Static and Dynamic Properties of DNA Confined in Nanochannels

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Damini Gupta

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: **Kevin D. Dorfman**

October, 2017

Acknowledgements

I would like to thank my advisor, Prof. Kevin Dorfman, for giving me the opportunity to work on a project that closely aligned with my interests, and for constant encouragement and guidance throughout the duration of the project. I have learnt a lot from him about polymer physics, in particular, and academic life in general. He has also provided me some wonderful learning opportunities with collaborators at both McGill University and Bionano Genomics. Kevin, thank you!

I would like to thank the committee members, Prof. David Wood, Prof. Xiang Cheng and Prof. Joseph Zasadzinski for taking the time to serve on the committee.

Prof. Chetan Gadgil and Prof. Günter Reiter are the two mentors that I would like to acknowledge for inspiring me to pursue a PhD. Thank you so much for your influence!

Thanks to the Prof. Walter Reisner and his former student Dr. Sara Mahshid for their contribution to the projects presented in this dissertation and teaching me nanofabrication.

Thanks to Dr. Jeff Reifenberger for teaching me how to build a laser based microscope.

I have had incredible coworkers during my time in the Dorfman research group. I feel so fortunate to have shared this experience with all of you: Jeremy Miller, Allanah Austin, Mark Pagkaliwangan, Derek Huang, Abhiram Muralidhar, Teresa Sells, Akash Arora, Scott White, Pranav Agrawal, Xiaolan Li, Seunghwan Shin, Sarit Dutta, Doug Tree, Scott King, Joel Thomas, Aashish Jain, Michael McGovern, Hui-Min Chuang, Mathew Thomas, Aditya Bhandari, Guo Kang Cheong, and Paridhi Agrawal. Special shoutout goes to Julian Sheats and Zhicheng Long, who mentored me and taught me valuable research skills when I first joined the group.

Thanks to my CEMS and out-of-department friends, Pratibha, Zeinab, Iyer, Sneha, Deeksha, Dhruv, Monalisa, Mayuree, Samrat, Devansh, Meera, Manjiri, and Palak for helping me keep some of my sanity and most of my dark hair.

I am really grateful to my dad and sisters, Pratiksha and Shivani, for being the rock that I could lean on for one of the most difficult times in our life. Thank you *papa* and *didi* for being alive and in my life, I love you all equally.

Finally, I would like to acknowledge financial support from the University of Minnesota's Doctoral Dissertation Fellowship, the National Science Foundation (DMR-1206045), the National Institutes of Health (R01-HG006851) and the department of Chemical Engineering & Materials Science. Some of the fabrication steps were performed in the Minnesota Nano Center at the University of Minnesota, which receives partial support from the NSF through the NNIN.

This dissertation is dedicated to loving memory of my mother,

Sunita K. Gupta

Although she was my inspiration to pursue a doctoral degree,

she was unable to see my graduation. This is for her.

Abstract

Next-generation sequencing (NGS) techniques have considerably reduced the cost of high-throughput DNA sequencing. However, it is challenging to detect large-scale genomic variations by NGS due to short read lengths. Genome mapping can easily detect large-scale structural variations because it operates on extremely large intact molecules of DNA with adequate resolution. One of the promising methods of genome mapping is based on confining large DNA molecules inside a nanochannel whose cross-sectional dimensions are approximately 50 nm. Even though this genome mapping technology has been commercialized, the current understanding of the polymer physics of DNA in nanochannel confinement is based on theories and lacks much needed experimental support. The results of this dissertation are aimed at providing a detailed experimental understanding of equilibrium properties of nanochannel-confined DNA molecules.

The results are divided into three parts. In first part, we evaluate the role of channel shape on thermodynamic properties of channel-confined DNA molecules using a combination of fluorescence microscopy and simulations. Specifically, we show that high aspect ratio of rectangular channels significantly alters the chain statistics as compared to an equivalent square channel with same cross-sectional area. In the second part, we present experimental evidence that weak excluded volume effects arise in DNA nanochannel confinement, which form the physical basis for the extended de Gennes regime. We also show how confinement spectroscopy and simulations can be combined to reduce molecular weight dispersity effects arising from shearing, photo-cleavage, and nonuniform staining of DNA. Finally, the third part of the thesis concerns the dynamic properties of nanochannel confined DNA. We directly measure the center-of-mass diffusivity of single DNA molecules in confinement and show that that it is necessary to modify the classical results of de Gennes to account for local chain stiffness of DNA in order to explain the experimental results.

In the end, we believe that our findings from the experimental test of the phase diagram for channel-confined DNA, with careful control over molecular weight dispersity, channel geometry, and electrostatic interactions, will provide a firm foundation for the emerging genome mapping technology.

Contents

List of Tables

List of Figures

Statement of the Author's Contributions

Much of the research contained in this dissertation was performed collaboratively. Some chapters have appeared (or will appear) as articles in various journals. Because many of the journal articles list multiple authors, I would like to clarify my particular contributions to the research.

I wrote most of **Chapter 2**. One figure in this Chapter is reproduced from E Werner, G. Cheong, D Gupta, K. Dorfman, and B Mehlig, "One-parameter theory for DNA extension in a nanochannel", arXiv preprint **1705.04619** (2017). I curated the data for this figure.

Chapter 3 is based on D. Gupta, J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, and W. Reisner, "Mixed confinement regimes during equilibrium confinement spectroscopy of DNA", J. Chem. Phys. **140**, 214901 (2014). In this publication, we performed experiments to study the effect of shape of nanochannels on thermodynamic properties of DNA. Next, we performed simulations to explain the experimental results. SM fabricated the nanofluidic devices. I conducted all the experiments and analyzed most of the data from them. AM performed simulations. I chose the simulations parameters and analyzed the results. All the authors shared interpretation of the results jointly.

Chapter 4 is based based on D. Gupta, J. J. Miller, A. Muralidhar, S. Mahshid, W. Reisner, and K. D. Dorfman, "Experimental evidence of weak excluded volume effects for nanochannel confined DNA", ACS Macro Lett. **4**, 759

(2015). In this publication, we performed experiments to investigate the extended de Gennes regime of confinement for DNA. Additionally, we used a combination of experiments and simulation to reduce the effect of polydispersity of DNA samples. SM fabricated the nanofluidic devices. JJM and I performed all the experiments. I analyzed all the data from them. AM performed simulations. I chose the simulations parameters and analyzed the results. All the authors shared interpretation of the results jointly.

Chapter 5 is based based on D. Gupta and K. D. Dorfman, "Diffusion of DNA in nanochannels", to be submitted (2017). I fabricated the devices, conducted all the experiments and analyzed all the data. KDD and I interpreted the results.

Chapter 1

Introduction

.

It is well known that the wealth of information required for the existence of life is encoded in the base pair sequence of a deoxyribonucleic acid (DNA) molecule. Accordingly, a massive effort has taken place to determine the sequence of these base pairs (bp) in the form of human genome project (HGP) (Venter et al., 2001). In United States alone, the total monetary contribution to the HGP was $2.7 billion (*The Cost of Sequencing a Human Genome* 2016). This project, launched in 1990 and completed in 2003, had a tremendous impact on our current understanding of how certain sequences of protein-coding DNA (genes) work together to direct the growth, development and maintenance of an entire organism. It also gave important insights into the genetic similarities and differences within humans and with other species (Przeworski et al., 2000). At the end of this endeavor, 99% of euchromatic portions of human DNA have been sequenced accurately with huge gaps remaining only in heterochromatic portions (estimated to be 200 mega bp)

(Eichler et al., 2004). The human genome draft developed in the HGP is used as a reference map for the present day sequencing of a human genome. The ultimate goal of the field is not to obtain a collection of consensus sequences, rather move beyond to catalogue genetic variation and draw connections to diseases and phenotypes. It was clear from the years of effort invested in this large-scale international project that the improvements in existing techniques to sequence DNA were required to detect variations for cheaper cost at much faster rates.

The major technique that contributed to the HGP was Sanger sequencing. This technique burst onto the genetics scene in the late 1970s and researchers came to regard it as the gold standard for DNA sequencing (Sanger et al., 1977). Its major advantages included high accuracy and reduction of toxic chemicals compared to the previous sequencing methods. However, Sanger sequencing is a costly and time-consuming process. To sequence a human genome, which is approximately 3 billion bp long, the DNA had to be broken down into 100 kilo base pairs (kbp) fragments. These fragments were amplified using molecular cloning and further broken down to 2-3 kbp for sequencing. The DNA fragments were physically separated by capillary electrophoresis (Karger and Guttman, 2009), and automation and parallelization was difficult and expensive.

1.1 Next and third generation sequencing

With goal of improving the drawbacks of the Sanger sequencing techniques, the next generation sequencing (NGS) techniques were developed. Rather than performing fragmentation-amplification steps twice in the Sanger sequencing meth-

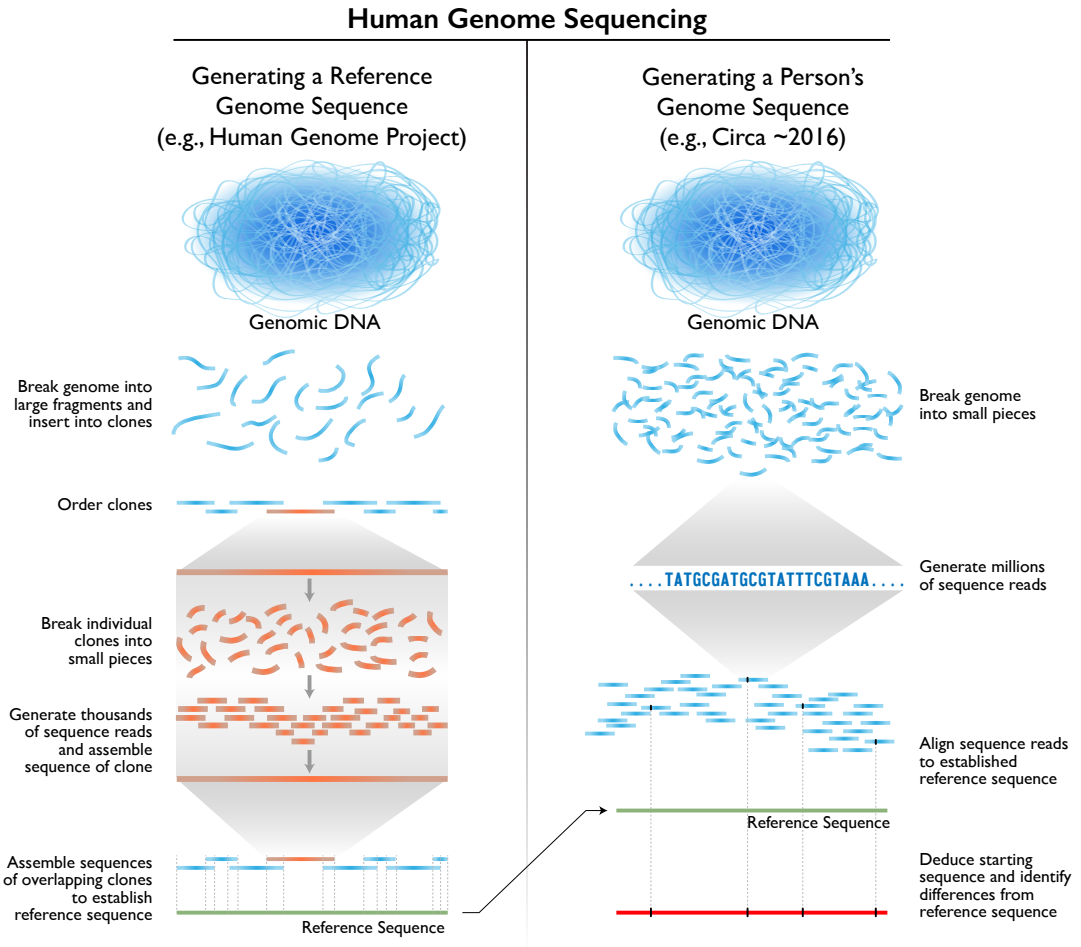

Figure 1.1. Comparison between sequencing workflow used in the human genome project and sequencing workflow circa 2016 based on the next generation sequencing (NGS) techniques. The NGS techniques rely on a single fragmentation-amplification process, followed by massively parallelized and automated sequencing. Reproduced from *The Cost of Sequencing a Human Genome* (2016).

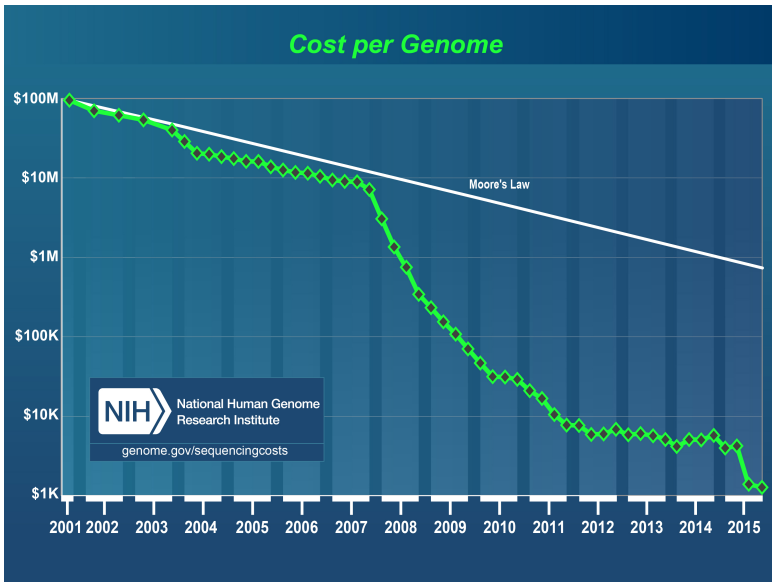

Figure 1.2. Cost of sequencing a genome in the last two decades. Reproduced from Wetterstrand (2016).

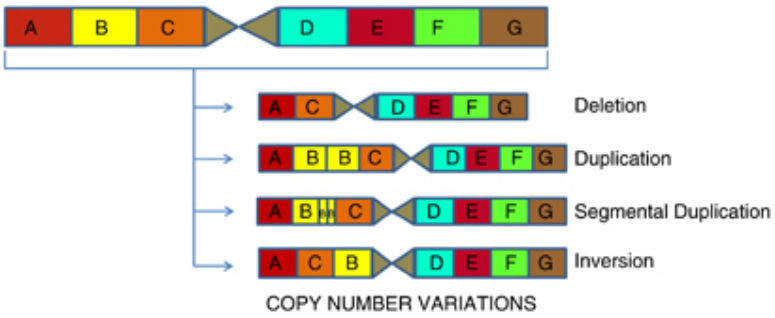

Figure 1.3. Genomic variants in form of copy number variations can be classified as deletion, duplication, segmental duplication and inversion. These variations can encompass the entire gene or a segment of a particular gene represented in the figure. Reproduced from Almal and Padh (2012).

ods, NGS uses a single fragmentation-amplification step as seen in Figure 1.1. Both the cost and the time for sequencing were reduced by a massive parallelization of both amplification and sequencing steps (Shendure and Ji, 2008; Mardis, 2008; Ansorge, 2009; Metzker, 2010). This was achieved by locally amplifying DNA fragments in a parallel fashion by ensuring that the amplified products remain spatially clustered. Then, sequencing by a synchronized synthesis approach was followed to locally read the sequence of amplified clusters, thus reducing the cost dramatically. The National Human Genome Research Institute reports a drastic five order of magnitude decrease in the cost of sequencing a genome in just a fifteen years span (Figure 1.2).

Despite the amazing achievements of the NGS techniques, they suffer from many limitations such as loss of the large-scale ordering of the genetic information due to inherent amplification bias, short read lengths and sub-kbp fragmentation process. In fact, the read lengths achieved by the NGS techniques ($<$ 250 bp) (Shendure and Ji, 2008; Metzker, 2010) are even smaller than the traditional Sanger sequencing techniques ($\sim$ 1000 bp) (Karger and Guttman, 2009). As a result, detection of long-range structural variations (Feuk et al., 2006) such as copy-number variations as shown in Figure 1.3, and *de novo* assembly can be challenging (Earl et al., 2011).

Third generation sequencing techniques, such as single molecule real time sequencing (commercialized by PacBio) (Levene et al., 2003), linked-read sequencing (commercialized by 10$\times$ Genomics) (Zheng et al., 2016; Eisenstein, 2015), or nanopore sequencing (commercialized by Oxford Nanopore Technology) (Deamer et al., 2016) can provide much longer read lengths. However, they can either ac-

curately read smaller fragments (up to 3 kbp) or have very high errors for longer reads (~ 15%) (Goodwin et al., 2016). Indeed, they will have a blind spot towards much larger genomic rearrangements of tens of kilobase pair to megabase pairs of DNA, which can result in multiple phenotypes and disease states (Feuk et al., 2006). Additionally, it is not clear that these novel technologies, with their base pair level precision, alone can efficiently illuminate the large-scale genomic oddities due to *data deluge* problem suffered by them (*DNA Sequencing Caught in Deluge of Data*).

1.2 Optical mapping

Alternatively, optical mapping techniques which were developed to provide a coarser view of the base pair sequence, can address drawbacks of the next, and the third generation sequencing techniques (Persson and Tegenfeldt, 2010; Levy and Craighead, 2010; Dorfman et al., 2013; Levy-Sakin and Ebenstein, 2013; Bogas et al., 2017; Müller and Westerlund, 2017). Pioneered by Schwartz et al. (1993), they have become a versatile tool to reveal large-scale sequence information of DNA. In its earliest form, a long DNA molecule of order of 200 kbp was stretched in an agarose gel and was cut using restriction enzymes. An optical map was created by observing the distance between repeats of a given sequence (recognized by restriction enzymes), under a fluorescence microscope. A main advantage with this technique is that the genomic information is read from very large pieces of DNA, up to mega-basepairs. Since then, the optical mapping techniques have been continuously evolving both in terms of the method used to identify the repeats of a

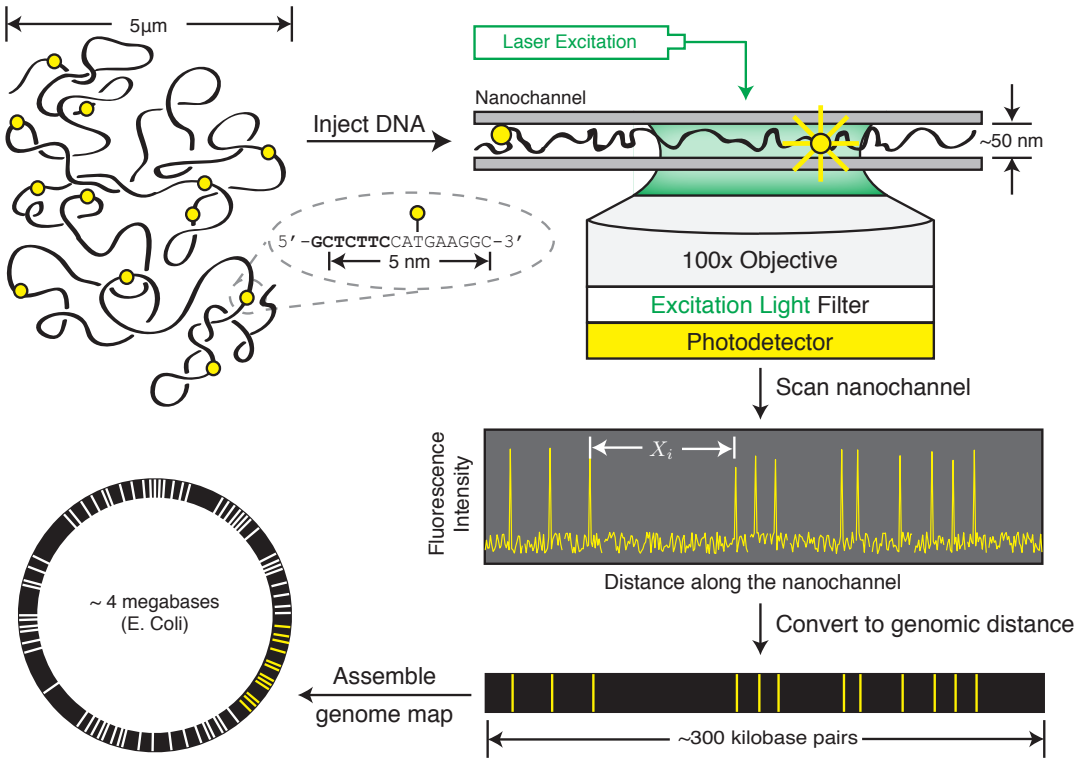

Figure 1.4. Schematic of optical mapping using nick-labelling via stretching DNA inside a nanochannel. Reproduced from Dorfman et al. (2014).

given sequence as well as the stretching techniques. Figure 1.4 illustrates optical mapping workflow for an E. Coli genome using nick-labelling method, in which a long fragment of DNA is stretched inside a nanochannel.

The most popular method to identify the repeats, called nick-labelling, was first demonstrated by Jo et al. (2007) in a nanofluidic environment. Since the output from nick-labelling resembles a barcode, it is also commonly known as DNA barcoding. The first step of the process is depicted in the upper left quadrant of Figure 1.4, where a fragment of DNA much larger than 1 kbp is taken. Then, a nicking enzyme is used to create sequence-specific single-strand nicks in the DNA. A DNA polymerase subsequently removes and at the same time replaces the nucleotides along the nicked strand. The incorporated nucleotides are linked with fluorophores that allows direct visualization of the nicking sites using fluorescence microscopy. After nick labeling, the entire DNA backbone is stained with a different fluorescent dye, to facilitate the detection of the nicking sites. The molecule of DNA can be stretched using various techniques so that the nicking sites can be imaged and resolved. For example in the upper right quadrant Figure 1.4, the DNA fragment is elongated by a very small channel with width on the order of 50 nm. The entire stretching process is done in a parallel manner by using an array of small channels. The stretched molecule is then imaged using fluorescence microscopy by sequentially illuminating the backbone and fluorophores. The signal of the fluorescence intensity is processed to obtain a measure of the genomic distance between nicking sites by measuring the physical distance between them along the nanochannel, as shown in the bottom right quadrant. Finally, in the bottom left quadrant, many overlapping DNA barcodes from different fragments

of genomic DNA are assembled into a genome map.

DNA molecules are present in a coiled state in their natural form and straightening them out through an external stimulus is the key physical process required for the optical mapping techniques to work successfully. The DNA could be straightened out using multiple ways: (i) dynamic molecular combing, where DNA is stretched by a receding contact line on a surface that preferentially absorbs DNA at one of its ends (commercialized by Genomic Vision) (Bensimon et al., 1994), (ii) extensional flow, wherein the DNA molecule is stretched by an elongational flow in a microfluidic device (commercialized by OpGen)(Chan et al., 2004), and (iii) nanochannel confinement, where the DNA molecule is stretched by injecting the DNA molecule in a nanochannel whose width is of order of 50 nm (commercialized by Bionano Genomics) (Lam et al., 2012). However, the nanochannel confinement offers the most homogeneous and efficient stretching of DNA, which is crucial to maximize the information that can be obtained from optical DNA maps (Müller and Westerlund, 2017). The DNA molecules in this method are in an equilibrium state as the stretching is due to confinement by the walls. In contrast, other methods involve measurements at a stretched state far from equilibrium, typically produced by some kind of flow.

The nanochannel based optical mapping technology, in part due to the availability of various commercialized platforms to the researchers from Bionano Genomics, has achieved multiple milestones in span of last five years. It is rapidly becoming an auxiliary tool which in combination with next and third generation sequencing techniques, is giving unprecedented window into genomic information. For more details, please see excellent review articles by Bogas et al. (2017) and

Müller and Westerlund (2017).

1.3 Research outline

The optical mapping techniques which are based on confining DNA inside nanochannels, have great potential to advance genome mapping, as is clear from both the wide range of applications and the commercialization of the technology by BioNano Genomics (Bogas et al., 2017; Müller and Westerlund, 2017). Unfortunately, the behavior of DNA in confinement is not thoroughly understood. Initially, it was believed that roughly four decade old theories for cases of strong confinement (Daoud and de Gennes, 1977) and weak confinement (Odijk, 1983) could completely describe the behavior of DNA for the entire range of confinement. The seminal experiments by Reisner et al. (2005) pointed out the inadequacy of the classical theories to explain the majority of the stretching behavior. Subsequently, a theoretical framework (Odijk, 2008), supported by intensive simulation data (Wang et al., 2011; Dai et al., 2014; Werner and Mehlig, 2014; Muralidhar et al., 2014b), was developed to explain this inconsistency. As a result, the existence of four distinct regimes as a function of channel size has been proposed.

Even though the theoretical regimes gives us insights into possible conformations and equilibrium properties of DNA in confinement, there is a lack of experimental data to substantiate them. Additionally, details of the regimes are very subtle and their translation to experimental data is not straightforward, owing to interplay of multiple length scales originating from electrostatics (charge on DNA backbone and walls of the channel), channel size and DNA molecule size. *As such,*

the goal of this dissertation is to provide an experimentally validated model for the entire confinement spectrum of DNA. In the process, we would like to find answers to some fundamental questions surrounding the progression of confinement of DNA from bulk to fully stretched state.

In line with the above goals, **Chapter 2** summarizes the current state of the literature surrounding the physics of DNA in confinement. We will start with the review of the reasons behind using DNA as a model polymer to answer fundamental questions in polymer physics. In order to compare the theories with experimental results obtained from DNA as the model system, the physical parameters describing DNA will be discussed. The focus will be on the challenges and the assumptions associated with the determination of these physical parameters. We then proceed with the presentation of the four distinct regimes which are proposed to describe the confinement process of DNA inside nanochannels. Finally, we will discuss the experiments done so far with DNA in nanochannels.

Fabrication of multiple channel sizes are required to experimentally test the recently proposed regimes. The ideal approach would be to make almost square channels of varied sizes since most theories are intended for the square channels. However, it can be time consuming and costly. Alternatively, confinement spectroscopy can provide multiple channel size on a single device with massive reduction in both time and associated cost. A potential drawback of such devices is the unknown consequences of the use of high aspect ratio channels to obtain precise channel sizes. As such, **Chapter 3** focus on the role of aspect ratio of the channel on the thermodynamic properties of confined DNA. We have used a combination of fluorescence microscopy experiments and Pruned Enriched Rosenbluth

Method simulations of a discrete wormlike chain model to measure the average extension and the fluctuation about that mean extension of λ-DNA in 100 nm deep nanochannels with widths ranging from 100 nm to 1000 nm in discrete 100 nm steps. The average extension is only weakly affected by the channel aspect ratio. In contrast, the fluctuations of the chain extension qualitatively differ between rectangular channels and square channels with the same cross-sectional area, owing to the "mixing" of different confinement regimes in the rectangular channels. As a result, confinement spectroscopy at high aspect ratio is probably not the best tool for probing the different theories for channel-confined DNA. Additionally, both our experiments and simulations clearly indicate that the fluctuations of the chain extension continues to increase with increase in aspect ratio. These findings have important consequences for use of nanochannels for high-throughput DNA mapping. While high aspect ratio channels are cheaper to fabricate, high fluctuations associated with them may adversely impact the barcode alignment accuracy.

With the new knowledge gained about the confinement spectroscopy device, **Chapter 4** is focused on probing a newly proposed regime, called the extended de Gennes regime for nanochannel confined DNA. We present experimental demonstration that weak excluded volume effects, that form the physical basis for the extended de Gennes regime, arise in DNA nanochannel confinement. In particular, by performing measurements of the variance in chain extension as a function of nanochannel dimension for effective channel size ranging from 305 nm to 453 nm, we show that the scaling of the variance in extension with channel size rejects the de Gennes scaling in favor of the extended de Gennes scaling at the 95% con-

fidence level. We also show how simulations and confinement spectroscopy can be combined to reduce molecular weight dispersity effects arising from shearing, photocleavage, and nonuniform staining of DNA.

In **Chapter 5**, we switch gears from chain statistics to dynamic properties of DNA confined in nanochannels. We have measured the diffusivity of λ-DNA molecules in approximately square nanochannels with effective size ranging from 123 to 260 nm at moderate ionic strength. We found that it is necessary to modify the classic results of de Gennes theory to account for local chain stiffness in order to explain the experimental results for DNA in practical channel sizes. Our results are in quantitative agreement with the predictions of previous simulations of the Kirkwood diffusivity of a discrete wormlike chain model, without the need for any fitting parameters, thereby lending confidence in usage of Kirkwood-Riseman approach, to simulate dynamics of channel-confined DNA.

Finally, **Chapter 6** restates the important points and conclusion drawn from the research presented in this dissertation. We then discuss possible avenues for future work. In particular, this chapter highlights the possibilities for practical use of the experimental tools developed in this dissertation to answer several outstanding questions in the field.

Chapter 2

Background

This chapter provides a brief background and reviews recent literature concerning the physics of confined DNA which is necessary to understand the material presented in the following chapters. For details, please see excellent review articles by Reisner et al. (2012) and Dai et al. (2016).

2.1 DNA as a model polymer

A polymer by definition is a macromolecule made up of smaller repeating chemical units. DNA, therefore, being entirely made up of the four nucleotides is a polymer. In the past few decades, double stranded DNA (dsDNA) molecules have been extensively used as a model system to investigate polymer physics at the single molecule level (Shaqfeh, 2005; Latinwo and Schroeder, 2011). Their effectiveness as a model polymer arises from their high compatibility with fluorescence microscopy (Glazer and Rye, 1992; Gurrieri et al., 1997), their mesoscopic length

scales and time scales, and their availability in relatively monodisperse form via viral genomes (λ-DNA, T4-DNA).

