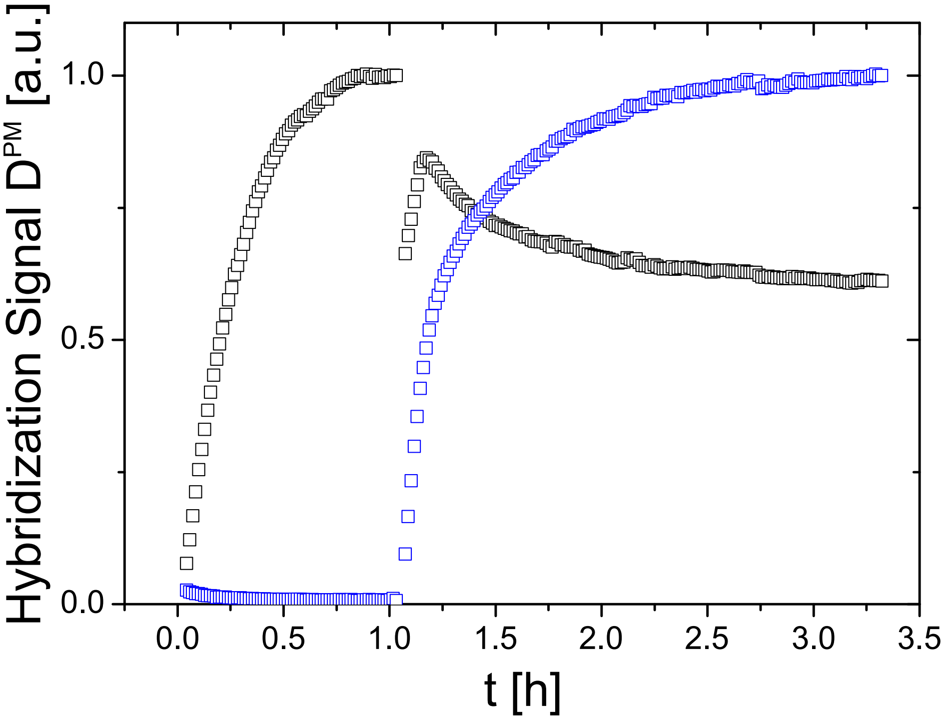


Figure 4.6.: Reversibility of the PM hybridization. The graph shows the hybridization signal D^{PM} of Cy-5 (black) and Cy-3 (blue) labeled PM molecules as a function of time at 44 °C. The Cy-5 and Cy-3 hybridization signal is independently normalized to the respective equilibrium value of D^{PM-Cy5} and D^{PM-Cy3}. Thus, the absolute values of the two channels cannot be compared with each other. After priming the probe molecules with PM-Cy5 targets ($[PM]_0 = 10$ nM) for about one hour, the target solution is replaced by an equivalent solution of PM-Cy3 targets. This experiment confirms the reversibility of the PM hybridization, since the introduction of Cy-3 labeled molecules leads to a decrease of the hybridization signal of the Cy-5 labeled targets. Note that the Cy5 signal does not drop to zero, since after the exchange of the target solution, some discriminated PM-Cy5 targets can still occupy free probe species.

Fig. 4.7 shows the results of the competitive hybridization experiments “PM vs. MM1”, measured with the TIRF method and 0.25xSSC at 22 °C, as well as “PM vs. MM2” and “PM vs. MM1”, measured with a DNA-microarray setup and 5xSSPE buffer at 44 °C [36, 40]. Also for these different ion concentrations, the system “PM vs. MM2” shows standard specificity while for “PM vs. MM1” the high specificity is clearly observable (compare Figs. 4.2 and 4.7). Even though the equilibrium duplex concentration of the PM is reduced in competition with 1 µM of MM1 targets (see Fig. 4.7 C for 0.25xSSC and 22 °C), one can conclude that the hybridization of the PM is still highly specific. For a quantitative conclusion it would be necessary to determine the binding affinities

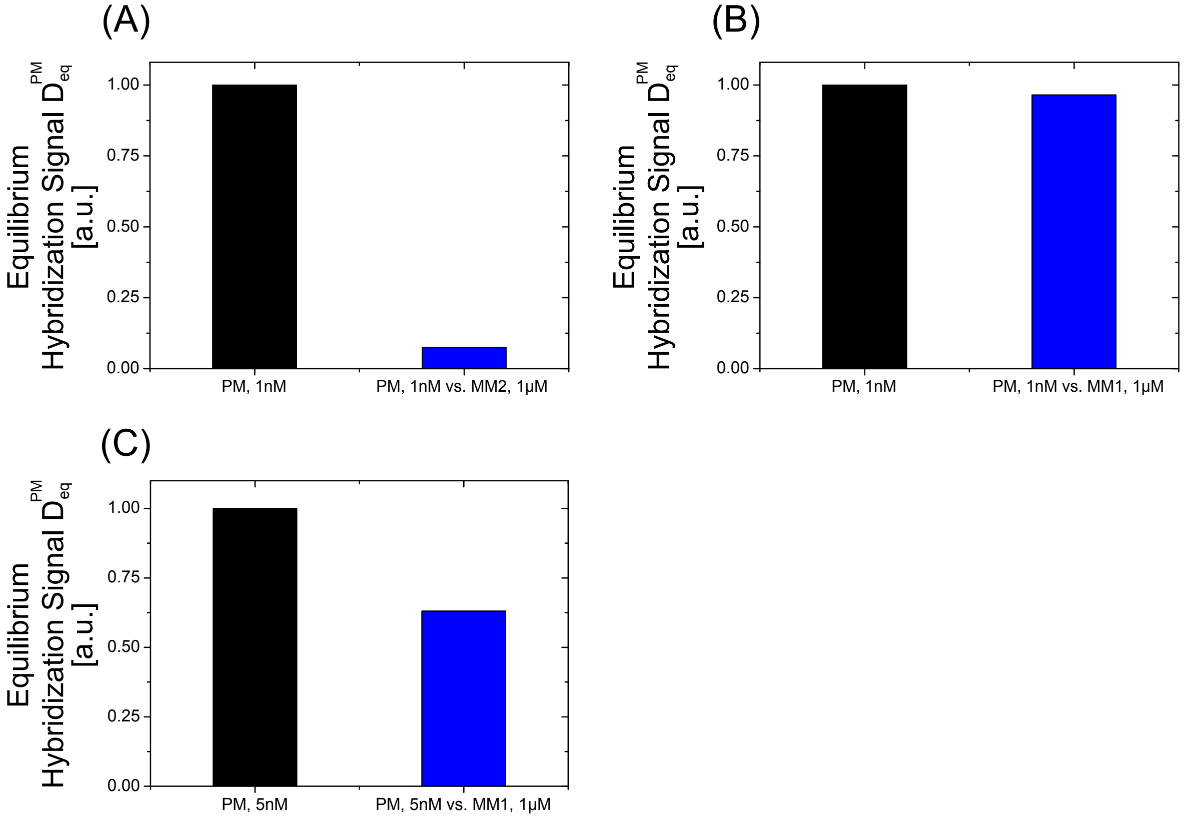


Figure 4.7.: Metal ion dependence of the high specificity. **(A)**, competitive system "PM vs. MM2". The graph shows the equilibrium values of the PM hybridization without competition (black bar) and in the presence of 1 µM of MM2 targets (blue bar). $[PM]_0 = 1$ nM and $T = 44$ °C in both cases. The experiment is performed by Christian Trapp on a DNA microarray setup using a 5xSSPE hybridization buffer (0.95 M monovalent ions) [36]. **(B)**, same as (A) but for the system "PM vs. MM1" [36]. **(C)**, same as (A) but for 0.25xSSC buffer (62.5 mM monovalent ions) and a temperature of 22 °C using the TIRF method. The results show that the existence of standard and highly specific systems does not depend on the ion concentration of the used hybridization buffer.

K^{PM} and K^{MM1} in experiments without competition for the conditions 0.25xSSC and 22 °C. This, however, is not done in the context of this work. Put together, one can conclude that the effect does not depend on the ion concentration of the used hybridization buffer. Additionally, the experiments shown in Fig. 4.7 A and B confirm the effect of the highly specific hybridization with a different experimental setup.

4.2.4. Measurements in bulk

In order to rule out that the high specificity is caused by surface artifacts, measurements in bulk are performed. In this case, the probe molecules are not immobilized to a surface and the FRET method is used to detect the hybridization between target and probe (see chapter 3.1.6 for details). The competitive systems “PM vs. MM1” (high specificity) and “PM vs. MM2” (standard specificity) are considered for this verification in bulk. First, the binding affinities are determined for the target molecules PM, MM1, and MM2 by measuring extended Langmuir isotherms in bulk, closely following the experimental approach presented in section 4.1.1. Since the hybridization signal D in bulk is not calibrated into units equivalent to the duplex concentration in the measurement chamber, a modified version of the extended Langmuir isotherm (Eq. (2.16)) is used to fit the data:

$$D_{eq} = \frac{A}{2}\left([T]_0 + [P]_0 + K^{-1}\right) - \frac{A}{2}\sqrt{\left([T]_0 + [P]_0 + K^{-1}\right)^2 - 4[T]_0[P]_0} \qquad (4.3)$$

where A is a calibration factor that is also a fit parameter. In order to analyze the competition in bulk between PM and MM1 as well as PM and MM2, two kinds of experiments are performed: i) The HAT and the probe constitute the FRET pair and the LAT remains unmodified. In this case the hybridization signal of the HAT is measured as a function of time for increasing concentrations of the LAT. ii) The LAT and the probe constitute the FRET pair and the HAT remains unmodified. In this case, the hybridization signal of the LAT is measured as a function of time for increasing concentrations of the HAT. The results in bulk are compared with the surface based experiments in order to investigate, if immobilization of the probe molecules has a significant influence on the measured binding affinities and on the occurrence of the high specificity itself.

Fig. 4.8 shows the extended Langmuir isotherms for the target molecules PM, MM1, and MM2. The extracted binding affinities are combined in Table 4.2. The results show that the immobilization of probe molecules to a surface does not lead to a significant modification of the binding affinities (compare Table 4.1). Fig. 4.9 shows the result of the competitive hybridization experiments in bulk. The measurements clearly confirm that the effect of highly specific DNA hybridization is not caused by surface artifacts. The findings of the surface based experiments shown in Fig. 4.2 can be reproduced in bulk. Thus, one can conclude that the effect of the high specificity is not only of technical importance for surface based hybridization on DNA bio-chips, but also has a general meaning for the molecular recognition of DNA.

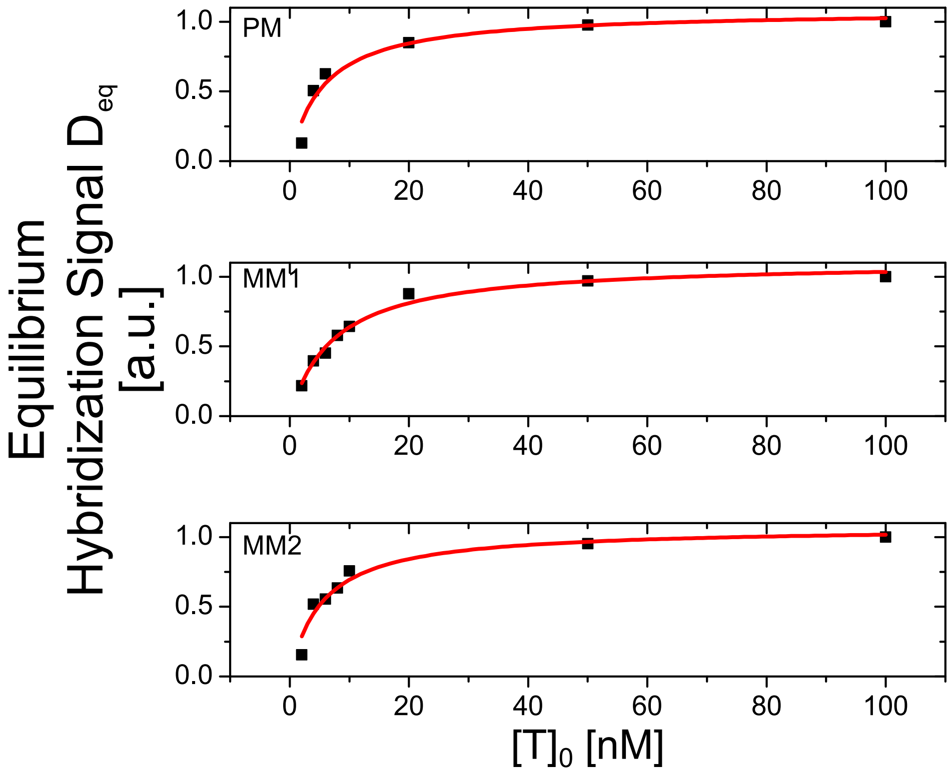


Figure 4.8.: Determination of binding affinities in bulk at 44 °C for the target molecules PM, MM1, and MM2 using the FRET method (see chapter 3.1.6 for details). The probe molecule and the respective target molecule constitute the FRET pair. The initial probe concentration $[P]_0$ is 10 nM in all measurements. The equilibrium hybridization signal D_{eq} is normalized to its maximum value reached for a target concentration of 100 nM. Fitting a modified version of the extended Langmuir isotherm (Eq. (4.3)) to this data reveals the binding affinity K of the investigated target molecule in bulk. The respective values of the binding constant are given in Table 4.2

Table 4.2.: Binding affinities in bulk for the target molecules PM, MM1, and MM2. The values for K at 44 °C are determined with the extended Langmuir isotherm (compare Fig. 4.8).

Target	**K [10^8 1/M]**
PM	2 ± 0.5
MM1	1.3 ± 0.5
MM2	2 ± 0.4

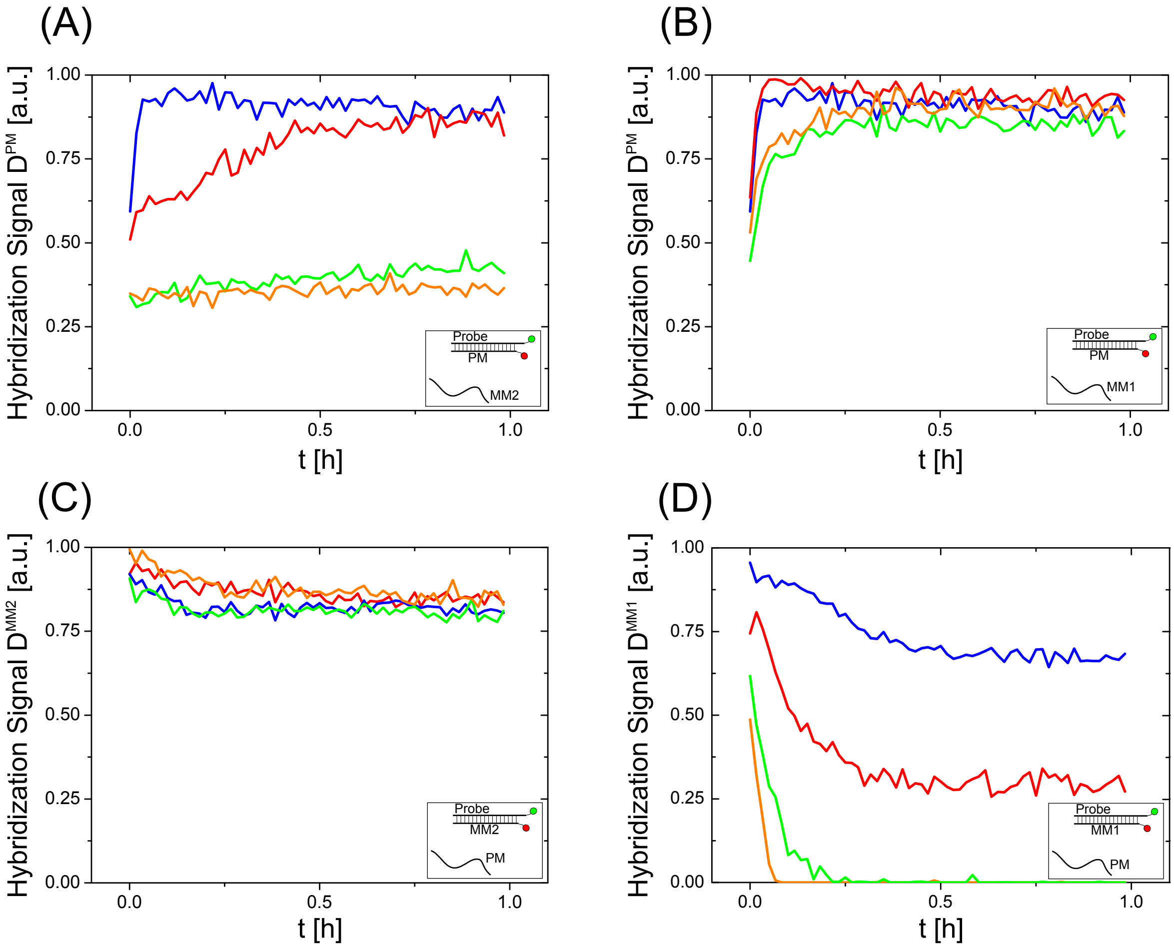


Figure 4.9.: Competitive DNA hybridization in bulk between PM and MM1 as well as PM and MM2 at 44 °C. The schemes in each graph show which DNA molecules are tagged with the fluorescent dyes and thus constitute the FRET pair. All data is normalized to the upper detection limit of the plate reader. The probe concentration is 10 nM in all measurements. **(A)**, competitive system "PM vs. MM2". PM and probe constitute the FRET pair while the MM2 is not labeled. Increasing the MM2 concentration ($[MM2]_0 = 0$ nM, blue; 10 nM, red; 100 nM, green and 1 µM, orange) results in a reduced PM hybridization signal D^{PM} ($[PM]_0 = 5$ nM in all cases). **(B)**, same as (A) but for the competitive system "PM vs. MM1". PM and probe constitute the FRET pair while the MM1 is unmodified. Even a MM1 concentration of 1 µM does not lead to a detectable decrease of D^{PM}. Thus, the measurements in bulk show the same behavior as the surface experiments (compare Fig. 4.3). **(C)**, competitive system "PM vs. MM2". MM2 and probe constitute the FRET pair, while the PM is unmodified. The PM concentration is slightly increased ($[PM]_0 = 0$ nM, blue; 5 nM, red; 10 nM, green and 20 nM, orange) while the MM2 concentration is kept constant at 100 nM. Independent of the PM target concentration, D^{MM2} always reaches the same value. **(D)**, same as (C) but for the competitive system "PM vs. MM1". MM1 and probe constitute

the FRET pair, while the PM is unmodified. Unlike the case presented in (C), a slight increase of the PM concentration results in a strong displacement of the MM1. For a concentration ratio $[MM1]_0 : [PM]_0 = 100 : 10$ (green) and $100 : 20$ (orange), D^{MM1} drops below the detection limit of the plate reader.

4.2.5. “Path independence” reveals thermodynamic equilibrium

In principle it is possible that the highly specific hybridization of the different high affinity targets (see Figs. 4.4 B and 4.5) is only a metastable state that relaxes on a long time scale. Thus, it is necessary to rule out such non-equilibrium effects and consequently verify thermodynamic equilibrium conditions. For the competitive system “PM. vs. MM1” as a representative of the highly specific systems, experiments with persistent variation of the hybridization temperature are performed. Here, a hysteresis of the measured equilibrium hybridization signals D_{eq}^{PM} and D_{eq}^{MM1} as a function of temperature points towards thermodynamic non-equilibrium. Furthermore, the order of target introduction into the measurement chamber must not have any influence on the signals after equilibration, and the signals have to be constant over a long time period. All these requirements for thermodynamic equilibrium are verified in the following experiments.

The results presented in Figs. 4.10 and 4.11 verify thermodynamic equilibrium conditions for the competitive hybridization between PM and MM1. The equilibrium signals D_{eq}^{PM} and D_{eq}^{MM1} show no hysteresis as a function of temperature. Priming the probe surface with MM1 targets does not affect the high specificity of the PM and the corresponding displacement of the MM1 after equilibration. Long term measurements in bulk using the FRET method (see chapter 3.1.6 for details) reveal the equilibrium signals D_{eq}^{PM} and D_{eq}^{MM1} to be constant over a time period of at least one week. Put together, one can conclude that the exclusive hybridization of the HAT indeed is the equilibrium state of any competitive system of high specificity, and that this effect is not caused by non-equilibrium artifacts.

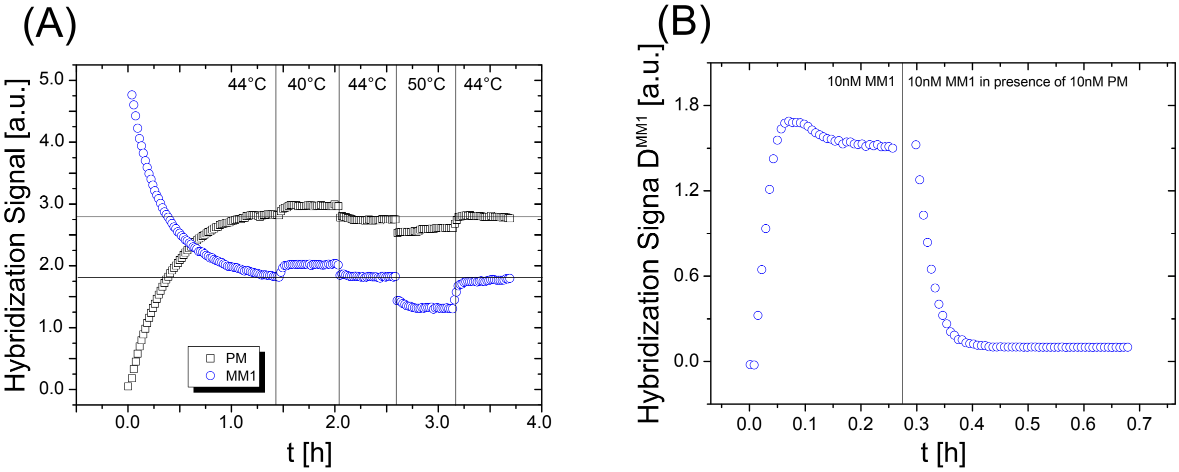


Figure 4.10.: Verification of thermodynamic equilibrium conditions. **(A)**, competitive hybridization between PM and MM1 at varying temperatures. The graph shows the hybridization signal of the PM (black) and MM1 (blue) as a function of time. At $t = 0$, the target solution is introduced into the measurement chamber. The target concentrations are $[PM]_0 = 5$ nM and $[MM1]_0 = 1$ µM, respectively. The hybridization signal of both targets at 44 °C are considered until they remain constant in time (horizontal lines). After that, the temperature is alternately decreased and increased (vertical lines, temperature values are given). The experiment reveals that the initially at 44 °C measured equilibrium values for PM and MM are reached again. Note that the varying hybridization signals at varying temperatures consist of two contributions, the amount of hybridized DNA and the emission characteristics of the dye. **(B)**, displacement experiment at 44 °C. After priming the probe molecule surface with MM1 targets ($[MM1]_0 = 10$ nM) the target solution is replaced by a two component mixture consisting of MM1 ($[MM1]_0 = 10$ nM) and PM ($[PM]_0 = 10$ nM). After that, the MM1 targets are releases as the PM is hybridizing with the probes. Thus, priming the surface does not affect the equilibrium values of the experiment.

4.3. Characterization of the high specificity

4.3.1. Concentration limit of the effect

The high specificity of the PM hybridization is indicated by the failure of Eq. (4.2) that predicts the complete displacement of the PM from the measured binding affinities and the concentration ratio (see Fig. 4.2 B). Thus, it is interesting to ask how much the MM1 concentration has to be increased in order to measure any influence on the equilibrium

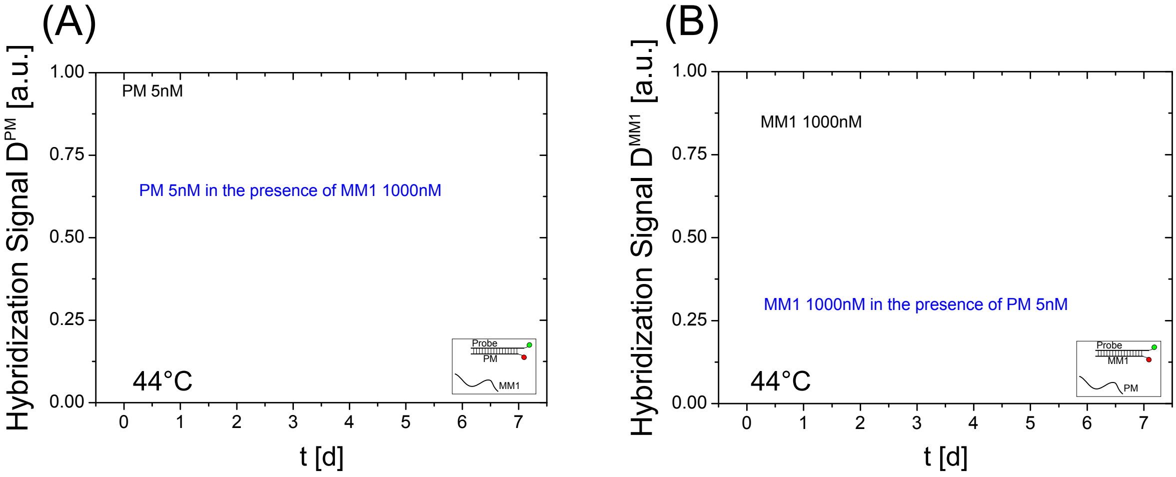


Figure 4.11.: Long term measurement of the competitive hybridization between PM and MM1 at 44 °C. Because surface grafted probes are released on such a long time scale the experiment is performed in bulk using the FRET method (see chapter 3.1.6 for details). Data is normalized to the upper detection limit of the plate reader. The probe concentration is 10 nM in all experiments. **(A)**, PM and probe constitute the FRET pair, while the MM1 is unmodified (see schematic in the box). The graph shows the hybridization signal of the PM as function of time. **(B)**, MM1 and probe constitute the FRET pair, while the PM is unmodified. The result shows that in both cases (A) and (B), the equilibrium values of the hybridization signals D_{eq}^{PM} and D_{eq}^{MM1} remain constant over a time period of one week.

duplex concentration of the PM. This is done in the following measurement. The experimental approach exactly follows the procedure shown in Fig. 4.4. The hybridization signal D^{PM} is measured for four different MM1 concentrations $[MM1]_0 = 0$ µM, 1 µM, 10 µM and 100 µM, while $[PM]_0$ is kept constant at 5 nM. A MM1 concentration of 100 µM defines a technical limit of the experiment. For higher concentrations, the target solution becomes too viscous in order to handle it without damaging the measurement chamber. Furthermore, the high concentration of the fluorescent dyes leads to fluorescence artifacts that invalidate the calibration of the hybridization signal.

The results are presented in Fig. 4.12. Even though the equilibrium duplex concentration of the PM can be reduced by about 40% compared to the case without competition, the effect is still predominant. Put together, even a concentration ratio of $[PM]_0 : [MM1]_0 = 1 : 20,000$ is not adequate to prevent the exclusive binding of the PM. This result underlines that the effect of the high specificity is seemingly dramatic and predominant over a broad concentration range.

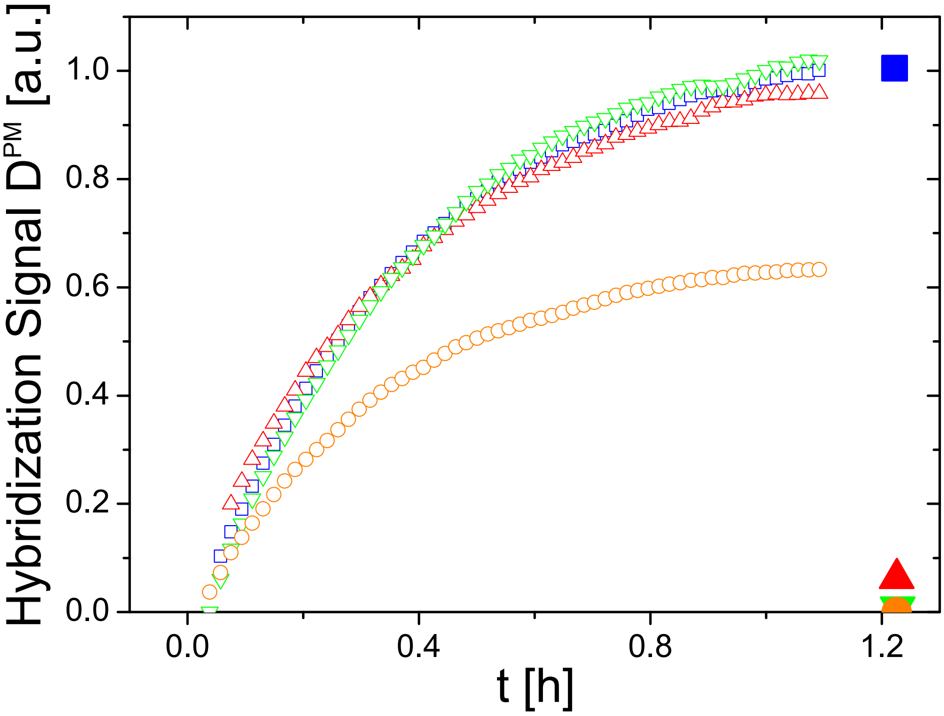


Figure 4.12.: Concentration limit of the high specificity using the example of the competitive system "PM vs. MM1" at 44 °C. The graph shows the hybridization signal of the PM as a function of time for four different concentrations of MM1 ($[MM1]_0 = 0$ μM, blue; 1 μM, red; 10 μM, green; and 100 μM, orange). $[PM]_0$ is kept constant at 5 nM for all measurements. D^{PM} is normalized to the equilibrium value of the single hybridization (blue curve). The respective filled symbols are the predictions of Eq. (4.2) for the equilibrium values, thereby using the binding affinities K^{PM} and K^{MM1} (see Table 4.1). Even though the high specificity is reduced for a concentration ratio $[PM]_0 : [MM1]_0 = 1 : 20,000$ (orange graph), the effect is still predominant.

4.3.2. Correlation between high specificity and LAT mismatch position

As shown in section 4.1.3, the existence of competitive systems of standard and high specificity is a result of a relatively weak perturbation, namely the slight shift in position of the erroneous base in the sequence of HAT or LAT. For instance, the PM completely looses its high specificity, observed for competition with a LAT with a mismatching base in the middle of the strand (MM1), when moving this mismatch position close to the end of the strand (MM2, compare Table 4.1). This raises the question, if there is a general dependence of the observed specificity of the HAT on the mismatch position

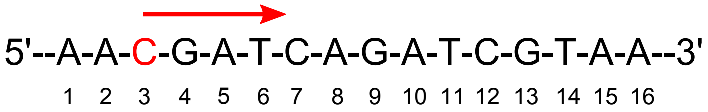


Figure 4.13.: Variation of the mismatch position within the LAT sequence. In order to analyze the correlation between the high specificity of the PM and the position of the non-matching base of the competitor LAT, the mismatch position of the LAT is varied. Starting at the third base, the position of the erroneous nucleotide (red) is shifted base by base to position fourteen. The mismatching base included in the target is always the same as the opposed base in the probe molecule (for instance, the mismatching base is C, if the opposed base in the probe is C). For all these twelve target molecules, the competition between PM and the respective LAT is investigated (Fig. 4.14).

within the competing LAT. This is analyzed in the following experiments. The PM is considered as the HAT that has to compete with a LAT, whose position of the erroneous base is varied systematically. Therefore, the target molecules MM1-MM14 are used (compare appendix Table A.1). Fig. 4.13 shows a schematic of the variation of the mismatch position. Starting at the third base (reading direction 5′ → 3′), the position of the mismatch is shifted base by base up to the fourteenth position. The mismatching base included in the LAT is always the same as the respective opposed base in the probe molecule. For all these twelve low affinity targets, the specificity of the PM hybridization is investigated in competitive measurements. The degree of specificity is defined as the ratio of the equilibrium hybridization signal of the PM with and without competition:

$$\frac{D_{eq}^{PM}\left([MM1]_0 = 1\ \mu\mathrm{M}\right)}{D_{eq}^{PM}\left([MM1]_0 = 0\ \mu\mathrm{M}\right)} \tag{4.4}$$

Fig. 4.14 shows the result of the competitive hybridization experiments between the PM and twelve versions of the LAT with varying positions of the erroneous base, following the scheme in Fig. 4.13. The experiments show that there is no systematic dependence of the specificity of the PM on the mismatch position of the respective competing LAT. The specificity varies from high specificity (observed for LATs with mismatch position 3, 6 and 9) and standard specificity (observed for mismatch position 14).

One has to mention that in general, quantitative conclusions about the specificity of a competitive system are only possible, if the binding affinities of both competitors K^{HAT} and K^{LAT} are known. In the present experiments this is not the case, because not all binding affinities of the twelve LAT versions are determined with extended Langmuir isotherms. However, the qualitative comparison of the measured specificity of all the

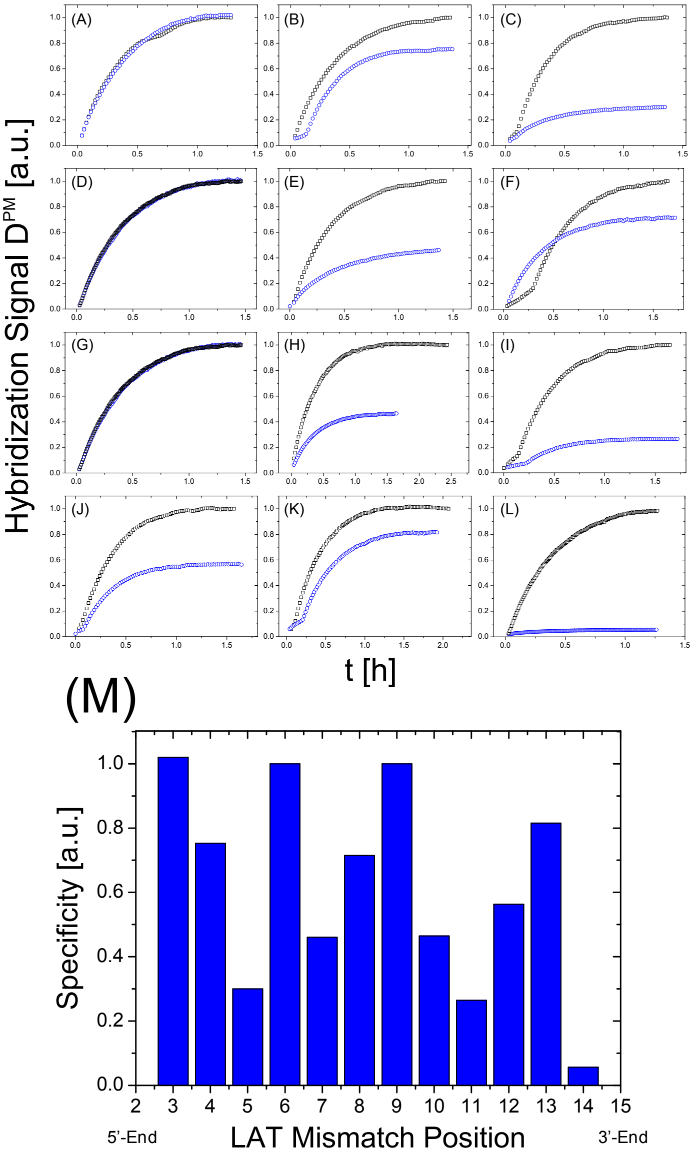


Figure 4.14.: Dependence of the high specificity of the PM on the position of the non-matching base within the sequence of the LAT at 44 °C. **(A)-(L)**, raw data of the experiment. All graphs show the hybridization signal of the PM as a function of time in the case of a single hybridization (no competition, black) and in the presence of 1 µM of the LAT (blue). The non-matching base is varied from position three (A) to position fourteen (L), following the scheme in Fig. 4.13. **(M)**, specificity of the PM hybridization as a function of the mismatch position of the respective LAT, using the data from (A)-(L). The specificity is defined in Eq. (4.4). The data shows no systematic correlation between the specificity of the PM and the mismatch position of the LAT.

twelve competitive systems is sufficient to exclude a systematic dependence of the high specificity on the mismatch position of the LAT.

4.3.3. Free dissociation vs. sequential displacement

The replacement of a particular target molecule by a competitive binder can follow two different pathways, the free dissociation or sequential displacement [41, 42]. After the hybridization of a particular target molecule B, a competing molecule A has two possibilities to conquer the probe (see Fig. 4.15). The first one is the spontaneous dissociation of target B followed by the association of target A (Fig. 4.15 A). The second option solely requires the existence of a pre-melted region within the duplex consisting of probe and target B (Fig. 4.15 B). This allows the simultaneous association of target A resulting in the temporary formation of a triplex structure wherein the DNA strands can interact. This interaction leads to the complete displacement of target B by migration of the branch point. In the following the existence of dissociation pathway and sequential displacement pathway is analyzed for the competitive system "PM vs. MM1" as a representative of the highly specific systems. Following [41], the observed dissociation rate k_{obs}^{MM1} of the MM1 in presence of the PM is given by

$$k_{obs}^{MM1}([PM]) = k_1 + k_2 \cdot [PM] \tag{4.5}$$

k_1 is the spontaneous dissociation rate, k_2 is the sequential displacement rate and $[PM]$ the target concentration of the PM. In the case of a MM1 single hybridization (no competition, $[PM] = 0$), the sequential displacement pathway cannot occur and Eq. (4.5) simplifies to

$$k_{obs}^{MM1} = k_1 \tag{4.6}$$

In other words, k_1 is the upper bound for the spontaneous dissociation without any competitor interactions. Consequently, in a competitive situation between PM and MM1, any observed MM1 dissociation rate $k_{obs}^{MM1} > k_1$ verifies that the PM is able to displace its competitor. The spontaneous dissociation rate k_1 of the MM1 is determined by measuring the hybridization kinetics of the MM1 in a single hybridization experiment without competition and fitting of Eq. (2.14) to this data. In order to determine k_{obs}^{MM1} in the presence of the PM, competitive experiments are performed. After priming the probe molecule surface with MM1 targets, the MM1 hybridization signal D^{MM1} is measured as a function of time for increasing concentrations of PM targets. Fitting these dissociation

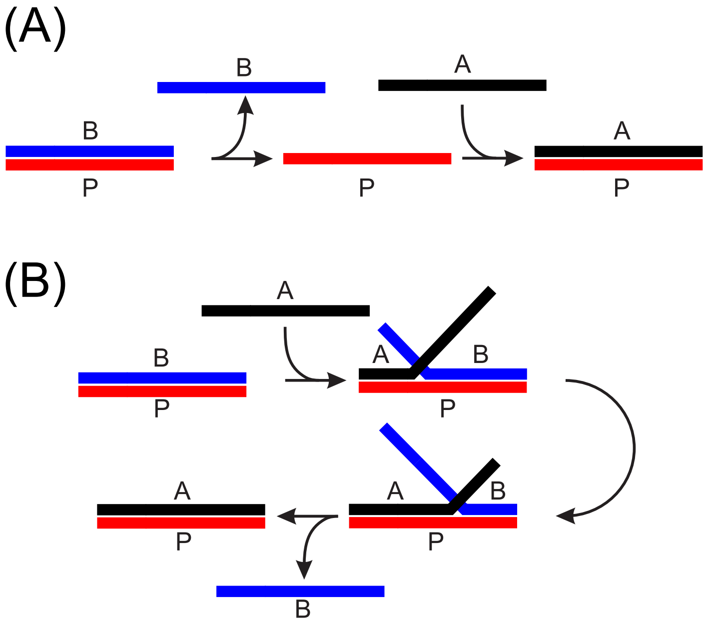


Figure 4.15.: Displacement mechanisms during competitive DNA hybridization [41, 42]. **(A)**, dissociation pathway. The spontaneous dissociation of target B (blue) from the probe P (red) is followed by the association of target A (black). **(B)**, sequential displacement pathway. A pre-melted region in the duplex consisting of probe and target B allows simultaneous binding of target A resulting in the temporary formation of a triplex structure. Migration of the branch point causes the complete replacement of target B by target A.

curves with the expression

$$D^{MM1}(t) = D_0 \cdot e^{-k_{obs}^{MM1} \cdot t} \tag{4.7}$$

reveals the observed MM1 dissociation rate k_{obs}^{MM1} in the presence of the PM. D_0 is the initial hybridization signal after priming the probe molecules. As mentioned above, for $k_{obs}^{MM1} > k_1$ the sequential displacement pathway is verified.

Fig. 4.16 shows the result of the experiments described above. The hybridization kinetics of the MM1 reveals a spontaneous dissociation rate of $k_1 = 3.1 \cdot 10^{-3}$ 1/s (Fig. 4.16 A). In competition, for increasing values of $[PM]$, the observed MM1 dissociation rate clearly exceeds the value of k_1: $k_{obs}^{MM1}(10\text{ nM}) = (10 \pm 0.3) \cdot 10^{-3}$ 1/s, $k_{obs}^{MM1}(25\text{ nM}) = (16 \pm 1) \cdot 10^{-3}$ 1/s, $k_{obs}^{MM1}(50\text{ nM}) = (21 \pm 2) \cdot 10^{-3}$ 1/s. This verifies the sequential displacement pathway for the competitive system "PM vs. MM1" and consequently the existence of an interaction between the two competitors in terms of the branch migration (Fig. 4.15 B).

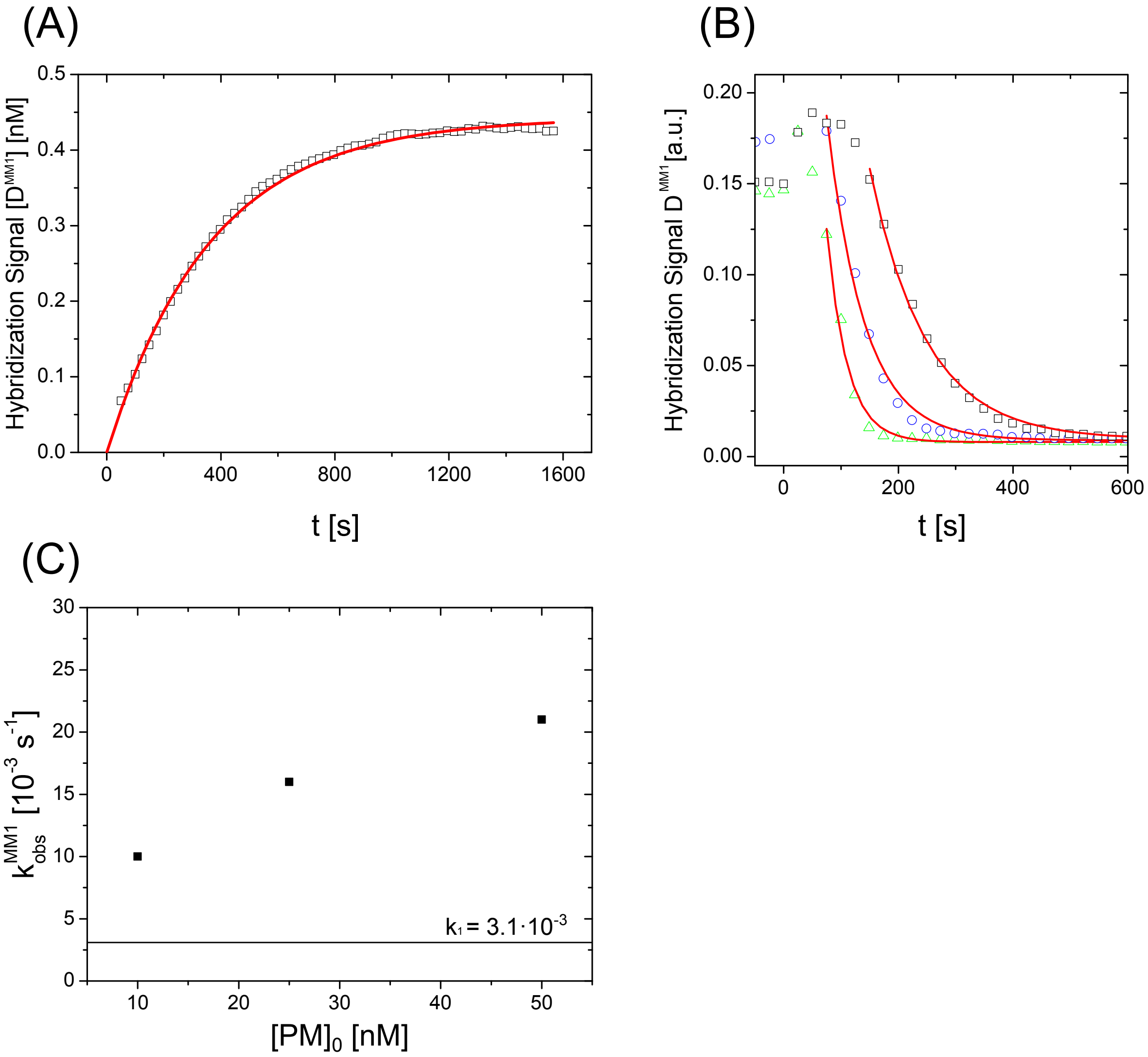


Figure 4.16.: The sequential displacement pathway applies to the competitive system "PM vs. MM1" (also presented in [30]). **(A)**, hybridization kinetics of the MM1 without competition at 44 °C. The graph shows the hybridization signal of the MM1, $[D^{MM1}]$ (equivalent to the duplex concentration in the measurement chamber) as a function of time ($[MM1]_0 = 2$ nM). Fitting the curve with Eq. (2.14) (red line) reveals the spontaneous dissociation rate k_1 to be $3.1 \cdot 10^{-3}$ 1/s. **(B)**, displacement experiment (performed by Timo Mai in earlier measurements). After priming the probe molecules with MM1 targets ($[MM1]_0 = 10$ nM) the target solution is replaced by a two component mixture consisting of MM1 (still concentrated at 10 nM) and PM (10 nM, □; 25 nM, ◯; 50 nM, △). In the following the dissociation of the MM1 is measured as a function of time. Fitting these curves with Eq. (4.7) (red lines) then reveals the observed MM1 dissociation rate, k_{obs}^{MM1}, for the three different PM concentrations: $k_{obs}^{MM1}(10\text{ nM}) = (10 \pm 0.3) \cdot 10^{-3}$ 1/s, $k_{obs}^{MM1}(25\text{ nM}) = (16 \pm 1) \cdot 10^{-3}$ 1/s, and $k_{obs}^{MM1}(50\text{ nM}) = (21 \pm 2)\cdot$

10^{-3} 1/s. **(C)**, verification of the sequential displacement pathway. The values for k_{obs}^{MM1} extracted in (B) are plotted against $[PM]$. The horizontal line indicates the spontaneous dissociation rate k_1 extracted in (A). For all PM concentrations, k_{obs}^{MM1} exceeds k_1 and thus the sequential displacement pathway applies to the competitive system "PM vs. MM1".

It is important to mention that just the existence of the sequential displacement pathway does not explain the occurrence of highly specific competitive systems. The underlying molecular process of the sequential displacement pathway is the migration of the duplex branch point (compare Fig. 4.15 B). However, this process is symmetric. This means that the branch point can migrate in two directions so that either target B is replaced by target A or vice versa. In this sense, the branch migration does not change the equilibrium state of the system, since the binding free energy of the probe symmetrically distributes between target A and B depending on how many bases of each strand are bound to the probe. Thus, Eq. (4.2) already accounts for the existence of such triplex structures as long as they appear symmetrically. In this sense, one can conclude from the experiments above that for the highly specific competitive systems, the binding free energy landscape in competition has to change asymmetrically in favor of one particular target molecule compared to the case without competition.

