DNA Denaturation Mapping in Nanochannels for Bacterial Identification

Master Thesis

M.S. in Physics, Nanoscience

Stefano Scaramuzza

Supervisors:
Prof. Jonas Tegenfeldt
Dr. Jason Beech

21. December 2014

Department of Physics
Division of Solid State Physics

Lund University

Preface

During this Master dissertation, I got the wonderful opportunity to spend two semesters on exploring partial denaturation mapping of DNA in nanochannels. The work was part of a bigger project with the long-term objective to develop a diagnostic tool for bacterial infections. The project combines various scientific disciplines including physics, nanotechnology, biology and computation. Contributing to the project are the research groups of Jonas O. Tegenfeldt, Lund University carrying out the experimental research; Tobias Ambjörnsson, Lund University, dealing with the data analysis, Sarah Brovall from the Birgitta Henriques Normark and Staffan Normark Group at Karolinska Institutet, Stockholm providing the bacterial DNA; and finally Joachim Fritzsche from Chalmers University fabricating the nanochannel devices. I further had the chance to present the progress of the project as a poster at the Micronano System Workshop in Uppsala (May 2014). It was a pleasure for me to work on this highly interdisciplinary and challenging thesis which enabled me to gain a wide knowledge in various scientific disciplines, and, in particular, motivated me to pursue a career as a researcher.

Stefano, 21. December 2014

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor Jonas O. Tegenfeldt for providing me the opportunity to work on this fascinating project, for his support and for being always open for my own Ideas. I am also deeply grateful to my co-supervisor Jason P. Beech for the technical assistance as well as for the creative inputs he has given me throughout the project.

Besides, I am thankful to Tobias Ambjörnsson, Charleston Noble, Erik Lagerstedt and Christoffer Pichler from the Department of Theoretical Physics for the support in the image data analysis and the numerous constructive discussions.

I appreciate my peers at the division for the comfortable and creative working environment without which this journey would not have been so smooth and memorable.

In particular, I would like to thank my parents for realizing my aspiration to pursue my education and research at Lund University, and the constant encouragement thereafter.

Abstract

In the current work, partial denaturation mapping of DNA in nanochannels is studied with the goal to use this technique as a tool for the identification of bacteria. This method has the potential to simplify and speed up the diagnosis of bacterial infections.

A partially denatured and YOYO-1-labeled DNA molecule displays a sequence specific pattern of fluorescent and non-fluorescent regions along the molecule. This pattern can be interpreted as a barcode. By stretching the DNA molecule inside nanofluidic channels, this barcode can easily be imaged with an optical microscope and compared with a database consisting of theoretically generated barcodes. The DNA is identified by finding the theoretical barcode that matches best to the experimental barcode.

In this project, first fundamental studies of the denaturation pattern formation have been performed. Pattern formation has been studied as a function of time, confinement and fluorescent dye concentration. Second, experimentally acquired barcodes were compared to their corresponding theoretical barcodes to demonstrate the agreement between experiment and theory. Finally, identification of the strain specific DNA of *S. pneumoniae* was attempted.

The time it takes for denaturation patterns to form upon heating the DNA has been observed to be approximately 10 minutes. The difference in melting temperature between nano- and microconfinement was found to be in the range of 5-10 °C. An effect of the fluorescent dye concentration on the melting temperature of DNA molecules was observed. Further, a good agreement between experimental and theoretical barcodes could be demonstrated. A large amount of denaturation maps of DNA fragments from different strains of *S. pneumoniae* were measured. However, successful identification of the bacterial DNA has not been realized yet. Potential ways to improve the results are discussed.

Contents

List of Figures

Abbreviations

BME	β-mercaptoethanol
bp	Basepair
DNA	Deoxyribonucleic acid
DRIE	Deep reactive-ion etching
dsDNA	Double-stranded DNA
EBL	Electron-beam lithography
EDTA	Ethylenediaminetetraacetic acid
EMCCD	Electron multiplying charge-coupled device
IS	Information score density
NA	Numerical aperture
PCR	Polymerase chain reaction
RIE	Reactive-ion etching
SEM	Scanning electron microscope
ssDNA	Single-stranded DNA
TBE	TRIS-borate-EDTA
TE	TRIS-EDTA
TRIS	Tris(hydroxymethyl)-aminomethan

Chapter 1

Introduction

Bacterial infections are a common cause of serious diseases and thus a threat to human health. It is crucial to diagnose an infection and identify the causative agent at an early stage in order to prescribe the right medication and thus increase the chance of a successful treatment. A precise diagnosis is particularly important when antibiotics are used. The choice of the right antibiotics from the beginning of a therapy helps reducing the evolution of antibiotic resistant bacteria.

Apart from diagnosis based on symptoms or clinical presentation a common way to identify the bacteria causing the infection is the controlled growth of a bacterial culture using isolated bacteria from the patient. The identification is done based on different characteristics of the culture. Another way to identify bacteria is by doing biochemical tests as, e.g., immunoassays. Polymerase chain reaction (PCR) is used to amplify bacterial DNA *in vitro* allowing a sensitive and fast detection of microbes.

While the conventional diagnostic tools are well established and reliable they also have some disadvantages. Biochemical tests can be expensive and only available for specific pathogens. Microbial cultures, while cheap and well studied, take a long time to grow (from days to weeks) and can be, for some specific bacteria, difficult or even impossible to grow. PCR, while fast and sensitive, requires specific primers (oligonucleotides) corresponding to the tested bacteria. These primers are not available for all existing bacteria. Further, the primers have to be preselected prior to a diagnosis, i.e., it is only possible to test a patient for some selected bacteria while unexpected pathogens may remain undetected.

DNA denaturation mapping in nanochannels has the potential to speed up the diagnosis and increase the sensitivity while keeping the costs low. The basic principle of this technique is to take advantage of the fact that DNA

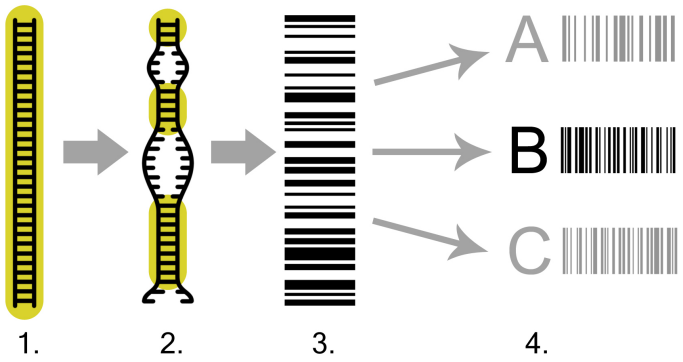


Figure 1.1: Illustration of the procedure of bacterial identification using DNA denaturation mapping. 1.) A fluorescently labeled DNA molecule is stretched out in a nanochannel. At room temperature the fluorescence is homogeneous along the molecule (yellow color). 2.) Heating of the molecule at a specific temperature causes AT rich regions to denature while GC regions stay intact. Denatured parts of the DNA loose their fluorescent dye molecules and become dark. 3.) The bright and dark regions along the molecule can be interpreted as a barcode containing the unique genomic information of the DNA molecule. 4.) The DNA molecule is identified by comparing the experimentally achieved barcode with a database of theoretical barcodes which are generated based on known sequences.

molecules automatically stretch out when being confined close to one dimension. Under these circumstances a labeled DNA molecule is easily observable with a microscope. Usually the intercalating fluorescent dye YOYO-1 is used for labelling the DNA. When heated at a certain temperature, AT rich regions become single-stranded while GC rich regions stay double-stranded. Since the dye molecules only bind to double-stranded DNA, the GC rich regions emit light while the AT rich regions are dark. The result is a barcode-like pattern along the molecule allowing a direct visualization of the genomic information [1]. An identification of the DNA can be made by comparing this experimentally obtained barcode with a database of theoretical barcodes which are generated based on known DNA sequences. The procedure of bacterial identification using DNA denaturation mapping is illustrated in Fig. 1.1.

The advantage of this technique is that in an ideal case only one DNA molecule of the pathogen is needed in order to identify it. Also the equipment is relatively cheap and simple to set up. Furthermore, the identification of the DNA can be automated and is therefore fast. Finally, given a big enough database of theoretical barcodes, every type of bacteria has the potential to be detected.

During the past decade single molecule DNA analysis in nanochannels for different applications has continuously been studied [2]. Alternatively to denaturation mapping other barcoding systems have been developed using

sequence specific labelling [3, 4]. The barcode formation caused by competitive binding of YOYO-1 and the AT-specific antibiotic netropsin has been used to identify a specific *E. coli* bacteria [5].

In this project the focus is on using denaturation mapping of DNA in nanochannels for bacterial identification. Denaturation mapping in nanochannels has already successfully been done in different studies [1, 6, 7]. However, no studies could be found that use denaturation mapping in nanochannels for bacterial identification.

In the first part of this project fundamental aspects of the denaturation pattern formation are studied. This is done by investigating the dependence of the pattern formation on time, confinement and dye load. In the second part experimental barcodes are compared to their corresponding theoretical ones. The gained knowledge from the first two parts is used in the final part, where denaturation maps of DNA of different strains from *Streptococcus pneumoniae*, or *S. pneumoniae*, are measured and attempt is made at identification.

The experiments were performed using nano- and microfluidic devices in combination with fluorescence microscopy. The data analysis was done using various scripts in MATLAB.

Before presenting the results, the two fundamental phenomena used in this thesis are discussed, i.e., the partial denaturation of DNA (see Section 2) and the physics of DNA in nanochannels (see Chapter 3). After describing the most relevant concepts of the method (see Chapter 4) the results are presented and discussed (see Chapter 5).

Chapter 2

Partial Denaturation of DNA

2.1 Introduction

An important phenomenon discussed in this thesis is the partial denaturation of DNA, since it makes it possible to visualize the information stored in a DNA molecule. It is important to know some basics about the structure of DNA before being able to understand partial denaturation. In this chapter first a brief overview of the relevant characteristics of the DNA is made. After that, the partial denaturation is introduced.

2.2 Structure of DNA

The DNA (deoxyribonucleic acid) is a molecule found in all living organisms and some viruses. It contains all the genetic information of its carrier. In order to describe the structure of the DNA it is convenient to start by introducing its building blocks: the nucleotides (see Fig. 2.1). They are molecules with a nitrogenous base (adenine, thymine, cytosine or guanine), a five-carbon sugar 2-deoxyribose and a phosphate.

The DNA is a linear polymer made of nucleotides connected via phosphodiester bonds. The amount of nucleotides varies strongly between different organisms. The sequence of the nucleotides along the DNA molecule encodes the genetic information of the organism carrying the DNA. The part of the molecule formed by the sugars and phosphates is called the backbone.

In most cases the DNA is not a single strand of nucleotides but a pair of two strands (or double strand) bound together via hydrogen bonds as shown in Fig. 2.1. If not explicitly mentioned to be single-stranded, the DNA is considered to be double-stranded. Double-stranded DNA is usually referred to as dsDNA and single-stranded DNA as ssDNA.

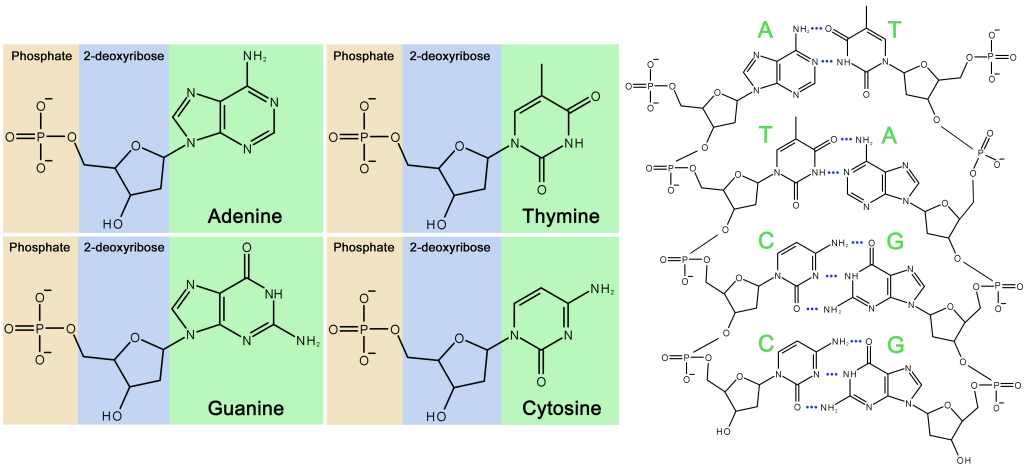


Figure 2.1: Structural formula of the 4 nucleotides found in DNA. Each of them is characterized by its nitrogenous base (adenine, thymine, guanine, cytosine). The structure of a double-stranded DNA molecule is shown in the left. The hydrogen bonds are visualized by the blue dotted lines.

The hydrogen bonds holding together two strands are formed between pairs of nitrogenous bases (see Fig. 2.1). Adenine binds to thymine via 2 hydrogen bonds (AT basepair) while guanine binds to cytosine via 3 hydrogen bonds (GC basepair). Since adenine only binds to thymine and guanine only binds to cytosine the order of the bases of one strand can be deduced from the other strand. The strands of the same DNA molecule are thus complementary and each of them contains the same information. The length of a DNA molecule can be estimated by considering that 1 basepair (bp) is approximately 0.332 nm long [8]. dsDNA is usually found in the so called B-DNA conformation, i.e., the backbone of each strand coils helically around the same axis forming a double helix.

dsDNA is more stable than ssDNA. The stability of dsDNA is partly given by the hydrogen bonds between the bases where AT basepairs, having only 2 hydrogen bonds, are bond weaker than GC basepairs which have 3 hydrogen bonds. A bigger contribution to the stability comes from the stacking interactions, i.e., attractive forces between the aromatic rings of two successive basepairs [8]. The attraction between two GC basepairs is bigger than the one between two AT basepairs. In total GC rich regions are thus more stable than AT rich regions. This is important when it comes to partial denaturation of DNA as discussed in the following section (see Section 2.3).

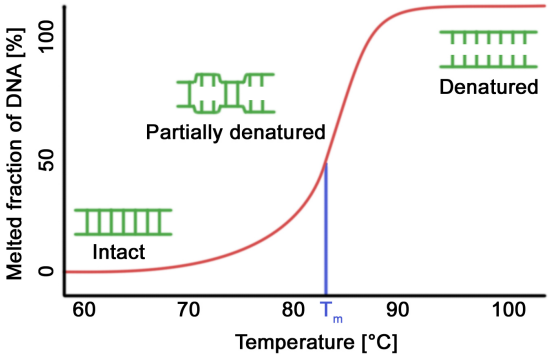


Figure 2.2: Qualitative behavior of a DNA melting curve [8]. At low temperature the molecule is intact. Close to T_m it is partially denatured and at high temperatures it is completely denatured.

2.3 Partial Denaturation

When enough energy in the form of heat is applied to a DNA molecule the hydrogen bonds are broken resulting in two single-stranded DNA molecules. This process is called denaturation or melting. The melting temperature T_m is defined as the temperature at which half of the DNA molecule is melted, i.e., the two strands are separated, and half of it is double-stranded (see Fig. 2.2). The melting temperature depends on the sequence, the length and the environment of the DNA. Typically T_m is roughly between 80 °C and 100 °C [8]. The process of slowly cooling down denatured DNA in order for the single strands to rebind is called annealing.

AT basepairs are less stable than GC basepairs (see Section 2.2). Thus AT basepairs break at lower temperatures than GC basepairs. This is why the transition between intact and denatured DNA is not abrupt but happens within a temperature interval of roughly 10 degrees. During this transition the DNA is partially denatured. When partially denatured, AT rich regions are separated, i.e., single-stranded, and GC rich regions are double-stranded. The conformation of a partially denatured DNA molecule depends therefore on the DNA sequence. By using a specific fluorescent dye (see Section 4.2.2) it is possible to fluorescently label double-stranded parts of the DNA while single-stranded parts stay unaffected. This leads to a pattern of fluorescent and non-fluorescent parts along the partially denatured DNA. This so called *melting pattern* allows to visualize the underlying sequence of the DNA. This will be discussed in more detail in Section 4.2.2.

The amount of GC pairs defines the melting temperature of a DNA molecule. It is however possible to influence the melting temperature by changing the environment of the DNA. Four relevant factors that influence the T_m are:

- Ionic strength of the buffer in which the DNA is suspended.
- Intercalated molecules on the DNA.
- Confinement of the DNA.
- Presence of formamide in the buffer in which the DNA is suspended.

A high ionic strength leads to a high melting temperature because positive ions screen the negative charged phosphates, reducing the repulsion between both strands of the DNA and making it thus more stable [8]. Intercalating molecules, i.e., molecules binding to the DNA between the basepairs, stabilize the DNA and thus result in a rise of the melting temperature [9]. A strong confinement of the DNA increases the melting temperature due to changes of its energy states [10]. The presence of formamide(CH_3NO) lowers the melting temperature for two reasons. First, formamide molecules reduce the ionic strength in close proximity of the DNA. Second, once the bonds between the basepairs are broken, the formamide molecules are able to form hydrogen bonds with the bases, preventing the bases to form pairs again [11]. The difference in T_m caused by the presence of formamide can be estimated with Eq. (2.1) [12].

$$\Delta T_m \approx -0.6 \cdot F \tag{2.1}$$

Here F is the amount of formamide in the buffer in percentage by volume. A common amount of formamide used in experiments is 50% (v/v) leading to a reduction of T_m by $\sim$30 °C.

Chapter 3

Confinement of DNA

3.1 Introduction

Another important concept of this thesis is the physics of DNA molecules under strong confinement, in particular close to one dimension. The confinement of DNA molecules close to one dimension allows to stretch DNA molecules in a simple and controlled way without the use of mechanical interactions or chemicals. This allows to observe DNA molecules in an easy way. In order to describe the behaviour of DNA in confinement some concepts from polymer physics are used. In this chapter first the physics of DNA in 3D is described which is needed in order to discuss the physics under strong confinement. The content of this chapter summarizes the relevant concepts from ref. [1, 2, 13].