DNA has been traditionally used as the model system for investigating equilibrium, and non-equilibrium dynamics of polymer chains in solutions (Shaqfeh, 2005). With advances in microfabrication, and nanofabrication techniques, it is now possible to fabricate channels and slits with well defined geometries, and characteristic dimensions ranging from hundreds of nanometers down to the 50 nm length scale corresponding to the persistence length of DNA. This has allowed a direct visualization of confined DNA. As a result, DNA has further been adopted as the model polymer for studying the physics of confined polymers. Indeed, DNA has been used to gain insights into equilibrium properties of polymers in channel confinement (Tegenfeldt et al., 2004; Reisner et al., 2005; Reisner et al., 2007; Zhang et al., 2008; Thamdrup et al., 2008; Persson et al., 2009; Utko et al., 2011; Kim et al., 2011; Su et al., 2011; Werner et al., 2012; van Kan et al., 2012; Nyberg et al., 2013; Frykholm et al., 2014; Gupta et al., 2014; Alizadehheidari et al., 2015; Iarko et al., 2015; Reinhart et al., 2015; Gupta et al., 2015; Sheats et al., 2015; Reifenberger et al., 2015), and slit confinement (Balducci et al., 2006; Hsieh et al., 2007; Hsieh et al., 2008; Strychalski et al., 2008; Tang et al., 2010; Lin et al., 2011) for a vast number of studies.

2.2 Fundamental physical properties of DNA and their estimation

The chemical and geometric structure of a DNA molecule is highly complex. For simplifying the modeling of DNA in nanochannel confinement while retaining the essential physics of the system, DNA is often treated as a neutral, homogenous, and semiflexible polymer. This model is fully specified by four parameters: contour length L, persistence length l_{p}, effective width w, and effective channel size D_{eff}.

2.2.1 Contour length

The contour length, L, is the length of the DNA molecule stretched from end-to-end. It is simply given by total number of base pairs in the molecule multiplied by the average rise of the base pair. For example, the most commonly used DNA for polymer physics studies, λ-DNA (48.5 kbp) has a contour length of 16.5 μm. This calculation is for an undyed DNA. In order to make the DNA visible in a fluorescence microscope, the molecules are typically stained with bisintercalating dye like YOYO-1. The dye is expected to increase L linearly in proportion to the amount of dye bound due to intercalation of dye between base pairs. It is estimated that every bound dye molecule increases L by about 0.5 nm (Günther et al., 2010; Kundukad et al., 2014).

2.2.2 Persistence length

The persistence length, l_{p} measures the stiffness of the molecule. There are multiple definitions associated with this length scale depending upon its mathematical description. The persistence length can be defined as the length scale at which the bending energy of the backbone is of the order of $k_{\mathrm{B}}T$, where k_{B} is the Boltzmann constant, and T is the absolute temperature (Rubinstein and Colby, 2003). It is also defined as the contour over which the DNA molecule forgets its orientation or more formally it determines the length scale over which the tangent-tangent correlation function along the molecule contour length decays (Doi and Edwards, 1988). Experimentally, it is measured using light scattering experiments, single-molecule elasticity studies or with atomic force microscopy.

The persistence length depends on the concentration of ions in the solvent, the amount of dye bound and the underlying base-pair sequence. While the effect of the dye and the underlying sequence on the persistence length are highly active areas of research, they are expected to change l_{p} by a very small amount (Reisner et al., 2012; Kundukad et al., 2014). On the other hand, the ionic strength can significantly change l_{p}. For reference, the persistence length of double stranded DNA in a high ionic strength buffer (about 100 mM of monovalent salt), $l_{\mathrm{p}} = 50$ nm and in a low ionic strength buffer (about 5 mM of monovalent salt), $l_{\mathrm{p}} = 73$ nm (Dobrynin, 2005). For our baseline scenario when comparing experiments to the theory, we assume that the persistence length is only affected by the ionic strength while neglecting other dependencies. For detailed discussion on calculation of the ionic strength and the persistence length, see Chapter 3.

2.2.3 Effective width

The effective width, w quantifies the repulsive interactions between different parts of the molecule and will be dependent upon the ionic strength of the solution. It basically maps the interactions between charged segments of DNA to an equivalent neutral polymer with purely hard-core repulsion. The standard calculation of the effective width was carried out by Stigter (1977) by relating the second-virial coefficient of hard cylinders to that arising from solving the Poisson-Boltzmann equation for DNA. In a high ionic strength buffer, we find this width is approximately 5 nm.

2.2.4 Effective channel size

There exist a repulsive interaction between the charged segment of DNA and the charged channel walls. As such, the actual distance available to DNA, to move freely inside the channel is less than the physical size of the channel, D. The effective channel size of the channel, D_{eff} is often approximated as $D - \delta$. The parameter δ is a wall-DNA depletion length that models the electrostatic interaction of the DNA with the walls. We estimate that $\delta \approx w$. The accuracy of this approximation is unknown.

2.3 Confinement of DNA in nanochannel

The statics and dynamics of a confined polymer were well described analytically by Daoud and de Gennes (1977) and Brochard and De Gennes (1977) quite some

time ago using scaling concepts. When a flexible polymer is confined in two dimensions in a channel size, D, which is smaller than its radius of gyration in solution, R_g ($D < R_g$), it forms a series of isometric compression blobs of diameter D (Daoud and de Gennes, 1977), as illustrated in Figure 2.1(d). This confinement regime is known as de Gennes regime. The average chain extension scales like $X \sim D^{-2/3}$. On the other hand, in the strongest confinement limit for $D \ll l_p$, a stiff chain ($l_p/w \gg 1$) is deflected back and forth by the channel walls (Odijk, 1983) as seen in Figure 2.1(a). The Odijk scaling theory predicts $X/L = [1 - 2\alpha(D/l_p)^{2/3}]$. The quantity α here is a constant, which has been determined numerically (Yang et al., 2007; Burkhardt et al., 2010) and estimated analytically (Jo et al., 2007).

The possibility of fabricating nanochannels at length scales of the order of l_p of DNA and the ability of a DNA molecule to be visualized with fluorescence microscopy, resulted in experimental efforts to directly test the Odijk and de Gennes regime predictions in nanochannels (Tegenfeldt et al., 2004; Reisner et al., 2005; Reisner et al., 2007; Zhang et al., 2008; Persson et al., 2009; Su et al., 2011; Werner et al., 2012). Seminal work by Reisner et al. (2005) measured an apparent scaling law for the average chain extension, $\langle X \rangle \sim D^{-0.85}$ for $D > l_p$ suggesting a possible shortcoming in the blob theory predictions, which predicted a scaling of $X \sim D^{-2/3}$ (Daoud and de Gennes, 1977). This discrepancy between theory and experiments resulted in a large amount of theoretical and simulation work in recent times for DNA confined in nanochannels.

The challenge in developing a theory for the extension of nanocofined DNA arises from the semiflexible nature of DNA. For reference, $l_\mathrm{p} \approx 50$ nm and $w \approx 5$ nm for DNA in high ionic strength solutions and thus, $l_\mathrm{p} \gg w$. This inequality

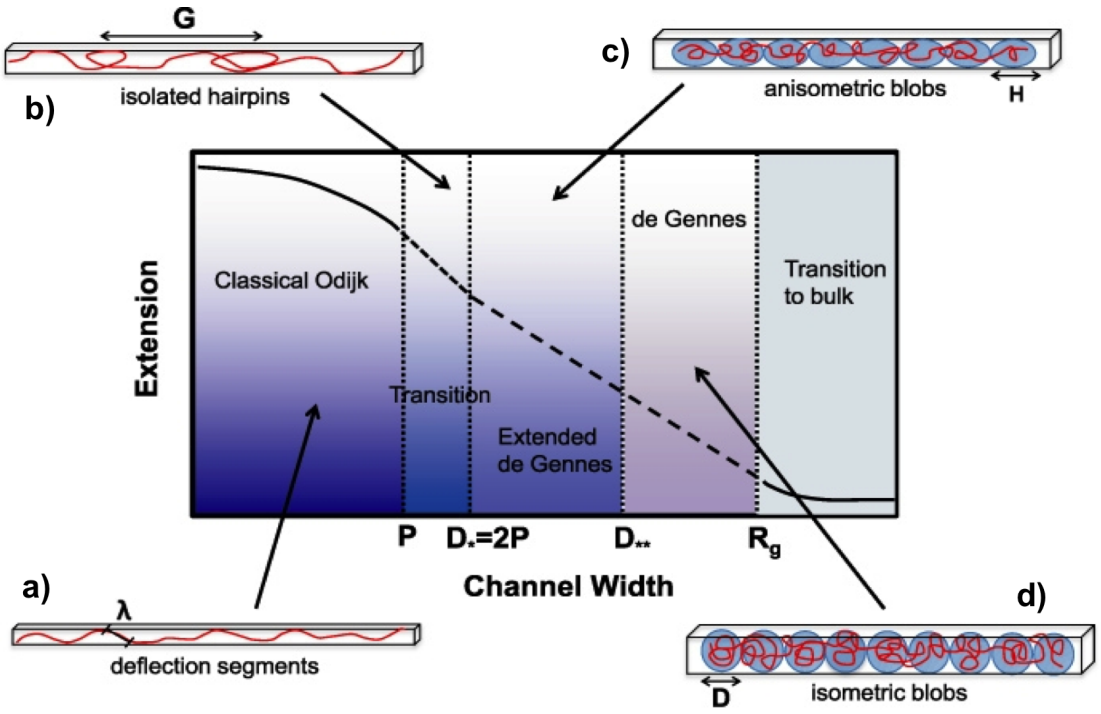

Figure 2.1. Schematic illustration of the different regimes of DNA confinement in a square nanochannel. The scaling arguments by Odijk (2008) refer to four distinct regimes in which the chain conformations are fundamentally different: (a) the Odijk regime, (b) the backfolded Odijk regime (transition regime with hairpins), (c) the extended de Gennes regime, and (d) the de Gennes regime. Reproduced from Reisner et al. (2012).

emphasizes the intrinsic difficulty of describing DNA in a wide range of situations. DNA is considerably stiffer than typical synthetic polymers ($l_\mathrm{p} \approx w$), yet the number of persistence lengths L/l_p in genomic DNA samples is large. Any theory for the conformational statistics of DNA must account for both the local stiffness of the polymer, as well as global excluded-volume interactions.

Recently, Odijk (2008) tried to explain the discrepancy between theory and experiments by suggesting the existence of two new regimes in between the Odijk regime of strong confinement and the de Gennes regime of weak confinement within a scaling analysis for a semiflexible polymer. As the channel width increases beyond the Odijk regime, the chain enters the backfolded Odijk regime. Here, the channel width is still small enough to produce deflection segments, but it is large enough to allow for formation of hairpin folds, as illustrated in Figure 2.1(b). When hairpins form, the polymer chain experiences excluded-volume interactions. Thereafter, with further increase in channel size, the chain enters the blob regime. If the channel width is not too large, $D < 2.24 l_\mathrm{p}^2/w$ (Odijk, 2008; Dai et al., 2014), the excluded volume interactions inside the blob are too weak for spherical blobs of diameter D to be mutually excluding. Instead, in this extended de Gennes regime, the statistics are analyzed in terms of oblong blobs as illustrated in Figure 2.1(c).

At the beginning of this dissertation research, the experimental data were at odds with theories, specially as a function of channel size. The reason behind this inconsistency were two fold. First, most experimental studies either did not have knowledge of these two new regimes or failed to recognize the dense phase-space of channel confined DNA. Indeed, the experiments done so far are mostly concentrated in the extended de Gennes regime followed by the backfolded Odijk

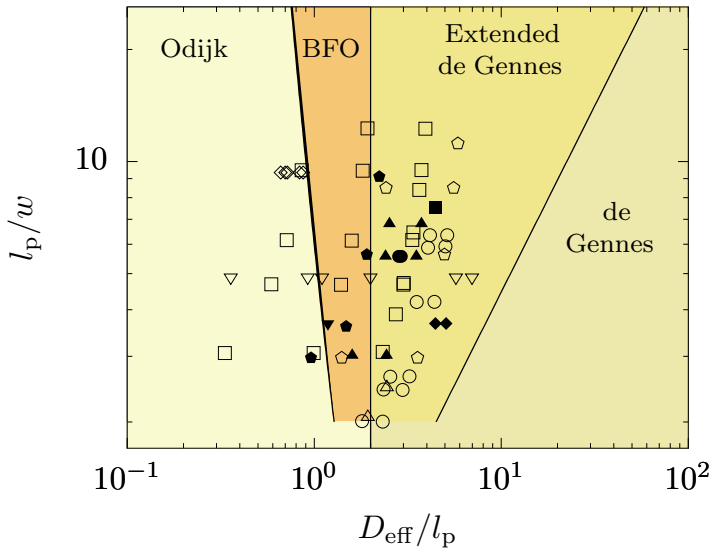

Figure 2.2. Phase diagram and experimental studies for DNA confinement within a square nanochannel: Reisner et al. (2005) (▽), Reisner et al. (2007) (□), Thamdrup et al. (2008) (■), Zhang et al. (2008) (❍), Utko et al. (2011) (●), Kim et al. (2011) (△), Werner et al. (2012) (▲), Gupta et al. (2014) (▼), Reinhart et al. (2015) (◇), Gupta et al. (2015) (◆). Some experiments use funnels (Utko et al., 2011; Werner et al., 2012; Gupta et al., 2015); only the maximum and minimum channel widths are indicated. The condition for the boundary between the Odijk regime and the backfolded Odijk (BFO) regime is described in Odijk (2008) and Muralidhar et al. (2014a), the boundary between backfolded Odijk and extended de Gennes regimes is $D_{\text{eff}} = 2l_{\text{p}}$ (Odijk, 2008) and the boundary between the extended de Gennes and de Gennes regimes is $D_{\text{eff}} = 2.24\, l_{\text{p}}^2/w$ (Odijk, 2008; Dai et al., 2014) Reproduced from Werner et al. (2017).

regime as seen in Figure 2.2. Second, most of the experimental studies done prior to this dissertation primarily focused on measuring the extension as a function of channel size. While extension is the simplest observable metric and quite relevant for genomic applications, measurements of both the extension and its variance (fluctuation about an average value over time) are required to resolve open questions surrounding the different regimes. For example, scaling of the extension as a function of channel size is similar for both extended de Gennes regime and classical de Gennes regime. However, scaling of the variance in extension as a function of channel size should be able to clearly distinguish between the two regimes. While Su et al. (2011) and Thamdrup et al. (2008) did measure the fluctuations, it was done over a channel size range which was too narrow to deduce anything significant about various confinement models.

Chapter 3

Effect of channel shape on thermodynamic properties of DNA in nanoscale confinement

3.1 Introduction

*The thermodynamic properties of DNA confined in a nanochannel have been the subject of substantial debate (Reisner et al., 2012) since experiments by Reisner et al. (2005) almost a decade ago suggested a possible shortcoming in the blob theory of Daoud and de Gennes (1977). Since that time, intense theoretical and simulation work on channel-confined DNA has produced a more nuanced understanding of confined semiflexible chains, but these new predictions now require

*This chapter is based on D. Gupta, J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, and W. Reisner, "Mixed confinement regimes during equilibrium confinement spectroscopy of DNA", J. Chem. Phys. **140**, 214901 (2014).

experimental investigation. Confinement spectroscopy (Persson et al., 2009) is a potentially powerful approach. This method uses a nanochannel with a fixed depth and a variable channel width, thereby probing many different channel sizes in a single device. The comparison of confinement spectroscopy data to theory implicitly assumes that the thermodynamics of a confined chain in a rectangular channel can be mapped to a square channel through the geometric mean of the channel size (Persson et al., 2009; Werner et al., 2012), as suggested by comparing the scaling results from Flory theories for blob-regimes in rectangular and square channels (Odijk, 2008). However, when one uses confinement spectroscopy to test the crossover between different regimes of confinement, the experimental design leads to "mixing" of confinement regimes, where the channel height corresponds to a fixed confinement regime while the channel width moves between different regimes (Figure 3.1). In this chapter, we use a combination of experiment and simulation to show that such regime "mixing" only has a small effect on the extension of the chain. In contrast, regime mixing leads to qualitatively different fluctuations in the extension for chains in square and rectangular channels of equal cross-sectional area.

There is now a relative consensus that four regimes characterize the thermodynamics of DNA confined in a nanochannel as a function of the channel size D, as suggested first by Odijk (Odijk, 2008). In the strongest confinement, known as the Odijk regime (Odijk, 1983), the chain is renormalized into a series of deflection segments. In this regime, hard-core repulsion exists between the chain and the channel walls but the chain stiffness prevents intrachain excluded volume interactions. From a theoretical perspective, the Odijk regime for nanochannel

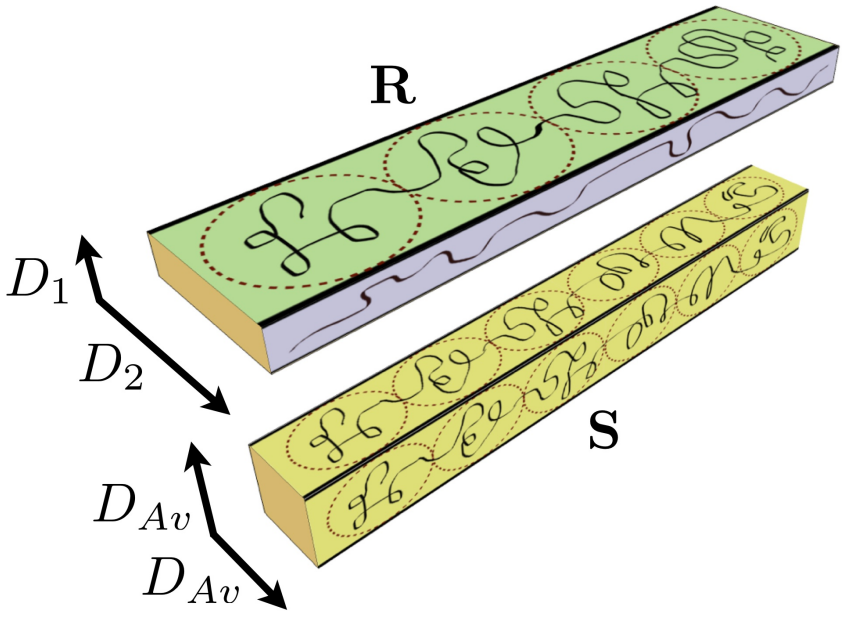

Figure 3.1. Schematic illustration of mixing of confinement regimes in confinement spectroscopy. R corresponds to a rectangular channel with depth $D_1 = 100$ nm and width $D_2 = 400$ nm. S shows an equivalent square channel of dimension $D_{\mathrm{Av}} = 200$ nm. The purple face indicates the projection of the DNA onto the wall of height D_1, which lies in the transition regime. The yellow face indicates the projection of the DNA onto the wall of height D_{Av} which lies in the extended de Gennes regime. The green face indicates the projection of the DNA onto the wall of height D_2 which lies in the de Gennes regime. The black curves represent polymer chains and red circles represent blobs.

confinement is very well established (Yang et al., 2007; Burkhardt et al., 2010). As the channel size increases further, the chain enters the backfolded Odijk regime. Here, the channel width is still small enough to produce deflection segments, but it is large enough to allow for formation of hairpin folds. When hairpins form, the polymer chain experiences excluded-volume interactions. Provided the chain is sufficiently long, numerous simulations of confined wormlike chains indicate that this regime ends when $D \approx 2l_p$, with l_p being the persistence length of the chain (Odijk, 2008; Cifra, 2009; Cifra et al., 2009; Wang et al., 2011; Dai et al., 2014; Muralidhar et al., 2014a). As the channel size increases even further, the chain enters into a so-called "extended de Gennes" regime (Wang et al., 2011) where the chain consists of sausage-like, anisometric blobs (Brochard et al., 2005; Odijk, 2008). At the largest values of D, the weak confinement results in the classical de Gennes regime (Daoud and de Gennes, 1977), where the chain consists of isometric blobs. These two blob regimes were given similar names (Wang et al., 2011) to highlight that the scaling for the extension in the isometric blob de Gennes regime, $X \sim D^{-2/3}$, "extends" into the anisometric blob regime. The latter scaling is based on a Flory exponent $\nu = 3/5$; the more accurate estimate $X \sim D^{(\nu-1)/\nu} \sim D^{-0.7015}$ is in excellent agreement with simulation data for long chains in the extended de Gennes and de Gennes regimes (Wang et al., 2011; Tree et al., 2012; Tree et al., 2013b; Dai and Doyle, 2013; Dai et al., 2014).

It is important to recognize that, while the extension of the confined chain is the simplest observable metric and quite relevant for genomic applications of confined DNA (Jo et al., 2007; Das et al., 2010; Welch et al., 2012; Lam et al., 2012; Hastie et al., 2013; Dorfman et al., 2013; Marie et al., 2013), this single thermodynamic

variable is likely insufficient to resolve many of the open questions surrounding different regimes. Rather, we require measurements of both the average chain extension and its variance, the latter being related to the effective spring constant of the chain. For example, the scaling for the chain extension with channel size cannot distinguish between the extended de Gennes and de Gennes regimes (Wang et al., 2011). However, the fluctuations of the chain should scale differently with channel size in these two regimes (Dai et al., 2013), allowing an unambiguous demarkation of the border between them. In a similar manner, the theory of the transition regime may ultimately be resolved through an analysis of the variance of the chain extension along with its mean extension.

Confinement spectroscopy is a particularly attractive approach to obtain such data, using the precision channel sizes afforded by electron-beam (e-beam) lithography and reactive ion etching while avoiding the high cost of producing many devices. The original confinement spectroscopy experiments (Persson et al., 2009) used a fixed depth D_1 and a continuously tapered width $D_2(x)$, where x is the axial distance down the channel. In principle, one can access a wide range of average channel sizes, $D_{\mathrm{Av}} \equiv \sqrt{D_1 D_2}$, in a single device, with the resolution limited by the step size used to pattern the taper. The result is a massive reduction in the cost and time relative to fabricating individual devices with square channels for each value of D_{Av}. However, there are two reasons to question the ability of confinement spectroscopy data to test the prevailing theories. First, while the mapping to an equivalent square channel via D_{Av} seems reasonable for small aspect ratios, confinement spectroscopy typically requires very large aspect ratios $D_2(x)/D_1$ to access multiple regimes of confinement (based on D_{Av}) at a single etch depth

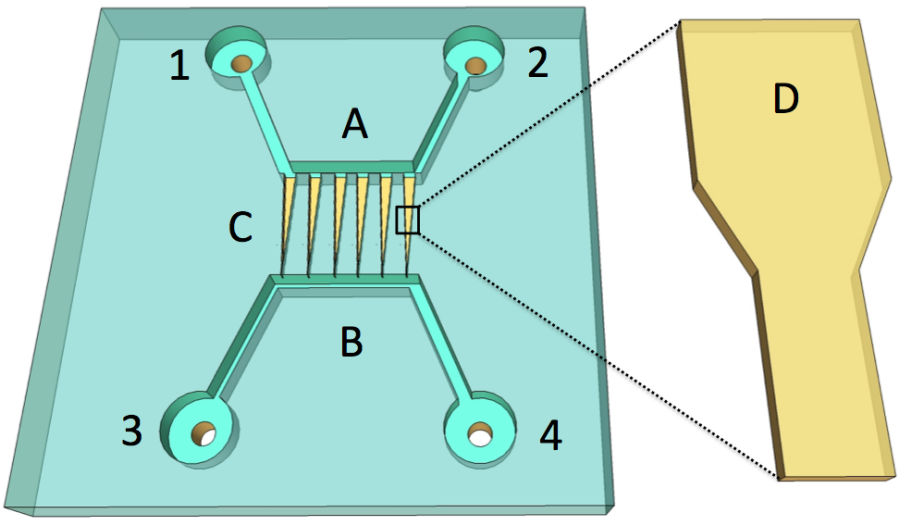

Figure 3.2. Schematic of the device for equilibrium confinement spectroscopy. Region C represents an array of nanochannel cascades of fixed depth $D_1 = 100$ nm and variable width D_2 from 800 nm down to 100 nm in 100 nm decrements with an additional 1000 nm wide region prior to the 800 nm region to facilitate loading of DNA with minimal shear. The nanochannels are connected to loading ports (1-4) through 50 μm wide, 1 μm deep microchannels (A and B). Region D shows two constant width regions which are 45 μm long each and connected by a 5 μm long tapered region.

D_1. Second, a continuous taper leads to a configurational entropy gradient in the channel, which can cause the DNA to gradually drift to the wider side of the taper. Trapping the DNA at a particular value of D_{Av} may require applying an opposing force, either an electric field or a fluid flow, to maintain the DNA center of mass position at a fixed location x. While this opposing force is generally weak (Persson et al., 2009; Werner et al., 2012), its presence implies that confinement spectroscopy in a tapered nanochannel is an inherently non-equilibrium measurement.

In this chapter, we employ the confinement spectroscopy concept (Persson

et al., 2009; Werner et al., 2012) for equilibrium measurements, thereby removing the second caveat and allowing us to address issues related to accessing multiple regimes in a confinement spectroscopy device. In our device, depicted schematically in Figure 3.2, we use a fixed depth $D_1 = 100$ nm and a piecewise constant function $D_2(x)$ that cascades from a maximum channel width of 800 nm to 100 nm in 100 nm increments, with an additional 1000 nm wide region that facilitates loading and alignment between the nanochannels and microchannels during device fabrication. Using the latter region for measurement as well gives us nine values of D_2, where $D_{\mathrm{Av}}/l_p \in [1.40, 4.70]$ traverses the extended de Gennes and transition regimes. Each fixed value of D_2 is maintained for a distance of 45 μm, which is much longer than the extension of λ-DNA in any of these channels. We use a transient pressure-driven flow to overcome the entropic penalty to move a DNA molecule to a new channel width, and then we make our measurements under force-free conditions in the absence of a pressure gradient. Our experiments thus span from the largest aspect ratios (Jo et al., 2007) used previously for channel-like confinement down to the ideal case of a square nanochannel, albeit with a data density that is much lower than the continuous taper case (Persson et al., 2009; Werner et al., 2012).

Since we do not have access to a set of square nanochannels with the same values of D_{Av} as our confinement spectroscopy device, we have adopted a simulation-based approach to assess the role of the aspect ratio. To this end, we use Pruned-Enriched Rosenbluth Method (PERM) simulations (Grassberger, 1997; Prellberg and Krawczyk, 2004) of an off-lattice, discrete wormlike chain model of confined DNA, following our recent work in this area (Tree et al., 2013b; Tree et al., 2013a;

Muralidhar et al., 2014b). We first simulated the mean extension and its variance for the rectangular channels used in our experiments, using the ionic strength of the buffer to compute the effective width (Stigter, 1977) and persistence length (Dobrynin, 2006; Hsieh et al., 2008) of the DNA. The increase of the backbone length due to the intercalating dye is thus the only free parameter. The rather good quantitative agreement between experiment and simulation thus allows us to use simulations to assess the effect of the aspect ratio through further PERM simulations in square channels of width $D = D_{\mathrm{Av}}$.

3.2 Experimental Methods

3.2.1 Device Fabrication

The devices were fabricated on 4 in diameter, 0.5 mm thick fused silica substrates (Hoya). The nanochannel funnels were created via electron beam lithography (Leica VB6) followed by a CF_4:CHF_3 reactive ion etching (RIE) step. Microchannels (50 μm wide, 1 μm deep) for interfacing the nanochannels to sand-blasted loading ports were created via a UV contact lithography followed by RIE. The etched substrates were directed bonded to 0.17 μm fused silica cover-glass via a direct bonding process based on RCA cleans.

The nanochannel depth is 97 nm, as measured by atomic force microscopy (AFM) in the 700 nm and 1000 nm wide region of a duplicate device from the same wafer. The widths of the nanochannels in the cascade were measured by scanning electron microscopy (SEM) in this duplicate device. The SEM images

and cross-section measurements are included in Appendix A.

3.2.2 DNA Preparation

λ-DNA (48.5 kilobase pairs, kbp, New England Biolabs) was stained with YOYO-1 fluorescent dye (Invitrogen) at a concentration of 1 dye molecule for every 10 bp of DNA. After staining at room temperature for an hour, the DNA samples were heated to 50°C for 12 hours to make the distribution of the dye molecules more homogeneous (Carlsson et al., 1995; Nyberg et al., 2013). DNA (10 μg/ml) samples were prepared in 1x (10 mM) Tris-base buffer. Prior to nanochannel experiments, an oxygen scavenger β-mercaptoethanol (BME, Sigma-Aldrich) was added to the solution (6% v/v) to suppress photobleaching.

To calculate the ionic strength of the solution, the system of chemical equilibrium equations was solved iteratively (Tris base: $pK_b = 5.96$, BME: $pK_a = 9.6$) using a program developed by Santiago and coworkers (Santiago, 2010), leading to an ionic strength $I = 7.18$ mM and pH = 7.49. The available on-line program (Santiago, 2010) was modified to calculate ionic strength of a buffer according to a method used by Hsieh et al. (2008); we confirmed the accuracy of the modified program by reproducing data previously appearing in the literature (Hsieh et al., 2008). Experimentally measured pH values were also comparable to the predicted value, further confirming the accuracy of the calculation. We chose to work at this low ionic strength to access relatively large values of the fractional extension of λ-DNA in these channel sizes (Reisner et al., 2007; Jo et al., 2007), which greatly simplifies the image analysis. The lowest value of ionic strength

was limited by the use of an anti-photobleaching agent (BME), which increases the ionic strength through protonic dissociation (Hsieh et al., 2008). We used the same method to compute the ionic strength for data appearing in the literature (Tegenfeldt et al., 2004; Werner et al., 2012), using the reported buffer conditions.

We use λ-DNA for convenience, as it is readily available and less susceptible to shear cleavage than longer DNA (Kovacic et al., 1995). Recently, we have shown that λ-DNA in a high ionic strength buffer (l_p = 53 nm, w = 4.6 nm) is long enough so that its mean span X is essentially in the asymptotic limit for long chains in square nanochannels up to D_{Av} = 400 nm (Muralidhar et al., 2014b). This conclusion extends to lower ionic strengths (Muralidhar et al., 2014b) because the effective width and persistence length have different dependencies on ionic strength (Tree et al., 2013a). A detailed justification for this argument is in the Appendix A.