4.3.4. Melting curves of PM and MM1

In order to gain further insight into the cause of the high specificity, it is necessary to determine the thermodynamic parameters of entropy change ΔS, enthalpy change ΔH, and melting temperature T_m (see chapter 2.2.5) for the target molecules PM and MM1. This is done by measuring the melting curves of the mentioned DNA molecules independently without competition. A slow temperature increase leads to denaturation of double-stranded DNA in thermodynamic quasi-equilibrium. After priming the whole probe molecule surface with target molecules (excess concentration of 1 μM) the fraction of occupied probes Θ is measured as a function of temperature. The contribution of the temperature dependent efficiency of the fluorescent dye to the hybridization signal is removed as explained in chapter 3.1.5 and appendix section A.7. The surface occupancy Θ can be linked to the binding constant K and the values of entropy (ΔS) and enthalpy change (ΔH) by combining the Langmuir isotherm

$$\Theta = \frac{K \cdot [T]_0}{K \cdot [T]_0 + 1} \tag{4.8}$$

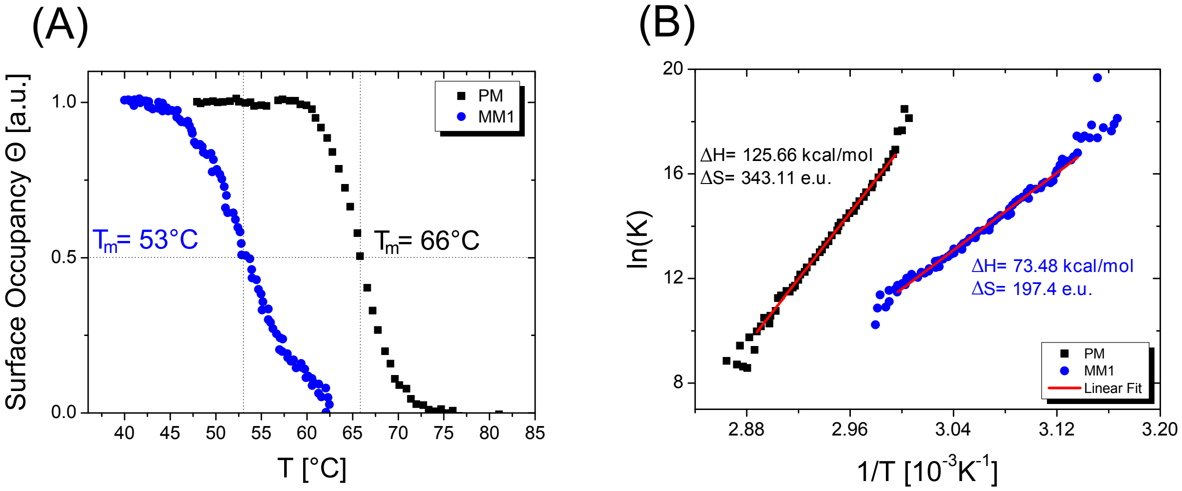


Figure 4.17.: Denaturation experiment for PM and MM1. **(A)**, melting curves. The graph shows the fraction of occupied probes Θ as a function of temperature T for PM (black) and MM1 (blue) in individual experiments without competition. The target concentration is 1 µM in both cases. The melting temperature T_m is defined as the temperature where the surface occupancy equals 50% (intersection of the dotted lines). **(B)**, Arrhenius plot of the melting transition of PM (black) and MM1 (blue) after converting the surface occupancy Θ (see A) into the binding constant K (Eq. (4.9)). The graph shows the logarithm of the binding constant K as a function of the inverse temperature $1/T$. The red lines are linear fits to the experimental data, thus revealing the values of enthalpy ΔH and entropy change ΔS (e.u. = "entropic units" = kcal/(mol · K)) of PM and MM1, respectively. The result presents a less important change in ΔS and ΔH for MM1 compared to the PM. A hybridized MM1 possesses an increased number of entropic degrees of freedom compared to PM.

and

$$K(T) \propto e^{\frac{\Delta G}{RT}} = \exp\left(\frac{\Delta H}{R} \cdot \frac{1}{T} - \frac{\Delta S}{R}\right) \tag{4.9}$$

After measuring the melting curve, the surface occupancy Θ is converted into the binding affinity K following Eq. (4.8). The plot of the logarithm of K over the inverse temperature $1/T$ (Arrhenius plot, see Eq. (4.9)) and the linear fit of these data reveal the thermodynamic parameters ΔS and ΔH.

Fig. 4.17 shows the melting curves of the PM and MM1. The extracted thermodynamic parameters of melting temperature T_m, entropy change ΔS, and enthalpy change ΔH for the two molecular species are combined in Table 4.3. The experiment reveals T_m to be 53 °C for the MM1 and 66 °C for the PM. The non-matching base of the MM1

Table 4.3.: Values of enthalpy change ΔH, entropy change ΔS and melting temperature T_m for target molecules PM and MM1. Data is obtained from the denaturation curves (Fig. 4.17).

Target	ΔH $[\text{kcal} \cdot \text{mol}^{-1}]$	ΔS $[\text{kcal} \cdot \text{mol}^{-1}] \cdot \text{K}^{-1}$	T_m [°C]
PM	125.66 ± 2	343.11 ± 3	66 ± 1
MM1	73.48 ± 1	197.4	53 ± 1

destabilizes the duplex and shifts the melting transition to smaller temperature values compared to the PM. Comparing the values of ΔH and ΔS for the two molecular species reveals the entropy loss due to duplex formation of PM and probe to be about 42% higher than the entropy loss due to duplex formation of MM1 and probe. Put together, the hybridized MM1 possesses an increased number of entropic degrees of freedom compared to the PM that binds to the probe molecule more stiffly.

4.3.5. Correlation between high specificity and melting temperature

The results of the denaturation experiments for PM and MM1 (section 4.3.4) raise the question, if there is a correlation between the occurrence of the high specificity and the melting temperatures of the competing target molecules. In the following, this is analyzed for all standard and highly specific systems presented in Figs. 4.4 and 4.5. T_m is measured experimentally for the target molecules PM and MM1 (Fig. 4.17). For all other molecules MM2, MM3, MM4, PM*, and MM* the values are determined by using the dinamelt web server for nucleic acid melting predictions [29]. Comparing the measured and predicted values of T_m for the molecules PM and MM1 reveals an offset of +8 °C between theory and experiment. This is due to the modification of the target strands with fluorescent dyes and the immobilization of the probe molecules to the surface, which is not considered in [29]. Therefore, 8 °C is added to the predicted values of T_m for the sequences MM2, MM3, MM4, PM* and MM*. The results are presented in Table 4.4. Plotting the melting temperature T_m^{HAT} of the HAT against the melting temperature T_m^{LAT} of the LAT for all competitive systems presented in Figs. 4.4 and 4.5 reveals a possible correlation between the occurrence of the high specificity and the value of T_m.

Table 4.4.: Melting temperatures of all target molecules hybridizing to the probe at a target concentration of 1 μM. The melting temperatures of PM and MM1 are measured values obtained from the respective denaturation curve (Fig. 4.17). The values for MM2, MM3, MM4, PM* and MM* are theoretically predicted with the dinamelt web server for nucleic acid melting predictions [29] for a concentration of 0.57 M monovalent ions (3xSSC buffer). Note that the target PM* and MM* hybridize to a different probe molecule than the other targets.

Target	T_m [°C]
PM	66
MM1	53
MM2	58
MM3	48
MM4	44
PM*	67
MM*	52

Fig. 4.18 shows that the melting temperatures T_m^{HAT} and T_m^{LAT} are key parameters that give information about if a competitive systems shows either standard specificity or high specificity. Standard specificity is observable if the melting temperatures of HAT and LAT are comparable ($T_m^{HAT} \approx T_m^{LAT}$). High specificity, on the other hand, occurs if the melting temperature of the HAT exceeds the melting temperature of the LAT by at least 10 °C, while the melting temperature of the LAT is comparable to the measurement temperature itself ($T_m^{HAT} > T_m^{LAT} \approx T$). Together with the results of the melting curves for PM and MM1 (section 4.3.4), one can conclude from the findings in Fig. 4.18 that high specificity occurs if the LAT hybridizes with a large number of entropic degrees of freedom, particularly pronounced at its melting temperature. The HAT, on the other hand, binds in a relatively stiff conformation for temperatures below its melting point.

4.3.6. Correlation between high specificity and temperature

The correlation between the melting temperatures of the competing target molecules T_m^{HAT} and T_m^{LAT} and the existence of standard and highly specific competitive systems (section 4.3.5) point towards a general temperature dependence of the effect. Because high specificity is observed for $T_m^{HAT} > T_m^{LAT} \approx T$ and standard specificity is observed for $T_m^{HAT} \approx T_m^{LAT}$ (see Fig. 4.18), it should be possible to convert any highly specific system "HAT vs. LAT" into standard specificity by varying the temperature T of the

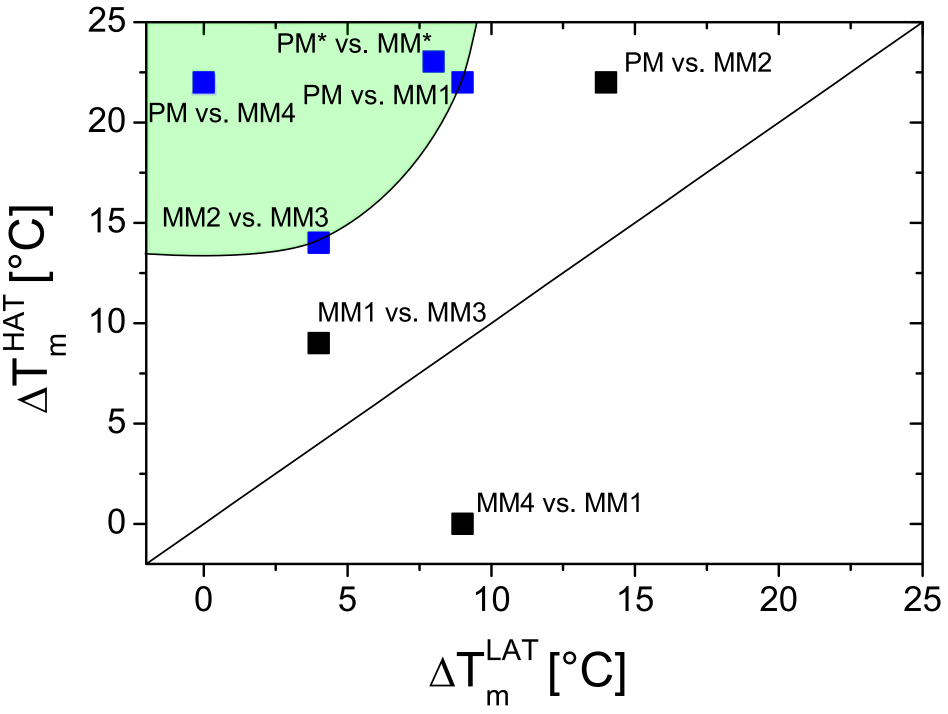


Figure 4.18.: Predicted melting temperatures of "HAT vs. LAT" for all competitive cases in Fig. 4.4 (also presented in [30]) and "PM* vs. MM*" (Fig. 4.5). Blue squares: highly specific systems, black squares: standard specific systems. For ease of viewing the hybridization temperature of 44 °C is subtracted from the melting temperature ($\Delta T_m = T_m - 44$ °C). The solid line indicates $\Delta T_m^{HAT} = \Delta T_m^{LAT}$.

experiment. This is analyzed in the following section. Considering the competitive hybridization between PM and MM1, the degree of specificity, as defined in Eq. (4.4), is measured as a function of temperature.

Fig. 4.19 shows the competitive hybridization between PM and MM1 at varying temperatures T. The experiment reveals that the high specificity of the PM diminishes with decreasing temperature. Thus, one can gradually change any competitive system from standard to high specificity by varying the temperature of the experiment. The results verify the findings of the previous section 4.3.5.

4.3.7. PM affinity in competition

The experimental results show that the high specificity is based on a temperature dependent (section 4.3.6) interaction (section 4.3.3) involving different entropic degrees

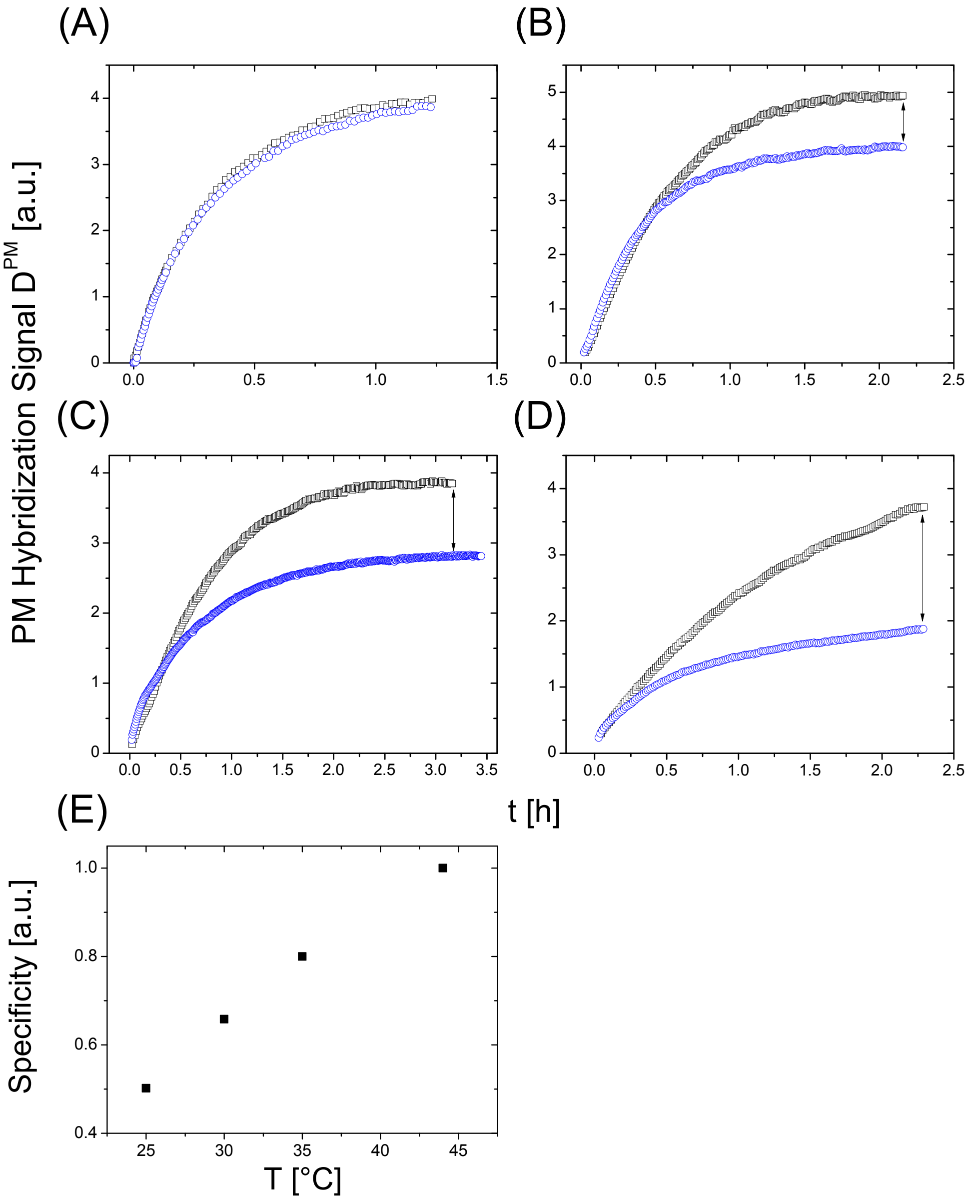


Figure 4.19.: Correlation between high specificity of the PM and temperature T (also presented in [30]). **(A)-(D)**, raw data of the experiment. The graphs show the hybridization signal of the PM as a function of time in the case of a single hybridization (black) and in the presence of 1 µM of MM1 targets (blue) for the temperatures 44 °C (A), 35 °C (B), 30 °C (C) and 25 °C (D). The PM concentration is 5 nM in all measurements. The arrows display the difference between D_{eq}^{PM} with and without MM1 competition (equilibrium not reached in the case of (D)). **(E)**, degree of specificity, as defined in Eq. (4.4), using the values from (A)-(D). The high specificity of the PM diminishes with decreasing temperatures.

of freedoms (section 4.3.4) that are particularly pronounced at the melting temperature (section 4.3.5). The binding affinities of the competing targets K^{HAT} and K^{LAT}, probed in experiments without competition (section 4.1.1), fail to describe the equilibrium duplex concentration of the HAT. This raises the question, if the high specificity is caused by a molecular interaction that improves K^{HAT} in a competitive environment by several orders of magnitude compared the value measured in the individual case. This is analyzed in the following experiments. For the competitive system "PM vs. MM1", the determination of the binding affinity K^{PM} (see section 4.1.1) is repeated in a background of MM1 targets, concentrated at 1 µM, for temperatures of 44 °C and 35 °C. By fitting the extended Langmuir isotherm to the data, one can estimate the binding affinity K^{PM} during competition with the MM1. The calibration of the hybridization signal into units equivalent to the duplex concentration in the measurement chamber (see appendix section A.3) is only valid for a temperature of 44 °C. Thus, for extracting the binding affinities from the experiments, the modified version of the Langmuir isotherm (Eq. (4.3)) is used. In order to validate the extracted binding affinities in competition they are compared with the respective values without competition at 44 °C (already determined in section 4.1.1) and at 35 °C (determined in this section).

All extracted binding affinities are combined in Table 4.5. Fig. 4.20 shows the extended Langmuir isotherms of PM and MM1 without competition, at a temperature of 35 °C. Both molecules improve their binding affinity for the probe at this temperature by about one order of magnitude compared to the values at 44 °C. This is not surprising, since the reduced number of temperature fluctuations enables the formation of more stable duplexes at 35 °C. Fig. 4.21 shows the extended Langmuir isotherm of the PM in the presence of 1 µM of the MM1 targets, for a temperature of 35 °C and 44 °C, respectively. In spite of the MM1 background, the PM hybridizes as if its competitor was not present at all. For both temperatures, 35 °C and 44 °C, the extracted binding affinities of the PM are of the same order of magnitude as the respective values probed without competition. The PM does not improve its hybridization affinity in the competitive environment. Put together, the experiments show that the presence of the LAT does not affect the affinity of the HAT for binding to the probe molecule. Thus, one can conclude that in the case of the highly specific hybridization, the HAT has to decreases the binding affinity of the LAT by several orders of magnitude. However, this requires an asymmetric interaction between the HAT and the LAT, since otherwise it would be impossible to change the binding free energy landscape in favor of the HAT.

One has to mention that, generally, the extraction of binding affinities with the extended Langmuir isotherm is only valid for data obtained from single hybridization experiments.

Table 4.5.: Binding affinities of PM and MM1 in single and competitive hybridization experiments at 44 °C and 35 °C. The values of K^{PM} and K^{MM1} at 44 °C in experiments without competition are determined in Fig. 4.1 using the extended Langmuir isotherm. The respective values at 35 °C are determined in Fig. 4.20. Compared to 44 °C, K^{PM} and K^{MM1} are increased by about one order of magnitude at 35 °C. The binding affinity of the PM in the presence of the MM1 is determined in Fig. 4.21. K^{PM} in competition is not reduced compared to the respective value without competition. The presence of the MM1 does not affect the affinity of the PM for the probe molecule.

	K @ 35°C [10^8 1/M]	K @ 44°C [10^8 1/M]
PM single hybridization	26 ± 5	9 ± 5
MM1 single hybridization	2.8 ± 0.8	0.6 ± 0.1
PM in the presence of MM1	12 ± 4	21 ± 9

However, the main objective of the present experiments is to estimate the binding affinity of the HAT in competition and to investigate, if the value of K^{HAT} does change dramatically compared to the case without competition. For this purpose, the approach described above is adequate, because the background of MM1 targets can simply be treated as a perturbation of the PM single hybridization.

4.3.8. Competitive denaturation experiments

In order to gain further insight into the temperature dependent interaction between HAT and LAT, additional experiments are performed. The conclusions of the previous section claim an interaction between the highly specific HAT and the LAT that reduces K^{LAT} by several orders of magnitude. In the case of a highly specific system, such an interaction should change the melting profile of the LAT compared to the competition free case. For standard specificity the melting transition of the LAT should not be affected. This is analyzed in the following measurements. For the competitive systems "MM4 vs. MM1" (standard specificity, compare Fig. 4.4 A) and "PM vs. MM1" (high specificity, compare Fig. 4.4 B) the melting curve of the MM1, which constitutes the LAT in both cases, is measured for increasing concentrations of the high affinity targets MM4 and PM, respectively. The general experimental approach for measuring melting curves is presented in section 4.3.4. By measuring the surface occupancy Θ^{MM1}, the binding affinities K^{MM1} at 44 °C are determined as a function of the concentration of the respective HAT, thereby using the Langmuir isotherm (compare Eq. (2.11))

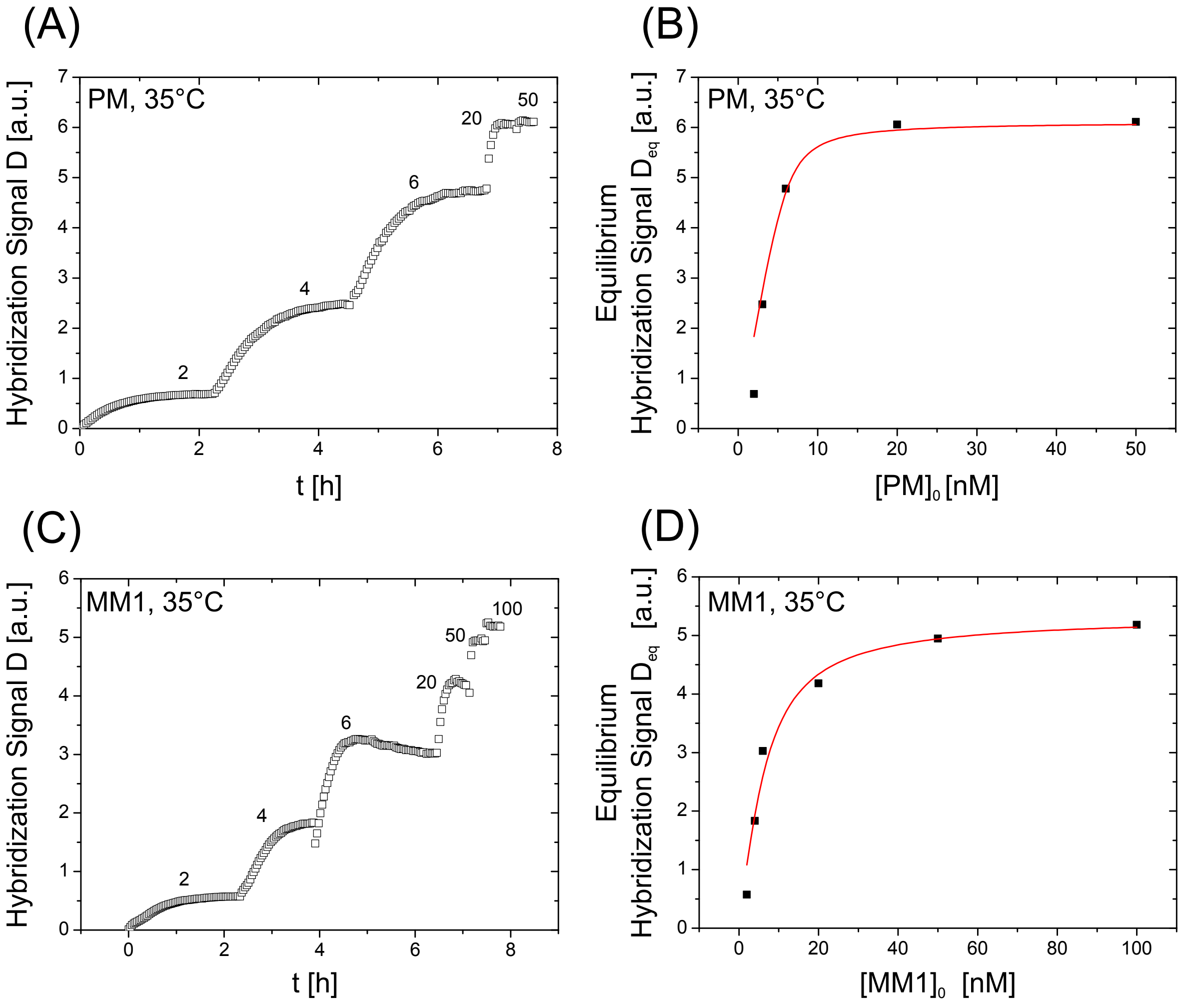


Figure 4.20.: Determination of binding affinities at 35 °C for target molecules PM and MM1. **(A), (C)**, raw data. The graphs show the hybridization signal D for PM (A) and MM1 (C) as a function of time. Following an increase of the target concentration $[T]_0$ (respective values in nM), D is measured until it reaches its equilibrium value D_{eq} and remains constant in time. Procedure is repeated until all probe molecules are occupied. **(B), (D)**, extended Langmuir isotherms. The respective values for D_{eq} from (A) and (C) are plotted against the target concentrations. Fitting of Eq. (4.3) to this data (red line) yields the binding constants K of the investigated target molecule in units 1/M. The values are shown in Table 4.5.

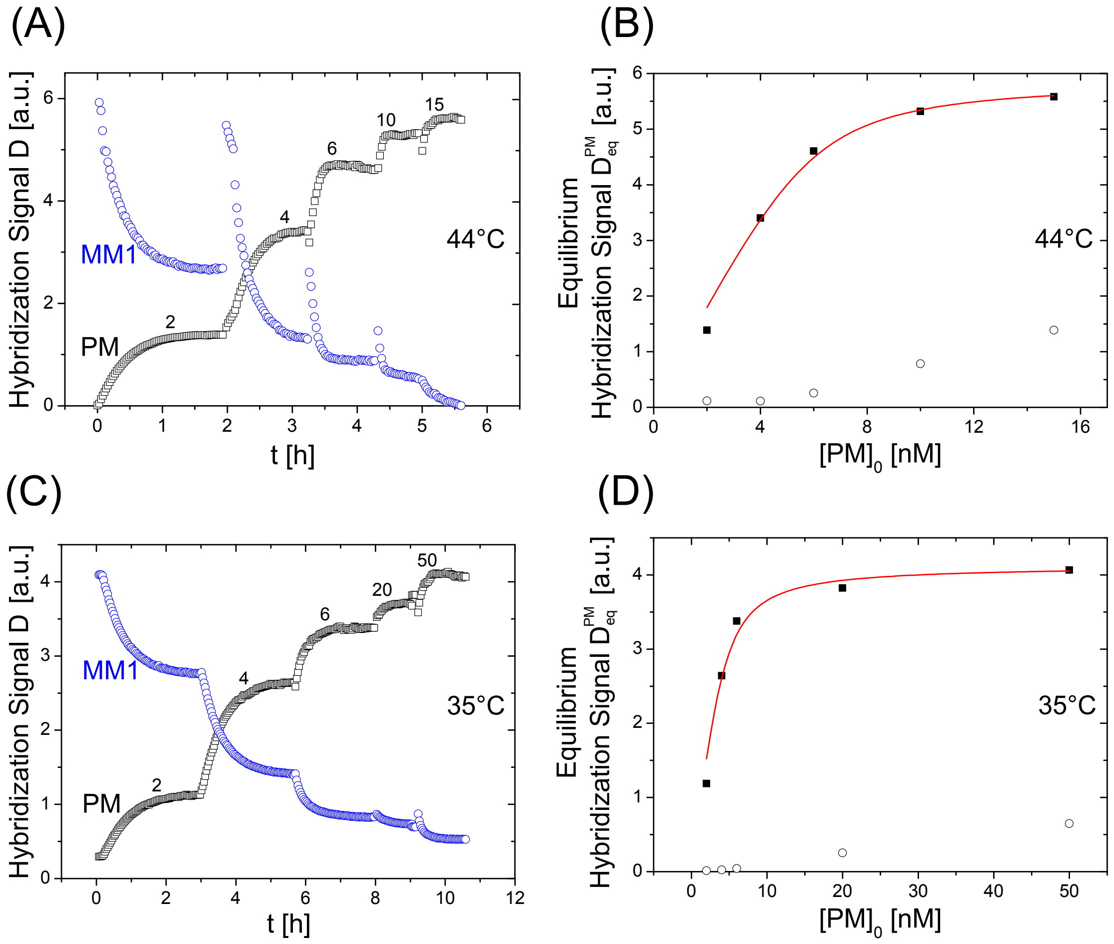


Figure 4.21.: Determination of the PM binding affinities K^{PM} at 44 °C and 35 °C in the presence of 1 µM MM1. **(A), (C)**, raw data. The graphs show the hybridization signals of the PM (black) and MM1 (blue) at 44 °C (A) and 35 °C (C) as a function of time with a background of 1 µM of MM1 targets. Following an increase of the PM target concentration $[PM]_0$ (respective values in nM), D is measured until it reaches its equilibrium value D_{eq} and remains constant in time. The procedure is repeated until all probe molecules are occupied by PM targets and all MM1 targets are released. **(B), (D)**, extended Langmuir isotherms for the PM. The respective values of D_{eq}^{PM} from (A) and (C) are plotted against $[PM]_0$. Fitting of Eq. (4.3) to this data (red line) yields the binding constants K of the PM in units 1/M at the respective temperature. The values are listed in Table 4.5. The open circles present the predictions of Eq. (4.2) for the equilibrium hybridization signals, thereby using the binding affinities of PM and MM1 at 44 °C and 35 °C probed without competition (see Table 4.5). In spite of the MM1 background, the PM hybridizes as if its competitor was not present at all.

$$\Theta^{MM1} = \frac{K^{MM1} \cdot [MM1]_0}{K^{MM1} \cdot [MM1]_0 + 1} \tag{4.10}$$

Fig. 4.22 shows the melting curves of the MM1 in competition with the targets MM4 and PM, respectively. In the case of the highly specific system "PM vs. MM1", the melting transition of the MM1, concentrated at 1 µM, clearly depends on the presence of the PM. By slightly increasing the PM concentration $[PM]_0$ in a stepwise manner, the melting temperature T_m of the MM1 can be shifted from 53 °C in the competition free case (also compare Fig. 4.17) to about 45 °C in the presence of 5 nM PM. This shift corresponds to a significant affinity loss of the MM1 by almost two orders of magnitude. However, performing the same experiment with the standard specific system "MM4 vs. MM1" reveals the MM1 melting profile to be completely unaffected by the presence of the competitor MM4. The binding affinity of the MM1 remains within the same order of magnitude as the one probed without competition, even for a concentration of 5 nM of the MM4 targets. In agreement with the conclusions of the previous section 4.3.7 the results confirm that the highly specific HAT interacts with the LAT and indeed reduces its binding affinity K^{LAT} by orders of magnitude. For a competitive system of standard specificity this interaction obviously does not exist and the presence of the HAT does not affect K^{LAT}.

One has to note that, generally, the extraction of the binding affinities from the melting curves with Eq. (4.10) is only valid for a single hybridization. In competition, the surface occupancy Θ of a specified target molecule A is also a function of the concentration and the binding affinity of the competing species B. Anyway, Eq. (4.10) is used to extract K^{MM1} from the competitive denaturation experiments, since the objective of these experiments is to investigate a general, qualitative dependence of the binding affinities on the presence of both, a highly specific (PM) and a standard specific (MM4) competitor.

4.3.9. Fluctuation dynamics of PM and MM1

In order to gain further insight into the temperature dependent interactions between the highly specific HAT and the LAT, fluorescence correlation spectroscopy (FCS) is performed. FCS is a single molecule technique that can address fluctuation dynamics in double-stranded DNA [34]. Thermal excitations lead to "DNA breathing" due to local denaturation and reclosing of the double-stranded structure. In order to evaluate the fluctuation dynamics of the duplexes PM-Probe and MM1-Probe, the FCS setup

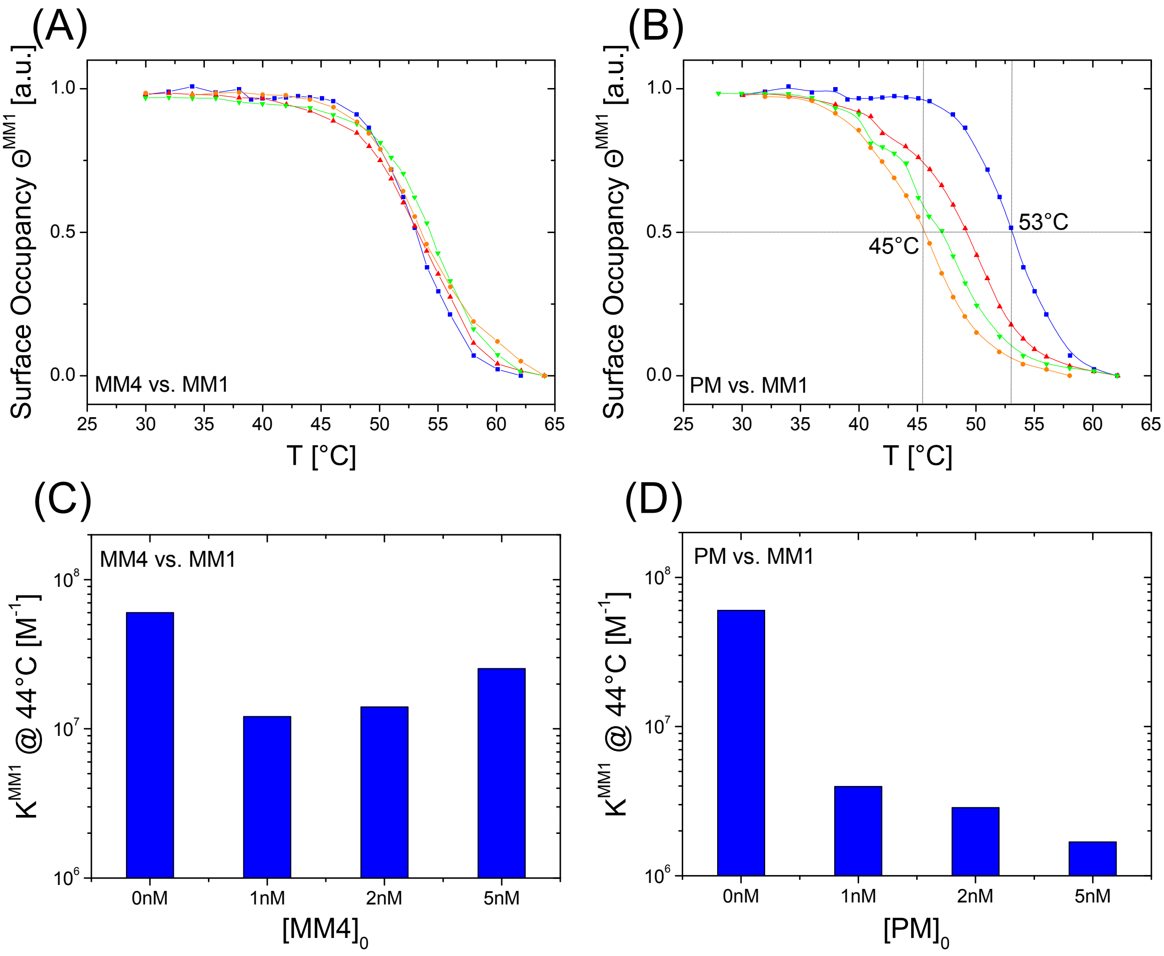


Figure 4.22.: Competitive denaturation experiments. **(A)**, standard specific system "MM4 vs. MM1". The graphs show the surface occupancy Θ^{MM1} of the low affinity target MM1, concentrated at 1 µM, as a function of temperature T, in the presence of the MM4, concentrated at 0 nM (no competition, blue), 1 nM (red), 2 nM (green) and 5 nM (orange). The presence of the MM4 does not lead to a measurable change of the MM1 melting profile. **(B)**, same as (A) but for the highly specific system "PM vs. MM1". In this case an equivalent concentration increase of the PM leads to a shift of the melting temperature $T_m = T\left(\Theta^{MM1} = 0.5\right)$ from 53 °C in the competition free case (blue) to 45 °C in the case of $[PM]_0 = 5$ nM (orange). **(C), (D)**, extracted binding affinities K^{MM1} at 44 °C from the data in (A) and (B) by using the Langmuir isotherm (Eq. (4.10)). For "PM vs. MM1" (D) the shift of the melting profile corresponds to a significant loss of the MM1 binding affinity by almost two orders of magnitude. However, for the standard specific system "MM4 vs. MM1" (C), K^{MM1} remains within the same order of magnitude for all PM concentrations.

described in chapter 3.1.7 is used. The respective target molecules and the probe molecule constitute the FRET pair and the correlation function $G(t)$ of the fluctuation dynamics of both duplexes PM-Probe and MM1-Probe is measured at room temperature and 44 °C, respectively. Additionally, it is analyzed, if the presence of the PM has any influence on the fluctuation dynamics of the MM1, as indicated by the melting curves in competition (section 4.3.8). For temperatures $T < T_m$ the shape of the correlation function $G(t)$ is mainly dominated by the reclosing dynamics. In this case, the fluctuation dynamics of double-stranded DNA can be described with the expression

$$G(t) \propto \left(1 + \frac{t}{2\tau}\right) \cdot \operatorname{erfc}\left(\sqrt{\frac{t}{4\tau}}\right) - \sqrt{\frac{t}{\pi\tau}} \cdot \exp\left(-\frac{t}{4\tau}\right) \tag{4.11}$$

with $\operatorname{erfc}(u) = 1 - \frac{2}{\sqrt{\pi}} \cdot \int_0^u \exp\left(-x^2\right)\, dx$ [34]. Fitting of the measured correlation functions with Eq. (4.11) reveals the characteristic time constant τ of the relaxation dynamics of the investigated double strand.

Fig. 4.23 shows the correlation function of the fluctuation dynamics of the duplexes PM-Probe and MM1-Probe at room temperature and 44 °C. The overall relaxation time constants τ^{PM} and τ^{MM1} are listed in Table 4.6. The experiment reveals that at room temperature the time constants for PM and MM1 are different but comparable. This indicates similar fluctuation dynamics of the two duplexes. At 44 °C, however, the relaxation dynamics of the MM1 is markedly slower. In agreement with the melting curves presented in section 4.3.4, one can conclude that the MM1 duplex exhibits an increased number of entropic degrees of freedom that leads to slower relaxation dynamics. Fig. 4.23 also shows that the presence of the PM, concentrated at 100 pM, modifies the correlation function of the MM1 compared to the competition free case. In agreement with the findings in section 4.3.3 and 4.3.8, this indicates an interaction between the highly specific PM and the MM1. Note that the overall relaxation time constant τ cannot be extracted from the measured correlation function in the case of competition, since the model in [34] does not account for a competitive situation.

One has to note that there are differences between the experimental conditions used in this work and the ones presented in [34]. The DNA duplexes PM-Probe and MM1-Probe are not hairpin loops and are lacking GC clamps. Additionally, the experiments in this work are performed at a higher salt concentration (0.57 M monovalent ions). Nevertheless, Eq. (4.11) is used to extract the corresponding time constants since a qualitative difference between the fluctuation dynamics of PM-Probe and MM1-Probe at room temperature and 44 °C is sufficient for further interpretation of the data.

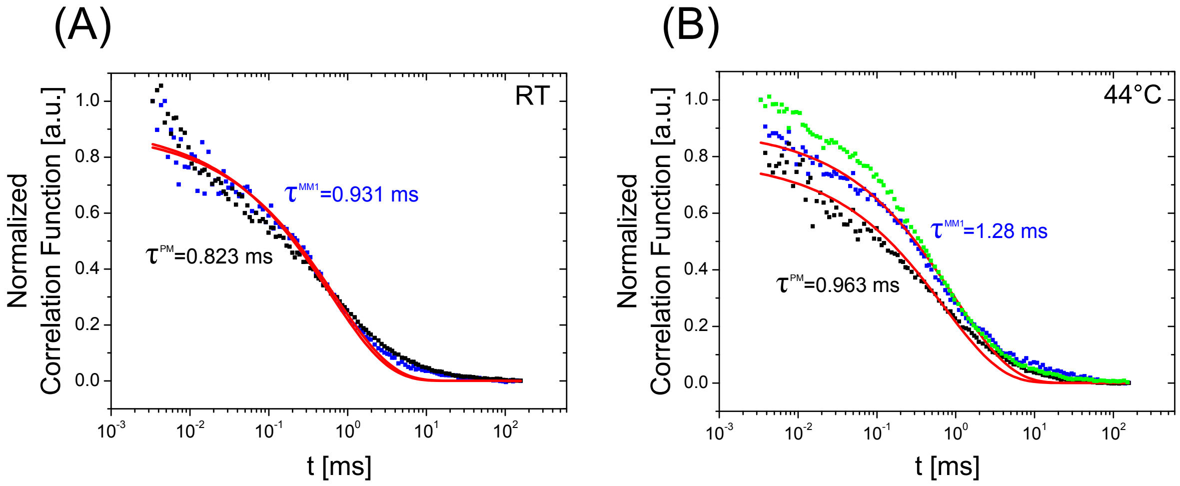


Figure 4.23.: Fluctuation spectra of double-stranded DNA molecules PM-Probe and MM1-Probe in bulk, obtained with the FCS setup (compare chapter 3.1.7). All correlation functions are the mean value of five independent measurements (see appendix section A.9). **(A)**, experiment performed at room temperature (RT). The graph shows the correlation functions as a function of time for the PM-Probe duplex (black) and the MM1-Probe duplex (blue). The concentration of duplexes is 10 nM in both cases. The red line is the fit of Eq. (4.11) to the data, thus revealing the overall relaxation time constants τ^{PM} and τ^{MM1}. The values are different but comparable. **(B)**, same as (A) but for a temperature of 44 °C. The time constants now differ more strongly, indicating different fluctuation dynamics of the two duplexes at this temperature. The green curve presents the correlation function of the MM1-Probe duplex within a background of PM targets, concentrated at 100 pM. The presence of the PM leads to modified fluctuation dynamics of the MM1-Probe duplex (compare the blue and the green curve), indicating an interaction between PM and MM1 on the probe strand.

Table 4.6.: Overall relaxation time constants τ of the fluctuation dynamics of DNA duplexes PM-Probe and MM1-Probe, for room temperature and 44 °C. The values are determined by fitting Eq. (4.11) to the correlation functions presented in Fig. 4.23. While the time constants of PM-Probe and MM1-Probe are similar at room temperature, they differ more strongly at 44 °C.

Duplex	**τ at room temperature [ms]**	**τ at 44°C [ms]**
PM-Probe	0.823 ± 0.06	0.963 ± 0.06
MM1-Probe	0.931 ± 0.05	1.28 ± 0.05

4.3.10. Limiting entropic degrees of freedom: Measurements with LNA

The experimental studies show that the MM1-Probe duplex exhibits a fluctuating molecule with an increased number of entropic degrees of freedom (sections 4.3.4 and 4.3.9). This leads to the asymmetric interaction that causes the high specificity of the PM (sections 4.3.3 and 4.3.8). Thus, it should be possible to reduce the effect of the highly specific hybridization of the PM by limiting the entropic degrees of freedom in the MM1-Probe duplex. This is analyzed in the following experiments. Locked nucleic acids (LNA) consist of a ribose moiety modified with an extra methylene bridge connecting the 2′ oxygen and 4′ carbon. Compared to a DNA nucleotide, the stiffness is increased. Consequently, the entropic degrees of freedom of a DNA strand are reduced when LNA monomers are incorporated into the polymer. For the competitive system "PM. vs. MM1" as a representative of the highly specific cases, an increasing number of DNA nucleotides within the MM1 sequence is replaced by LNA nucleotides. Therefore, the LNA molecules MM1-LNA-1 and MM1-LNA-2 listed in Table 4.7 are designed. In competitive hybridization experiments, the high specificity of the PM in competition with the MM1 is investigated as a function of incorporated LNA monomers. The LNA molecules can only be tagged with the dyes provided by the vendor *Exiqon*. Compared to the usually employed dyes Cy-3 and Cy-5, these modifications additionally stabilize double-stranded DNA (see appendix section A.8). Thus, for the following experiments, the DNA targets PM and MM1 are also tagged with the dyes provided by *Exiqon* in order to compare the experimental data of all molecules.

Fig. 4.24 shows the result of the experiments. Indeed, by replacing DNA bases by the corresponding LNA bases within the MM1 target, the high specificity of the PM can be diminished. For the case of competition between PM and MM1-LNA-2, the high specificity of the PM is no longer observable. Thus, the experimental results confirm

Table 4.7.: MM1 target molecules with incorporated LNA bases. The non-matching base is underlined. DNA bases that are replaced by the corresponding LNA bases are shown in green.

Target	**Sequence (5′ → 3′)**
MM1	AAG-GAT-CAC-ATC-GTA-A
MM1-LNA-1	AAG-GAT-CAC-ATC-GTA-A
MM1-LNA-2	AAG-GAT-CAC-ATC-GTA-A

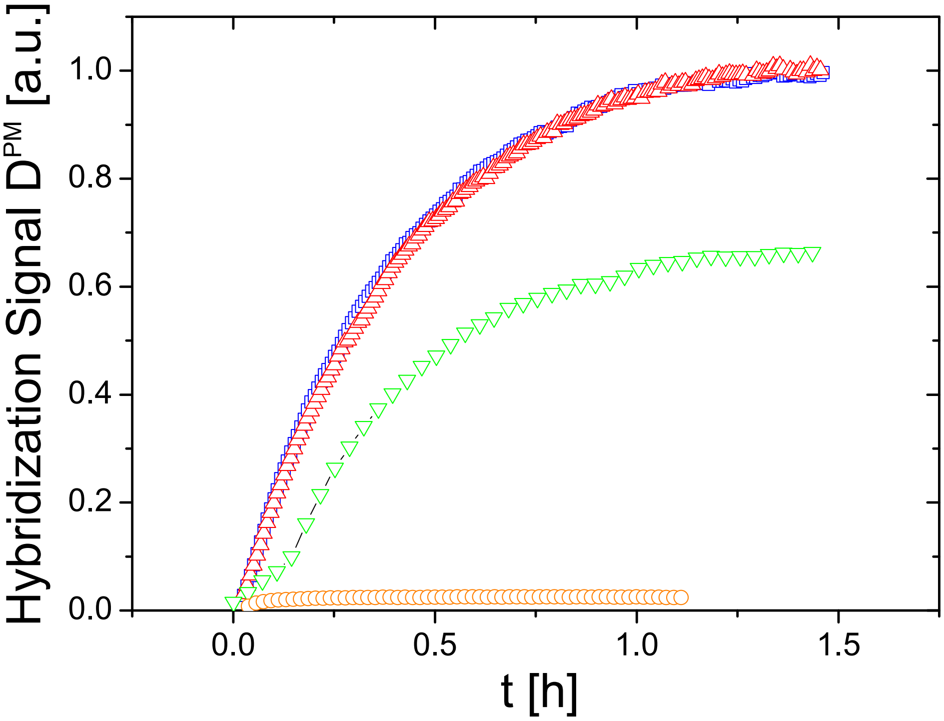


Figure 4.24.: Competitive hybridization between PM and different versions of the MM1 target at 44 °C. The graphs shows the hybridization signal of the PM (labeled with the fluorescent dye 665 provided by *Exiqon*) as a function of time for a single hybridization (blue), and for the case of competition with the MM1 (red), MM1-LNA-1 (green) and MM1-LNA-2 (orange). All the competitors are labeled with the dye 563, also provided by *Exiqon*. The PM concentration is 5 nM and the concentration of the competitor is 1 µM in all measurements. D^{PM} is normalized to the equilibrium value of the single hybridization (blue curve). By replacing more and more DNA bases of the MM1 strand by LNA bases, the high specificity of the PM diminishes.

the conclusions of the previous section 4.3.9. The entropic degrees of freedom in the MM1-Probe duplex at 44 °C are responsible for the occurrence of the highly specific PM hybridization. By replacing DNA nucleotides by LNA nucleotides within the MM1 sequence, these fluctuations can be reduced and the high specificity of the PM diminishes.

As discussed in section 4.3.2, one has to mention that quantitative conclusions about the specificity of a competitive system are only possible, if the binding affinities of both competitors K^{HAT} and K^{LAT} are known. However, in the present experiments, the binding affinities of the LNA molecules MM1-LNA-1 and MM1-LNA-2 are not determined. LNA is reported to increase the binding affinity of a DNA target for binding to a complementary probe molecule [43] and also improves mismatch discrimination [44]. Thus, the decrease of the high specificity of the PM could basically be the result of the increased binding affinity of the target molecules MM1-LNA-1 and MM1-LNA-2. Nevertheless, also in this case, the results confirm the conclusions of the previous section 4.3.9. The qualitative observation that 1 μM of the MM1 targets are discriminated by only 5 nM of the PM targets, but 1 μM of MM1-LNA-2 targets obviously discriminate 5 nM of the PM targets, supports the hypothesis that the entropic degrees of freedom in the MM1-Probe duplex play an essential role for the occurrence of the high specificity.