3.2 Physics of DNA in 3D

The physics of a DNA molecule is described by considering the molecule to be a chain, which is characterized by the following parameters:

- **Contour length**: End to end length L of the molecule.
- **Persistence length**: Length P along the chain over which it can be considered to be rigid. This tells how stiff the chain is.
- **Effective width**: Width w_{eff} of the chain caused by physical and electrostatic interactions.

The contour length equals to the amount of basepairs of the DNA multiplied by 0.332 nm (see Section 2.2). For T4 phage DNA (see Section 5.1) the contour length is $L \approx 55\,\mu m$. The effective width and the persistence length depend on the ionic strength of the buffer, as explained later on. The effective width ranges from 5 nm for an ionic strength of 0.1 M to 20 nm for an ionic strength of 0.005 M [13]. The persistence length is roughly 50 nm

for ionic strengths above 0.01 M. For ionic strengths below 0.01 M it can become up to 80 nm [13].

Further, the DNA is considered to be self-avoiding, i.e., every segment of a chain occupies a certain volume that can not be occupied by the rest of the chain.

A DNA molecule in free solution tends to coil into a blob of finite volume because of a competition between self-avoidance effects causing the DNA to expand and entropic effects causing the DNA to coil (see Fig. 3.1). The radius of gyration of the coil is in this case expressed by the so called Flory radius R_F in Eq. (3.1).

$$R_F \approx (w_{eff} 2P)^{\frac{1}{5}} L^{\frac{3}{5}} \tag{3.1}$$

It is intuitive that an increase in w_{eff}, P or L also increases R_F.

3.3 Physics of DNA in Confinement

If a DNA molecule is strongly confined it can no longer be described as a coil. In our case, by strong confinement we mean a confinement close to one dimension, i.e., a channel with $\sqrt{\sigma} < R_F$, where σ is the area of the cross-section of the channel. In a channel with a rectangular cross-section, $\sqrt{\sigma}$ corresponds to the geometric average of the width and height of the channel.

The conformation of the DNA depends on the level of confinement. We differ between two regimes. The de Gennes regime where $P < \sqrt{\sigma} < R_F$ and the Odijk regime where $\sqrt{\sigma} < P$.

In the de Gennes regime the DNA is still able to fold back since $P < \sqrt{\sigma}$. Thus a coil formation is possible inside the channel. However, since the radius of a coil has to be smaller than R_F the DNA is modeled as a chain of smaller coils (see Fig. 3.1a) where each coil behaves as it would be without any confinement. The resulting length of the chain of coils, or extension r of the DNA, is expressed by Eq. (3.2).

$$\frac{r}{L} \approx A \left(\frac{w_{eff} P}{\sigma} \right)^{\frac{1}{3}} \tag{3.2}$$

A is a factor of proportionality with a value close to 1. We can tell from Eq. (3.2) that an increase of w_{eff} or P results in an increase of r while an increase of the channel cross-section reduces the extension. Smaller channels thus lead to a bigger extension.

In the Odijk regime the DNA can not fold back any more and thus the

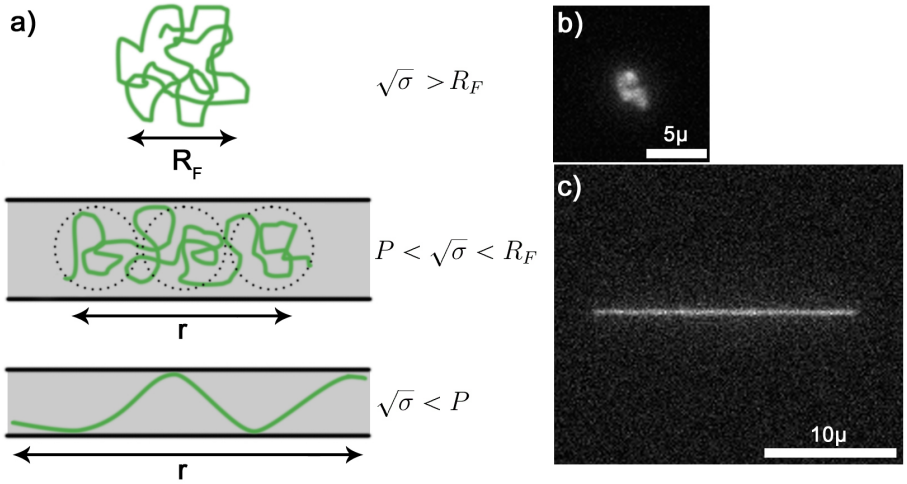


Figure 3.1: (a) Illustration of a DNA molecule in free solution (single coil), in the de Gennes regime (chain of coils) and in the Odijk regime (no coiling). (b) Microscope image of a fluorescently labeled DNA molecule in free solution. (c) Microscope image of a fluorescently labeled DNA molecule extended in a nanochannel in the de Gennes regime.

formation of a coil inside the channel is not possible. Thus the effective with of the DNA is of no importance any more. The DNA molecule undulates along the channel as shown in Fig. 3.1a. The resulting extension r of the DNA in a channel with rectangular cross-section (width w and height h) is expressed by Eq. (3.3).

$$\frac{r}{L} \approx 1 - B\left[\left(\frac{w}{P}\right)^{\frac{2}{3}} + \left(\frac{h}{P}\right)^{\frac{2}{3}}\right] \tag{3.3}$$

$B \approx 0.091$ is a factor of proportionality. Same as in the de Gennes regime an increase of P results in an increase of r. Fig. 3.1b-c shows microscope images of real DNA molecules in free solution and confinement.

Generally the extension r depends also on the environment, namely the ionic strength. The reason for that is that the persistence length and the effective width are both dependent on the ionic strength. A decrease in ionic strength reduces the screening of the negative charged phosphates of the DNA by positive ions. This results in an increases of the electrostatic repulsion of neighboring segments of the DNA leading to an increase of w_{eff} and P which, according to Eq. (3.2) and Eq. (3.3), also increases r.

It is interesting to note that in both regimes the extension r of the DNA is directly proportional to the contour length L. This simplifies the analysis of observational data from DNA in nanochannels.

A very important thing to point out is that once the DNA is extended, it is at equilibrium and its length r will stay constant over time, apart from thermal fluctuations. According to ref. [13] it can be assumed that in the experiments made in this thesis we deal with the de Gennes regime of scaling.

Chapter 4

Materials and Methods

4.1 Introduction

In the previous chapters (chapter 2 and 3) the two most important concepts on which the work of this thesis is based on have been introduced, namely partial denaturation of DNA and stretching of DNA under confinement. These two concepts are experimentally combined by using nanochannels in order to stretch out partially denatured DNA which can then be observed and analyzed with a microscope. How this is realized is explained in this chapter. First, the experimental components are discussed, i.e., the design and fabrication of the nanochannels, the fluorescent dye YOYO-1, which is used to stain the DNA molecules, and the experimental setup including the microscope. Second, the acquisition of a barcode from a DNA molecule is presented step by step. This includes the staining and melting of the DNA, the handling of the setup, the image acquisition and the image processing. The used experimental setup and the method are mostly based on work done in ref. [1, 2, 14].

4.2 Experimental Components

4.2.1 Nanochannels

The nanochannels in which the DNA molecules will be stretched and observed are inside a nanofluidic device made of silicon. We also refer to it as the chip. A schematic of the chip is shown in Fig. 4.1. The main parts of the chip are the nanochannels, the microchannels, the inlets and the glass lid which is used to seal the channels but still keeping them visible from the outside. The liquid DNA sample is introduced into the chip through the inlets and is transported through the microchannels to the nanochannels where it stretches out and is observed. The DNA sample is moved in the channels by applying pressure differences to the inlets (see Section 4.3.3).

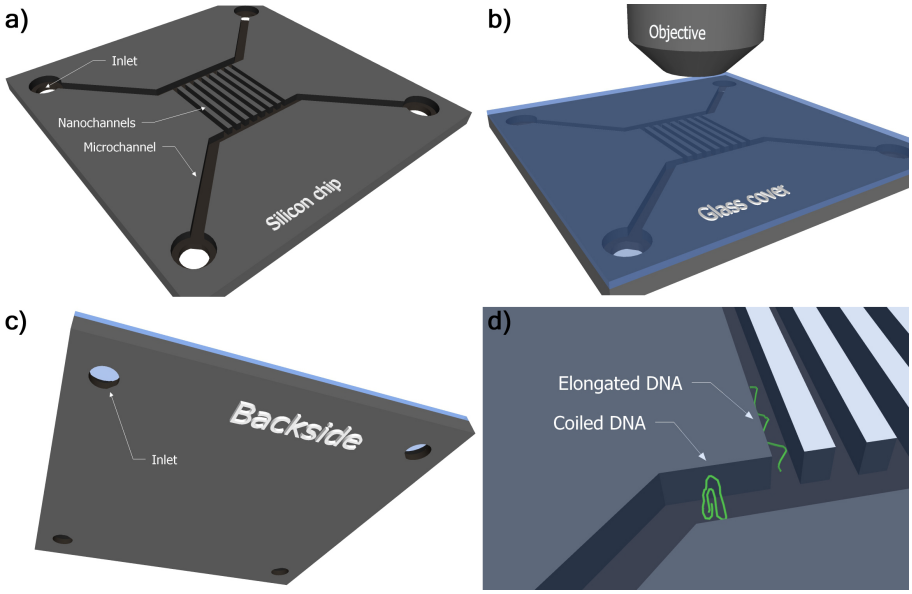


Figure 4.1: Schematic of the nanochannel device. The proportions do not represent the real case. (a) Unsealed silicon chip featuring nano- and microchannels plus inlets. (b) Chip sealed with borosilicate glass ready to be observed. (c) Backside of chip featuring the inlets. (d) Close up of the nano- and microchannels. In the microchannels the DNA is coiled while in the nanochannels it stretches out.

The nanochannels used in this project have a width and depth of 150 nm while the microchannels have a width of 50 µm and a height of 150 nm.

The chip is made of silicon and is fabricated using conventional techniques from semiconductor device fabrication. There are many different ways to fabricate nanochannels [2]. In the following the main steps of the procedure for the fabrication of the devices used in this thesis is presented[1]:

1. **Thermal oxidation**: Wet oxidation of a silicon wafer at high temperature. The channels will be etched into the formed oxide layer.

2. **Reactive-ion etching (RIE) of nano- and microchannels**: The wafer is spin coated with resist. The resist is exposed by electron-beam lithography (EBL) for the nanochannels and optical lithography for the microchannels. After development of the resist the channels are formed by RIE.

3. **Deep reactive-ion etching (DRIE) of the inlets**: First an aluminum mask is sputtered onto the wafer. Then the mask is spin coated with resist, exposed optically in the areas of the inlets and developed. After wet etching the aluminum, the hard mask is ready and the inlets

[1]The fabrication was done by Joachim Fritzsche at Chalmers University.

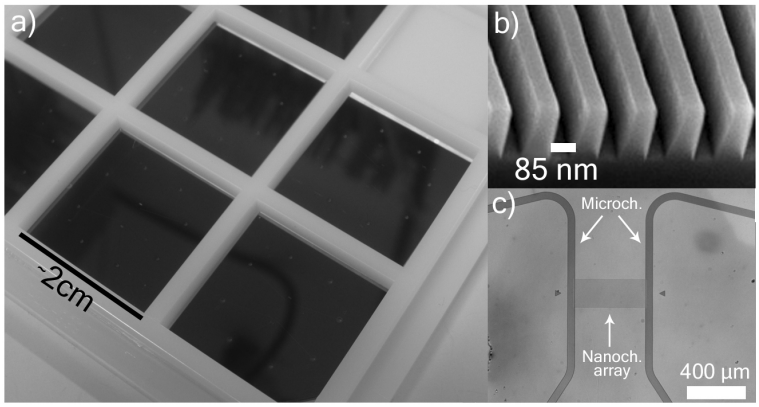


Figure 4.2: (a) Nanochannel devices. 8 inlets per chip are visible since each chip contains 2 separate nanochannel arrays with the corresponding microchannels and inlets. (b) SEM image of a nanochannel array (this design was not used in this project). (Image: Courtesy of Stephen Y. Chou, NanoStructure Laboratory, Princeton University) (c) Brightfield image of the nano- and microchannels (top view). The nanochannels are too small to be distinguished individually.

can be etched by DRIE. After the etching the inlets, the hard mask is removed.

4. **Fusion bonding of glass wafer**: The channels are sealed by fusion bonding of the silicon wafer with a borosilicate glass wafer.

The chip and some scanning electron microscope (SEM) pictures of nanochannels are shown in Fig. 4.2. A description of how chips are used is given in Section 4.2.3.

4.2.2 Fluorescent Dye

Bare DNA molecules in free solution are not visible under an optical microscope. In order to make DNA visible it is usually labeled with a fluorescent dye. There are many different types of fluorescent dyes. In this project the synthetic cyanine dye YOYO-1 is used.

YOYO-1 molecules are bis-intercalating, i.e., the two planar chromophores of the molecules bind between two neighbouring basepairs of the DNA. When binding in this mode, there is a space of 2 basepairs between the two chromophores. Hence, each dye molecule occupies 4 basepairs in total. The maximal ratio of dye molecule per basepairs, also denoted as dye:bp ratio, is thus 1:4. If this ratio is exceeded, the dye molecules might bind in different and unexpected modes [1].

In free solution the maximum absorption of the molecules is at 458 nm and

the maximum emission is at 564 nm. When bound to DNA the fluorescence intensity is increased by a factor of 3200 and the energy states of the molecules change resulting in an absorption maximum at 489 nm and emission maximum at 509 nm [15]. Figs. 3.1b-c show DNA molecules labeled with YOYO-1.

The dye has two notable effects on the DNA. First, as discussed in Section 2.3, it stabilizes the DNA leading to an increases in melting temperature. Second, it also leads to an increases in contour length since the molecules occupy space between the basepairs. The increase in length is 0.51 nm per intercalated dye molecule [16].

On ssDNA the YOYO-1 molecules either behave like in free solution or they do not bind at all [17]. In both cases the fluorescence intensity would be low compared to when bound to dsDNA. This is important since it allows to distinguish between single- and double-stranded parts of a DNA molecule. In a partially denatured DNA molecule (see Section2.3) the double-stranded parts would fluoresce strongly while the single-stranded parts would be almost not detectable. In other words GC rich regions would appear bright and AT regions would stay dark. This leads to a unique pattern of bright and dark regions along the molecule which can be interpreted as a barcode.

4.2.3 Experimental Setup

The fluorescently labeled DNA was imaged using a fluorescence microscope (Eclipse Ti, Nikon) with a 100x oil immersion objective (Plan Apo VC, NA 1.4, Nikon). In an ordinary optical microscope, light that is either transmitted through or scattered from a sample is observed. In a fluorescence microscope however, the fluorescent sample is excited at a certain wavelength and only the emitted light from the sample is detected. To do that, a specific light filter system is used. The optical light path and the filter system of a fluorescent microscope is depicted and described in Fig. 4.3. The used light engine was a Sola SM lamp (Lumencor) and the image acquisition was done with an iXon EMCCD camera (Andor). The recorded images are greyscale images where the grey value corresponds to the detected light intensity. The optical resolution of a microscope is estimated by Eq. (4.1).

$$d = \frac{0.61\lambda}{NA} \tag{4.1}$$

Here, λ is the emission wavelength and NA is the numerical aperture of the objective. In the used setup the resolution is therefore estimated to be $d = 220\,\text{nm}$ or, more relevant for DNA studies, $d = 653\,\text{bp}$. The size of the field of view using this setup is $80 \times 80\,\mu\text{m}$.

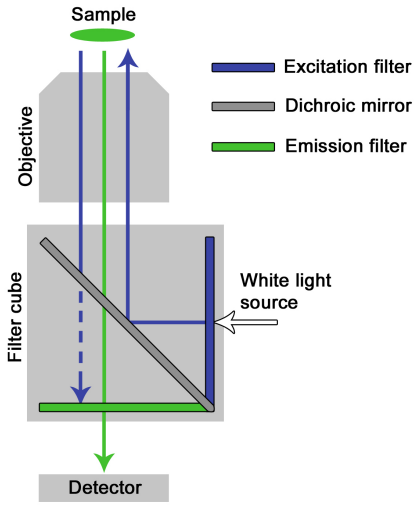


Figure 4.3: Optical lightpath of a fluorescence microscope. White light, generated by a light engine, is filtered by the excitation filter which is transparent only for wavelengths that excite the sample (here around 490 nm). Then the light hits a dichroic mirror which reflects the excitation light but is transparent to the emission light of the sample. The emitted light from the sample goes through the diachronic mirror and is filtered by the emission filter which blocks any light not originating from the fluorescence of the sample (mostly scattered excitation light). This way only emission light from the sample (here around 510 nm) hits the detector.

As stated in Section 4.2.1, the DNA molecules inside the chip are moved by applying pressure differences to the inlets. As an interface between the pressure control unit and the chip a device holder made of plastic was used. Fig. 4.4 shows a schematic and pictures of the device holder. The device holder also allows to place the chip on the microscope stage and contains an access to the chip for a cartridge heater which can be used to heat and partially denature DNA directly on the chip. How the DNA is moved and how the cartridge heater is used is described in more detail in the next section (see Section 4.3).

4.3 Method — From Pure DNA to Barcodes

4.3.1 Sample Preparation

Prior to every experiment a DNA sample had to be prepared. This included staining of the DNA with YOYO-1 and adding specific chemicals to the buffer in which the DNA is suspended.