3.2.3 Nanochannel Experiments

Two parallel microchannels were loaded with DNA solution by capillary action. Each reservoir on the device was then sealed with one of the top ends of a T-junction. The other top end was used for both filling up the reservoir and for introducing fresh DNA solution, which was done every three hours. The bottom end of the T-junction was used to apply pressure by connecting it to a two-way valve having a pressurized air line on one side and atmospheric pressure on the other side. Four T-junctions were assembled into a chuck (Persson and Tegenfeldt, 2010). After filling up the reservoirs with DNA solution and sealing the

other top end of T-junction with the help of O-rings, DNA molecules were drawn inside the nanochannels by applying pressure on the bottom end of two T-junctions directly connected to the reservoirs adjacent to the 1000 nm wide nanochannel region. Once the DNA reached a desired location inside the nanochannels, the pressure in all the four reservoirs was brought down to atmospheric pressure by using all four valves.

The locations of each 45 μm long section of constant width D_2 was recorded from calibration marks and accessed using a Prior II stage controller, which has a precision of 2 μm. To ensure we are not at the interface, we only use data in the central 40 μm of a region of constant width. To make a measurement in a given channel width D_2, a fresh group of molecules were brought from the reservoir to the center of that region of the nanochannel. The DNA thermal fluctuations were recorded with an EM-CCD camera (Photometrics, Cascade II) using a 100x (1.4 N.A.) oil immersion objective on an inverted epifluorescence microscope (Leica DMI 4000B) with a 120 W metal halide light source at 75% of maximum intensity. We used Micro-Manager to record 1500 images at a speed of 10 frames per second with 100 ms exposure time. We chose to observe many molecules at a fixed channel size, rather than tracking single DNA molecules as they move through the cascade, to increase the throughput. We also found that tracking single molecules with such long exposure times led to significant photobleaching after 150 seconds.

3.2.4 Data Processing

Molecule extensions were extracted from the images using a fitting program developed previousl (Reisner et al., 2005). The intensity profiles parallel to the channel axis were assumed to follow a double error function profile, a convolution of an error function and a box function. Using this fit, the molecule extension was estimated at the full width at half max (FWHM) of the fluorescence intensity. This method leads to an uncertainty of roughly 5 pixels to the FWHM, which corresponds to approximately 800 nm. The uncertainty varies from frame to frame and between channel sizes, and we use the latter conservative error estimate for all of our analysis. The mean of the extension for each channel size was calculated by pooling the molecules together under the assumption of homogeneous contour length. This also added uncertainty since there were undoubtably some fragmented molecules, due to fractionation within the stock solution, photocleavage, and shear cleavage. Obviously fractionated molecules were removed from the ensemble, but fluctuations in the chain extension make it difficult to determine if some molecules are slightly too short or simply tend to sample the smaller range of extensions during the measurement. There is also some uncertainty from the channel roughness, but we have not included this in our estimate of the error since the fitting uncertainty is the dominant factor. The fluctuations of the extension were calculated by first computing the variance for each individual molecule and then taking the average of these data.

3.2.5 Relaxation time

There are two possible approaches to assess the fluctuations of DNA confined in the channel: (i) the variance of the extension and (ii) the longest relaxation time of the chain. We prefer the former approach, since it directly accesses information about the effective spring constant of the confined chain while circumventing subtle details surrounding the hydrodynamics of confined DNA (Tree et al., 2012). Indeed, the hydrodynamics of confined DNA in rectangular channels remains unexplored, and the coupling between the chain configurations and hydrodynamic interactions in rectangular channels should be even more complicated than the case for a square channel. Nevertheless, it is still essential that we obtain an estimate of the relaxation time of the chain to ensure that the data in our ensemble of configurations is uncorrelated. While we could use correlated data and then correct for the correlation error (Chodera et al., 2007), the uncorrelated approach is simpler and does not require an accurate measurement of the relaxation time to estimate the additional error in the data analysis due to these correlations.

To compute the relaxation (correlation) time for the extension data, we first computed the autocorrelation function for the instantaneous chain extension,

$$G(\tau) = \frac{\langle \delta X(t)\, \delta X(t+\tau) \rangle_t}{\langle (\delta X(t))^2 \rangle_t} \tag{3.1}$$

using all of the data for a particular DNA molecule. In the latter, $\delta X = X - \langle X \rangle_t$ where $\langle \cdots \rangle_t$ denotes a time average. The longest relaxation time, τ_R, was obtained by fitting the short-time decay of the autocorrelation function with an exponential

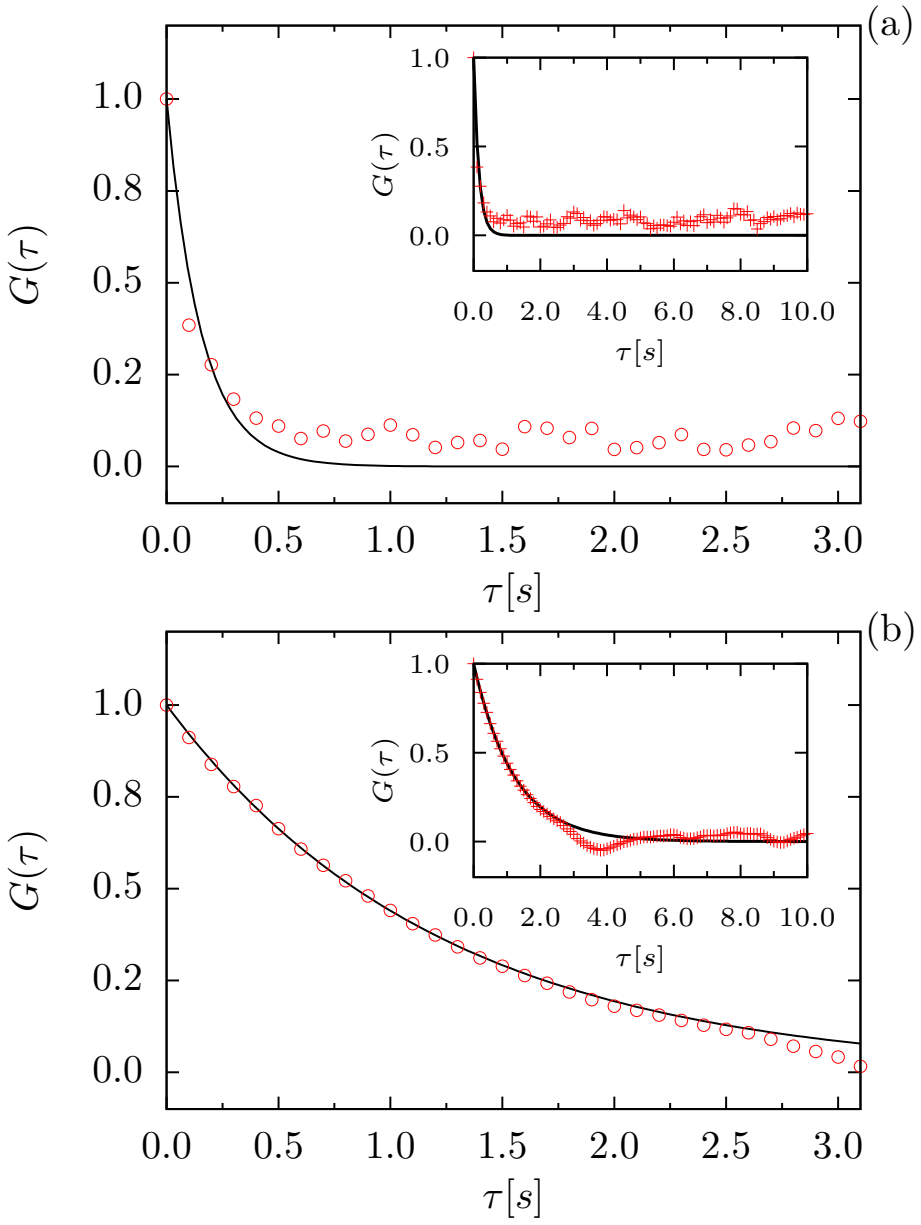

Figure 3.3. (a) Example of the autocorrelation function $G(\tau)$ versus the time lag τ for one λ-DNA molecule in the 100 nm wide channel. (b) Same data for another λ-DNA molecule in the 1000 nm wide channel.

function, where the τ_R is the time constant for the exponential. The fit was constrained such that $G(0) = 1$.

Two examples of the autocorrelation are plotted in Figure 3.3(a) ($D_2 = 100$ nm) and Figure 3.3(b) ($D_2 = 1000$ nm). It is clear that the frame rate used in these experiments, which is dictated by the intensity of the illumination, the field of view of the camera, and the camera sensitivity, is much too slow to make an accurate measurement of the short relaxation time in the 100 nm channel. We chose to use a wide field of view to allow us to image multiple molecules at the same time, which is a compromise between camera frame rate and experimental expediency. As the channel size increases, the data acquisition rate appears to lead to a reasonably good estimate of the relaxation time. The insets of these figures show the autocorrelation time over first 10 s of the fitted data.

To ensure that our ensembles of extension measurements consist of uncorrelated data, we used frames spaced by ten times the maximum relaxation time of 1 s. This corresponds roughly to the largest time lags in the insets of Figs. 3.3(a) and (b), which are clearly decorrelated even for the largest channel. With this conservative estimate for the decorrelation time, each DNA molecule contributes 14 extension measurements. The uncorrelated extension data obtained for all the molecules were then pooled to calculate the mean extension for each channel size.

3.3 Simulation Methods

3.3.1 Discrete Wormlike Chain Model

We used an off-lattice, touching-bead, discrete wormlike chain (DWLC) model to perform our PERM simulations (Tree et al., 2013a; Wang and Gao, 2005). The model consists of a discretized chain of contour length L with $N_\mathrm{b}+1$ beads connected by rigid rods of length l_b. The bond angles between the rods are constrained using the bending potential

$$\beta U_\mathrm{b} = \kappa \sum_{j=1}^{N_\mathrm{b}-1} \left(1-\cos\phi_j\right), \tag{3.2}$$

where ϕ_j is the bond angle at the $j+1^{\mathrm{th}}$ bead and κ is the bending constant. Here $\beta = (k_\mathrm{B}T)^{-1}$ is the inverse of the thermal energy. The bending constant for a desired persistence length l_p was obtained by solving the non-linear equation (Schellman, 1974)

$$\frac{2l_\mathrm{p}}{l_\mathrm{b}} = \frac{\kappa - 1 + \kappa\coth\kappa}{\kappa + 1 - \kappa\coth\kappa}. \tag{3.3}$$

Excluded volume interactions in the model are accounted for by the use of a hard-core repulsive potential given by

$$\beta U_\mathrm{EV} = \begin{cases} \infty & \text{if } r_{ij} \le w, \\ 0 & \text{if } r_{ij} > w, \end{cases} \tag{3.4}$$

where r_{ij} is the distance between beads with indices i and j. The minimum distance between the walls and the center of the beads is kept at $w/2$, thereby accounting for the repulsion between the chain and the walls. We used $l_{\mathrm{b}} = w$, as this value of l_{b} yields sufficient coarse-graining without compromising the physical properties of the model (Tree et al., 2013a; Muralidhar et al., 2014b).

3.3.2 Parameter Estimation

In order to map our simulation results with the DWLC model to experimental results for DNA confined in channels, choosing appropriate parameters for l_{p} and w at a given ionic strength is of utmost importance. The intercalating dyes in these experiments introduce an additional unknown parameter, the contour length in the presence of the dye. Although many studies have investigated the physical properties of DNA in the presence of these dyes (Nyberg et al., 2013; Günther et al., 2010; Bennink et al., 1999), there exists a lack of consensus on how l_{p} and w vary as a function of dye concentration. Therefore, we chose the simplest approach and estimated the parameters, l_{p} and w using established theories (Hsieh et al., 2008; Dobrynin, 2006; Stigter, 1977) for undyed DNA. The use of such parameters for l_{p} and w in the past led to simulation results which agree with experimental data for unconfined, undyed DNA, although there exist systematic errors in predicting properties of dyed DNA (Tree et al., 2013a). This disagreement is conceivably due to the modified contour length, which we use here as a fitting parameter to match our experiments.

The persistence length of DNA is often estimated using either the predic-

tions from Odijk (1977) and Skolnick and Fixman (1977) (OSF) or Dobrynin (Dobrynin, 2006). Although both predictions are in agreement for high ionic strengths ($I > 100$ mM), the persistence length values from the two theories are distinctly different for lower ionic strengths (Hsieh et al., 2008). The disparity is due to the difference in the predictions for the electrostatic contribution to the persistence length. However, Dobrynin's theory conforms with experimental observations that the electrostatic persistence length $l_{\mathrm{p}}^{\mathrm{el}} \sim k^{-1}$, where k is the Debye length (Hsieh et al., 2008). Therefore, we made use of the expression for DNA given in Eq. (7) of Hsieh et al. (2008),

$$l_{\mathrm{p}} = 46.1 + \frac{1.9195}{\sqrt{I}}\,\mathrm{nm}, \tag{3.5}$$

which is based on Dobrynin's theory (Dobrynin, 2006). Equation (3.5) yields a value of $l_{\mathrm{p}} = 68.7$ nm for the ionic strength corresponding to our experiments.

To model the repulsive electrostatic interactions between distal segments of the chain, Stigter (1977) obtained an expression for an effective hard core width of DNA by relating the second-virial coefficient of hard cylinders to that arising from solving the Poisson-Boltzmann equation for DNA. We employed Stigter's theory (Stigter, 1977; Hsieh et al., 2008) to obtain the width $w = 18.7$ nm of the DNA molecule at our ionic strength.

3.3.3 Pruned-Enriched Rosenbluth Method

We used the DWLC model with the parameters thus obtained to carry out Pruned-Enriched Rosenbluth Method (PERM) simulations (Grassberger, 1997; Prellberg

and Krawczyk, 2004; Tree et al., 2013a; Muralidhar et al., 2014b). We performed chain growth starting with one bead chosen at a random position inside the channel, to a maximum length of $N_b + 1$ beads. Our simulations consisted of 400 000 such tours (Tree et al., 2013a; Muralidhar et al., 2014b) for each channel size. All averages of metrics such as the extension and its variance were computed on the fly during run-time, which enables growth of long chains, circumventing the clock time required for read-write of large amounts of data. For more details on the implementation of our simulations, such as weighting of configurations and strategies for pruning and enrichment, we direct the reader to previous publications from our group (Muralidhar et al., 2014b; Tree et al., 2013a).

In all of the simulations, the axis of the channel passing through its center was set as the x-axis of our coordinate system. Accordingly, the extension of DNA inside nanochannels is the mean span of the chain parallel to the channel axis

$$X \equiv \langle \max(x_j) - \min(x_j) \rangle, \text{ where } (j \in [1, N_b + 1]), \tag{3.6}$$

where $\langle \cdots \rangle$ stands for the weighted average over all the tours.(Tree et al., 2013a; Muralidhar et al., 2014b) Similarly, we calculated the variance of the extension,

$$\delta^2 X = \langle X^2 \rangle - \langle X \rangle^2 \tag{3.7}$$

using an online incremental algorithm (West, 1979) to obtain an estimate of the fluctuations of the extension in one pass.

3.4 Results

Figure 3.4 plots both the experimental and simulation data for the average chain extension X as a function of the channel size. The figure illustrates our main points: (i) the experimental data and simulation data are in quantitative agreement to within the experimental uncertainty using a dyed contour length of 18.6 μm, corresponding to a 13% increase compared to naked DNA; and (ii) the confinement spectroscopy approach introduces only a small systematic error in X as the channel aspect ratio increases up to a relatively large value of D_2/D_1 = 10. We have also provided the data in Figure 3.4 in tabulated form in Appendix A for clarity.

To compare with previous experiments, we have computed the apparent power law exponents for the rectangular channel data over the full range of channel sizes. From a theory perspective, the use of apparent power laws is fraught with peril, as the values of D_{Av} used here should span from the transition regime, $X \sim D_{\mathrm{Av}}^{-1}$, into the extended de Gennes regime, $X \sim D_{\mathrm{Av}}^{-0.7}$. Indeed, recognizing the existence of two power laws over the relatively narrow range of channel sizes used in DNA experiments was critical to resolving (Wang et al., 2011) the discrepancy between the early experiments (Reisner et al., 2005) and blob theory (Daoud and de Gennes, 1977). Nevertheless, apparent power law fits are a useful *empirical* approach to compare different experimental data sets. Our data produce a fit $X \sim D_{\mathrm{Av}}^{-0.86\pm0.08}$ for the experiments and $X \sim D_{\mathrm{Av}}^{-0.83\pm0.02}$ for the simulations. As we might expect, the experimental data are scattered about the trendline with $R^2 = 0.97$, whereas the simulation data are much less noisy ($R^2 = 0.99$). The apparent

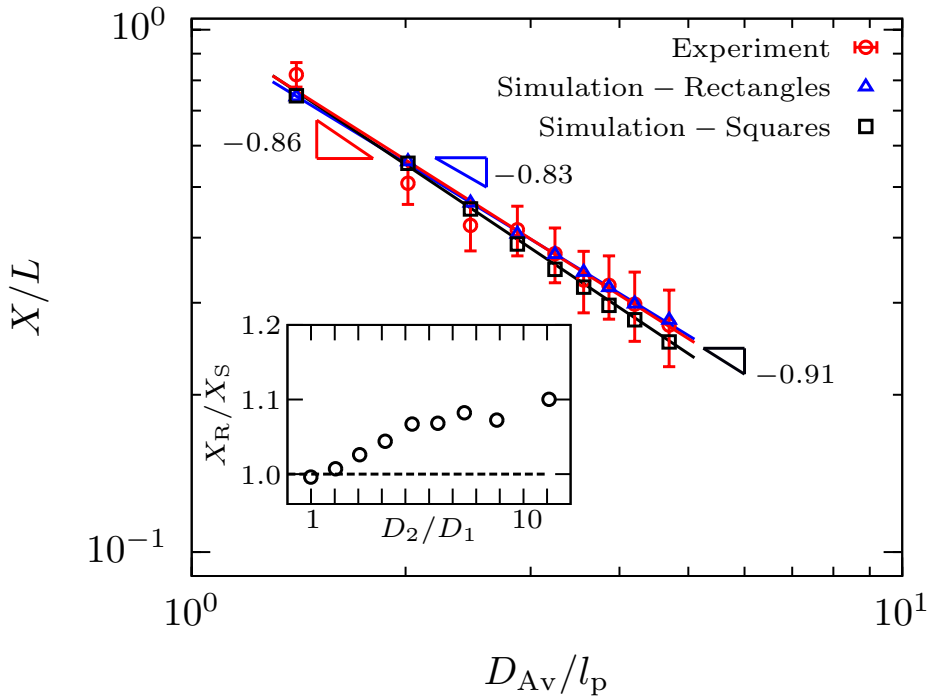

Figure 3.4. Molecule extension in channels of depth 100 nm and widths 100 to 1000 nm. The mean extension is scaled to a fractional extension using a dyed DNA contour length of 18.6 μm. Black squares and trendline were calculated using the DWLC model for squares; blue triangles and trendline were calculated using the DWLC model for rectangles; red circles and trendline are experimental data. Experimental data points are the mean of 196 measurements made from 14 molecules. Error bars for experimental data represent the standard error of the mean plus the measurement uncertainty in §3.2.4; mean extension simulation error bars are within the size of the markers. Slopes shown were calculated using a linear regression of the means, with the exponent for the experiments being -0.86 ± 0.08, the rectangular simulations -0.83 ± 0.02 and the square simulations -0.91 ± 0.03. Extension is shown on a log-log plot. For clarity, these data are also provided in tabulated form in Supporting Material A. The inset shows the ratio of the simulated extension in the rectangular channel, X_R, relative to the square channel, X_S, as a function of the aspect ratio, D_2/D_1.

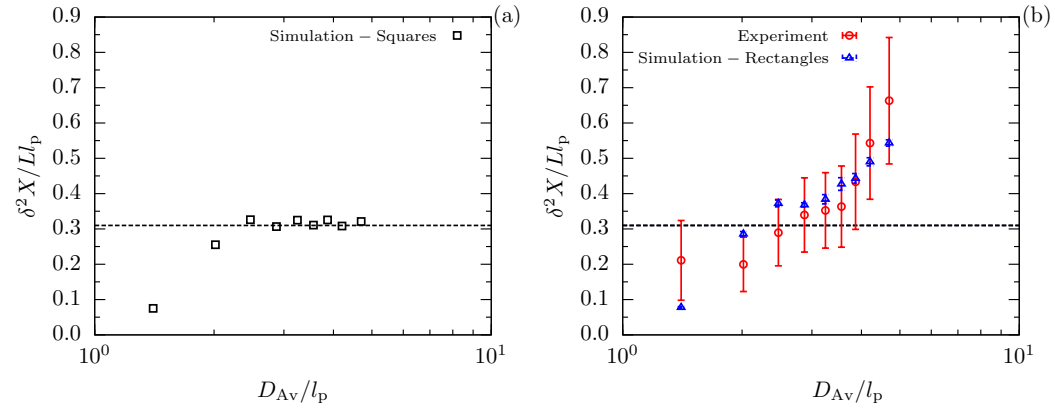

Figure 3.5. Molecule extension fluctuations in channels of depth 100 nm and widths 100 to 1000 nm. Mean extension variance has been scaled to a contour length of $L = 18.6$ μm. (a) Black squares were calculated using the DWLC model for square channels. Error bars are smaller than the symbol size. (b) Blue triangles were calculated using the DWLC model for rectangular channels; red circles are experimental data. Experimental data points are the mean of 14 molecules. Error bars represent the standard error of the mean for rectangle simulations plus the measurement uncertainty in §3.2.4. Both panels are semi-log plots. The dashed line in both panels represents the approximate plateau of the variance for simulated square channels in panel (a).

exponents here are quite similar to results reported previously. For example, the experiments by Reisner et al. (2005) in relatively low aspect ratio channels were fit by $X \sim D_{\mathrm{Av}}^{-0.85}$. Previous confinement spectroscopy experiments in tapered nanochannels also produced $X \sim D_{\mathrm{Av}}^{-0.85}$ for λ-DNA (Persson et al., 2009; Werner et al., 2012), as well as $X \sim D_{\mathrm{Av}}^{-0.83}$ and $X \sim D_{\mathrm{Av}}^{-0.78}$ for T4 DNA (169 kbp) (Persson et al., 2009).

Figure 3.5(a) shows the simulated variance in the chain extension, $\delta^2 X$, as a function of the channel size for the square nanochannels. The shape of this curve is essentially the same as obtained previously (Tree et al., 2013c) in a model of experiments in high ionic strength buffers (Reisner et al., 2005). The variance drops precipitously as the channel size decreases below $D_{\mathrm{Av}} \approx 2l_p$. For the larger

channels, the insensitivity of $\delta^2 X$ to channel size, $\delta^2 X \sim D_{\mathrm{Av}}^0$, is indicative of the scaling predicted for the extended de Gennes regime (Wang et al., 2011) and consistent with previous simulations for these channel sizes (Tree et al., 2013c; Dai et al., 2013). Indeed, the observation of a scaling $X \sim D_{\mathrm{Av}}^{-0.7}$ and $\delta^2 X \sim D_{\mathrm{Av}}^0$ over the same range of channel sizes is compelling evidence of an extended de Gennes regime (Dai et al., 2013).

3.5 Discussion

3.5.1 Regime mixing during confinement spectroscopy

The most interesting part of our analysis is the role of the aspect ratio. It is apparent in Figure 3.4 that there is a gradual deviation between the simulation data for the average extension in rectangular nanochannels and the "equivalent" square nanochannels as the aspect ratio increases. However, this deviation is not especially large and somewhat exaggerated by the logarithmic scale used for the power law fits. The inset of Figure 3.4 shows how the ratio of the extension in the rectangular channel deviates from its square counterpart as the aspect ratio increases. While there is some error introduced by the aspect ratio, this error is an acceptable compromise with respect to the experimental expediency of confinement spectroscopy. Moreover, the congruence between the extension data we have obtained here, previous data for relatively small aspect ratios (Reisner et al., 2005), and previous data for tapered nanochannels (Persson et al., 2009; Werner et al., 2012) indicates that the non-equilibrium nature of the tapered nanochannels is proba-

bly not a critical factor for measurements of the average extension. However, one should be cautious when interpreting the significance of an apparent scaling exponent extracted from confinement spectroscopy; in the present case, the square channels were best fit by $X \sim D_{\mathrm{Av}}^{-0.91} \pm 0.03$ ($R^2 = 0.99$), which is substantially different than the exponent -0.83 ± 0.02 obtained for simulations in the rectangular channels.

While the average extension, X, is not strongly affected by the channel aspect ratio, this is not the case for its variance, $\delta^2 X$. Scaling theories can provide insights into this increase in variance in rectangular channels with increasing channel width. Odijk (2008) suggested the square channel size D_{**} for the cross-over between the extended de Gennes and the de Gennes regimes is of the order of l_p^2/w. Assuming a prefactor of unity, this would correspond to a value of $D_{**} \approx 250$ nm for our ionic strength. We also know that the crossover from the transition regime to the extended de Gennes regime should be at $D_* \approx 2l_p = 135$ nm. Recall that all but the two smallest square channels (Supplementary Material A) used in Figure 3.5 should be in the extended de Gennes regime, where $\delta^2 X \sim D_{\mathrm{Av}}^0$ (Dai et al., 2013). Our data in Figure 3.5(a) show excellent agreement with this prediction.

However, recall that for the rectangular channels used in our experiments, the channel width varies between $D_2 = 100$ nm to 1000 nm, while the depth is kept constant at $D_1 = 100$ nm. This would mean that while the depth of the channels severely constrains bending of the molecule, as we would expect in the transition regime, the width of most of the channels used here ($D_2 > 300$ nm) impose constraints such as those in the de Gennes regime (Figure 3.1). In fact,

a power law fit to our simulation data for rectangular channels, included in the Appendix A, shows that $\delta^2 X \sim D_2^{0.32\pm0.12}$ which is strikingly close to the 1/3 exponent that one would expect in the de Gennes regime.

This mixing of regimes leads to a monotonic increase in the variance for the rectangular channels. In high aspect ratio channels, a fluctuation leading to an increase in the chain extension can be resolved without an increase in excluded volume interactions by swelling of the chain in the direction perpendicular to the channel axis. For these 100 nm deep channels, there is minimal bending penalty when avoiding excluded volume interactions in the wider direction. As a result, the chain can experience large fluctuations in X without incurring a significant excluded volume penalty. In contrast, for an isotropic channel, similar magnitude density fluctuations in the axial direction must lead to substantial excluded volume interactions, which are energetically unfavorable. Moreover, the excluded volume interactions could only be resolved by bending the chain, which is more difficult when both directions are small versus the case where one direction is small and the other is large. The excluded volume interactions, coupled to the bending energy, thereby suppress large fluctuations in X in the smaller aspect ratios.

3.5.2 Comparing simulation and experiment

One of the most challenging aspects of simulating DNA thermodynamics in nanochannels is determining the persistence length, effective width and contour length of dyed DNA. In our analysis, we treat the contour length of the DNA backbone as the sole free parameter, with the persistence length (Dobrynin, 2006) and effective

width (Stigter, 1977) obtained theoretically using the buffer ionic strength. This one degree of freedom allows us to shift the ordinate value of all of the simulation data by a fixed amount, which requires that the overall trend in the simulation data match that of the experiments. Previous studies of DNA extension under dynamic conditions, such as in flow (Perkins et al., 1995) or electrokinetic stretching around solid obstacles (Bakajin et al., 1998; Randall and Doyle, 2005), suggest that the backbone length should increase by 25% to 30% at full dye loading, corresponding to an increase from the naked contour length of 16.5 μm, based on a rise of 0.34 nm per base pair, to a stained contour length around 22 μm (Tegenfeldt et al., 2004). At the dye to base pair ratio used here of 1:10, others (Günther et al., 2010; Persson et al., 2009) suggested that the increased contour length due to dye intercalation is approximately 19.8 μm. Alternatively, if we assume that the increase in the backbone length is linear in dye loading up to the maximum loading, then the 40% loading we used here should lead to a dyed contour length of $L = 18.3$ μm (Reisner et al., 2005).

Arguably the best approach to determine the increased extension due to intercalation is to work with concatemers of λ-DNA, on the assumption that the dye increases the contour length uniformly along the chain. While we do not have such data for our experiments, Tegenfeldt et al. (2004) previously reported data for the extension of λ-DNA concatemers $n = 1, 2, \ldots, 8$ in 100 nm $\times$ 200 nm channels. In the latter reference, the authors suggested that their dye loading should be close to full intercalation, and assumed a contour length of 22 μm per concatamer. By using PERM simulation results, we found excellent agreement between the simulations and these experiments (Tegenfeldt et al., 2004), using a dyed contour

length of L = 20.1 μm per λ-DNA concatemer. The alternate choice of $L = 22$ μm leads to systematic deviations between the simulations and experiments. A detailed comparison is included in the Appendix A.

Nyberg et al. (2013) provide two possible explanations for this smaller than expected extension due to intercalation. First, their findings suggest that the amount of DNA extension as a function of YOYO loading depends on the ionic strength of the buffer solution. Intriguingly, DNA extension at lower ionic strengths shows a stronger than linear dependence with YOYO concentration, while the dependence is weaker than linear at higher ionic strengths. This might explain why the modified contour length in the data from Tegenfeldt et al. (2004) is lower than the expected value of $L = 22$ μm, as the ionic strength in their work is high. Second, Nyberg et al. (2013) have shown that the heating protocol that makes the dye ratios homogeneous is more effective at low dye concentrations than at high dye concentrations. Considering that Tegenfeldt et al. (2004) worked with dye ratios close to maximum intercalation, we could expect more heterogeneity leading an average contour length smaller than the maximum value at complete intercalation.