4.4. Explanation of the high specificity: Model and numerical assessment

4.4.1. Conclusion of the experimental results

The experimental results presented in the previous section point towards a molecular interaction between the competitors HAT and LAT as the reason for the observed high specificity (sections 4.3.3, 4.3.8, 4.3.9). The temperature dependence (sections 4.3.5, 4.3.6 and reference [30]) underlines that different thermal fluctuations (section 4.3.9), involving different entropic degrees of freedom of the competing strands (sections 4.3.4, 4.3.10) play an essential role for the occurrence of the effect. Put together, one can conclude from the experimental results that in the highly specific cases, the competing strands interact via the mentioned thermal fluctuations and the effective binding free energies change in such a way that solely the HAT is able to bind to the probe molecules by reducing the binding affinity of the LAT (section 4.3.8). The HAT is not affected by the molecular competition with the LAT (section 4.3.7), which underlines that the interaction has to be asymmetric. Otherwise, it would be impossible to tilt the free energy landscape in favor of one of the two competitors. The effect also requires a highly collective reaction in order to generate the high non-linearities which are necessary to produce the observed strong change in the hybridization specificity of the HAT as a response to the shift of the mismatch position within the LAT sequence (section 4.3.2).

Although the interactions between the highly specific HAT and the LAT are a matter of thermal fluctuations that are difficult to picture in detail, in the following a numerical assessment is developed that is based on the experimental results. The objection of the numerical assessment is to model the observed high specificity in thermodynamic equilibrium with a reasonable choice of parameters. The numerical assessment is developed in strong collaboration with Christian Trapp and therefore also presented in [36].

4.4.2. Pre-melted and helicoidal binding configurations

The experimental results point towards the existence of different fluctuating binding configurations of HAT and LAT in the highly specific cases. For temperatures close to the melting point T_m, the existence of pre-melted configurations in double-stranded DNA is well known [45, 46]. Two different binding configurations are considered that may coexist on the same double strand (see Fig. 4.25 A):

1. Pre-melted configuration
   The pre-melted configurations occur predominantly close to the melting temperature of the DNA double strand. It is a flexible configuration with many entropic degrees of freedom. A closed base pair generates the enthalpy Δh. The overall enthalpic contribution of a given pre-melted configuration is simply the sum over all closed bonds. Additionally, a next neighbor coupling energy j is considered between adjacent closed bonds.

2. Helix configuration
   In this case the DNA duplex forms the well known double helix with low entropy and flexibility. Two adjacent pre-melted base pairs gain the next neighbor coupling energy $J > j$ when they transit from a pre-melted configuration to a helix configuration. This is, however, at the cost of a reduced number of entropic degrees of freedom of the strand.

In the following, the implementation of the two binding configurations is explained in detail.

4.4.3. Free energy of pre-melted configurations

In order to calculate the enthalpy $\Delta H^p_{\vec{v}}$ of a pre-melted configuration, a collective interaction between adjacent bases is considered using an ISING description. This is because a specified base possesses an increased closing probability if its neighbor is already bound. This leads to the following expression:

$$\Delta H^p_{\vec{v}} = \Delta H^p_{\vec{v}}(\vec{v}) = \sum_{i=1}^{M} j \cdot v_i \cdot v_{i+1} + \sum_{i=1}^{M+1} \delta(v_i - 1) \cdot \Delta h \tag{4.12}$$

M is the maximum number of nearest neighbor pairs in the DNA duplex [5, 47]. For instance, a DNA duplex with 16 bases possesses $M = 15$ next neighbor pairs. $\vec{v}$ is a vector of length $M + 1$ with entries 0 and 1. It represents a pre-melted binding configuration between target and probe, where 1 stands for a closed base pair and 0 for an open base pair. j is the coupling constant between two adjacent closed base pairs and Δh is the enthalpy gain for the closing of a base pair. The values of Δh are positive, assuming that the reference state is the totally denatured state, where all base pairs are open.

The transition from the totally denatured strand to a pre-melted configuration causes a entropy change of the strands, because any pre-melted configuration has less degrees of freedom compared to the fully denatured state. The entropy change is determined by treating the conformations of a strand as a self-avoiding walk (SAW) on a lattice [36, 48, 49]. For simplicity it is considered that one step in the lattice corresponds to one unbound base of the DNA strand. Thus, the number of configurations $\#_{SAW}$ of length x is

$$\#_{SAW}(x) = \mu^x \cdot x^{\gamma - 1} \tag{4.13}$$

$\gamma = 1.157 \pm 3 \cdot 10^{-3}$ is the universal entropic constant and $\mu = 4.684$ depends on the considered geometry [48]. Thus, the entropy change $\Delta S^p(y)$ between the totally denatured state and a pre-melted configuration with y closed base pairs is

$$\Delta S^p(y) = R \cdot \begin{cases} \ln\left[\frac{\#_{SAW}(M+1)}{\#_{SAW}(M+1-y)}\right] & \text{, if } y < M + 1 \\ \ln\left[\#_{SAW}(M+1)\right] & \text{, if } y = M + 1 \end{cases} \tag{4.14}$$

where R is the molar gas constant. The pre-melted configuration where all base pairs are closed corresponds to the largest entropy change $\Delta S^p(y)$. Ultimately, the binding free energy $\Delta G^p_{\vec{v}}$ of a pre-melted configuration at temperature T is

$$\Delta G^p_{\vec{v}} = \Delta H^p_{\vec{v}} - T \cdot \Delta S^p \tag{4.15}$$

following the Gibbs equation (compare for instance Eq. (2.26)).

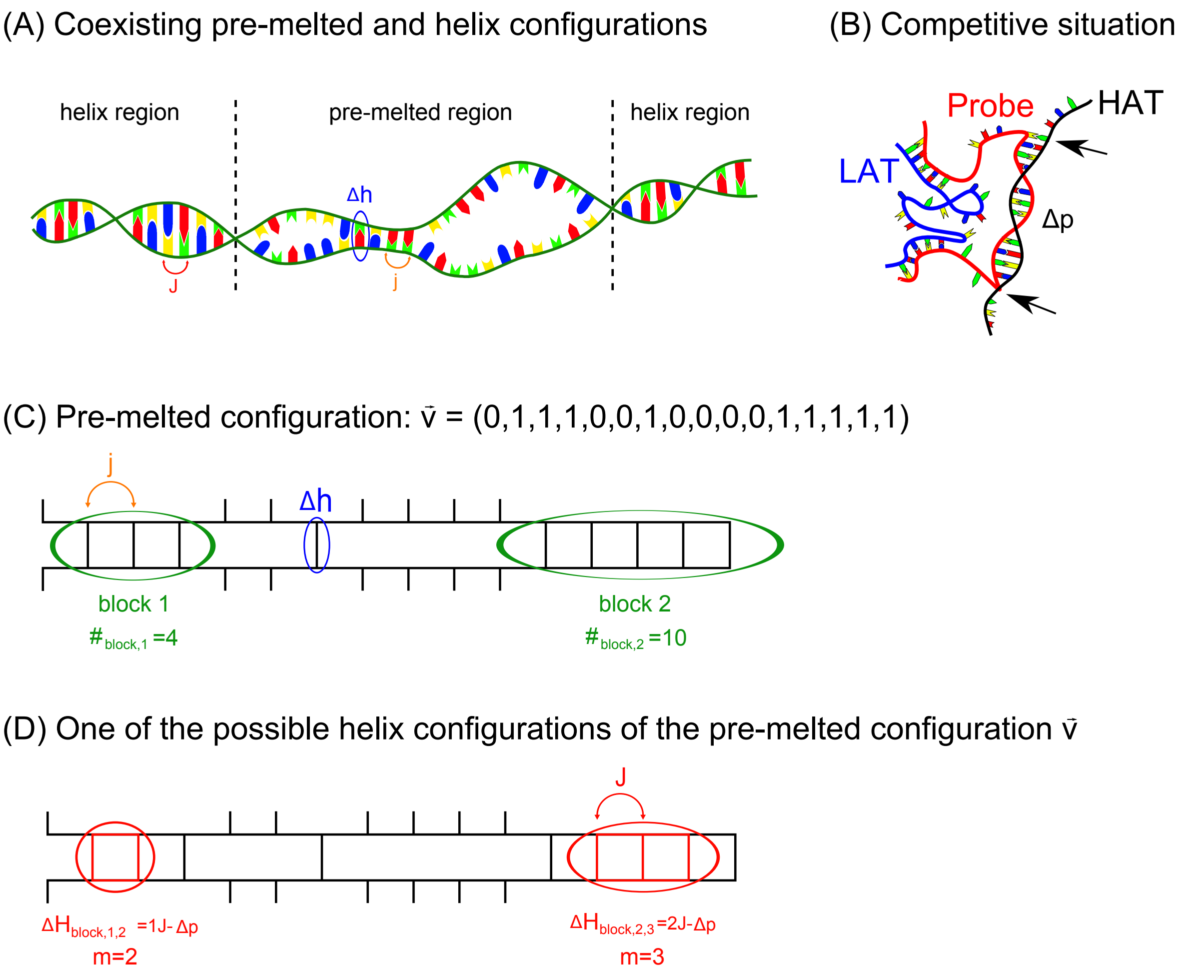


Figure 4.25.: Model and numerical assessment (compare text for details). **(A)**, coexisting pre-melted and helix configurations. The flexible pre-melted configuration (middle region) possesses many entropic degrees of freedom. A closed base pair gains the enthalpy Δh and two adjacent closed base pairs gain the coupling energy j. The stiff helix configurations (outer regions) possess a low number of entropic degrees of freedom. Two adjacent closed base pairs in a pre-melted configuration gain the additional next neighbor coupling energy $J > j$ when they transit to a helix configuration. **(B)**, competitive situation. Formation of a triplex structure consisting of HAT, LAT and the probe causes an interaction between the DNA molecules. If one specified target molecule resists the deformation that a competitor has to apply to bind to the probe itself, in a mean field picture this can be described by using an average energy penalty Δp that increases the required degree of cooperativity of helix configurations. **(C, D)**, numerical assessment. (C), pre-melted configuration $\vec{v} = (0,1,1,1,0,0,1,0,0,0,0,1,1,1,1,1)$. The enthalpy $\Delta H^p_{\vec{v}}$ of this configuration is determined with Eq. (4.12). Together with the entropy change ΔS^p between the totally denatured state and this pre-melted configuration (see Eq. (4.14)), the binding free energy $\Delta G^p_{\vec{v}}$ is determined with Eq. (4.15). The number of

blocks (segments of closed bases next to each other without interruption) for this pre-melted configuration $\vec{v}$ is 2 (green circles). According to Eq. (4.17) the number of possible helix configurations k of this specific pre-melted configuration $\vec{v}$ is $\#_{block,1} \cdot \#_{block,2} = 4 \cdot 10 = 40$. (D), one of the 40 possible helix configurations of the pre-melted configuration $\vec{v}$ in (C). The red parts form a helix. The additional enthalpy $\Delta H_{block,i,m}$ of such a helix follows Eq. (4.18). Together with the entropy change S^h, $\Delta G^h_{\vec{v},k}$ is determined with Eq. (4.21). The combination of $\Delta G^p_{\vec{v}}$ and $\Delta G^h_{\vec{v},k}$ gives the total free energy $\Delta G_{\vec{v},k}$ of a specified binding configuration (see Eq. (4.22)).

4.4.4. Free energy of coexisting pre-melted and helix configurations

Two adjacent closed base pairs in the pre-melted configuration can transit in pairs to a helicoidal configuration, thereby generating the additional next neighbor coupling energy $J > j$. In order to calculate the number k of all possible helix configurations of a specified pre-melted configuration $\vec{v}$, one has to determine the total number $\#_{blocks}$ of individual "blocks" i of adjacent closed base pairs in the pre-melted configuration of length $l_i > 2$. The individual blocks are independent of each other and can form a helix configuration. For instance, the pre-melted configuration shown in Fig. 4.25 C, possesses two individual blocks of length $l_1 = 3$ and $l_2 = 5$. In Fig. 4.25 D, two of the bases of block 1 and three of the bases of block 2 have formed a helix. By considering the bases of individual blocks follow the double ended Zipper binding statistics [5, 50], the number of possible configurations $\#_{block,i}$ of one individual block i of length l_i is

$$\#_{block,i} = \frac{l_i \cdot (l_i - 1)}{2} + 1 \tag{4.16}$$

With this expression, the total number k of all possible helix configurations of one specific pre-melted configuration $\vec{v}$ is

$$k = \prod_{i=1}^{\#_{block}} \#_{block,i} \tag{4.17}$$

If $2 \leq m \leq l_i$ bases of an individual block i are in helix configurations, the enthalpy gain is

$$\Delta H_{block,i,m} = (m-1) \cdot J - \Delta p \tag{4.18}$$

As mentioned in section 4.4.1, the experimental results point towards an asymmetric interaction between the highly specific HAT and the LAT. For the numerical assessment it is assumed that these interactions are the result of the formation of a triplex structure of short lifetime, consisting of HAT, LAT, and the probe molecule (see Fig. 4.25 B). In order to account for this interaction, the energy penalty Δp is included. Δp reflects the energy required to form a helicoidal configuration in the presence of a competing target that is bound to the probe and resists the deformation. To compensate for Δp, a larger number of adjacent bases in the helix configuration are needed. In this sense, Δp increases the required degree of cooperativity to form a helix configuration in the presence of a competitor. From this it follows that Δp is zero for the completely developed helix configuration and for the competition free case (single hybridization). The latter aspect is also in agreement with the nearest neighbor standard model of DNA hybridization [11]. Furthermore, Δp is zero, if the interaction between HAT and LAT is symmetric and the triplex population does not favor one of the two competitors. For the numerical assessment, Δp is expressed as a multiple integer of the coupling energy J of the helix configuration. For the calculation of the partition function only blocks with $\Delta H_{block,i,m} > 0$ need to be considered. If in Eq. (4.18)

$$(m-1)\cdot J \leq \Delta p \Leftrightarrow \Delta H_{block,i,m} \leq 0 \tag{4.19}$$

the corresponding configuration does not reach the required degree of cooperativity and is not taken into account. Therefore Eq. (4.16) is a function of Δp. The additional enthalpy $\Delta H^h_{\vec{v},k}$ generated by the helix configurations is obtained by summing over all individual blocks

$$\Delta H^h_{\vec{v},k} = \sum_{i=1}^{\#_{block}} \Delta H_{block,i,m} \tag{4.20}$$

If there are no helix configurations, $\Delta H^h_{\vec{v},k}$ is zero and the stability of the duplex is only governed by the energy of the pre-melted configurations following Eq. (4.15). The transition from a pre-melted configuration to a helix configuration causes an additional entropy change $S^h(z)$ due to the increased persistence length. z is the number of bases in helix configurations and $S^h(z)$ is calculated as described in Eq. (4.13). Please note that the reference state for the calculation of $S^h(z)$ is still the totally denatured state. Thus, the completely developed helix configuration possesses the largest change in entropy. The binding free energy of the helix configurations is

$$\Delta G^h_{\vec{v},k} = \Delta H^h_{\vec{v},k} - T\cdot S^h \tag{4.21}$$

Ultimately, the overall binding free energy of any bound configuration can be calculated by summing the two contributions of pre-melted configurations (Eq. (4.15)) and helix configurations (Eq. (4.21)):

$$\Delta G_{\vec{v},k} = \Delta G^{p}_{\vec{v}} + \Delta G^{h}_{\vec{v},k} \tag{4.22}$$

In order to determine the binding affinity K of a specified target molecule, one has to determine the partition function Z_{duplex} of a DNA duplex, which is the sum over all possible pre-melted and helix configurations

$$K = Z_{duplex} = \sum_{\vec{v}} \sum_{k(\vec{v})} e^{\beta \Delta G_{\vec{v},k}} \tag{4.23}$$

The first sum runs over all possible pre-melted configurations $\vec{v}$ of the duplex and the second sum runs over all possible helix configurations k corresponding to one specific pre-melted configuration $\vec{v}$.

4.4.5. Numerical assessment of the single hybridization

Consider an ensemble with N^A molecules of species A and N binding sites with $N > N^A$. A binding site can be occupied by exactly one molecule and there is no interaction between adjacent binding sites. Binding of a molecule A to a binding site gives an energy $\Delta G^A + \mu^A$, where $\mu^A < 0$ is the chemical potential. The grand canonic partition function for this ensemble is

$$\mathcal{Z} = \sum_{N^A} \binom{N}{N^A} e^{\beta N^A \left(\Delta G^A + \mu^A\right)} = \left[1 + e^{\beta\left(\Delta G^A + \mu^A\right)}\right]^N \tag{4.24}$$

where $\beta = 1/RT$ and $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$. The term in squared brackets on the right side of Eq. (4.24) corresponds to the partition function of one single binding site. The average number $\langle N^A \rangle$ of occupied binding sites can be determined with

$$\langle N^A \rangle = \frac{\partial \ln Z}{\partial \left(\beta \mu^A\right)} = N \cdot \frac{e^{\beta\left(\Delta G^A + \mu^A\right)}}{1 + e^{\beta\left(\Delta G^A + \mu^A\right)}} = N \cdot \frac{K^A \cdot [A]_{eq}}{1 + K^A \cdot [A]_{eq}} \tag{4.25}$$

Here,

$$[A]_{eq} = e^{\beta \mu^A} \tag{4.26}$$

is the equilibrium concentration of molecules of species A and

$$K^A = e^{\beta \Delta G^A} \tag{4.27}$$

Table 4.8.: Individual theoretical binding constants K^{HAT} and K^{LAT} calculated with Eq. (4.23) and their ratio for a given parameter set $(j, J, \Delta h)$ and a temperature of 44 °C.

$\mathrm{K^{HAT}}$	$\mathrm{K^{LAT}}$	$\mathrm{K^{HAT}/K^{LAT}}$	**j [kcal/mol]**	**J [kcal/mol]**	**Δh [kcal/mol]**
$3.20 \cdot 10^7$	$1.88 \cdot 10^6$	17.02	0.3	1.7	0.7

is the binding constant. Please note that the right side of Eq. (4.25) is identical to the Langmuir isotherm in Eq. (2.7).

In the following, the N^A molecules of species A and the N binding sites are considered to be complementary DNA oligonucleotides (targets and probes). For simplicity, in the numerical assessment, the DNA strands constitute homopolymers with 16 bases in length. Following the model presented in sections 4.4.3 and 4.4.4, DNA molecules have a distribution of pre-melted and helix configurations. Thus, K^A in Eq. (4.25) is the partition function Z_{duplex} of all possible pre-melted and helix configurations using Eq. (4.23). The individual binding of two different target molecules is considered: a HAT that constitutes a perfect match and a LAT that possesses a single erroneous base at the ninth position of the strand (reading direction 5' → 3'). As the mismatching base pair of the LAT remains open, the LAT possess only 2^{15} pre-melted configurations plus the respective helix configurations, while the HAT has 2^{16} pre-melted configurations plus the respective helix configurations. The values for the next neighbor coupling energies j and J and the value for the enthalpy gain of a closed base pair Δh are needed for the calculation of $\Delta G_{\vec{v},k}$ (compare sections 4.4.3 and 4.4.4). They are unknown and therefore considered as free parameters. However, they need to be in qualitative agreement with the nearest neighbor standard model of DNA hybridization [11]. The mentioned parameters are chosen in such a way that the theoretical individual binding affinities K^{HAT} and K^{LAT}, calculated with Eq. (4.23), differ by one order of magnitude, as observed experimentally for the target molecules PM and MM1 (compare Table 4.1). Furthermore, as explained in section 4.4.4, Δp is zero, since K^{HAT} and K^{LAT} are calculated for the individual case without competition. Table 4.8 shows one reasonable parameter set and the resulting binding constants for a temperature of 44 °C. Fig. 4.26 shows how helix configurations of different sizes contribute to the calculated binding affinities K^{HAT} and K^{LAT} presented in Table 4.8. The numerical assessment reveals that for the same conditions, both targets, HAT and LAT hybridize to the probe molecules in a mixture of pre-melted and helix configurations. However, the LAT generates the major part of its binding affinity through a large number of short helix configurations

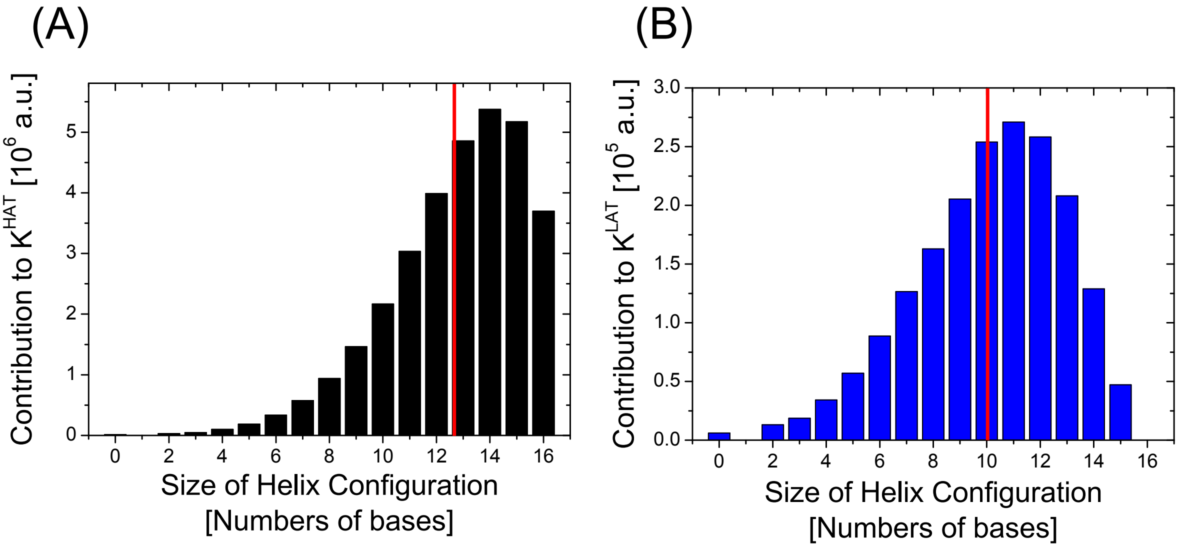


Figure 4.26.: Contribution of helix configurations of different sizes to the binding affinity K for HAT **(A)** and LAT **(B)** considered as homopolymers of 16 bases in length, hybridizing at a temperature of 44 °C. The numerical assessment is performed for the parameter set presented in Table 4.8. Summing over all the contributions of the helix configurations reveals the binding constants $K^{HAT} = 3.20 \cdot 10^7$ and $K^{HAT} = 1.88 \cdot 10^5$ (compare Table 4.8). The first bar (zero bases in helix configurations) reflects the contribution of all pre-melted configurations. There is no contribution of size one helices since a helix configuration consists of at least two bases. K^{HAT} is generated through almost completely closed helix configurations (mean size 12.7 bases, red line), while K^{LAT} stems from a larger number of short helix configurations (mean size 10 bases, red line).

of about 10 bases in length, while for the HAT more of the binding affinity stems from almost completely developed helix configurations. Thus, in agreement with the experimental results summarized in section 4.4.1, the numerical assessment confirms that LAT and HAT possess different numbers of entropic degrees of freedom involving different fluctuation dynamics at a temperature of 44 °C. This difference is required for the already claimed asymmetric interaction of HAT and LAT that leads to the observed high specificity. This is explained in detail in the next section 4.4.6.

4.4.6. Numerical assessment of the competitive hybridization

Consider an ensemble with N^A molecules of species A, N^B molecules of species B and N binding sites with $N^A + N^B < N$. A binding site can be occupied by exactly one molecule and there is no interaction between adjacent binding sites. Binding of a molecule A to a binding site gives the energy $\Delta G^A + \mu^A$, binding of molecule B the energy $\Delta G^B + \mu^B$. The grand canonic partition function of this ensemble is

$$\begin{aligned}
\mathcal{Z} &= \sum_{N^A} \sum_{N^B} \binom{N}{N^A} \binom{N - N^A}{N^B} e^{\beta\left[N^A\left(\Delta G^A + \mu^A\right) + N^B\left(\Delta G^B + \mu^B\right)\right]} \\
&= \sum_{N^A} \binom{N}{N^A} e^{\beta N^A\left(\Delta G^A + \mu^A\right)} \cdot \sum_{N^B} \binom{N - N^A}{N^B} e^{\beta N^B\left(\Delta G^B + \mu^B\right)} \\
&= \sum_{N^A} \binom{N}{N^A} e^{\beta N^A\left(\Delta G^A + \mu^A\right)} \cdot \left[1 + e^{\beta\left(\Delta G^B + \mu^B\right)}\right]^{N - N^A} \\
&= \left[1 + e^{\beta\left(\Delta G^A + \mu^A\right)} + e^{\beta\left(\Delta G^B + \mu^B\right)}\right]^N
\end{aligned} \tag{4.28}$$

The term in squared brackets in the last line of Eq. (4.28) corresponds to the partition function of one single binding site. In analogy to Eq. (4.25), the average number of binding sites occupied by molecule A is

$$\langle N^A \rangle = N \cdot \frac{e^{\beta\left(\Delta G^A + \mu^A\right)}}{1 + e^{\beta\left(\Delta G^A + \mu^A\right)} + e^{\beta\left(\Delta G^B + \mu^B\right)}} \tag{4.29}$$

and the average number of binding sites occupied by molecule B is

$$\langle N^B \rangle = N \cdot \frac{e^{\beta\left(\Delta G^B + \mu^B\right)}}{1 + e^{\beta\left(\Delta G^B + \mu^B\right)} + e^{\beta\left(\Delta G^A + \mu^A\right)}} \tag{4.30}$$

With the Eqs. (4.29) and (4.30) as well as the combined Eqs. (4.26) and (4.27), the ratio $\langle N^A \rangle / \langle N^B \rangle$ of the average number of occupied binding sites by molecules A and B is

$$\frac{\langle N^A \rangle}{\langle N^B \rangle} = \frac{e^{\beta\left(\Delta G^A + \mu^A\right)}}{e^{\beta\left(\Delta G^B + \mu^B\right)}} = \frac{K^A \cdot}{K^B \cdot} \cdot \frac{[A]_{eq}}{[B]_{eq}} = \frac{[A]_{eq}}{[B]_{eq}} \cdot e^{\beta \Delta\Delta G} \tag{4.31}$$

where K^A and K^B are the equilibrium binding constants and $\Delta\Delta G = \Delta G^A - \Delta G^B$ is the difference between the binding free energies between molecule A and B. Please note that for $[D^A]_{eq}/[D^B]_{eq} \propto \langle N^A \rangle / \langle N^B \rangle$ Eq. (4.31) is identical to Eq. (2.23). In the following, the molecules A and B are considered to be the DNA homopolymers HAT and LAT, both described in the previous section 4.4.5. The binding affinities K^{HAT} and

K^{LAT} in Eq. (4.31) can be determined again with the partition function Z_{duplex} using Eq. (4.23). However, in the competitive case, two different scenarios, concerning the kind of interactions between HAT and LAT, have to be distinguished:

1. Symmetric interactions between HAT and LAT

In this case, as explained in section 4.4.4, the population of the triplex consisting of HAT, LAT, and probe does not favor one of the two competitors, since the thermal fluctuations of HAT and LAT are comparable. In the numerical assessment this is reflected by an energy penalty of $\Delta p = 0$ equivalent to the single hybridization (section 4.4.5). Thus, the binding affinities K^{HAT} and K^{LAT} calculated with Eq. (4.23) are identical to the respective values determined for the competition free case (see Table 4.8). Put together, this means that without antagonistic interaction reflected by $\Delta p = 0$, the competitive hybridization is fully described with the single hybridization affinities K^{HAT} and K^{LAT} using Eq. (4.31). Based on the experimental results this scenario applies to the competitive systems of standard specificity (Fig. 4.4 A).

2. Antagonistic interactions between HAT and LAT

In this case, the interactions between HAT and LAT are asymmetric due to different entropic degrees of freedom involving different thermal fluctuations (compare section 4.4.4). In the numerical assessment, this asymmetry is reflected by energy penalty values of $\Delta p > 0$ because one specified target molecule resists the deformation that a competitor has to apply to bind to the probe itself. For the same parameter set as introduced in Table 4.8, Table 4.9 and Fig. 4.27 show the binding affinities K^{HAT} and K^{LAT} calculated with Eq. (4.23) as a function of Δp. The numerical assessment confirms that under identical conditions, the LAT looses the major part of its binding affinity while the effect on the HAT is also there, but negligible. This is because, the LAT, with its non-matching base pair, generates the most part of its binding affinity through a large number of short helix configurations with low cooperativity (see Fig. 4.26 B). However, the penalty in particular increases the required size of even these helix configurations and this causes the loss of K^{LAT} by three orders of magnitude. On the other hand, for the HAT more of the binding affinity stems from almost completely developed helix configurations (see Fig. 4.26 A). These configurations, however, are still able to overcome the barrier of Δp so that the change of K^{HAT} is negligible. Based

Table 4.9.: Theoretical binding affinities K^{HAT}, K^{LAT} in competition as a function of the energy penalty Δp, expressed in multiple integers of J. The binding constants are calculated with Eq. (4.23) for a temperature of 44 °C using the same parameter set $(j, J, \Delta h)$ as presented in Table 4.8.

Δp	K^{HAT}	K^{LAT}	K^{HAT}/K^{LAT}
0	$3.20 \cdot 10^7$	$1.88 \cdot 10^6$	17.02
1J	$9.01 \cdot 10^6$	$2.43 \cdot 10^4$	370.26
2J	$8.04 \cdot 10^6$	$6.81 \cdot 10^3$	1180.84
3J	$7.98 \cdot 10^6$	$6.10 \cdot 10^3$	1308.99
4J	$7.97 \cdot 10^6$	$6.06 \cdot 10^3$	1316.84
5J	$7.97 \cdot 10^6$	$6.05 \cdot 10^3$	1317.22
6J	$7.97 \cdot 10^6$	$6.05 \cdot 10^3$	1317.22

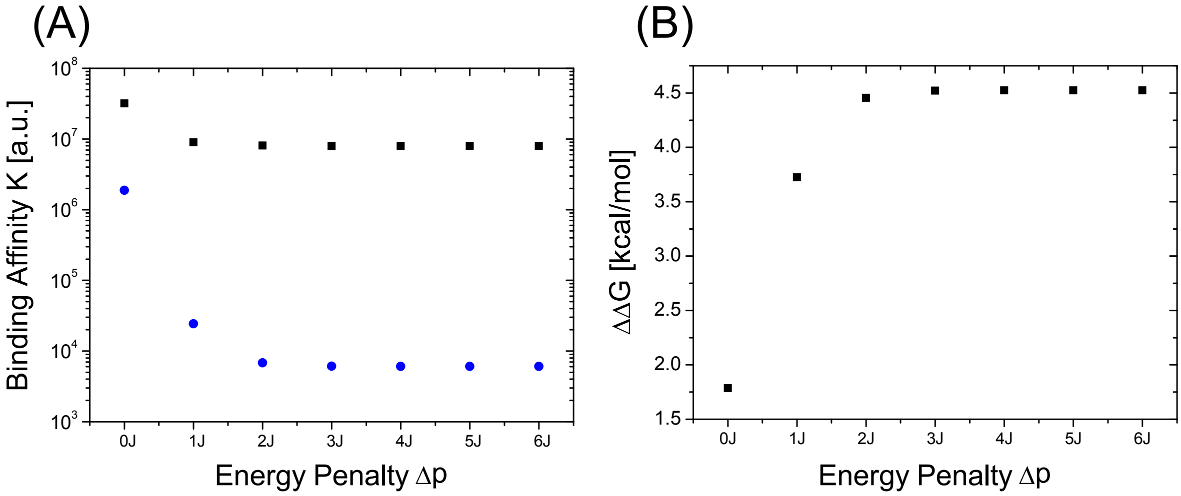


Figure 4.27.: Theoretical prediction of the high specificity. **(A)**, binding affinities K^{HAT} and K^{LAT} listed in Table 4.9 as a function of the energy penalty Δp, expressed in multiple integers of the next neighbor coupling energy of the helix configurations J at 44 °C. While K^{LAT} (blue) decreases by about three orders of magnitude, the effect on K^{HAT} (black) is also there but negligible. **(B)**, binding free energy difference $\Delta\Delta G = \Delta G^{HAT} - \Delta G^{LAT}$ as a function of Δp. ΔG^{HAT} and ΔG^{LAT} are determined with the values K^{HAT} and K^{LAT} from (A) using Eq. (4.27). The graph illustrates the asymmetric change of the binding free energy landscape in favor of the HAT.

on the experimental results, this scenario applies to the competitive systems of high specificity (Fig. 4.4 B).

4.4.7. Free energy of the total system

The results presented in Fig. 4.27 and Table 4.9 are now used to determine the overall Gibbs free energy G^{total} of a system consisting of N binding sites and two competing binding species HAT (high affinity target) and LAT (low affinity target). Each binding site can be empty, occupied by a HAT, occupied by a LAT, or occupied by both molecules HAT and LAT at the same time. Formation of HAT-HAT-probe and LAT-LAT-probe triplexes are not considered. This is because these configurations may well occur in situations without competition, and they do not play a role in achieving high specificity with the present formalism. The fraction of probes occupied by a HAT (LAT) is c^{HAT} (c^{LAT}). The fraction of probes occupied by a HAT and a LAT simultaneously is c_T. Following this convention, all binding sites are occupied as triplexes simultaneously by a HAT and a LAT, if $c^{HAT} = c^{LAT} = c_T = 1$. The effective Gibbs free energy G^{total} per site N is:

$$
\begin{aligned}
\frac{G^{total}}{N} &= H - R \cdot T \cdot S \\
&= c^{HAT} \cdot \left(\epsilon_0^{HAT} + \mu^{HAT}\right) - c_T \cdot \left(\epsilon_0^{HAT} + \mu^{HAT} - \epsilon_T^{HAT} - \mu^{HAT}\right) \\
&+ c^{LAT} \cdot \left(\epsilon_0^{LAT} + \mu^{LAT}\right) - c_T \cdot \left(\epsilon_0^{LAT} + \mu^{LAT} - \epsilon_T^{LAT} - \mu^{LAT}\right) - R \cdot T \cdot S \\
&= c^{HAT} \cdot \left(\epsilon_0^{HAT} + \mu^{HAT}\right) - c_T \cdot \tilde{\epsilon}^{HAT} \\
&+ c^{LAT} \cdot \left(\epsilon_0^{LAT} + \mu^{LAT}\right) - c_T \cdot \tilde{\epsilon}^{LAT} - R \cdot T \cdot S
\end{aligned}
\tag{4.32}
$$

$\epsilon_0^{HAT} = R \cdot T \cdot \ln\left(K^{HAT}\right)$ and $\epsilon_0^{LAT} = R \cdot T \cdot \ln\left(K^{LAT}\right)$ are the energies that are gained when a HAT or a LAT hybridizes with a probe molecule. $\mu^{HAT} = R \cdot T \cdot \ln\left([HAT]\right)$ and $\mu^{LAT} = R \cdot T \cdot \ln\left([LAT]\right)$ are the chemical potentials of HAT and LAT, respectively. $\epsilon_T^{HAT} = R \cdot T \cdot \ln\left(K_T^{HAT}\right)$ is the energy that is gained when a HAT hybridizes with a probe molecule that is already occupied by a LAT (triplex formation). The same applies for $\epsilon_T^{LAT} = R \cdot T \cdot \ln\left(K_T^{LAT}\right)$, when a LAT hybridizes with a probe molecule that is already occupied by a HAT. $\tilde{\epsilon}^{HAT}$ and $\tilde{\epsilon}^{LAT}$ are the energy corrections that reflect the interaction of HAT and LAT when both are bound to one and the same probe. The entropy S divided by the gas constant R is:

$$
S = S_0 - \ln\left[\binom{N}{N \cdot c^{HAT}} \cdot \binom{N}{N \cdot c^{LAT}}\right]
\tag{4.33}
$$

The constant value of S_0 in Eq. (4.33) is neglected in the following. The fraction c_T of probes occupied by a HAT and a LAT simultaneously is a function of the conditional probability κ_T^{HAT} (κ_T^{LAT}) that a given LAT (HAT) shares its probe molecule with a LAT

(HAT):

$$c_T = c^{HAT} \cdot \kappa_T^{LAT} = c^{LAT} \cdot \kappa_T^{HAT} \tag{4.34}$$

With this expression, Eq. (4.32) can be written as

$$\begin{aligned}\frac{G^{total}}{N} &= c^{HAT} \cdot \left(\epsilon_0^{HAT} + \mu^{HAT}\right) - c^{LAT} \cdot \kappa_T^{HAT} \cdot \tilde{\epsilon}^{HAT} \\ &+ c^{LAT} \cdot \left(\epsilon_0^{LAT} + \mu^{LAT}\right) \cdot \left(1 - \frac{\kappa_T^{HAT} \cdot \tilde{\epsilon}^{LAT}}{\epsilon_0^{LAT} + \mu^{LAT}}\right) - R \cdot T \cdot S\end{aligned} \tag{4.35}$$

If all available probes are occupied by a LAT, there is always a fraction that share the probe with a HAT. This is because the HAT can achieve binding states that the LAT cannot reach (see Fig. 4.26). This probability $\kappa_T^{HAT,min}$ is a property of the competing molecules, and for a given fraction c^{LAT} there must be the fraction

$$c^{HAT,min} = c^{LAT} \cdot \kappa_T^{HAT,min} \tag{4.36}$$

of HAT molecules on the LAT occupied probes. If all probe molecules are occupied by a HAT, all probe bound LAT molecules must share their probe with a HAT:

$$\kappa_T^{HAT}\left(c^{HAT} = 1\right) = 1 \tag{4.37}$$

κ_T^{HAT} is a linear function of c^{HAT}. For a given c^{LAT}, κ_T^{HAT} depends on c^{HAT} as

$$\kappa_T^{HAT} = \frac{c^{HAT,min}}{c^{LAT}} + \frac{c^{LAT} - c^{HAT,min}}{c^{LAT}(1 - c^{HAT,min})} \cdot \left(c^{HAT} - c^{HAT,min}\right) \tag{4.38}$$

In the following, the binding constants K^{HAT} and K^{LAT} are taken according to experimental values. They differ by about one order of magnitude, as observed for the targets PM and MM1 (compare Table 4.1). HAT and LAT are considered in concentrations of $[HAT]_0 = 5$ nM and $[LAT]_0 = 1$ µM, respectively. These concentrations are used in most of the competitive hybridization experiments (compare for instance section 4.1.2). For a temperature of 44 °C, the energies and chemical potentials in Eq. (4.35) are determined as

$$\epsilon_0^{HAT} = R \cdot T \cdot \ln\left(K^{HAT}\right) = 12.99 \text{ kcal/mol} \tag{4.39}$$

$$\epsilon_0^{LAT} = R \cdot T \cdot \ln\left(K^{LAT}\right) = 11.28 \text{ kcal/mol} \tag{4.40}$$

$$\mu^{HAT} = R \cdot T \cdot \ln\left([HAT]_0\right) = -12.04 \text{ kcal/mol} \tag{4.41}$$

$$\mu^{LAT} = R \cdot T \cdot \ln\left([LAT]_0\right) = -8.7 \text{ kcal/mol} \tag{4.42}$$

The results of the numerical assessment listed in Table 4.9 suggest that for an energy penalty of $\Delta p = 4J$ the binding affinities K^{HAT} and K^{LAT} decrease by factors of 4 and 310, respectively. These reduced binding affinities can be translated to average triplex binding energies ϵ_T^{HAT} and ϵ_T^{LAT} that apply when a particular target has to bind to a probe molecule that is already occupied by the competitor:

$$\epsilon_T^{HAT} = R \cdot T \cdot \ln\left(K^{HAT}/4\right) = 12.11 \text{ kcal/mol} \tag{4.43}$$

$$\epsilon_T^{LAT} = R \cdot T \cdot \ln\left(K^{LAT}/310\right) = 7.67 \text{ kcal/mol} \tag{4.44}$$

With this, one obtains

$$\tilde{\epsilon}^{HAT} = \epsilon_0^{HAT} - \epsilon_T^{HAT} = 0.88 \text{ kcal/mol} \tag{4.45}$$

$$\tilde{\epsilon}^{LAT} = \epsilon_0^{LAT} - \epsilon_T^{LAT} = 3.61 \text{ kcal/mol} \tag{4.46}$$

With these parameters, Fig. 4.28 visualizes the effective Gibbs free energy landscape determined by Eq. (4.35) as a function of c^{HAT} and c^{LAT}, as well as different values of κ_T^{HAT}. For low triplex probabilities $\kappa_T^{HAT} \leq 0.6$, the Gibbs free energy reaches its maximum for high values of c^{LAT} but low values of c^{HAT}. The system tends to a configuration where it is almost exclusively the LAT that binds to the probes. This case reflects the experimental finding of standard specificity, where the much higher concentrated LAT prevails over the HAT. However, triplex probabilities of $\kappa_T^{HAT} > 0.6$ lead to a sharp transition of the binding free energy maximum to high values of c^{HAT} but very low values of c^{LAT}. This result reflects the competitive hybridization systems of high specificity. Fig. 4.28 further shows that the probe molecules are mostly occupied by only one target species, HAT or LAT (consider the sum of c^{HAT} and c^{LAT} in Fig. 4.28). One understands that due to the triplex interaction that depends on a product of concentrations, there is a sharp, asymmetric modification of the binding free energy landscape in favor of the HAT. The formalism resembles the Landau description of phase transition. The case where both molecular species are bound to the same probe as a triplex does not occur for a large fraction of the probes. This is because the triplex does not constitute a minimum of the energy landscape. Although the existence of triplex states is required for the observation of the high specificity, there are no experimental hints that these states are occupied in thermodynamic equilibrium. This is like in a phase separation, where the molecules separate to avoid the interaction that causes their separation.

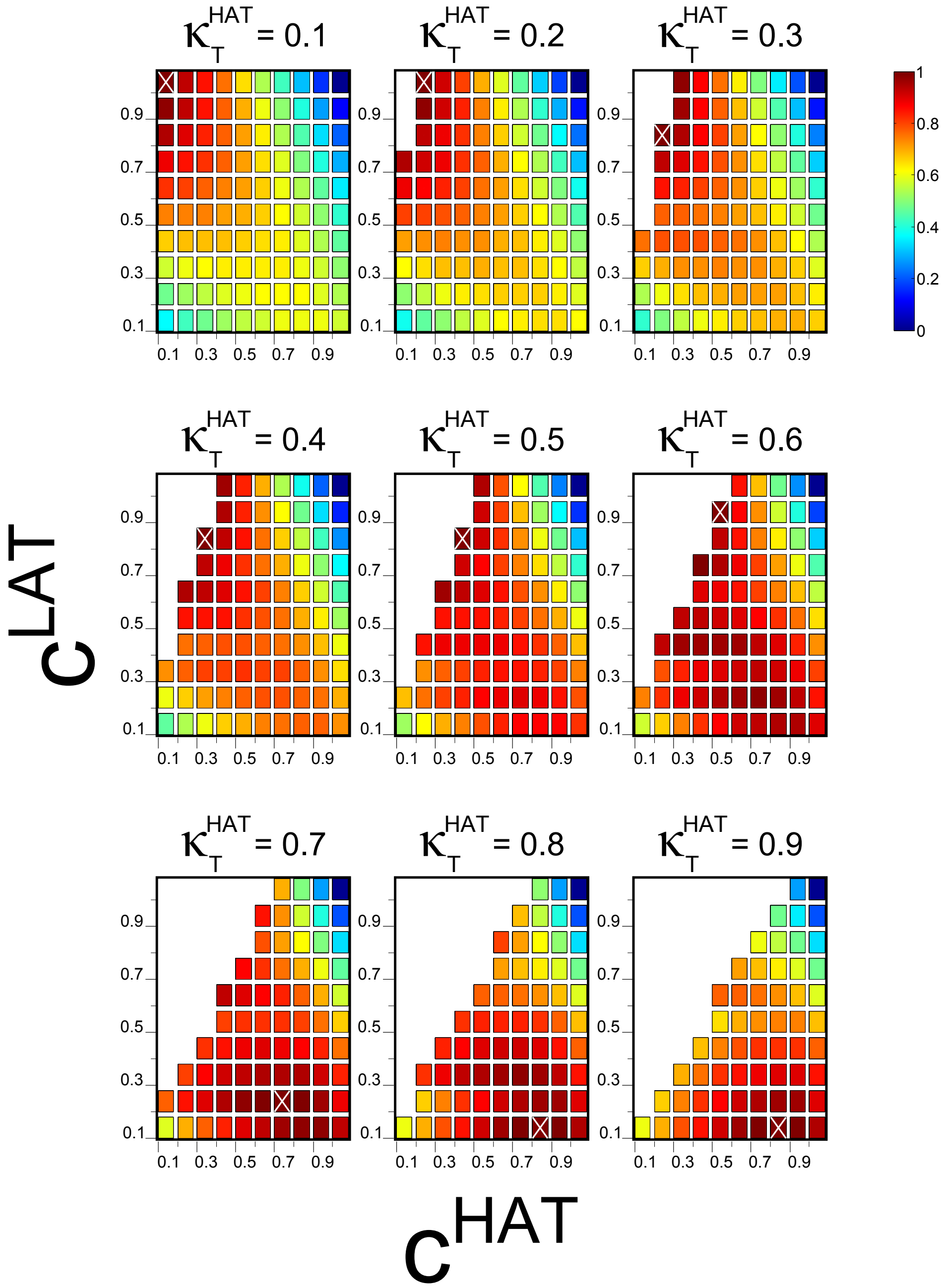


Figure 4.28.: Effective Gibbs free energy landscape of a competitive hybridization, following Eq. (4.35). All parameters are listed in Eqs. (4.42), (4.43), and (4.44). The graphs show the effective free energy G^{total}/N as a function of c^{HAT} and c^{LAT} and for different values of the conditional triplex probability κ_T^{HAT}. Maximum and minimum values of G^{total}/N are normalized to 1 and 0, respectively. In each sub-plot, the white "x" marks the maximum of the energy landscape. Triplex probabilities of $\kappa_T^{HAT} > 0.6$ lead to a sharp shift of the binding free energy maximum to high values of c^{HAT} but low values of c^{LAT}.

4.4.8. Specificity increase in theory and experiment

In the following, the theoretically predicted specificity increase of the HAT in competition with the LAT is compared with the experimental data. Therefore, the surface occupancy Θ^{HAT} is evaluated as a function of the concentration of the LAT, following

$$\Theta^{HAT} = \frac{[D^{HAT}]_{eq}}{[P]} = \frac{K^{HAT} \cdot [HAT]_{eq}}{K^{HAT} \cdot [HAT]_{eq} + K^{LAT} \cdot [LAT]_{eq} + 1} \tag{4.47}$$

(compare Eq. 4.2). For the binding affinities K^{HAT} and K^{LAT}, the theoretically predicted values listed in Table 4.9 are used. Two cases are considered: The case of the single hybridization ($\Delta p = 0J$), where the binding affinities K^{HAT} and K^{LAT} differ by one order of magnitude as observed experimentally for PM and MM1, and the competitive case $\Delta p = 4J$ where K^{HAT} and K^{LAT} differ by three orders of magnitude. The latter case reflects the assumption that the additional triplex state modifies the binding free energy landscape in such a way that the system can basically be described with the triplex binding energies ϵ_T^{HAT} and ϵ_T^{LAT}, while the individual energies ϵ_0^{HAT} und ϵ_0^{LAT} can be neglected (compare section 4.4.7). The theoretical data of Θ^{HAT} as a function of $[LAT]$ is compared with the experimental data found for the concentration limit of the highly specific system “PM vs. MM1” (section 4.3.1). The result is presented in Fig. 4.29. It reveals that the high specificity of the PM can be predicted from the theoretical binding affinities reached for an energy penalty of $\Delta p = 4J$ within the technical concentration limits of the LAT. The asymmetric modification of the binding free energy landscape leads to a shift of the transition of Θ^{HAT} to increased values of $[LAT]$.