The provided DNA was already purified and no additional purification steps were needed. As a buffer for the DNA the chemical TRIS-Borate-EDTA (TBE) was used. TRIS stands for Tris(hydroxymethyl)-aminomethan. The

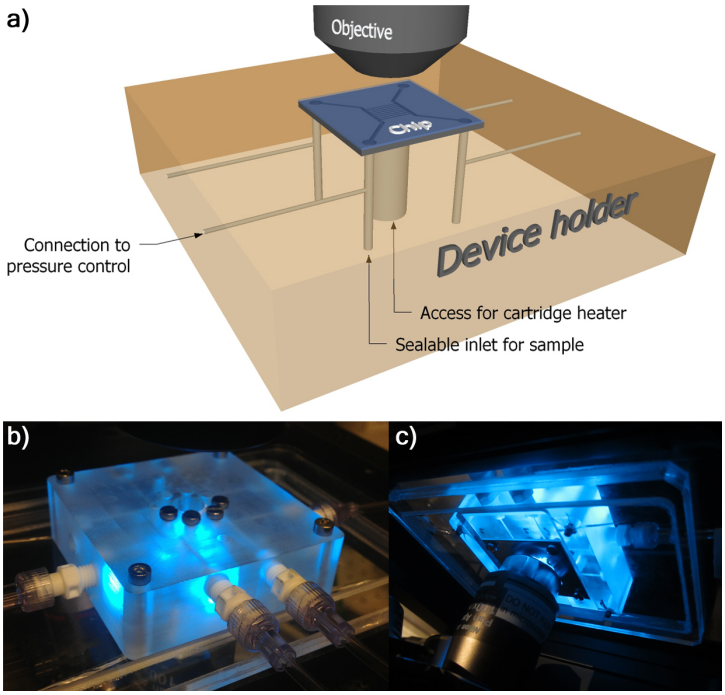


Figure 4.4: (a) Illustration of the device holder. (b) Device holder in use. The inlets are sealed with screws. O-rings are placed between the screws and the holder. The connection of the device holder with the pressure control is established with tubes. (c) Backside of device holder. The chip is mounted to the device holder with a metal steel frame which is screwed to the holder. At the inlets between chip and holder o-rings are placed. In this picture the chip is covered by the oil immersion objective which is almost in contact with the chip.

Ethylenediaminetetraacetic acid (EDTA) in the TBE scavenges multivalent metal ions which are common cofactors for enzymes. A reduction of these ions thus minimizes the enzymatic degradation of the DNA. The concentration of the TBE is usually stated as $A\times$TBE, where A stands for the concentration. A common concentration is $1\times$TBE, which stands for 89 mM TRIS, 89 mM Boric acid and 2 mM EDTA. A high TBE concentration leads to a high salt concentration, i.e., a high ionic strength, of the buffer. Additionally to the TBE, 3% (v/v) of β-mercaptoethanol (BME) was added to the buffer. BME reduces photobleaching (photo-chemical damage of the dye molecules) and photoinduced damage of the DNA. BME scavenges the oxygen in the buffer. Oxygen is known to favor photobleaching and photodamaging [18, 19]. BME also slightly increases the ionic strength of the buffer [20].

The DNA was stained with YOYO-1 at dye:bp ratios above 1:4 (see Section 4.2.2). A common dye:bp ratio was 1:10. The staining of the DNA was done at rather high salt concentrations, e.g., $2\times$TBE. A high salt concentration allows the dye molecules to diffuse easily along the DNA allowing a uniform staining of the molecule [21].

After the dye was added to the DNA, the sample was heated at 40 °C for 30 minutes in order to enhance the diffusion of the dye molecules along the DNA molecule, leading to a uniform staining. Then 50% (v/v) formamide was added in order to lower the melting temperature of the DNA. After this step, the DNA sample was ready to be partially denatured.

A detailed protocol of the DNA sample preparation is presented in Appendix C.

4.3.2 Partial Denaturation

To create the melting patterns the stained DNA had to be heated at a specific temperature for a specific time. We differ between two techniques: *On-chip heating* and *off-chip heating*.[2] This two techniques are explained in the following

In on-chip heating the DNA was heated directly inside the chip. The prepared DNA sample was introduced into the chip through the inlets of the device holder at room temperature. The chip containing the DNA was then heated by a cartridge heater with thermocouple (336 Temperature Controller, Lake Shore Cryotronics) which was introduced into the device holder and brought into contact with the chip. In order to optimize the thermal

[2] Off-chip heating was mainly developed by J. P. Beech in J. O. Tegenfeldt's group at Lund University.

conductivity a thermal paste was applied between chip and cartridge heater. According to different test measurements it took < 1 minute for the chip to reach a stable temperature. Further, the temperature difference between the inlets and the area of the nanochannels was $\sim 10\,^{\circ}\mathrm{C}$. This allowed to partially denature DNA inside and close to the nanochannels while the DNA that was still in the inlets remained unaffected by the heat.

In off-chip heating the prepared DNA sample was heated inside a PCR machine. Immediately after the heating the sample was quickly cooled down by adding a relatively big volume of ice-cold milliQ water (containing 3% (v/v) of BME) to the sample. This quenching procedure serves different purposes. First, it quickly cools down the DNA ensuring that the DNA will not properly reanneal but stay in a metastable state, conserving the denaturation patterns that have been created during the heating. Second, it reduces the concentration of the dye molecules that have fallen off the DNA during the melting, decreasing the probability that they would rebind to the DNA. Third, it reduces the salt concentration. This slows down the diffusion of the dye molecules along the DNA and also increases the extension of the DNA once they are inside the nanochannels (See section 3.3). Once the DNA sample was cooled down and diluted it was ready to be inserted into the chip and to be observed.

While the on-chip heating has various advantages, such as good temperature control, it also has disadvantages. For example it is not feasible for the measurement of a big amount of molecules ($\sim$1000) in short time, which would be used for the bacterial identification, because it requires waiting steps between each movie acquisition in order for the DNA to partially denature. Furthermore, a drop in image quality with time has been observed during the experiments performed in this thesis. The increased temperature of the chip or the high concentration of formamide (no dilution during cooling step) are assumed to cause these problems. Because of these disadvantages on-chip heating is only used for fundamental studies of partial denaturation while off-chip heating is used for the bacterial identification.

4.3.3 Moving and Imaging the DNA

After inserting the DNA into the chip it had to be transported from the inlets through the microchannels into the nanochannels where it was observed. The sample was moved by applying pressurized nitrogen to the inlets at different pressures. Fig. 4.5 describes in detail how the DNA is moved inside the channels. Nitrogen was used because oxygen is known to enhance the damage of stained DNA (see Section 5.2.3). To reduce the amount of oxygen even more, it was important to wait for $\sim$1 hour after introducing the DNA into the chip in order to let the oxygen diffuse out of the sample.

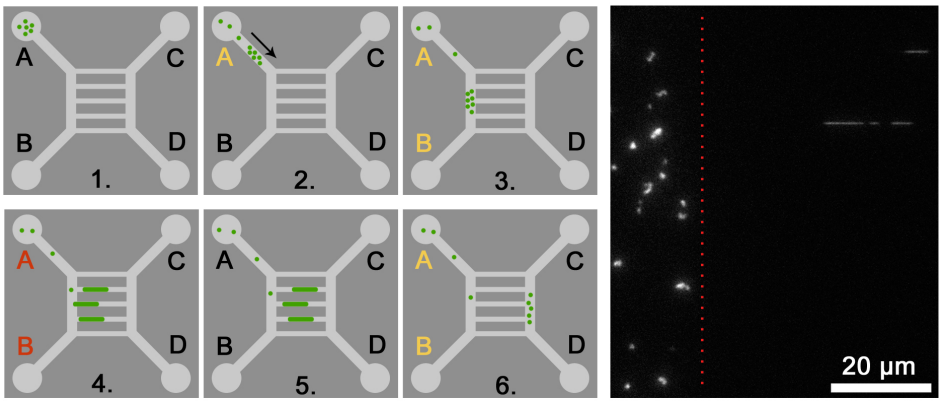


Figure 4.5: (Left) Illustration on how DNA is moved inside the chip. 1.) Sample is introduced in inlet A. 2.) Low pressure (300mbar) is applied to inlet A, making the DNA move through the microchannels towards the nanochannels. 3.) Low pressure is applied to inlet A and B. The fluid moves through the nanochannels. For the DNA it is energetically more favourable to stay in the microchannels than to go into the nanochannels. Hence the DNA accumulates in front of the nanochannels. 4.) A short pulse of high pressure (2000mbar) applied to A and B, forces the DNA to enter the nanochannels. 5.) Once the DNA is in the nanochannels all pressure is turned off. The DNA stretches out and is imaged. 6.) The imaged DNA is flushed away by applying again low pressure to A and B. New DNA is imaged by repeating all these steps. (Right) Fluorescence microscope image of DNA in the chip. In the microchannels (left to dotted line) the DNA is coiled while in the nanochannels (right to dotted line) it is stretched out.

The DNA was imaged by recording a time lapse of 10 or 20 seconds at a frame rate of 10 frames per second. Since many nanochannels fit into the field of view of the microscope, several molecules could be imaged simultaneously. How the images look like and how they were processed is discussed in the next section.

4.3.4 Image Processing and Analysis

A single frame from a recorded movie of partially denatured DNA in nanochannels is shown in Fig. 4.6a. DNA molecules with melting patterns are clearly visible. The time development of a single DNA molecule is displayed by a so called kymograph (see Fig. 4.6b). A kymograph is created by first averaging the intensity of the DNA molecule over its width, leading to a one dimensional intensity profile. This is done for each frame of the movie. The intensity profiles of all frames are then placed next to each other, allowing to visualize the timetrace of the molecule over time in one single image.

From the kymograph in Fig. 4.6b it can be seen that the molecule slightly moves and stretches on short timescales due to thermal motion. In order to eliminate this effect from the kymographs, they are aligned using a computa-

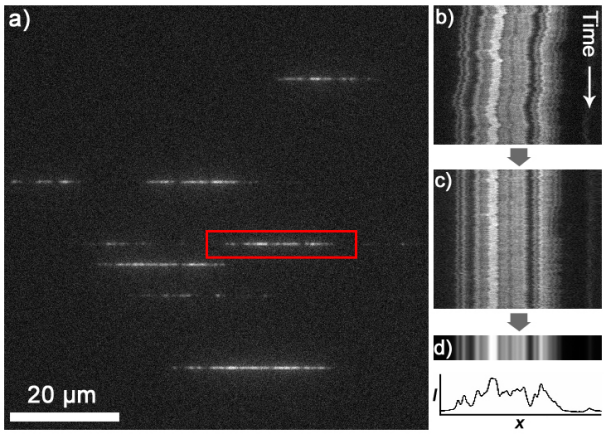


Figure 4.6: (a) Partially denatured T4GT7 Phage DNA extended in nanochannels. The denaturation patterns along the molecules are clearly visible. This image shows one frame out of a 15 seconds long recording. (b) Kymograph of the red marked molecule. Here the denaturation patterns are even better visible. The extension of the molecule varies slightly with time due to thermal motion. (c) Aligned kymograph. (d) Time average of the aligned kymograph (barcode) and its intensity profile.

tional method from ref. [22]. The basic principle of the alignment algorithm is to first detect prominent features on the kymograph, i.e., very bright or dark "columns" of the kymograph. Then the average position over time of this features is calculated. At every time step of the kymograph the position of the feature is then shifted towards its average position. This is done for each feature. When doing that, the pixels between the features are stretched or compressed linearly. By averaging the features over time, the thermodynamic equilibrium of the molecule is found. Fig. 4.6c shows an aligned kymograph. Finally, a time average of the aligned kymograph is computed, resulting in a one dimensional intensity profile (or barcode) of the partially denatured molecule at equilibrium (see Fig. 4.6d). The final intensity curve is ready to be compared with a theoretical intensity curve.

A theoretical barcode is generated by using the known sequence of the DNA, which has been determined using conventional sequencing methods. The Fixman-Freire approximation of the Poland–Scheraga model for the helix–coil transition of DNA is used in order to calculate the melting probabilities for each basepair in the sequence [23]. This model is based on a one-dimensional Ising model with two components, one for unbound basepairs and one for bound basepairs. After calculating the probabilities, the unit for the length of the DNA is translated from bp to nm by considering that single-stranded parts of the DNA are shorter than double-stranded parts [7]. Finally, the probability curve is smoothed by convoluting it with a Gaussian curve with a standard deviation that corresponds to the resolution of the microscope (see Section 4.2.3). The result can be interpreted as an

intensity curve of the expected fluorescence along the molecule.

Before comparing the theoretical intensity curve with the experimental one, the curves have to be rescaled because they have different and arbitrary intensity scales. It is common in image analysis and signal recognition to rescale compared curves in a way that both have mean 0 and standard deviation 1. The rescaling of a curve $f(x)$ with standard deviation σ_f and mean $\overline{f}$ is shown in Eq. (4.2).

$$f_r(x) = \frac{f(x) - \overline{f}}{\sigma_f} \tag{4.2}$$

The detection of the molecules in the recorded movies, the generation and alignment of the kymographs as well as the generation of a theoretical barcode is done with a collection of custom-written MATLAB (MathWorks) scripts named *DNABarcodeMatchmaker*, which is developed and constantly improved by T. Ambjörnsson et al.[3] This software is also used for the bacterial identification, which will be discussed in more detail in Section 5.4.2. Also separate MATLAB scripts are used for the data analysis. These scripts, if not specifically referred to be part of *DNABarcodeMatchmaker*, have been written by the author of this thesis himself.

[3]The main contributors of *DNABarcodeMatchmaker* are T. Ambjörnsson, S. Hammarberg, E. Lagerstedt, A. Nilsson, C. Noble, C. Pichler and M. Reiter-Schad from the Department of Theoretical Physics, Lund University.

Chapter 5

Results and Discussion

5.1 Introduction

In this chapter the results of the performed experiments are presented and discussed. The chapter is divided into three main parts.

In the first part (see Section 5.2), fundamental characteristics of partial denaturation are studied. The goal of this part is to gain more knowledge about the denaturation mapping of DNA in nanochannels. The timescale of the melting pattern formation is investigated as well as the influence of confinement on the pattern formation. Further, the influence of the YOYO-1 dye load on the DNA in off-chip heating is studied.

In the second part (see Section 5.3), experimental barcodes are compared to the corresponding theoretical barcodes. The aim is to show the agreement between experiment and theory. A good agreement is necessary for the later identification of bacteria. A comparison of experimental and theoretical barcodes is made at a single temperature as well as at multiple temperatures, leading to the creation of an experimental temperature dependent denaturation map.

Finally, in the third part (see Section 5.4), the gained knowledge from part one and two is used to measure denaturation maps of DNA from *S. pneumoniae* and attempt is made at identifying the measured DNA using *DNABarcodeMatchmaker*.

In the experiments from the first two parts the DNA from T4GT7 phage was used. T4GT7 phage is a derivate of the T4 phage which is a virus that infects *Escherichia coli* bacteria. The T4GT7 DNA is particularly suitable for the type of investigation performed in these experiments because its sequence is known and monodisperse and purified solutions of it are commercially

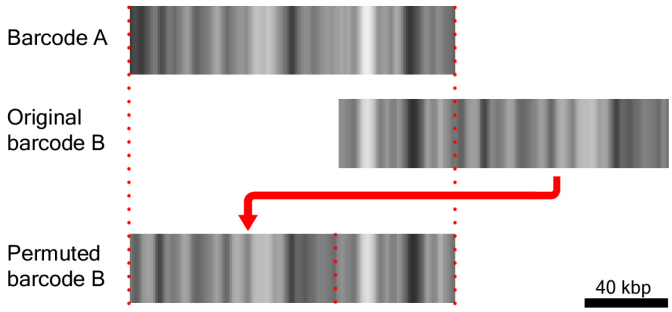


Figure 5.1: Comparison of experimentally acquired barcodes of two different and complete T4GT7 phage DNA molecules. By circularly permuting the barcode B it can easily be aligned with barcode A.

available (Nippon Gene, Toyama, Japan). Further, its total length of 166 kbp [24] fits nicely into the field of view of the used microscope, even when using high magnification objectives. A notable characteristic of this DNA is that its sequence is circularly permuted as shown in Fig. 5.1. In order to simplify the notation the DNA from T4GT7 phage will from now on be called T4 DNA.

5.2 Fundamental Studies of Melting Pattern Formation

5.2.1 Time Dependence of Pattern Formation

To better understand the timescale of the denaturation pattern formation several on-chip measurements have been performed. The found timescale will be used to set a minimal heating time for the partial denaturation of DNA.

T4 DNA was used that has been stained with YOYO-1 at a dye:bp ratio of 1:10. The sample contained 50% (v/v) formamide and 3% (v/v) BME. The final TBE concentration was 0.18×TBE which, using ref. [20], corresponds to an estimated ionic strength of 0.02 M.

The on-chip heating was done in a specific order. First, the chip was loaded with the sample at room temperature. Then, still at room temperature, the DNA was accumulated in the microchannel close to the nanochannels. After that, the temperature of the cartridge heater was set to 50 °C which corresponds to an estimated temperature of 45 °C inside the chip. While keeping the temperature of the cartridge heater constant at 50 °C, every ∼2 minutes a small amount from the accumulated DNA was inserted into the nanochannels and a movie of 10 seconds was recorded. After ∼20 minutes the accumulated DNA in the microchannel was flushed away and the heating system was switched off for about half an hour in order for it to cool down to room temperature. This experiment was then repeated three more times to collect enough data. In total 296 molecules were imaged at different times during the heating.

According to Eq. (2.1) the used heating temperature of 45 °C with 50% (v/v) of formamide corresponds to 75 °C without formamide. 75 °C is although ∼6 °C higher than the temperature where one would expect denaturation patterns to be observed without using formamide (compare Fig. 5.2). Nonetheless, clear denaturation patterns could be seen in the experiment. This temperature difference could be explained by the fact that YOYO-1 stabilizes the DNA, resulting in a rise of the melting temperature [9]. Another reason for the temperature difference might be that the theoretical model assumes the DNA to be in free solution while the DNA inside the microchannels is slightly confined. The change in melting temperature due to confinement will be discussed in more detail in section 5.2.2. Finally, the ionic strength and the temperature inside the chip have not been directly measured but estimated, leading to a possible small systematic error.