Given these competing arguments, we chose to run a series of simulations for contour lengths around these estimates and found that a value of $L = 18.6$ μm most closely agrees with our experimental data in Figure 3.4. The difference between our result and other estimates (Reisner et al., 2005; Günther et al., 2010; Persson et al., 2009) is small if we consider the error in measuring the chain extension from the double-error function fit, especially at the smaller values of the extension.

Overall, the quantitative agreement between the simulations and experiments

is excellent considering that there is a single degree of freedom, namely the increased extension due to intercalation. We use this single fitting parameter to shift the ordinate value for the extension data. Thus, the experimental data and simulation results for the average extension need to have the same functional form to produce quantitative agreement. While the apparent exponents are slightly different, the overall shapes of the curves in Figure 3.4 are similar enough to provide satisfactory agreement. For the variance data, the two data sets are clearly in qualitative agreement in Figure 3.5(b), and the quantitative agreement is very satisfying when we consider that there are no adjustable parameters remaining to fit the data; we have used the one degree of freedom to set the value of L in Figure 3.4. Note that the simulations and experiments only agree for rectangular channels; the data for the variance in square channels in Figure 3.5(a) is qualitatively different than for rectangular channels.

3.6 Conclusions

In this chapter, we have used an equilibrium confinement spectroscopy approach to assess the role of the aspect ratio on the mean extension and variance of DNA confined to a nanochannel. The average extension is only weakly affected by the aspect ratio, even out to aspect ratios of ten, making confinement spectroscopy a powerful tool for probing DNA extension by confinement. Moreover, the non-equilibrium nature of a tapered nanochannel is a small perturbation for measurement of the average extension.

Interestingly, our analysis also showed that confinement spectroscopy, even at

equilibrium, leads to qualitative differences in the fluctuations of the chain extension as a function of aspect ratio. Indeed, for our highest aspect ratio channels, the "equivalent" square channel should be in the extended de Gennes regime and have scaling $\delta X^2 \sim D_{\mathrm{Av}}^0$ (Dai et al., 2013). Both our experiments and simulations clearly indicate the variance in the rectangular nanochannels continues to increase with D_{Av}. As a result, confinement spectroscopy is probably not the best tool for probing the different theories for channel-confined DNA; such experiments would require crossing between different regimes of confinement in the same device, and thus experience the "regime mixing" effects we observed here. These theories could be modified to account for aspect ratio (Odijk, 2008) and then compared to equilibrium confinement spectroscopy data. However, this does not seem to be an optimal approach since it adds another degree of freedom (aspect ratio) onto an already contentious area of research. It seems more promising to pursue ionic strength as the independent thermodynamic variable (Reisner et al., 2007; Zhang et al., 2008; Jo et al., 2007), rather than channel size, with a modest number of approximately square channel of different width to provide overlapping data. It is also possible to create a single device with many square channels, for example by using a focused ion beam (Menard and Ramsey, 2013).

These findings have important consequences for use of nanochannels for high-throughout DNA mapping. All other considerations being equal, when comparing devices that yield the same DNA extension, the device with the lower fluctuations is to be preferred as fluctuations may adversely impact barcode alignment. This is one factor that might make aspect ratio unity nanochannels preferable to slit-like nanochannels. Naturally, there are a number of other factors involved in choice of

nanochannel design, including one's available fabrication resources/process cost, the confinement free energy and the hydrodynamic impedance.

While our focus here is on the role of the aspect ratio on the thermodynamic properties of confined DNA, we should also point out the very good quantitative agreement between experiment and simulation for the rectangular nanochannels. The only inputs to the model are the ionic strength of the buffer, which sets the chain persistence length and effective width via established theories (Dobrynin, 2006; Stigter, 1977), and the dyed contour length of the chain, which we determined from a best fit to the experimental data and obtained a result that is well within the range of other independent experiments. Importantly, there are no adjustable parameters to fit our experimental data for the variance. The results thus obtained lend confidence to simulation predictions and theories of confined DNA based on the wormlike chain model. We are thus optimistic that experiments that carefully control for the role of the channel aspect ratio will be able to resolve many of the outstanding questions surrounding the theory of channel confined DNA.

Chapter 4

Experimental evidence of weak excluded volume effects for nanochannel confined DNA

4.1 Introduction

*Determining the equilibrium conformation of nanochannel-confined DNA has attracted substantial attention as a fundamental problem in polymer physics (Reisner et al., 2012) connected to an emerging technology for genomics (Persson and Tegenfeldt, 2010; Lam et al., 2012; Welch et al., 2012; Hastie et al., 2013; Cao et al., 2014; Nilsson et al., 2014). Initially, it was proposed (Reisner et al., 2005) that the de Gennes blob theory for weak confinement (Daoud and de Gennes, 1977) and

*This chapter is based on D. Gupta, J. J. Miller, A. Muralidhar, S. Mahshid, W. Reisner, and K. D. Dorfman, "Experimental evidence of weak excluded volume effects for nanochannel confined DNA", ACS Macro Lett. **4**, 759 (2015).

Odijk's deflection segment theory for strong confinement (Odijk, 1983) provided a complete description of the thermodynamics of a semiflexible chain of persistence length l_p confined in a channel of size D. Intense simulation and theoretical efforts over the past few years (Odijk, 2008; Wang et al., 2011; Dai et al., 2013; Dai et al., 2014; Werner and Mehlig, 2014; Werner and Mehlig, 2015) suggest that excluded volume in confinement plays a more subtle role than previously thought. In particular, in addition to the classic de Gennes blob regime, an "extended de Gennes" regime has been postulated, holding for channel sizes $l_\mathrm{p} \lesssim D \lesssim l_\mathrm{p}^2/w$, where w is the effective width of the polymer backbone. While both the theory of nanochannel-confined DNA (Brochard et al., 2005; Odijk, 2008; Wang et al., 2011; Dai et al., 2013; Dai et al., 2014; Werner and Mehlig, 2014; Muralidhar et al., 2014a; Werner and Mehlig, 2015) and practical applications of the technology in genomics (Persson and Tegenfeldt, 2010; Lam et al., 2012; Welch et al., 2012; Hastie et al., 2013; Cao et al., 2014; Nilsson et al., 2014) have advanced rapidly, the experimental evidence in support of the emerging theoretical framework is lacking. In this chapter, we provide evidence in support of the weak excluded volume effects that form the physical basis for the extended de Gennes regime. In doing so, we also show how a detailed simulation combined with confinement spectroscopy can reduce molecular weight dispersity effects that are introduced into an otherwise monodisperse DNA sample due to shear (Kovacic et al., 1995), photocleavage (Akerman and Tuite, 1996; Alizadehheidari et al., 2015), or heterogeneous staining (Nyberg et al., 2013).

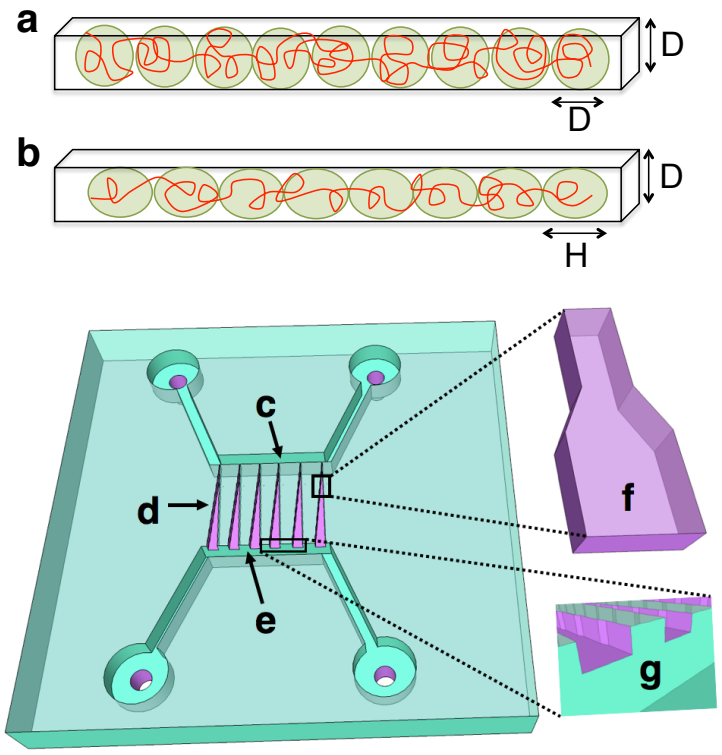

Figure 4.1. (a) In the classic de Gennes regime a confined self-avoiding chain corresponds to a string of isometric blobs. (b) In the extended de Gennes regime a confined and weakly self-avoiding chain corresponds to a string of anisometric blobs. (c-g) Schematic illustration of the confinement spectroscopy device. The device consists of two 1 μm deep and 50 μm wide microchannels (c,e), bridged by an array of nanofunnels (d). A single nanofunnel (f) has ten connected 300 nm deep and 45 μm long nanochannels; the width of the nanochannel varies from 350 nm to 750 nm in steps of 50 nm. The DNA are loaded in the nanochannel with pneumatic pressure across the intersection of the microchannel and the nanofunnel region (g). The detailed operation of this type of device is described in Chapter 3.

4.2 Theory

Classic de Gennes scaling theory suggests that a self-avoiding chain confined in a channel corresponds to a string of isometric blobs of diameter D (see Figure 4.1a) (Reisner et al., 2012). Excluded-volume effects are strong, determining both the polymer physics *within* a blob (each blob is a self-avoiding coil obeying Flory statistics) and *between* blobs (self-exclusion interactions between blobs lead to their linear ordering, swelling the chain along the channel). In contrast, the extended de Gennes regime is characterized by weak excluded volume effects (Werner and Mehlig, 2014; Dai et al., 2014). In this regime, a chain of persistence length l_p and width w, confined in a channel of size D, can be envisioned as a series of *anisometric* blobs (Odijk, 2008) of diameter D and length $H \cong (Dl_\mathrm{p})^{2/3}w^{-1/3}$ (see Figure 4.1b) (Brochard et al., 2005). Excluded volume between blobs lead to swelling of the chain along the channel axis but the excluded volume interactions within a blob are marginal (Dai et al., 2014). The swelling along the channel axis leads to a fractional extension X/L

$$X/L = 1.176(62)(l_\mathrm{p}w/D^2)^{1/3} \tag{4.1}$$

where X is the mean span of a polymer of contour length L (Wang et al., 2011; Dai et al., 2014; Werner and Mehlig, 2014). The prefactor comes from the theory by Werner and Mehlig, which maps the extended de Gennes regime onto a one-dimensional, weakly self-avoiding walk (Werner and Mehlig, 2014). The latter scaling is identical to that in the de Gennes regime (Daoud and de Gennes, 1977)

and hence the moniker "extended" de Gennes (Wang et al., 2011). The weakness of the excluded volume interactions manifests itself in the variance in the fractional extension (Odijk, 2008; Wang et al., 2011; Dai and Doyle, 2013; Dai et al., 2014; Werner and Mehlig, 2014). Since the blobs themselves are at the crossover between real and ideal chains, the variance in the fractional extension is independent of the channel size (Wang et al., 2011; Dai and Doyle, 2013; Dai et al., 2014; Werner and Mehlig, 2014),

$$\delta^2 X / L l_{\mathrm{p}} = 0.264(99) \tag{4.2}$$

where again the prefactor comes from the theory of Werner and Mehlig (Werner and Mehlig, 2014). In contrast, in the classic de Gennes regime, $\delta^2 X \sim D^{1/3}$ (Reisner et al., 2005). Note that Equations 4.1 and 4.2 are expressed in terms of the persistence length, which leads to different numerical values than the equivalent expressions in terms of Kuhn length (Werner and Mehlig, 2014).

4.3 Methods

Demonstrating the existence of weak excluded volume effects through the scaling $\delta^2 X \sim D^0$ in the range $l_{\mathrm{p}} \lesssim D \lesssim l_{\mathrm{p}}^2/w$ is a key step towards establishing the existence of an extended de Gennes regime. The experimental approach seems straightforward: make a series of channels in this size range, introduce DNA into these channels, and measure the variance in the chain extension. Unfortunately, molecular weight dispersity of the DNA represents a critical obstacle to obtaining accurate measurements of the scaling exponents and, even worse, the prefactors to

the scaling laws. In many cases very large DNA molecules (sometimes on the order of megabases in size) are required to reach the long chain limit (Tree et al., 2013a; Mansfield and Douglas, 2013; Dai et al., 2013; Muralidhar et al., 2014b). While genomic DNA samples are monodisperse, such long DNA molecules are prone to shear breakage (Kovacic et al., 1995). Moreover, the DNA can photocleave during measurements (Akerman and Tuite, 1996). Heterogeneous staining (Nyberg et al., 2013) can lead to differences in L due to variations in the increased contour length as a result of intercalation between molecules. In all of these cases, the associated uncertainty in L can lead to large uncertainties in the measurement of $\delta^2 X$, thereby making the tests of the theories challenging.

We have developed an approach that combines equilibrium confinement spectroscopy (Gupta et al., 2014) with Pruned-Enriched Rosenbluth Method (PERM) simulations of a confined, discrete wormlike chain (Tree et al., 2013b) to reduce the effects of molecular weight dispersity. The nanochannel device, illustrated schematically in Figure 4.1c-g, consists of a channel of depth $D_1 = 300$ nm that cascades in discrete steps in width from $D_2 = 350$ nm to $D_2 = 750$ nm. It is thus similar to previous confinement spectroscopy devices with a continuous taper in D_2 (Persson et al., 2009), but the 45 μm long regions of constant width D_2 allow us to make measurements at equilibrium (Gupta et al., 2014). The buffer for our experiments has an ionic strength of 7.18 mM. Accounting for the effect of the anti-photobleaching agent β-mercaptoethanol at 6% v/v, theory (Hsieh et al., 2008) predicts a persistence length of $l_\mathrm{p} = 69$ nm and effective width $w = 19$ nm (Stigter, 1977; Dobrynin, 2006; Gupta et al., 2014), although the accuracy of this persistence length should be viewed in light of recent data suggesting a lower

persistence length of DNA (Brunet et al., 2015).

Since our channels are rectangular, we use the effective channel width (Reisner et al., 2005; Wang et al., 2011; Werner and Mehlig, 2015),

$$D_{\text{eff}} = \sqrt{(D_1 - \delta)(D_2 - \delta)} \tag{4.3}$$

to map the results back to an equivalent square channel of size $D_{\text{eff}} + \delta$. The parameter δ is a wall-DNA depletion length that models the electrostatic interaction of the DNA with the walls (Reisner et al., 2012). We estimate that $\delta \approx w$ (Wang et al., 2011). Our effective channel sizes range from 305 nm to 453 nm, which spans much of the extended de Gennes regime; simulations by Dai et al. (2014) suggest that the lower bound of the extended de Gennes regime for our conditions is 276 nm and the upper bound is 562 nm. Most of our channel sizes are also below the upper limit $D_2 = 550$ nm (assuming a prefactor of unity) where Werner and Mehlig (2015) predict that the channel anisotropy will affect measurements of the variance in chain extension for our value of D_1 (Gupta et al., 2014), and we check the magnitude of the anisotropy effect via simulations. Additional information about the device fabrication and buffer conditions are provided in Appendix B.

The key advantage of the equilibrium spectroscopy approach (Gupta et al., 2014) is that the same molecule can be studied in all the channel sizes at equilibrium, making it a single-molecule study in the true sense (Frykholm et al., 2014). In other words, time-domain measurements for some molecule i can be used to determine the ensemble average from an ergodic hypothesis, and repeating this measurement for different molecules provides information about the experimental un-

certainty, much of which arises from a very conservative estimate of the error in the image analysis (see Appendix B). The approach is identical to typical simulation approaches, where independent replicas (the different molecules) are used to assess the sampling errors in expectation values (the measurements from a single molecule). In our case, we first pneumatically load a T4 GT7-DNA molecule (166 kbp, Nippon Gene) from the microchannel into the $D_2 = 350$ nm region and take $n = 40$ stroboscopic measurements of the span of that molecule with a minimum sampling interval of 5 seconds. Let us denote the j^{th} measurements on molecule i as $X_i^{(j)}$. The time lapse corresponds to a small value of the autocorrelation in the chain extension (see Appendix B), which ensures that the $X_i^{(j)}$ are statistically independent. When we compute the average span for molecule i and its variance,

$$X_i = (1/n)\sum_{j=1}^{n} X_i^{(j)} \tag{4.4}$$

$$\delta^2 X_i = [1/(n-1)]\sum_{j=1}^{n}(X_i^{(j)} - X_i)^2 \tag{4.5}$$

we assume that these values are estimates of the ensemble average, i.e., $X \approx X_i$ and $\delta^2 X \approx \delta^2 X_i$. To obtain the values of X_i and $\delta^2 X_i$ at other channel sizes, we move the same molecule successively through each value of D_2 and repeat the process. At the conclusion of this single-molecule experiment, we return the molecule to the $D_2 = 350$ nm region and image it again to confirm that the extension of the molecule has not become significantly shorter due to photocleavage or shear cleavage. We performed this single-molecule protocol with 40 molecules, of which 29 did not shorten during the long measurement period. The end result is 1160 measurements of the chain extension at each channel size, blocked into 29 replicas for

statistical averaging. The detailed criteria for rejecting fragmented molecules and additional information about the experimental protocol are provided in Appendix B.

We found a broad distribution of extensions X_i when we binned the results in the 1 μm bins shown in Figure 4.2a. To see if this wide distribution was a result of molecular weight dispersity, we took advantage of our ability to systematically move a given molecule between different values of the channel width. To simplify our analysis, we considered how the average of the extension of the w_{bin} molecules within a given bin, $X_{\text{bin}} = (1/w_{\text{bin}}) \sum_{i=1}^{w_{\text{bin}}} X_i$, evolves as a function of D_{eff}. Figure 4.2b shows the results for two representative bins; the data for the other bins are included in Appendix B. Overall, it is clear that when we bin the molecules based on their extension for $D_2 = 350$ nm, plots of X_{bin} as a function of the effective width D_{eff} are parallel curves. Since we know that $X \sim L$ for sufficiently long chains (see Supporting Information for Muralidhar et al. (2014b)) we can conclude that the distribution in Figure 4.2b results from molecular weight dispersity, most likely a combination of shear breakage and non-uniform staining (which leads to non-uniform extension due to intercalation) (Nyberg et al., 2013).

To compute the nominal contour length L_{bin} for those molecules in a given bin of Figure 4.2a, we performed PERM simulations of a discrete wormlike chain in a rectangular channel for each value of D_2. These simulations follow directly from Chapter 3, and a brief description is provided in Appendix B. Importantly, PERM provides data for the simulated extension, X_{sim}, in a given rectangular channel size as a function of simulated contour length, L_{sim}. For each channel, we ran our simulations out to approximately double the contour length of unstained T4 DNA,

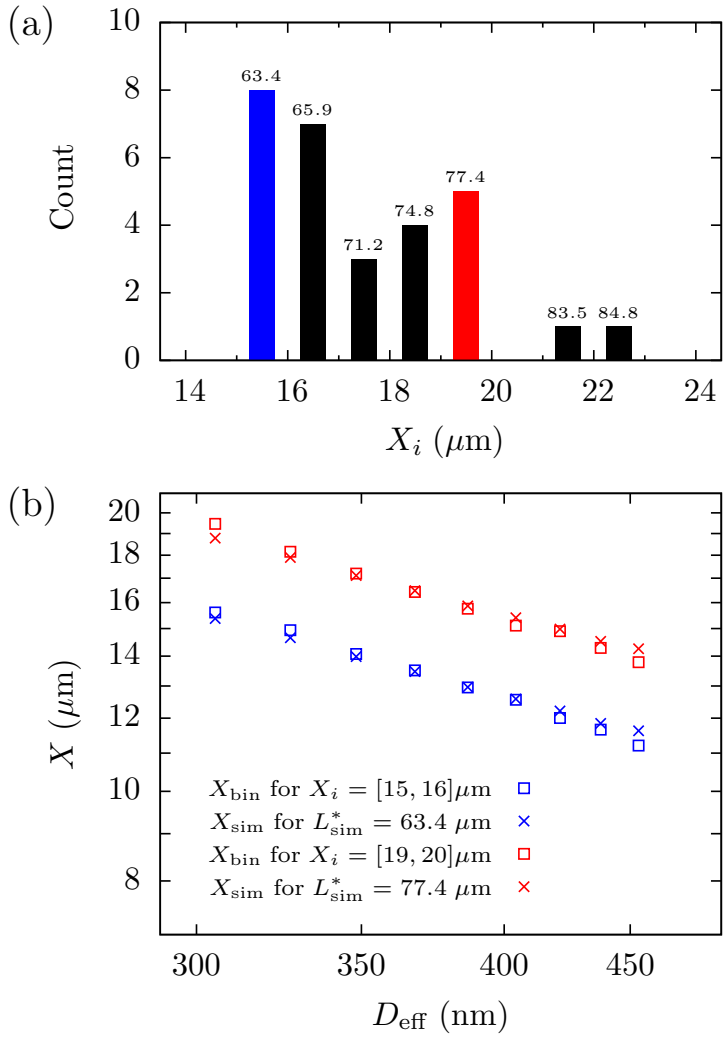

Figure 4.2. (a) Probability distribution of average extension, X_i, for individual T4 DNA molecules from 40 uncorrelated measurements of their span for $D_2 = 350$ nm. The bin size is 1 μm. The number above each bar is the value of the contour length in that bin, $L_{\rm bin}$ (μm), obtained by comparison of the extension to PERM simulations. Linear regression of $L_{\rm bin}$ to X_i gives a correlation coefficient $R^2 = 0.98$. (b) Log-log plot of the average value of the extension, $X_{\rm bin}$, for molecules that reside in two different bins (X_i = [15,16] μm in blue and X_i = [19,20] μm in red) as a function of effective channel size (□ symbols). To determine the corresponding molecular weight, we fit these curves to PERM simulations of the chain extension (× symbols) using the contour length as a free parameter. A plot similar to panel (b) for all of the bins is included in Appendix B.

well beyond any possible effect due to intercalation (Bakajin et al., 1998; Randall and Doyle, 2005). Then, for each bin in Figure 4.2a, we determined the value of L^*_{sim} that minimized the sum of squared error in $X_{\mathrm{sim}} - X_{\mathrm{bin}}$ over all the channel sizes and set $L_{\mathrm{bin}} = L^*_{\mathrm{sim}}$. Figure 4.2b reports the result of this analysis for two bins, and the corresponding analysis for the other bins is included in Appendix B. The data in Figure 4.2b indicate that there is some disagreement between the simulation and experimental data. However, the L_{bin} values are linearly dependent on X_i to within 2% error. There may also be a systematic error in L_{bin}, but this will only affect the prefactor and not the exponent in the scaling law (see Appendix B).

The ultimate result of our binning-and-simulation protocol are estimates for the contour length of molecules within each bin, which are reported in Figure 4.2a. In what follows, we use the value of L_{bin} as the contour length L_i for any molecule within that bin, thereby compensating for the effects of molecular weight dispersity by rescaling each molecule by its L_i value. We are now in a position to determine the validity of Equation 4.2 for the variance in chain extension, which is the key prediction arising from weak excluded volume effects in the extended de Gennes regime.

4.4 Results and Discussion

The independence of the variance in extension from channel size in Figure 4.3 demonstrates that weak excluded volume effects are manifest for DNA in nanochannel confinement. This is the key result of this chapter. We find experimentally

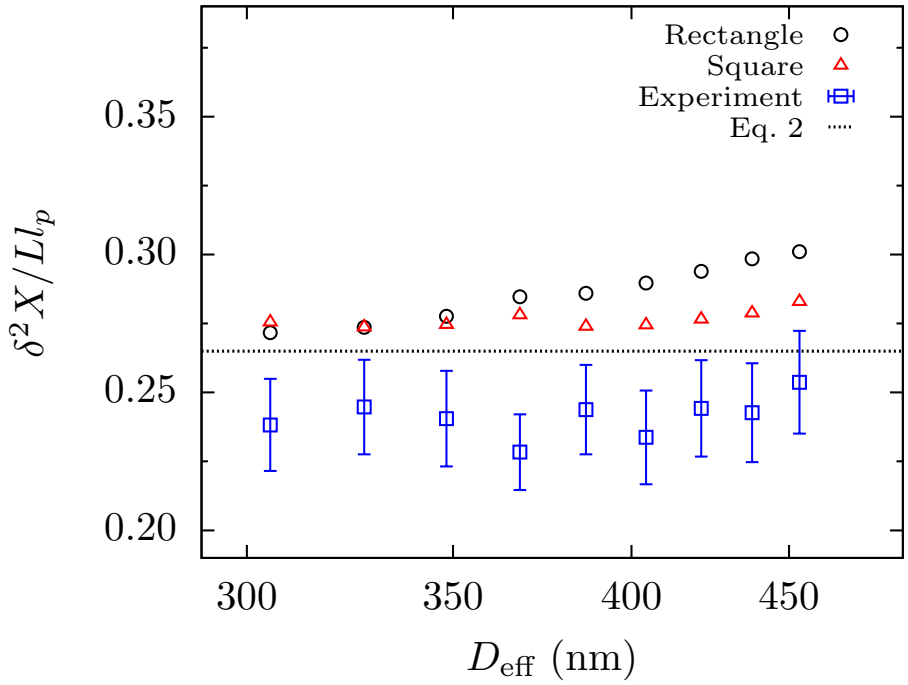

Figure 4.3. Semilog plot of the average variance in extension, $\delta^2 X$ as a function of effective channel size, D_{eff}, for rectangular channels with depth $D_1 = 300$ nm and widths D_2 ranging from 350 nm to 750 nm. The experimental mean extension variance (blue □) are the mean of 29 molecules. Error bars represent the combination of standard error of the mean and propagation of measurement uncertainty (see Appendix B). PERM simulations in the rectangular channels of dimension $D_1 = 300$ nm and $D_2 = 350$ nm to 750 nm for a contour length $L = 70.2$ μm are included for rectangular channels (black 5) and equivalent square channels (red △) of size $D_{\text{eff}} + w$. Error bars for the simulation data are smaller than the symbol size. The black dashed line is Equation 4.2.

that the variance in chain extension has the scaling $\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{0.08\pm0.18}$, where the uncertainty in the exponent is at the 95% confidence level. We also did an alternate analysis, where we obtained the exponent for each molecule and then computed their average. This led to $\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{0.05\pm0.27}$ at the same confidence. Note that this alternate method provides the exponent but no information about the prefactor, since it does not require estimating L or l_p. A hypothesis test at the 5% significance level rejects the $\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{1/3}$ hypothesis in favor of the $\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{0}$ hypothesis by both analysis methods. Our variance data agree to within 8% of the theoretical prefactor predicted by Werner and Mehlig and to within 16% of the PERM data. However, while we believe the agreement for the scaling exponent is robust to our data analysis method, the agreement for the prefactor must be viewed in light of the method to determine L_bin. A discussion of this subtle point is included in Appendix B. Moreover, the error estimates in Figure 4.3 do not include uncertainties in L or l_p. These sources of correlated error affect the prefactor, but do not affect the slope. Thus, it seems reasonable to conclude that the agreement for the prefactor is semi-quantitative.

We also checked for anisotropy effects due to the rectangular shape of the channel (Gupta et al., 2014; Werner and Mehlig, 2015) by performing PERM simulations in square channels with size $D_\mathrm{eff} + w$. Figure 4.3 shows no systematic deviation in the simulated values of $\delta^2 X/Ll_\mathrm{p}$ between square channels and rectangles up to the aspect ratio $D_2/D_1 = 1.7$ of the widest rectangular channel. There does appear to be a small difference for $D_\mathrm{eff} > 370$ nm. However, the maximum difference between the rectangular channels and the equivalent square channel was close to 6%, which is of same order of magnitude as the uncertainty in the

experimental measurement. We thus conclude that any anisotropy effect in these experiments is much smaller than the circa 200% increase that occurs in the 100 nm deep rectangular channels that exhibit regime mixing (Gupta et al., 2014).

While the simulations in square channels produce an exponent similar to the experiments ($\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{0.05\pm0.05}$ at 95% confidence), the simulations for the rectangular channels rejects both classical and extended de Gennes regime hypothesis at the 5% significance level ($\delta^2 X/Ll_\mathrm{p} \sim D_\mathrm{eff}^{0.27\pm0.03}$). The discrepancy between simulation and experiments likely arises from the uncertainty in the estimation of parameters which goes into the model used for PERM simulations. We will revisit this uncertainty later in our discussion of the extension results.

Figure 4.4 summarizes the different results for the average fractional extension of the chain, X/L, as a function of effective channel size, D_eff. The relatively good quantitative agreement between the PERM simulations and experimental data results from using L as a fitting parameter, whereupon the only significant metric for comparison is the power law exponent (Gupta et al., 2014). The experimental data produce a scaling of $X \sim D_\mathrm{eff}^{\alpha}$ with $\alpha = -0.86 \pm 0.02$. Similar to many previous studies of nanochannel confined DNA (Reisner et al., 2005; Persson et al., 2009; Werner et al., 2012), we find that the apparent exponent is more negative than the predictions from weakly self-avoiding random walk theory ($\alpha = -2/3$, Equation 4.1) (Werner and Mehlig, 2014), predictions based on a corrected Flory exponent ($\alpha = -0.7015$) (Wang et al., 2011; Dai et al., 2014), or our simulations in rectangular channels ($\alpha = -0.71 \pm 0.01$).