4.5. Conclusion and discussion

The results of the numerical assessment show that in general, the binding affinities K^{HAT} and K^{LAT} can change in a competitive environment compared to the competition free case. In the model presented in the previous sections, this is achieved with an asymmetric interaction in the form of short-lived triplex structures consisting of HAT, LAT, and the probe molecule. The binding free energy landscape changes by several orders of magnitude in favor of the HAT but without the observation that the triplex states are indeed occupied in a large number. It is important to mention that the reduction of K^{LAT} during competition with the HAT does not contradict thermodynamics. Binding in competition is thermodynamically a different situation that is not necessarily

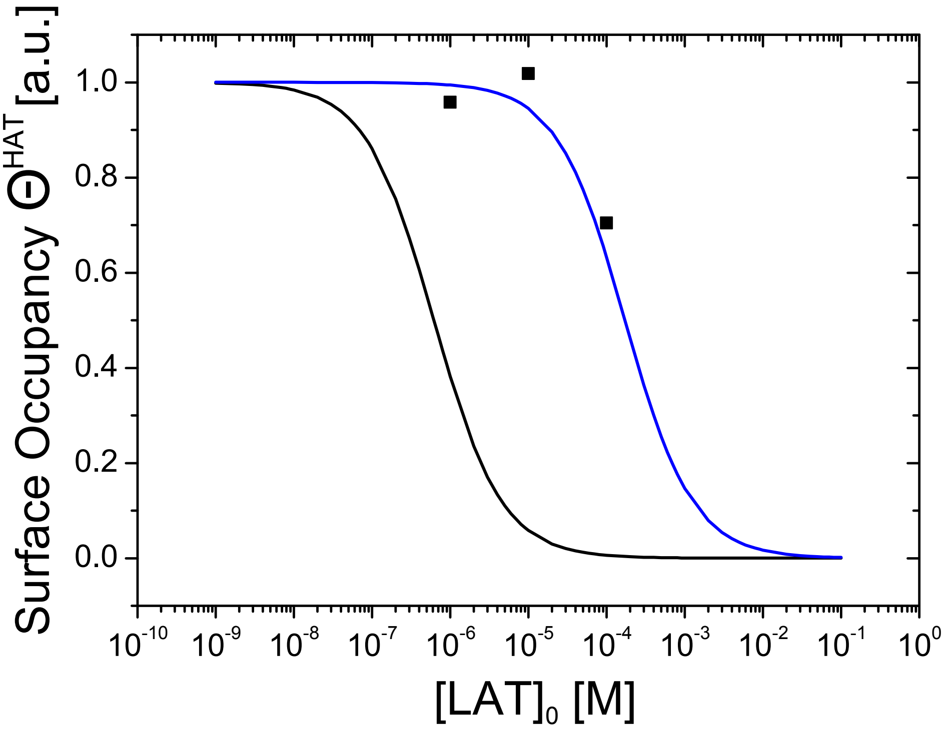


Figure 4.29.: Evaluation of the numerical assessment. The graph shows the surface occupancy Θ^{HAT} as a function of the LAT concentration $[LAT]$ following Eq. 4.47. All data is normalized to the single hybridization ($[LAT] = 0$). The solid lines are the predictions of the numerical assessment for Θ^{HAT}, thereby using two different values for the predicted binding affinities K^{HAT} and K^{LAT} (compare Table 4.9): single hybridization (black line, $\Delta p = 0$, $K^{HAT} = 3.20 \cdot 10^7$, $K^{LAT} = 1.88 \cdot 10^6$) and competitive hybridization (blue line, $\Delta p = 4J$, $K^{HAT} = 7.97 \cdot 10^6$, $K^{LAT} = 6.06 \cdot 10^3$). The filled symbols are the experimentally measured values for the highly specific competitive system "PM vs. MM1", using the data presented in section 4.3.1. Θ^{PM} is determined with Eq. (4.47) using the values of D^{PM} in Fig. 4.12. The increase of the hybridization specificity of the PM observed in the experiments can be qualitatively predicted with the theoretically determined binding affinities in competition for $\Delta p = 4J$.

described by the mean energies of individual binding. Rather, in competition, the different fluctuating binding states of HAT and LAT can combine in new ways and change the effective binding free energy landscape. At thermodynamic equilibrium, DNA strands form transient triplexes as a result of thermal fluctuations, regardless whether in competition or not. These exchange dynamics are particularly pronounced at the melting transition. Large fluctuations do not necessarily affect the macroscopic binding constant K, which reflect a mean value. However, if the binding conformations of the competitors are antagonistic, as expressed in the model with the energy penalty Δp, and the triplex population favors one competitor, the triplex fluctuations will modify the effective binding free energy landscape.

Two remarks may apply on the approximation of the intermolecular interactions with the energy penalty Δp:

1. Δp is taken as a constant but is likely to be small for a very low and very large number of base pairs that form a helix configuration with a maximum between the two cases. This, however, will only enhance the tendency to favor the HAT.
2. There must also be an energy penalty that applies to pre-melted configurations as long as the molecules form a triplex. This penalty will work in the same direction as Δp and increase the likelihood of more collective binding states, also favoring the HAT.

Put together, an improvement of the approximation for the penalty Δp is likely to increase the specificity of the HAT. Extensive molecular dynamic simulations may gain further insight into the complex situation of highly fluctuating DNA strands and help to develop a more detailed model.

5. Influence of DNA methylation on the DNA duplex stability

The following chapter deals with the influence of methylated and hydroxymethylated cytosine bases within short oligonucleotide strands on the specificity of DNA hybridization. The main focus of the investigation lies on the modification of the binding affinity K of a specified DNA target of 16 bases in length due to the methylation or hydroxymethylation of two cytosine bases. Furthermore, competitive hybridization experiments between methylated and non-methylated target molecules are the subject of interest. The experimental methods used for this study are the same as the ones introduced in the previous chapter 4. One has to mention that the experiments constitute preliminary studies, since the main focus of this work lies on the ultrahigh DNA hybridization specificity due to molecular competition (chapter 4) as well as the *in vitro* study of the transcriptional regulation of epigenetic switches (chapter 6).

5.1. Hybridization affinity of methylated DNA targets

In order to analyze the influence of methylated and hydroxymethylated bases on the binding affinity K of a particular target molecule, Langmuir isotherms are determined for the molecules PM^{m}, $\mathrm{MM1}^{\mathrm{m}}$, and $\mathrm{MM1}^{\mathrm{hm}}$ (see appendix Table A.1). Compared to their non-methylated analogons PM and MM1, these targets contain two bases of methylcytosine or hydroxymethylcytosine within their sequence. A hydroxymethylated version of the PM is not investigated in the context of this work. The experimental approach for determining the binding constant K of these targets exactly follows the procedure explained in chapter 4.1.1. The probe molecules remain non-methylated. In order to detect a possible difference in binding affinity between the methylated single strands and their non-methylated analogons, the experiments are performed at temperatures close to the melting point T_m. T_m is predicted with the dinamelt web server for nucleic acid melting predictions [29] and also by measuring melting profiles (section 5.2). Close to the melting point, the double strand is additionally destabilized by the temperature fluctuations and a possible influence of DNA methylation on the duplex stability is easier to detect. In the case of MM1, $\mathrm{MM1^{m}}$, and $\mathrm{MM1^{hm}}$ the usually employed hybridization temperature of 44 °C is adequate. However, for the targets PM and $\mathrm{PM^{m}}$, the temperature is increased to 60 °C. Thus, the binding affinity of the non-methylated PM (see Fig. 4.1) has to be determined again for a temperature of 60 °C. Since the calibration of the hybridization signal D into units equivalent to the duplex concentration in the measurement chamber (see appendix section A.3) is only valid for a temperature of 44 °C, the data has to be fitted with the modified version of the extended Langmuir isotherm

$$D_{eq} = \frac{A}{2}\left([T]_0 + [P]_0 + K^{-1}\right) - \frac{A}{2}\sqrt{\left([T]_0 + [P]_0 + K^{-1}\right)^2 - 4[T]_0[P]_0} \qquad (5.1)$$

(compare Eq. (4.3)) in order to extract the binding affinity K of the investigated target molecule.

Fig. 5.1 shows the raw data and the extended Langmuir isotherms for the target molecules PM and $\mathrm{PM^{m}}$ at 60 °C as well as MM1, $\mathrm{MM1^{m}}$, and $\mathrm{MM1^{hm}}$ at 44 °C. The respective binding constants K are combined in Table 5.1. The experiment shows that methylating only two of the sixteen bases of the target molecules has significant influence on the duplex stability although the probe molecule remains non-methylated. In all cases, the binding affinity of the methylated single strands increases by about one order of magnitude compared to the non-methylated case. Hydroxymethylation of the same bases further stabilizes the duplex.

5.2. Methylation dependent modification of the DNA melting profile

As presented in the previous section, the methylation of only a few bases of a target molecule has significant influence on its binding affinity K. In order to gain further insight into this stabilizing effect of DNA methylation, denaturation curves are measured for the target molecules $\mathrm{MM1^{m}}$ and $\mathrm{MM1^{hm}}$. The data is compared with the experimental results of the non-methylated analogons. The denaturation curve of the target $\mathrm{PM^{m}}$ is not measured, since the temperature that is required to completely melt the double strand consisting of $\mathrm{PM^{m}}$ and probe exceeds the technical limit of the experimental

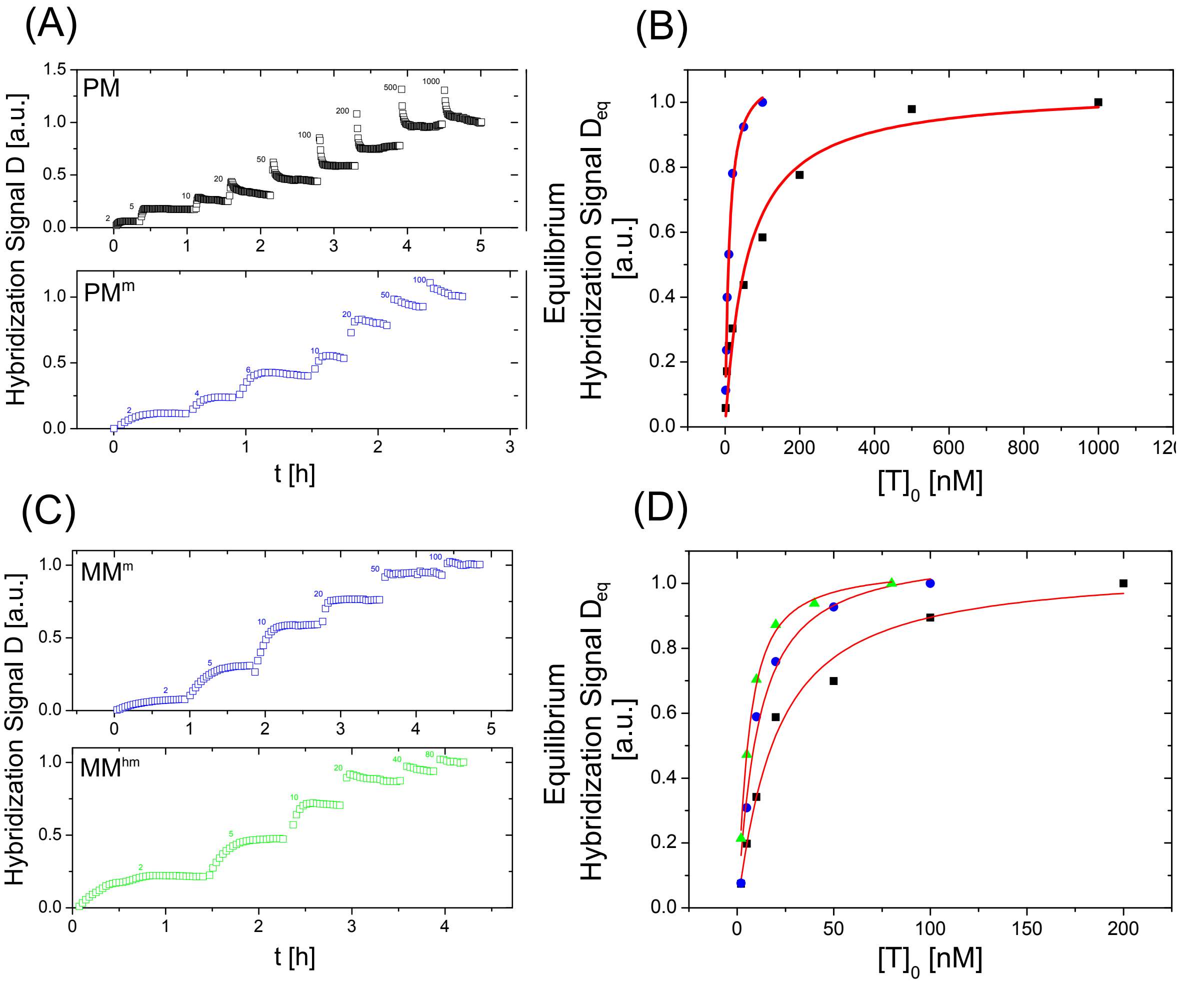


Figure 5.1.: Determination of binding affinities for the target molecules PM and PM^m at 60 °C (A, B) and for the target molecules MM1, MM1^m, and MM1^{hm} at 44 °C (C, D). **(A, C)**, raw data. The graph shows the hybridization signal D as a function of time. Following an increase of the target concentration $[T]_0$ (respective values in nM), D is measured until it reaches its equilibrium value D_{eq} and remains constant in time. Procedure is repeated until all probe molecules are occupied. **(B), (D)**, extended Langmuir isotherms. The respective values for D_{eq} from (A) and (B) are plotted against $[T]_0$. The values for the MM1 (black dots) stem from the measurement presented in Fig. 4.1. Fitting of Eq. (5.1) to this data (red line) yields the binding constants K of the investigated target molecule in units 1/M. The values are listed in Table 5.1.

setup. The approach for measuring denaturation curves is explained in detail in chapter 4.3.4. The contribution of the temperature dependent efficiency of the fluorescent dye to the hybridization signal is removed as explained in chapter 3.1.5 and A.7. The surface

Table 5.1.: Sequences and binding affinities of target molecules used for the investigation of the influence of DNA methylation on the duplex stability. Non-complementary bases are shown underlined and in red. The targets PM^m and MM1^m contain two bases of methylcytosine (C_{CH_3}) and the target MM1^{hm} contains two bases of hydroxymethylcytosine (C_{OCH_3}). The values for K at 60 °C and 44 °C are determined with the extended Langmuir isotherm (compare Fig. 5.1).

Target	Sequence (5′ → 3′)	K [10^7 1/M]
		60°C:
PM	AAG-GAT-CAG-ATC-GTA-A	1.7 ± 0.5
PM^m	AAG-GAT-C_{CH_3}AG-ATC_{CH_3}-GTA-A	11 ± 0.5
		44°C:
MM1	AAG-GAT-CA$\underline{C}$-ATC-GTA-A	6 ± 1
MM1^m	AAG-GAT-C_{CH_3}A$\underline{C}$-ATC_{CH_3}-GTA-A	11 ± 0.5
MM1^{hm}	AAG-GAT-C_{OCH_3}A$\underline{C}$-ATC_{OCH_3}-GTA-A	24 ± 3

occupancy Θ is linked to the binding constant K and the values of entropy and enthalpy change by combining the Eqs. (5.2) and (5.3):

$$\Theta = \frac{K \cdot [T]_0}{K \cdot [T]_0 + 1} \tag{5.2}$$

$$K(T) \propto e^{\frac{\Delta G}{RT}} = \exp\left(\frac{\Delta H}{R} \cdot \frac{1}{T} - \frac{\Delta S}{R}\right) \tag{5.3}$$

After measuring the melting curve, the surface occupancy Θ is converted into the binding affinity K following Eq. (5.2). Plotting the logarithm of K over the inverse temperature $1/T$ (Arrhenius plot, see Eq. (5.3)) and linear fitting of these data then reveals the thermodynamic parameters ΔS and ΔH.

Fig. 4.17 shows the melting curves of MM1^m and MM1^{hm}. The extracted thermodynamic parameters of melting temperature T_m, entropy change ΔS, and enthalpy change ΔH for the two molecular species are combined in Table 5.2. The experiment confirms the stabilizing effect of methylated bases incorporated into the sequence. Compared to the non-methylated target MM1, the melting temperature of the methylated strands is increased by 7 °C (MM1^m) and 9 °C (MM1^{hm}), respectively. This indicates an increased duplex stability of the methylated targets MM1^m and MM1^{hm} of comparable magnitude. In the case of MM1^m this result is also confirmed by a particularly pronounced change in enthalpy and entropy compared to the non-methylated MM1. However, for the MM1^{hm}, the values of ΔS and ΔH make no sense. Due to the comparable melting temperatures of

Table 5.2.: Values of enthalpy change ΔH, entropy change ΔS, and melting temperature T_m for target molecules MM1, MM1$^{\mathrm{m}}$, and MM1$^{\mathrm{hm}}$. Data is obtained from the denaturation curves presented in Fig. 5.2 (MM1$^{\mathrm{m}}$, MM1$^{\mathrm{hm}}$) and Fig. 4.17 (MM1).

Target	ΔH $[\mathrm{kcal} \cdot \mathrm{mol}^{-1}]$	ΔS $[\mathrm{kcal} \cdot \mathrm{mol}^{-1}] \cdot \mathrm{K}^{-1}$	$\mathrm{T_m}$ [°C]
MM1	73.48 ± 1	197.4	53 ± 1
MM1$^{\mathrm{m}}$	200.44 ± 9	574.95 ± 23	60 ± 1
MM1$^{\mathrm{hm}}$	83.48 ± 3	220.73 ± 9	62 ± 1

MM1$^{\mathrm{m}}$ and MM1$^{\mathrm{hm}}$ one would also expect comparable changes in entropy and enthalpy for the two molecular species. However, the values ΔH and ΔS of the MM1$^{\mathrm{hm}}$ do not differ significantly compared to the values extracted for the non-methylated MM1. This result is in conflict with the conclusions of the observed melting point shift (Fig. 5.2). An explanation for this seemingly paradoxical result could be the shape of the denaturation curve of the MM1$^{\mathrm{hm}}$. Compared to PM and MM1 (Fig.4.17) as well as MM1$^{\mathrm{m}}$, the melting transition of the MM1$^{\mathrm{hm}}$ (Fig4.17) is less sharp, indicating the existence of additional binding conformations within the pre-melted DNA duplex. This conformation invalidates the assumptions of a two-state description like the Langmuir model [51] and has to be a direct consequence of the hydroxymethylated DNA bases. Thus, the extraction of the thermodynamic parameters ΔS and ΔH with the combined Eqs. (5.2) and (5.3) seems to be unreasonable for the target MM1$^{\mathrm{hm}}$, since this assumes the validity of the two-state description. On the other hand, the observation of an increased melting temperature of the MM1$^{\mathrm{hm}}$ compared to the MM1 is an experimental fact that does not need any interpretation by a model.

5.3. Competition between methylated and non-methylated targets

As shown in the previous sections, methylation of some bases of a DNA target causes a significant increase in the binding affinity and therefore enhances the stability of the formed double strand. Considering the results of competitive DNA hybridization presented in the previous chapter 4, it is therefore interesting to investigate if DNA methylation also plays a role for the occurrence of the high specificity. Just like in the case of LNA (compare chapter 4.3.10), one could imagine that methylated DNA bases do also limit the entropic degrees of freedom of a DNA duplex and thus are essen-

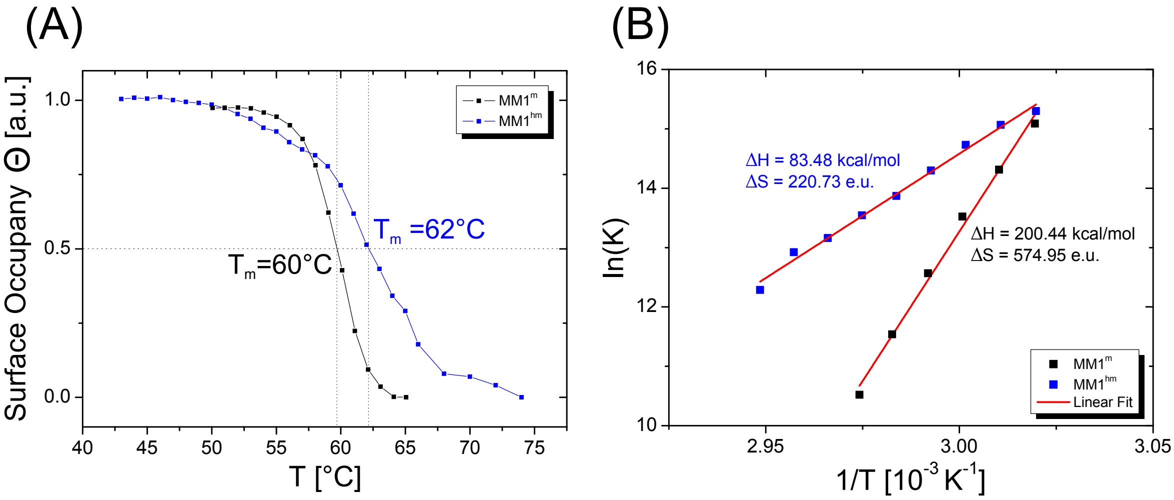


Figure 5.2.: Denaturation experiment for MM1^{m} and MM1^{hm}. **(A)**, melting curves. The graph shows the fraction of occupied probes Θ as a function of temperature T for MM1^{m} (black) and MM1^{hm} (blue). The target concentration is 1 µM in both cases. The melting temperature T_m is defined as the temperature where the surface occupancy equals 50% (intersection of the dotted lines). **(B)**, Arrhenius plot of the melting transition of MM1^{m} (black) and MM1^{hm} (blue) after converting the surface occupancy Θ (see A) into the binding constant K (Eq. (5.2)). The graph shows the logarithm of the binding constant K as a function of the inverse temperature $1/T$. The red lines are linear fits to the experimental data, thus revealing the values of enthalpy ΔH and entropy change ΔS (e.u. = "entropic units" = kcal/(mol · K)) of MM1^{m} and MM1^{hm}, respectively.

tial for the occurrence of standard and high specificity (compare chapter 4.4). Thus, in the following section, competitive hybridization experiments between methylated and non-methylated targets are performed. The experimental approach exactly follows the procedure introduced in chapter 4.1.2. Two systems are investigated: First, the competition between PM and MM1^{m}, in order to analyze if the PM hybridization is still highly specific even if the MM1 is methylated, and second, "PM^{m} vs. PM", in order to analyze if the methylated version of the PM shows a high specificity during competition with its non-methylated analogon. Both experiments are performed at a temperature of 44 °C. Thus, only the results of "PM vs. MM1^{m}" can be compared with the predictions of Eq. (2.21), since the binding affinity $K^{PM^{m}}$ is solely determined for a temperature of 60 °C (compare Fig. 5.1)

Fig. 5.3 shows the results of the competitive hybridization experiment between PM and MM1^{m} as well as "PM^{m} vs. PM". The experiment shows that the hybridization

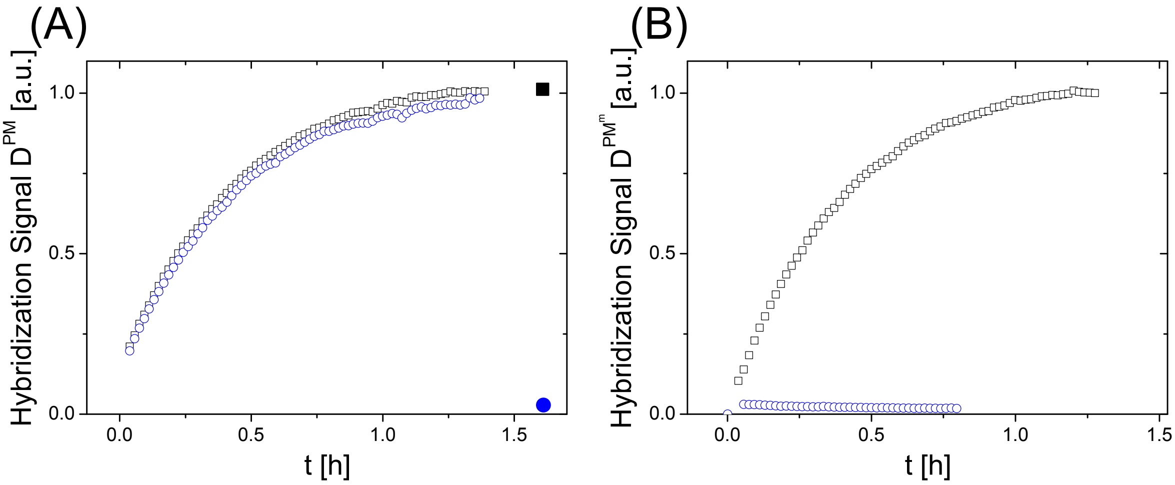


Figure 5.3.: Competitive hybridization between methylated and non-methylated target molecules. **(A)**, competition between PM and MM1^m at 44 °C. The graph shows the hybridization signal of the PM as a function of time in the case of a single hybridization (no competition, black) and in the presence of 1 µM of MM1^m. The target concentration of the PM is 5 nM in both measurements. D^{PM} is normalized to the final value of the single hybridization. The filled symbols show the equilibrium duplex concentration, D_{eq}^{PM}, predicted by Eq. (2.21) thereby using the respective target concentrations and binding affinities (compare Tables 4.1 and 5.1). The result is the same as observed for the competitive system "PM vs. MM1" (Fig. 4.2 A). In spite of the increased binding affinity of the MM1^m, a 200-fold excess concentration of MM1^m targets over PM targets does not lead to a reduced PM hybridization signal. The system shows high specificity (compare chapter 4.1.3). **(B)**, same experiment as shown in (A) but for competition between PM^m and PM. The presence of 1 µM PM completely releases the PM^m from the probe molecules. The predictions of Eq. (2.21) are not shown since K^{PM^m} is not determined for a temperature of 44 °C.

of the PM remains highly specific, no matter if the MM1 possesses methylated bases or not (compare Fig. 4.2). In both situations the predictions of Eq. (2.21) for the equilibrium hybridization signals D_{eq}^{PM} are wrong. In spite of the fact that the binding affinity K^{PM^m} is not determined for a temperature of 44 °C, one can conclude for the competitive system "PM^m vs. PM" that the hybridization of the PM^m shows standard specificity. A 200-fold excess concentration of the non-methylated version of the PM completely releases the methylated analogons from the probe molecules. Put together, DNA methylation does not directly play an essential role for the occurrence of standard or highly specific DNA hybridization (see chapter 4 for a definition of high and standard specificity).

5.4. Conclusion and discussion

In this chapter it is shown that methylation of only a few bases of a DNA oligonucleotide has significant influence on the DNA duplex stability. By methylating two bases of a single-stranded DNA molecule of 16 bases in length, the binding affinity can be increased by about one order of magnitude, no matter if the molecule possesses a non-matching base (MM1^{m}) or not (PM^{m}). The hydroxymethylation of the same bases leads to a comparable increase in binding affinity. The increased duplex stability can also be shown by measuring the melting curves of the methylated DNA molecules. The melting point T_m is shifted to higher temperatures (7 °C for MM1^{m}, 9 °C for MM1^{hm}), which also leads to a reduced number of entropic degrees of freedom in the case of the MM1^{m} compared to its non-methylated analogon. The melting profile of the MM1^{hm} indicates the existence of additional binding conformations that invalidate the assumption of a simple two-state description. Competitive hybridization experiments between methylated and non-methylated target molecules do not give further insight into the effect of highly specific DNA hybridization investigated in detail in the previous chapter 4. The competitive system“PM vs. MM1^{m}” shows the same high specificity as observed for the non-methylated case “PM vs. MM1”, while the system “PM^{m} vs. PM” shows standard specificity.

Like mentioned before, the hybridization experiments shown in the present chapter constitute preliminary studies. The main focus of this work lies on the study of the highly specific DNA hybridization (chapter 4) and the epigenetic regulation of *in vitro* gene switches (chapter 6). Consequently, the results presented in this chapter should be considered as a starting point for further investigations about how DNA methylation alters the properties of single-stranded DNA hybridization.

6. In vitro study of the transcriptional regulation of epigenetic switches

In the present chapter, the experimental results of the *in vitro* study of the epigenetic regulation of protein expression are presented and discussed. In the context of this work, two transcriptional regulation systems are investigated. Section 6.1 deals with the study of the pap regulator operons and the DNA binding proteins Lrp and PapI (compare chapter 2.4.1). In section 6.2 the methylation binding domain of the human repressor MeCP2 is the subject of interest (compare chapter 2.4.2). Its function as an epigenetically regulated repressor is studied in detail. In section 6.3 the functionality of the two epigenetic regulation systems are merged in order to design a two state switch regulated by Dam and CpG methylation. A conclusion of all experimental findings is presented in section 6.4. The objective of the experiments presented in this chapter is to gain further insight into the metabolic functionality of the studied binding proteins and their role as transcriptional repressors as a function of the methylation state of the DNA operons. Therefore, the cell free extract is an ideal tool, since it enables the study of the transcriptional regulation within a totally endogenous system, including all the relevant molecular players, but under highly controlled experimental conditions. On the other hand, the experiments are also motivated by approaches of structural biology. Thus, the second objective of the studies is also the characterization of the cell free extract as a tool for designing synthetic gene circuits based on methylation binding proteins and methylated DNA.

6.1. Gene switching based on Dam methylation, Lrp and PapI

In the following, the *in vitro* expression results of the epigenetic gene switch based on Dam methylation, Lrp, and PapI are presented. The study involves cloning of the different binding operons for Lrp (compare chapter 6.1 and Table 2.1) as regulatory parts for the expression of the reporter gene coding for eGFP. The methylation dependent binding of Lrp and its co-repressor PapI is investigated by measuring the fluorescence intensity of the recombined reporter. The methylation dependent regulator proteins PapI and Lrp are also expressed in the crude extract from plasmids pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 and pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, respectively. The concentration of recombined proteins as a function of the respective plasmid concentration is investigated in [35] for the promoter OR2-OR1-Pr. Based on these results, the yield of expressed PapI and Lrp proteins in the cell free extract is about 5 μM for a plasmid concentration of 1 nM. Lrp and PapI are not known to interact with the consensus sequence of OR2-OR1-Pr and thus no feedback loops should occur. Dam methylation of all reporter plasmids or PCR products is performed as explained in chapter 3.2.2. Plasmids free of Dam methylation are produced by using Dam-/Dcm- competent cells for bacterial amplification.

6.1.1. The pap123 operon

The pap123 operon is designed to investigate the cooperative binding of PapI-Lrp to the proximal promoter consisting of Lrp binding sites 1, 2, and 3 (compare chapter 2.4.1). Fig. 6.1 schematically illustrates the pap123 operon regulating the expression of the reporter gene coding for eGFP. Binding site 2 contains the GATC binding site for Dam methylase and can thus be either methylated or non-methylated. Operon 2 is flanked by Lrp binding sites 1 and 3. According to literature [16–18, 22], PapI-Lrp can bind cooperatively to Lrp sites 1, 2, and 3. This means, that the methylation dependent binding of Lrp to binding site 2 should also mediate the binding to sites 1 and 3. Since site 3 is overlapping the -35 -10 region of the promoter of the pap123 operon, the expression of the reporter should be repressed.

eGFP expression as a function of the reporter plasmid concentration

In order to determine the promoter strength of the pap123 operon, batch mode expression experiments with increasing concentrations of the reporter plasmid pBEST-pap123-UTR1-eGFP-T500 are performed. This is necessary to determine the optimal plasmid concentration for subsequent expression experiments and also to study a possible difference in the expression efficiency of the extract for Dam methylated and non-methylated

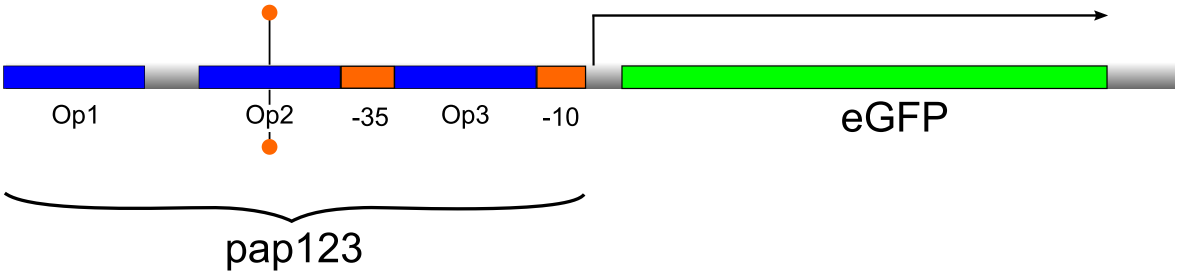


Figure 6.1.: Schematic illustration of the pap123 operon. The Lrp binding sites 1, 2, and 3 (compare Table 2.1) regulate the expression of the reporter gene coding for eGFP. The operon 2 contains a GATC site that can either be methylated or non-methylated (orange pin). The operon 3 is cloned in-between the -35 -10 recognition sequence for σ^{70}.

DNA templates.

Fig. 6.2 shows the result of a batch mode expression of pBEST-pap123-UTR1-eGFP-T500 as a function of the reporter plasmid concentration. The experiment reveals an efficient eGFP expression for methylated and non-methylated reporter plasmids, up to a concentration of 20 nM. In most of the subsequent experiments the reaction also contains template DNA for the expression of the regulator proteins Lrp and PapI. This expression also dissipates the resources of the *in vitro* expression machinery. Therefore, it is necessary to balance a high reporter protein yield, required for the quantitative analysis of the experimental results, with the possibility to add further plasmids for the expression of the molecular regulators. Based on the results presented in Fig. 6.2, the standard reporter plasmid concentration is adjusted to 5 nM for all subsequent experiments, if not stated otherwise.

eGFP expression as a function of the amino acid concentration

The efficiency of the cell free expression does highly depend on the concentration of amino acids added to the extract [35]. The optimal amino acid concentration to reach the best possible yield of the reporter protein also depends on the number and concentrations of other plasmids added to the extract (for instance the template DNA for the expression of the regulator proteins Lrp and PapI). Therefore, it is necessary to determine the optimal amino acid concentration in a batch mode reaction that expresses

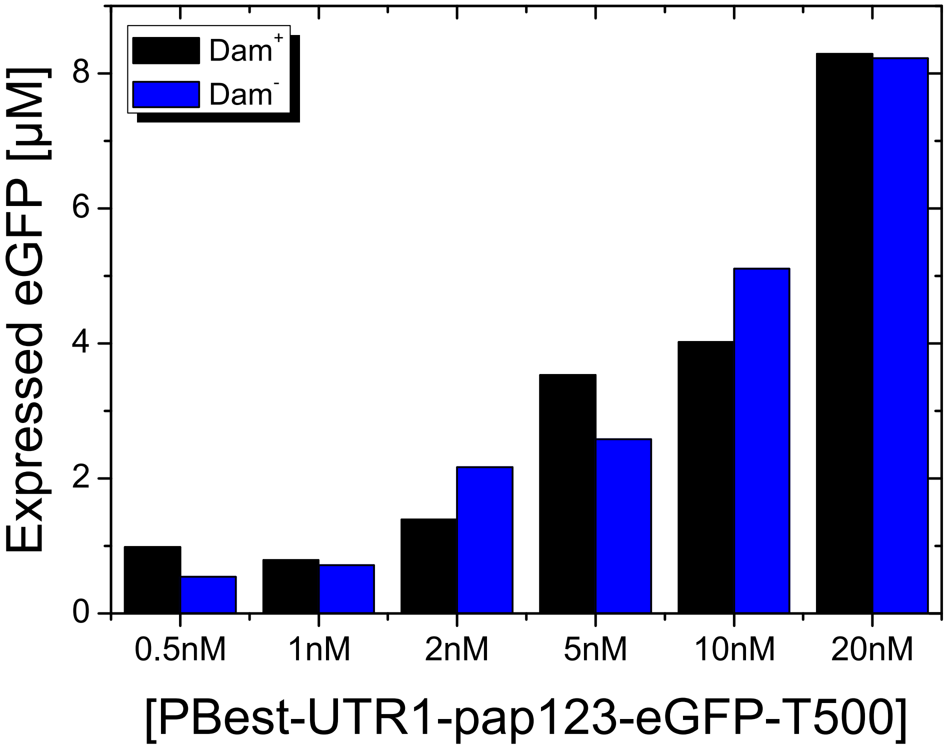


Figure 6.2.: Batch mode expression of eGFP regulated by the pap123 operon. The graph shows the concentration of the recombined eGFP as a function of the concentration of the reporter plasmid pBEST-pap123-UTR1-eGFP-T500 in the case of Dam methylated DNA (Dam$^+$, black) and non-methylated DNA (Dam$^-$, blue). The concentration of amino acids added to the reaction is 0.5 mM.

all three molecular players eGFP, Lrp, and PapI at the same time, at typically employed plasmid concentrations. This is done in the following experiments for the cases of both, a Dam methylated and a non-methylated version of the reporter pBEST-pap123-UTR1-eGFP-T500.

Fig. 6.3 shows the respective results. The experiment reveals an optimal amino acid concentration of 0.5 mM for expressing eGFP, Lrp and PapI within the same cell free reaction. In this case, the eGFP expression level of Dam methylated and non-methylated reporter plasmids is maximal. Thus, the amino acid concentration is adjusted to 0.5 mM in all subsequent experiments if not stated otherwise.

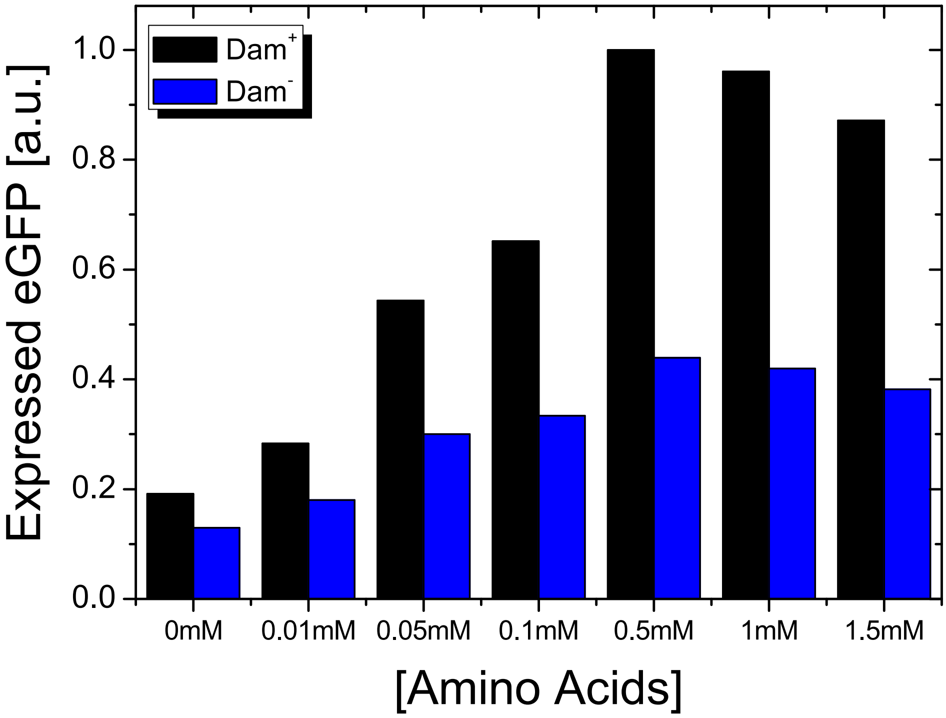


Figure 6.3.: Batch mode expression of pBEST-pap123-UTR1-eGFP-T500 as a function of the amino acid concentration in the cell free reaction. The graph shows the concentration of the recombined eGFP as a function of the amino acid concentration in the case of Dam methylated DNA (Dam$^+$, black) and non-methylated DNA (Dam$^-$, blue). Data is normalized to the Dam$^+$ expression level reached for an amino acid concentration of 0.5 mM. The values on the x-axis refer to the concentration of each of the 20 native amino acids. The concentration of the reporter plasmid is 5 nM. Additionally, the reaction contains 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500, expressing the regulator Lrp, as well as 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, expressing the co-regulator PapI.

Lrp mediated repression of pap123-eGFP

In order to analyze the Lrp mediated repression of the eGFP expression regulated by the pap123 operon, batch mode expression experiments with increasing concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are performed for Dam methylated and non-methylated reporter plasmids.

The results are presented in Fig. 6.4. Increasing the Lrp concentration in the cell free extract by adding higher concentrations of the Lrp plasmid leads to a decreased expression level of eGFP by about 80% in the case of Dam$^+$ reporter plasmids, and by about 85% in the case of Dam$^-$ reporter plasmids. Obviously, Dam methylation of Lrp

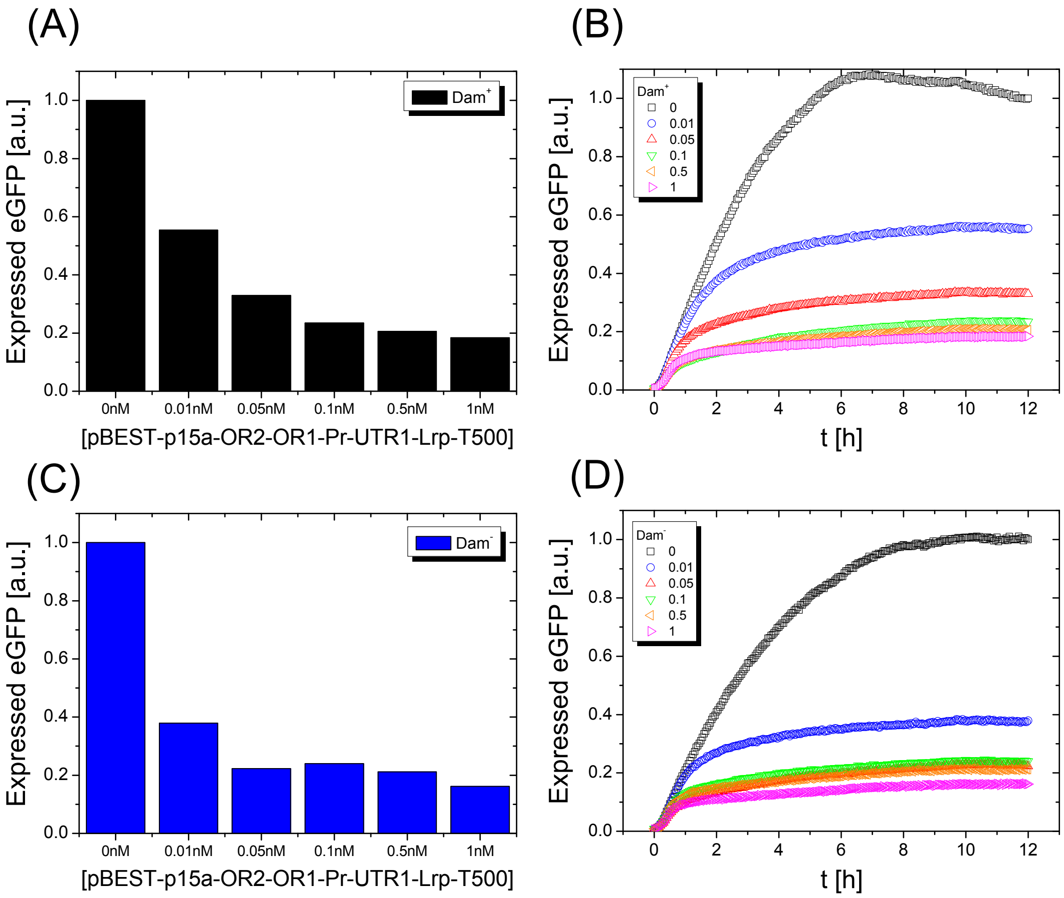


Figure 6.4.: Methylation response of the Lrp mediated repression of pBEST-pap123-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap123 operon in pBEST-pap123-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter plasmid is 5 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). The eGFP expression level decreases with increasing concentrations of Lrp by a factor of 80% (for 1 nM of the Lrp plasmid) in the presence of PapI. **(C), (D)**, same as (A) and (B) but for the non-methylated pap123 operon. In the presence of PapI, Lrp represses the system by a factor 85% (for 1 nM of the Lrp plasmid).

binding site 2 does not essentially influence the cooperative binding of PapI-Lrp to site 3 overlapping the -35 -10 region of the pap123 promoter. The methylation response of the system is very low, since only a difference of 5% between Dam^+ and Dam^- expression is observable for a pBEST-pa15a-OR2-OR1-Pr-UTR1-Lrp-T500 plasmid concentration of 1 nM. This result is overlapping with the 5% error bar, determined for the reproducibility of the expression level from batch to batch [39]. The kinetics of the eGFP expression, also presented in Fig. 6.4, highlights the role of Lrp as a repressor. After an initial phase where the eGFP expression is a linear function of time, the concentration of recombined Lrp within the cell free reaction is high enough to hinder further expression of the reporter protein. The duration of the linear eGFP production phase decreases with increasing concentrations of the Lrp template DNA. However, the experiment also reveals that the binding of Lrp to the pap123 operon does not completely stop the expression of the reporter. After the linear phase, the fluorescence signal of recombined eGFP further increases with a smaller slope, indicating a “leaky” repression of the system.

PapI response of the Lrp mediated repression of pap123-eGFP

In the following, the role of PapI as a co-repressor of Lrp is studied in the case of the pap123 operon. Therefore, the Lrp mediated repression experiments presented in Fig. 6.4 are repeated without the addition of 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500.

Fig. 6.5 shows the results of the Lrp mediated repression of pap123-eGFP in the absence of PapI. Comparing the results with the findings in Fig. 6.4 reveals a less important repression efficiency of Lrp in the absence of PapI (Dam$^+$/PapI$^+$: 80%, Dam$^-$/PapI$^+$: 85%; Dam+/PapI$^-$: 75%, Dam−/PapI$^-$: 75%). This is in agreement with the results of the literature that PapI slightly increases the affinity of Lrp for the cooperative binding to sites 1, 2, and 3 [17]. A methylation response of the Lrp binding to the pap123 operon is no longer observable in the absence of PapI, indicating that the methylation sensitivity of Lrp is mediated by the co-repressor PapI.

6.1.2. The pap2 operon

The pap2 operon (Fig. 6.6) is designed to study the PapI mediated binding of Lrp without any cooperative effects. Compared to the pap123 operon, the flanking Lrp

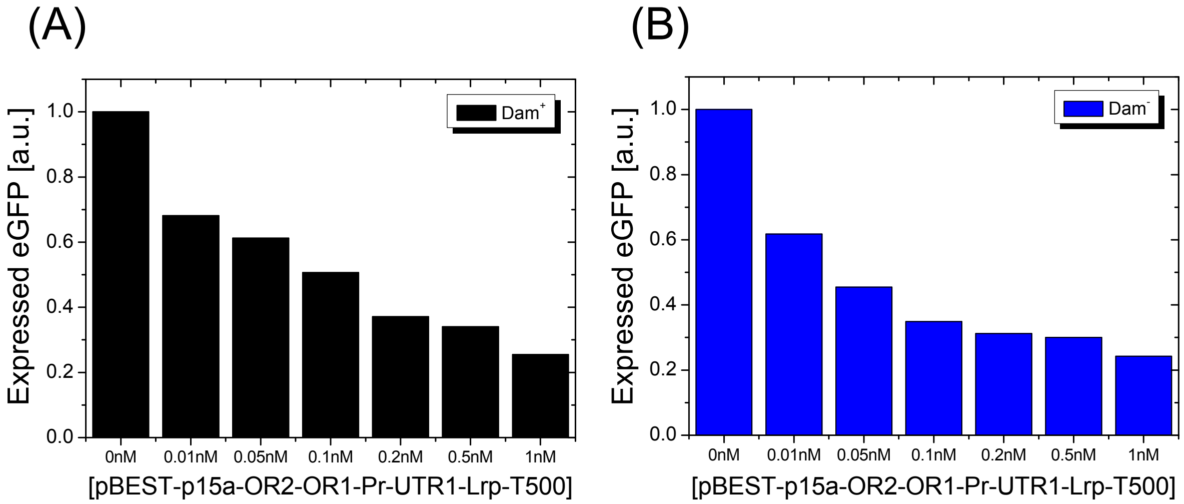


Figure 6.5.: Methylation response of the Lrp mediated repression of pBEST-pap123-UTR1-eGFP-T500 without expressing the co-repressor PapI. **(A)**, Dam methylated reporter plasmids. The graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of reporter plasmid and amino acids is 5 nM and 0.5 mM, respectively. Data is normalized to the equilibrium value of the eGFP expression without the addition of the Lrp template DNA. **(B)**, same as (A) but for the case of non-methylated reporter plasmids.

binding sites 1 and 3 are removed in pap2 and only binding site 2 is cloned in-between the -35 -10 region of the promoter. Thus, it is possible to study the direct methylation response of PapI-Lrp binding to this operon. The binding of Lrp to site 2 highly depends on the presence of the co-repressor PapI [17]. The reporter plasmid pBEST-pap2-UTR1-eGFP-T500 is determined to be toxic during plasmid amplification when using the standard -10 and -35 consensus sequences for σ^{70} ATAAT and TTGACA, respectively. Thus, the recognition sites are changed to GTAAT and TGGACA, respectively, in order to decrease the affinity of the transcription factor. In this case, the reporter plasmid is stable during bacterial amplification.