The results of the measurements are visualized in Fig. 5.3 showing ky-

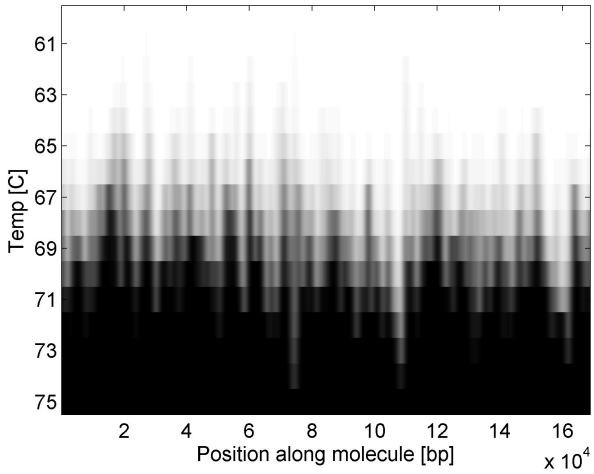


Figure 5.2: Theoretical temperature dependent denaturation map of T4 DNA. This map shows the theoretical barcodes of T4 DNA heated at different temperatures without Formamide and at an ionic strength of 0.02 M. Black corresponds to fully melted DNA and white correspond to fully double-stranded DNA. Melting patterns are visible roughly between 64 °C and 72 °C.

mographs of different T4 DNA molecules at different times of heating. The denaturation patterns appear already within the first minutes of heating, although the contrast of the patterns is not very strong in this early stage. After 10 minutes the quality of the patterns seems to stabilize.

To have a closer look at the denaturation pattern formation over time the intensity profile of the barcodes from selected molecules at different times of heating is plotted in Fig. 5.4. The intensities are background-corrected, making them comparable. It can be seen that after 9 minutes the pattern formation stabilizes.

A good way to quantify the pattern quality of a barcode is by using the so called *information score density* or simply *information score* (IS) defined in Eq. (5.1) [5, 22].

$$\mathrm{IS} = -\sum_{k} \log\left(\frac{1}{\sqrt{2\pi\log(\sigma^2+1)}} \cdot \exp\left(-\frac{\log(|\Delta I(k)|)^2}{2\log(\sigma^2+1)}\right)\right) \cdot \frac{1}{r} \qquad (5.1)$$

Here $\Delta I(k)$ is the difference in intensity between two prominent peaks and valleys of the intensity profile of the barcode. A peak is considered to be prominent if it can be distinguished from the background noise (σ^2 is the variance of the background noise). The division by r, the barcode length, is made in order to be able to compare the IS of molecules with different length. The more distinct peaks and valleys an intensity profile has, the

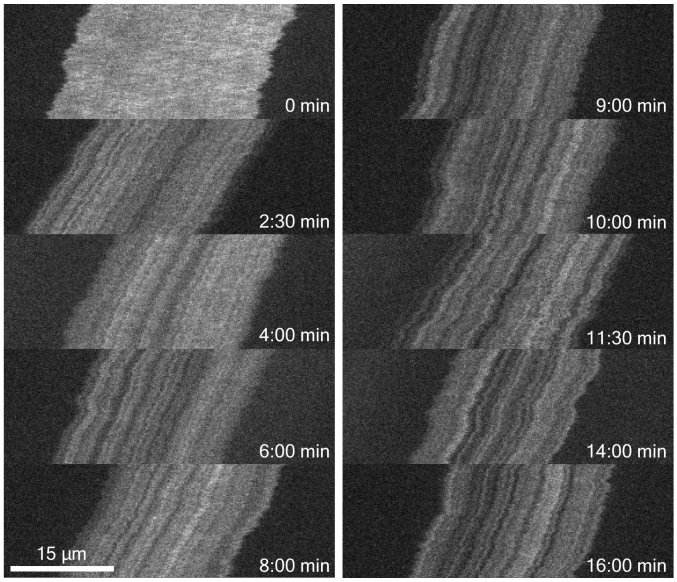


Figure 5.3: Kymographs of T4 molecules heated in microchannels and imaged at different times after turning on the heating system. The duration of the kymographs is 10 seconds. The patterns start appearing within the first minutes. The contrast is automatically adapted for each kymograph.

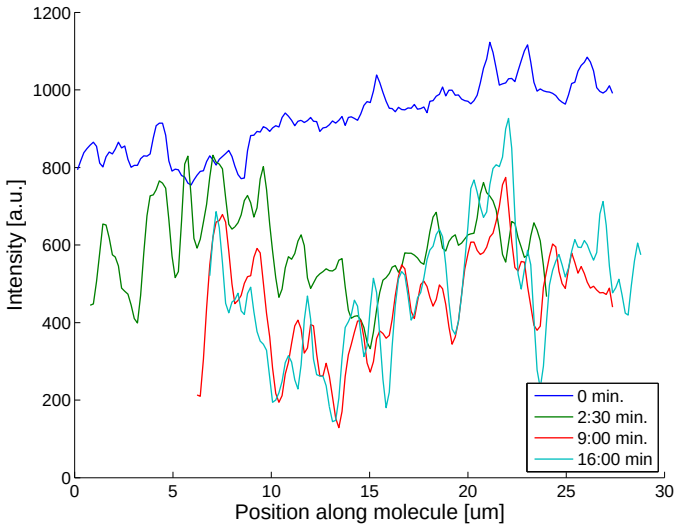


Figure 5.4: Background-corrected intensity profile of T4 DNA molecules (heated in microchannels) at different times of heating. The molecules do not have all the same length because they might have been photocut or shrunk with time because of dye dissociation [21, 25].

higher the IS. The IS allows thus to measure how well developed a melting pattern is because fully developed melting patterns have more distinct features than they would at an early stage of heating. The IS is calculated using *DNABarcodeMatchmaker*.

Fig. 5.5a shows the IS of the same molecules from Fig. 5.4. There is a strong increase of the IS within the first 5 minutes, meaning that the patterns are developing, and then it seems to stabilize. Fig. 5.5b shows the IS of all 296 measured molecules. Here the IS of every molecule was first calculated. Then the IS of all molecules within time intervals of 1 minute was averaged and plotted. Within the first 7 minutes there is a constant increase in the IS. After 7 minutes the development stops. Some of the time intervals show a big standard deviation. This might be caused by the fact that some of the observed molecules were not intact. Depending on the sequence of a fragment it can have a particularly high or low IS.

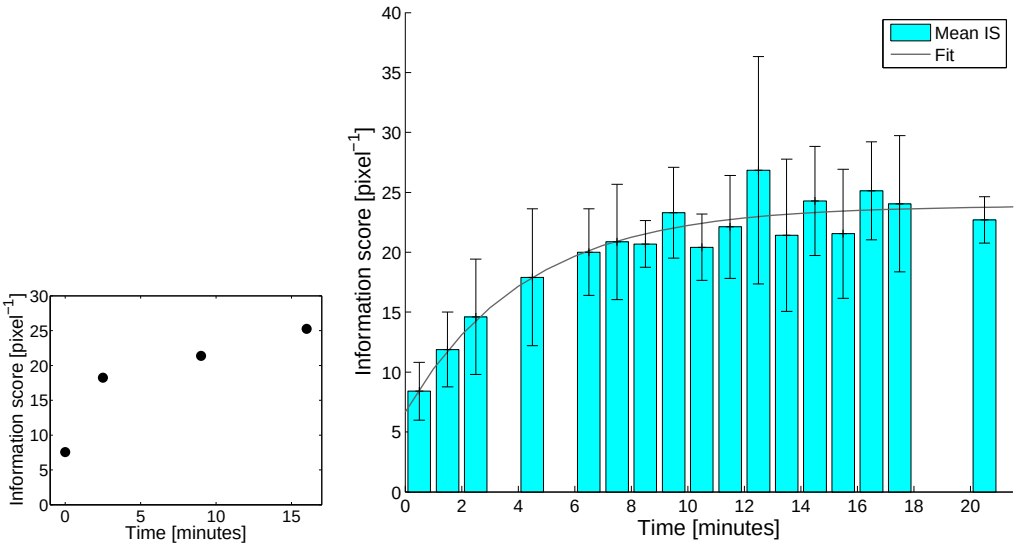
The timescale of the pattern development can also be estimated by assuming that the time development of the IS is exponential and follows the model shown in Eq. (5.2).

$$\mathrm{IS}(t) = \mathrm{IS}_0 + \mathrm{IS}_{max} \cdot \left(1 - e^{\frac{-t}{\tau}}\right) \tag{5.2}$$

Here IS_0 is the IS at time zero. Since DNA molecules at room temperature still show some slight variation of emission intensity along the molecule the value of IS_0 is bigger than zero. $\mathrm{IS}_{max} + \mathrm{IS}_0$ is the IS of a completely developed pattern and τ gives the time at which the patterns are developed by $\sim$63%. Fitting Eq. (5.2) to the data in Fig. 5.5b leads to $\mathrm{IS}_0 = 6.7$ pixel^{-1}, $\mathrm{IS}_{max} = 17.2$ pixel^{-1} and a time constant $\tau = 260$ sec = 4 min 20 sec.

So far the pattern formation was either observed on different molecules for different times or on an average of many molecules. In order to observe the formation on a single molecule, i.e., the same molecule at different times, it has to be kept inside the nanochannels during the whole heating process and imaged at different times. This has been done by first inserting the molecules into the nanochannels and then imaging them every $\sim$2 minutes after turning on the heating system. A higher heating temperature than the previous one had to be used (see Section 5.2.2). Here 55 °C was used. In order to avoid an overlap of the molecule with its own ghost image[1] the molecule was slightly shifted inside the nanochannel previous to each observation. Between the video acquisition the microscope lamp was switched off to avoid photobleaching or photocuts.

[1] A ghost image is caused by fluorescence of dissociated dye molecules sticking to the channel walls. This effect is only notable for DNA heated inside nanochannels.



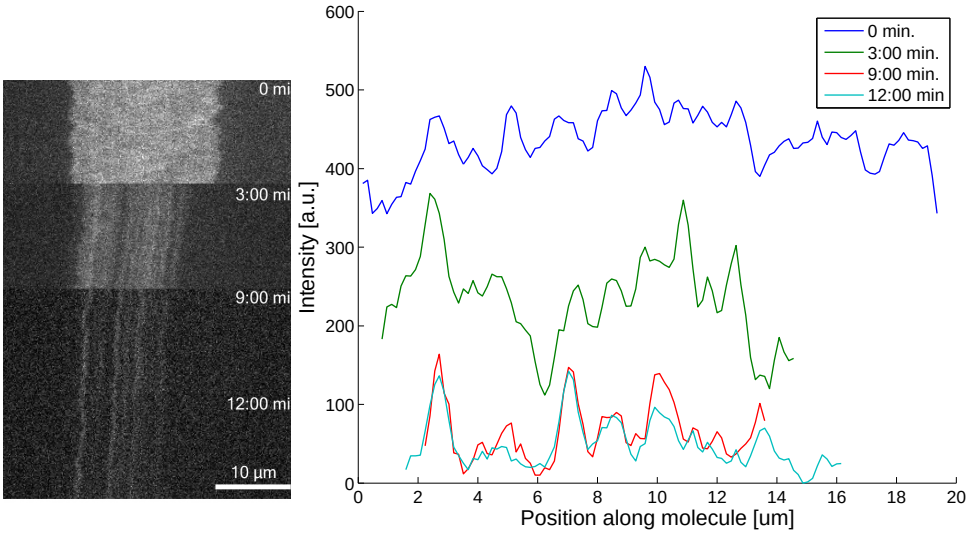
(a) (b)

Figure 5.5: (a) IS of the molecules shown in Fig. 5.4. (b) Average of the IS of molecules for time intervals of 1 minute. The total length of the error bars is twice the standard deviation. In total 296 molecules were analysed. The grey curve shows the fit of Eq. (5.2).

The time development of one molecule is shown in Fig. 5.6. Since it is always the same molecule observed at different times of heating it can be seen more clearly that from minute 9 to 12 the pattern does not change much. The same can be seen by looking at the IS development in Fig. 5.7. The pattern development is similar as the one for DNA in microchannels.

The intensity curve in Fig. 5.6b at 9 and 12 minutes corresponds to a melting temperature of 71.5 °C when compared to the theory (see Fig. A.1). The curve at 3 minutes however could not be fit to any temperature. This could be a sign that evolving melting patterns do not take shape of melting patterns that one would expect at lower temperatures but must follow other unknown rules.

It can be concluded that the denaturation pattern formation starts within the first 5 minutes of heating and stabilizes after roughly 10 minutes. A minimum heating time of 10 minutes is thus suggested. This also confirms similar observations made in other studies [26]. Further, the shape of the melting patterns during the development appears to be unpredictable using the present theory.



(a) (b)

Figure 5.6: (a) 10 seconds long Kymographs of a single T4 DNA molecule melted in nanochannels at 55 °C and observed at different times of heating. The contrast is automatically adapted for each shown kymograph. (b) Background-corrected intensity curve of the barcode generated from the same kymographs. The pattern formation stabilizes after 9 minutes.

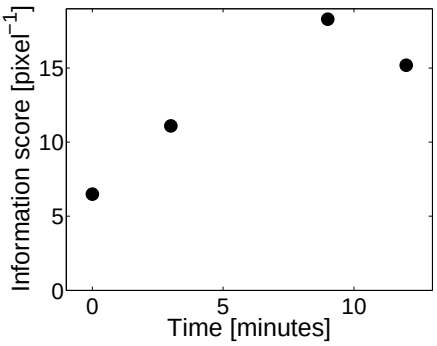


Figure 5.7: IS of the molecules shown in Fig. 5.6.

5.2.2 Confinement Dependence of Pattern Formation

According to the observations made in Section 5.2.1 and ref. [10] a confinement of the DNA raises the melting temperature. The reason for that is a different energy and entropy of confined DNA. In this section, an attempt is made at quantifying this difference in melting temperature between nano- and microchannels. The results can be helpful when predicting melting temperatures of confined DNA.

T4 DNA was prepared and imaged in the same way as the DNA heated inside the nanochannels from Section 5.2.1. Additionally, after the 15-20 minutes of heating, the molecules in the nanochannels were flushed away and the molecules that were inside the microchannel close to the nanochannels during the whole time of heating were inserted into the nanochannels. These molecules have also been heated during the time the molecules in the nanochannels were observed and are used as a control group showing what would have happened if the molecule in the nanochannels would have been melted in the microchannels. The whole procedure was repeated at 40 °C, 45 °C, 50 °C and 55 °C. At 40 °C only measurements from the microchannel were taken since we are sure that, based on previous measurements, no patterns would appear inside the nanochannels. For each temperature and time 5-30 molecules were observed.

Fig. 5.8 shows kymographs at different temperatures and times at which according to the results in Section 5.2.1 the melting patterns should be clearly visible. It shows molecules heated inside the nanochannels and the control group heated inside the microchannels. It can be clearly seen that when melted inside microchannels the patterns are already visible starting from 45 °C while the molecules melted inside the nanochannels contain patterns starting from 50 °C.

To quantitatively describe this observation the IS was again used as a measure for how developed the melting patterns are. The results are presented in Fig. 5.9. As a comparison, this figure also shows again the same data from Fig. 5.5b of the molecules melted in microchannels (grey curve).

It can be seen in Fig. 5.9 that molecules heated at 40 °C in the microchannels have a low IS. This is expected since no patterns were observed at that temperature. At 45 °C for the nanochannels the IS is low all the time while for the microchannels it is, as expected, close to the values from the grey curve. This means that inside the nanochannels no patterns developed while in the microchannels they developed, reproducing similar values as in Fig. 5.5b. At higher temperatures the IS for the microchannels stays always close to the one from the grey curve. It is important to note that some

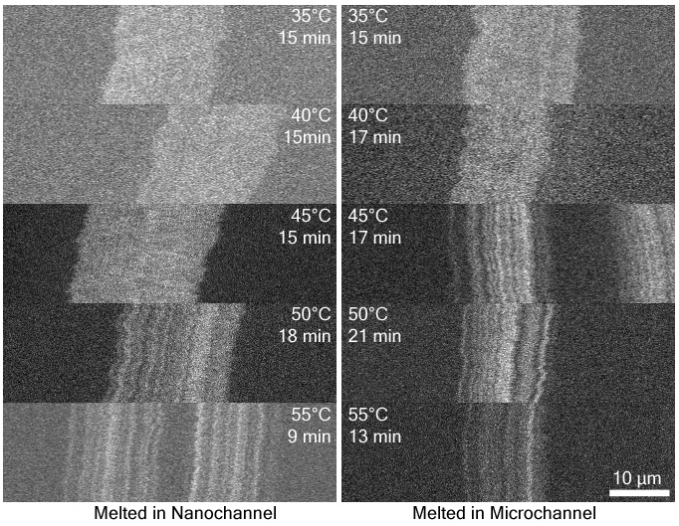


Figure 5.8: Comparison of 10 seconds long kymographs of T4 DNA heated in nanochannels (left) and microchannels (right) at different temperatures and at a time where the pattern formation is expected to be complete. The contrast is automatically adapted for each kymograph.

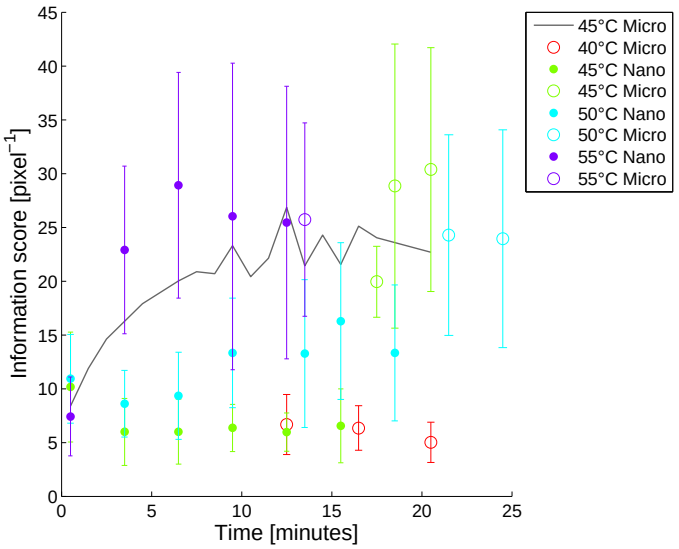


Figure 5.9: Evolution of the IS from T4 DNA heated inside nanochannels (full circles) over time at different temperatures compared with heating inside microchannels (empty circles). The grey curve represents the same values from Fig. 5.5b and is used as a comparison. Each data point represents an average of the IS of 5-30 molecules. The total length of the error bars is twice the standard deviation.

of the molecules from the microchannels were completely melted at 55 °C. Those molecules are not easily detected and thus the plotted values at that temperature only represent the still intact molecules. The time development of the patterns from the molecules inside the nanochannels starts changing at temperatures above 45 °C. A slight increase with time is observable at 50 °C. However the IS after 10 minutes is still not as high as the one in the molecules from the microchannels. Only at 55 °C the IS from the molecules in the nanochannels follows the curve from the molecules in the microchannels (grey curve) and stabilizes within 10 minutes.