Disagreement between theory and simulation results in Figure 4.4 likely arises from the finite molecular weight. As shown in Appendix B, PERM simulations

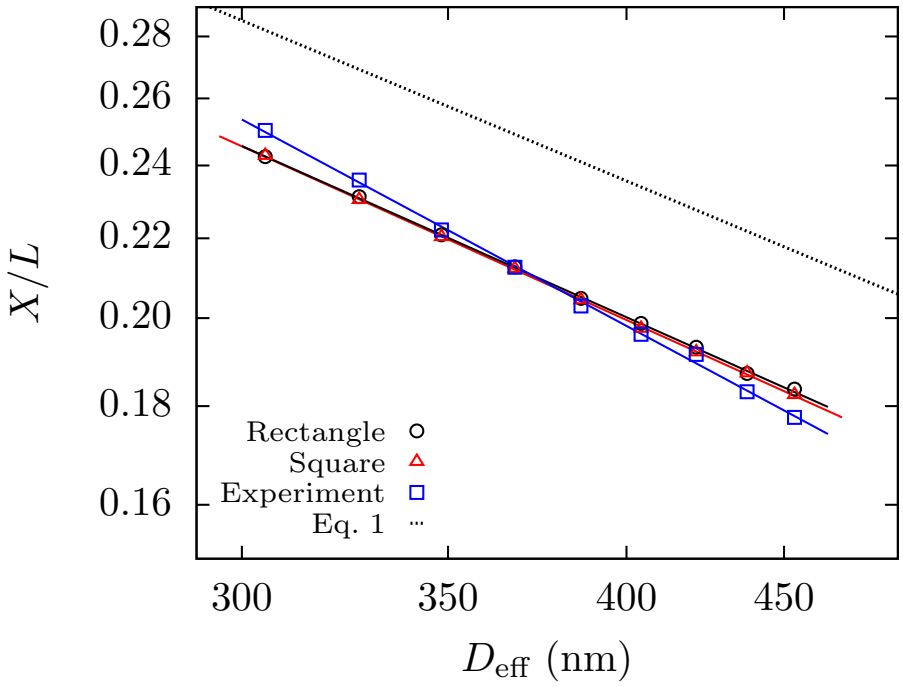

Figure 4.4. Log-log plot of the average fractional extension, X/L, as a function of effective channel size, D_{eff}. The experimental data (blue □) are the average of 29 molecules. Power law fit to the data yields an exponent $\alpha = -0.86 \pm 0.03$. The uncertainty in the exponent is determined from combination of propagation of 95% confidence interval for the exponent of an individual molecule and its standard error of the mean. PERM simulations in the rectangular channels (black 5) and square channels (red △) were calculated for $L = 70.2$ μm. Power law fit to the data yields an exponent $\alpha = -0.71 \pm 0.01$ for the rectangular channel and $\alpha = -0.72 \pm 0.01$ for the square channel. The uncertainties in the exponent for the simulations are determined from 95% confidence intervals. The error bars for the extension are smaller than the symbol size. The black dashed line is Equation4.1 (Werner and Mehlig, 2014). The prefactor for the simulated extension is similar to previous work (Dai and Doyle, 2013; Dai et al., 2014).

for our rectangular channels suggest that the magnitude of the apparent exponent α slowly declines with molecular weight once the chain is large enough to form several blobs, only reaching the asymptotic value for the experimentally impractical scenario of DNA in the megabase range. Disagreement between simulation and experimental data in Figures 4.3 and 4.4 likely arises from uncertainties on the precise value of the persistence length at our salt concentration (Brunet et al., 2015), possible changes in ionic strength during the experiment, and the exact value of the wall depletion length δ. We feel in particular that errors arising from the depletion length might explain the steeper dependence of the experimental extension on channel size compared with simulation. The variance data, independent of channel size, should be affected only by errors on the persistence length, explaining why the variance data agrees better with simulation than the extension. We also see quantitative disagreement between Equation 4.1 and our experiments and simulations, as was the case with previous simulations (Dai et al., 2014).

4.5 Conclusion

In summary, we have provided experimental evidence supporting the theoretical prediction of weak excluded volume in the extended de Gennes regime. Simultaneously, we developed an approach to account for the effect of molecular weight dispersity effects. These experiments represent the first test of the emerging theoretical framework describing channel-confined DNA (Odijk, 2008). The complete experimental test of the phase diagram for channel-confined DNA, with careful control over molecular weight dispersity, will provide a firm foundation for the

emerging genomic technology (Persson and Tegenfeldt, 2010; Lam et al., 2012; Welch et al., 2012; Hastie et al., 2013; Cao et al., 2014; Nilsson et al., 2014).

Chapter 5

Diffusion of DNA in nanochannels

5.1 Introduction

*The equilibrium response of DNA to nanochannel confinement has received considerable attention as a fundamental problem in polymer physics (Reisner et al., 2012; Dai et al., 2016) as well as the basis behind a genome mapping technology (Persson and Tegenfeldt, 2010; Lam et al., 2012; Michaeli and Ebenstein, 2012). Experimentally, the statistical mechanical properties of DNA in nanochannel confinement (such as its average extension and the variance about that average extension) have been investigated in detail by varying the degree of confinement, the flexibility of the chain, the length of DNA molecules and the amount of bound fluorescent dye (Tegenfeldt et al., 2004; Reisner et al., 2005; Reisner et al., 2007; Zhang et al., 2008; Thamdrup et al., 2008; Persson et al., 2009; Utko et al., 2011; Kim et al., 2011; Su et al., 2011; Werner et al., 2012; van Kan et al., 2012; Nyberg

*This chapter is based on D. Gupta and K. D. Dorfman, "Diffusion of DNA in nanochannels", to be submitted (2017).

et al., 2013; Frykholm et al., 2014; Gupta et al., 2014; Alizadehheidari et al., 2015; Iarko et al., 2015; Reinhart et al., 2015; Gupta et al., 2015; Sheats et al., 2015; Reifenberger et al., 2015). However, the dynamic response of DNA in nanochannel confinement still remains relatively unexplored, in particularly with respect to the center-of-mass diffusion coefficient.

The salient open question about DNA dynamics in nanochannel confinement is whether the diffusivity scaling predicted by the blob theory of Brochard and De Gennes (1977) can be applied to practical situations involving DNA, which typically take place in the so-called extended de Gennes regime (Odijk, 2008; Wang et al., 2011; Werner and Mehlig, 2014; Dai et al., 2014). Even though no experimental data for diffusivity in nanochannels are available, there is a substantial body of experimental data on DNA diffusion in nanoslits that contradict the prediction from the classical blob theory (Balducci et al., 2006; Hsieh et al., 2007; Strychalski et al., 2008; Tang et al., 2010; Lin et al., 2011). This discrepancy in the nanoslit geometry was reconciled by accounting for local correlations arising in a semiflexible chain, which introduce a correction to the blob theory (Dai et al., 2013). Recently, Muralidhar and Dorfman (2015) proposed that a similar correction for local chain correlations would also work in nanochannels. Indeed, their mobility data, obtained from simulations of a wormlike chain model in channel sizes corresponding to the extended de Gennes regime, agree well with a modified blob theory (Muralidhar and Dorfman, 2015). In this chapter, we confirm that these computational predictions hold true in experiments. In addition to showing that the classical blob theory fails to describe the experimental data, we observe quantitative agreement between experimental data and simulation data (Muralidhar

and Dorfman, 2015) without the need for any fitting parameters.

5.2 Theory

We are concerned here with the diffusivity of a semiflexible polymer with a persistence length l_p, effective width w and contour length L, confined in a square channel of size D. In the classical de Gennes regime, corresponding to $D \gg l_\mathrm{p}^2/w$ (Odijk, 2008), the chain is comprised of a series of isometric blobs of dimension proportional to D. The chain segments inside a blob comprise a three-dimensional, self-avoiding coil in a good solvent, obeying Flory statistics. The contour length in each blob is then $L_\mathrm{blob} = D^{5/3}(wl_\mathrm{p})^{-1/3}$, where we use standard Flory exponent, $\nu = 3/5$ (Schaefer et al., 1980). The extension is then (Daoud and de Gennes, 1977),

$$X \sim \frac{L}{L_\mathrm{blob}} D = L\left(\frac{wl_\mathrm{p}}{D^2}\right)^{1/3} \tag{5.1}$$

where L/L_blob is the total number of blobs in the confined chain.

One way to determine the diffusion coefficient, D_t is by estimating first the friction factor of the confined chain, ζ_chain, and then using Einstein's relation, $D_\mathrm{t} = k_\mathrm{B}T/\zeta_\mathrm{chain}$, where k_B is the Boltzmann constant and T is the absolute temperature. The blob theory (Brochard and De Gennes, 1977) assumes that hydrodynamic interactions (HI) are screened over length scales greater than D. Consequently, the total friction factor of the chain is the sum of the individual friction factor of the blobs, $\zeta_\mathrm{chain} = (L/L_\mathrm{blob})\zeta_\mathrm{blob}$. The friction factor of the blob can be estimated as $\zeta_\mathrm{blob} \simeq 6\pi\eta D$, where η is the solvent viscosity. The diffusion coefficient is then

(Brochard and De Gennes, 1977),

$$D_{\mathrm{t}} \simeq D_{\mathrm{R}}\left(\frac{D^2}{w l_{\mathrm{p}}}\right)^{1/3} \tag{5.2}$$

where $D_{\mathrm{R}} = k_{\mathrm{B}}T/(6\pi\eta L)$ is the (Rouse) diffusivity of a freely-draining chain. Comparing Equation 5.1 and Equation 5.2 leads to the convenient relation,

$$D_{\mathrm{t}} \simeq D_{\mathrm{R}}\left(\frac{X}{L}\right)^{-1} \tag{5.3}$$

In other words, the confined chain is a non-draining object (Zimm) whose friction is proportional to its size.

An equivalent way to determine the scaling law in Equation 5.2 is by using the pre-averaged Kirkwood approximation (Tree et al., 2012),

$$D_{\mathrm{t}} = \frac{k_{\mathrm{B}}T}{L}\int_0^{D/2} h(r)\Omega(r)dr, \tag{5.4}$$

where $\Omega(r) = 1/(6\pi r\eta)$ is the angle-averaged Oseen tensor in free solution, and $h(r) \equiv 4\pi r^2 l_{\mathrm{p}} g(r)$ is a dimensionless form of the pair correlation function $g(r)$. The limited HI outside the length scale D sets the upper bound in the integral in Equation 5.4. The pair correlation function inside a de Gennes blob is

$$h(r) \sim \left(\frac{r^2}{l_{\mathrm{p}} w}\right)^{1/3} \tag{5.5}$$

Equations 5.4 and 5.5 produce the same scaling relation as Equation 5.2.

For channel sizes $l_{\mathrm{p}} \ll D \ll l_{\mathrm{p}}^2/w$, the conformation of semiflexible chain in

confinement takes shape of a series of anisometric blobs of length $(Dl_{\mathrm{p}})^{2/3}w^{-1/3}$ and diameter D in extended de Gennes regime (Odijk, 2008; Dai et al., 2014). The excluded volume interactions within a blob in this regime are marginal. Note that the extended de Gennes regime is very similar to the marginal condition of semiflexible polymers in semidilute solution proposed by Schaefer et al. (1980). Both the average extension and variance about that average extension in this regime are well established from both experimental (Gupta et al., 2015; Iarko et al., 2015) and theoretical (Dai et al., 2014; Werner and Mehlig, 2014; Wang et al., 2011; Odijk, 2008) perspectives. While the early work suggested that the blob theory of Brochard and de Gennes (Brochard and De Gennes, 1977) for $D \gg l_{\mathrm{p}}^2/w$ would produce the appropriate diffusivity scaling in the extended de Gennes regime (Tree et al., 2012; Reisner et al., 2012), recent simulations of the Kirkwood diffusivity of a wormlike chain model by Muralidhar and Dorfman (2015) show that a modified form of blob theory, inspired by similar work on diffusion in nanoslits (Dai et al., 2013; Dorfman et al., 2014), is necessary. The classical blob theory neglects the effect of local stiffness of the chain while computing the pair correlation function in Equation 5.5. In its modified form, it is assumed that the local chain stiffness results in strong correlation along the backbone of the chain at length scales smaller than l_{p}. Consequently, the pair correlation function is approximately (Dai et al., 2013),

$$h(r) \approx \begin{cases} 2 & \text{if } r < l_{\mathrm{p}}/2 \\ \left(\dfrac{r^2}{l_p w}\right)^{1/3} & \text{if } r \geq l_{\mathrm{p}}/2. \end{cases} \tag{5.6}$$

Using this modified pair correlation function in Equation 5.4 and casting the final result in terms of the chain extension, we obtain (Muralidhar and Dorfman, 2015),

$$D_{\mathrm{t}} \simeq D_{\mathrm{R}} \left[c_1 + c_2 \left(\frac{X}{L} \right)^{-1} \right] \tag{5.7}$$

where c_1 and c_2 are functions of l_{p}/w. As a result of accounting for local chain stiffness, an additional term in the diffusivity scaling appears in Equation 5.7 when compared to Equation 5.3 for the classical blob theory. The goal of our experimental study is to determine whether the additional contribution in Equation 5.7 is required.

5.3 Experimental Methods

5.3.1 Device fabrication

The nanochannel devices were fabricated from 500 μm thick fused silica wafers (University Wafers) using a layer of electron beam lithography to define the nanochannels (100 μm long) and a layer of contact photolithography to define the microchannels (50 μm wide, 1 μm deep). We used a standard design where two parallel microchannels are connected with an array of approximately square nanochannels (Tegenfeldt et al., 2004). Each lithography step was followed by a reactive ion etching (RIE: $CF_4/CHF_3/Ar$) step to transfer the pattern from the resist layer to the wafer. Note that we have used a cyclic RIE method to improve the etch selectively of fused silica when compared to the PMMA resist used in the electron beam lithography step (Wüest et al., 2005). This led to relatively straight

Table 5.1: Summary of channel dimensions and statistics.

Width (nm)	Height (nm)	D_{eff} (nm)	Number of molecules
137	118	117	57
137	171	143	66
176	171	163	38
205	215	200	62
280	260	260	74

walls when compared to the slanted walls observed in some previous nanochannel studies (Gupta et al., 2015; Iarko et al., 2015). The access ports to the microchannels were created using sand-blasting. Finally, the devices were fusion-bonded to 170 μm thick fused-silica coverslips. The channel depth in devices was characterized prior to the bonding using atomic force microscopy (AFM). The widths of the channel were measured by scanning electron microscopy (SEM) in a duplicate device. The nanochannel cross-sectional dimensions are listed in table 5.1.

5.3.2 DNA preparation

The λ-DNA molecules (48.5 kilobase pairs, New England Biolabs) were dissolved in 0.5× TBE buffer and labeled with YOYO-1 fluorescent dye (Invitrogen) at a concentration of 1 dye molecule for every 10 base pairs (bp) of DNA. The DNA solution was kept at room temperature for an hour followed by heating at 50° C for 12 hours to achieve higher degree of homogenous staining than the samples stained at only room temperature (Nyberg et al., 2013). Prior to start of the experiments, an anti-photobleaching agent β-mercaptoethanol (Sigma-Aldrich, 6% v/v) was

added to the solution. The ionic strength of this solution has been calculated to be 28 mM using a method developed previously (Gupta et al., 2014).

The reasonably moderate molecular weight of λ-DNA, compared to short alternatives (typically 1/2 λ-DNA) and long alternatives (typically T4-DNA), was critical to the success of our experiments. On one hand, the diffusivity of shorter molecules is higher, making the measurement simpler, but they have an extension in the channel comparable to the measurement uncertainty in the largest channel size used here. On the other hand, longer molecules are prone to shear cleavage and photo-nicking. Moreover, longer molecules have smaller diffusion constant requiring longer exposures for accurate measurements. The moderate molecular weight of λ-DNA is an attractive compromise between these experimental limitations.

5.3.3 Parameter estimation

In order to draw comparison between our results and previous studies, we need to estimate three parameters, namely the contour length L, the persistence length l_{p}, and the effective width w. The bisintercalating dye (YOYO-1) used for visualization of DNA is expected to increase L linearly in proportion to the amount of dye bound. We assume that every bound dye molecule adds 0.51 nm to the bare contour length of 0.34 nm per base pair which gives $L = 19$ μm for λ-DNA (Kundukad et al., 2014; Günther et al., 2010).

Although many studies in the past have investigated the effect of dyes on mechanical properties of DNA (Nyberg et al., 2013; Bennink et al., 1999; Günther

et al., 2010), the results are contradictory. Moreover, a recent study by Kundukad et al. (2014) found that the persistence length of DNA is at best weakly affected with the intercalation for a well-equilibrated system. Therefore, we chose to neglect the effect of dye and use the experimentally validated (Hsieh et al., 2008) theory of Dobrynin (2005) to estimate $l_{\mathrm{p}} = 58$ nm at our moderate ionic strength of 28 mM. We employed theory of Stigter (1977) to obtain the effective width $w = 10$ nm.

5.3.4 Nanochannel experiments

The two parallel microchannels on a device were filled with DNA solution by capillary action. The DNA molecules then were driven to the center of a 100 μm long nanochannel using pressure-driven flow and allowed to relax for 120 s before video acquisition. We observed laser excited (Coherent OBIS 473 nm), single DNA molecules using an inverted epifluorescene microscope (Olympus IX73) with a 100× (1.4 N.A.) oil-immersed objective. The images were taken with an EMCCD camera (Photometerics, Cascade II:512) at 2 fps with a 200 ms exposure time. A total of 2000 frames were acquired for each molecule. Each filling resulted in the acquisition of approximately 30 molecules. The total number of molecules observed in each channel size are reported in Table 5.1, which correspond to multiple fillings of the device. After the experiments, the device was cleaned with a standard RCA-1 step and heated to 1000 °C for 6 hours to allow its reuse. The experiments were conducted at room temperature.

5.3.5 Data processing

The center-of-mass of an individual molecule in a movie was tracked using a custom-written Matlab program (Reisner et al., 2005). First, the location of two extreme ends of the molecule (x_1, x_2) were extracted by fitting the molecule's intensity profile, $I(x)$, inside a nanochannel to a linear combination of two error functions, which is a convolution of a Gaussian point-spread function and a box function. Then, the intensity-weighted center-of-mass of the molecule (Strychalski et al., 2008), $x_{\mathrm{com}}(t)$, for a given frame recorded at time t was computed as

$$x_{\mathrm{com}}(t) = \frac{\int_{x_1}^{x_2} xI(x,t)\,dx}{\int_{x_1}^{x_2} I(x,t)\,dx}\,. \tag{5.8}$$

The one-dimensional diffusivity, D_{t} was obtained from mean squared displacement (MSD),

$$\mathrm{MSD}(\delta t) = \left\langle [x_{\mathrm{com}}(t) - x_{\mathrm{com}}(t-\delta t)]^2 \right\rangle_{\mathrm{n,t}} = 2D_{\mathrm{t}}\delta t, \tag{5.9}$$

where $\langle \dots \rangle_{\mathrm{n,t}}$ represent an ensemble and a time average and δt is the time lag between images.

In general, the MSD increased linearly with time, as seen in Figure 5.1. However, the dynamic diffusion coefficient, $\mathrm{MSD}/2\delta t$, decays continuously up to a maximum time lag of 150 s in the smallest channel size before reaching a fairly constant value. Similar behavior was observed in nanoslits (Hsieh et al., 2007) for $\delta t < 0.2$ s and was investigated in detail for colloidal particle tracking experiments (Savin and Doyle, 2005). At larger time lags ($\delta t > 400$ s), $\mathrm{MSD}/2\delta t$ fluctuates substantially due to limited statistics. The one-dimensional axial diffusivity, D_{t},

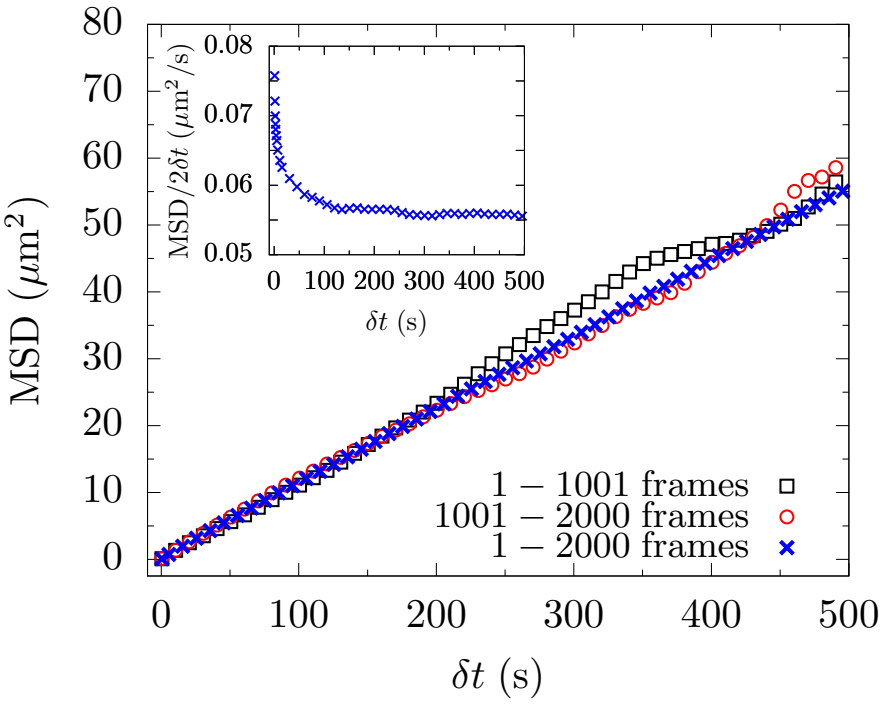

Figure 5.1. Mean squared displacement of 57 λ-DNA molecules inside a 137 × 118 nm^2 nanochannel as a function of time lag from different segments of recorded images. The power law fit to data from the complete movie (blue ×) for δt (s) ∈ [150, 400] yields an exponent 0.98, indicating normal diffusive behavior. The inset shows the dynamic diffusion coefficient, MSD/2δt as a function of time lag. After an initial decay till δt ~ 150 s, the MSD/2δt value is reasonably constant. A linear fit to MSD for δt (s) ∈ [150, 400] in this nanochannel gives the diffusion coefficient $D_{\rm t}$ = 0.05596 ± 0.00003 μm^2/s. Similar plots for the other channel sizes are shown in Figure 5.2.

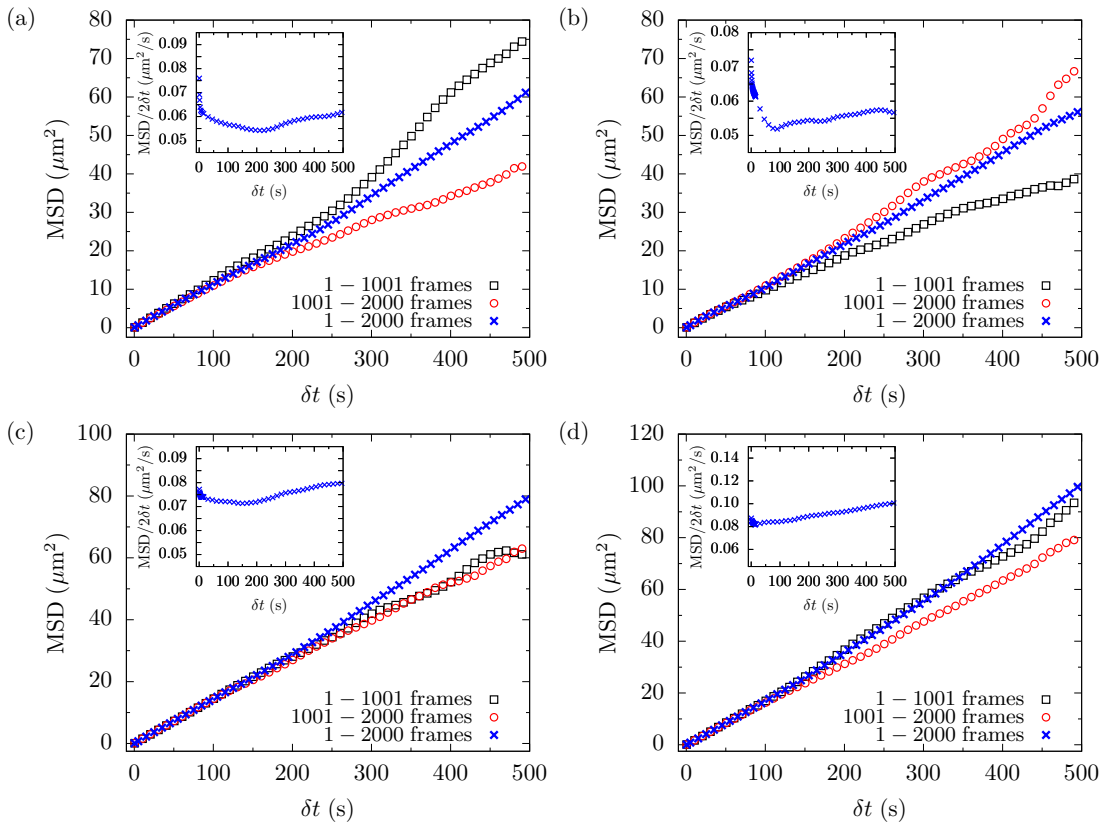

Figure 5.2. Mean squared displacement of λ-DNA molecules as a function of time lag for D_{eff} = (a) 143 nm, (b) 163 nm, (c) 200 nm and (d) 260 nm.

was obtained by fitting the MSD curve with a linear function for δt (s) $\in$ [150, 400] to take advantage of the stable, well-sampled region of the dynamic diffusion coefficient. The fit was constrained such that MSD (0) = 0.

In the course of the experiments, there is a possibility of either adsorption of DNA molecules on the channel surface or induced motion of DNA molecules due to spurious fluid flow. To filter out the systematic errors that would occur in either case, we computed the scaling of MSD $\sim \delta t^{\beta}$ for ensemble averaged trajectories of all the molecules recorded in a single filling of a device with DNA solution.

Any DNA molecules from the single filling of a device ($\approx$ 30 molecules per filling) with anomalously diffusive trajectories ($|\beta - 1| > 0.1$) were not included in the final ensemble. The total number of acceptable λ-DNA molecules acquired in each channel size is listed in Table 5.1.

Single DNA molecules also photobleach, especially with the relatively long exposure time used in the experiments here when compared to our previous experiments in a similar setup (Gupta et al., 2015). Moreover, photochemical reactions can lead to either shortening of the contour length or fragmentation of the molecule (Akerman and Tuite, 1996; Alizadehheidari et al., 2015). Photocleaved molecules become systematically shorter in length and diffuse faster over time. In order to make sure that the physical properties of the DNA are consistent over the time period of image acquisition, we used the following steps. First, we divided single movies (2000 frames) into five sub-movies of equal length. Then, we compared the extension of molecules among sub-movies. We did not find any statistically significant shortening of molecules over time, indicating a negligible effect of either photobleaching or photocleaving. We also used a method developed by Hsieh et al. (2007) for DNA studies in nanoslits where they compare MSD of short movies as a secondary test. Figure 5.1 shows one such analysis done inside a nanochannel with cross-sectional dimension as 137×118 nm^2. The MSD curves are very close to each other for $\delta t < 200$ s and no systematic deviation is observed for $\delta t > 200$ s beyond that caused by fluctuations due to limited statistics, thus lending confidence in absence of any substantial effect caused by prolonged light exposure times.

5.4 Results and Discussion

The first step in our analysis is to check whether the experimental data we are analyzing belongs to the extended de Gennes regime. Since the channel sizes used here are almost square, we calculate the effective channel size as $D_{\text{eff}} = \sqrt{(D_1 - \delta)(D_2 - \delta)}$, where D_1 and D_2 are channel height and width, respectively. The parameter, δ is a wall-DNA depletion length to account for electrostatic interaction. We estimate $\delta = w$, following previous studies (Wang et al., 2011; Gupta et al., 2015; Iarko et al., 2015). Our effective channel sizes range from 117 nm to 260 nm (table 5.1) which indeed span much of the extended de Gennes regime; the scaling theory by Odijk (2008) suggests that the range of the extended de Gennes regime for our conditions is 116 nm to 336 nm, which is quite narrow. However, simulations by Dai et al. (2014) suggest that the lower bound is 232 nm and upper bound is 754 nm. The possible consequences of this contradiction will be discussed later.

Figure 5.3a shows the result for the average extension of DNA molecules, X, as a function of effective channel size, D_{eff}. The extension and the channel size are normalized in a manner to produce universal curves in the extended de Gennes regime. Our results agree with previous experimental studies and are very close to predictions by the theory Werner and Mehlig (2014), $X/L = 1.176(62)(D_{\text{eff}}^2/(l_{\text{p}}w))^{-1/3}$. The deviation between experiments and theory have been explained in the past as a likely result of the inaccuracy in estimation of D_{eff} (Gupta et al., 2015; Iarko et al., 2015).