Lrp mediated repression of pap2-eGFP

Fig. 6.7 shows the results of the Lrp mediated repression of pap2-eGFP for Dam$^+$ and Dam$^-$ reporter plasmids. Due to the reduced binding affinity of σ^{70} for the changed -35 and -10 consensus sequences of the pap2 promoter, the concentration of the plasmid

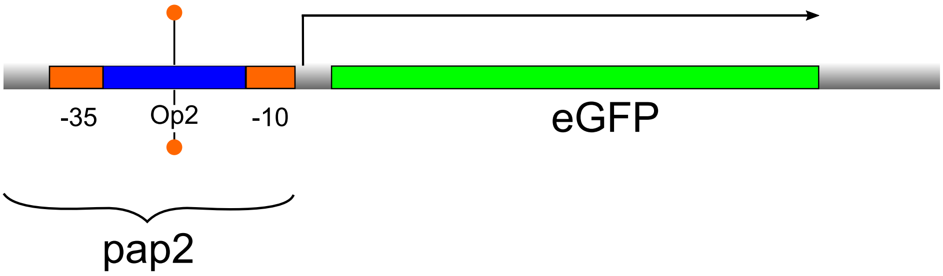


Figure 6.6.: Schematic illustration of the pap2 operon. The Lrp binding site 2 (compare Table 2.1) is cloned in-between the -35 -10 recognition sequence for σ^{70} and regulates the expression of the reporter gene coding for eGFP. The operon 2 contains a GATC site that can either be methylated or non-methylated (orange pin). Due to a toxicity of the plasmid pBEST-pap2-UTR1-eGFP-T500, the -10 consensus sequence of σ^{70} is changed from ATAAT to GTAAT, while the -35 consensus sequence is changed from TTGACA to TGGACA, decreasing the affinity of the sigma factor. However, the sequence of Lrp binding site 3 is not changed.

pBEST-pap2-UTR1-eGFP-T500 has to be increased from 5 nM to 10 nM in order to improve the signal to noise ratio. The experiment shows that increasing the Lrp plasmid concentration up to 1 nM leads to a reduced eGFP expression level. However, the repression efficiency of PapI-Lrp is markedly lower (Dam$^+$: 42%, Dam$^-$: 55%) compared to the case of the pap123 operon (Dam$^+$: 80%, Dam$^-$: 85%), where the regulator proteins were able to bind cooperatively to Lrp binding sites 1, 2, and 3. This supports the assumption of [22] and [52] that cooperative binding of PapI-Lrp to the proximal sites 1, 2, and 3 involves bending of the double-stranded DNA, which more efficiently blocks the annealing of the RNA polymerase than just the occupation of the binding site 2 without DNA bending. This is also supported by the results of the expression kinetics, since after the linear phase of eGFP expression, the fluorescence intensity further increases with higher slopes compared to the case of pap123-eGFP (compare Figs. 6.4 B and D with Figs. 6.7 B and D). This indicates a more “leaky” repression of the system regulated by the pap2 compared to the pap123 operon. Thus, one can conclude that cooperative binding of PapI-Lrp to sites 1, 2, and 3 is required for stable phase-OFF cells, not expressing the pap gene *in vivo*. Furthermore, the results show a methylation response of the repression efficiency of PapI-Lrp. This confirms that Dam methylation of Lrp binding site 2 hinders binding of PapI-Lrp to this site (compare Fig. 6.4).

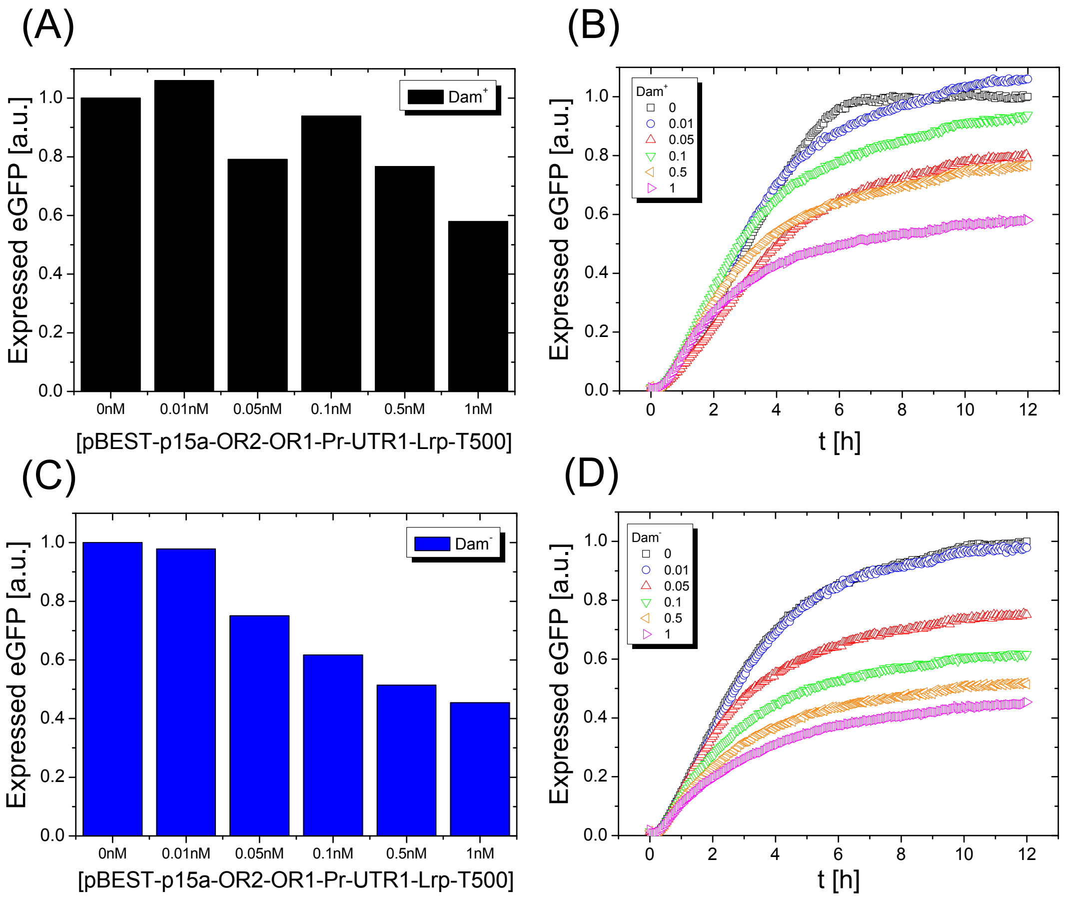


Figure 6.7.: Methylation response of the Lrp mediated repression of pBEST-pap2-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap2 operon in pBEST-pap2-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter plasmid is 10 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). The eGFP expression level decreases with increasing concentrations of Lrp by a factor of 42% (for 1 nM of the Lrp plasmid) in the presence of PapI. **(C), (D)**, same as (A) and (B) but for the non-methylated pap2 operon. In the presence of PapI, Lrp represses the system by a factor 55% (for 1 nM of the Lrp plasmid).

6.1.3. The pap222 operon

The pap222 operon is schematically illustrated in Fig. 6.8. It contains three Lrp binding sites 2 that regulate the expression of the reporter eGFP. The binding sites can either be Dam methylated or non-methylated. Primarily, the operon was designed to improve the repression efficiency and methylation sensitivity of the pap2 operon by trippling the Lrp binding site 2. This approach is defined by the structural biological motivation of designing an epigenetic regulation system that switches between the ON and OFF state of gene expression more efficiently. However, studying the pap222 operon and comparing the results with the findings of pap123 and pap2 also gives information about the cooperative binding of PapI-Lrp and the methylation sensitivity of the system. The plasmid pBEST-pap222-UTR1-eGFP-T500 is determined to be toxic during bacterial amplification. Changing the -35 and -10 consensus sequences in order to decrease the binding affinity of σ^{70} could not solve this problem. Thus, instead of plasmid DNA, a linear PCR product, containing the pap222 operon, the untranslated region 1, the gene coding for eGFP, and the transcription terminator T500 is used to study the Lrp mediated repression of pap222-eGFP. *In vitro* gene expression from linear template DNA is known to be less efficient compared to an expression from plasmids. For the study of pap222-eGFP it is assumed that using linear PCR products for gene expression instead of plasmid DNA does not influence the methylation dependent binding affinity of Lrp and PapI. This, however, is not confirmed experimentally in the context of this work.

Lrp mediated repression of pap222-eGFP

Fig. 6.9 shows the results of the Lrp mediated repression of pap222-UTR1-eGFP-T500. Since linear DNA is degraded in the cell free extract, the batch mode expression experiment is performed with the addition of 3.3 μM of the lambda phage protein gamS, protecting the exposed 3' end of the reporter DNA templates. In the case of Dam methylated pap222-eGFP, PapI-Lrp hardly represses the reporter expression. Compared to the Lrp free case, the expression level of eGFP is even improved by about 15% when adding 1 nM of the plasmid pBEST-OR2-OR1-Pr-UTR1-Lrp-T500. This result is surprising when noting the experimental findings of the pap2 operon (section 6.1.2). By trippling the Lrp binding site 2 (pap222 operon), one would simply expect an amplification of the PapI-Lrp repression efficiency, measured for the case of just using one Lrp binding site 2 (pap2 operon). Thus, in agreement with the results presented

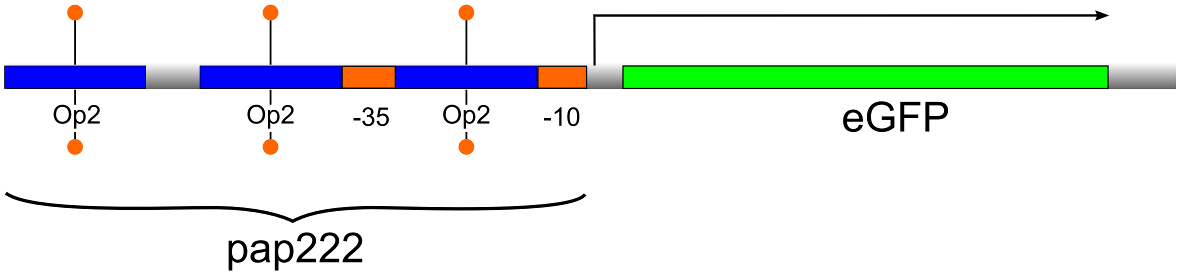


Figure 6.8.: Schematic illustration of the pap222 operon. Three Lrp binding sites 2 (compare Table 2.1) regulate the expression of the reporter gene coding for eGFP. The operon 2 contains a GATC site that can either be methylated or non-methylated (orange pin). One of the three operons is cloned in-between the -35 -10 recognition sequence for σ^{70}. Due to a toxicity of pBEST-pap222-UTR1-eGFP-T500, the plasmid could not be used as template DNA for the study of the pap222 regulated eGFP expression. Instead, the linear PCR product pap222-UTR1-eGFP-T500, containing the sequence for pap222, the untranslated region 1, the gene coding for eGFP, and the transcription terminator T500 is used.

in section 6.1.2, the experiments suggest that the functionality of pap gene switching cannot be explained by the simple, uncooperative attachment of PapI-Lrp complexes to double-stranded DNA and the resulting blocking of the RNA polymerase. In this sense it is also surprising that PapI-Lrp is still capable of repressing the system in the case of non-methylated pap222-eGFP template DNA. However, the repression efficiency is reduced compared to pap2 by a factor of 15%. Although pap222-eGFP shows a PapI-Lrp mediated repression only in the Dam^{-} case but not in the Dam^{+} case, the repression efficiency in the case of non-methylated reporter DNA is not very high. Thus, the pap222 operon cannot be used as an efficient epigenetic switch for *in vitro* gene circuits.

6.1.4. The pap456 operon

The pap456 operon is designed to study the cooperative binding of PapI-Lrp to the distal pap promoter consisting of Lrp binding sites 4, 5, and 6 (compare chapter 2.4.1). In the native system, the binding of Lrp to either the proximal promoter (sites 1, 2, and 3) or the distal promoter (sites 4, 5, and 6) is dictated by the PapI level, since PapI increases the Lrp affinity for binding to the distal sites rather than the affinity for binding to the proximal sites [17]. The operon is schematically illustrated in Fig. 6.10. It

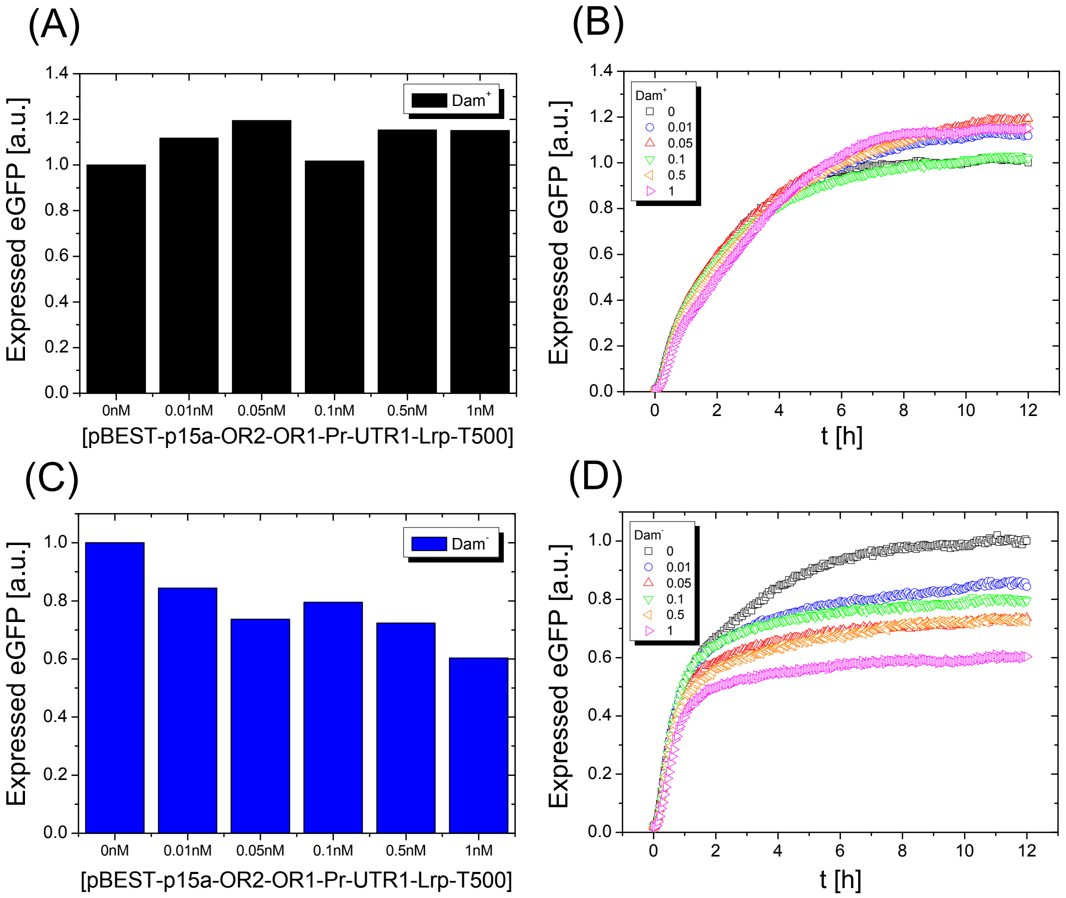


Figure 6.9.: Methylation response of the Lrp mediated repression of the PCR template DNA pap222-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap222 operon in pap222-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter DNA is 5 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI and 3.3 µM of the lambda phage protein gamS to protect the linear PCR product from degradation. The eGFP expression level is hardly influenced by the presence of Lrp. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). **(C), (D)**, same as (A) and (B) but for the non-methylated pap222 operon. Increasing the Lrp concentration leads to a reduced eGFP expression level by a factor of 40%.

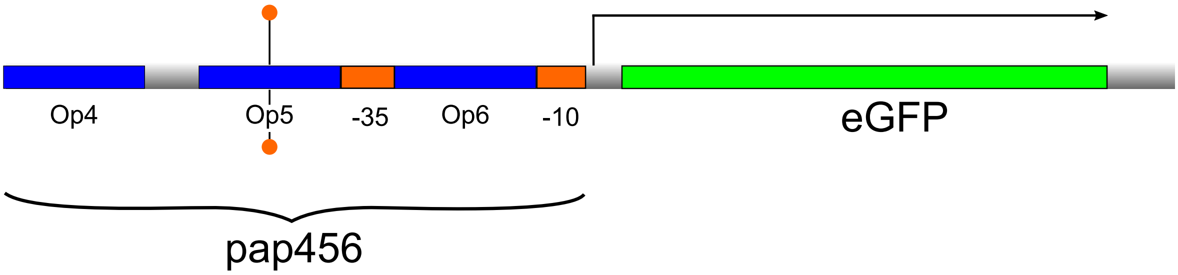


Figure 6.10.: Schematic illustration of the pap456 operon. The Lrp binding sites 4, 5, and 6 (compare Table 2.1) regulate the expression of the reporter gene coding for eGFP. The operon 5 contains a GATC site that can either be methylated or non-methylated (orange pin). The operon 6 is cloned in-between the -35 -10 recognition sequence for σ^{70}. Due to a toxicity of the plasmid pBEST-pap456-UTR1-eGFP-T500, the -10 consensus sequence of σ^{70} is changed from ATAAT to ATATT, decreasing the affinity of the sigma factor. However, the sequence of Lrp binding site 6 is not changed.

contains the Lrp binding site 5, which can be either Dam methylated or non-methylated. Site 5 is flanked by sites 4 and 6, while 6 is overlapping with the -35 -10 region of the promoter. Due to a toxicity during plasmid amplification, the -10 consensus sequence is changed to ATATT in order to decrease the affinity of the transcription factor σ^{70}. In this case the plasmid is stable during amplification.

eGFP expression as a function of the reporter plasmid and the PapI concentration

PapI is known to be a protein that can unspecifically bind to DNA also in the absence of Lrp [17]. Both, Lrp and PapI, acquire enhanced affinity to bind to the distal pap promoter sites 4, 5, and 6, to which neither of them binds to an appreciable extent without the other [21]. Thus, it is necessary to analyze if PapI alone is capable of repressing the system regulated by the pap456 operon, also in the absence of Lrp. This is done in the following experiment.

Fig. 6.11 shows the eGFP expression regulated by the pap456 operon as a function of both, the eGFP reporter plasmid concentration and the concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500. The experiment shows that for Dam methylated and non-methylated reporter plasmids the eGFP expression is hardly modified as a function of the PapI concentration in the cell free extract. Only in the case of very high reporter

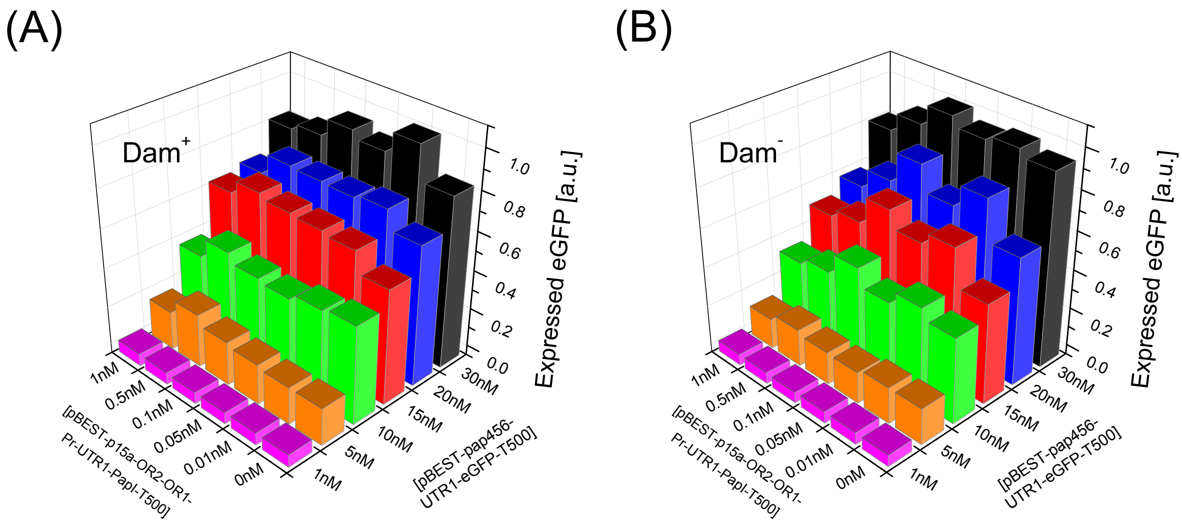


Figure 6.11.: PapI response of the expression efficiency of eGFP regulated by the pap456 operon for the case of Dam methylated **(A)** and non-methylated **(B)** reporter plasmids pBEST-pap456-UTR1-eGFP-T500. The graph shows the expression level of eGFP normalized to its highest value as a function of the concentration of the reporter plasmid pBEST-pap456-UTR1-eGFP-T500 as well as the concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, expressing PapI. The amino acid concentration is 0.5 mM. In the case of the Dam$^-$ results, a decreased expression can be observed for reporter plasmid concentrations of $15 - 30$ nM and PapI plasmid concentrations of $0.5 - 1$ nM, most probably caused by a sharing effect of the expression machinery at high plasmid concentrations.

plasmid concentrations of 15 nM, 20 nM, and 30 nM as well as high concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 of 1 nM, a decrease of the expression level can be observed, particularly in the case of non-methylated reporters. However, this is most likely due to a sharing effect of the expression machinery observed for very high total plasmid concentrations. This can be neglected for a total plasmid concentration less than about 10 nM. For the subsequent experiments one can conclude that even though PapI is in principle able to bind unspecifically to DNA [21], it does not influence the expression efficiency of eGFP regulated by the pap456 operon, when using typical reporter plasmid concentrations of 5 nM.

Lrp mediated repression of pap456-eGFP

In order to analyze the Lrp mediated and cooperative repression of eGFP regulated by the pap456 operon, batch mode expression experiments with increasing concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are performed for Dam methylated and non-methylated reporter plasmids in the presence of PapI.

Fig. 6.12 shows the respective results. The cooperative binding of PapI-Lrp leads to a repressed eGFP expression level for both, Dam methylated (by 80%) and non-methylated (by about 70%) reporter plasmids. The observed methylation response is only weak and works in the opposite direction as observed for the pap123 operon. This indicates that the methylation state of Lrp binding site 5 has no significant hindering effect on the cooperative binding of Lrp to the distal promoter sites 4, 5, and 6 in the presence of PapI, since only site 6 is overlapping with the -35 -10 region of the promoter. This is in contrast to the *in vitro* findings in literature [17] that Dam methylation of site 5 decreases the affinity of PapI-Lrp for sites 4-6 and thus stabilizing the phase-OFF cells not expressing pap. The methylation state of site 5 does not seem to play an essential role for the cooperative locking of the phase-OFF state, although the expression regulated by Lrp binding site 5 alone (see pap5 operon, section 6.1.5) shows a significant methylation response (see section 6.1.5).

PapI response of the Lrp mediated repression of pap456-eGFP

In order to gain further insight into the role of PapI for mediating the Lrp binding to the distal promoter sites 4, 5, and 6, the repression experiments presented in Fig. 6.12 are repeated without the addition of 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500.

Fig. 6.13 shows the results of the Lrp mediated repression of pap456-eGFP in the absence of PapI. Comparing the results with the findings in Fig. 6.12 reveals a less important repression efficiency of Lrp in the absence of PapI in the case of Dam^+ reporters ($\mathrm{Dam}^+/\mathrm{PapI}^+$: 80%, $\mathrm{Dam}^+/\mathrm{PapI}^-$: 70%) but comparable repression efficiency in the case of Dam^- reporters (Dam—/PapI^+: 70%, Dam—/PapI^-: 75%). Only for the Dam^+ case, this result is in agreement with the findings of the literature that PapI slightly increases the affinity for the cooperative binding of LRP to the distal sites 4, 5, and 6 [17]. However, the result is in contrast to the finding that PapI increases the affinity for all distal sites 4, 5, and 6 rather than for proximal sites 1, 2, and 3 (compare section 6.1.1). Based on the batch mode expression results, the presence of PapI does

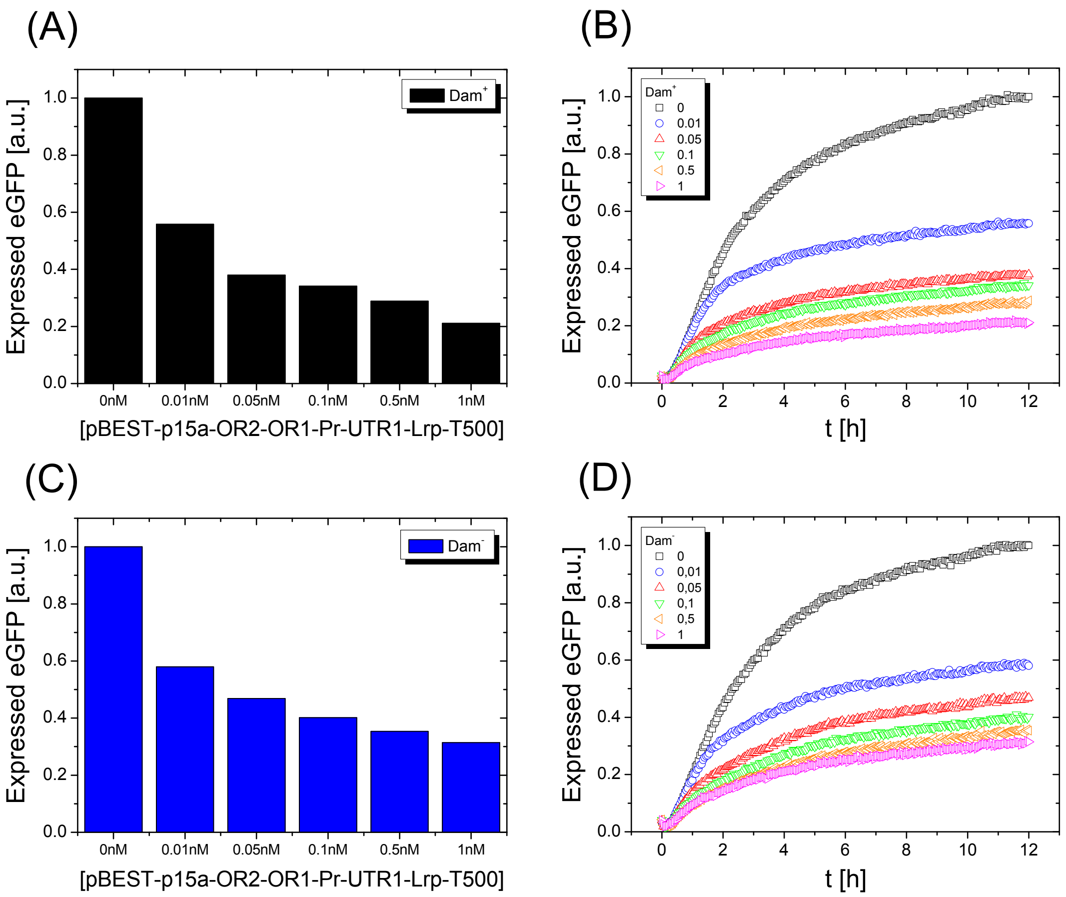


Figure 6.12.: Methylation response of the Lrp mediated repression of pBEST-pap456-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap456 operon in pBEST-pap456-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter plasmid is 5 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). The eGFP expression level decreases with increasing concentrations of Lrp by a factor of 80% (for 1 nM of the Lrp plasmid) in the presence of PapI. **(C), (D)**, same as (A) and (B) but for the non-methylated pap456 operon. In the presence of PapI, Lrp represses the system by a factor of 70% (for 1 nM of the Lrp plasmid).

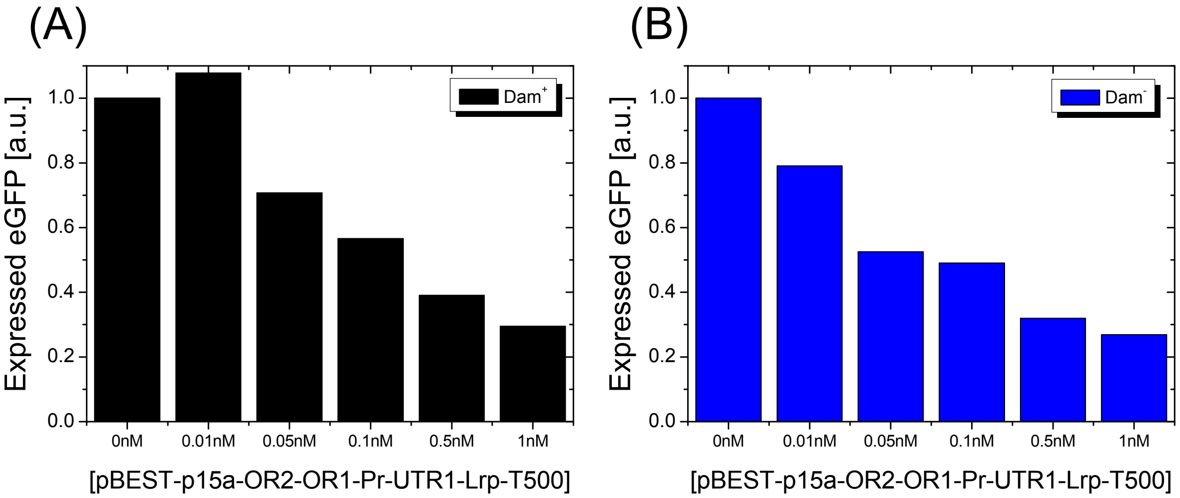


Figure 6.13.: Methylation response of the Lrp mediated repression of pBEST-pap456-UTR1-eGFP-T500 without expressing the co-repressor PapI. **(A)**, Dam methylated reporter plasmids. The graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of reporter plasmid and amino acids is 5 nM and 0.5 mM, respectively. Data is normalized to the equilibrium value of the eGFP expression without the addition of the Lrp template DNA. **(B)**, same as (A) but for the case of non-methylated reporter plasmids.

not alter the cooperative binding of PapI-Lrp to all distal sites 4, 5, and 6, although the expression regulated by the Lrp binding site 5 alone (no cooperativity) shows a significant PapI response (see pap5 operon, section 6.1.5).

6.1.5. The pap5 operon

The pap5 operon (Fig. 6.14) is designed to study the PapI mediated binding of Lrp without any cooperative effects. Compared to the pap456 operon, the flanking Lrp binding sites 4 and 6 are removed in pap5 and only binding site 5 is cloned in-between the -35 -10 region of the promoter. Thus, it is possible to study the direct methylation response of the binding of PapI-Lrp to this operon. According to literature [17] the Lrp binding site 5 shows the highest methylation response of the binding of Lrp. Additionally, it is known that Lrp forms more stable complexes with its binding site 5 in the presence and absence of PapI, compared to respective complexes formed with binding site 2.

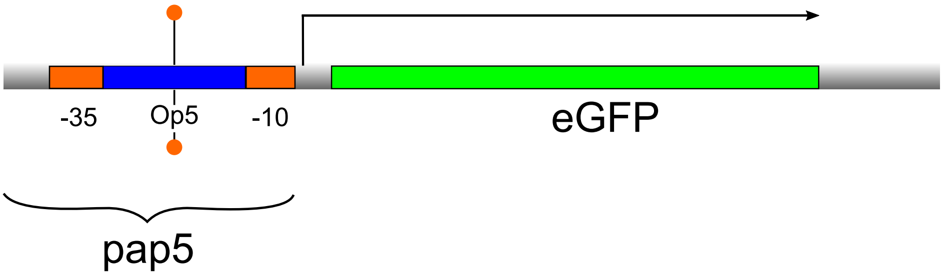


Figure 6.14.: Schematic illustration of the pap5 operon. The Lrp binding sites 5 (compare Table 2.1) is cloned in-between the -35 -10 recognition sequence for σ^{70} and regulates the expression of the reporter gene coding for eGFP. The operon 5 contains a GATC site that can either be methylated or non-methylated (orange pin).

eGFP expression as a function of the reporter plasmid and the PapI concentration

Lrp is known to show a strong PapI response when binding to its recognition site 5 [17]. Just like in the case of the pap456 operon, it is therefore necessary to analyze if PapI alone is capable of repressing the system regulated by the pap5 operon before studying the PapI-Lrp mediated repression of the system. This is done in the following experiments.
By analogy with Fig. 6.11, Fig. 6.15 shows the eGFP expression regulated by the pap5 operon as a function of both, the eGFP reporter plasmid concentration and the concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500. The experiment reveals a similar result as shown in section 6.1.4 for the pap456 operon. For the case of Dam methylated and non-methylated reporter plasmids, the eGFP expression is hardly modified when changing the PapI concentration in the cell free extract. Only for a reporter plasmid concentration of 10 nM and a pBEST-OR2-OR1-Pr-UTR1-PapI-T500 concentration of $0.5 - 1$ nM, a decrease of the eGFP repression level can be observed in the case of non-methylated reporter plasmids. Like supposed for the case of the pap456 operon, this is most likely due to a sharing effect of the expression machinery for high plasmid concentrations. Put together, one can conclude, that PapI is not able to repress the system regulated by the pap5 operon in the absence of Lrp.

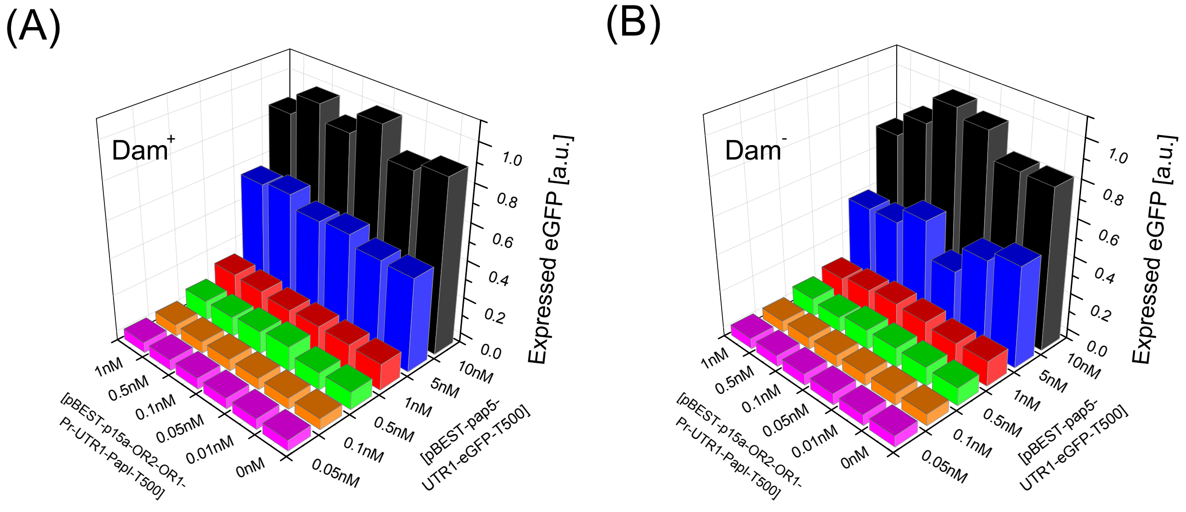


Figure 6.15.: PapI response of the expression efficiency of eGFP regulated by the pap5 operon for the case of Dam methylated **(A)** and non-methylated **(B)** reporter plasmids pBEST-pap5-UTR1-eGFP-T500. The graph shows the expression level of eGFP normalized to its highest value as a function of the concentration of the reporter plasmid pBEST-pap5-UTR1-eGFP-T500 as well as the concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, expressing PapI. The amino acid concentration is 0.5 mM. In the case of the Dam^- results, a decreased expression can be observed for reporter plasmid concentration of 10 nM and PapI plasmid concentrations of 0.5—1 nM, most probably caused by a sharing effect of the expression machinery at high plasmid concentrations.

Lrp mediated repression of pap5-eGFP

Fig. 6.16 shows the results of the Lrp mediated repression of pap5-eGFP for Dam^+ and Dam^- reporter plasmids. Dam methylation of the GATC consensus sequence located in the binding site 5 causes the most significant change in repression efficiency of PapI-Lrp of all operons studied in the context of this work. In the case of Dam^+ reporters, the expression level of eGFP increases as a function of the Lrp plasmid concentration by a factor of 54%. This effect is seemingly dramatic, when noting that PapI-Lrp is only known to act as a repressor in the context of the epigenetic regulation of pap expression, but never as an activator. In the case of Dam^- reporters, the presence of PapI-Lrp represses the system by a factor of about 75%, like it has already been observed for other pap operons. However, the kinetics reveals a particularly pronounced “leakproof” repression, since the eGFP expression level does not increase on a long timescale after

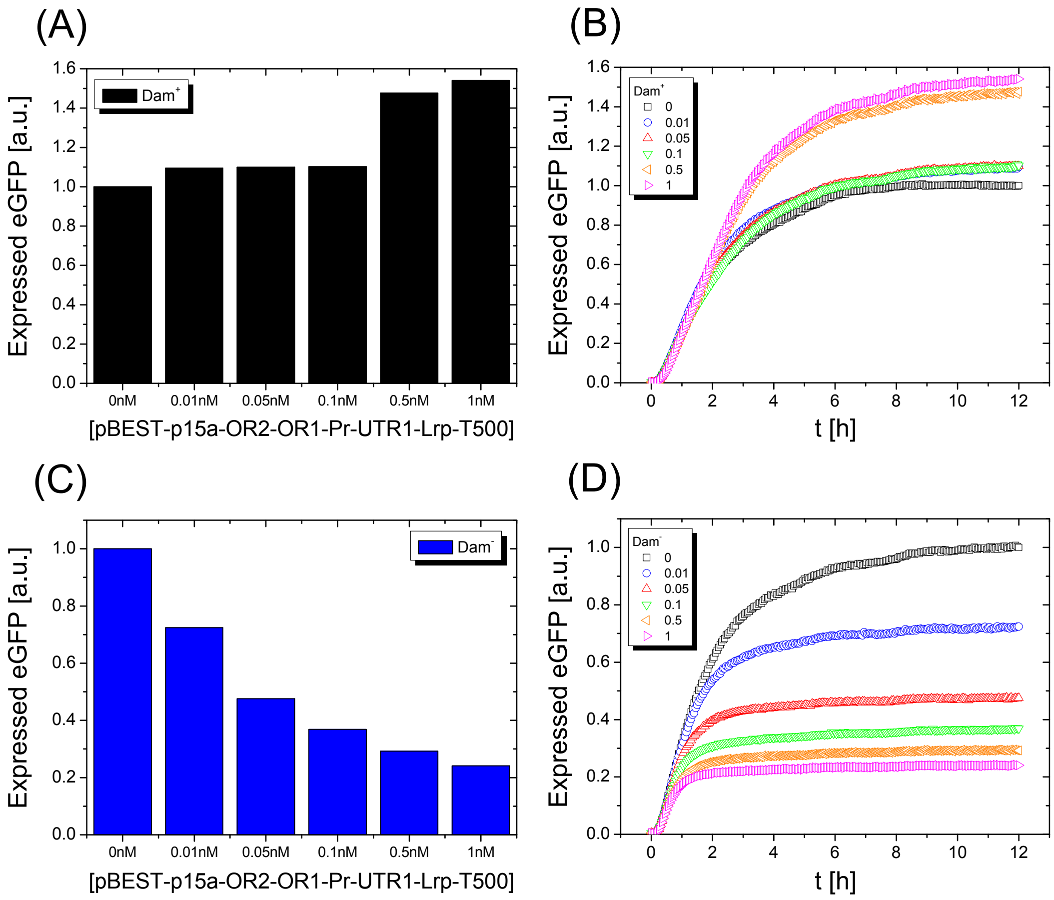


Figure 6.16.: Methylation response of the Lrp mediated repression of pBEST-pap5-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap5 operon in pBEST-pap5-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter plasmid is 5 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). The eGFP expression level increases with increasing concentrations of Lrp by a factor of 54% (for 1 nM of the Lrp plasmid) in the presence of PapI. **(C), (D)**, same as (A) and (B) but for the non-methylated pap5 operon. In the presence of PapI, Lrp represses the system by a factor of 75% (for 1 nM of the Lrp plasmid).

the linear phase. One can conclude from the experiments that methylation of the pap5 operon obviously changes the role of PapI-Lrp from a repressor (in the case of the non-methylated Lrp site 5) to the role of an activator (in the case of the Dam methylated Lrp site 5). The mechanism by which the presence of PapI-Lrp could improve the expression efficiency of eGFP is unknown. One can speculate that binding of PapI-Lrp to the methylated version of the pap5 operon produces a more accessible DNA conformation for the annealing of σ^{70}, while the non-methylated DNA hinders this annealing of the sigma factor. For the purpose of constructing an epigenetic switch, the pap5 operon shows the best switching behavior of all the constructs studied in the context of this work. For a concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 of 1 nM, the Dam^{+} expression exceeds the Dam^{-} expression by one order of magnitude.

PapI response of the Lrp mediated repression of pap5-eGFP

According to literature, the affinity of Lrp for binding to the pap operon 5 is markedly increased in the presence of PapI [17]. It is further known that Lrp specifically interacts with the binding site 5 only in the presence of PapI and that Dam methylation of site 5 inhibits the formation of the ternary complex PapI-Lrp-DNA [21]. All this is observed for *in vitro* experiments. In the following it is analyzed if the presence of PapI influences the eGFP expression regulated by the pap5 operon. Therefore, the Lrp mediated repression of eGFP is analyzed in the presence and absence of the co-repressor PapI for both, the methylated and non-methylated version of the pap5 operon.

The results are presented in Fig. 6.17. In agreement with [21], the experiment reveals that PapI is necessary for the methylation sensitivity of the system. In the presence of PapI, increasing the plasmid concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 leads to a reduced expression level of about 75% in the case of Dam^{-} reporters and an enhanced expression level of about 50% in the case of Dam^{+} reporters. This behavior is already described in Fig. 6.16. However, in the absence of PapI, the expression level increases in both, the Dam^{+} and Dam^{-} case, as a function of the Lrp plasmid concentration. The improvement of the expression is of comparable magnitude as in the case of Dam^{+}/PapI^{+}. Obviously, the presence of PapI is required for the Lrp mediated repression of the system in the case of the non-methylated version of the pap5 operon. One can conclude that only the formation of the ternary complex consisting of PapI, Lrp and the pap5 DNA leads to a repression of the system. On the other hand, formation of the binary complex consisting of Lrp and the pap5 DNA leads to an enhanced expression efficiency of eGFP. Dam methylation of the pap5 operon

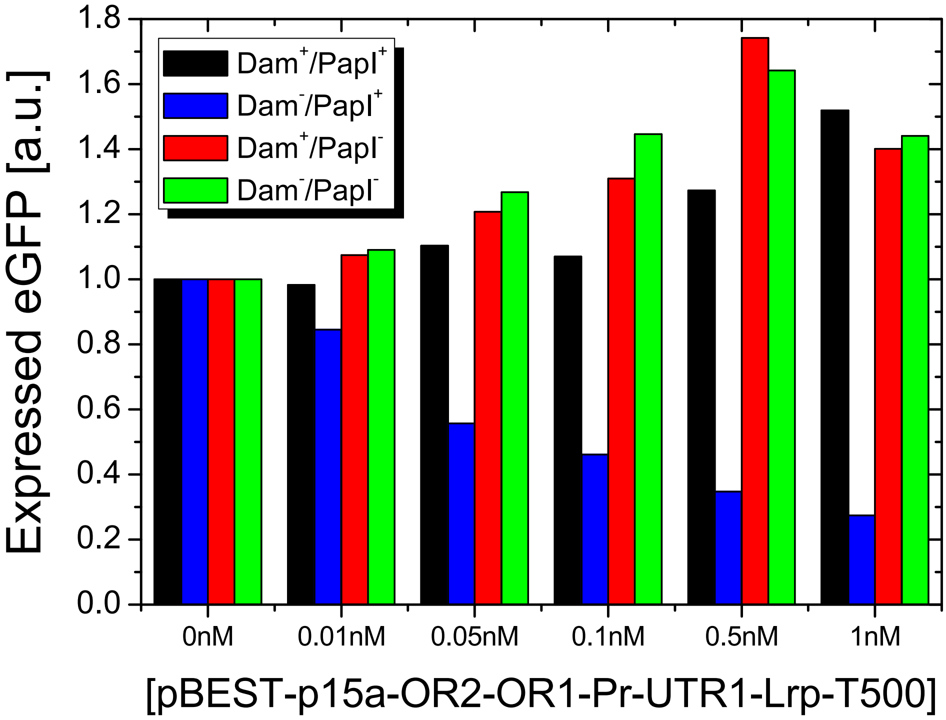


Figure 6.17.: PapI response of the Lrp mediated repression of eGFP regulated by the pap5 operon as a function of the methylation state of the reporter plasmid. The graph shows the eGFP expression level as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 for Dam methylated (black and red) and non-methylated reporters (blue and green), in the presence (black and blue) and absence (red and green) of 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500. Data is normalized to the equilibrium value of the eGFP expression without the addition of the Lrp plasmid. The presence of PapI is required for the Lrp mediated repression of the system in the case of the non-methylated version of the pap5 operon (blue bars).

inhibits formation of the ternary complex. For the purpose of constructing an epigenetic switch, all experiments should therefore be performed only in the presence of PapI.

eGFP expression as a function of the amino acid concentration

Since the pap5 operon shows the best switching behavior as a function of Dam methylation, it is interesting to analyze if the switching efficiency can be increased for different amino acid concentrations within the cell free reaction. This is done in the following experiment.

Fig. 6.18 shows the eGFP expression regulated by the Dam methylated and non-

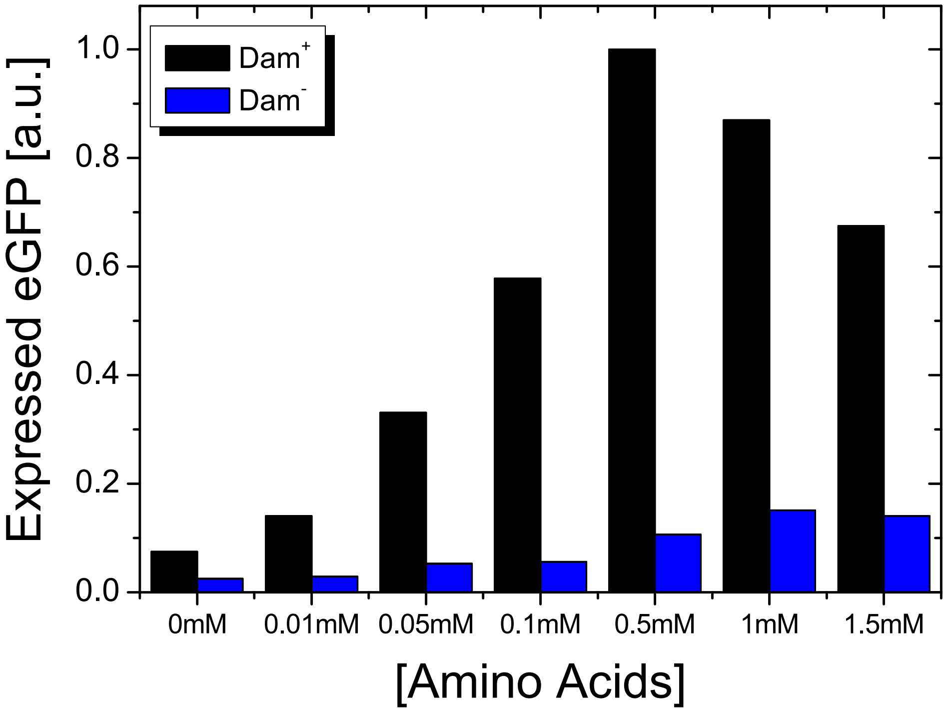


Figure 6.18.: Batch mode expression of pBEST-pap5-UTR1-eGFP-T500 as a function of the amino acid concentration in the cell free reaction. The graph shows the concentration of the recombined eGFP as a function of the amino acids in the case of the Dam methylated pap5 operon (Dam$^+$, black) and the non-methylated operon (Dam$^-$, blue). Data is normalized to the Dam$^+$ expression level reached for an amino acid concentration of 0.5 mM. The value on the x-axis refer to the concentration of each of the 20 native amino acids. The concentration of the reporter plasmid is 5 nM. Additionally, the reaction contains 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500, expressing the regulator Lrp, as well as 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, expressing the co-regulator PapI. The best switching efficiency is reached for an amino acid concentration of 0.5 mM (90% difference between Dam$^+$ and Dam$^-$ expression).

methylated pap5 operon, as a function of the amino acid concentration within the cell free extract. The experiment reveals the best switching efficiency of the system to be reached for an amino acid concentration of 0.5 mM, a reporter plasmid concentration of 5 nM, and 1 nM of each plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 and pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500, respectively. Under this experimental conditions the difference between Dam$^+$ and Dam$^-$ expression is determined to be about 90%.