The big standard deviation in Fig. 5.9 might be caused by two reasons. One might be the same as the one for Fig. 5.5b, i.e., broken molecules representing parts of the DNA with an either very high or very low IS. It can however be seen that the standard deviation increases with temperature. The staining of the DNA is always slightly heterogeneous [21] and thus the dye:bp ratio differs between the molecules. Since the melting temperature depends on the dye:bp ratio [9], this leads to a spread in melting temperature, i.e., some molecules show melting patterns at lower and some at higher temperatures. Finally this leads to a spread of the IS. For low temperatures where the melting patterns are not developed at all or only weak this effect is smaller.

We can conclude that the pattern formation inside nanochannels slowly starts at 50 °C. At 55 °C it is comparable to the one in the microchannels. In the microchannels the patterns already appear at 45 °C. The difference in melting temperature between the nano- and microchannels is thus estimated to be in the range of 5-10 °C. This shows that depending on the type of experiment, the influence on the melting temperature caused by confinement can not be neglected. In order to make a more precise statement however more measurements at more different temperatures are needed. It can also be concluded that it is important to make sure that the staining of the DNA is done carefully in order to minimize any spread in melting temperature between the molecules.

5.2.3 Influence of Dye Load in Off-Chip Heating

In this section we study how the influence of the dye load on DNA affects the melting pattern formation in off-chip heating. It is important to study off-chip heating since, contrary to the on-chip heating, it will be used when performing the experiments for the bacterial identification. The investigation was done by heating T4 DNA with different dye loads at different temperatures and with different amounts of formamide. The DNA was heated off-chip and observed on glass slides.

Each observed sample was heated in a PCR machine during 10 minutes. During heating, each sample had a salt concentration corresponding to 1×TBE and contained 3% (v/v) BME. The formamide concentration, dye:bp ratio and temperature were changed from experiment to experiment. The ionic strength during heating was estimated to be 0.028 M.

2 µL of the melted and cooled sample were put on a poly-L-lysine coated glass slide (Poly-Prep Slides, Sigma). The sample was then covered with a cover glass. By applying some pressure to the cover glass and slightly sliding it, the DNA stretched out and stuck to the glass slide. The poly-L-lysine enhanced the sticking of the DNA to the glass. After that, the sample was ready to be observed under the microscope. Since the observation was made on glass slides and not in the nanochannels only a qualitative description of the results can be made. The results are summarized in Fig. 5.10.

T4 DNA stained at a dye:bp ratio of 1:10 and heated at 45 °C with 50% (v/v) of formamide is expected to show well developed denaturation patterns according the previous on-chip measurements and theory. However, when melting off-chip, the melting patterns are very weak (see Fig. 5.10a). Melting under the same circumstances but with a strong decrease of the dye:bp ratio to 1:100 leads to well developed patterns (see Fig. 5.10b). We could observe in Section 5.2.1 that the dye might lead to an increase in melting temperature. Apparently this effect is much stronger when melted off-chip. An increase of the heating temperature to 71 °C allows to form melting patterns at a dye:bp ratio of 1:10 (see Fig. 5.10c). Note that according to Fig. A.2 this is the temperature at which melting patterns would be expected. It seems that the dye cancels out the effect of the formamide. In fact, by leaving away the formamide completely but keeping the temperature of 71 °C and dye:bp ratio of 1:10 only very weak patterns are visible on the DNA (see Fig. 5.10d). At 95 °C and still with a dye load of 1:10 and no formamide it was possible to reproduce patterns of a similar quality as the ones melted at 45 °C with 50% (v/v) of formamide and a dye load of 1:100 (see Fig. 5.10e). The adding of formamide at 95 °C lead to an almost complete denaturation of all the molecules (see Fig. 5.10f).

It is known from ref. [9] and it has also been observed in Section 5.2.1 that an increase in dye:bp ratio increases the melting temperature. This has also been observed so far in the on-chip heating. When melting off-chip however, this effect seems to be much grater. The cooling step from the off-chip heating must therefore somehow influence the pattern formation. An explanation for this phenomena could not be found.

It can be concluded that the influence of the dye:bp ratio on the pattern formation is much stronger when the DNA is heated off-chip. A low dye:bp

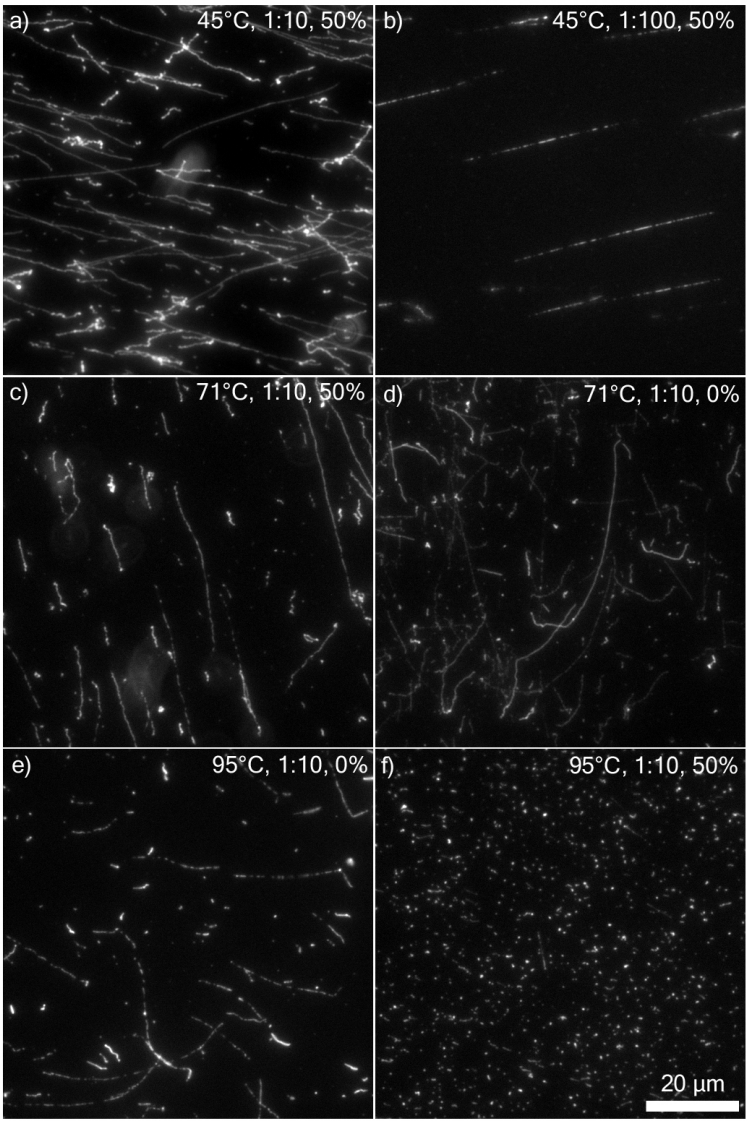


Figure 5.10: T4 DNA heated off-chip at different temperatures, dye:bp ratios and amounts of formamide. The DNA is stretched out on glass slides.

ratio reduces this effect and allows a better control over the temperature at which melting patterns occur. More experiments are needed in order to quantify the difference in melting temperature and also to study why the cooling step has this influence.

5.2.4 Summary

The investigation of the time dependence of the pattern formation lead to the conclusion that the denaturation pattern formation starts within the first 5 minutes of heating and stabilizes after roughly 10 minutes. The time constant of the pattern formation is estimated to be $\tau = 4$ min 20 sec by taking the IS as a measure of the melting pattern development and assuming that that it follows an exponential rise. The minimum heating time is thus suggested to be set to 10 minutes. The experiments on the confinement dependence of the pattern formation showed that the melting temperature difference between nano- and microchannels is in the range of 5-10 °C. These experiments also revealed that a spread of the dye load among the molecules leads to a spread of melting temperatures. Finally, it could be shown that the increase of the melting temperature with dye:bp ratio is much stronger in off-chip heating compared to on-chip heating.

5.3 Comparison of Experiment to Theory

5.3.1 Comparison to Theory at one Temperature

In this section the agreement between experimentally acquired barcodes of T4 DNA and their corresponding theory at the specific heating temperature is shown and discussed. A good agreement is needed for the later bacterial identification.

The used T4 DNA was stained at a dye:bp ratio of 1:10 and heated on-chip at 45 °C in the presence of 50% (v/v) formamide and 3% (v/v) BME. The TBE concentration during heating was 0.18×TBE and the ionic strength is estimated to be 0.02 M. Several movies of 20 seconds duration of the DNA in the nanochannels were measured and the barcodes of the molecules were generated.

Using *DNABarcodeMatchmaker* a theoretical barcode was created. The theoretical barcode needed to be adjusted in order for it to be comparable with the experimental barcodes, i.e., the length and resolution had to be adapted. The length was set to 23 µm. This length was determined by measuring the length of intact T4 molecules inside the channels. At the given channel width, salt concentration and dye load we would expect a stretching factor of the DNA of ∼0.45 according to ref. [27]. This leads to a length of 23.2 µm which is close to the measured one. The resolution of the theoretical map was set to 320 nm, corresponding to an estimated resolution of 2 pixels. A lower resolution than the theoretically expected one of the microscope (see Section 4.2.3) was chosen because the alignment of the kymographs reduces slightly the final resolution of the barcode. The theoretical barcodes were generated for a heating temperature of 69 °C. This temperature was found to fit best the experimental barcodes. Again this temperature does not correspond to the one expected using Eq. (2.1) for the same reasons discussed in Section5.2. Before comparing the experimental and theoretical barcode, their mean was set to 0 and their standard deviation to 1 (see Section 4.3.4).

Fig. 5.11 shows the comparison of two experimentally acquired barcodes to their corresponding theory. The experiment fits well to the theory. The observed differences could be minimized by fine tuning the input parameters when generating the theoretical curve, i.e., temperature, ionic strength, optical resolution or length of the molecule. It is thus important to have good control over the experiment in order to determine these parameters accurately.

In Fig. 5.12 an average of 9 complete T4 molecules has been made and compared with theory. The experimental barcode still fits well to the the-

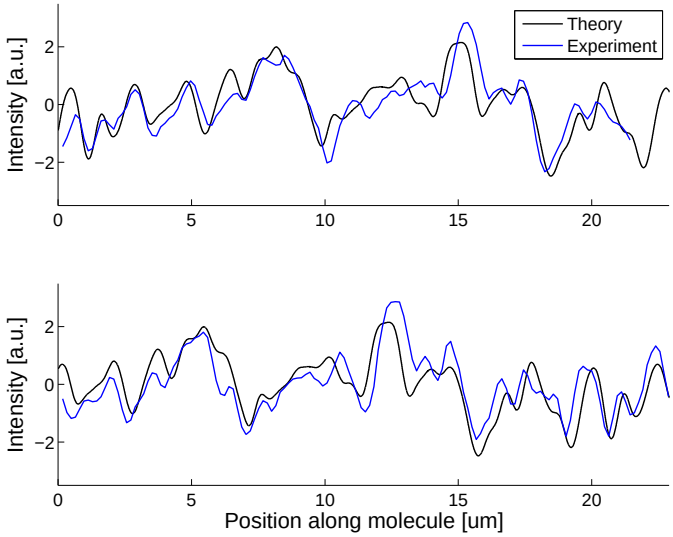


Figure 5.11: Comparison of experimental (blue) and theoretical (black) intensity curves of two different partially denatured T4 DNA molecules. The theoretical barcodes corresponds to a temperature of 69 °C. The theoretical and experimental barcodes have all mean 0 and standard deviation 1.

ory. However, some features of the experimental curve have been averaged out. Although the curves match well, the standard deviation is rather big. The reason for that could be variations in the sample. Another reason might be that not all experimental barcodes have been aligned perfectly to each other before being averaged. Finally, a variation in dye load between the molecules resulting in different melting temperatures (see Section 5.2) might also be a reason for the big standard deviation. This problem will be avoided in the next section (see Section 5.3.2) by fitting the experimental barcodes to different temperatures instead of just one.

It could be shown that there is a good agreement between theory and experiment and that the melting patterns are theoretically predictable. This is important since an identification of a DNA molecule relies on a good match between experiment and theory. Similar results have also been obtained in [1, 6, 7].

5.3.2 Comparison to Theory at Multiple Temperatures

In the previous section (see Section 5.3.1) experimental barcodes have been compared to theoretical ones at one specific temperature. In this section experimental barcodes from a whole temperature range are compared to the theory. This is done by creating a temperature dependent denaturation map as the one in Fig. 5.2 from experiments and comparing it to the correspond-

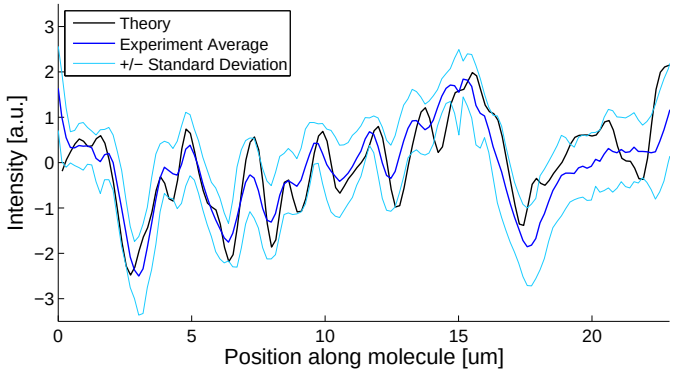


Figure 5.12: Comparison of experimental (blue) and theoretical (black) intensity curves of partially denatured T4 DNA. The experimental curve consists of an average of 9 different T4 molecules. The bright blue curves show $\pm$ the standard deviation of the mean intensity at each position of the molecule. The theoretical barcode corresponds to a temperature of 69 °C. Both curves have mean 0 and standard deviation 1.

ing theoretical one.

One way to create an experimental temperature dependent denaturation map would be to first heat and image stained T4 molecules at different temperatures. Here however, instead of heating the DNA at different temperatures, advantage was taken of the fact that heterogeneous staining of the molecules leads to a spread in melting temperature (see section 5.2). This allowed us to collect data of molecules with denaturation patterns corresponding to different temperatures within a single experiment.

T4 DNA was stained at a dye:bp ratio of 1:100. A low ratio was used to enhance heterogeneous staining [21] and in order to keep the heating temperature low (see Section 5.2.3). While preparing the sample, a heterogeneous staining was additionally provoked by skipping the heating process at low temperatures of the freshly stained DNA (see Section 4.3.1) and by avoiding careful mixing steps and waiting times. The DNA was then heated off-chip during 10 minutes at 42 °C with 50% (v/v) formamide and 3% (v/v) BME. The TBE concentration was 1×TBE and the ionic strength is estimated to be 0.028 M. Using Eq. (2.1) this corresponds to a heating temperature of 72 °C without formamide. At that temperature melting patterns are expected to occur according to the theory (compare Fig. A.2). Other than in the previous experiments (see section 5.2) the dye has only a small influence on the melting temperature due to the low dye:bp ratio.

After inserting the sample into the nanochannels many videos of the molecules with a duration of 20 seconds were acquired. Fig. 5.13 shows some T4 molecules in the nanochannels, their kymographs and the resulting bar-

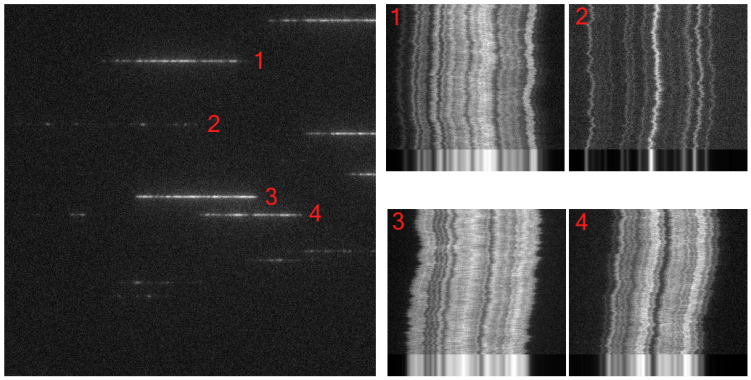


Figure 5.13: Left: T4 DNA in nanochannels stained at a dye:bp ratio of 1:100 and heated at 42 °C. Right: Kymographs of 20 seconds duration from 4 selected molecules and their resulting barcode. The barcodes correspond to different temperatures.

codes. It can be seen that the staining was successfully heterogeneous since the molecules show denaturation patterns corresponding to different temperatures.

In total 512 molecules were imaged. It is important to note that not all molecules were intact. Fragments of molecules (minimum length of 8 µm) are therefore also considered in the data analysis. The kymographs of the molecules were extracted from the videos and aligned using *DNABarcodeMatchmaker*. Since a big amount of molecules was analyzed the process of generating barcodes from the aligned kymographs, fitting the barcodes to the right temperatures and averaging them had to be automated. This was done by writing a separate set of MATLAB scripts that processed the aligned kymographs in the following way:

1. Make a time average of the aligned kymograph. The result is the intensity profile of the molecule and its surrounding background noise.
2. Temporarily smooth the intensity curve by making a convolution of it with a Gaussian function. This allows a better edge detection in the next step.
3. Identify edges of molecule using Otsu's method for thresholding, a common technique for feature recognition in image analysis [28].
4. Extract the molecule.
5. Subtract background from intensity profile of molecule. The background is estimated to be the average of the background noise.

The result is a background subtracted intensity curve for each molecule, ready to be compared with a theoretical map.