Figure 5.3b shows the result for variance about average extension, σ^2; the

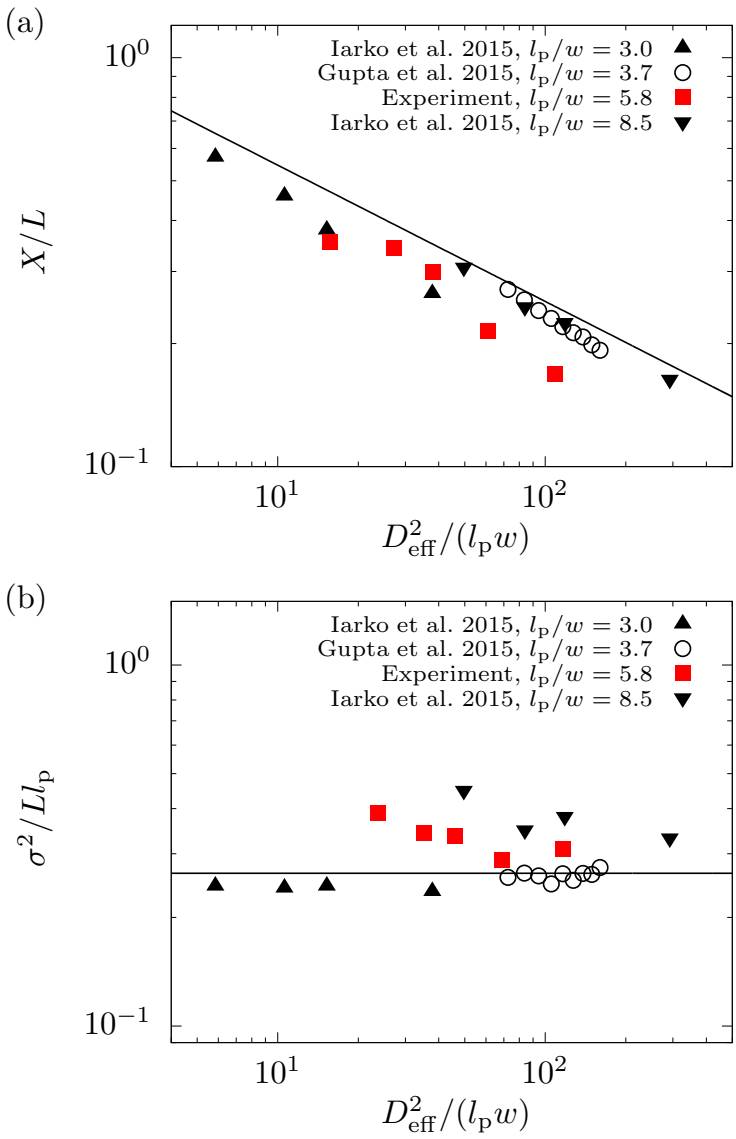

Figure 5.3. (a) Average fractional extension, X/L and (b) normalized variance in extension, σ^2/Ll_p, from the current contribution (red ■) and previous experimental studies (Gupta et al., 2015; Iarko et al., 2015) (black 5, ▲, ▼). The experimental results are compared to the extended de Gennes theory (Werner and Mehlig, 2014) (black solid line). The standard errors for our experiments are smaller than the size of data points.

values are normalized to collapse the experiments done with DNA of different sizes in different ionic strength environments. As expected, σ^2 from our experiments does not depend upon D_{eff} similar to previous studies in the extended de Gennes regime (Gupta et al., 2015; Iarko et al., 2015). When compared to the exact value of variance predicted by theory in this regime (Werner and Mehlig, 2014), $\sigma^2/(Ll_{\text{p}}) = 0.264(99)$, we observe a systematic deviation in experimental value as a function of chain semiflexibility l_{p}/w; σ^2/Ll_{p} increases with increase in l_{p}/w. This can be attributed to either inaccuracy in the methodology used to determine l_{p} as a function of ionic strength or the failure of the experimental system to satisfy the strong inequalities assumed to formulate the extended de Gennes theory. Indeed, the lower bound for the regime suggested by Dai et al. (2014), $D_{\text{eff}} = 4l_{\text{p}}$, is not satisfied by most of the experimental studies.

While there remain discrepancies between our experiments and theory, Figure 5.3 shows that the thermodynamic data in the present experiments are in good agreement with previous experimental work. We thus proceeded to investigate the dependence of the axial diffusivity of DNA, D_{t}, as a function of average extension, X, shown in Figure 5.4. We chose to plot the data in terms of the extension X because it can be measured directly in the experiments. As a result, we avoid any systematic errors that arise from using D_{eff}, whose accuracy has been questioned in previous nanochannel experiments (Gupta et al., 2015; Iarko et al., 2015). In order to test the validity of predictions from the blob theories, the simplest approach would be to compute a power law for $D_{\text{t}} \sim X^{\alpha}$. We obtain $\alpha = -0.72 \pm 0.25$ at 95% confidence level in Figure 5.4a. The apparent exponent is weaker than the -1 scaling predicted from the classical blob theory. Moreover, the range for X used

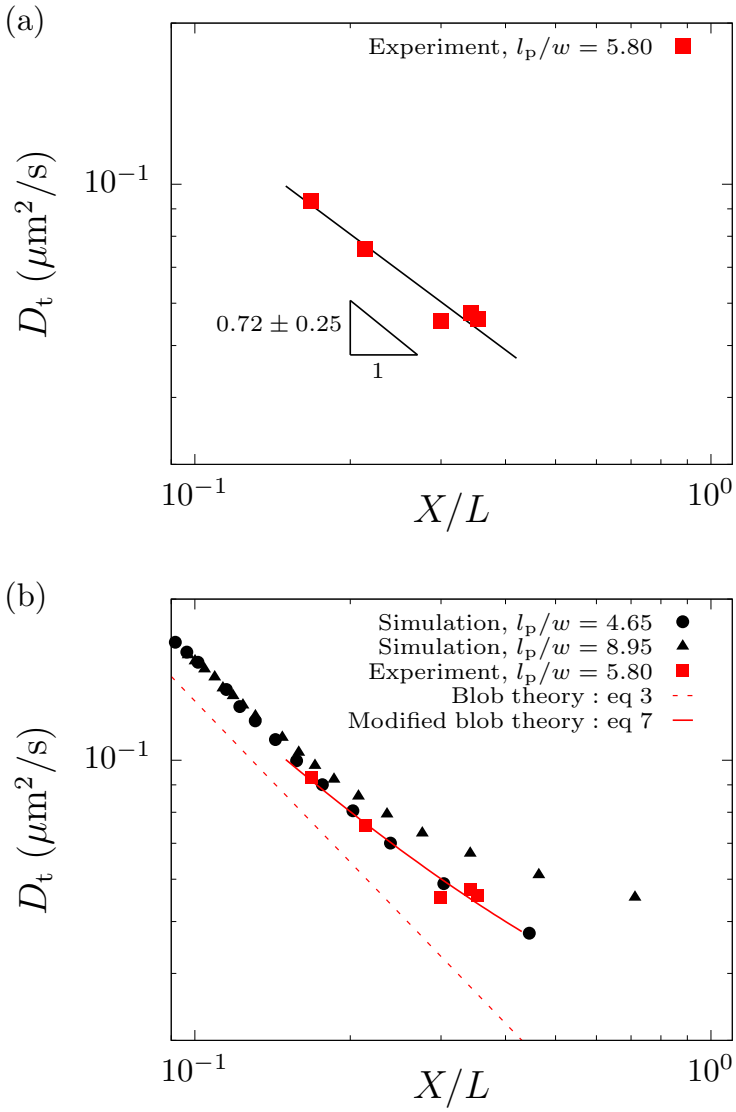

Figure 5.4. Log-log plot of diffusion constant, D_t, as a function of the average fractional extension, X/L. (a) Power law fit (solid black line) to the experimental data (red ■) yields an exponent −0.72±0.25. (b) The simulation (Muralidhar and Dorfman, 2015) and theoretical predictions (Equations 5.3 and 5.7) were computed with Rouse diffusivity, $D_R = 0.0129$ μm^2/s for λ-DNA ($L = 19$ μm) in water ($\eta = 0.89$ cP) at room temperature ($T = 25°$ C). The line calculated from Equation 5.7 gives $c_1 = 1.5 \pm 1.3$ and $c_2 = 0.9 \pm 0.3$. The error bars for the experiments are calculated from uncertainty in linear fit to uncorrelated MSD vs δt and are smaller than symbol size. All uncertainties in the parameters correspond to a 95% confidence level.

here is too small to accurately test a power law dependence and the error bars are substantial. Overall, the outcome of the power law analysis is inconclusive.

Instead, in Figure 5.4b, we directly compare the predictions from the blob theories in Equations 5.3 and 5.7 to our experimental data. Note that the value of the Rouse diffusivity appearing in both eqs, $D_{\mathrm{R}} = 0.0129\ \mu\mathrm{m}^2/\mathrm{s}$, is estimated for λ-DNA ($L = 19\ \mu\mathrm{m}$) in water ($\eta = 0.89$ cP) at room temperature ($T = 25^\circ$ C). Clearly, the modified blob theory expression in Equation 5.7 agrees well with our data for the entire range of X used here when compared to the classical blob theory prediction from Equation 5.3. A more robust method to compare two theories is to investigate the statistical significance of the additional term c_1 appearing in eq 7 via a hypothesis test. The best fit to Equation 5.7 gives $c_1 = 1.5 \pm 1.3$ and $c_2 = 0.9 \pm 0.3$ at 95% confidence level. A hypothesis test at the 5% significance level rejects the $c_1 = 0$ hypothesis (classical blob theory) in favor of the $c_1 > 0$ hypothesis (modified blob theory). This is the key result of this chapter.

Finally, we compare the experimental data to simulations of the Kirkwood diffusivity of a discrete wormlike chain model by Muralidhar and Dorfman (2015). In order to simulate DNA, it is essential to establish its l_{p}, w, and L values. Since simulation results for the l_{p}/w value corresponding to our experimental conditions are not available, we plot the simulation data for two closest values of l_{p}/w. Thereafter, the diffusion coefficients were computed from the dimensionless simulation results using the estimated Rouse diffusivity, $D_{\mathrm{R}} = 0.0129\ \mu\mathrm{m}^2/\mathrm{s}$ for $L = 19\ \mu\mathrm{m}$. The quantitative agreement between the simulation data at $l_{\mathrm{p}}/w = 4.95$ and the experimental data at $l_{\mathrm{p}}/w = 5.8$ in Figure 5.4b is excellent considering that this analysis did not require any fitting parameters. Indeed, as an alternate test,

we fit Equation 5.7 to the simulation data and obtained constants identical to the experiments; the constants for simulation data set are c_1 = 1.4±0.4 and c_2 = 1.0±0.1 for the X range commensurate to the experimental data.

5.5 Conclusion

We have presented measurements of center-of-mass diffusivity of DNA confined in approximately square nanochannels with effective sizes ranging from 117 to 260 nm at moderate ionic strength. We found that it is necessary to modify the blob theory to account for local chain stiffness in order to explain the experimental results in the extended de Gennes regime. This modification is very similar to that done in nanoslits albeit over a wider range of nanoslit height. Our results are in quantitative agreement with previous simulations of the Kirkwood diffusivity of a discrete wormlike chain model without need for fitting parameters, thus lending confidence to the simulation predictions.

From methodology point of view, our work will serve as a valuable reference for future experimental studies of DNA diffusivity over wider range of channel sizes via careful control of the ionic strength, which can probe the role of local chain stiffness for DNA in confinement. Another area of interest would be to see how the aspect ratio of rectangular channels alters the dynamic response of DNA, as the rectangular channel shape affects the hydrodynamics and polymer configurations differently.

Chapter 6

Conclusion

At the beginning, we identified that our goal is to provide an experimentally validated model for the entire confinement spectrum of DNA inside channel geometry. Additionally, we expressed our desire to find answers to some fundamental questions surrounding the progression of confinement of DNA from coiled to fully stretched state. While we have made considerable progress on both the fronts, we have also faced some interesting challenges along the way. As such, we made appropriate pit stops to address them. Finally, some open questions remain, which will be discussed at the end of this chapter.

Much of the thesis has been focussed on measuring the thermodynamic properties of single DNA molecules when confined in nanochannel of varying size via fluorescence microscopy, specifically the fluctuation about average extension of single DNA molecules. When a DNA molecule is inserted inside a nanochannel, it spreads out along the channel. The measure of this spread is the extension. While the extension is the simplest observable metric and quite relevant for genomic ap-

plications, measurements of both the extension and its variance (fluctuation about an average value over time) are required to resolve open questions surrounding the different regimes. For example, scaling of the extension as a function of channel size is similar for both extended de Gennes regime and classical de Gennes regime. However, scaling of the variance in extension as a function of channel size should be able to clearly distinguish between the two regimes. Most of the experimental studies done prior to this thesis were primarily focused on measuring the extension as a function of channel size, as summarized by Reisner et al. (2012). While Su et al. (2011) and Thamdrup et al. (2008) did measure the fluctuations, their measurements were done over a channel size range that was too narrow to deduce anything significant about various confinement models.

It was clear from the beginning that fabrication of multiple channel sizes were required to experimentally test different models for confinement of DNA inside nanochannels. We started with exploring nanochannel device designs that could afford multiple channel sizes quickly in Chapter 3. Confinement spectroscopy is a potentially powerful approach but uses rectangular channels of very high aspect ratio. With the combination of fluorescence microscopy experiments and PERM simulations, we were able to show that the average extension is only weakly affected by the channel aspect ratio. In contrast, the fluctuations of the chain extension qualitatively differ between rectangular channels and square channels with the same cross-sectional area. In the rectangular channels, the two channel dimensions lie in different regimes, leading to the mixing of different confinement regimes. Additionally, both our experiments and simulations clearly indicate that the fluctuation about average extension continued to increase with increase in as-

pect ratio. As a result, confinement spectroscopy with high aspect ratio is probably not the best tool for probing the different theories for channel-confined DNA. One should be mindful about the effect of the aspect ratio on thermodynamic properties. These findings also have important consequences for use of nanochannels for high-throughout DNA mapping. All other considerations being equal, when comparing devices that yield the same DNA extension, the device with the lower fluctuations is to be preferred as fluctuations may adversely impact barcode alignment.

In Chapter 4, we used a confinement spectroscopy device with low aspect ratio to probe the extended de Gennes regime. By performing the measurements of the fluctuation in chain extension as a function of channel size, we have provided experimental evidence supporting the theoretical prediction of weak excluded volume in the extended de Gennes regime. Simultaneously, we developed an approach to account for the effect of molecular weight dispersity of long DNA samples. These experiments represent the first test of the emerging theoretical framework describing channel-confined DNA.

While we have mainly focused on investigating the extended de Gennes regime, the transitional response of DNA from weak confinement to strong confinement conditions remained an open question. A recent theory by Odijk (2008) backed up by simulations (Muralidhar et al., 2014a), suggest formation of deflection segments at smaller length scales accompanied by formation of hairpins at larger length scales in the backfolded Odijk regime. Experimentally testing this regime is challenging as it is narrow in terms of the channel size range and will require very high molecular weight chains. Alternatively, Werner et al. (2017) came up with

a theory which is based on one-dimension telegraph process and does not require going into microscopic details of the chain configuration. Indeed, we were able to show that this theory collapsed experimental data from thirteen different studies involving vast range of square channel sizes and other experimentally relevant parameter space, on to a single master curve. The most important consequence of such a synergy between experiments and theory is that better devices could be designed in-silico for genomic applications with desired experimentally relevant parameters. Additionally, the universal mechanisms of channel-confinement of DNA could be understood in a better manner.

For the purpose of this work, we ignored the dependence of the persistence length on the underlying base pair sequence. The recent reconciliation of experimental observations and theory for nanochannel confined DNA can now be used as a tool to investigate the dependence of the persistence length on the base pair sequence.

DNA is a polyelectrolyte, and the electrostatic interactions between charged walls, and the DNA molecule can dramatically alter the behavior of a confined chain, especially in narrow channels whose size is less than 50 nm. In our work, these interactions were modeled by using a wall-DNA depletion length, which was assumed to be equal to the effective width of the DNA. The accuracy of such an assumption is unknown. One way to accurately determine the wall-depletion length could be to do experiments using DNA in nanochannels with careful control of ionic strength; the electrostatic interactions are dependent upon the ionic strength. For example, Iarko et al. (2015) observed that the accuracy of estimation of wall-DNA depletion length for channels, decreases with decrease

in ionic strength via only two ionic strength conditions. Alternatively, the full non-linear Poisson-Boltzmann equation could be solved near the walls and the corners of the channel. Indeed, a combination of both should be able to resolve this open question.

While the statistical mechanical properties of DNA in nanochannel confinement have been investigated in detail, the dynamic response still remains relatively unexplored, in particularly with respect to the center-of-mass diffusion coefficient. Reisner et al. (2005) explored the dynamic behavior indirectly by measuring the relaxation time of chain extension. In Chapter 5, we have directly measured the diffusivity of λ-DNA confined in approximately square nanochannels. We found that it is necessary to modify the classic results of de Gennes theory to account for local chain stiffness in order to explain the experimental results for DNA in practical channel sizes. This modification is very similar to that done in a nanoslit geometry albeit over a wider range of nanoslit height. Additionally, our results are in quantitative agreement with previous simulations of the Kirkwood diffusivity of a discrete wormlike chain model without need of fitting parameters, thereby lending confidence in usage of Kirkwood-Riseman approach, to simulate dynamics of channel-confined DNA.

From a methodology point of view, our work will serve as a valuable reference for future experimental studies of DNA diffusivity over wider range of channel sizes via careful control of the ionic strength, which can probe the role of local chain stiffness for DNA in confinement. Certainly, another area of interest would be to see how high aspect ratio of rectangular channels affects the dynamic response of DNA. We have indirectly measured dynamic response via relaxation

time measurements for rectangular channel in Chapter 4 and found them to be an order of magnitude smaller than expected from similar measurements done in square channel (Reisner et al., 2005). The aspect ratio considerably altered the chain statistics in Chapter 3, which in combination with the asymmetric nature of hydrodynamics in rectangular channels, will lead to complicated behavior in the dynamic properties.

In the end, we believe that the complete experimental test of the phase diagram for channel-confined DNA, with careful control over molecular weight dispersity, channel geometry, electrostatic interactions and sequence dependent length scales, will provide a firm foundation for the emerging genomic technology.

Bibliography

Akerman, B. and E Tuite, "Single and double strand photocleavage of DNA by YO, YOYO and TOTO", Nucleic Acids Res. **24**, 1080 (1996).

Alizadehheidari, M., E. Werner, C. Noble, M. Reiter-Schad, L. K. Nyberg, J. Fritzsche, B. Mehlig, J. O. Tegenfeldt, T. Ambjornsson, F. Persson, and F. Westerlund, "Nanoconfined circular and linear DNA: Equilibrium conformations and unfolding kinetics", Macromolecules **48**, 871 (2015).

Almal, S. H. and H. Padh, "Implications of gene copy-number variation in health and diseases", J. Hum. Genet. **57**, 6 (2012).

Ansorge, W. J., "Next-generation DNA sequencing techniques", New Biotechnol. **25**, 195 (2009).

Bakajin, O. B., T. A. J. Duke, C. F. Chou, S. S. Chan, R. H. Austin, and E. C. Cox, "Electrohydrodynamic stretching of DNA in confined environments", Phys. Rev. Lett. **80**, 2737 (1998).

Balducci, A. G., P Mao, J Han, and P. S. Doyle, "Double-stranded DNA diffusion in slitlike nanochannels", Macromolecules **39**, 6273 (2006).

Behrens, S. H. and D. G. Grier, "The charge of glass and silica surfaces", J. Chem. Phys. **115**, 6716 (2001).

Bennink, M. L., O. D. Schärer, R. Kanaar, K. Sakata-Sogawa, J. M. Schins, J. S. Kanger, B. G. de Grooth, and J. Greve, "Single-molecule manipulation of double-stranded DNA using optical tweezers: Interaction studies of DNA with RecA and YOYO-1", Cytometry **36**, 200 (1999).

Bensimon, A, A Simon, A Chiffaudel, V Croquette, F Heslot, D Bensimon, et al., "Alignment and sensitive detection of DNA by a moving interface", Science **265**, 2096 (1994).

Bogas, D., L. Nyberg, R. Pacheco, N. F. Azevedo, J. P. Beech, M. Gomila, J. Lalucat, C. M. Manaia, O. C. Nunes, J. O. Tegenfeldt, et al., "Applications of optical DNA mapping in microbiology.", BioTechniques **62**, 255 (2017).

Brochard, F and P.-G. De Gennes, "Dynamics of confined polymer-chains", J. Chem. Phys. **67**, 52 (1977).

Brochard, F., T. Tanaka, N. Borghi, and P.-G. de Gennes, "Semiflexible polymers confined in soft tubes", Langmuir **21**, 4144 (2005).

Brunet, A., C. Tardin, L. Salomé, P. Rousseau, N. Destainville, and M. Manghi, "Dependence of DNA persistence length on ionic strength of solutions with monovalent and divalent salts: A joint theory–experiment study", Macromolecules **48**, 3641 (2015).

Burkhardt, T. W., Y. Yang, and G. Gompper, "Fluctuations of a long, semiflexible polymer in a narrow channel", Phys. Rev. E **82**, 041801 (2010).

Cao, H., A. R. Hastie, D. Cao, E. T. Lam, Y. Sun, H. Huang, X. Liu, L. Lin, W. Andrews, S. Chan, S Huang, X Tong, R Requa, T Anantharaman, A Krogh, H Yang, H Cao, and X Xu, "Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology", GigaScience **3**, 34 (2014).

Carlsson, C., M. Jonsson, and B. Åkerman, "Double bands in DNA gel electrophoresis caused by bis-intercalating dyes", Nucleic Acids Res. **23**, 2413 (1995).

Chan, E. Y., N. M. Goncalves, R. A. Haeusler, A. J. Hatch, J. W. Larson, A. M. Maletta, G. R. Yantz, E. D. Carstea, M. Fuchs, G. G. Wong, et al., "DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags", Genome Res. **14**, 1137 (2004).

Chodera, J. D., W. C. Swope, J. W. Pitera, C Seok, and K. A. Dill, "Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations", J. Chem. Theory Comput. **3**, 26 (2007).

Cifra, P., "Channel confinement of flexible and semiflexible macromolecules", J. Chem. Phys. **131**, 224903 (2009).

Cifra, P., Z. Benková, and T. Bleha, "Chain extension of DNA confined in channels", J. Phys. Chem. B **113**, 1843 (2009).

Dai, L. and P. S. Doyle, "Comparisons of a polymer in confinement versus applied force", Macromolecules **46**, 6336 (2013).

Dai, L., D. R. Tree, J. R. C. Van Der Maarel, K. D. Dorfman, and P. S. Doyle, "Revisiting blob theory for DNA diffusivity in slitlike confinement", Phys. Rev. Lett. **110**, 168105 (2013).

Dai, L., J. van der Maarel, and P. S. Doyle, "Extended de Gennes regime of DNA confined in a nanochannel", Macromolecules **47**, 2445 (2014).

Dai, L., C. B. Renner, and P. S. Doyle, "The polymer physics of single DNA confined in nanochannels", Adv. Colloid Interface Sci. **232**, 80 (2016).

Daoud, M. and P.-G. de Gennes, "Statistics of macromolecular solutions trapped in small pores", J. Phys. Paris **38**, 85 (1977).

Das, S. K., M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao, "Single molecule linear analysis of DNA in nano-channel labeled with specific sequence specific fluorescent probes", Nucleic Acids Res. **38**, e177 (2010).

Deamer, D., M. Akeson, and D. Branton, "Three decades of nanopore sequencing", Nat. Biotechnol. **34**, 518 (2016).

Dobrynin, A. V., "Electrostatic persistence length of semiflexible and flexible polyelectrolytes", Macromolecules **38**, 9304 (2005).

Dobrynin, A. V., "Effect of counterion condensation on rigidity of semiflexible polyelectrolytes", Macromolecules **39**, 9519 (2006).

Doi, M. and S. F. Edwards, *The theory of polymer dynamics*, Vol. 73 (oxford university press).

Dorfman, K., D Gupta, A Jain, A Muralidhar, and D. Tree, "Hydrodynamics of DNA confined in nanoslits and nanochannels", Eur. Phys. J.: Spec. Top. **223**, 3179 (2014).

Dorfman, K. D., S. B. King, D. W. Olson, J. D. P. Thomas, and D. R. Tree, "Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis and DNA stretching", Chem. Rev. **113**, 2584 (2013).

Earl, D., K. Bradnam, J. S. John, A. Darling, D. Lin, J. Fass, H. O. K. Yu, V. Buffalo, D. R. Zerbino, M. Diekhans, et al., "Assemblathon 1: a competitive assessment of de novo short read assembly methods", Genome Res. **21**, 2224 (2011).

Eichler, E. E., R. A. Clark, and X. She, "An assessment of the sequence gaps: unfinished business in a finished human genome", Nat. Rev. Genet. **5**, 345 (2004).

Eisenstein, M., *Startups use short-read data to expand long-read sequencing market.*

Feuk, L., A. R. Carson, and S. W. Scherer, "Structural variation in the human genome", Nat. Rev. Genet. **7**, 85 (2006).

Frykholm, K., M. Alizadehheidari, J. Fritzsche, J. Wigenius, M. Modesti, F. Persson, and F. Westerlund, "Probing physical properties of a DNA-protein complex using nanofluidic channels", Small **10**, 884 (2014).

Glazer, A. N. and H. S. Rye, "Stable dye–DNA intercalation complexes as reagents for high-sensitivity fluorescence detection", Nature **359**, 859 (1992).

Goodwin, S., J. D. McPherson, and W. R. McCombie, "Coming of age: ten years of next-generation sequencing technologies", Nat. Rev. Genet. **17**, 333 (2016).

Grassberger, P., "Pruned-enriched Rosenbluth method: Simulations of theta polymers of chain length up to 1,000,000", Phys. Rev. E **56**, 3682 (1997).

Günther, K., M. Mertig, and R. Seidel, "Mechanical and structural properties of YOYO-1 complexed DNA", Nucleic Acids Res. **38**, 6526 (2010).

Gupta, D. and K. D. Dorfman, "Diffusion of DNA in nanochannels", to be submitted (2017).

Gupta, D., J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, and W. Reisner, "Mixed confinement regimes during equilibrium confinement spectroscopy of DNA", J. Chem. Phys. **140**, 214901 (2014).

Gupta, D., J. J. Miller, A. Muralidhar, S. Mahshid, W. Reisner, and K. D. Dorfman, "Experimental evidence of weak excluded volume effects for nanochannel confined DNA", ACS Macro Lett. **4**, 759 (2015).

Gurrieri, S., K. S. Wells, I. D. Johnson, and C. Bustamante, "Direct visualization of individual DNA molecules by fluorescence microscopy: characterization of the factors affecting signal/background and optimization of imaging conditions using YOYO", Anal. Biochem. **249**, 44 (1997).

Hastie, A. R., L. Dong, A. Smith, J. Finklestein, E. T. Lam, N. Huo, H. Cao, P.-Y. Kwok, K. R. Deal, J. Dvorak, M.-C. Luo, Y. Gu, and M. Xiao, "Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex aegilops tauschii genome", PLoS ONE **8**, e55864 (2013).

Hsieh, C.-C., A. Balducci, and P. S. Doyle, "An experimental study of DNA rotational relaxation time in nanoslits", Macromolecules **40**, 5196 (2007).

Hsieh, C.-C., A. Balducci, and P. S. Doyle, "Ionic effects on the equilibrium dynamics of DNA confined in nanoslits", Nano Lett. **8**, 1683 (2008).

Iarko, V., E. Werner, L. K. Nyberg, V. Müller, J. Fritzsche, T. Ambjörnsson, J. P. Beech, J. O. Tegenfeldt, K. Mehlig, F. Westerlund, and B. Mehlig, "Extension of nanoconfined DNA: Quantitative comparison between experiment and theory", Phys. Rev. E **92**, 062701 (2015).

Jo, K., 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", Proc. Natl. Acad. Sci. USA **104**, 2673 (2007).

Karger, B. L. and A. Guttman, "DNA sequencing by CE", Electrophoresis **30** (2009).

Kim, Y., K. S. Kim, K. L. Kounovsky, R. Chang, G. Y. Jung, J. J. DePablo, K. Jo, and D. C. Schwartz, "Nanochannel confinement: DNA stretch approaching full contour length", Lab Chip **11**, 1721 (2011).

Kovacic, R. T., L. Comai, and A. J. Bendich, "Protection of megabase DNA from shearing", Nucleic Acids Res. **23**, 3999 (1995).

Ku, H. H., "Notes on the use of propagation of error formulas", J. Res. Nat. Stand. C **70**, 263 (1966).

Kundukad, B., J. Yan, and P. S. Doyle, "Effect of YOYO-1 on the mechanical properties of DNA.", Soft Matter **10**, 9721 (2014).

Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, and P.-Y. Kwok, "Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly", Nat. Biotechnol. **30**, 771 (2012).

Latinwo, F. and C. M. Schroeder, "Model systems for single molecule polymer dynamics", Soft Matter **7**, 7907 (2011).

Levene, M. J., J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb, "Zero-mode waveguides for single-molecule analysis at high concentrations", Science **299**, 682 (2003).

Levy-Sakin, M. and Y. Ebenstein, "Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy", Curr. Opin. Biotech. **24**, 690 (2013).

Levy, S. L. and H. G. Craighead, "DNA manipulation, sorting, and mapping in nanofluidic systems", Chem. Soc. Rev. **39**, 1133 (2010).

Lin, P. K., J. F. Chang, C. H. Wei, P Tsao, W Fann, and Y.-L. Chen, "Partial hydrodynamic screening of confined linear and circular double-stranded DNA dynamics", English, Phys. Rev. E **84**, 031917 (2011).

Mansfield, M. L. and J. F. Douglas, "Is duplex DNA a swollen random coil?", Soft Matter **9**, 8914 (2013).

Mardis, E. R., "Next-generation DNA sequencing methods", Annu. Rev. Genomics Hum. Genet. **9**, 387 (2008).

Marie, R., J. N. Pederson, D. L. V. Bauer, K. H. Rasmussen, M Yusuf, E Volpi, H. Flyvbjerg, A. Kristensen, and K. U. Mir, "Integrated view of genome structure and sequence of a single DNA molecule in a nanofluidic device", Proc. Natl. Acad. Sci. USA **110**, 4893 (2013).

Menard, L. D. and J. M. Ramsey, "Nanofluidic device architectures for the controlled transport and high throughput analysis of single DNA molecules in nanochannels", in Proceedings of microtas 2013 ().

Metzker, M. L., "Sequencing technologiesâăťthe next generation", Nat. Rev. Genet. **11**, 31 (2010).

Michaeli, Y. and Y. Ebenstein, "Channeling DNA for optical mapping", Nat. Biotechnol. **30**, 762 (2012).

Müller, V. and F. Westerlund, "Optical DNA mapping in nanofluidic devices: principles and applications", Lab Chip **17**, 579 (2017).

Muralidhar, A. and K. D. Dorfman, "Kirkwood diffusivity of long semiflexible chains in nanochannel confinement", Macromolecules **48**, 2829 (2015).

Muralidhar, A., D. R. Tree, and K. D. Dorfman, "Backfolding of wormlike chains confined in nanochannels", Macromolecules **47**, 8446 (2014).

Muralidhar, A., D. R. Tree, Y. Wang, and K. D. Dorfman, "Interplay between chain stiffness and excluded volume of semiflexible polymers confined in nanochannels", J. Chem. Phys. **140**, 084905 (2014).