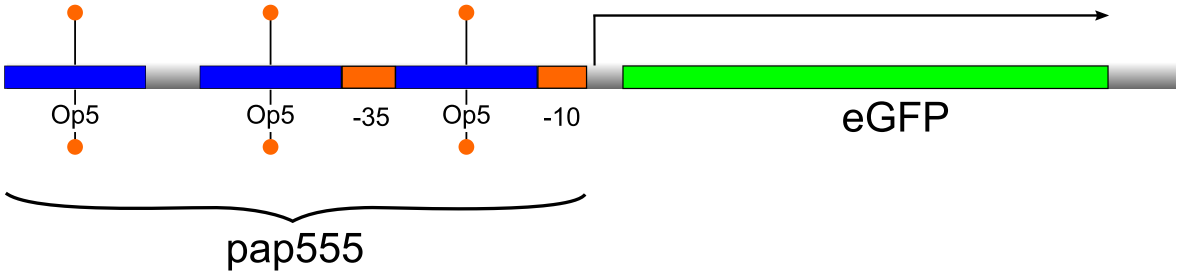


Figure 6.19.: Schematic illustration of the pap555 operon. Three Lrp binding sites 5 (compare Table 2.1) regulate the expression of the reporter gene coding for eGFP. The operon 5 contains a GATC site that can either be methylated or non-methylated (orange pin). One of the three operons is cloned in-between the -35 -10 recognition sequence for σ^{70}. Due to a toxicity of pBEST-pap555-UTR1-eGFP-T500, the plasmid could not be used as template DNA for the study of the pap555 regulated eGFP expression. Instead, the linear PCR product pap555-UTR1-eGFP-T500, containing the sequence for pap555, the untranslated region 1, the gene coding for eGFP, and the transcription terminator T500 is used.

6.1.6. The pap555 operon

The pap555 operon is constructed in order to further amplify the switching properties of the pap5 operon. The idea is that if using one Lrp binding site 5 as a regulator for the expression of eGFP already shows an efficient switching behavior, tripling this binding site 5 should further improve the binding affinity of the Pap-Lrp complex and thus lead to an even more efficient switching. On the other hand, studying the pap555 operon and comparing the results with the findings of pap456 and pap5 also gives information about the cooperative binding of PapI-Lrp and the methylation sensitivity of the system. The pap555 operon is schematically illustrated in Fig. 6.19. It contains three Lrp binding sites 5 that regulate the expression of the reporter. All binding sites can either be Dam methylated or non-methylated. By analogy to the pap222 operon (section 6.1.3), the plasmid pBEST-pap555-UTR1-eGFP-T500 is determined to be toxic during bacterial amplification and changing the -35 and -10 consensus sequences in order to decreasing the binding affinity of σ^{70} could not solve this problem. Thus, a linear PCR product consisting of the pap555 operon, the untranslated region 1, the gene coding for eGFP, and the transcription terminator T500 is used to study the Lrp mediated repression of pap555-eGFP. The assumptions for expression experiments with linear DNA templates explained in section 6.1.3 have also to be considered in the case of pap555-eGFP.

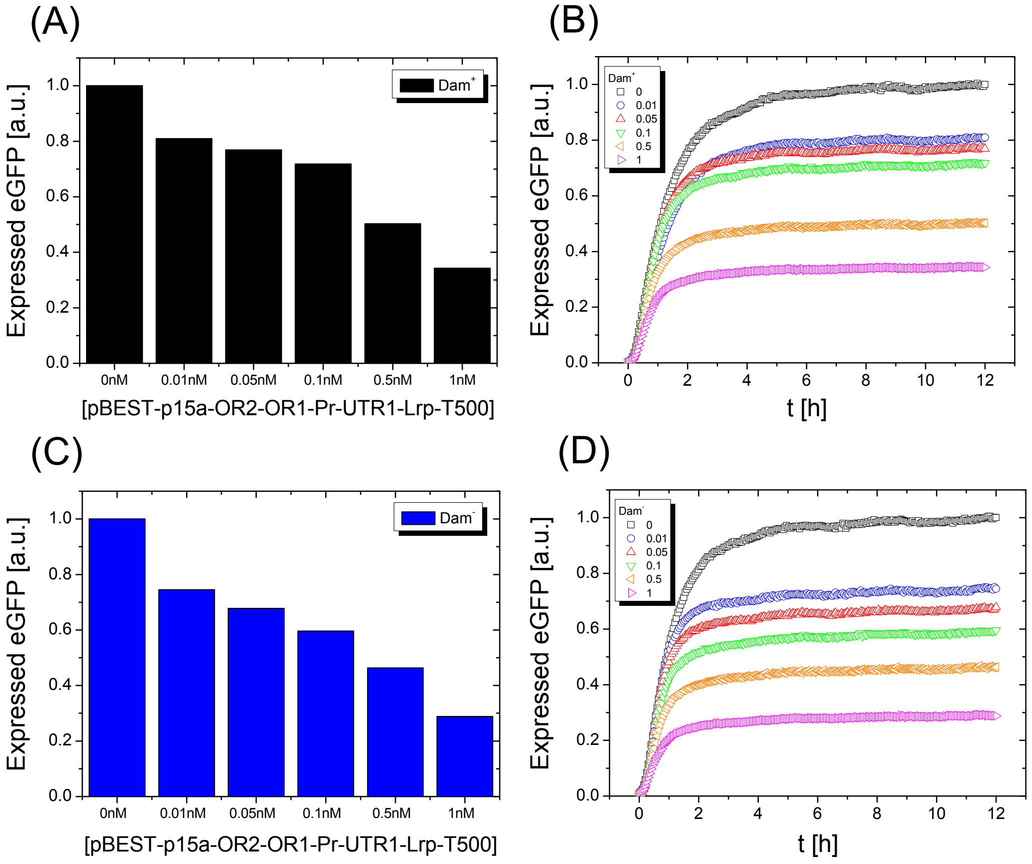


Figure 6.20.: Methylation response of the Lrp mediated repression of the PCR template DNA pap555-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the Dam methylated pap555 operon in pap555-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the Lrp template DNA. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500. The concentration of the reporter DNA is 5 nM, and the amino acid concentration 0.5 mM. The reaction further contains 1 nM of pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 expressing the co-regulator PapI and 3.3 µM of the lambda phage protein gamS to protect the linear PCR product from degradation. The eGFP expression level decreases with increasing concentrations of Lrp by a factor of 66% (for 1 nM of the Lrp plasmid) in the presence of PapI. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 are given in the legend (values in nM). **(C), (D)**, same as (A) and (B) but for the non-methylated pap555 operon. Increasing the Lrp concentration leads to a reduced eGFP expression level by a factor of 70%.

Lrp mediated repression of pap555-eGFP

Fig. 6.20 shows the result of the Lrp mediated repression of pap555-UTR1-eGFP-T500. By analogy with pap222-UTR1-eGFP-T500, the expression experiment is performed with the addition of 3.3 μM of the lambda phage protein gamS, protecting the exposed 3′ end of the reporter DNA templates. The experiment reveals that trippling the Lrp binding site 5 leads to a loss of the methylation sensitivity of the system observed in the case of the pap5 operon. The presence of PapI-Lrp represses the system by a factor of 66% (Dam$^+$) and 70% (Dam$^-$). The repression efficiency is less pronounced compared to the case of the pap456 operon (compare section 6.1.4). This result verifies the conclusions presented in section 6.1.3 that regulation of the pap expression cannot be explained with the simple linear binding of PapI-Lrp and the subsequent blocking of the RNA polymerase. In this case, trippling the Lrp binding site 5 should simply lead to an amplification of the effects observed for the pap5 operon, which is, however, not the case. The results underline that the cooperative interaction of Dam methylation, PapI-Lrp, and all six pap regulatory operons is absolutely necessary for the efficient regulation of the pap expression. The fact that the eGFP expression regulated by the pap555 operon no longer shows any methylation response excludes the construction for using it as an epigenetic switch for *in vitro* gene circuits.

6.2. Gene switching based on CpG methylation and MBD

In the following, the *in vitro* expression results of the epigenetic gene switch based on CpG methylation and the methylation binding domain (MBD) of the human repressor MeCP2 are presented. By analogy with the experiments performed on the pap switch (section 6.1), the experiments involve cloning of the MBD recognition sequence (see Fig. 2.11) as the regulatory part for the expression of the reporter proteins eGFP or mCherry. Measuring the fluorescence intensity of the recombined reporters gives information about the binding of MBD. The methylation binding domain is expressed from the vector pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 within the cell free extract. The yield of recombined MBD is about 20 μM for a plasmid concentration of 4 nM [35]. For further investigation of the methylation dependent binding of MBD, the protein is also obtained in its recombined form and directly introduced into the cell free reaction. By analogy to the pap switch (section 6.1), the second objective of the experimental study is the design of an epigenetic switch on the basis of MeCP2 and CpG methylated DNA. This is interesting because MeCP2 is known to possess exactly the opposite binding properties than PapI-Lrp for the case of the pap5 operon (section 6.1.5). While Dam methylation of the DNA blocks high affinity binding of the PapI-Lrp complex, CpG methylation is required for the high affinity binding of MeCP2. CpG methylation of the regulatory part of all reporter plasmids is performed as explained in chapter 3.2.2. The extract is produced from the mcr positive E. coli strain BL21. Thus, CpG methylated DNA can be subject to degradation by the mcr restriction system if it contains a specified recognition sequence (see appendix section B.7). However, all CpG methylated reporter plasmids used in the context of this work (compare appendix Table B.1) do not contain this recognition site of the mcr system and can therefore be expressed by the crude extract.

6.2.1. The mbd operon

The mbd operon is designed to study the methylation dependent gene repression by the methylation binding domain of the human protein MeCP2. The design of the mbd operon involves cloning of the promoter III of the mouse-brain-derived neurodrophic factor (BDNF) gene as an artificial promoter for the regulation of the expression of eGFP and mCherry. Fig. 6.21 shows a scheme of the mbd operon. The DNA regulatory sequence contains a central CpG site that can be either methylated or non-methylated. It is known that MeCP2 possesses high affinity for binding to this sequence motif as a function of the methylation state of the operon [25, 26].

eGFP and mCherry expression as a function of the reporter plasmid concentration

In order to analyze the promoter strength of the mbd operon, batch mode expression experiments with increasing concentrations of the reporter plasmids pBEST-mbd-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500 are performed. These measurements are also performed in order to determine the optimal reporter plasmid concentrations for all subsequent expression experiments and also for comparing the expression efficiency of the two reporter proteins with each other. Furthermore, the experiment is performed to analyze if CpG methylated reporter plasmids are not subject to degradation by the mcr restriction system of the cell free extract.

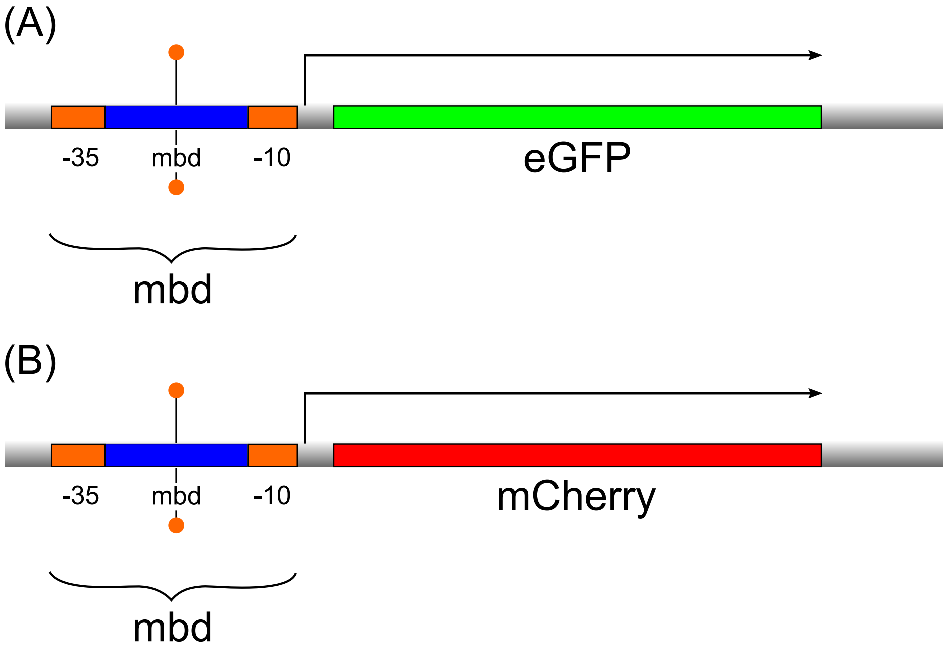


Figure 6.21.: Schematic illustration of the mbd operon. The promoter III of the mouse-brain-derived neurodrophic factor (BDNF) gene (compare Fig. 2.11) is cloned in-between the -35 -10 recognition sequence for σ^{70} and regulates the expression of the reporter gene coding for eGFP **(A)** and mCherry **(B)**. The sequence contains a CpG site that can either be methylated or non-methylated (orange pin).

Fig. 6.22 shows the yield of recombined reporters eGFP and mCherry as a function of the respective plasmid concentration. The experiment reveals a comparable expression level of both reporter proteins for any given plasmid concentration. Based on these results and by analogy with the experiments performed on the pap switch (section 6.1), the standard reporter plasmid concentration is adjusted to 5 nM for subsequent experiments, when using the mbd operon as a promoter. Choosing this concentration enables the simultaneous expression of other molecular players from plasmid DNA, like for instance MBD, within the same cell free reaction without considering any sharing artifacts. Here, the concentration of all plasmids used in one single batch mode reaction should not exceed a total value of 10 nM. The experimental results also demonstrate that both plasmids, pBEST-mbd-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500 are obviously mcr negative and thus not subject to degradation within the cell free extract. Comparing the reporter expression levels of the CpG^{+} case with the respective ones of the CpG^{-} case reveals expression from CpG methylated plasmids to be less efficient than expression from plasmids free of CpG methylation (compare for instance the

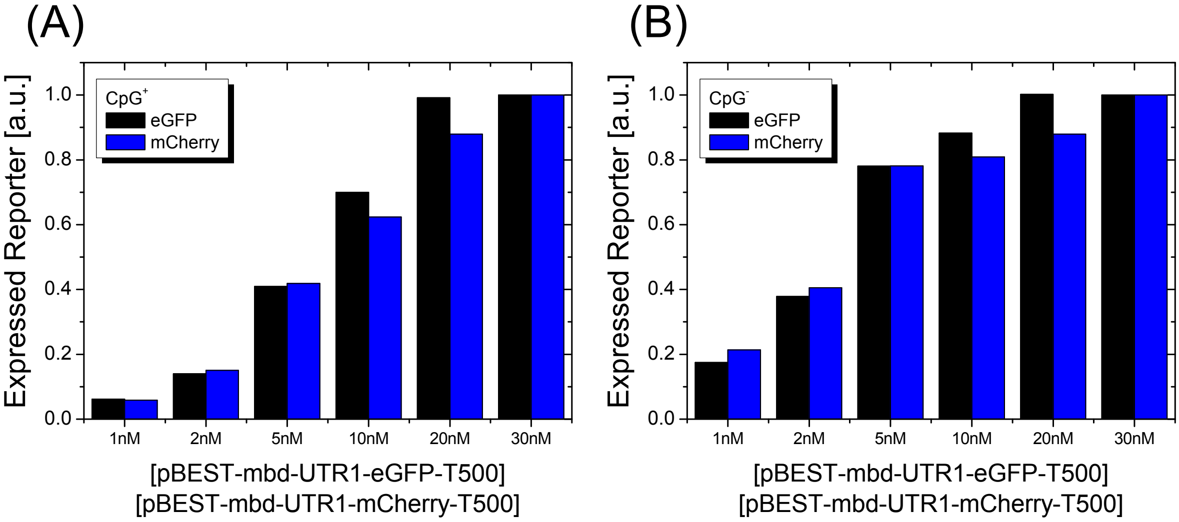


Figure 6.22.: Batch mode expression of eGFP and mCherry regulated by the mbd operon as a function of the reporter plasmid concentration. **(A)**, expression from CpG methylated reporter plasmids. **(B)**, expression from plasmids free of CpG methylation. The graphs show the concentration of the recombined eGFP (black) and mCherry (blue) as a function of the concentration of the reporter plasmids pBEST-mbd-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500, respectively. Data is normalized to the eGFP and mCherry expression level reached for a reporter plasmid concentration of 30 nM. The concentration of amino acids added to the reaction is 0.5 mM.

expression levels for reporter plasmid concentrations from 1 — 10 nM in Fig. 6.22 A with the respective values in Fig. 6.22 B). The reason for this behavior is unknown. However, degradation of the CpG methylated DNA by the mcr restriction system of the extract can be excluded as the reason for the decreased expression efficiency. Using an mcr positive plasmid as the template for the reporter expression would lead to a vanishing fluorescence signal due to complete degradation of the DNA (compare appendix section B.7). For the interpretation of the experimental data, the difference between the CpG^+ and CpG^- can be neglected. All subsequent results are presented in arbitrary units, what means that the expression level of the reporter protein is shown as a function of a predetermined reference measurement.

MBD mediated repression of mbd-eGFP

Fig. 6.23 shows the result of the MBD mediated repression of mbd-eGFP as a function of the plasmid concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500, for CpG methylated and non-methylated reporter plasmids. In the case of the CpG$^+$ reporter, the eGFP expression level decreases by a factor of 85% when using a pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 plasmid concentration of 1 nM. The annealing of the methylation binding domain to the methylated mbd operon efficiently represses the system. However, there is also a measurable repression of eGFP in the case of non-methylated reporter plasmids by a factor of 43%. Two reasons may apply for this experimental finding: i) the methylation binding domain does also bind to the non-methylated mbd operon but with a lower affinity than for CpG methylated DNA. ii) the measured repression is an artifact caused by a sharing effect of the expression machinery of the cell free extract due to the high total plasmid concentration (5 nM of pBEST-mbd-UT1-eGFP-500 + 4 nM of pBEST-p15a-OR2-OR1-Pr-MBD-UTR1-T500). In order to rule out a sharing effect as the reason for the MBD mediated repression of eGFP in the case of the methylated and non-methylated mbd operon, the experiment presented in Fig 6.23 is repeated with a recombined version of the methylation binding domain of MeCP2. In this case, MBD is not expressed within the cell free reaction from the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-T500. All resources of the expression machinery are available for the reporter expression and sharing artifacts are excluded. The results are presented in Fig. 6.24. The experiment confirms that the repression in the CpG$^+$ and CpG$^-$ case is indeed caused by the specific interaction between MBD and the mbd operon. The repression efficiency of the recombined version of MBD is comparable to the results presented in Fig. 6.23. However, there is a difference between the two experiments concerning the final concentration of recombined MBD. In the case of the plasmid expression of the methylation binding domain, 4 nM of pBEST-p15a-OR2-OR1-Pr-MBD-UTR1-T500 should lead to recombined protein concentration of about 20 μM [35]. This protein concentration leads to a repression efficiency of 85% in the case of the CpG$^+$ reporter. However, when using the recombined version of MBD, 6 μM of the protein lead to a repression efficiency of 92% in the case of the CpG$^+$ reporter. The reason for this difference is the simultaneous expression of eGFP and MBD in the experiments presented in Fig. 6.23. Considering the kinetics of the CpG$^+$ reporter expression for a pBEST-p15a-OR2-OR1-Pr-MBD-UTR1-T500 plasmid concentration of 4 nM reveals the final eGFP expression level to be reached after approximately two hours. After this time, the expression level hardly increases as a function of time until

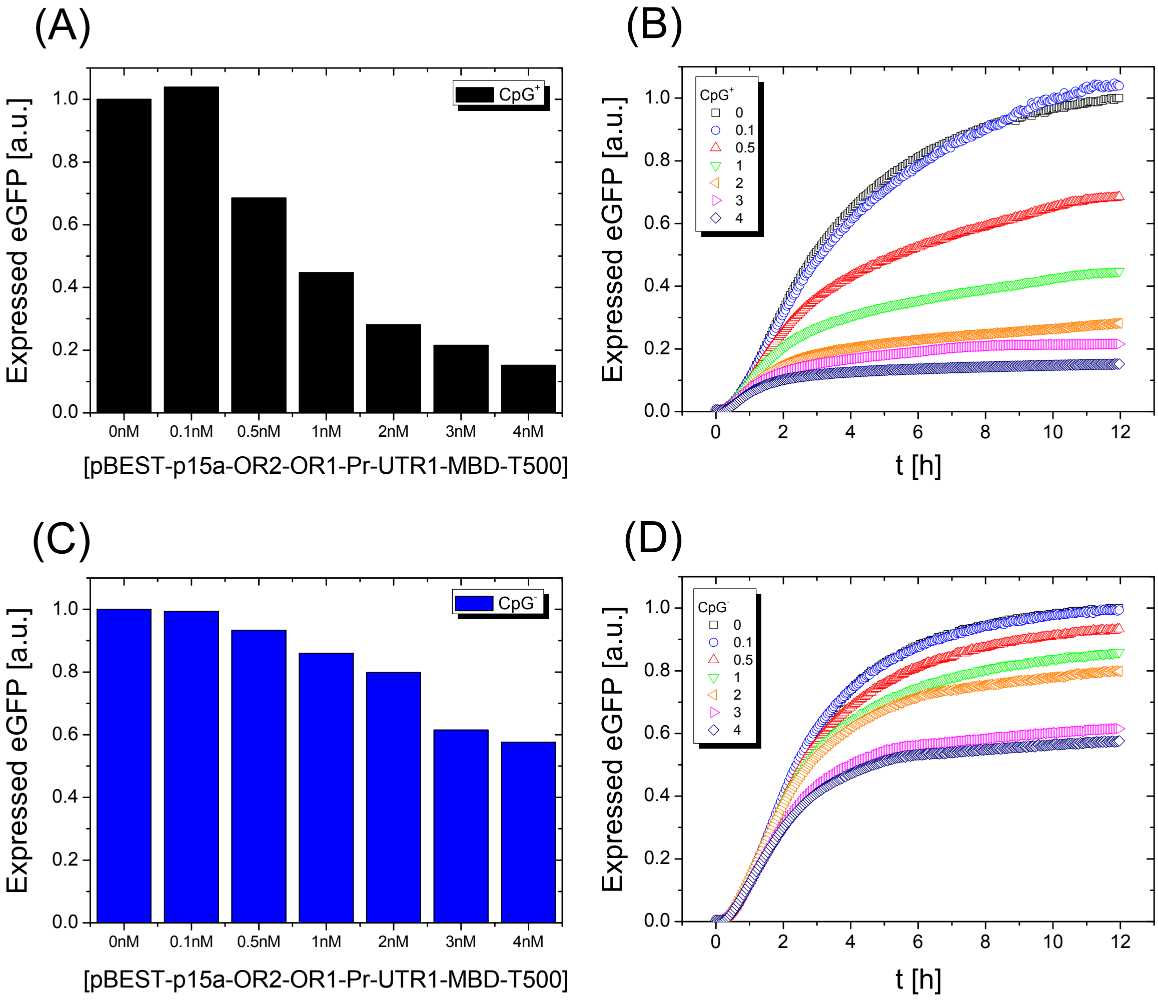


Figure 6.23.: Methylation response of the MBD mediated repression of pBEST-mbd-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the CpG methylated mbd operon in pBEST-mbd-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the recombined MBD protein. In (A) the graph shows the eGFP expression as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500. The concentration of the reporter plasmid is 5 nM, and the amino acid concentration 0.5 mM. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The plasmid concentrations of pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 are shown in the legend (values in nM). The eGFP expression level decreases with increasing concentrations of MBD by a factor of 85% (for 1 nM of the MBD plasmid). **(C), (D)**, same as (A) and (B) but for the non-methylated mbd operon. The eGFP expression level decreases with increasing concentrations of MBD by a factor of 43% (for 1 nM of the MBD plasmid).

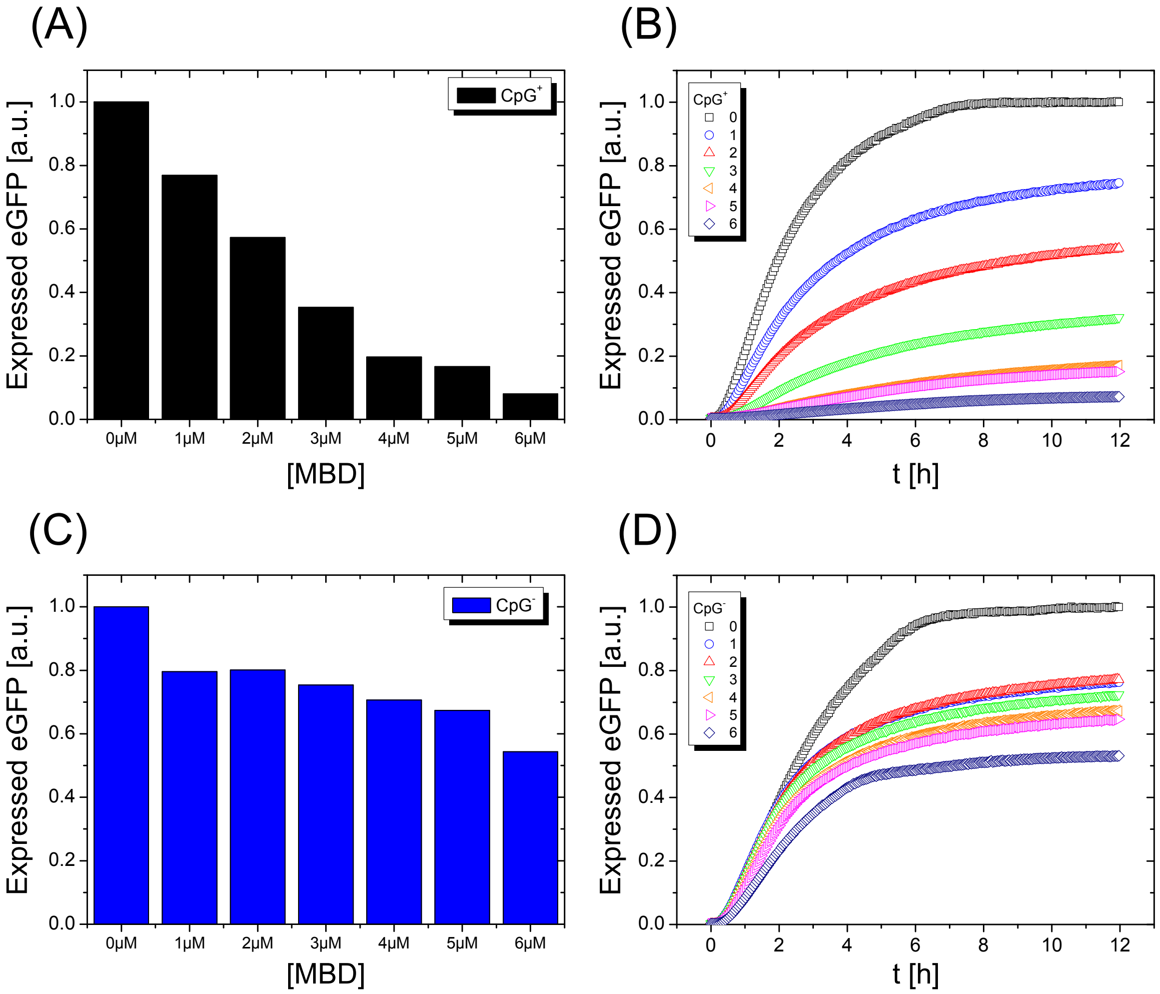


Figure 6.24.: Methylation response of the recombined MBD mediated repression of pBEST-mbd-UTR1-eGFP-T500. **(A), (B)**, equilibrium values (A) and kinetics (B) for the case of the CpG methylated mbd operon in pBEST-mbd-UTR1-eGFP-T500. All data is normalized to the respective equilibrium value of the eGFP expression without the addition of the recombined MBD protein. In (A) the graph shows the eGFP expression as a function of the concentration of the recombined MBD. The concentration of the reporter plasmid is 5 nM, and the amino acid concentration 0.5 mM. (B) shows the kinetics of the eGFP expression as a function of time under the same conditions as described in (A). The protein concentrations of MBD are given in the legend (values in µM). The eGFP expression level decreases with increasing concentrations of MBD by a factor of 92% (for 6 µM of MBD). **(C), (D)**, same as (A) and (B) but for the non-methylated mbd operon. The eGFP expression level decreases with increasing concentrations of MBD by a factor of 46% (for 6 µM of hMeCP2).

the end of the measurement. This suggests that it takes about two hours until the extract is enriched with recombined MBD proteins in an adequate concentration to bind to the mbd operon and consequently repress the system. During this time, eGFP is still expressed. On the other hand, when using the recombined version of MBD (compare Fig. 6.24 (D), for a protein concentration of 6 μM), the domain is capable of repressing the system from the beginning of the experiment. A pre-incubation time of the MBD plasmid within the cell free reaction for about two hours should in principle lead to an even more complete repression, which is similar to the results of the recombined protein. Put together, one can conclude that MBD specifically binds to the CpG methylated version of the mbd operon but also interacts with the non-methylated promoter. However, since the repression efficiency is particularly pronounced in the case of CpG methylated reporters, the mbd construct is adequate for using it as an epigenetic switch in gene circuits.

eGFP expression as a function of the amino acid concentration

By analogy with the experiments performed with the pap5 operon (section 6.1.5), in the following the expression efficiency of eGFP regulated by the mbd operon is measured as a function of the amino acid concentration in the cell free extract. This experiment is performed to determine the optimal amino acid concentration for the simultaneous expression of both, the reporter protein eGFP from pBEST-mbd-UTR1-eGFP-T500 and the MBD protein from pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 within one single batch mode reaction at typically employed plasmid concentrations. This is done in the following experiments for the cases of both, a CpG methylated and a non-methylated version of the reporter pBEST-mbd-UTR1-eGFP-T500.

Fig. 6.25 shows the respective results. By analogy with the results of the pap switch (compare Figs. 6.3 and 6.18) the experiment reveals an adequate amino acid concentration of 0.5 mM for expressing eGFP and MBD within the same cell free reaction. Even if the maximum expression efficiency for CpG methylated reporters is reached for an amino acid concentration of 1 mM, the respective expression level is only increased by a factor of 15% compared to the yield reached for an amino acid concentration of 0.5 mM. Thus, the concentration is adjusted to 0.5 mM in all subsequent experiments in order to make the results comparable with those of the pap switch (compare Fig. 6.18). For an amino acid concentration of 0.5 mM, a reporter plasmid concentration of 5 nM, and a pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 concentration of 4 nM, the difference between the CpG^{+} and CpG^{-} expression is determined to be about 80%.

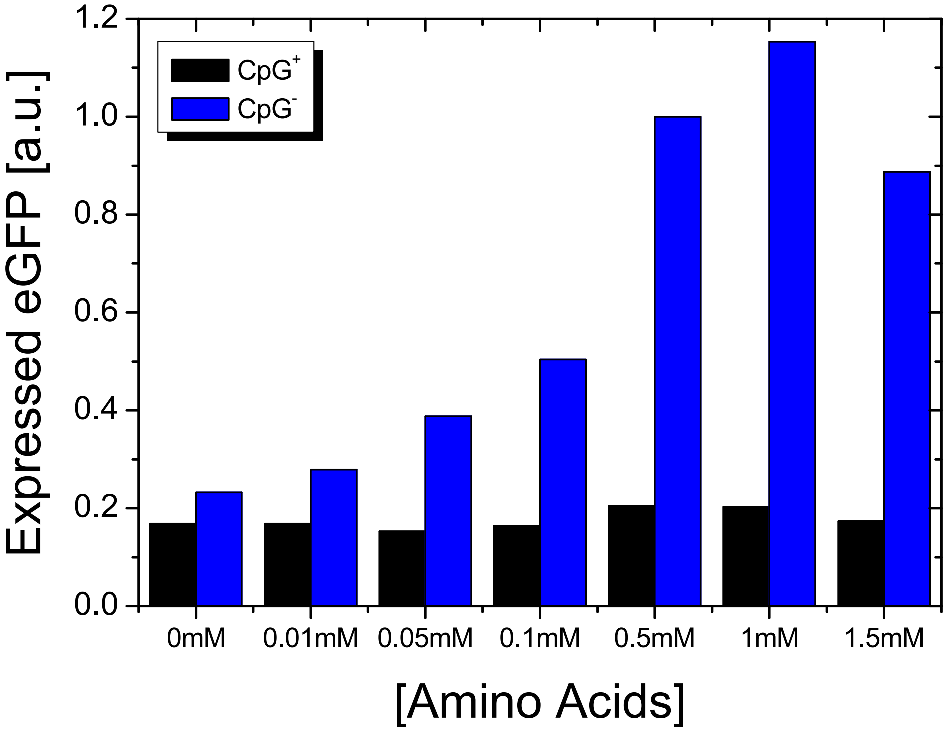


Figure 6.25.: Batch mode expression of pBEST-mbd-UTR1-eGFP-T500 as a function of the amino acid concentration in the cell free reaction. The graph shows the concentration of the recombined eGFP as a function of the amino acids in the case of the CpG methylated mbd operon (CpG$^+$, black) and the CpG non-methylated operon (CpG$^-$, blue). Data is normalized to the CpG$^+$ expression level reached for an amino acid concentration of 0.5 mM. The value on the x-axis refers to the concentration of each of the 20 native amino acids. The concentration of the reporter plasmid is 5 nM. Additionally, the reaction contains 4 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500, expressing the regulator MBD. For an amino acid concentration of 0.5 mM, the difference between CpG$^+$ and CpG$^-$ expression is determined to be 80%.

6.2.2. Sequence mutation of the mbd operon

In order to gain further insight into the MBD mediated repression of gene expression as a function of CpG methylation, in the following, four mutated versions of the mbd operon are the subject of interest. The sequences of the mutants are listed in Table 6.1. The operons mbdM1 and mbdM3 are designed to study the influence of mutating the flanking bases of the central CpG motif. mbdM2 breaks the run of AT bases adjacent to the CpG motif. This AT run is reported to be essential for the DNA binding selectivity of the MBD [27]. mbdM4 combines the motivations for mbdM2 and mbdM3 within

Table 6.1.: Overview of the mutated versions of the mbd operon studied in the context of this work. Only the top strand sequence of the regulatory DNAs is shown. All constructs contain the central CpG motif (red), which can either be methylated or non-methylated. The bases shown in blue indicate the sequence that is mutated compared to the native promoter III of the mouse-brain-derived neurodrophic factor (BDNF) gene (mbd-operon).

Operon	Sequence (5′ → 3′)
mbd	CTG-GAA-CGG-AAT-TCT-TTC
mbdM1	CTG-GAA-CGC-AAT-TCT-TTC
mbdM2	CTG-GAA-CGG-AGC-CCT-TTC
mbdM3	CTG-GGG-CGG-AAT-TCT-TTC
mbdM4	CTG-GGG-CGG-AGC-CCT-TTC

one single construct. The central CpG motif of the mbd operon, which can either be methylated or non-methylated, is conserved in all four mutants. All the operons listed in Table 6.1 are used as regulatory parts for the expression of the reporter eGFP in the plasmids pBETS-mbdM1-UTR1-eGFP-T500, pBETS-mbdM2-UTR1-eGFP-T500, pBETS-mbdM3-UTR1-eGFP-T500, and pBETS-mbdM4-UTR1-eGFP-T500.

MBD mediated repression of mbdM1-M4-eGFP

For all four mutants listed in Table 6.1, Fig. 6.26 shows the MBD mediated repression of eGFP as a function of the concentration of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500. For a MBD plasmid concentration of 1 nM, the experiment reveals a repression efficiency from 54% to 69% in the CpG^{+} case (mbdM1, 54%; mbdM2, 66%; mbdM3, 60%; and mbdM4, 69%), and from 24% to 31% in the CpG^{-} case (mbdM1, 28%; mbdM2, 31%; mbdM3, 24%; and mbdM4, 27%). Thus, the MBD mediated repression efficiency is comparable for all four mutants. Neither the mutation of the CpG flanking sites in mbdM1 and mbdM3, nor the breakup of the AT run in mbdM2 leads to a significant modification of the methylation specific repressor properties of MBD. The reduction of the eGFP expression level is comparable with the native mbd operon in the methylated and non-methylated case (compare Fig. 6.23 for the case of 1 nm of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500). This experimental finding is in agreement with the results of literature that MBD binds to CpG methylated DNA in any sequence context due to contacts between a hydrophobic patch on the MBD that interacts directly with the DNA methyl groups [24]. However, it is surprising that

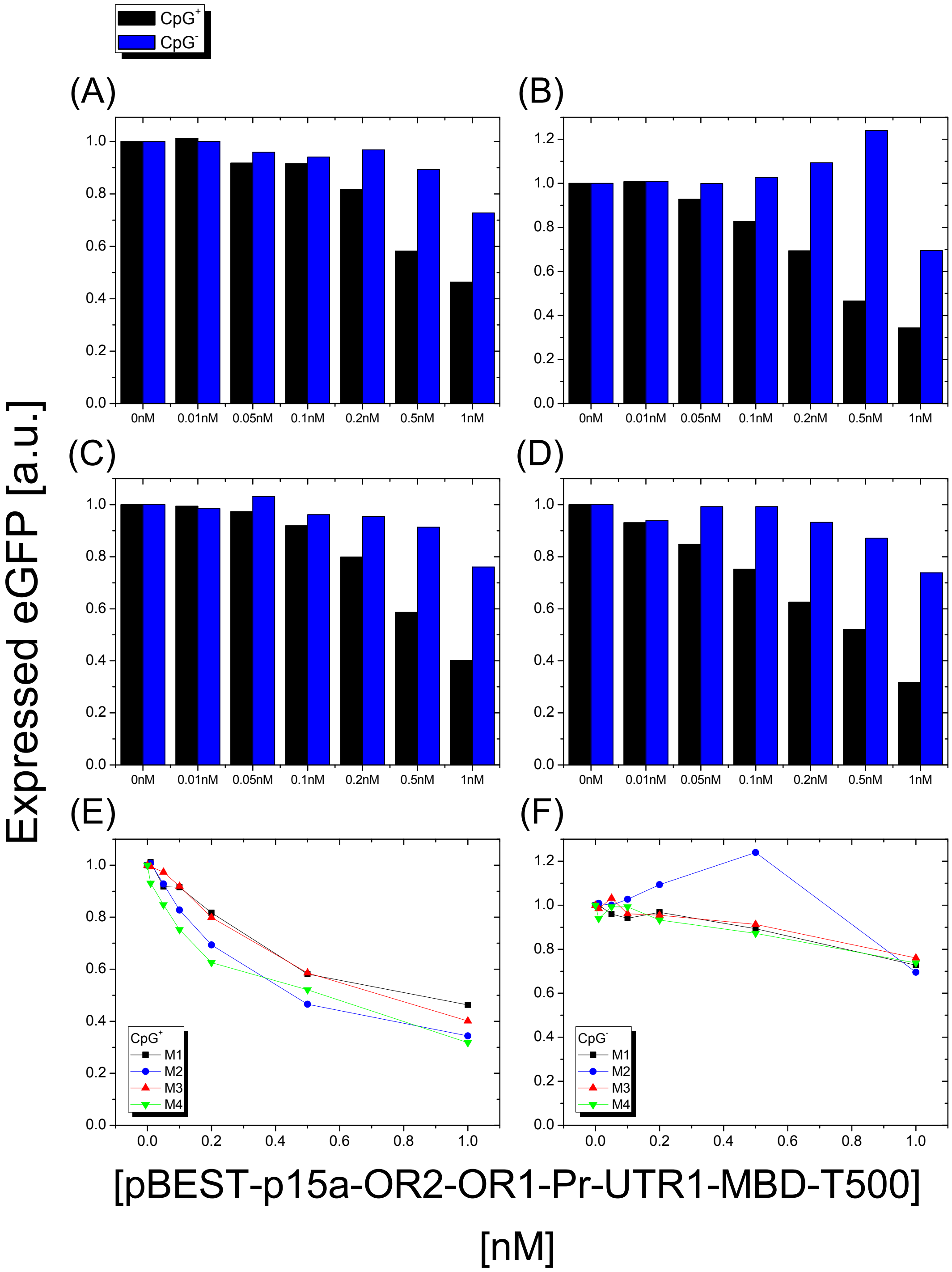


Figure 6.26.: Methylation response of the MBD mediated eGFP repression, using mutated versions of the mbd operon (Table 6.1), included in the reporter plasmids pBEST-mbdM1-UTR1-eGFP-T500 **(A)**, pBEST-mbdM2-UTR1-eGFP-T500 **(B)**, pBEST-mbdM3-UTR1-eGFP-T500 **(C)**, and pBEST-mbdM4-UTR1-eGFP-T500 **(D)**, in the case of CpG methylated (black) and non-methylated DNA (blue). All data is norma-

lized to the respective equilibrium value of the eGFP expression without the addition of the MBD plasmid. The graphs show the eGFP expression as a function of the concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500. The concentration of the reporter plasmids are 5 nM, and the amino acid concentration is 0.5 mM in each case. **(E), (F)**, eGFP expression level for all mutants as a function of the MBD plasmid concentration for the CpG$^+$ (E) and the CpG$^-$ (F) case. For 1 nM of the MBD plasmid, the eGFP level decreases by factors of 54% (mbdM1), 66% (mbdM2), 60% (mbdM3) and 69% (mbdM4) in the CpG$^+$ case and by factors of 28% (mbdM1), 31% (mbdM2), 24% (mbdM3) and 27% (mbdM4) in the CpG$^-$ case.

breaking up the AT run adjacent to the central CpG motif (mbdM3) does not lead to a more important modification of the MBD mediated repression efficiency as a function of the methylation state of the mbdM3 operon compared to the native regulatory part [27]. On the other hand, it is observed that increasing the MBD plasmid concentrations up to 0.5 nM leads to an increased eGFP expression level in the case of the CpG methylated mbdM3 operon, which is not observable for the non-methylated DNA. The reason for this activator-like behavior of MBD in the absence of the AT run is currently unknown. Not till the plasmid concentration of pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 exceeds the value of 0.5 nM, the eGFP level is repressed compared to the MBD free case.

6.3. Two-state gene switching based on MBD and PapI-Lrp

For the purpose of constructing an epigenetic two state switch, in the following, the properties of the pap5 operon (section 6.1.5) with its Dam methylation dependent regulators PapI and Lrp are merged with the properties of the mbd operon (section 6.2.1) with its CpG methylation dependent regulator MBD. The pap5 operon regulates the expression of the reporter eGFP, using the plasmid pBEST-pap5-UTR1-eGFP-T500. The mbd operon regulates the expression of the reporter mCherry, using the plasmid pBEST-mbd-UTR1-mCherry-T500. The Dam dependent regulatory proteins PapI and Lrp are expressed from the plasmids pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 and pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500, respectively. The CpG methylation dependent regulator MBD is added to the reaction in its recombined form in order to save resources of the expression machinery. The objective of the experiment is the methylation dependent switching between the logical states ON and OFF by adding

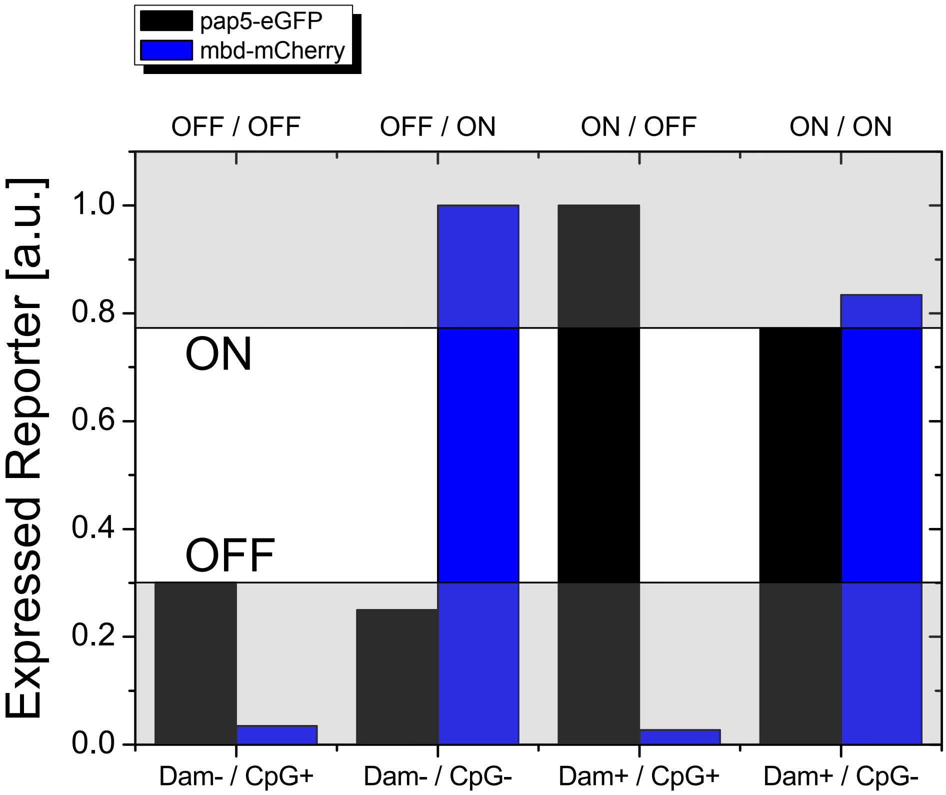


Figure 6.27.: Epigenetic two state switch based on Dam and CpG methylation of the operons pap5 and mbd, and the regulator proteins PapI, Lrp, and MBD. The graph shows the expression levels of eGFP and mCherry. The bottom axis shows the methylation state of the pap5 and mbd operon. Here, Dam^+ and Dam^- refer to the Dam methylated and non-methylated pap5 operon, while CpG^+ and CpG^- refer to the CpG methylated and non-methylated mbd operon, respectively. The top axis shows the corresponding logical states "ON" and "OFF". The eGFP expression level is normalized to the data obtained for the logical state ON/OFF. The mCherry expression level is normalized to the data obtained for the logical state OFF/ON. All reactions contain 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500, 1 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500, 6 µM of the recombined MBD, and 0.5 mM of amino acids. The concentrations of Dam, CpG, and non-methylated reporter plasmids pBEST-pap5-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500 are 5 nM in each case.

the respective Dam methylated, CpG methylated, or non-methylated reporter plasmids in the presence of all regulator proteins PapI, Lrp, and MBD.

The results are shown in Fig. 6.27. The experiment reveals successful switching between the logical states ON and OFF for both, the pap5-eGFP expression and the mbd-mCherry expression.

The simultaneous expression of both reporters, eGFP and mCherry, within one single reaction tube might lead to an unwanted, radiationless energy transfer between eGFP (donor) and mCherry (acceptor) based on FRET (compare chapter 3.2.4). By considering Eq. (3.1) and a Förster radius of 51 Å [38], the distance between the reporter proteins, at which the FRET efficiency is reduced to 10%, is determined to be about 6 nm. Due to the repulsive interaction, it is unlikely that the proteins can overcome this distance during a standard batch mode reaction. Additionally, the study of the energy transfer between eGFP and mCherry requires fusion of the two proteins by a short polypeptide chain of only a few amino acids in length [38]. Both aspects point towards a negligible contribution of FRET during a batch mode reaction when expressing both, eGFP and mCherry, at the same time. In order to definitely rule out FRET based artifacts, the fluorescence signal of eGFP and mCherry, simultaneously expressed from plasmids pBEST-pap5-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500 within one single batch mode reaction is compared with the respective fluorescence signal of an individual reporter expression in the absence of the other. The result is presented in Fig. 6.28. It confirms that FRET can be neglected during batch mode reactions involving simultaneous expression of both reporter proteins. The observed reduction of the fluorescence signals does not exceed a factor of about 25%.