Using *DNABarcodeMatchmaker* a theoretical map was created for the temperature range of 69-75 °C in steps of 1 °C. Same as in Section 5.3.1, the map was adjusted to be comparable with experimental molecules. Its length was set to 32 µm. This length was again based on the measurement of the length of intact T4 DNA molecules in the nanochannels. The stretching factor according to ref. [27] was 0.55, leading to a length of 30.7 µm. This length is close to the measured one. The resolution of the theoretical map was again set to 320 nm. The obtained theoretical map is shown in Fig. 5.14 on the left.

Using again a separate set of MATLAB scripts, the experimentally acquired intensity curves were compared with the theoretical map. They were first all placed along the theoretical map where they were most similar to the theoretical map. Before measuring the similarity between the experimental and theoretical curve both of the curves were rescaled to have their minimum at 0 and maxima at 1. This rescaling was chosen because setting the mean of the curves at 0 and standard deviation 1, as done so far (see Eq. (4.2)), lead to unsatisfying results, i.e., many curves had their best match at the wrong place. One reason for that could be that the theoretical intensity curves at high temperatures have many 0 values leading to complications.

As a measure for the similarity between the curves the *Sum of Absolute Difference*, or SAD, between the rescaled curves was computed as shown in Eq. (5.3). Here x_{ex} is the experimental curve and x_{th} the theoretical. The SAD was chosen over other metrics because it is robust and lead to the best results.

$$SAD = \sum_i |x_{ex_i} - x_{th_i}| \tag{5.3}$$

Each of the 512 molecules was placed at the positions along the theoretical map where they minimized the SAD. Once all barcodes were placed where they fit best to the theory, they were averaged.

In order to visualize the results the averaged experimental barcode at each temperature was set to have the same minimum and maximum as the original, not rescaled, theoretical barcode. The result is illustrated as a greyscale image in Fig. 5.14. The experimental map is very similar to the theoretical one. A better overview gives the plot of the intensities of each temperature in Fig. 5.15.

A lack of coverage at the highest and lowest temperature is notable. This can be explained by Fig. 5.16 which shows the amount of fitted barcodes for each temperature. It shows a distribution similar to the normal one with the highest value at 71 °C. The center of the distribution can be explained

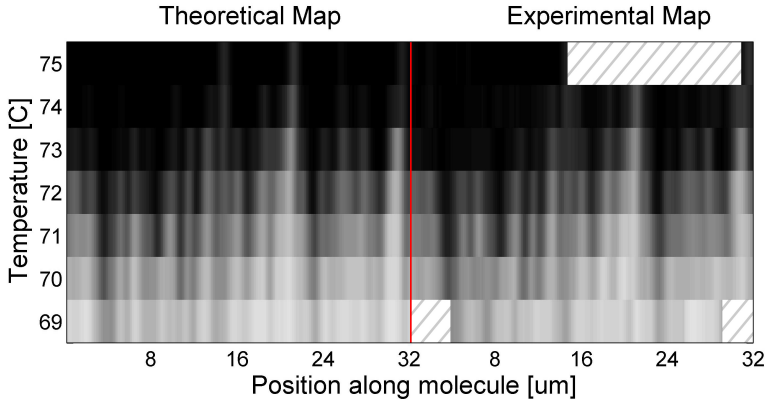


Figure 5.14: Comparison of the theoretical (left) and experimental (right) temperature dependent denaturation map of T4 DNA in greyscale. The hatched areas represent sections that are not covered by any molecules.

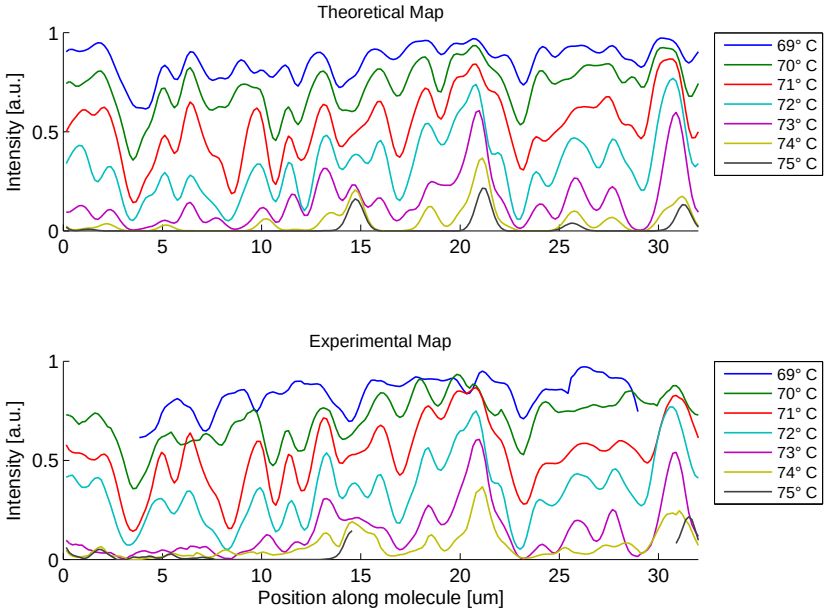


Figure 5.15: Intensity curves of the theoretical (top) and experimental (bottom) temperature dependent denaturation map of T4 DNA.

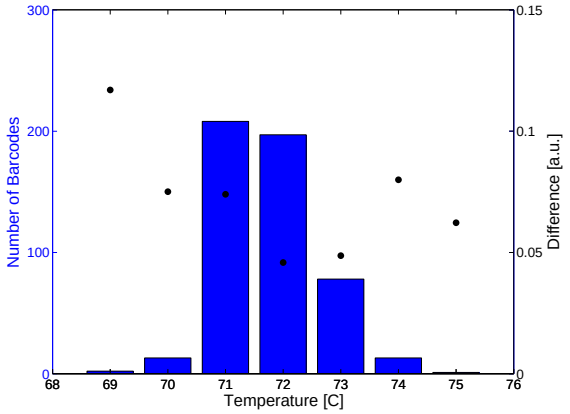


Figure 5.16: Histogram showing the amount of molecules for each temperature over which the average for the temperature dependent denaturation map was made. The difference (*SAD* divided by length) between theory and experiment for each temperature is shown by the black dots.

by the fact that the corresponding heating temperature without formamide is at 72 °C while its width is due to the heterogeneous staining. It can be seen that the temperature spread of the molecules is smaller than the temperature range they were compared to, causing the observed lack of coverage at the highest and lowest temperatures. The lack of coverage at 75 °C can additionally be explained by the fact that according to theory this region is very dark, i.e., completely melted, making these molecules almost undetectable.

In order to quantify the similarity of the experimental map with the theoretical map the *SAD* at each temperature was computed. The result is shown in Fig. 5.16. The amount of molecules per temperature seems not to influence the *SAD*. The *SAD* also seems to be best at 72 °C. However, in order to make a better statement about the similarity a more sophisticated analysis needs to be done.

It can be concluded that the theoretical predictions of the partial denaturation could be experimentally reproduced. It could also be shown that an automatic alignment of experimental barcodes to the corresponding theory is possible. This is required in the first steps towards bacterial DNA identification. The next steps would be to describe the results quantitatively and also to compare the similarity of the experimental barcodes to theories of different DNA strains. This is implemented and constantly improved in *DNABarcodeMatchmaker* which will be used in the chapter 5.4 for the bacterial identification.

5.3.3 Summary

It could be shown in this section that there is a good agreement between experimental and theoretical barcodes. By comparing an experimental temperature dependent denaturation map to its corresponding theory it could further be shown that an automatic fitting of experimental barcodes to the theory leads to satisfying results. A good match of experiment and theory is an important prerequisite for the bacterial identification. A more advanced way of matching experimental barcodes to their theory is done in Section 5.4 for the bacterial identification using *DNABarcodeMatchmaker*.

The experiments from Section 5.3.2 further revealed that a heterogeneous staining of the DNA can be used to create a spread in melting temperature within the same sample. This allows us to create melting patterns corresponding to different temperatures in one single experiment.

5.4 Bacterial identification

5.4.1 Introduction

In this section the gained knowledge the previous experiments (see Sections 5.2 and 5.3) is used to measure barcodes of the DNA from different *Streptococcus pneumoniae* bacteria. Using *DNABarcodeMatchmaker* attempt is made at identifying the DNA.

S. pneumoniae are a common cause for many severe bacterial infections [29]. The main diseases caused by *S. pneumoniae* are pneumonia, meningitis, sepsis and otitis media (middle ear inflammation). *S. pneumoniae* are a main reason for high morbidity and mortality and their antibiotic resistance has been increasing over the past years [30]. Thus, a lot of research is done on these bacteria, including the development of better diagnostic tools.

In this thesis the *S. pneumoniae* strains TIGR4, D39 and R6 have been analyzed. They have been chosen because of their availability and known sequence. The DNA was extracted from the bacteria, purified and suspended in TE buffer (10 mM TRIS, 1 mM EDTA) using standard procedures[2].

The DNA extraction does not provide complete DNA molecules but only fragments of a maximum length of $\sim 1.5 \cdot 10^5$ bp. Complete DNA molecules are difficult to obtain and they would be too long to fit the field of view of the used microscopes since their length varies between 2-2.1 Mbp [31]. Because of that, the barcodes of many fragments of the same strain have to be measured. These barcodes are then assembled to build a complete experimental barcode of the total DNA, which is then compared to the corresponding theory. According to calculations made in Appendix B it is recommended to measure a minimum amount of $\sim$200 fragments per strain in order to be sure that the complete length of the DNA can be assembled by the fragments.

In the next section (see Section 5.4.2) first the results from the strain TIGR4 are analyzed in detail. After that, the results for the strain D39 and R6 are discussed. The figures of these two strains are shown in the appendix D. For each DNA strain several hundreds of fragments have been measured and analyzed.

5.4.2 Identification of TIGR4, D39 and R6

First, the barcodes of many fragments from the strain TIGR4 had to be acquired. The DNA was stained at a dye:bp ratio of 1:10 and was heated

[2]The DNA extraction was done by Sarah Brovall from the Birgitta Henriques Normark and Staffan Normark Group at Karolinska Institutet, Stockholm

off-chip during 10 minutes (see Section 5.2.1) at 70 °C in the presence of 50% (v/v) formamide and 3% (v/v) BME. The relatively high temperature was used based on the findings from Section 5.2.3. The TBE concentration during heating was 1×TBE and the ionic strength is estimated to be 0.028M. A detailed protocol on how the TIGR4 DNA has been prepared is shown in Appendix C.

A total of ∼2000 fragments were measured. All kymographs showed well developed denaturation patterns. Kymographs showing a damage of the molecule caused by photocutting were excluded manually. Further, *DNABarcodeMatchmaker* was used to automatically exclude molecules that were shorter than 8 µm. However, not all of the fragments shorter than 8 µm have been rejected because the filtering by length done in *DNABarcodeMatchmaker* is done before the Kymographs are aligned and is thus not very accurate. A total of 619 kymographs have been selected for further analysis.

The length distribution of the barcodes from the fragments is shown in Fig. 5.17. It is similar to a normal distribution and has a mean of 13.4 µm and a standard deviation of 4.0 µm. According to calculations made in Appendix B the length and amount of the barcodes analyzed are high enough to assemble a complete DNA molecule.

Fig. 5.18 shows the distribution of the IS. The distribution is also similar to a normal one and has a mean of 30.3 pixel^{-1} and a standard deviation of 9.7 pixel^{-1}. Compared to the measurements from Section 5.2 the measured IS of the TIGR4 fragments is considered to represent fragments with well developed melting patterns.

To identify the DNA strain from which the measured fragments originated, their barcodes were compared with the theory. This procedure was completely done by *DNABarcodeMatchmaker* and is briefly explained in the following.

First a complete experimental barcode had to be assembled using the measured barcodes from the fragments. This was done by first generating the complete theoretical barcode for the strain that was planned to be tested, using the usual input parameters such as temperature, ionic strength, optical resolution and stretching factor. Then the experimental barcodes from the fragments were placed at the position along the theoretical barcode where they fit best. This position was found by computing the normalized cross-correlation [5] between the experimental and theoretical barcode. The position at which the normalized cross-correlation was maximized was decided to be the position where the experimental barcodes fit best to the theoretical barcode. Once all experimental barcodes had been placed at

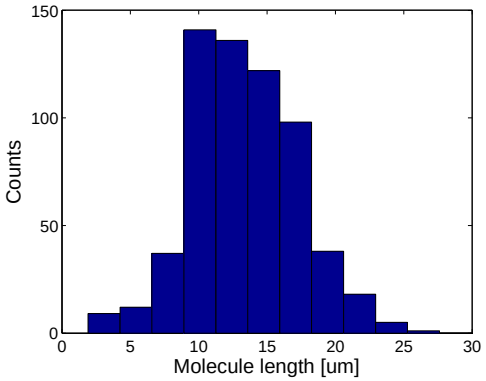


Figure 5.17: Histogram showing the length distribution of the 619 analyzed TIGR4 DNA fragments. The mean is 13.4 µm and the standard deviation is 4.0 µm. Note that this distribution does not represent the natural distribution of the fragment length since short fragments (<8 µm) are mostly ignored in the data analysis.

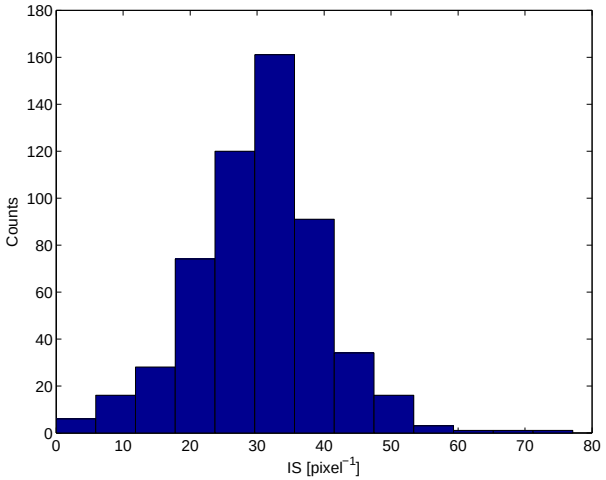


Figure 5.18: Histogram showing the IS distribution of the 619 analyzed TIGR4 DNA fragments. The mean is 30.3 pixel^{-1} and the standard deviation is 9.7 pixel^{-1}.

their position where they fit best to the theoretical barcode, an average over all the experimental barcodes was taken resulting in one single and complete experimental barcode. We refer to this barcode as the assembled barcode. If enough fragments to cover the whole length of the theoretical barcode were used, the assembled barcode would have the same length as the theoretical barcode.

Next step was to calculate how good the assembled barcode fit the theoretical barcode. This was done by computing the cross-correlation C between the rescaled assembled barcode and the rescaled theoretical barcode, as shown in Eq. (5.4).

$$C = \frac{1}{l} \sum_{n} p(n) \cdot t(n) \tag{5.4}$$

Here l is the total length of the assembled barcode, n is the position along the barcode, p is the rescaled assembled barcode and t is the rescaled theoretical barcode. If the assembled barcode had gaps, i.e., the fragments did not cover the whole theoretical barcode, C was computed only for the regions where the two barcodes overlapped. A higher value of C means a better match. Hence, if fragments from TIGR4 were matched to the theory of TIGR4 we would expect a high C. However, if fragments from TIGR4 would be matched to the theory of D39 instead, we would expect a lower C.

From the fragments of the TIGR4 DNA an assembled barcode was generated using theoretical barcodes from 14 different strains, including TIGR4. For each strain C was computed. Because the heating of the DNA was done off-chip, the temperature that has to be used to generate the theoretical barcode was not exactly known (see Section 5.2.3). Thus the value of C of the different strains was not computed at a single temperature but for different temperatures in the interval 60-79 °C. The values of C for each strain at each temperature are shown in Fig. 5.19.

First, a general observation of the plot in Fig. 5.19 is discussed. One can see that the cross-correlations of all strains have a maximum at temperatures close to 75 °C and a minimum at temperatures close to 65 °C and 80 °C. This general dependence on the temperature can be explained by looking at the theoretical temperature dependent denaturation map of TIGR4 in Fig. 5.20. The denaturation map of the other strains is similar to the one of TIGR4. At around 75 °C the barcode is most developed, i.e., it shows the most features. This makes it easy for fragments to match to the theoretical barcode at that temperature since there are many possibilities to find a good fit, resulting in an overall high cross-correlation. At higher and lower temperatures however, big regions of the DNA are either completely melted or completely double-stranded, making it difficult for fragments to find a place

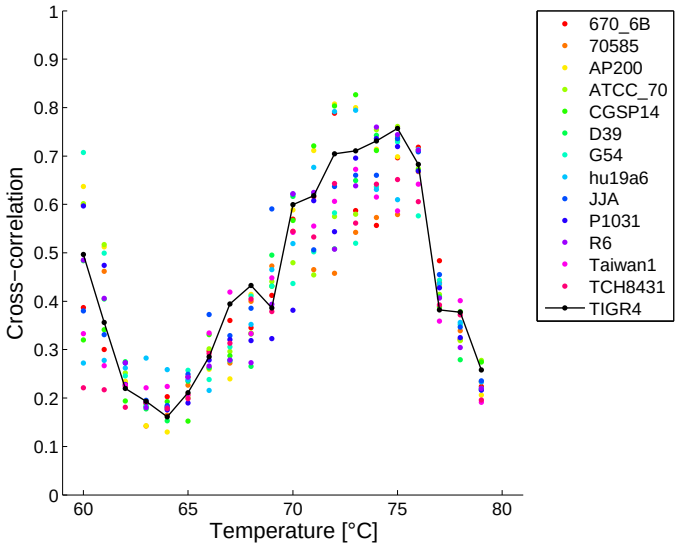


Figure 5.19: Cross-correlation C of 619 fragments from TIGR4 DNA matched to theories of different strains calculated for different temperatures.

to match to the theory. This results in an overall low cross-correlation. The reason why at low temperatures (60 °C) the cross-correlation of some strains raises again could not be explained.