Nilsson, A. N., 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 Res., e118 (2014).

Nyberg, L., 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 Res. **41**, e184 (2013).

Odijk, T., "Polyelectrolytes near the rod limit", J. Polym. Sci. B **15**, 477 (1977).

Odijk, T., "Scaling theory of DNA confined in nanochannels and nanoslits", Phys. Rev. E **77**, 060901(R) (2008).

Odijk, T., "On the statistics and dynamics of confined or entangled stiff polymers", English, Macromolecules **16**, 1340 (1983).

Perkins, T. T., D. E. Smith, R. G. Larson, and S. Chu, "Stretching of a single tethered polymer in a uniform flow", Science **268**, 83 (1995).

Persson, F. and J. O. Tegenfeldt, "DNA in nanochannels - directly visualizing genomic information", Chem. Soc. Rev. **39**, 985 (2010).

Persson, F., P. Utko, W. Reisner, N. B. Larsen, and A. Kristensen, "Confinement spectroscopy: Probing single DNA molecules with tapered nanochannels", Nano Lett. **9**, 1382 (2009).

Pollack, A., *Dna sequencing caught in deluge of data*, `http://www.nytimes.com/2011/12/01/business/dna-sequencing-caught-in-deluge-of-data.html?mcubz=1`, 11-30-2011.

Prellberg, T. and J. Krawczyk, "Flat histogram version of the pruned and enriched rosenbluth method", Phys. Rev. Lett. **92**, 120602 (2004).

Przeworski, M., R. R. Hudson, and A. D. Rienzo, "Adjusting the focus on human variation", Trends Genet. **16**, 296 (2000).

Randall, G. C. and P. S. Doyle, "DNA deformation in electric fields: DNA driven past a cylindrical obstacle", Macromolecules **38**, 2410 (2005).

Reifenberger, J. G., K. D. Dorfman, and H Cao, "Topological events in single molecules of E. coli DNA confined in nanochannels", Analyst **140**, 4887 (2015).

Reinhart, W. F., J. G. Reifenberger, D. Gupta, A. Muralidhar, J. Sheats, H. Cao, and K. D. Dorfman, "Distribution of distances between DNA barcode labels in nanochannels close to the persistence length", J. Chem. Phys. **142**, 064902 (2015).

Reisner, W., K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin, "Statics and dynamics of single DNA molecules confined in nanochannels", Phys. Rev. Lett. **94**, 196101 (2005).

Reisner, W., 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", Phys. Rev. Lett. **99**, 058302 (2007).

Reisner, W., J. N. Pedersen, and R. H. Austin, "DNA confinement in nanochannels: Physics and biological applications", Rep. Prog. Phys. **75**, 106601 (2012).

Rubinstein, M. and R. H. Colby, *Polymer physics*, Vol. 23 (Oxford University Press New York).

Sanger, F., S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors", Proc. Natl. Acad. Sci. USA **74**, 5463 (1977).

Santiago, J. G., *Buffer calculator*, `http://microfluidics.stanford.edu/download/index.html`.

Savin, T. and P. S. Doyle, "Static and dynamic errors in particle tracking microrheology", Biophys. J. **88**, 623 (2005).

Schaefer, D., J. Joanny, and P Pincus, "Dynamics of semiflexible polymers in solution", Macromolecules **13**, 1280 (1980).

Schellman, J., "Flexibility of DNA", Biopolymers **13**, 217 (1974).

Schwartz, D. C., X. Li, L. I. Hernandez, S. P. Ramnarain, E. J. Huff, and Y.-K. Wang, "Ordered restriction maps of saccharomyces cerevisiae chromosomes constructed by optical mapping", Science **262**, 110 (1993).

Shaqfeh, E. S., "The dynamics of single-molecule DNA in flow", J. Non-Newtonian Fluid **130**, 1 (2005).

Sheats, J., J. G. Reifenberger, H. Cao, and K. D. Dorfman, "Measurements of DNA barcode label separations in nanochannels from time-series data", Biomicrofluidics **9**, 064119 (2015).

Shendure, J. and H. Ji, "Next-generation DNA sequencing", Nat. Biotechnol. **26**, 1135 (2008).

Skolnick, J. and M. Fixman, "Electrostatic persistence length of a wormlike polyelectrolyte", Macromolecules **10**, 944 (1977).

Stigter, D., "Interactions of highly charged colloidal cylinders with applications to double-stranded DNA", Biopolymers **16**, 1435 (1977).

Strychalski, E. A., S. L. Levy, and H. G. Craighead, "Diffusion of DNA in nanoslits", English, Macromolecules **41**, 7716 (2008).

Su, T, S. K. Das, M. Xiao, and P. K. Purohit, "Transition between two regimes describing internal fluctuation of DNA in a nanochannel", PLoS ONE **6**, e16890 (2011).

Tang, J., S. L. Levy, D. W. Trahan, H. G. Craighead, and P. S. Doyle, "Revisiting the conformation and dynamics of DNA in slitlike confinement", Macromolecules **43**, 7368 (2010).

Tegenfeldt, J. O., C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin, "The dynamics of genomic-length DNA molecules in 100-nm channels", Proc. Natl. Acad. Sci. USA **101**, 10979 (2004).

Thamdrup, L. H., A. Klukowska, and A. Kristensen, "Stretching DNA in polymer nanochannels fabricated by thermal imprint in PMMA", Nanotechnology **19**, 125301 (2008).

The Cost of Sequencing a Human Genome (2016), `https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/`.

Tree, D. R., A. Muralidhar, P. S. Doyle, and K. D. Dorfman, "Is DNA a good model polymer?", Macromolecules **46**, 8369 (2013).

Tree, D. R., Y. Wang, and K. D. Dorfman, "Mobility of a semiflexible chain confined in a nanochannel", Phys. Rev. Lett. **108**, 228105 (2012).

Tree, D. R., Y. Wang, and K. D. Dorfman, "Extension of DNA in a nanochannel as a rod-to-coil transition", Phys. Rev. Lett. **110**, 208103 (2013).

Tree, D. R., Y. Wang, and K. D. Dorfman, "Modeling the relaxation time of DNA confined in a nanochannel", Biomicrofluidics **7**, 054118 (2013).

Tree, D. R., W. F. Reinhart, and K. D. Dorfman, "The Odijk regime in slits", Macromolecules **47**, 3672 (2014).

Utko, P., F. Persson, A. Kristensen, and N. B. Larsen, "Injection molded nanofluidic chips: Fabrication method and functional tests using single-molecule DNA experiments", Lab Chip **11**, 303 (2011).

Van Kan, J. A., C. Zhang, P. Perumal Malar, and J. R. C. van der Maarel, "High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies", Biomicrofluidics **6**, 036502 (2012).

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al., "The sequence of the human genome", science **291**, 1304 (2001).

Wang, J. and H. Gao, "A generalized bead-rod model for brownian dynamics simulations of wormlike chains under strong confinement", J. Chem. Phys. **123**, 084906 (2005).

Wang, Y., D. R. Tree, and K. D. Dorfman, "Simulation of DNA extension in nanochannels", Macromolecules **44**, 6594 (2011).

Welch, R. L., R. Sladek, K. Dewar, and W. W. Reisner, "Denaturation mapping of saccharomyces cerevisiae", Lab Chip **12**, 3314 (2012).

Werner, E. and B. Mehlig, "Confined polymers in the extended de Gennes regime", Phys. Rev. E **90**, 062602 (2014).

Werner, E and B Mehlig, "Scaling regimes of a semiflexible polymer in a rectangular channel", Phys. Rev. E **91**, 050601 (2015).

Werner, E., F. Persson, F. Westerlund, J. O. Tegenfeldt, and B. Mehlig, "Orientational correlations in confined DNA", Phys. Rev. E **75**, 041802 (2012).

Werner, E, G. Cheong, D Gupta, K. Dorfman, and B Mehlig, "One-parameter theory for DNA extension in a nanochannel", arXiv preprint **1705.04619** (2017).

West, D. H. D., "Updating mean and variance estimates: an improved method", Commun. ACM **22**, 532 (1979).

Wetterstrand, K. A., *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*, `www.genome.gov/sequencingcostsdata`, Accessed: 08-28-2017.

Wüest, R, P Strasser, F Robin, D Erni, and H Jäckel, "Fabrication of a hard mask for inp based photonic crystals: increasing the plasma-etch selectivity of poly (methyl methacrylate) versus SiO2 and SiNx", J. Vac. Sci. Technol. B **23**, 3197 (2005).

Yang, Y., T. W. Burkhardt, and G. Gompper, "Free energy and extension of a semiflexible polymer in cylindrical confining geometries", Phys. Rev. E **76**, 011804 (2007).

Zhang, C., F. Zhang, J. A. van Kan, and J. R. C. van der Maarel, "Effects of electrostatic screening on the conformation of single DNA molecules confined in a nanochannel", J. Chem. Phys. **128**, 225109 (2008).

Zheng, G. X., B. T. Lau, M. Schnall-Levin, M. Jarosz, J. M. Bell, C. M. Hindson, S. Kyriazopoulou-Panagiotopoulou, D. A. Masquelier, L. Merrill, J. M. Terry, et al., "Haplotyping germline and cancer genomes using high-throughput linked-read sequencing", Nat. Biotech. **34**, 303 (2016).

Appendix A

Supporting Information to Chapter 3

A.1 SEM images of channel cross sections in Chapter 3.

Figure A.1 contains scanning electron microscopy (SEM) images of the channel cross-section obtained using a device from the same wafer as the device used for the experiments. Prior to SEM, the fused silica wafer was coated in a thin layer of gold to improve contrast.

Table A.1 contains a list of the target nanochannel widths and the actual widths in the final device. All simulations were performed using the actual width of the channel and the 97 nm depth obtained by atomic force microscopy (AFM).

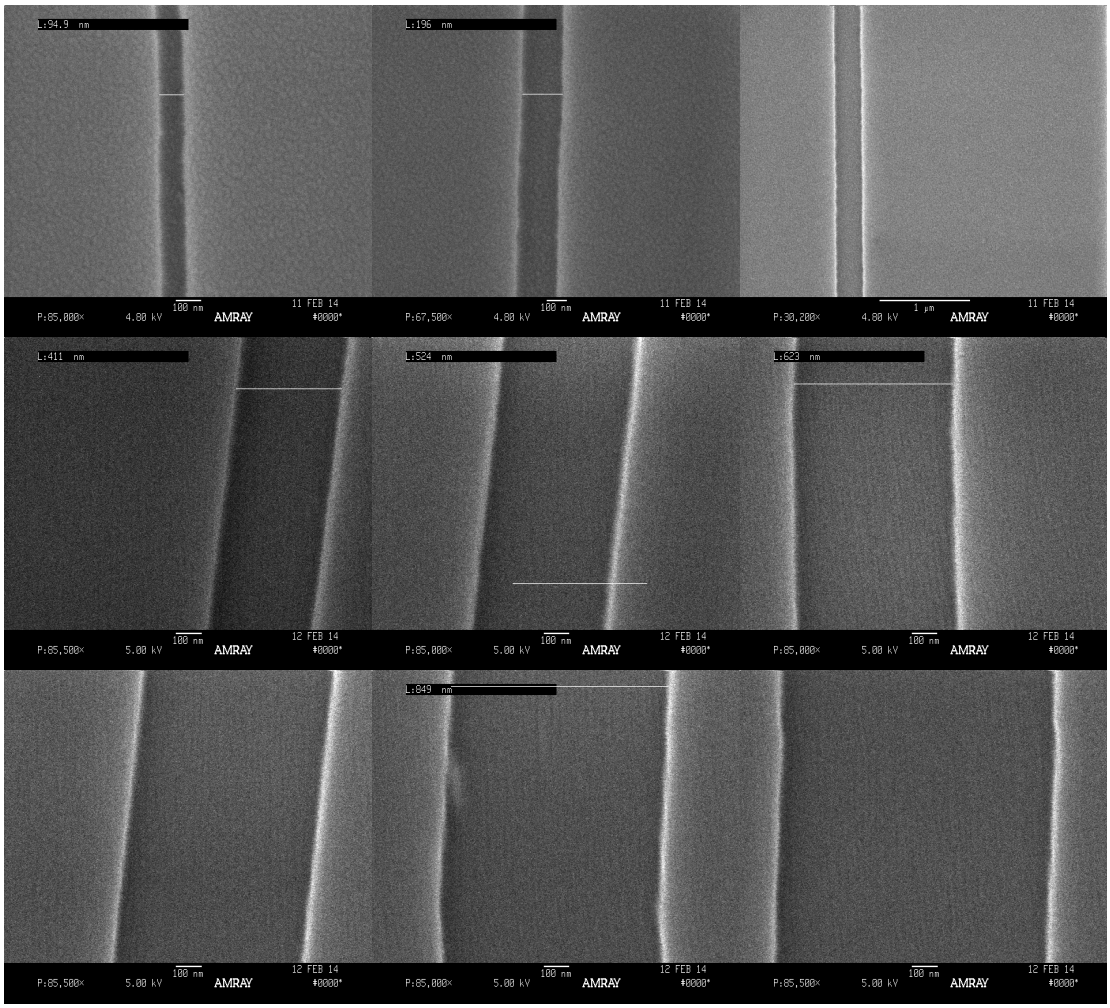

Figure A.1. SEM images of the channel cross sections.

A.2 Role of molecular weight

In the present study, we chose to use a relatively low ionic strength buffer to facilitate image analysis, as the stretching of the DNA in these channels is greatly increased by lowering the ionic strength (Reisner et al., 2007; Jo et al., 2007) to achieve a persistence length of $l_p = 68.7$ nm. The increased stretching occurs be-

Target size (nm)	Actual size (nm)
100	96
200	198
300	297
400	402
500	513
600	618
700	728
800	861
1000	1076

Table A.1: Target sizes and final width of the nanochannels after electron-beam patterning and reactive ion etching. The actual sizes are obtained from three randomly chosen channels that are defect-free.

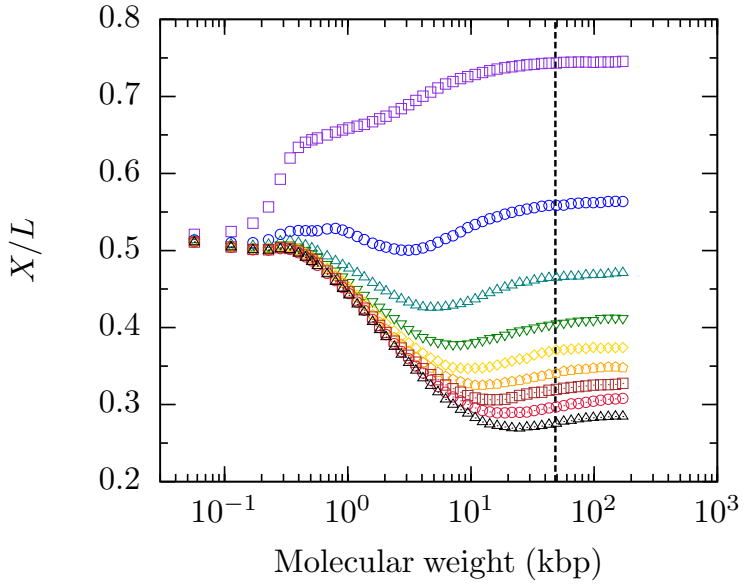

Figure A.2. Simulation results for the fractional extension of undyed DNA in a 7.18 mM ionic strength buffer as a function of molecular weight for rectangular channels of fixed depth (D_1 = 100 nm) and variable width, D_2. Violet □ = 100 nm, and black △ = 1000 nm. The dashed line denotes the molecular weight of λ-DNA.

cause the relevant dimensionless parameter, D_{Av}/l_p, decreases with ionic strength at a fixed value of D_{Av}. The tradeoff for this strategy is that the relevant dimensionless parameter describing the contour length of the chain, L/l_p, also decreases by 23% relative to its value in a high ionic strength buffer.

Before engaging in our analysis, it seems prudent to confirm that λ-DNA is long enough to provide meaningful extension data. We thus performed PERM simulations for all of the rectangular channel sizes in our experiments out to approximately three times the length of λ-DNA. The mean extensions are shown in Figure A.2, with the vertical dashed line indicating the position of λ-DNA. The trends in molecular weight are identical to previous work (Muralidhar et al., 2014b), and we refer the reader to the latter reference for a detailed discussion.

For our purposes here, we simply note that, although we have reduced the number of statistical segments by lowering the ionic strength, λ-DNA should be close to the asymptotic regime even out to $D_{\mathrm{Av}} = 315$ nm, which is the largest channel size ($D_2 = 1073$ nm) used in our experiments. The results obtained in Figure A.2 correspond to undyed DNA. However, if we further assume that the intercalating dye increases the backbone length L while leaving the persistence length l_p and effective width w unchanged, then dyed λ-DNA will also be in the long chain limit.

The reason why λ-DNA is still sufficiently long, even with the reduction in the number of persistence lengths, is that the effective width is 18.7 nm, a more than 4-fold increase over the 4.6 nm effective width in a high ionic strength buffer. The larger effective width leads to increased excluded volume interactions, thereby reducing the number of persistence lengths required to form a self-avoiding blob. Another way to think about this result is that, somewhat counter-intuitively, polyelectrolytes like DNA become more flexible as the ionic strength decreases (Tree et al., 2013a); while the persistence length increases and the DNA becomes stiffer, the ratio of the persistence length to the effective width decreases with decreasing ionic strength so these larger statistical segments interact more easily. In the present circumstances, the ratio l_p/w has been reduced from its high ionic strength value of 11.5 to a much more isotropic value of 3.68. In a coarse-grained sense, the flexibility of a polymer is characterized by the monomer anisotropy parameter l_p/w, not by the absolute value of the persistence length.

A.3 Tabulated data for Figure 3.4 of Chapter 3

D_{Av}/l_p	X/L_{Exp}	$X/L_{\mathrm{Sim\ (Rect)}}$	$X/L_{\mathrm{Sim\ (Sq)}}$
1.40	0.81	0.74	0.75
2.01	0.50	0.56	0.55
2.46	0.42	0.46	0.45
2.87	0.41	0.41	0.39
3.24	0.37	0.37	0.35
3.56	0.33	0.34	0.32
3.86	0.32	0.32	0.30
4.20	0.29	0.30	0.28
4.70	0.27	0.28	0.25

Table A.2: Tabulated data for Figure 3.4 of Chapter 3.

A.4 Power law fit for the fluctuations in rectangular channels

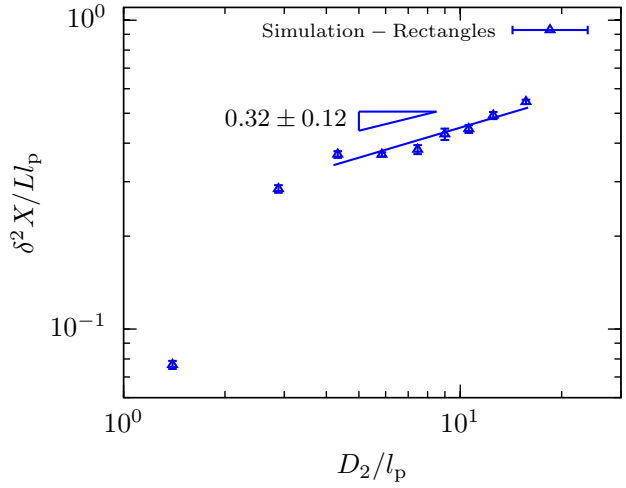

Figure A.3. Power law fit to the simulation data in rectangular channels from Figure 3.5 of Chapter 3. Only channels where D_2 is expected to provide de Gennes-like constraints are included in the fit.

A.5 Increased extension due to intercalation

To determine the increased extension due to intercalation, we compared Tegenfeldt et al. (2004) previously reported data for the extension of λ-DNA concatemers in 100 nm × 200 nm channels to PERM simulation results. Figure A.4 shows the results of the simulations using a persistence length of $l_p = 59.4$ nm and an effective width $w = 11$ nm, which correspond to the ionic strength, I = 20.8 mM, for the 0.5× TBE buffer (Hsieh et al., 2008) with 100 mM DL dithiothreitol (DTT) used

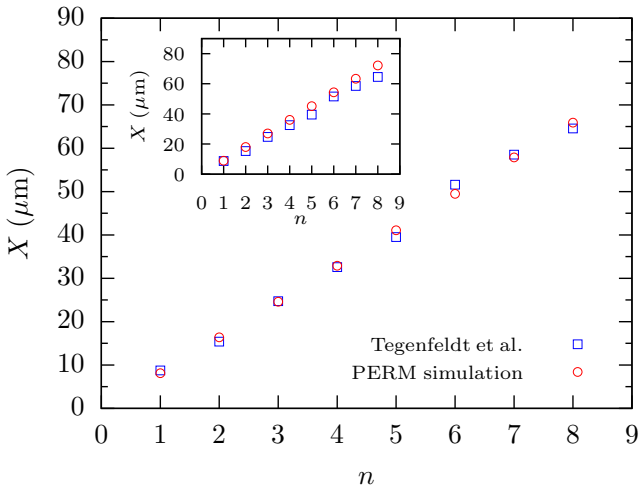

Figure A.4. Comparison of simulation results with experimental data from Tegenfeldt et al. (2004) for concatemers of λ-DNA ($n = 1, 2, \ldots, 8$) confined in a 100 nm wide and 200 nm deep channel. The plot shows good agreement for a dye-modified DNA contour length of $L = 20.1$ μm per λ-DNA concatamer. The inset shows simulation results for $L = 22$ μm, the contour length reported by Tegenfeldt et al. (2004), The simulation input parameters $l_p = 59.4$ nm and $w = 11.0$ nm were obtained from Equation 3.5 and Stigter's theory (Stigter, 1977), respectively, for the reported buffer conditions (Tegenfeldt et al., 2004) and 100 mM DL dithiothreitol (DTT) used in their experiments.

in these experiments (Tegenfeldt et al., 2004). We found good agreement between the simulations and these experiments using a dyed contour length of $L = 20.1$ μm per λ-DNA concatemer. The inset shows that the alternate choice of $L = 22$ μm, as suggested by Tegenfeldt et al. (2004), leads to systematic deviations between the simulations and experiments. Figure A.4 also shows the substantial utility of making these measurements as a function of molecular weight; while the difference between the simulations and experiments for λ-DNA ($n = 1$) is relatively small, the effect of varying the parameter L becomes apparent as the molecular weight increases because we know that $X \sim L$ for sufficiently long chains.

Appendix B

Supporting Information to Chapter 4

B.1 Experimental Methods and Results

B.1.1 Device Fabrication

The devices were fabricated using the same approach as our previous publication (Gupta et al., 2014). The only difference lies in the channel dimensions chosen. The devices used here consist of two 1 μm deep and 50 μm wide microchannels bridged by an array of nanofunnels. The nanofunnels have a target depth D_1 = 300 nm. The funnels are constructed stepwise with sections 45 μm long having constant width dimension D_2. The width of each section ranges from D_2 = 350 nm to 750 nm in increments of 50 nm.

To get a more accurate estimate of the channel sizes, one non-bonded device

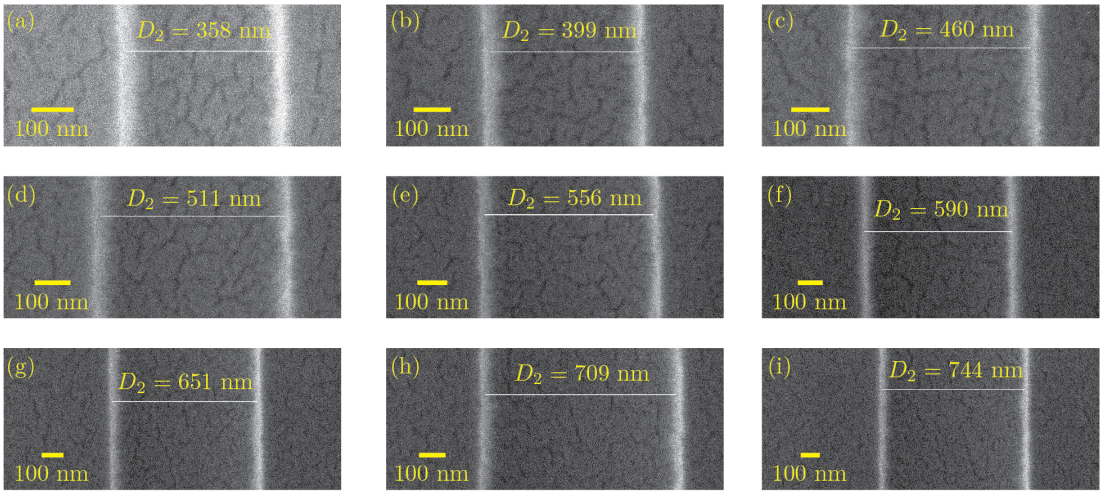

Figure B.1. SEM images of the channel widths. Note that the scale bars are different in each image due to different magnification levels. The value of D_2 corresponds to the length indicated by the solid white line.

from the wafer was characterized using scanning electron microscopy (SEM). Figure B.1 contains SEM images of the channel widths. Table 1 contains a list of the target nanochannel widths and the actual widths in the device. The depth of the channels was determined to be 294 nm using atomic force microscopy (AFM) in a 2 μm wide region at the exit of the nanofunnel.

B.1.2 DNA Preparation

The T4 GT7-DNA molecules (166 kbp, Nippon Gene) were dissolved in 10 mM Tris-base with 0.1% w/v PVP-40 (Polyvinylpyrrolidone, Sigma-Aldrich) to suppress DNA-wall interactions. The DNA were labeled with YOYO-1 fluorescent dye (Invitrogen) at a 1:10 ratio (dye molecule/base pairs). The solution was kept at room temperature for an hour and then heated to 50°C for 12 hours to ensure uniform labeling of all the molecules (Carlsson et al., 1995; Nyberg et

Target size (nm)	Actual size (nm)
350	352
400	401
450	453
500	505
550	554
600	600
650	652
700	707
750	758

Table B.1: Target sizes and final width D_2 of the nanochannels after electron-beam patterning and reactive ion etching. The actual sizes are obtained at least 3 measurements at different locations within a region of constant width. To obtain an estimate of the measurement uncertainty, we made six measurements in the smallest channel width and obtained a standard deviation of 6 nm.

al., 2013). In addition, the anti-photobleaching agent β-mercaptoethanol (BME, Sigma-Aldrich, 6% v/v) was added to the solution immediately before the start of the experiments. The ionic strength for this solution was determined previously to be 7.18 mM (Gupta et al., 2014).

The relatively high molecular weight of T4 DNA, compared to λ DNA, is critical to the success of our experiments. In our previous studies using λ DNA in smaller channels (Gupta et al., 2014), we found that the fractional extension for $D_{\text{eff}} = 281$ nm was as low as $X/L = 0.25$. The latter channel size is close to the smallest channel size in the current study. For the small fractional extensions characterizing the extended de Gennes regime, we need to use long DNA molecules so that their extension is much larger than the measurement uncertainty from the microscope optics.

B.1.3 Experimental Procedure

Molecule images were captured using an optical setup consisting of an EMCCD camera (Photometerics, Cascade II: 512) mounted on an inverted epifluorescence microscope (Olympus IX73) with a 100x, 1.4 NA oil immersion objective using laser excitation (Coherent OBIS 473 nm) as a light source. The stroboscopic exposure of the sample was controlled using a shutter controller (Thorlabs), which was triggered using the output of the camera. In-plane drift of the device was controlled using an automated microscope stage (Applied Scientific Instrumentation, accuracy 0.25 μm/mm). Lastly, the camera and the stage were controlled using Micro-Manager.

B.1.4 Image Analysis

Experimental image data analysis used a custom-written Matlab program (Reisner et al., 2005). The extension of each molecule was measured from its kymograph by fitting the intensity profile inside the nanochannel to a linear combination of two error functions, which is a convolution of a Gaussian point-spread function (PSF) and a box function. A best common value for the full width at half maximum (FWHM) is used for all the frames of a single molecule at a particular channel size. The maximum uncertainty in the measurement of the extension, σ_{u}, has been estimated to be 0.8 μm from the FWHM of the PSF part of the function. This is a very conservative estimate of the uncertainty, as the fitting algorithm is likely to be more accurate that the PSF.

B.1.5 Relaxation Time

In addition to time-lapse data, we also took continuous time series data in the smallest five channel sizes to determine the longest relaxation time of the molecule. A single T4 DNA molecule was recorded for total of 40 s at an exposure time of 40 ms and an interval of 40 ms between exposures.