6.4. Conclusion and discussion

The present chapter deals with the *in vitro* study of the transcriptional regulation of epigenetic switches. An E. coli based cell free expression system is used (see chapter 3.2.4 and reference [35]). This cytoplasmic extract contains the relevant molecular machinery for protein expression within a test tube and was developed at the institute of biophysics at the University of Minnesota. It constitutes one of the most efficient *in vitro* expression systems. A large amount of recombinant proteins can be synthesized in a few hours under highly controlled experimental conditions [35]. Compared to *in vivo* methods, the system provides highly accessible expression data without the need of complex data evaluation. The present study involves cloning of different regulatory DNA operons into circular plasmids, using methods of molecular biology. These operons regulate the measurable expression of a fluorescent reporter protein within a test tube. The energy for the expression is provided by a complex biochemical reaction buffer based on PGA (see appendix section B.3). DNA binding proteins, simultaneously expressed within the cell free system, can bind to the template DNA and consequently activate or hinder the expression of the reporter (compare Fig. 3.13). The first objective

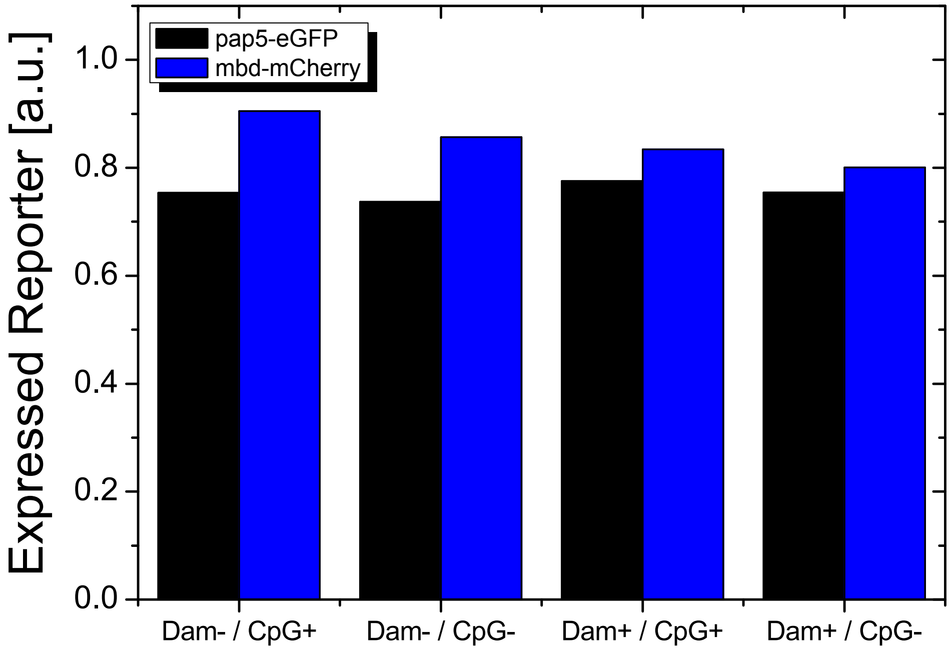


Figure 6.28.: Simultaneous expression of pap5-eGFP (black) and mbd-mCherry within one single batch mode reaction. The graph shows the eGFP and mCherry expression level. Dam^+ and Dam^- refers to the Dam methylated and non-methylated pap5 operon, respectively, while CpG^+ and CpG^- refers to the CpG methylated and non-methylated mbd operon, respectively. All data is normalized to the respective individual expression level in the absence of the respective second reporter. The concentration of the reporter plasmids pBEST-pap5-UTR1-eGFP-T500 and pBEST-mbd-UTR1-mCherry-T500 is 5 nM. The amino acid concentration is 0.5 mM. Compared to the individual case, the simultaneous expression of both reporters leads to a reduced reporter fluorescence by factors of 27% (Dam^-/CpG^-, eGFP channel) to 10% (Dam^-/CpG^+, mCherry channel).

of the work was the establishment of the cell free expression system at the institute of biological experimental physics at Saarland University. Due to the complex protocol for the preparation of the crude extract (see appendix section B.5), which involves a diversity of biochemical approaches, this has turned out to be very challenging. In the context of the present work, the cell free extract is used for the first time to investigate epigenetically regulated transcription systems. Here, particular proteins are allowed to bind to the template DNA as a function of its methylation state and consequently change the expression efficiency of the reporter.

The first part of the work deals with the procaryotic epigenetic switch based on Dam methylation (see section 6.1). The extract enables the study of the native E. coli DNA binding proteins Lrp and PapI and how they regulate gene expression on the transcriptional level. Changing the methylation state of the plasmid DNA often led to very high mutation rates and thus complicated the bacterial amplification. To overcome this problem, the promoter strength of some operons is reduced by changing the -35 and/or -10 consensus sequence for binding of the sigma factor. The different repression efficiencies of the different pap operons cannot be explained by the simple methylation dependent annealing of PapI and Lrp to these operons and the resulting blocking of the RNA polymerase. In this sense, one of the most important findings of this study is that an increased affinity of the regulator proteins PapI and Lrp for binding to DNA does not necessarily mean an increased repression efficiency of the system. The cooperativity of Lrp for binding to sites 1, 2, and 3 (pap123 operon, section 6.1.1) as well as sites 4, 5, and 6 (pap456 operon, section 6.1.4) is different than previously thought [17]. The methylation response of the Lrp mediated repression of the reporter is completely different in the cases of the pap2 (section 6.1.2) and the pap5 operon (section 6.1.5), although Lrp is reported to possess comparable affinity for binding to both operons *in vitro* [17]. Trippling of Lrp binding site 2 or 5 leads to a completely different methylation response of the repression efficiency compared to the case when using only one of these binding sites (see sections 6.1.2 and 6.1.3, as well as 6.1.5 and 6.1.6). The best Dam methylation dependent switching between the logical states ON and OFF is found for the pap5 operon (see section 6.1.5).

The second part of the work deals with the eucaryotic epigenetic switch based on CpG methylation (see section 6.2). Here, the methylation binding domain (MBD) of the human repressor MeCP2 regulates the reporter expression as a function of the methylation state of the DNA. For the first time it is achieved to study an eucaryotic epigenetic regulation system with a cytoplasmic extract that is totally based on a procaryotic expression machinery. The results reveal a significantly improved repression efficiency of the methylation binding domain when using a CpG methylated version of the native promoter III of the mouse-brain-derived neurodrophic factor gene (mbd operon). The results reveal an efficient CpG methylation dependent switching between the logical expression states ON and OFF (see section 6.2.1). Mutating the CpG flanking base pairs and adjacent AT runs of this promoter does not lead to a significant modification of the repression efficiency, although these mutations were reported to essentially change the binding affinity of MeCP2 *in vitro* [25–27].

In the last part of the study the procaryotic epigenetic switch based on Dam methylation and the regulators Lrp and PapI is merged with the eucaryotic switch based on CpG methylation and the regulator MBD. The two reporter proteins eGFP and mCherry, as well as all regulator proteins are present in one test tube to produce an epigenetically regulated two-state switch (see section 6.3). Although the two regulation systems stem from different organisms, the genes can efficiently by turned ON and OFF by changing the methylation state of the template DNA.

7. Summary

The present thesis deals with the experimental and numerical investigation of cooperative molecular binding and recognition of DNA and proteins. The main focus of the work lies on the competitive hybridization of single stranded DNA molecules (chapter 4), the influence of DNA methylation on the stability of double-stranded DNA (chapter 5), and the epigenetic regulation of *in vitro* gene switches (chapter 6).

The first part of the present thesis deals with the experimentally observed ultrahigh specificity of a single-stranded DNA molecule (target A) during competition with a second strand with slight sequence variations (target B) for binding to a complementary molecule (probe) in thermodynamic equilibrium. In spite of comparable binding affinities K^A and K^B as well as an excess concentration of B over A by at least two orders of magnitude, solely target A binds to the probe and discriminates the competing species B (chapter 4.1). The effect was firstly observed by Timo Mai in the context of an ongoing effort of the institute of biological experimental physics at Saarland University to understand DNA hybridization. The optical methods total internal reflection fluorescence (TIRF, chapter 3.1.5), Förster resonance energy transfer (FRET, chapter 3.1.6), and fluorescence correlation spectroscopy (FCS, chapter 3.1.7) are used to detect hybridization and base pair fluctuation dynamics of the fluorescently labeled DNAs. The effect of the high specificity emerges independently from the DNA sequence motif (chapter 4.2.1) and is observable over a broad ion concentration range of the hybridization buffer (chapter 4.2.3). High specificity can be observed in surface based measurements and also in bulk (chapter 4.2.4). Thermodynamic equilibrium conditions are verified for all experiments (chapters 4.2.5) and consequently non-equilibrium effects are excluded. The experimental results point towards a molecular interaction between the competitors A and B as the reason for the observed high specificity (chapters 4.3.3, 4.3.8, 4.3.9). The temperature dependence (chapters 4.3.5, 4.3.6 and reference [30]) underlines that different thermal fluctuations (chapter 4.3.9), involving different entropic degrees of freedom of the competing strands (chapters 4.3.4, 4.3.10) play an essential role for the occurrence of the effect. Put together, one can conclude from the experimental results that in the highly specific cases, the competing strands interact via thermal fluctuations and the effective binding free energies change in such a way that solely target A is able to bind to the probe molecules by reducing the binding affinity of target B (section 4.3.8). Target A is not affected by the molecular competition with target B (chapter 4.3.7), which underlines that the interaction has to be asymmetric. Otherwise, it would be impossible to tilt the free energy landscape in favor of one of the two competitors. Based on the experimental findings, a numerical assessment is developed with the purpose of modeling the observed high specificity in thermodynamic equilibrium with a reasonable choice of parameters (chapter 4.4). This has been done in strong collaboration with Christian Trapp [36]. The detailed numerical assessment reveals that both, target A and B bind to the probe in a mixture of pre-melted and helicoidal binding conformations. Target B generates its affinity through a large number of short helicoidal conformations. The increased affinity of target A stems from additional, almost completely closed conformations that target B cannot produce (Fig. 4.26). During competition, target A and B form a triplex structure with the probe molecule. In this case, the entropic cost of a binding conformation is augmented, since a particular target will be stretched by the presence of the competitor. Due to its non-matching base, target B cannot reach these highly cooperative states and consequently looses the major part of its binding free energy (Fig. 4.27). This leads to a sharp, asymmetric modification of the binding free energy landscape in favor of target A (Fig. 4.28).

The second study of the present thesis opens a new subject about the influence of cytosine methylation on the stability of double stranded DNA. It is shown that methylation of only two out of 16 bases of a target molecule produces an order of magnitude increased affinity for binding of that target to the probe (chapter 5.1). This is also underlined by the shift of the melting temperature of a methylated double strand compared to its non-methylated analogon (chapter 5.2). Comparing the melting profile of hydroxymethylated DNA with the respective profile of non-methylated DNA (Fig. 5.2) points towards the existence of additional, pre-melted binding states, which are a direct consequence of the hydroxymethylation.

The third topic of the present thesis deals with the epigenetic regulation of *in vitro* gene expression. Therefore, an E. coli based cell free expression system is used (see chapter 3.2.4 and reference [35]). The cytoplasmic extract contains the relevant molecular machinery for the expression of proteins from a DNA template and was developed at the institute of biophysics at the University of Minnesota. It constitutes one of the most efficient *in vitro* expression systems. A large amount of recombinant proteins can be synthesized in a few hours under highly controlled experimental conditions [35]. The study involves cloning of different regulatory DNA operons into circular plasmids, using methods of molecular biology. In the context of the present work, the cell free extract is used for the first time to investigate epigenetically regulated transcription systems. Here, particular methylation dependent binding proteins, simultaneously expressed in the cell free system, can bind to the mentioned DNA operons and consequently activate or hinder the measurable expression of a reporter protein within a test tube (chapter 3.2.4). Two regulation systems are studied: i) a procaryotic epigenetic switch based on Dam methylation and the DNA binding proteins Lrp and PapI (chapter 6.1), and ii) an eucaryotic epigenetic switch based on CpG methylation and the methylation binding domain (MBD) of the human repressor MeCP2 (chapter 6.2). In the case of i), the experiments reveal that an increased binding affinity of the regulator proteins PapI and Lrp for binding to DNA does not necessarily mean an increased repression efficiency of the system. The different repression efficiencies cannot be explained by the simple methylation dependent annealing of PapI and Lrp to DNA and the resulting blocking of the RNA polymerase. The cooperativity of Lrp for binding to the different regulatory DNAs, as well as the methylation response of this binding are different than previously thought (chapters 6.1.1, 6.1.2, 6.1.4, 6.1.5). In the case of ii), it is achieved for the first time to study an eucaryotic epigenetic regulation system with a cytoplasmic extract that is totally based on a procaryotic expression machinery. The results reveal an efficient switching between the logical expression states ON and OFF, as a function of the methylation state of the regulatory DNA (chapter 6.2.1). Mutating the CpG flanking base pairs and adjacent AT runs of the DNA does not lead to a significant modification of the repression efficiency, although these mutations were reported to essentially change the binding affinity of MeCP2 *in vitro* (chapter 6.2.2). In a final experiment, both epigenetic regulation systems i) and ii) are merged in order to produce an epigenetic two state switch based on Dam and CpG methylation (chapter 6.3). Although the two regulation systems stem from different organisms, the genes can efficiently by turned ON and OFF by changing the methylation state of the template DNA.

Put together, the results of the present work underline that evolution produces biological molecular recognizers as an ensemble so that they enhance each other's function. In the case of the high specificity of competitive DNA hybridization, the environment of several similar strands competing for the same partner adjusts the free energy landscape in such a way that oligonucleotides constitute a molecular "ONLY IF (best matching binding partner not available) THEN (bind to second best match)". This happens under thermodynamic equilibrium conditions, at the only energetic cost of creating the molecular environment. In the case of the epigenetic regulation of gene expression, the experimental findings show that the transcription machinery is much more collective and complex than previously thought. The experimental results may contribute to understand why the properties of a system studied *in vivo* are often no longer observable *in vitro*, when the relevant molecular players are decoupled from their complex molecular environment. Self-organizing objects could react to the presence of other substances in complex, predictable ways without the need for classical, molecular non-equilibrium thermodynamic machines.

Appendix

A. Appendix: DNA hybridization

A.1. DNA oligonucleotides

Table A.1 gives an overview of the DNA sequences and modifications of all oligonucleotides used in the context of this work for the study of competitive DNA hybridization (chaper 4) as well as the influence of DNA methylation on the DNA duplex stability (chapter 5).

A.2. Derivation of the evanescent field

The complete derivation of the evanescent field is adapted from [53] and also presented in [30]. Considering Maxwell's equations of electrodynamics

$$\begin{aligned}
\operatorname{rot}\vec{E}(\vec{r},t) &= -\frac{\partial\vec{B}(\vec{r},t)}{\partial t} && \text{(A.1)}\\
\operatorname{rot}\vec{H}(\vec{r},t) &= \frac{\partial\vec{D}(\vec{r},t)}{\partial t}+\vec{j}(\vec{r},t) && \text{(A.2)}\\
\operatorname{div}\vec{D}(\vec{r},t) &= \rho(\vec{r},t) && \text{(A.3)}\\
\operatorname{div}\vec{B}(\vec{r},t) &= 0 && \text{(A.4)}
\end{aligned}$$

where $\vec{E}(\vec{r},t)$ is the electric field, $\vec{H}(\vec{r},t)$ is the magnetic field, $\vec{D}(\vec{r},t)$ is the dielectric displacement, $\vec{B}(\vec{r},t)$ is the magnetic induction, $\vec{j}(\vec{r},t)$ is the current density, and $\rho(\vec{r},t)$ is the charge density. The optical materials used in the TIRF experiment are linear. Thus, $\vec{D}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$ can be expressed by using the following equations

$$\begin{aligned}
\vec{D}(\vec{r},t) &= \epsilon_0\epsilon_r(\vec{r})\cdot\vec{E}(\vec{r},t) && \text{(A.5)}\\
\vec{B}(\vec{r},t) &= \mu_0\mu_r(\vec{r})\cdot\vec{H}(\vec{r},t) && \text{(A.6)}
\end{aligned}$$

Table A.1.: Overview of all DNA oligonucleotides and sequences used for the investigation of DNA hybridization. The immobilized sequences IS and IS* are modified with a C6 amino linker at the 5′ end in order to allow covalent binding to the dendrimer coated surfaces (see chapter 3.1.1). A spacer of 15 thymine bases increases the distance to the surface. There is also a version of IS that is modified with the fluorescent dye cyanine (Cy) for measurements in bulk. All target molecules are modified with Cy at the 5′ end, except for the unlabeled and the internally tagged versions. Non-complementary bases are shown underlined and in red (e. g. $\underline{\text{A}}$). The targets PM$^{\text{m}}$ and MM1$^{\text{m}}$ contain two bases of methylcytosine (C_{CH_3}) and the target MM1$^{\text{hm}}$ contains two bases of hydroxymethylcytosine (C_{OCH_3}). The LNA targets MM1-LNA-1 and MM1-LNA-2 contain different numbers of LNA bases (e.g. G). They are provided with a special fluorescent dye (ExiqonDye).

Probe	**Sequence (5′ → 3′)**
IS	NH2-C6-15T-TTA-CGA-TCT-GAT-CCT-T
IS (for bulk measurement)	TTA-CGA-TCT-GAT-CCT-T-TTT-Cy
IS*	NH2-C6-15T-GTA-CTA-TTG-CTG-CCC

Target	**Sequence (5′ → 3′)**
PM (unlabeled)	AAG-GAT-CAG-ATC-GTA-A
PM	Cy-AAG-GAT-CAG-ATC-GTA-A
PM (internally tagged)	AAG-GAT-C-Cy-AG-ATC-GTA-A
PM$^{\text{m}}$	Cy-AAG-GAT-C_{CH_3}AG-ATC_{CH_3}-GTA-A
MM1 (unlabeled)	AAG-GAT-CA$\underline{\text{C}}$-ATC-GTA-A
MM1	Cy-AAG-GAT-CA$\underline{\text{C}}$-ATC-GTA-A
MM1 (internally tagged)	AAG-GAT-C-Cy-A$\underline{\text{C}}$-ATC-GTA-A
MM1$^{\text{m}}$	Cy-AAG-GAT-C_{CH_3}A$\underline{\text{C}}$-ATC_{CH_3}-GTA-A
MM1$^{\text{hm}}$	Cy-AAG-GAT-C_{OCH_3}A$\underline{\text{C}}$-ATC_{OCH_3}-GTA-A
MM1-LNA-1	ExiqonDye-AAG-GAT-CA$\underline{\text{C}}$-ATC-GTA-A
MM1-LNA-2	ExiqonDye-AAG-GAT-CA$\underline{\text{C}}$-ATC-GTA-A
MM2 (unlabeled)	AAG-GAT-CAG-ATC-G$\underline{\text{C}}$A-A
MM2	Cy-AAG-GAT-CAG-ATC-G$\underline{\text{C}}$A-A
MM3	Cy-AAG-GAT-C$\underline{\text{TC}}$-ATC-GTA-A
MM4	Cy-AA$\underline{\text{A}}$-GAT-CAG-ATC-G$\underline{\text{A}}$A-A
MM5	Cy-AA$\underline{\text{C}}$-GAT-CAG-ATC-GTA-A
MM6	Cy-AAG-$\underline{\text{C}}$AT-CAG-ATC-GTA-A
MM7	Cy-AAG-G$\underline{\text{T}}$T-CAG-ATC-GTA-A
MM8	Cy-AAG-GA$\underline{\text{A}}$-CAG-ATC-GTA-A
MM9	Cy-AAG-GAT-$\underline{\text{G}}$AG-ATC-GTA-A
MM10	Cy-AAG-GAT-C$\underline{\text{T}}$G-ATC-GTA-A
MM11	Cy-AAG-GAT-CAG-$\underline{\text{T}}$TC-GTA-A
MM12	Cy-AAG-GAT-CAG-A$\underline{\text{A}}$C-GTA-A
MM13	Cy-AAG-GAT-CAG-AT$\underline{\text{G}}$-GTA-A
MM14	Cy-AAG-GAT-CAG-ATC-$\underline{\text{C}}$TA-A
PM*	Cy-GGG-CAG-CAA-TAG-TAC
MM*	Cy-GGG-CAG-C$\underline{\text{TT}}$-TAG-TAC

where $\epsilon_r(\vec{r})$ and $\mu_r(\vec{r})$ are the electrical and magnetic permeability, which both depend on the used optical materials. All materials can be considered to be free of electrical charges and electrical currents:

$$\rho(\vec{r},t) = \vec{j}(\vec{r},t) \equiv 0 \tag{A.7}$$

Furthermore, all used materials are dielectric, so that $\mu_r(\vec{r}) = 1$. After the definition of the refraction index

$$n = \sqrt{\epsilon_r(\vec{r})\mu_r(\vec{r})} \tag{A.8}$$

which is considered to be independent of $\vec{r}$ in the following, Maxwell's equations simplify to

$$\begin{aligned} \operatorname{rot} \vec{E}(\vec{r},t) &= -\mu_0 \frac{\partial \vec{H}(\vec{r},t)}{\partial t} && \text{(A.9)} \\ \operatorname{rot} \vec{H}(\vec{r},t) &= \epsilon_0 n^2 \frac{\partial \vec{E}(\vec{r},t)}{\partial t} && \text{(A.10)} \\ n^2 \operatorname{div} \vec{E}(\vec{r},t) &= 0 && \text{(A.11)} \\ \operatorname{div} \vec{H}(\vec{r},t) &= 0 && \text{(A.12)} \end{aligned}$$

With Eq. (A.9) and Eq. (A.10) the magnetic field component can be eliminated

$$\begin{aligned} \operatorname{rot}\operatorname{rot} \vec{E}(\vec{r},t) &= -\mu_0 \frac{\partial \operatorname{rot} \vec{H}(\vec{r},t)}{\partial t} \\ &= -\mu_0 \epsilon_0 n^2 \frac{\partial^2 \vec{E}(\vec{r},t)}{\partial t^2} && \text{(A.13)} \end{aligned}$$

The vector identity

$$\operatorname{rot}\operatorname{rot} \vec{A} = \operatorname{grad}\operatorname{div} \vec{A} - \Delta \vec{A} \tag{A.14}$$

as well as Eqs. (A.11) and (A.13) then reveal the electrical component of the wave equation:

$$\left(\Delta - \mu_0 \epsilon_0 n^2 \frac{\partial^2}{\partial t^2} \right) \vec{E}(\vec{r},t) = 0 \tag{A.15}$$

To eliminate the time dependence, the following approach is used:

$$\vec{E}(\vec{r},t) = \vec{\tilde{E}}(\vec{r}) \cdot e^{\pm i\omega t} \tag{A.16}$$

Adopting Eq. (A.16) in Eq. (A.15) reveals

$$\Delta \vec{\tilde{E}}(\vec{r}) + \mu_0 \epsilon_0 \omega^2 n^2 \vec{\tilde{E}}(\vec{r}) = 0 \tag{A.17}$$

In the following, the electromagnetic fields caused by total reflection are described quantitatively. Therefore, a substrate with refraction index n_1 and a surrounding medium with refraction index n_2 are considered. Thus, the following equation have to be solved for both medias

$$\begin{aligned} \Delta\vec{\tilde{E}}(\vec{r}) + \mu_0\epsilon_0\omega^2 n_1^2\vec{\tilde{E}}(\vec{r}) &= 0 \qquad \text{(A.18)} \\ \Delta\vec{\tilde{E}}(\vec{r}) + \mu_0\epsilon_0\omega^2 n_2^2\vec{\tilde{E}}(\vec{r}) &= 0 \qquad \text{(A.19)} \end{aligned}$$

The x-z plane is considered as the plane of incidence. For the amplitude $\vec{\tilde{E}}(\vec{r})$ in Eq. (A.16) one has to consider the approach of a plane wave

$$\vec{\tilde{E}}(\vec{r}) = \vec{E}_0 \cdot e^{i(\vec{k}\cdot\vec{r})} = \vec{E}_0 \cdot e^{i(k_x\cdot x + k_z\cdot z)} \tag{A.20}$$

with

$$\begin{aligned} \vec{r} &= (x, y, z) = (x, 0, z) \qquad \text{(A.21)} \\ \vec{k} &= (k_x, k_y, k_z) = \left|\vec{k}\right| (\sin(\Theta_1), 0, \cos(\Theta_1)) \qquad \text{(A.22)} \end{aligned}$$

For $\vec{k} = |k|$ the approach (A.20) reveals the dispersion relation for the substrate and the surrounding media

$$\begin{aligned} k^2 &= \epsilon_0\mu_0 \cdot n_1^2 \cdot \omega^2 \qquad \text{(A.23)} \\ k^2 &= \epsilon_0\mu_0 \cdot n_2^2 \cdot \omega^2 \qquad \text{(A.24)} \end{aligned}$$

With the relation $c = \sqrt{\epsilon_0\mu_0}$ as well as $\omega = 2\pi\lambda c^{-1}$ the dispersion relation reveals the expressions of k for substrate and surrounding media

$$\begin{aligned} k_1 &= \frac{2\cdot\pi\cdot n_1}{\lambda_0} \qquad \text{(A.25)} \\ k_2 &= \frac{2\cdot\pi\cdot n_2}{\lambda_0} \qquad \text{(A.26)} \end{aligned}$$

where λ_0 is the wavelength of the incident wave. Through adopting the expressions (A.21), (A.22), (A.25) and (A.26) in the approach (A.20), one reveals the expressions for the respective transversal electrical and transversal magnetic polarization of the incident, the transmitted, and the reflected wave

$$\vec{\tilde{E}}_0(\vec{r}) = \vec{E}_0 \cdot \exp\left\{\left(-i\frac{2\pi n_1}{\lambda_0}\right) \cdot (x\cdot\sin(\Theta_1) + z\cdot\cos(\Theta_1))\right\} \tag{A.27}$$

$$\vec{\tilde{E}}_r(\vec{r}) = R \cdot \vec{E}_0 \cdot \exp\left\{\left(-i\frac{2\pi n_1}{\lambda_0}\right) \cdot (x \cdot \sin(\Theta_1) - z \cdot \cos(\Theta_1))\right\} \tag{A.28}$$

$$\vec{\tilde{E}}_t(\vec{r}) = T \cdot \vec{E}_0 \cdot \exp\left\{\left(-i\frac{2\pi n_2}{\lambda_0}\right) \cdot (x \cdot \sin(\Theta_2) + z \cdot \cos(\Theta_2))\right\} \tag{A.29}$$

The expressions for the reflection coefficient R and the transmission coefficient T can be quantified with Fresnell's equations

$$R_{TM} = \frac{n_2 \cdot \cos(\Theta_1) - n_1 \cdot \cos(\Theta_2)}{n_2 \cdot \cos(\Theta_1) + n_1 \cdot \cos(\Theta_2)} \tag{A.30}$$

$$R_{TE} = \frac{n_1 \cdot \cos(\Theta_1) - n_2 \cdot \cos(\Theta_2)}{n_1 \cdot \cos(\Theta_1) + n_2 \cdot \cos(\Theta_2)} \tag{A.31}$$

$$T_{TM} = \frac{2n_1 \cdot \cos(\Theta_1)}{n_2 \cdot \cos(\Theta_1) + n_1 \cdot \cos(\Theta_2)} \tag{A.32}$$

$$T_{TE} = \frac{2n_1 \cdot \cos(\Theta_1)}{n_1 \cdot \cos(\Theta_1) + n_2 \cdot \cos(\Theta_2)} \tag{A.33}$$

The correlation between the angle of incident and transmitted wave can be determined with Snell's law

$$\frac{\sin(\Theta_1)}{\sin(\Theta_2)} = \frac{n_2}{n_1} \tag{A.34}$$

Total reflection can be observed for $\sin(\Theta_1) \geq n_2/n_1$. This means that the total energy of the incident wave is reflected. Additionally, an evanescent wave travels in z-direction at the point of total reflection. Considering the expression $\sin^2(\Theta) + \cos^2(\Theta) = 1$, the limiting angle for the occurrence of total reflection ($\sin(\Theta_g) = n_2/n_1$), and Eq. (A.34) one reveals

$$\cos(\Theta_2) = \sqrt{1 - sin^2(\Theta_2)} = \sqrt{1 - \left(\frac{n_1^2}{n_2^2}\right) \cdot \sin^2(\Theta_2)} = \sqrt{1 - \left(\frac{\sin(\Theta_1)}{\sin(\Theta_g)}\right)^2} \tag{A.35}$$

Adopting Eq. (A.34) and (A.35) in Eq. (A.29) reveals the expression for the transmitted wave

$$\vec{\tilde{E}}_t(\vec{r}) = T \cdot \vec{E}_0 \cdot e^{-i \cdot x \cdot \frac{2\pi n_2}{\lambda_0} \cdot \sin(\Theta_1)} \cdot e^{-z \cdot \frac{2\pi n_2}{\lambda_0} \cdot \sqrt{\left(\frac{n_1^2}{n_2^2}\right) \cdot \sin^2(\Theta_1) - 1}} \tag{A.36}$$

The measurable intensity of an electromagnetic wave is determined with the timely weighted average of the pointing vector

$$I = \left|\left\langle \vec{S} \right\rangle\right| = \left(\frac{c}{8\pi}\right) \cdot \left|\Re e[\vec{E} \times \vec{H}]\right| \tag{A.37}$$

One can show that the timely weighted average of the pointing vector $\vec{S}_z$ equals 0 for the z-component. This means that in a mean field picture no energy is transferred into the medium of lower optical density. However, Eq. (A.37) constitutes a stationary state. In the beginning of an experiment, there indeed is an energy flow into the medium of lower optical density that causes an electromagnetic field. During the duration of the experiment, the energy in both medias does not reach detailed balance. This explains the seeming conflict that in spite of a totally reflected laser beam an evanescent wave can penetrate into the medium of lower optical density.
In the case of polarized waves the intensities in Eq. (A.37) can be expressed without the magnetic field component

$$I = \left(\frac{c}{8\pi}\right) \cdot \left|\Re e[\vec{E} \times \vec{E^*}]\right| \tag{A.38}$$

Considering the equations (A.27), (A.28), and (A.36) the intensities of incident, reflected and transmitted wave can be determined to be

$$\begin{aligned} I_0 &= n_1 \cdot E_0^2 && \text{(A.39)} \\ I_r &= |R|^2 \cdot n_1 \cdot E_0^2 && \text{(A.40)} \\ I_t &= |T|^2 \cdot n_2 \cdot E_0^2 \cdot \exp\left\{-z\frac{4\pi n_2}{\lambda_0} \cdot \sqrt{\left(\frac{n_1^2}{n_2^2}\right) \cdot \sin^2(\Theta_1) - 1}\right\} && \text{(A.41)} \end{aligned}$$

This means that the intensity of the evanescent field E_t decreases in an exponential fashion. In the following, the penetration depth d_p is introduced to express the intensity decrease in factors of $1/e$

$$d_p = \frac{\lambda_0}{4\pi n_1} \cdot \sqrt{\left(\frac{n_1^2}{n_2^2}\right) \cdot \sin^2(\Theta_1) - 1} \tag{A.42}$$

Thus, the intensity of the evanescent field as a function of the penetration depth d_p can be determined with Eq. (A.39)

$$I_{ev}(z) = I_0 \cdot \left(\frac{n_2}{n_1}\right) \cdot |T|^2 \cdot e^{-z/d_p} = A \cdot e^{-z/d_p} \tag{A.43}$$

where $A = I_0 \cdot \left(\frac{n_2}{n_1}\right) \cdot |T|^2$

A.3. Calibration of the TIRF hybridization signal

As explained in chapter 3.1.5, the hybridization signal D in surface based measurements is proportional to the amount of duplexes. In order to quantify the surface based hybridization experiments with the Langmuir formalism introduced in chapter 2, it is necessary to calibrate the hybridization signal into units equivalent to the duplex concentration in the hybridization chamber (see chapter 3.1.5). This is done with a calibration experiment which is described in the following.
Consider D_{eq}^{const} as the equilibrium hybridization signal of a single hybridization where the target concentration $[T](t)$ remains constant at the value $[T]^{const}$:

$$[T](t) = [T]^{const} \tag{A.44}$$

Furthermore, let D_{eq} be the equilibrium hybridization signal of a single hybridization experiment where $[T](t)$ decreases with the concentration $[D](t)$ of already hybridized targets:

$$[T](t) = [T]_0 - [D](t) \tag{A.45}$$

Here, $[T]_0$ is the target concentration for $t = 0$. Replacing the duplex concentration $[D]_{eq}$ in Eq. (2.24) by the measured hybridization signal D_{eq} in thermodynamic equilibrium leads to

$$K \cdot \left([T]_0 - D_{eq}\right) = \frac{D_{eq}}{\left([P]_0 - D_{eq}\right)} \tag{A.46}$$

In the case of a constant target concentration $[T]^{const}$ one reveals

$$K \cdot [T]^{const} = \frac{D_{eq}^{const}}{\left([P]_0 - D_{eq}^{const}\right)} \tag{A.47}$$

In order to calibrate the hybridization signal D into unit equivalent to the duplex concentration in the hybridization chamber, two experiments are performed. First, a single hybridization with constant target concentration $[T]^{const}$ reaching the value of D_{eq}^{const} in thermodynamic equilibrium. Second, a single hybridization, wherein the target concentration is not constant and $[T]_0$ is chosen in such a way that the equilibrium value D_{eq} is identical to D_{eq}^{const}. In this case the expression (A.46) and (A.47) are identical. One reveals

$$K \cdot [T]^{const} = K \left([T]_0 - D_{eq}\right) \tag{A.48}$$

or

$$D_{eq} = [T]_0 - [T]^{const} := [D]_{eq} \tag{A.49}$$

The difference in target concentrations $[T]_0 - [T]^{const}$ of the two measurements determines the duplex concentration $[D]_{eq}$ in the hybridization chamber. Since Eq. (A.49) is not a function of the binding affinity K of the target molecule, the calibration is valid for all molecular species at the chosen temperature. In the context of the work presented in [30], the calibration experiment is performed for the target molecule PM at a temperature of 44 °C. For a constant target concentration of $[PM]^{const} = 2$ nM an equilibrium signal of $D_{eq}^{const} = 5.58$ V is measured. For the second part of the experiment, a target concentration of $[PM]_0 = 5.7$ nM is determined to produce an identical equilibrium signal of 5.58 V. Thus, the concentration of target molecules hybridized to the probes is

$$D_{eq} = 5.7\ \text{nM} - 2\ \text{nM} = 3.7\ \text{nM} \tag{A.50}$$

From the above it follows that the calibration factor for converting the measured hybridization signal D into units equivalent to the duplex concentration in the hybridization chamber equals

$$\frac{[D]}{D} = \frac{5.58\ \text{V}}{3.7\ \text{nM}} \approx 0.65 \frac{\text{nM}}{V} \tag{A.51}$$

This factor has been verified in multiple calibration experiments with a standard derivation of about 5% (results not shown). The factor is used for converting the hybridization signal of all surface based measurements that require quantitative analysis. As explained in chapter 3.1.5, the calibration is only performed for the Cy-5 channel of the experiment, because in a competitive hybridization measurement, the fluorescence emission of the Cy-3 dyes can be quenched in close proximity to the Cy-5 labeled targets (FRET).

A.4. Regeneration of the probe molecule surfaces

A cover slip with grafted probe DNA can be used for multiple hybridization experiments. Before the start of a new experiment, hybridized target molecules can be removed by filling the hybridization chamber with a 10 mM sodium hydroxide (NaOH) solution for five seconds at 44 °C (NaOH regeneration). Fig. A.1 shows four hybridization kinetics of the PM at 44 °C with intermediary regeneration steps. Since treating the surface with NaOH also reduces the number of probe molecules immobilized to the dendrimer coated surface, the equilibrium duplex concentration decreases by about 7% per regeneration step. Thus, after regenerating a surface four times, a new substrate is used.

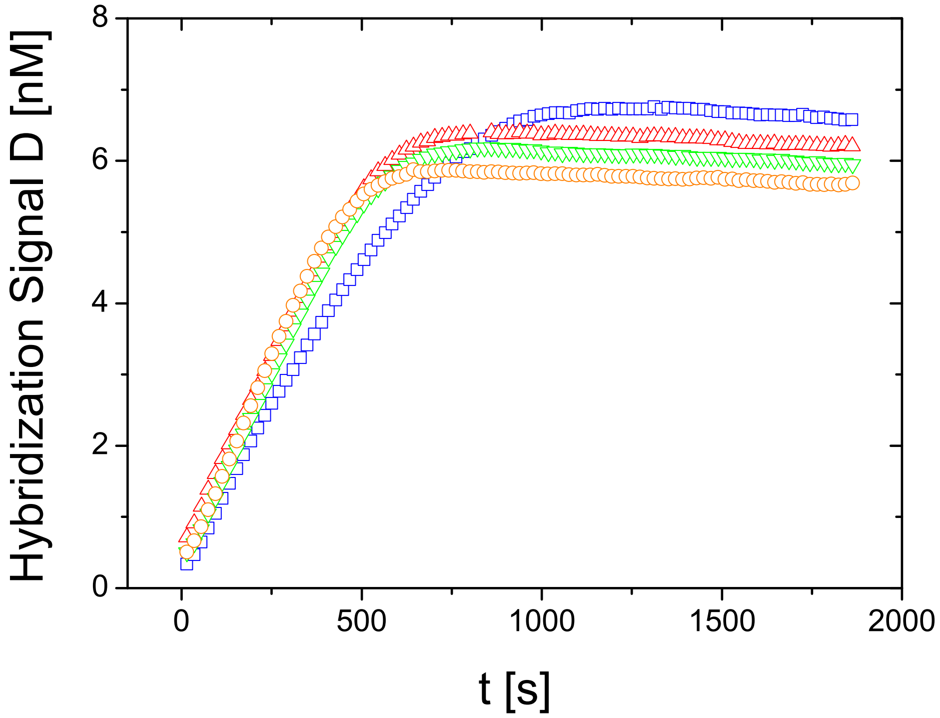


Figure A.1.: NaOH regeneration of a probe molecule surface (already presented in [30]). The graph shows the hybridization signal *D* of the PM as a function of time at 44 °C after 0 (blue), 1 (red), 2 (green) and 3 (orange) regeneration steps. The target concentration of the PM is 20 nM in each case. After each individual experiment the hybridized target molecules can be released from the probes by filling the hybridization chamber with a 10 mM sodium hydroxide (NaOH) solution for 5 seconds at 44 °C. After that the chamber is filled again with the 3xSSC buffer and the measurement can be started over. Since the NaOH treatment also releases some probe molecules, the equilibrium value of a specified hybridization measurement is reduced by about 7% compared to the previous measurement. Thus, after regenerating a surface four times, a new substrate is used.

A.5. Non-specific adsorption in surface based measurements

As explained in chapter 3.1.5, hybridized DNA targets are excited by the evanescent field, typically penetrating into the measurement chamber by about 100 nM. Besides the wanted excitation of hybridized targets it is also possible that DNA molecules are adsorbed by the dendrimer coated surface. This non-specific adsorption would also contribute to the measured hybridization signal and hence invalidate the experimental

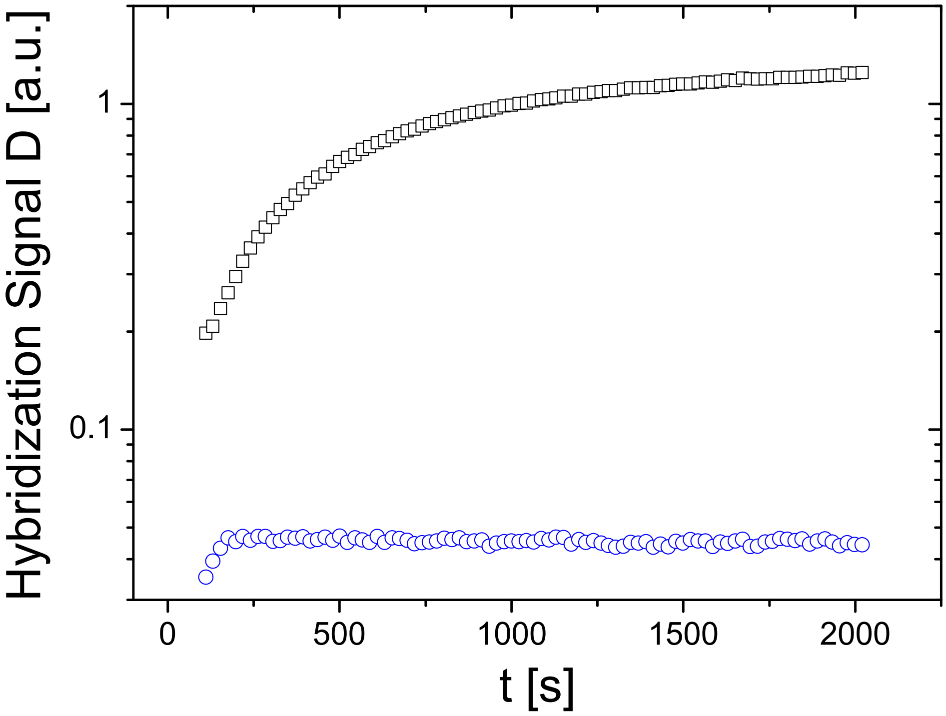


Figure A.2.: Non-specific adsorption during a DNA hybridization experiment (already presented in [30]). The graph shows the hybridization signal *D* of the PM as a function of time for a temperature of 44 °C in the case of a dendrimer coated substrate without immobilized DNA probes (blue) and a regular substrate with immobilized probes. The concentration of PM targets is 2 nM in both cases. The specific hybridization signal of the PM exceeds the contribution of the non-specific adsorption by at least one order of magnitude. Thus, the non-specific adsorption is neglected in all surface bases hybridization experiments.

data, since the target is also excited by the evanescent wave. In order to show that the contribution of a non-specific adsorption of target molecules at the dendrimer coated surface can be neglected during the hybridization experiments, the following measurement is performed (Fig. A.2). A dendrimer coated surface without immobilized DNA molecules is used as a substrate to analyze the non-specific adsorption of the PM, concentrated at 2 nM for a temperature of 44 °C. In this case, the measured fluorescence signal only stems from non-specifically adsorbed target molecules, since the complementary probes are missing. In order to directly compare the non-specific adsorption with the specific binding of a DNA target to its complementary probe, Fig. A.2 also contains the regular hybridization signal of the PM when using a regular substrate with immobilized probe molecules. The experiment shows that the non-specific adsorption is orders of magnitude lower than the specific hybridization. Thus, the contribution of non-specifically adsorbed targets is neglected in all hybridization measurements.

A.6. Sequence specific resolution of hybridization measurements

In order to verify the sequence specific resolution of DNA hybridization, the hybridization kinetics of the competitive system "PM vs. MM1" is measured. Comparable experiments were already performed in [30]. The PM is concentrated at 10 nM while the MM1 is concentrated at 50 nM. As already mentioned in chapter 3.1.5, due to energy transfer from Cy-3 to Cy-5, only the Cy-5 emission signal is proportional to the duplex concentration in the measurement chamber during competition. Thus, in order to make the hybridization signals of both species comparable, the experiment is performed twice; first, with labeling the PM with Cy5 and MM1 with Cy3 and second, with labels exchanged. The hybridization curves of both species are therefore based on two separate measurements with exchanged labels. Figure A.3 shows the result of the experiment. It reveals different hybridization kinetics for PM and MM1 during competition and consequently verifies the sequence specific resolution of the experimental setup.

A.7. Temperature dependence of the fluorescent dyes

The temperature dependent efficiency of the fluorescent dyes Cy-3 and Cy-5 can be considered as a linear function of temperature. This has been shown in earlier measurements performed by Timo Mai with a Cy labeled version of the probe molecule (results not shown). In order to determine the linear base line BL for eliminating the temperature dependent efficiency of the dye, the fluorescence signal D is measured as a function of temperature for a target excess concentration of 1 µM. Far below the melting temperature T_m, D does only depend on the temperature dependent efficiency of the dye, since the thermal fluctuations are too weak to pre-melt the duplex. The linear fit of these data points reveals the base line BL of the denaturation experiment. The surface occupancy Θ is determined by eliminating BL according to

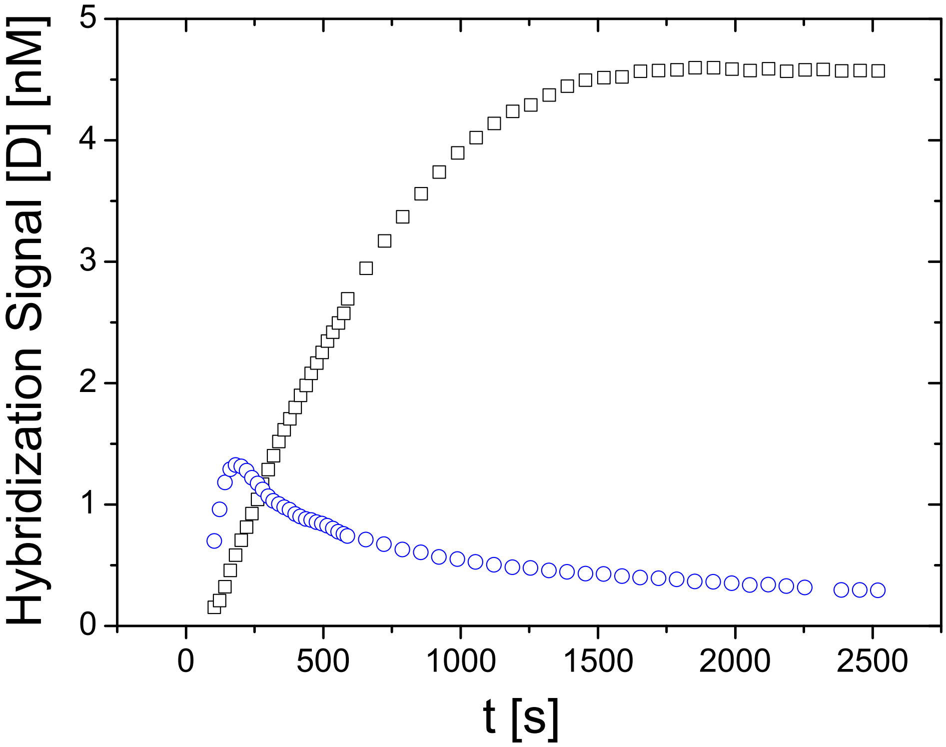


Figure A.3.: Verification of the sequence specificity of competitive hybridization using the example of 'PM vs. MM1' (already presented in [30]). The graph shows the fluorescence signal of hybridized PM (black, initial target concentration 10 nM) and hybridized MM1 (blue, initial target concentration 50 nM) at a temperature of 44 °C measured with the TIRF setup. Due to energy transfer between Cy-3 and Cy-5, the data is based on two separate experiments, first one with labeling PM with Cy5 and MM1 with Cy3 and the second one with labels exchanged. The result reveals the sequence specific resolution of the experimental setup.

$$\Theta = 1 - \frac{D}{BL} \tag{A.52}$$

Using the example of a MM1 denaturation curve, Fig A.4 shows the elimination of the temperature dependent efficiency of the dye as explained above.

A.8. Comparison between cyanine and Exiqon dyes

Fluorescent labels are known to stabilize double-stranded DNA [54]. Most DNA oligonucleotides used in this work are labeled with the cyanine dyes Cy-3 and Cy-

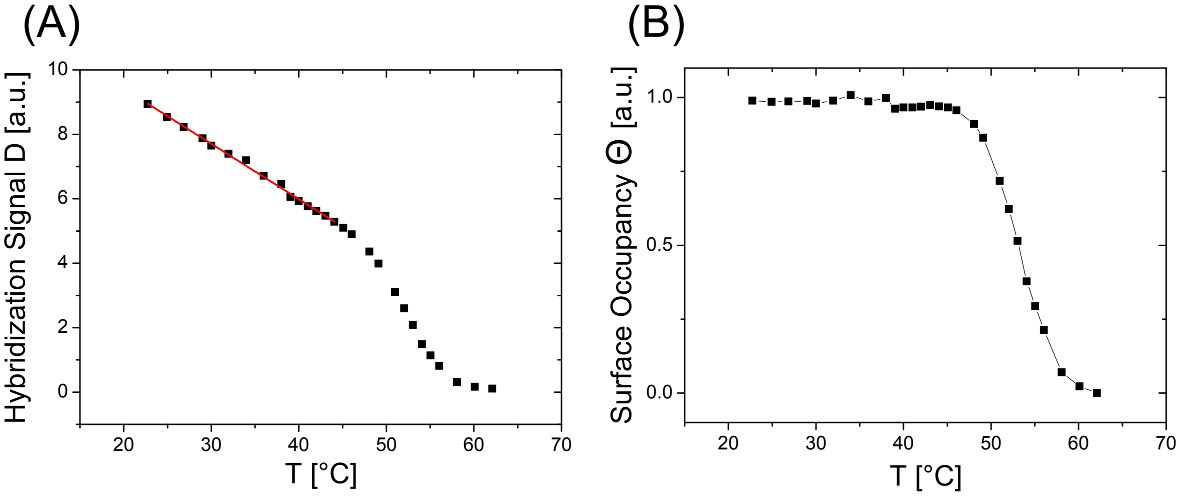


Figure A.4.: Elimination of the temperature dependent efficiency of the fluorescent dye in a denaturation experiment, using the example of the MM1. **(A)**, raw data. The graph shows the hybridization signal D as a function of temperature. Since an excess target concentration of 1 μM is used in this experiment, all probe molecules are occupied by MM1 targets at low temperatures. Far below the melting point T_m, D does only depend on the temperature dependent efficiency of the dye. The linear fit of these data points (red lines) reveals the base line BL of the denaturation experiment. **(B)**, denaturation curve of the MM1 after the correction of the experimental data with Eq. (A.52) thereby using the base line determined in (A).