In the plot in Fig. 5.19 at a specific temperature, the cross-correlation for TIGR4 is expected to be significantly higher than compared to all other strains. Seeing that would allow us to identify the TIGR4 DNA among the 14 different strains. The temperature at which one would see the peak would correspond to the correct heating temperature from the off-chip heating. However, a significant peak can not be seen. On the other hand, there are some temperatures at which the value of C from TIGR4 is among the highest (70 °C, 75 °C and 68 °C). Since we do not know which temperature should be considered for the comparison of C, attempt is made at determining the correct temperature.

To determine the temperature that should be considered, the fragment distribution of the assembled barcode is studied. This distribution is computed by considering the length of the assembled barcode in units of pixels. For each pixel the amount of fragments covering it is counted and plotted. For the right temperature this distribution is expected to be uniform. At a wrong temperature all the fragments might crowd at one specific region. This would lead to a non-uniform distribution. Note that the value of C does not necessarily decrease by much in such a case. As a rough measure of the uniformity of the distribution, the standard deviation of the distribution

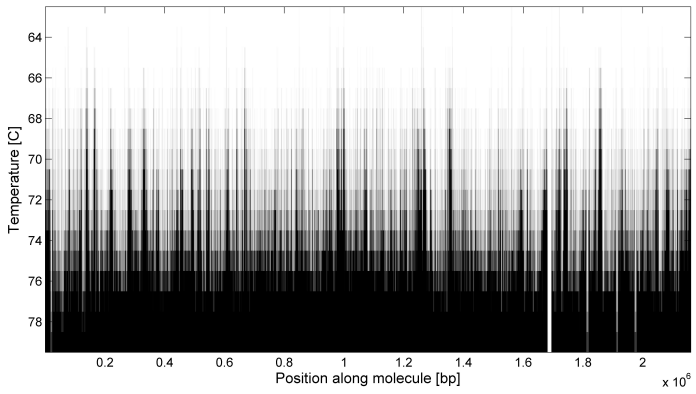


Figure 5.20: Theoretical temperature dependent denaturation map of TIGR4 DNA at an ionic strength of 0.028M.

is computed.

The standard deviation of the fragment distribution is calculated for the strain TIGR4[3] at all temperatures and is shown in Fig. 5.21. The minimum is at 72 °C. In Fig. 5.19 one can see that there are 4 strains with a higher C than TIGR4 at that temperature. The TIGR4 strain can thus not be distinguished from the other strains. It could however be, that the strains that have a higher C than TIGR4 have a similar sequence to the one of TIGR4. An analysis of the sequences would give more insight on that.

A closer look at the fragment distribution is taken in Fig. 5.22 which shows the fragment distribution, the theoretical barcode and the assembled barcode of TIGR4 at 72 °C. The first thing to notice is that the fragment distribution is not uniform. One reason for that could be that the measured fragments are not completely random but fragments of some parts of the DNA occur more than others. By comparing the distribution with the theoretical barcode one can however see that the fragments tend to match better in regions with strong negative peaks, i.e., strong melted areas. The reason for the non-uniform distribution is thus suggested to originate from a not optimized assembling process where fragments tend to match better in strong negative areas of the theoretical barcode. Nevertheless, the assembled code is similar to the theoretical one. However, the prominent negative peaks seen in the theoretical barcode are less prominent in the assembled one. The rescaling and averaging process done by *DNABarcodeMatchmaker* could be the cause for that.

[3]For a real DNA identification process one would have to consider the distribution of all strains since one would not know what strain had been measured.

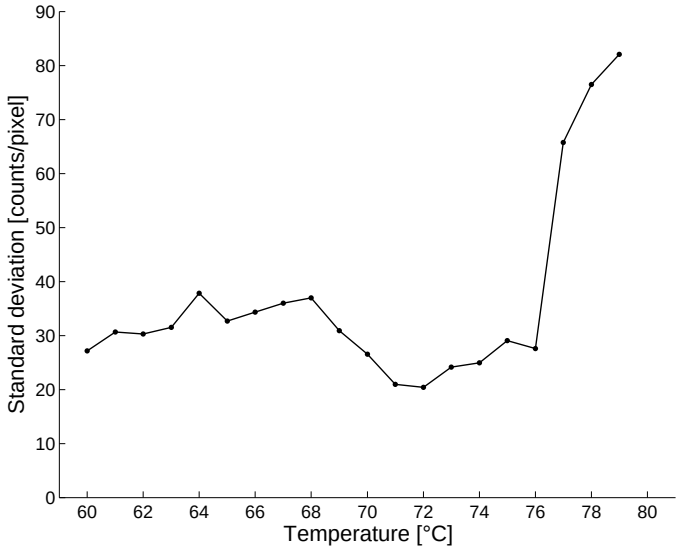


Figure 5.21: Standard deviation of the fragments distribution of TIGR4 DNA at different temperatures. The minimum is at 72 °C.

The same measurement and analysis as for the strain TIGR4 has been done for the strain D39 and R6. The protocol for the DNA sample preparation was analogous to the one for TIGR4 shown in Appendix C. For both strains ∼340 fragments have been analyzed. The plots of the cross-correlation and fragment distribution for the strains D39 and R6 are presented in the Appendix D. The results are very similar to the ones for TIGR4. In both cases no strain could be identified from the others by comparing the cross-correlation. Also the fragment distribution with the lowest standard deviation was still highly non-uniform with many fragments matching at strong negative peaks. It is however interesting to point out, that the cross-correlation of the R6 strain in Fig. D.4 is among the highest at almost every temperature. The fragments must have some properties that makes it especially easy for them to fit to the R6 strain at any temperature. However, for the temperature at which the fragment distribution is most uniform, there are 2 other strains with a higher cross-correlation.

5.4.3 Discussion

It can be concluded that, although a high enough number of DNA fragments with good melting pattern quality have been measured, a successful identification of the measured DNA strain could not be done yet. An optimization in the experimental procedure as well as processes from *DNABarcodeMatchmaker* could improve the results. Since a discussion on the improvement of *DNABarcodeMatchmaker* is beyond the scope of this thesis, only possible improvements from an experimental point of view are considered in the fol-

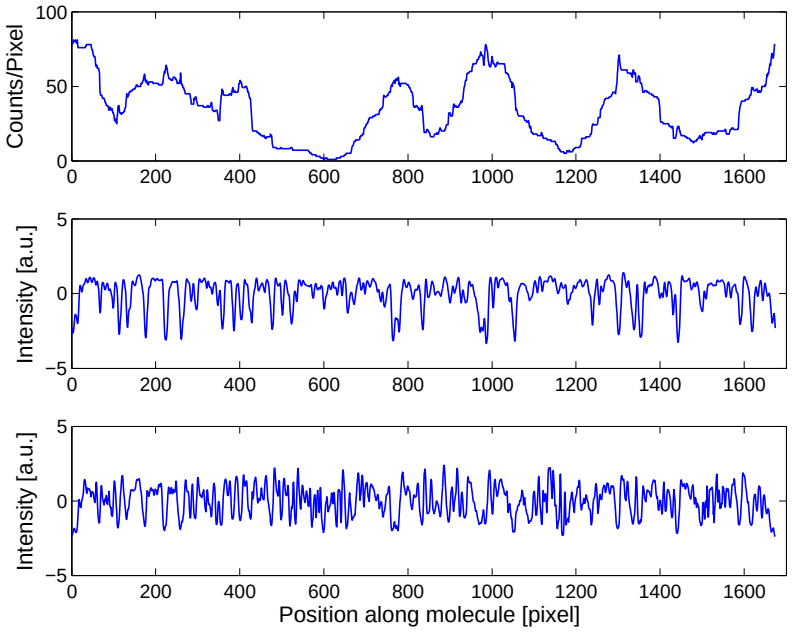


Figure 5.22: TIGR4 DNA fragments matched to corresponding theory at 72 °C (top) Counts of coverage per pixel (fragment distribution along theory). The total length of the fragments is 57652 pixel. An ideal case would have ∼35 counts per pixel. (middle) Theoretical barcode at 72 °C. (bottom) Assembled barcode.

lowing.

- The most important point that should be improved is the control over the temperature. Once the right temperature is known, a more careful analysis can be done. More studies on how the off-chip heating influences the temperature are therefore needed. The development of a new chip and device holder design that allows an on-chip heating without any reduction in throughput and image quality (see Section 4.3.2) might also lead to better control over the temperature.

- The use of longer DNA fragments might improve the identification since it would be easier for *DNABarcodeMatchmaker* to find the correct fit. Longer fragments can be achieved with alternative methods for the DNA extraction from bacteria but also by careful handling of the DNA samples during the experiments.

- A spread in melting temperature, as discussed in Section5.2, might also influence the results in a negative way. An optimization of the staining procedure is therefore suggested.

- The input parameters that are needed in order to create the theoretical barcode, such as optical resolution and stretching factor, might be

not accurate enough. A careful measurement of these parameters is therefore also expected to lead to better results.

Chapter 6

Conclusion and Outlook

In this thesis partial denaturation mapping of DNA in nanochannels has been studied with the goal to use this technique as a tool for bacterial identification. To do that, micro- and nanofluidic devices were used in combination with fluorescence microscopy. The data analysis was done using custom written MATLAB scripts.

Studies of the melting pattern formation could show that the formation of denaturation patterns starts within the first 5 minutes of heating and stabilizes after roughly 10 minutes. By taking the IS (see Eq. (5.1)) as a measure of the melting pattern development and by assuming that the IS follows an exponential rise, the time constant of the melting pattern formation is estimated to be $\tau = 4$ min 20 sec (see Section 5.2.1). A minimum heating time of 10 minutes is thus suggested when denaturing DNA.

By exploring partial denaturation under different levels of confinement of the DNA it could be demonstrated that the melting temperature increases with stronger confinement. Studies of the development of the IS in micro- and nanochennels lead to the conclusion that the difference in melting temperature between micro- and nanochannels is in the range of 5-10 °C (see Section 5.2.2). It is thus important to consider the influence of the confinement on the melting temperature when conducting experiments within strong confinement. More measurements are however needed to determine the exact difference in temperature.

An investigation of the influence of the dye load on the melting temperature in off-chip heating showed that the increase in melting temperature with dye:bp ratio is much stronger when heated off-chip than when heated on-chip (see Section 5.2.3). This should be therefore considered when performing off-chip experiments. More studies on the off-chip heating are needed to answer the question why the dye has this effect.

A comparison between experimentally acquired barcodes and their corresponding theoretical barcodes showed a good agreement between experiment and theory (see Section 5.3.1). The comparison between experiment and theory was completed in Section 5.3.2by automatically generating an experimental temperature dependent denaturation map of T4 DNA and by comparing it to the theoretical one. From these experiments it can thus be concluded that the partial denaturation can satisfyingly be predicted by theoretical models. This is an important prerequisite for the bacterial identification.

The experiments from Section 5.3.2 additionally showed that a spread in the dye load among the DNA molecules leads to a spread in the melting temperature, allowing us to create melting patterns corresponding to different temperatures within the same sample. This can be used as a tool for different future studies.

Several hundreds partially denatured DNA fragments from the *S. pneumoniae* strains TIGR4, D39 and R6 have been measured. Using *DNABarcodeMatchmaker* an attempt was made at identifying the strains. However, an identification of the DNA could not be achieved. An improvement in the experimental procedures as well as *DNABarcodeMatchmaker* might lead to better results. From an experimental point of view, a better control over the heating temperature is necessary. This can be achieved by studying in more detail how exactly the off-chip heating influences the melting temperature. Another way to gain more control over the heating temperature would be by developing a new chip and device holder that allows on-chip heating without any decrease in throughput or image quality. The use of longer DNA fragments and a more careful staining protocol might further improve the results. Finally, determining the input parameters that are needed to generate the theoretical barcode in an accurate way might also lead to better results.

Once a successful identification of the bacteria is achieved, an automation of the whole process would be the next goal to pursue. An outlook into the far future would be to integrate all processes, from the DNA extraction to the identification, into one single commercially available device.

Bibliography

[1] F. Persson and J. O. Tegenfeldt. "DNA in nanochannels—directly visualizing genomic information." *Chemical Society Reviews* 39.3 (2010), pp. 985–999. DOI: 10.1039/B912918A.

[2] W. Reisner, J. N. Pedersen, and R. H. Austin. "DNA confinement in nanochannels: physics and biological applications." *Reports on Progress in Physics* 75.10 (2012), p. 106601. DOI: 10.1088/0034-4885/75/10/106601.

[3] K. Jo, D. M. Dhingra, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest, and D. C. Schwartz. "A single-molecule barcoding system using nanoslits for DNA analysis." *Proceedings of the National Academy of Sciences* 104.8 (2007), pp. 2673–2678. DOI: 10.1073/pnas.0611151104.

[4] S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao. "Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes." *Nucleic acids research* 38.18 (2010), e177–e177. DOI: 10.1093/nar/gkq673.

[5] A. N. Nilsson, G. Emilsson, L. K. Nyberg, C. Noble, L. S. Stadler, J. Fritzsche, E. R. Moore, J. O. Tegenfeldt, T. Ambjörnsson, and F. Westerlund. "Competitive binding-based optical DNA mapping for fast identification of bacteria-multi-ligand transfer matrix theory and experimental applications on Escherichia coli." *Nucleic acids research* 42.15 (2014), e118. DOI: 10.1093/nar/gku556.

[6] R. L. Welch, R. Sladek, K. Dewar, and W. W. Reisner. "Denaturation mapping of Saccharomyces cerevisiae." *Lab on a Chip* 12.18 (2012), pp. 3314–3321. DOI: 10.1039/C2LC40504K.

[7] W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt, and H. Flyvbjerg. "Single-molecule denaturation mapping of DNA in nanofluidic channels." *Proceedings of the National Academy of Sciences* 107.30 (2010), pp. 13294–13299. DOI: 10.1073/pnas.1007081107.

[8] R. H. Garrett and C. M. Grisham. *Biochemistry*. 4th ed. Cengage Learning, 2010. ISBN: 978-0-495-10935-8.

[9] M. T. Bjorndal and D. K. Fygenson. "DNA melting in the presence of fluorescent intercalating oxazole yellow dyes measured with a gel-based assay." *Biopolymers* 65.1 (2002), pp. 40–44. DOI: 10.1002/bip.10220.

[10] H. Li, Z. Wang, N. Li, X. He, and H. Liang. "Denaturation and renaturation behaviors of short DNA in a confined space." *The Journal of chemical physics* 141.4 (2014), p. 044911. DOI: 10.1063/1.4891219.

[11] R. D. Blake and S. G. Delcourt. "Thermodynamic effects of formamide on DNA stability." *Nucleic acids research* 24.11 (1996), pp. 2095–103. DOI: 10.1093/nar/24.11.2095.

[12] C. Sadhu, S. Dutta, and K. Gopinathan. "Influence of formamide on the thermal stability of DNA." *Journal of Biosciences* 6.6 (1984), pp. 817–821. DOI: 10.1007/BF02716841.

[13] W. Reisner, J. Beech, N. Larsen, H. Flyvbjerg, A. Kristensen, and J. Tegenfeldt. "Nanoconfinement-Enhanced Conformational Response of Single DNA Molecules to Changes in Ionic Environment." *Physical Review Letters* 99.5 (2007), p. 058302. DOI: 10.1103/PhysRevLett.99.058302.

[14] F. Persson, F. Westerlund, and J. O. Tegenfeldt. "Fluorescence Microscopy of Nanochannel-Confined DNA." *Single Molecule Analysis*. Vol. 783. Springer, 2011. Chap. 9, pp. 159–179. ISBN: 978-1-61779-282-3. DOI: 10.1007/978-1-61779-282-3_9.

[15] H. S. Rye, S. Yue, D. E. Wemmer, M. A. Quesada, R. P. Haugland, R. A. Mathies, and A. N. Glazer. "Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications." *Nucleic acids research* 20.11 (1992), pp. 2803–2812. DOI: 10.1093/nar/20.11.2803.

[16] K. Günther, M. Mertig, and R. Seidel. "Mechanical and structural properties of YOYO-1 complexed DNA." *Nucleic acids research* 38.19 (2010), pp. 6526–6532. DOI: 10.1093/nar/gkq434.

[17] C. U. Murade, V. Subramaniam, C. Otto, and M. L. Bennink. "Force spectroscopy and fluorescence microscopy of dsDNA–YOYO-1 complexes: implications for the structure of dsDNA in the overstretching region." *Nucleic acids research* 38.10 (2010), pp. 3423–3431. DOI: 10.1093/nar/gkq034.

[18] J. Beech, L. Nyberg, J. Fritzsche, F. Westerlund, and J. O. Tegenfeldt. "What do photons do to fluorescently stained DNA in confinement?" *17th International Conference on Miniaturized Systems for Chemistry and Life Sciences, MicroTAS 2013; Freiburg; Germany; 27 October 2013 through 31 October 2013*. 2014, pp. 5–7. ISBN: 978-163266624-6.

[19] C. E. Aitken, R. A. Marshall, and J. D. Puglisi. "An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments." *Biophysical journal* 94.5 (2008), pp. 1826–1835. DOI: 10.1529/biophysj.107.117689.

[20] C.-C. Hsieh, A. Balducci, and P. S. Doyle. "Ionic effects on the equilibrium dynamics of DNA confined in nanoslits." *Nano letters* 8.6 (2008), pp. 1683–8. DOI: 10.1021/nl080605+.

[21] L. Nyberg, F. Persson, B. Åkerman, and F. Westerlund. "Heterogeneous staining: a tool for studies of how fluorescent dyes affect the physical properties of DNA." *Nucleic acids research* 41.19 (2013), e184–e184. DOI: 10.1093/nar/gkt755.

[22] C. Noble, A. N. Nilsson, C. Freitag, J. P. Beech, J. O. Tegenfeldt, and T. Ambjörnsson. "A Fast and Scalable Algorithm for Alignment of Optical DNA Mappings." *arXiv preprint: 1311.6379* (2013).

[23] R. M. Wartell and A. S. Benight. "Thermal denaturation of DNA molecules: a comparison of theory with experiment." *Physics Reports* 126.2 (1985), pp. 67–107. DOI: 10.1016/0370-1573(85)90060-2.