5. Thus, in competitive hybridization experiments, both target molecules are stabilized in the same way by their respective dye so that the stabilization effect does not have to be considered. However, the LNA targets (compare appendix Table A.1) cannot be ordered with the cyanine dyes, since the vendor *Exiqon* does not provide them. Instead, the LNA targets are labeled with *Exiqon* dyes named “563” and “667” indicating the fluorescence emission wavelength of 563 nm and 667 nm. According to *Exiqon* their dyes can be used as a substitute for Cy-3 and Cy-5. However, the stabilizing effect of these dyes and the cyanine dyes have to be compared. This is done in the following denaturation experiment. The melting curve of the target MM1, modified with the fluorescent dye Cy-5 is compared with the melting curve of the MM1, modified with the *Exiqon* dye “667”. The result is presented in Fig. A.5. It becomes apparent that the *Exiqon* dyes additionally stabilize double-stranded DNA compared to the usually employed cyanine dyes. The contribution of the temperature dependent efficiency of the fluorescent dyes to the hybridization signals is removed like explained in appendix

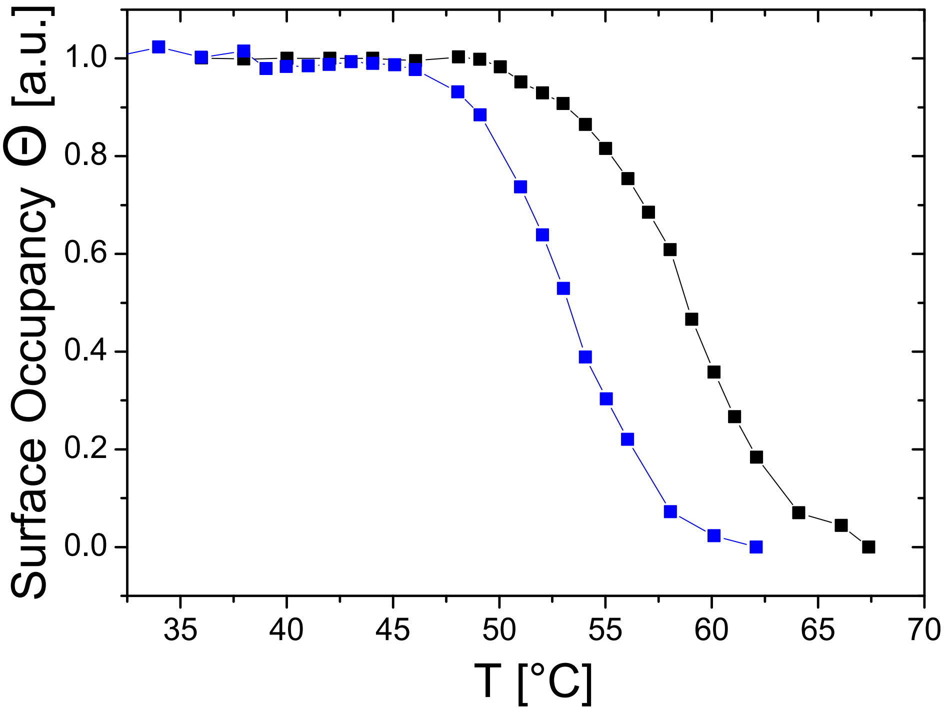


Figure A.5.: Denaturation experiment for two versions of the MM1 labeled with the fluorescent dye Cy-5 and the *Exiqon* dye "667". The graphs show the fraction of occupied probes Θ as a function of temperature T for MM1-Cy5 (black) and MM1-"667" (blue) in individual experiments without competition. The target concentration is 1 µM in both cases. The melting temperature T_m is defined as the temperature where the surface occupancy equals 50 percent. T_m is determined to be 53 °C for the Cy-5 labeled version of the MM1 and 58 °C for the "667" labeled version of the MM1. The *Exiqon* dyes additionally stabilize double-stranded DNA compared to molecules modified with cyanine dyes.

section A.7 thereby assuming that the temperature efficiency of the *Exiqon* dye "667" is also a linear function of temperature. The melting temperature of the "667" labeled version of the MM1 is significantly increased by about 5 °C compared to the Cy-5 labeled version. Thus, in competitive hybridization experiments between LNA and DNA targets, both, LNA and DNA molecules are labeled with the *Exiqon* dyes in order to avoid differently pronounced stabilization effects that may lead to misinterpretation of the data.

A.9. Averaging correlation functions and diffusion elimination in FCS

As explained in chapter 3.1.7, the correlation functions obtained with the FCS setup are mean values of at least four individual experiments. A measured correlation function $G_{meas}(t)$ consists of two contributions, the correlation $G_{diff}(t)$ due to the diffusion of the DNA molecules through the focused laser beam in the measurement chamber and the wanted correlation $G(t)$ of the base pair fluctuation dynamics. In order to eliminate the contribution of diffusion, $G_{meas}(t)$ is divided by G_{diff} measured on double-stranded molecules lacking the quencher Cy-5:

$$G(t) = \frac{G_{meas}(t)}{G_{diff}(t)} \tag{A.53}$$

Using the example of the MM1 hybridizing to the probe molecules at 44 °C, Fig. A.6 shows the approach for averaging the correlation functions and eliminating the contribution of diffusion.

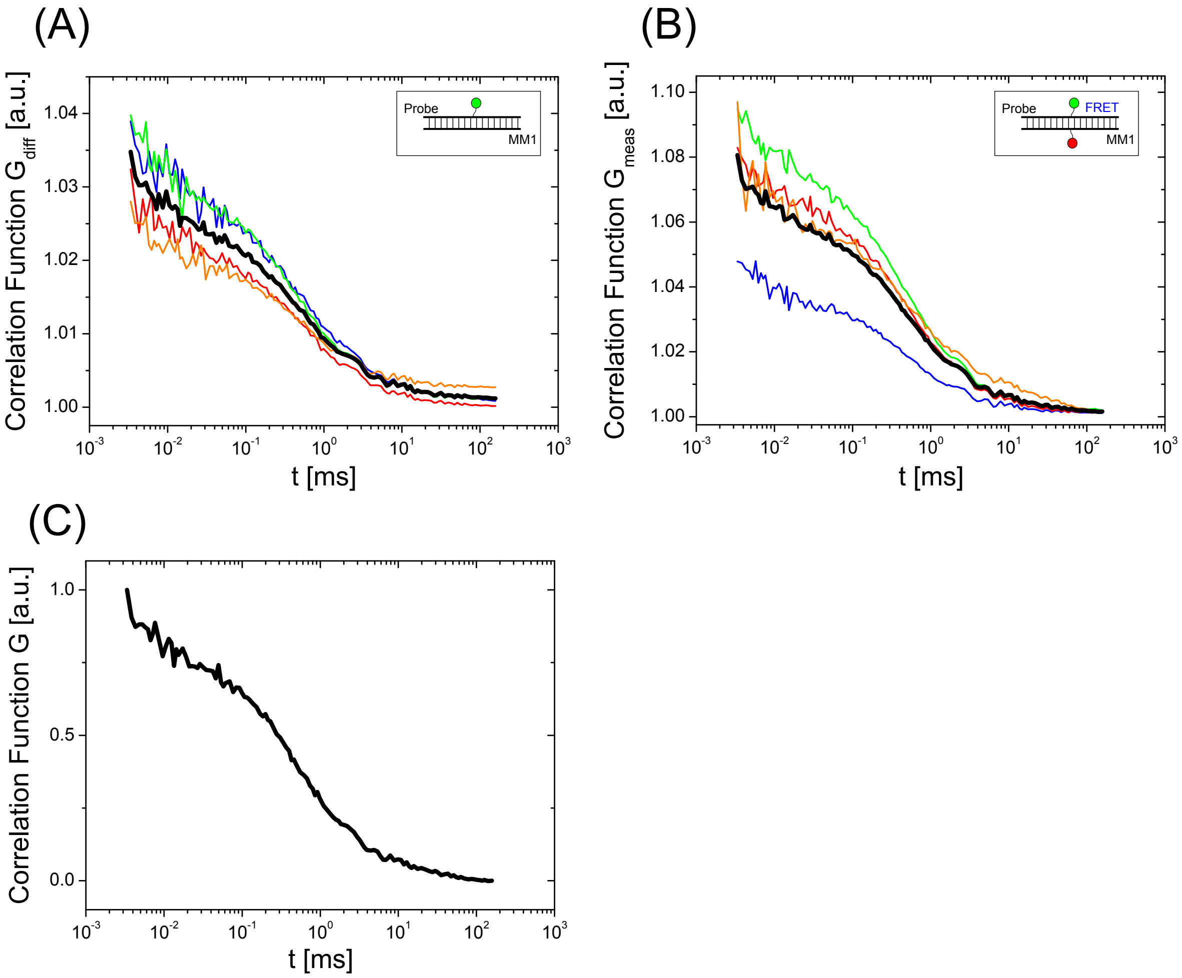


Figure A.6.: Averaging correlation functions and eliminating the contribution of diffusion. **(A)**, autocorrelation function G_{diff} of the diffusion of MM1:Probe duplexes through the focused laser beam as a function of time at 44 °C. The MM1 targets are lacking the Cy-5 quencher. The mean value of $G_{diff}(t)$ (black curve) is determined by averaging the correlation functions measured for at least four individual DNA solutions (red, blue, green and orange curve). **(B)**, autocorrelation function $G_{meas}(t)$ of the diffusion of MM1:Probe duplexes through the focused laser beam including the wanted correlation $G(t)$ of the opening and closing dynamics of the base pairs as a function of time. The correlation of the base pair fluctuations are detected by FRET between the fluorescent dyes Cy-3 and Cy-5 (colored graphs: individual measurements, black graph: mean value). **(C)**, correlation function $G(t)$ of the base pair fluctuations determined with Eq. (A.53) using the data from (A) and (B). Data is normalized to 1.

B. Appendix: In vitro gene expression

B.1. DNA vectors and PCR products

Table B.1 gives an overview of all DNA templates used in this work for the study of the epigenetic regulation of *in vitro* gene expression (chapter 6).

B.2. In vitro methylation

Enzymatic *in vitro* methylation is performed to methylate the reporter plasmids or reporter PCR products. The methylases M.SssI and HhaI are used for CpG methylation, while Dam is used for GATC methylation. Compared to the protocols provided by the vendor of methyltransferases (*New England Biolabs*, USA), the duration of the reaction is increased to three hours at 37 °C in order to reveal complete methylation of the template DNA. 25 µl of the enzyme is used to methylate 10 µg DNA. In order to varify complete methylation of the DNA templates, restriction digest with methylation dependent endonucleases HhaI (blocked by CpG methylation), DpnI (only active at Dam methylated GATC sites), and MboI (blocked by Dam methylation) is performed. Fig. B.1 shows three examples of analyzing the methylation state of the DNA templates.

Table B.1.: Overview of all DNA vectors and PCR products. All plasmids used in the context of this work stem from the vector pBEST-Luc (*Promega*, USA) containing a ColE1 origin of replication. It can be replaced by a p15a origin, involving a reduced number of plasmid copies per E. coli cell. OR2-OR1-PR indicates the strong lambda promoter Pr flanked by the operons OR2 and OR1 used for highly efficient expression of the regulator proteins Lrp, PapI, and MBD. UTR1 and T-500 stand for the untranslated region and the highly efficient transcription terminator T-500, respectively. The different pap promoters refer to the regulators of the expression of Pyelonephritis-associated pili in E. coli. The number behind pap indicates which Lrp binding sites has been included. The promoter mbd contains the recognition sequence for binding of the MBD. eGFP and mCherry stand for the reporters enhanced green fluorescent protein and mCherry, respectively.

Plasmids and PCR products
pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500
pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500
pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500
pBEST-pap123-UTR1-eGFP-T500
pBEST-pap2-UTR1-eGFP-T500
pap222-UTR1-eGFP-T500 (PCR product)
pBEST-pap456-UTR1-eGFP-T500
pBEST-pap5-UTR1-eGFP-T500
pap555-UTR1-eGFP-T500 (PCR product)
pBEST-mbd-UTR1-eGFP-T500
pBEST-mbd-UTR1-mCherry-T500
pBEST-mbdM1-UTR1-eGFP-T500
pBEST-mbdM2-UTR1-eGFP-T500
pBEST-mbdM3-UTR1-eGFP-T500
pBEST-mbdM4-UTR1-eGFP-T500

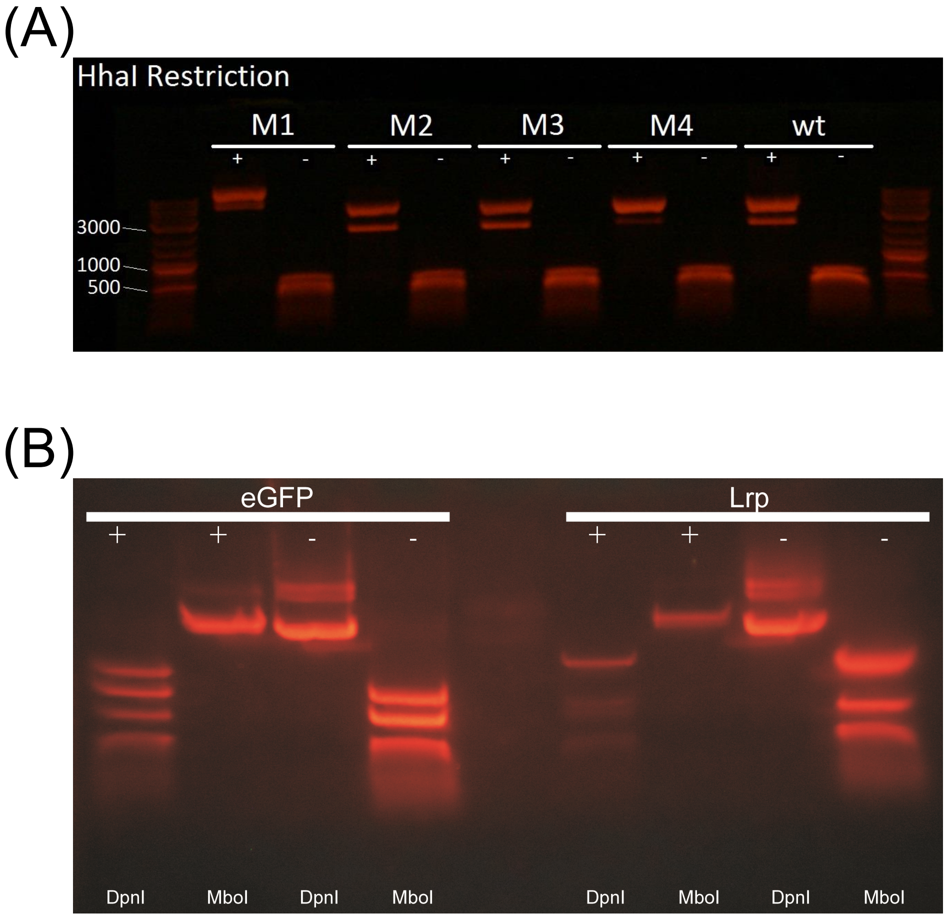


Figure B.1.: Analyzing the methylation state of DNA templates. **(A)**, restriction digest of the plasmids pBEST-mbd-UTR1-eGFP-T500 involving the four mutations (M1-M4) and the wild type (wt) studied in the context of this work, using the methylation dependent restriction endonuclease HhaI. The "+" stands for template DNA, previously incubated for three hours with the methyltransferase M.SssI, which methylates all CpG sites of the DNA. The "-" DNA has not been incubated with M.SssI. The first and the last lane show a DNA ladder and the size of the three major fluorescent bands are given (number of bases). **(B)**, restriction digest of the plasmids pBEST-pap123-UTR1-eGFP-T500 (eGFP) and pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 (Lrp) with the restriction endonucleases DpnI and MboI after gel electrophoresis and staining with ethidium bromide. The "+" stands for template DNA previously incubated with the Dam methyltransferase for three hours, while the "-" stands for template DNA free of Dam methylation.

B.3. Protocol for the preparation of the PGA buffer

Each component of the PGA buffer is prepared individually and stored at −80 °C. Deionized water is used for the preparation of all solution.

- Hepes (*Sigma-Aldrich, Switzerland*)
  476.62 mg/ml is solved in water. The final concentration of the stock solution is 2 M. pH is adjusted to a value of 8 with a potassium hydroxide solution.

- Nucleotides (*Sigma-Aldrich, Switzerland*)
  500 mg of ATP dipotassium salt dihydrate is solved in a total volume of 1495 µl of water (540 mM), and aliquoted by 77 µl. 500 mg of GTP disodium salt is solved in a total volume of 1160 µl of water (760 mM), and aliquoted by 56 µl. 300 mg of UTP trisodium salt dihydrate is solved in a total volume of 630 µl of water (813 mM), and aliquoted by 33 µl. 300 mg of CTP disodium salt dihydrate is solved in a total volume of 858 µl of water (621 mM), and aliquoted by 42 µl. 74 µl of the 540 mM ATP solution, 52.8 µl of the 760 mM GTP solution, 29.6 µl of the 813 mM UTP solution, and 38.8 µl of the 621 mM CTP solution are mixed with 60 µl of a 15% (v/v) potassium hydroxide solution.

- tRNA E. coli (*Roche*, Switzerland)
  50 mg of the tRNA are dissolved in 1 ml water. The solution is aliquoted by 100 µl.

- Coenzyme A (*Sigma-Aldrich, Switzerland*)
  50 mg of coenzyme A hydrate yeast is solved in a total volume of 1 ml of water. The final concentration of the stock solution is 65 mM. The solution is aliquoted by 100 µl.

- NAD (*Sigma-Aldrich, Switzerland*)
  100 mg of beta-nicotinamide adenine dinucleotide hydrate is solved in 727 µl of water with the addition of 77 µl of a 2M Tris solution. The final concentration of the stock solution is 175 mM. The solution is aliquoted by 50 µl.

- cAMP (*Sigma-Aldrich, Switzerland*)
  150 mg of adenosine 3′, 5′ cyclic monophosphate is solved in 348 µl of water with the addition of 257 µl of a 2M Tris solution. The final concentration of the stock solution is 650 mM. The solution is aliquoted by 32 µl.

- Folinic acid (*Sigma-Aldrich, Switzerland*)
  20 mg of folinic acid calcium salt is solved in a total volume 11534 µl. The final concentration of the stock solution is 33.9 mM. The solution is aliquoted by 51 µl.

- Spermidine (*Sigma-Aldrich, Switzerland*)
  100 µl is solved in 537 µl of water. The final concentration of the stock solution is 1 mM. The solution is aliquoted by 27 µl.

- 3-PGA (*Sigma-Aldrich, Switzerland*)
  3 g of D-3-phosphoglyceric acid disodium salt is solved in 3156 µl of water with the addition of 5.04 ml of a 2M Tris solution. The final concentration of the stock solution is 1.4 M. The solution is aliquoted by 520 µl.

The 4-fold 3-PGA buffer is prepared by mixing the following volumes of each stock solution: 600 µl 3-PGA, 24 µl water, 231.2 µl nucleotides, 96 µl tRNA E. coli, 96 µl coenzyme A, 46 µl NAD, 28.4 µl cAMP, 48 µl folinic acid, 24 µl spermidine, and 514.4 µl spermidine. The total volume of 1708 µl is aliquoted by 7 µl and shock frozen in liquid nitrogen. 6.43 µl of the 3-PGA buffer are used for a batch mode cell free reaction.

B.4. Protocol for the preparation of the amino acid stock solution

All 20 amino acids are part of the RTS amino acid sampler (*Sigma-Aldrich, Switzerland*). They are concentrated at 168 mM (except Leucine at 140 mM) in a volume of 1.5 ml. All acids are mixed together with the addition of 12 ml water. The final concentration of the solution is 6 mM for each individual amino acid (except Leucine at 5 mM). The solution is aliquoted by 10 µl and shock frozen in liquid nitrogen.

B.5. Protocol for the preparation of the crude extract

List of materials

- 2xYT broth (*Sigma-Aldrich*, Switzerland)

- Agar (*Sigma-Aldrich*, Switzerland)

- 0.1 mm glass beads (*Biospecs*, USA)
- Mini bead beater (*Biospecs*, USA)
- Bead beating tubes (*Biospecs*, USA)
- Micro bio-spin chromatography columns (*Biospecs*, USA)
- Allegra 25R centrifuge (*Beckman Coulter*, USA)
- Table centrifuge (*Eppendorf*, Germany)
- 200 ml centrifuge bottle (*Beckman Coulter*, USA)
- Chloramphenicol (*Sigma-Aldrich*, Switzerland)
- Standard disposable cuvettes (*Fisher Scientific*, USA)
- 10 kD MWCO dialysis cassettes (*Slide-A-Lyzer*, *Thermo Life Science*, USA)
- DL-Dithiotheritol (DTT), (*Sigma-Aldrich*, Switzerland)
- BL21 Rosetta 2 competent E. coli cells (*Novagen*, Germany)
- Standard pH meter
- 3 4000 ml Erlenmeyer glass flasks, (*Schott*, Germany)
- Incubator Shaker (*Edmund Bühler*, Germany)
- Ethanol $>$ 99% (*Sigma-Aldrich*, Switzerland)
- L-Glutamic acid potassium salt mono hydrate (*Sigma-Aldrich*, Switzerland)
- L-Glutamic acid hemimagnesium salt tetra hydrate (*Sigma-Aldrich*, Switzerland)
- Potassium phosphate mono basic solution (*Sigma-Aldrich*, Switzerland)
- Potassium phosphate di basic solution (*Sigma-Aldrich*, Switzerland)
- Tris base (*Sigma-Aldrich*, Switzerland)
- Acidic acid (*Fisher Scientific*, USA)
- Protein assay dye reagent concentrate (*Bio-rad*, USA)
- Spectrometer (*Varian*, USA)

- Table vortexer (*IKA*, Germany)
- Special accuracy weighing machine (*Ohaus*, Germany)

Preparation of the solutions and materials

All solutions are prepared with autoclaved deionized water.

- 2xYT broth
  62 g of the 2xYT powder is solved in a total volume of 1.8 l of water. The solution is autoclaved at 120 °C for 15 minutes and stored at 4 °C.
- Phosphate solution
  100 ml of the di basic solution is mixed with 55 ml of the mono basic solution. The mixture is stored at 4 °C.
- Chloramphenicol
  6.8 mg of chloramphenicol powder is solved in a total volume of 200 µl of ethanol. The solution is stored at −20 °C.
- 2xYT plate
  1.24 g of 2xYT powder and 0.44 g of agar are solved in a total volume of 40 ml of water and then autoclaved at 120 °C for 15 minutes. After autoclaving, the solution is allowed to cool down to a temperature of about 50 °C. 2.5 ml of the phosphate solution and 60 µl of chloramphenicol at a concentration of 34 mg/ml is added. 25 ml of the solution is pipetted into a standard petri dish that is incubated for about one hour at room temperature until the agar has gelified.
- DTT
  5 g of DTT powder is solved in a total volume of 32.84 ml of water. The solution is aliquoted by 1000 µl and stored at −20 °C.
- Tris solution
  60.57 g of the Tris powder is solved in a total volume of 250 ml of water and autoclaved at 120 °C for 15 minutes. The final concentration of the solution is 2 M.
- S30 buffer A
  10.88 g of the Mg-glutamate and 24.4 g of the K-glutamate is solved in at total volume of 2 l of water with the addition of 50 ml of the 2 M Tris solution. pH

is adjusted to a value of 7.7 by using acidic acid. The solution is autoclaved at 120 °C for 15 minutes and stored at 4 °C.

- S30 buffer B
  10.88 g of the Mg-glutamate and 24.4 g of the K-glutamate is solved in at total volume of 2 l of water. pH is adjusted to a value of 8.2 by using the 2 M Tris solution. The solution is autoclaved at 120 °C for 15 minutes and stored at 4 °C.

All the equipment that comes in contact with organic material is washed thoroughly and sterilized.

Transformation

A pipette tip of the BL21 cells is solved in 500 µl of the 2xYT. 10 µl of the solution is transferred on the 2xYT agar plate. The plate is incubated for one night at 37 °C.

Mini Culture

One bacterial colony from the 2xYT plate is transferred into 4 ml of 2xYT with the addition 0.27 ml of the phosphate solution and 4 µl of chloramphenicol. The culture is incubated at 37 °C for about eight hours under permanent agitation. After that 100 µl of this culture is transferred into 50 ml of 2xYT with the addition 3.3 ml of the phosphate solution and 50 µl. The culture is also incubated for about eight hours at 37 °C.

Extract preparation

248 ml of the phosphate solution is mixed with 3.6 l of 2xYT. Six 4 l Erlenmeyer flasks are filled with 660 ml per flask of this mixture. After that, 6.6 ml of the last mini culture is transferred into each flask. Bacterias are grown at 37 °C to a optical density of $OD_{600} = 1.5$. All the subsequent steps are performed at temperatures below 4 °C if not stated otherwise. The solution is transferred into four 200 ml centrifuge bottles and cells are harvested by centrifuging at 4425 x g for 12 minutes. The cells are washed twice with 200 ml of S30 buffer A containing 4000 µl of the DTT solution and resuspended in 37.5 ml of the same buffer. The mixture is transferred into four 50 ml Falcons and then again centrifuged at 1935 x g for eight minutes. The supernatant is discarded and the cell pellet is centrifuged again with the same parameters in order to completely

remove the buffer. After that, the cell pellet is weighed and then resuspended with S30 buffer A (mass of the cells times 0.9, [ml]) and the glass beads (mass of cells times 5.0, [g]). The viscous mixture is homogenized by thoroughly vortexing and then transferred into 1.5 ml bead beating tubes. Bead beating is performed twice at 4600 rpm for 30 seconds with the micro bead beater in order to crack the cell membrane. The lysate is then centrifuged at 6805 x *g* for ten minutes through chromatography columns in order to separate the beads from the cell lysate. The clear supernatant is then transferred into fresh bead beating tubes that are in the following centrifuged again at 12000 x *g* for 20 minutes. The clear supernatant is transferred into standard reaction tubes (500 µl per tube) and then incubated for 80 minutes at 37 °C under permanent agitation. The extract is again centrifuged at 12000 x *g* for ten minutes and then recovered into a 15 ml Falcon. Subsequently, the extract is transferred into a dialysis cassette by using a syringe and then dialyzed for three hours, under permanent agitation, against S30 buffer B containing 2000 µl of the DTT solution. The recovered extract is aliquoted by a volume of 30 µl in standard reaction tubes and shock frozen in liquid nitrogen. The protein concentration of the extract is determined by preparing a Bradford assay that is analyzed with the photospectrometer. The typical protein concentration of the extract is 27 – 30 mg/ml. The extract can be stored at −80 °C over a time period of at least one year.

B.6. Protocol for the preparation of a batch mode reaction

The preparation of a batch mode reaction is performed on ice. A typical reaction consists of

- 30 µl of the crude extract
- 2.7 µl of Mg-glutamate concentrated at 100 mM
- 1.5 µl of K-glutamate concentrated at 3 M
- 6.43 µl of the PGA-buffer
- 7.5 µl of all amino acids concentrated at 6 mM
- 4.5 µl 40% (v/v) poly ethylene glycol 8000 (*Sigma-Aldrich*, Switzerland)

- A specified volume of the DNA stock solution to reach the wanted DNA concentration
- A specified volume of deionized water to reach a total volume of 81 µl

In the case of expressing linear DNA templates, 3.3 µl of the recombined lambda phage protein gamS, concentrated at 99 µM, is added to the reaction. After homogenizing the solution by vortexing, a volume of 10.8 µl is pipetted into standard reaction tubes. A total volume of 81 µl is adequate to fill seven tubes. After that, 1.2 µl of the 10-fold concentrated DNA template coding for the regulator protein is added into each tube. For example, to reach a final concentration of 0, 0.5, 1, 2, 3, and 4 nM of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-MBD-T500 within the reaction, stock solutions of 0, 5, 10, 20, 30, and 40 nM have to be pipetted into the tubes. In the case of an end-point measurement, the reaction volume of 12 µl is divided (6 µl/6 µl) in order to allow an efficient diffusion of oxygen into the reaction. In the case of measuring the expression kinetics, 10 µl of the solution is directly pipetted into the wells of a multi well plate, which is in the following placed on the pre-incubated plate reader. The reaction is incubated for at least 12 hours at 29 °C.

B.7. Degradation of CpG methylated DNA in the cell free reaction

The E. coli strain BL21, which is used for the preparation of the crude extract (see chapter 3.2.3), is a mcr$^+$ E. coli strain. The mcr restriction system, in the presence of GTP can cleave CpG methylated DNA [55]. Thus, it is important to analyze if a CpG methylated plasmid does not contain the mcr recognition sequence, since otherwise the DNA is degraded by the restriction system in the cell free extract. For the purpose of demonstrating the degradation of CpG methylated DNA by the mcr system, a CpG methylated reporter plasmid containing the mcr recognition sequence is expressed in the standard BL21 extract as well as in an extract produced of a *NEB-10-beta* E. coli strain (*New England Biolabs*, USA) lacking the mcr system. The results are presented in Fig. B.2. In the context of this work, all CpG methylated plasmids (pBEST-mbd-UTR1-eGFP-T500 and its four mutated versions as well as pBEST-mbd-UTR1-mCherry-T500 (compare appendix Table B.1) are not degraded by the mcr system and can therefore be expressed in the standard BL21 extract.

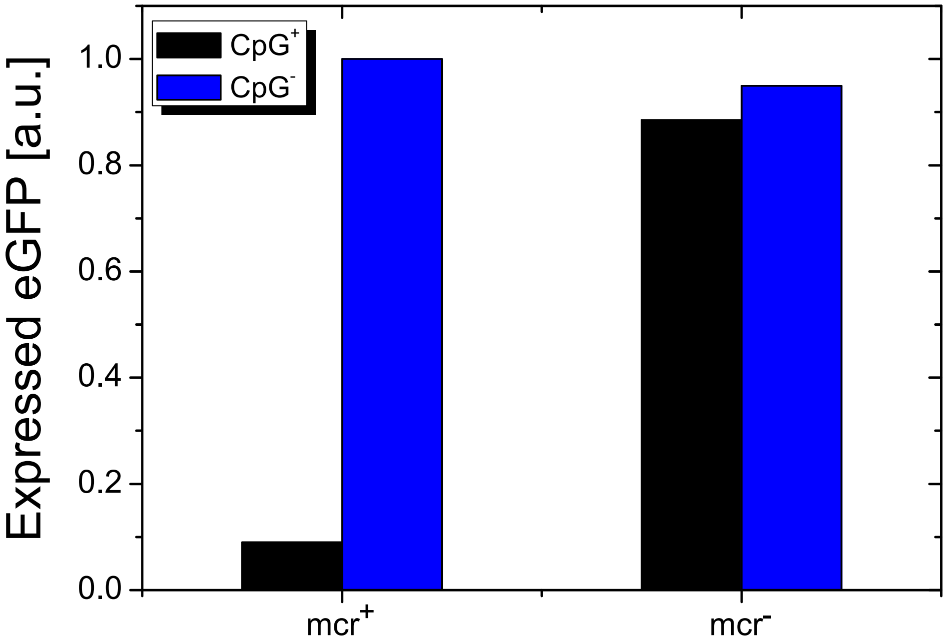


Figure B.2.: Degradation of CpG methylated DNA in the cell free reaction. The graph shows the eGFP expression level using 5 nM of a CpG methylated (black) and non-methylated (blue) mcr positive reporter plasmid, using the standard BL21 extract (mcr^+) and a NEB-10-beta extract (mcr^-) not containing the mcr restriction system. Data is normalized to the $\text{mcr}^+/\text{CpG}^-$ case. The amino acid concentration is 0.5 mM in all measurements. CpG methylated DNA is degraded in the mcr^+ extract.

Bibliography

[1] N. A. Campbell and J. B. Reece. *Biologie*. Pearson Studium, 2009.

[2] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72:1335–1346, 1997.

[3] S. A. Rice and P. Doty. The thermal denaturation of desoxyribose nucleic acid. *Journal of the American Chemical Society*, 79:3937–3947, 1957.

[4] D. Pörschke and M. Eigen. Co-operative non-enzymic base recognition III. Kinetics of the helix-coil transition of the oligoribouridylic oligoriboadenylic acid system and of oligoriboadenylic acid alone at acid pH. *J. Mol. Biol.*, 62:361–381, 1971.

[5] Thomas Naiser. *Characterization of oligonucleotide microarry hybridization*. Dissertation, Universität Bayreuth, 2007.

[6] W. Michel, T. Mai, T. Naiser, and A. Ott. Optical study of DNA surface hybridization reveals DNA surface density as a key parameter for microarray hybridization kinetics. *Biophysical Journal*, 92:999–1004, 2007.

[7] J. Bishop, S. Blair, and A. M. Chagovetz. A competitive kinetic model of nucleic acid surface hybridization in the presence of point mutants. *Biophysical Journal*, 90:831–840, 2006.

[8] J. Bishop, M. Chagovetz, and S. Blair. Kinetics of multiplex hybridization: Mechanism and implications. *Biophysical Journal*, 94:1726–1734, 2008.

[9] M. T. Horne, D. J. Fish, and A. S. Benight. Statistical thermodynamics and kinetics of DNA multiplex hybridization reactions. *Biophysical Journal*, 91:4133–4153, 2006.

[10] H. T. Allawi and J. Santalucia. Nearest neighbor thermodynamic parameters for internal G-A mismatches in DNA. *Biochemistry*, 37:2170–2179, 1998.

[11] John Santalucia and Donald Hicks. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

[12] F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[13] M. Ehrlich, M. A. Gama-Sosa, L. H. Huan, R. M. Midgett, K. C. Kuo, R. A. McCune, and C. Gehrke. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, 10(8):2709–2721, 1982.

[14] S. Kriaucionis and N. Heintz. The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and enriched in purkinje neurons. *Science*, 324:929–930, 2009.

[15] M. Tahiliani, K. P. Koh Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324:930–935, 2009.

[16] B. A. Braaten, X. Nou, L. S. Kaltenbach, and D. A. Low. Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in E. coli. *Cell*, 76:577–588, 1994.

[17] A. D. Hernday, B. A. Braaten, and D. A. Low. The mechanism by which DNA adenine methylase and Papi activate the pap epigenetic switch. *Molecular Cell*, 12:947–957, 2003.

[18] A. D. Hernday, B. A. Braaten, G. Broitman-Maduro, P. Engelberts, and D. A. Low. Regulation of the pap epigenetic switch by CpxAR: Phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. *Molecular Cell*, 16:537–547, 2004.

[19] M. Totsika, D. G. Moriel, A. Idris, B. A. Rogers, D. J. Wurpel, M.-D. Phan, D. L. Paterson, and M. A. Schembri. Uropathogenic Escherichia coli mediated urinary tract infection. *Current Drug Targets*, 13:1386–1399, 2012.

[20] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. Genbank. *Nucleic Acids Research*, 39:D32–D37, 2011.

[21] T. Kawamura, A. S. Vartanian, H. Zhou, and F. W. Dahlquist. The design involved in Papi and Lrp regulation of the pap operon. *J. Mol*, 409:311–332, 2011.

[22] L. S. Kaltenbach, B. A. Braaten, and D. A. Lo. Specific binding of papi to Lrp-pap DNA complexes. *Journal of Bacteriology*, 177:22:6449–6455, 1995.

[23] H. Berman, K. Henrick, H. Nakamura, and J. L. Markley. The worldwide protein data bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35:D301–D303, 2007.

[24] R. I. D. Wakefield, B. O. Smith, X. Nan, A. Free, A. Soteriou, D. Uhrin, A. P. Bird, and P. N. Barlow. The solution structure of the domain from MeCP2 that binds to methylated DNA. *Journal of Molecular Biology*, 291:1055–1065, 1999.

[25] W. G. Chen, Q. Chang, Y. Lin, A. Meissner, A. E. West, E. C. Griffith, R. Jaenisch, and M. E. Greenberg. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science*, 302:885–889, 2003.

[26] K. L. Ho, I. W. McNae, L. Schmiedeberg, R. J. Klose, A. P. Bird, and M. D. Walkinshaw. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Molecular Cell*, 29:525–531, 2008.

[27] R. J. Klose, S. A. Sarraf, L. Schmiedeberg, S. M. McDermott, I. Stancheva, and A. P. Bird. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Molecular Cell*, 19:667–678, 2005.

[28] X. Zou, W. Ma, I. A. Solov'yov, C. Chipot, and K. Schulten. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Research*, 40(6):2747–2758, 2012.

[29] N. Markham and M. Zuker. Dinamelt webserver for nucleic acid melting prediction. *Nucleic Acids Research*, 33:W577–W581, 2005.

[30] Marc Schenkelberger. *Die physikalische Beschreibung der DNA Konkurrenzhybridisierung an Oberflächen*. Diplomarbeit, Universität des Saarlandes, 2009.

[31] Phillip Baaske. *Messungen der DNA-Hybridisierungskinetik an funktionalisierten Oberflächen mittels Förster Resonanz Energie Transfer und evaneszentem Feld*. Diplomarbeit, Universität Bayreuth, 2005.

[32] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Kluwer Academic / Plenum Publishers, 1999.

[33] E. Trévisiol, V. Le-Berre-Anton, and J. Leclaire. Dendrislides, dendrichips: a simple chemical functionalization of glass slides with phosphorus dendrimers as an effective means for the preparation of biochips. *New J. Chem*, 27:1713–1719, 2003.

[34] G. Altan-Bonnet, Albert Libchaber, and Oleg Krichevsky. Bubble dynamics in double-stranded DNA. *Physical Review Letters*, 90:138101 1–4, 2003.

[35] J. Shin and V. Noireaux. Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. *Journal of Biological Engineering*, 4(9), 2010.

[36] Christian Trapp. *Molekulare Erkennung am Beispiel der DNA-Hybridisierung*. Dissertation, Universität des Saarlandes, 2013.

[37] T. Kigawa, T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, A. Tanaka, and S. Yokoyama. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics*, 5:63–68, 2004.

[38] L. Albertazzi, D. Arosio, L. Marchetti, F. Ricci, and F. Beltram. Quantitative FRET analysis with the EGFP-mCherry fluorescent protein pair. *Photochemistry and Photobiology*, 85:287–297, 2009.

[39] J. Shin and V. Noireaux. An E. coli cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. *ACS Synthetic Biology*, 1:29–41, 2012.

[40] Christian Trapp. *Kodierung und DNA-Microarrays*. Diplomarbeit, Universität des Saarlandes, 2009.

[41] Luis P. Reynaldo, Alexander V. Vologoskii, Bruce P. Neri, and Victor I. Lyamichev. The kinetics of oligonucleotide replacements. *J. Mol. Biol.*, 297:511–520, 2000.

[42] R. S. Quartin, M. Plewinska, and J. G. Wetmur. Branch migration mediated DNA labeling and cloning. *Biochemistry*, 23:8676–8682, 1989.

[43] H. Kaur, A. Arora, J. Wengel, and S. Maiti. Thermodynamics, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry*, 45:7347–7355, 2006.

[44] Y. You, B. G. Moreira, M. A. Behlke, and R. Owczarzy. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Research*, 34(8):1–11, 2006.

[45] 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. USA*, 104(3):712–716, 2007.

[46] D. Xu, K. O. E., and T. M. Nordlund. Melting and premelting transition of an oligomer measured by DNA base fluorescence and absorption. *Biochemist*, 33:9592–9599, 1994.

[47] J. Santalucia. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95:1460–1465, 1998.

[48] C. Vanderzande. *Lattice models of polymers*. Cambridge University Press, 1998.

[49] C. Trapp, M. Schenkelberger, and A. Ott. Stability of double-stranded oligonucleotide DNA with a bulged loop: a microarray study. *BMC Biophysics*, 4:20:1–13, 2011.

[50] 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. *Physical Review Letters*, 102:218301 1–4, 2009.

[51] Y. Zeng, A. Montrichok, and G. Zocchi. Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.*, 339:67–75, 2004.

[52] M. W. van der Woude, B. A. Braaten, and D. A. Low. Evidence for global regulatory control of pilus expression in Escherichia coli by Lrp and DNA methylation: model building based on analysis of pap. *Molecular Microbiology*, 6(17):2429–2435, 1992.

[53] Hans Peter Lehr. *Entwicklung eines fluoreszent - optischen Evaneszenzfeldmesssystems für die Echtzeitanalyse von DNA - Microarrays*. Dissertation, Albert Ludwigs Universität Freiburg im Breisgau, 2002.

[54] B. Moreira, Y. You, M. Behlke, and R. Owczarzy. Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochemical and Biophysical Research Communications*, 327(2):473–484, February 2005.

[55] H. Gowher, O. Leismann, and A. Jeltsch. DNA of Drosophila melanogaster contains 5-methylcytosine. *The EMBO Journal*, 19(24):6918–6923, 2000.

[56] S. E. Aiyar, S. M. McLeod, W. Ross, C. A. Hirvonen, M. S. Thomas, R. C. Johnson, and R. L. Gourse. Architecture of Fis-activated transcription complexes at the Escherichia coli rrnb p1 and rrne p1 promoters. *J. Mol. Biol.*, 316:501–516, 2002.

[57] H. T. Allawi and J. Santalucia. Nearest-neighbor thermodynamics of internal A-C mismatches in DNA: Sequence dependence and pH effects. *Biochemistry*, 37:9435–9444, 1998.

[58] H. T. Allawi and J. Santalucia. Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry*, 36:10581–10594, 1997.

[59] B. A. Baker and V. T. Milam. Hybridization kinetics between immobilized double-stranded DNA probes and targets containing embedded recognition segments. *Nucleic Acids Research*, pages 1–13, 2011.

[60] Oliver Ehler. *Analyse des Informationsgehalts eines DNA - Arrays*. Diplomarbeit, Universität Bayreuth, 2007.

[61] D. J. Fish, M. T. Horne, and R. P. Searles. Multiplex SNP discrimination. *Biophysical Journal*, 92:L89–L91, 2007.

[62] R. S. Gitan, H. Shi, C. M. Chen, P. S., and T. Hui-Ming Huang. Methylation-specific oligonucleotide microarray: A new potential for high-throughput methylation analysis. *Genome Research*, 12:158–164, 2001.

[63] E. Hering and R. Martin. *Photonik - Grundlagen, Technologie und Anwendung*. Springer Berlin Heidelberg, 2006.

[64] G. Heusipp, S. Fälker, and M. A. Schmidt. DNA adenine methylation and bacterial pathogenesis. *International Journal of Medical Microbiology*, 297:1–7, 2007.

[65] J. J. Hopfield. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specifity. *Proc. Natl. Acad. Sci. USA*, 71(10):4135–4139, 1974.

[66] R. Jaenisch and A. Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics supplement*, 33:245–254, 2003.

[67] Jona Kayser. *Parallele Messverfahren mit DNA Microarrays*. Master's thesis, Universität Bayreuth, 2008.

[68] C. Kittel. Phase transition of a molecular zipper. *American Journal of Physics*, 37(9):917–920, 1969.

[69] M. Kuhlmann and W. Nellen. RNAinterferenz. *Biologie unserer Zeit*, 3:142–150, 2004.

[70] P. Landini and S. J. W. Busby. Expression of the Escherichia coli Ada regulation in stationary phase: Evidence for rpoS-dependent negative regulation of alkA transcription. *Journal of Bacteriology*, 181:6836–6839, 1999.

[71] C. Li, J.-C. Hus, L. J. Sun, and P. Zhou. A methylation-dependent electrostatic switch controls DNA repair and transcriptional activation by E. coli ada. *Molecular Cell*, 20:117–129, 2005.

[72] S. Li, A. Pozhitkov, and M. Brouwer. A competitive hybridization model predicts probe signal intensity on high density DNA microarrays. *Nucleic Acids Research*, 36:6585–6591, 2008.

[73] M. A. Lieberman. The Langmuir isotherm and the standard model of ion-assisted etching. *Plasma Source Science and Technology*, 18, 2009.

[74] A. Lobner-Olesen, O. Skovgaard, and M. G. Marinus. Dam methylation: coordinating cellular processes. *Current Opinion in Microbiology*, 8:154–160, 2005.

[75] Wolfgang Michel. *Untersuchungen an Biopolymeren, Charakterisierung von vernetzten Aktinlösungen, Untersuchung der DNA Hybridisierung an Glasoberflächen*. Dissertation, Universität Bayreuth, 2008.

[76] H. M. Moody, M. H. P. van Genderen, L. H. Koole, H. J. M. Kocken, E. M. Meijer, and H. M. Buck. Regiospecific inhibition of DNA duplication by antisense phosphate-methylated oligodeoxynucleotides. *Nucleic Acids Research*, 17:4769–4782, 1989.

[77] N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. Santalucia. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A-A, C-C, G-G, and T-T mismatches. *Biochemistry*, 38:3468–3477, 1999.

[78] S. de los Rios and J. J. Perona. Structure of the Escherichia coli leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *J. Mol. Biol.*, 366:1589–1602, 2007.

[79] Y. Savir and T. Tlusty. Conformational proofreading: The impact of conformational changes on the specificity of molecular recognition. *PLoS ONE*, 2(5):e468, 2007.

[80] J. Scheuermann, F. Viti, and D. Neri. Unexpected observation of concentration-dependent dissociation rates for antibody-antigen complexes and other macro-

molekular complexes in competition experiments. *Journal of Immunological Methods*, 276:129–134, 2003.

[81] E. Schneider, G. Pliushch, N. El Hajj, and D. Galetzka. Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. *Nucleic Acids Research*, 38:3880–3890, 2010.

[82] U. V. Schneider, J. K. Severinsen, I. Géci, L. M. Okkels, N. Jøhnk, and N. D. Mikkelsen. A novel FRET pair for detection of parallel DNA triplex by the LightCycler. *BMC Biotechnology*, 10(4):1–12, 2010.

[83] Holger Schöneborn. *Selbstreplizierende Oligonucleotide: Zum Einfluss der Matrizenlänge bei kinetischen Untersuchungen auf der Basis des Fluoreszenz-Resonanz-Energie-Transfers*. Dissertation, Ruhr-Universität Bochum, 2002.

[84] A. S. N. Seshasayee. An assessment of the role of DNA adenine methyltranferase on gene expression regulation in E. coli. *PLoS One*, 3:e273, 2007.

[85] L. Shen and R. A. Waterland. Methods of DNA methylation analysis. *Current Opinion in Clinical Nutrition and Metabloic Care*, 10:576–581, 2007.

[86] Jonghyeon Shin. *Molecular programming with a transcription and translation cell-free toolbox: from elementary gene circuits to phage synthesis*. PhD thesis, University of Minnesota, 2012.

[87] J. Shin and V. Noireaux. Efficient cell-free expression with the endogenous E. coli RNA polymerase and sigma factor 70. *Journal of Biological Engineering*, 4(8):1–9, 2010.

[88] A. J. Turberfield and J. C. Mitchell. DNA fuel for free-running nanomachines. *Physical Review Letters*, 90:118102, 1–4, 2003.

[89] Y. Zeng, A. Montrichok, and G. Zocchi. Length and statistical weight of bubbles in DNA melting. *Physical Review Letters*, 91(14), September 2003.

[90] Y. Zhang, D. A. Hammer, and D. J. Graves. Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophysical Journal*, 89:2950–2959, 2005.

List of publications

- Marc Schenkelberger, Christian Trapp, Timo Mai, and Albrecht Ott. Competition for binding entails ultrahigh molecular recognition specificity. *to be published*.
- Christian Trapp, Marc Schenkelberger, and Albrecht Ott. Stability of double-stranded oligonucleotide DNA with a bulged loop: a microarray study. *BMC Biophysics*, 4(20):1-13, 2011.
- Marc Schenkelberger, Siba Shanak, Volkhard Helms, Vincent Noireaux, and Albrecht Ott. The MBD of MeCP2 stabilizes the Z-conformation of double-stranded DNA. *to be published*.
- Marc Schenkelberger, Jonghyeon Shin, Volkhard Helms, Vincent Noireaux, and Albrecht Ott. In vitro expression analysis of the pap epigenetic switch reveals unexpected behavior patterns. *in preparation*.