[24] *T4 GT7 DNA product description*. Nippon Gene. 2012. URL: http://www.nippongene.com/pages/products/clomod/dna_vec/t4gt7/t4gt7_english.htm (visited on 11/04/2014).

[25] M. Reuter and D. T. Dryden. "The kinetics of YOYO-1 intercalation into single molecules of double-stranded DNA." *Biochemical and biophysical research communications* 403.2 (2010), pp. 225–229. DOI: 10.1016/j.bbrc.2010.11.015.

[26] C. Freitag. "Single molecule studies of DNA in fluidic systems." Licentiate Thesis. Lund University, 2013.

[27] W. Reisner, J. P. Beech, N. B. Larsen, H. Flyvbjerg, A. Kristensen, and J. O. Tegenfeldt. "Nanoconfinement-enhanced conformational response of single DNA molecules to changes in ionic environment." *Physical review letters* 99.5 (2007), p. 058302. DOI: 10.1103/PhysRevLett.99.058302.

[28] N. Otsu. "A threshold selection method from gray level histograms." *IEEE Trans. Syst. Man Cybern.* 9.1 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

[29] D. Chiavolini, G. Pozzi, and S. Ricci. "Animal models of Streptococcus pneumoniae disease." *Clinical microbiology reviews* 21.4 (2008), pp. 666–685. DOI: 10.1128/CMR.00012-08.

[30] R. M. Mera, L. A. Miller, J. J. Daniels, J. G. Weil, and A. R. White. "Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States over a 10-year period: Alexander Project." *Diagnostic microbiology and infectious disease* 51.3 (2005), pp. 195–200. DOI: 10.1128/CMR.00012-08.

[31] M. Kanehisa and S. Goto. "KEGG: kyoto encyclopedia of genes and genomes." *Nucleic acids research* 28.1 (2000), pp. 27–30. DOI: 10.1093/nar/28.1.27.

Appendix A

Additional Figures

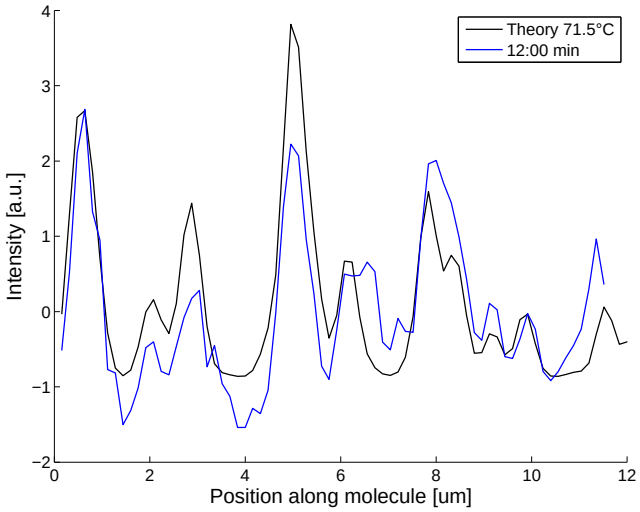


Figure A.1: Comparison of the intensity curve the T4GT7 molecule from Fig. 5.6 at 12 minutes to the theoretical curve corresponding to a heating temperature of 71.5 °C at an ionic strength of 0.02 M. Both curves have mean 0 and standard deviation 1.

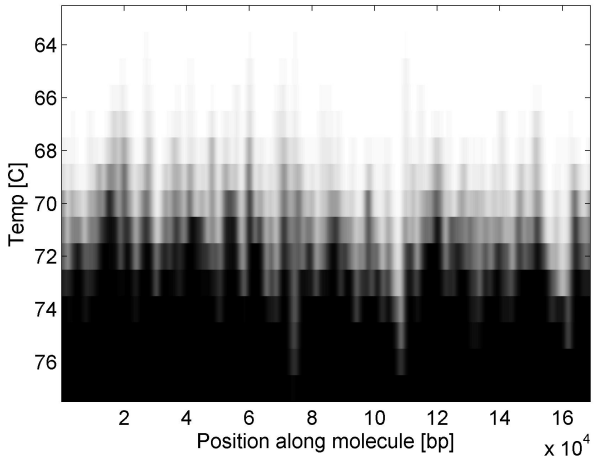


Figure A.2: Theoretical temperature dependent denaturation map of T4GT7 Phage DNA at an ionic strength of 0.028M.

Appendix B

Estimated Amount of Needed Fragments

An estimate of how many DNA fragments have to be measured in order to represent the whole length of a DNA molecule can be done in the following way. Let us assume that all the fragments have the same length l and the total length of the DNA is L. The unit of length is bp and every fragment has the same probability of being measured. An arbitrary basepair along the DNA has the probability of $\frac{l}{L}$ to appear within a fragment of the length l and the probability of not appearing is thus $\frac{L-l}{L}$. The probability that it is not included in n fragments is $\left(\frac{L-l}{L}\right)^n$. Hence, the probability of appearing in at least one of the n fragments is $p = 1 - \left(\frac{L-l}{L}\right)^n$. The value of p can also be interpreted as the average percentage of the total DNA that can be reproduced with n fragments. We can solve for n (see Eq. (B.1)) in order to make an estimate for how many fragments need to measured at least.

$$n = \frac{\log(1-p)}{\log\left(\frac{L-l}{L}\right)} \tag{B.1}$$

A fragment in the nanochannels with a length of 50 pixel (as seen through the microscope) corresponds to $l = 4.28 \cdot 10^4$ bp by considering that 1 pixel = 0.16 µm, 1 bp = 0.34 nm and a stretching factor inside the nanochannels of 0.55 (see Section 5.3.2). For $L = 2$ Mbp (see Section 5.4.1) and $p = 0.99$, i.e., 99% of the DNA represented by the fragments, we get $n = 213$. This means that in average at least 213 fragments of the length of 50 pixel have to be measured in order to represent 99% of the complete DNA. We note that this estimate is rather conservative since the length of 50 pixel is at the lower limit of the observed fragment length.

Appendix C

Protocol for DNA Sample Preparation

This protocol is used to stain and partially denature DNA. The result is a sample of partially denatured DNA ready to be either observed on a glass slide or inside the nanochannels. The values and numbers shown in this protocol are the ones that have been used for the preparation of the TIGR4 DNA fragments in Section 5.4. The procedure for the preparation of DNA from T4GT7, D39 and R6 used in the other experiments is analogous to the one for TIGR4. The protocol is mainly based on ref. [1, 14] and experience from J. P. Beech from J. O. Tegenfeldt's group at Lund University. The protocol describes the off-chip heating procedure (see Section 4.3.2). For on-chip heating simply stop at step 10 and introduce the sample into the chip where it is ready to be heated an observed.

The basic steps of the protocol are:

1. Reduce concentration of original DNA sample to 50 µM by mixing it with TBE buffer.
2. Stain the DNA with YOYO-1 at a dye:bp ratio of 1:10.
3. Add formamide to stained sample in order to have a solution with 50% (v/v) of formamide.
4. In on-chip heating: Introduce the sample into the chip and heat during 10 minutes. Stop the protocol here. In off-chip heating: heat the sample during 10 minutes in a separate heating block or PCR machine. Continue with next step.
5. Cool down and dilute the hot sample to a solution with a salt concentration corresponding to 0.05×TBE by mixing it with an ice-cold solution of milliQ containing 3% (v/v) of BME.

In the following a detailed description of the previous these steps is shown. First different volumes have to be calculated:

1. Calculate how much YOYO-1 (50 µM in 2×TBE) is needed in order to stain 2 µL of the original DNA (334 µM in 1×TE for TIGR4) at the wanted dye:bp ratio (1:10). Here we need 1.34 µL of YOYO-1.

2. Calculate how much TBE buffer (3×TBE) needs to be added to the stained DNA, i.e., 2 µL of DNA (334 µM in 1×TE) + 1.34 µLof YOYO-1 (50 µM in 2×TBE), in order to get a solution with salt concentration corresponding to 2×TBE. Assuming 1×TE≈1×TBE we need 2 µL of 3×TBE.

3. Calculate how much TBE buffer (2×TBE) needs to be added to the solution containing the original DNA solution (334 µM) + the 2 µL of 3×TBE in order to dilute it to 50 µM. Here we need 9.36 µL 2×TBE.

4. After the heating procedure the hot sample will be cooled down with a "cooling solution" (only in off-chip heating). The heated sample will have a volume of 8 µL and a salt concentration corresponding to 1×TBE. The goal is to dilute this salt concentration to 0.05×TBE by adding milliQ water. In addition we add 3% (v/v) of BME to the "cooling solution". Here the cooling solution consists of 174.2 µL of milliQ water and 4.8 µL of BME.

Once all these volumes are calculated we mix them in the following way:

5. Pipette 4.8 µL of BME into a test tube containing 147.2 µL of milliQ water. This will be the "cooling solution". Keep it in ice water until it will be used (only needed in off-chip heating).

6. Pipette 2 µL of the original DNA (334 µM) in tube *A*.

7. Pipette all the calculated volumes of TBE buffer, i.e., 2 µL of 3×TBE + 9.36 µL 2×TBE, in tube *B* and then add the 1.34 µL of YOYO-1 (50 µM). Mix well, vortex, centrifuge and then wait 5 minutes for the dye to evenly distribute.

8. Pipette content of tube *B* slowly into tube *A*.

9. Heat tube *A* for 30 minutes at 40 °C in order for the DNA to get evenly stained. After heating wait for 15 minutes until the tube is back to room temperature.

10. Add formamide to tube *A* with the same volume of then content of the tube in order to get a solution with 50% (v/v) of formamide. Here we need 16.03 µL. Fill and empty the pipette tip a couple of times with the content of tube *A* in order to mix the solution well. Do this very slowly. Then wait 5 minutes in order for the formamide to distribute well.

11. Pipette 8 µL from tube *A* into a PCR tube and heat it with a PCR machine (here at 70 °C) for 10 minutes.

12. Immediately after the heating procedure slowly pipette 152 µL of the "cooling solution" into the PCR tube. Slowly fill and empty the pipette tip a couple of times in order to mix the contents and then put the PCR tube back into the ice water. After 5 minutes the sample is ready to be observed.

When pipetting solutions containing DNA always use pipette tips with wide orifice and pipette very slowly in order to not damage the DNA. Never vortex or centrifuge tubes containing DNA. All tubes containing YOYO-1 must be either non-transparent or wrapped in aluminum foil in order to protect the YOYO-1 from the light. Always use gloves and sterile tools throughout the whole process in order to protect the DNA from any contamination.

Appendix D

Results of D39 and R6 Identification

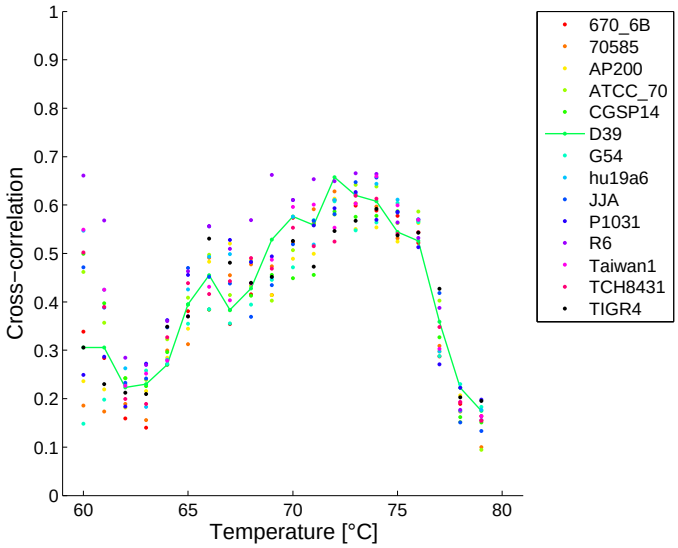


Figure D.1: Cross-correlation C of 343 fragments from D39 DNA matched to theories of different strains calculated for different temperatures.

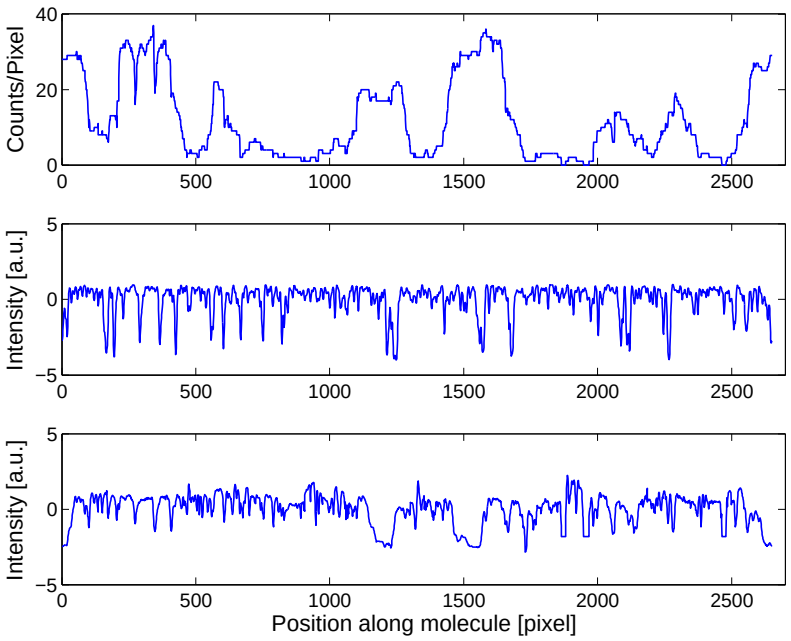


Figure D.2: D39 DNA fragments matched to corresponding theory at 71 °C. The minimum standard deviation is at 66 °C. The temperature at the second lowest standard deviation is chosen, i.e., 71 °C, because at 66 °C almost no melting patterns are developed according to theory. (top) Counts of coverage per pixel (fragment distribution along theory). The total length of the fragments is 32012 pixel. An ideal case would have ∼12 counts per pixel. (middle) Theoretical barcode at 71 °C. (bottom) Assembled barcode.

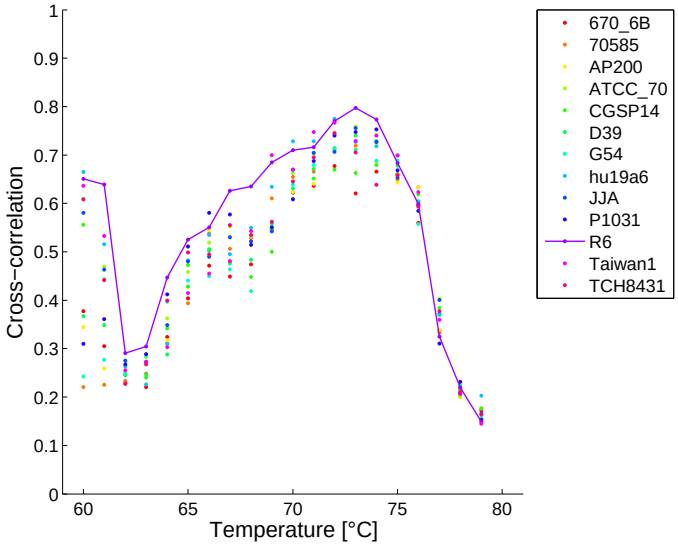


Figure D.3: Cross-correlation C of 343 fragments from R6 DNA matched to theories of different strains calculated for different temperatures.

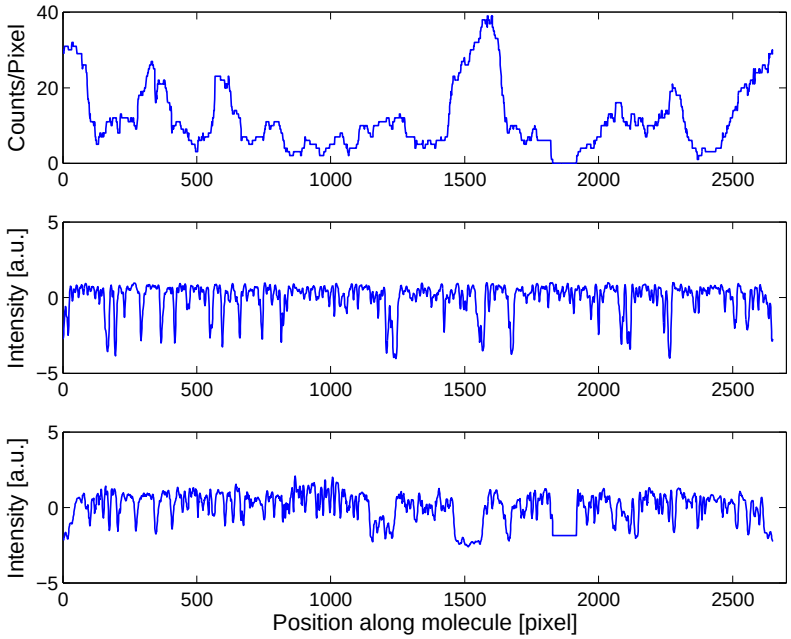


Figure D.4: R6 DNA fragments matched to corresponding theory at 71 °C. (top) Counts of coverage per pixel (fragment distribution along theory). The total length of the fragments is 31504 pixel. An ideal case would have ∼12 counts per pixel. (middle) Theoretical barcode at 71 °C. (bottom) Assembled barcode.

Self-reflection

I enjoyed that my thesis was part of a bigger project with the final objective to develop a diagnostic tool for bacterial infections. This gave me the opportunity to be part of the experienced research team working on this project, allowing me to learn a lot on how to conduct scientific work. Due to the highly interdisciplinary nature of the project I had insight on many different aspects of nanotechnology, biology, physics and computation. Hence, I obtained a broad set of new skills including the preparation of DNA samples, the operation of a fluorescent microscope and the corresponding equipment as well as the handling of microfluidic devices. I also had the chance to develop several MATLAB scripts for the image processing and data analysis. Throughout the thesis I got involved in many interesting scientific discussions which helped me to broaden my horizon extensively. One of the highlights of the thesis was when I had the chance to present the progress of the project as a poster at the Micronano System Workshop in Uppsala (May 2014). There I learned how to present scientific work in a compact an comprehensive way. I am satisfied with the results of my thesis and I am convinced that the learned skills will be useful for my future career as a researcher.