To compute the longest relaxation time for the span data, we first computed the autocorrelation function for the instantaneous chain span,

$$G_k(\tau) = \frac{\langle \delta X_k(t)\, \delta X_k(t+\tau) \rangle_t}{\langle (\delta X_k(t))^2 \rangle_t} \tag{B.1}$$

using all of the extension data, $X_k(t)$ for a particular DNA molecule k. Here,

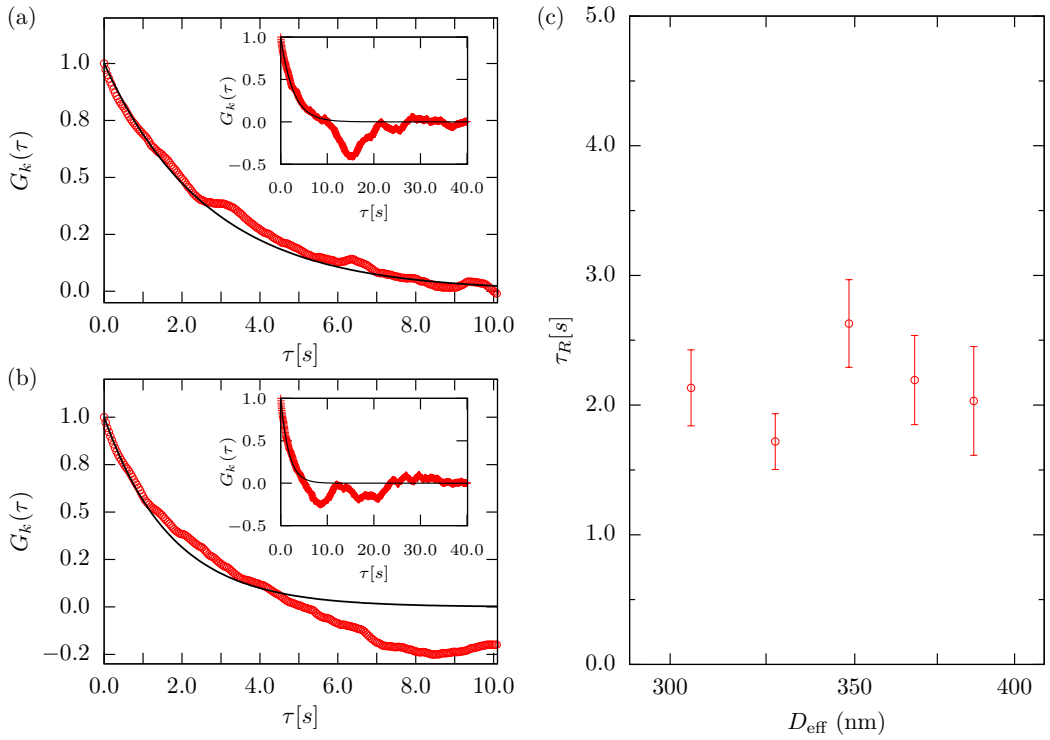

Figure B.2. Example of an exponential fit (black line) to the autocorrelation function $G_k(\tau)$ (red 5) versus the time lag τ for one T4 DNA molecule (a) in the 350 nm wide channel and (b) in the 550 nm wide channel. (c) The longest relaxation time, τ_{R} of molecule span as a function of effective channel size, D_{eff}.

$\delta X_k(t) = X_k(t) - \langle X_k \rangle_t$ where $\langle \cdots \rangle_t$ denotes a time average. The longest relaxation time, τ_k, was obtained by fitting the short-time decay of the autocorrelation function with an exponential function, where the τ_k is the time constant for the exponential. The fit was constrained such that $G_k(0) = 1$. Two examples of this fit are shown in Figure B.2a and B.2b for $D_2 = 350$ nm and $D_2 = 550$ nm respectively. The average relaxation time, τ_R, was calculated from the average of τ_k obtained from a minimum of 10 molecules in each channel size.

Figure B.2c shows that the maximum relaxation time is close to 2.5 s and in-

dependent of the channel size. For making measurements of the span, the stroboscopic time interval of 5 s means that each measurement should be uncorrelated. Note that the magnitude of the relaxation time for T4 DNA (166 kbp) here is roughly five times smaller than what would be expected according to scaling laws for the contour length in the extended de Gennes regime (Reisner et al., 2012) from a previous measurement of τ_R for λ-DNA (48.5 kbp) in a similar channel size (300 nm, 0.5x TBE buffer, $\tau_R = 0.8$ s) (Reisner et al., 2005). The source of this discrepancy is not obvious, although we note that the present measurements were made in rectangular channels and the previous measurements (Reisner et al., 2005) were made in (approximately) square channels. The details of the hydrodynamics of DNA confined in rectangular channels remains an open question.

B.1.6 Molecule Selection Criterion

We measured the same molecule twice in the channel with width, $D_2 = 350$ nm, first at the start of the experiment and again at the end of the experiment. For these two measurements for a molecule i, an unpaired two tail t-test was conducted at a 5% significance level. Only those molecules that statistically belonged to the same population were accepted. We imaged a total of 40 molecules but only 29 of them passed this test. Figure B.3 compares the average span, X_i for the molecules that passed the selection test. Rejected molecules had obvious shortening during the experiment, typically around 4 μm.

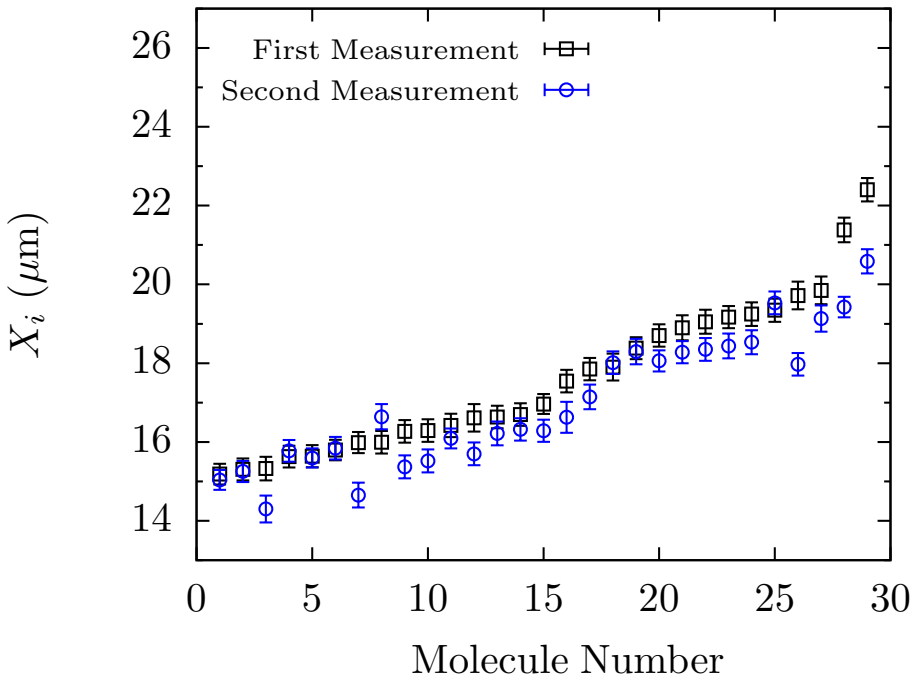

Figure B.3. The measurement of average span, X_i for the accepted molecules in the channel width, $D_2 = 350$ nm at the start and at the end of the experiment. The error bars represent the standard error of the mean.

B.1.7 Determining the DNA Length

Figure B.4 is equivalent of Figure 4.2 but with the matching to simulation data for all of the bins. The molecules were assembled into 7 bins according to their average span, X_i, in the channel with $D_2 = 350$ nm. Next, the average span in a bin,

$$X_{\text{bin}} = (1/w_{\text{bin}}) \sum_{i=1}^{w_{\text{bin}}} X_i \tag{B.2}$$

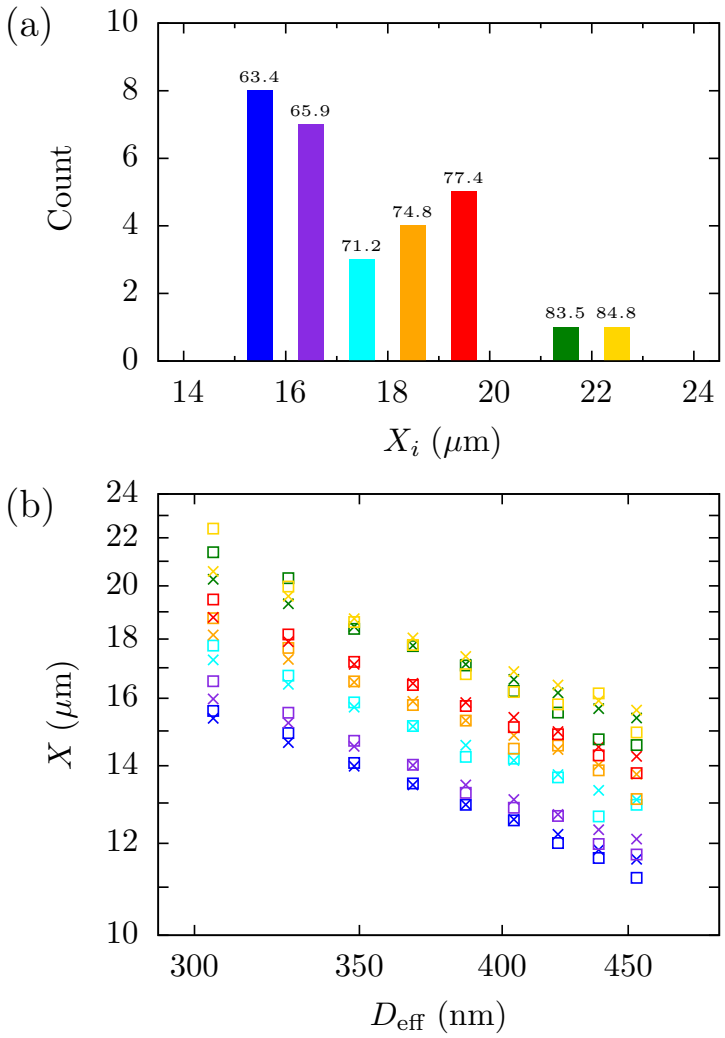

Figure B.4. (a) Probability distribution of average span, X_i, for individual T4 DNA molecules from 40 uncorrelated measurements of their span for $D_2 = 350$ nm. The number above each bar is the value of the contour length, $L_{\rm bin}$ (μm), obtained by comparison of the span to PERM simulations. (b) Plot of average value of span, $X_{\rm bin}$, for molecules that reside in a particular bin as a function of effective channel size, $D_{\rm eff}$ ($\square$ symbol). To determine the corresponding contour length, we fit these curves to PERM simulations of the chain span ($\times$ symbol) using the contour length as a free parameter. The color code for each symbol is same as that of the corresponding bin in panel (a).

was calculated for $w_{\rm bin}$ molecules in that bin. To estimate the contour length of the molecules in each bin, $L_{\rm bin}$, we minimized the sum of squared error,

$$\epsilon = \sum_{D_2=350}^{750} [X_{\rm sim}(D_2, L_{\rm sim}) - X_{\rm bin}(D_2)]^2 \tag{B.3}$$

of the average extension in that bin, $X_{\rm bin}$, to the PERM simulation span data, $X_{\rm sim}$, at the contour length $L_{\rm sim}$ for all channel widths, $D_2 = 350, 400, \ldots, 750$ nm. The value of $L^*_{\rm sim}$ that minimizes ϵ is set to be the value of $L_{\rm bin}$.

The contour length increased on average by 24% from the bare contour length of T4 DNA (67.4 μm). Additionally, the most probable increase in the contour length is 14%, which is close to our previous calculation for λ-DNA (13%) at the same dye ratio (Gupta et al., 2014).

The average extension, X, and variance in extension, $\delta^2 X$, were calculated by rescaling the measurement from molecule i by their contour length $L_i = L_{\rm bin}$ according to

$$\frac{X}{L} = \frac{1}{N}\sum_{i=1}^{N}\frac{X_i}{L_i} \tag{B.4}$$

$$\frac{\delta^2 X}{L l_p} = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta^2 X_i}{L_i l_p} \tag{B.5}$$

with $N = 29$ molecules in each channel size.

B.1.8 Error Analysis

In our previous work (Gupta et al., 2014), we identified the major sources of uncertainty in measurement of $j^{\rm th}$ span of molecule i, $X_i^{(j)}$ as (i) the fitting error

of the intensity profile to a double error function and (ii) non-uniform contour length of the molecule. Since we have systematically removed the second source of error, we only use the first source of error for further propagation. The maximum value of this error, $\delta X_i^{(j)} = \sigma_u$ has already been estimated to be 0.8 μm in Section B.1.4. Note that this is a very conservative estimate for the error. The error bars reported for the three relevant measurements in this letter (extension, variance extension and apparent exponent) are the probable error estimate, which combines the propagated measurement uncertainty (δ_1) and the standard error of the mean from sampling errors in expectation values from different molecules (δ_2). The probable error is computed as

$$\delta = \sqrt{\delta_1^2 + \delta_2^2} \tag{B.6}$$

B.1.8.1 Extension and Variance in Extension

The standard error of the mean for sampling error from Equation B.4 will be

$$\delta(X/L)_{SE} = \sqrt{\frac{\delta^2(X_i/L_i)}{N}} \tag{B.7}$$

To propagate the uncertainty in independent variable x,y,$\cdots$ for a dependent variable f, we use (Ku, 1966)

$$\delta f(x, y, \cdots) = \sqrt{\left(\frac{\delta f}{\delta x}\right)^2 \delta x^2 + \left(\frac{\delta f}{\delta y}\right)^2 \delta y^2 + \cdots} \tag{B.8}$$

The measurement uncertainty in X_i would then be $(\delta X_i)_u = (\sigma_u/\sqrt{n})$, which gives the measurement uncertainty in X/L,

$$\delta(X/L)_u = \frac{\sigma_u}{\sqrt{nN}}\sqrt{\frac{\sum_{i=1}^{N} 1/L_i^2}{N}} \tag{B.9}$$

By combining Equations B.7 and B.9, the total probable error in X/L is given by

$$\delta(X/L) = \sqrt{\delta(X/L)_{SE}^2 + \delta(X/L)_u^2} \tag{B.10}$$

The error in variance of span, $\delta(\delta^2 X/Ll_p)$ has been calculated using the same procedure. The uncertainty (δ_1) from propagation of a constant conservative estimate of the error in the point spread function (σ_u) is 63% of the total error. The uncertainty in the exponent at 95% confidence arises primarily from the scatter in the mean values of the data, with the uncertainty in the individual measurements making a negligible contribution.

B.1.8.2 Apparent Power Law Exponent

The apparent power law exponent for a single molecule i, α_i, is calculated by

$$\alpha_i = \frac{d \ln X_i}{d \ln D_{eff}} \tag{B.11}$$

The error bounds on α_i have been estimated by using 95% confidence intervals. The average exponent, α has been calculated using

$$\alpha = (1/N) \sum_{i=1}^{N} \alpha_i \tag{B.12}$$

Finally, the total error in α has been estimated by combining the standard error of the mean similar to Equation B.7 and propagating the uncertainty in α_i (using a combination of Equations B.8 and B.12).

Note that this method of determination of exponent and uncertainty was done in addition to the aforementioned calculation of exponent and uncertainty by using the average value of the variance along with its error bars.

B.2 Evaluation of Quantitative Agreement with Theory

In Fig. 3 of the main text, we see excellent quantitative agreement (to within 8%) between our experimental data and the theory of Werner and Mehlig (2014) for the variance in chain extension. However, in Fig. 4 of the main text, the agreement between the experiments and the theory (Werner and Mehlig, 2014) for the extension is not as good, although the experiments are relatively close to other simulation data (Dai et al., 2014). Note that Werner and Mehlig (2014) did point out potential issues with the prefactor for the extension in their theory. In the context of our experiments, a subtle issue in the comparison of theory with experiment is our approach to determine the absolute value of L_{bin} by matching

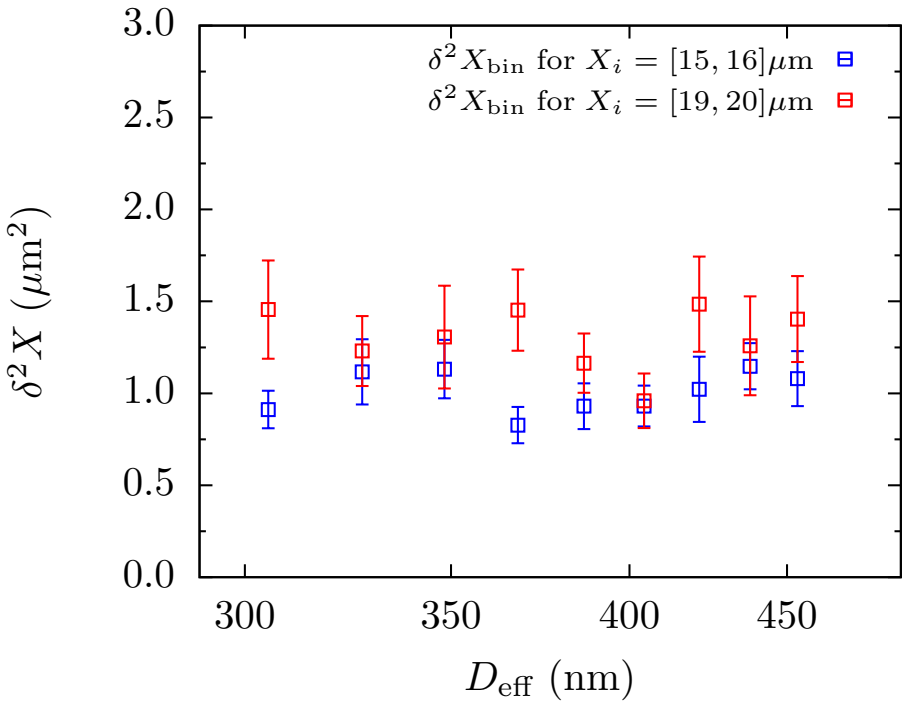

Figure B.5. Plot of the average value of the variance in extension, $\delta^2 X_{\mathrm{bin}}$, for molecules that reside in two different bins (X_i = [15,16] μm in blue and X_i = [19,20] μm in red) as a function of effective channel size (□ symbols). The error bars are the combination of the standard error of the mean and measurement uncertainty.

the extension data in a particular bin to PERM simulation data. As mentioned in the main text, we posit that this approach will not affect our conclusions about the scaling exponents, but it could affect the prefactor.

It is clear from the experimental data that each bin exhibits similar slopes for the extension. As seen in Figure B.5, the variance data in two representative bins are also parallel lines with almost zero slope. As a result, we have confidence that our conclusions regarding the scaling exponents are robust to the method we used to remove the effects of molecular weight dispersity. Indeed, if we only want to know the values of the scaling exponents, there is no need to collapse the data as

a function of molecular weight — we just need to compute the scaling exponents for each individual molecule. We did this alternate analysis for the variance data and found a scaling $\delta^2 X/Ll_p \sim D_{\text{eff}}^{0.05\pm0.27}$ at 95% confidence.

If we want to assess the value of the prefactors, however, we need to know the contour length of the chain. Our method may introduce a systematic error in the value of L_{bin} through (i) the use of theory to determine the persistence length and the effective width, rather than a separate experimental determination and (ii) the assumption that the DNA-wall depletion length δ is the same as the effective width w. Although the parallel lines in Figure B.4 can be collapsed by assuming a relative molecular weight between different bins, determining the absolute value of that molecular weight requires the matching to PERM experiments. So long as the experimental data and simulations have roughly the same scaling exponent (which is the case, especially given the narrow range of experimentally accessible channel sizes in the extended de Gennes regime), the absolute value of the molecular weight can be obtained by shifting the curves in the vertical direction (by rescaling L). However, if there is an uncertainty in δ, which seems reasonable, then our values of L_{bin} have a systematic error. As a result, when we plot X/L or $\delta^2 X/Ll_p$, the systematic error will propagate to the prefactor for the scaling. Note that changing δ will change the scaling exponents, but realistic changes in δ lead to changes in the scaling exponent that are of the order of the measurement uncertainty.

B.3 Simulation Methods and Results

B.3.1 Model and Methodology

We use a discrete wormlike chain model (DWLC) of touching beads each with a hard-core effective width of size w. The stiffness of the backbone is accounted for by a bending potential U_{bend} imposed on every contiguous triplet of beads forming an angle θ given by

$$\beta U_{\text{bend}} = \kappa(1 - \cos\theta), \tag{B.13}$$

where κ is the bending constant and $\beta = 1/kT$, the inverse thermal energy. This results in a persistence length of the form (Muralidhar and Dorfman, 2015)

$$\frac{l_p}{w} = \frac{\kappa}{\kappa - \coth(\kappa) + 1}. \tag{B.14}$$

The persistence length and the effective hard core width of DNA for the given buffer is calculated from Dobrynin's theory (Dobrynin, 2006) and Stigter's method (Stigter, 1977) respectively as was done in our previous paper (Gupta et al., 2014). As in our previous work (Gupta et al., 2014), we made a simplifying assumption that the depletion width for DNA-wall interactions is same as that for DNA-DNA interactions. Therefore, for a channel with cross-sectional dimensions D_1 and D_2, the effective dimensions that the centre of the backbone can explore are $D_1 - w$ and $D_2 - w$.

With the aforementioned model, we ran Pruned-Enriched Rosenbluth Method (PERM) simulations for a total of 12 million tours for a given channel size. Our

simulations were divided into 6 independent replicas of 2 million tours each. We were thus able to calculate thermodynamic properties such as the mean extension and the variance of extension of the confined moleucle. For more details on our PERM algorithm and methodology, we direct the readers to previous papers from our group (Tree et al., 2013a; Tree et al., 2014; Muralidhar et al., 2014b; Muralidhar et al., 2014a; Muralidhar and Dorfman, 2015).

B.3.2 Molecular Weight Dependence

In a trio of figures, we plot how the average fractional extension (Figure B.6a), the variance in average extension (Figure B.6b), and the apparent exponent for the mean span (Figure B.7) vary as a function of molecular weight of unstained DNA. The span data in Figure B.6a also support the claim in the main text that our chains are long enough to reach the asymptotic scaling $X \sim L$. Note that the Figure B.7 data has been obtained using only 1 million tours for the PERM simulations.

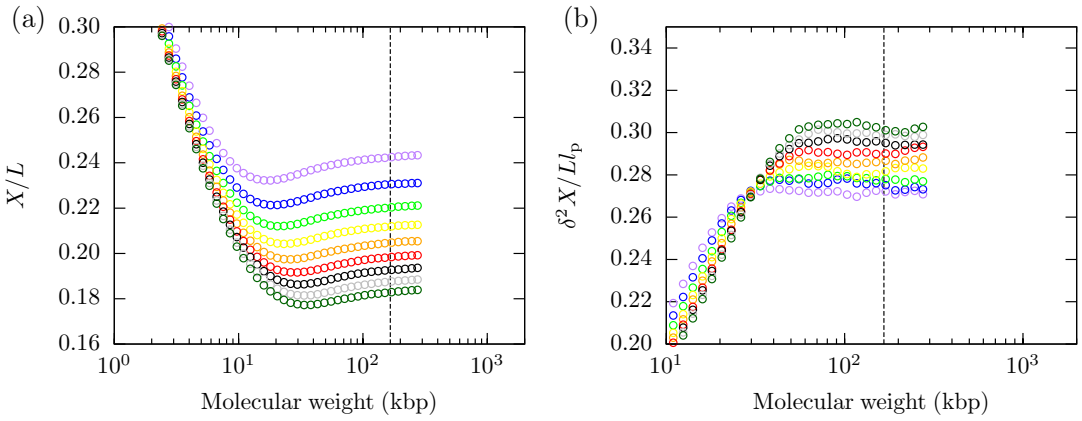

Figure B.6. Semilogarithmic plots of (a) the average fractional span, X/L, and (b) variance in the average extension, $\delta^2 X/Ll_p$, obtained from PERM simulations for a rectangular channel with fixed depth, $D_1 = 300$ nm, and variable width, D_2, as a function of the molecular weight of DNA. D_2 for purple 5 = 350 nm, blue 5 = 400 nm, green 5 = 450 nm, yellow 5 = 500 nm, orange 5 = 550 nm, red 5 = 600 nm, black 5 = 650 nm, gray 5 = 700 nm and dark-green 5 = 750 nm. The vertical black dashed line represents the molecular weight of T4 DNA, and the comparison to PERM data is made in the absence of any extension due to intercalation.

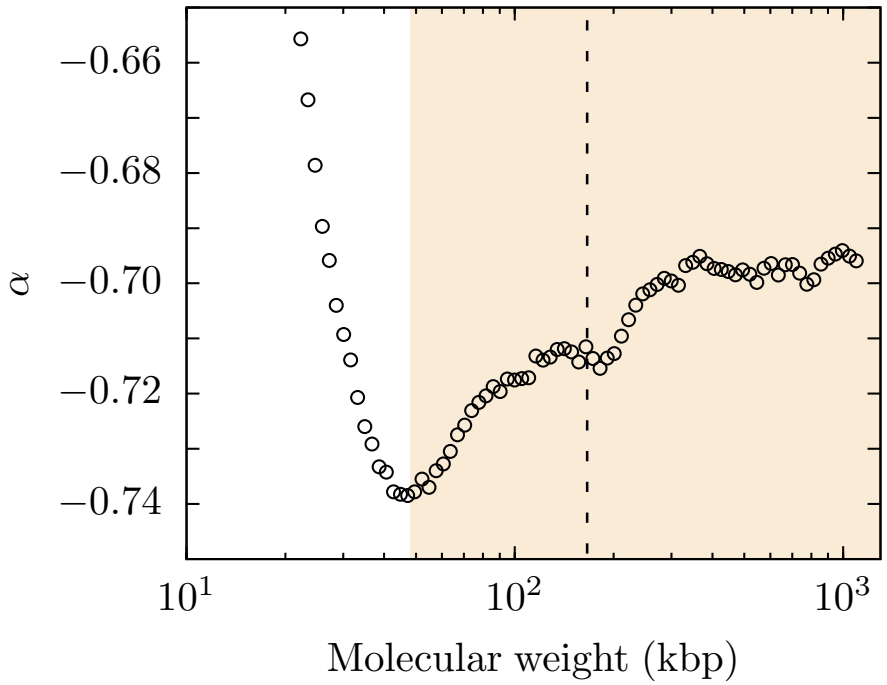

Figure B.7. The apparent power law exponent, α, for the rectangular channels over the full range of the channel size as a function of the molecular weight of DNA using the PERM simulations. The vertical black dashed line represents the molecular weight of T4 DNA, and the comparison to PERM data is made in the absence of any extension due to intercalation. The orange shaded region represents the experimentally relevant molecular weights starting from λ-DNA to about 1 megabase pair.

Appendix C

Fabrication Protocols for Chapter 5

This appendix describes the steps used to build the nanofluidic devices in fused silica substrates used for the experiments in Chapter 5. The fabrication workflow is outlined in Figure C.1. The fused silica substrates (SiO_2) have been used widely for biological applications due to their know surface properties, biocompatibility and low auto-fluorescence. In particular for DNA, the electrical insulating properties of fused silica and negative charges on its surface in aqueous solution (Behrens and Grier, 2001) are essential for carrying out high-throughput equilibrium studies. We also found out that the devices made in fused silica substrates can be reused multiple times.

Much of the fabrication work was accomplished in the Minnesota Nano Center (MNC) at the University of Minnesota. So the recipes described here will work best for the machines at the MNC.

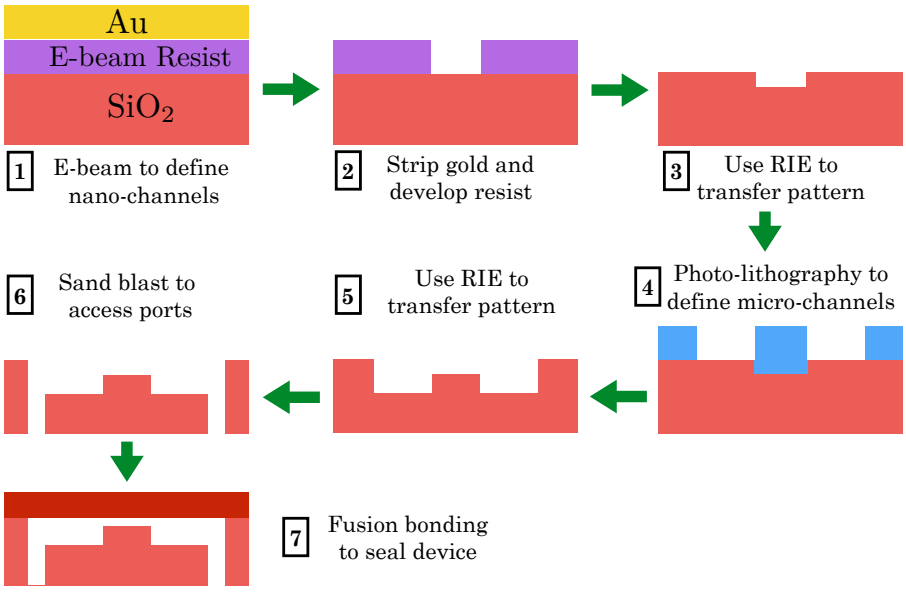

Figure C.1. Workflow used in fabrication of nanofluidic devices.

C.1 Nanochannel fabrication

The nanochannels have been fabricated using e-beam lithography (EBL) followed by transfer of the pattern from the resist to substrate via a reactive ion etching (RIE) step.

1) Design the nanochannel pattern on CAD software. The CAD files should be created in GDS format. Clewin software provided by the MNC has been able to reliably create desired patterns. When creating a nanochannel pattern, ensure that the total size of the pattern is less than 500 μm long in each direction. The EBL machine can only write a maximum distance of 500 μm at a time. Beyond this distance, the patterns will suffer from stitching error.

The device design used here is very similar to the one illustrated in Figure 3.2. I have created an array of 90 nanofunnels at a pitch distance of 3 μm to avoid over exposure of the pattern during the EBL step. Each nanofunnel has four, 100 μm long connected nanochannels with variable widths (100 nm, 150 nm, 200 nm, and 240 nm). Additionally, I have created a post-array block adjacent to the 240 nm wide nanofunnel region to facilitate loading of longer DNA molecules. The post-array design was inspired from the BioNanoGenomics design obtained from their website. It consists of a 270 μm by 20 μm block of post-array with spacing between a post ranging from 0.2 μm near the nanochannel entrance to 1 μm close to the microchannel region. The post-array depth is similar to the depth of the nanochannels as they are created along with the nanochannels. Finally, the design was extended by 25 μm on each side to have an overlap region with the microchannels. This increases the robustness of the design and the device.

2) Design the global alignment marks on CAD software. These marks can be used to align nanochannels with microchannels. The CAD files should again be created in GDS format. These marks can be placed at size greater than 500 μm size, the maximum write distance of EBL at a time, if a 25 μm overlap region between microchannel and nanochannel was already included in the design.

3) Fracture the CAD files using "Layout Beamer" software. Import the GDS file using the default configuration of the "Layout Beamer" software provided by the MNC. Add a layer of "Heal" followed by a layer of "PEC" (proximity error correction) in the default configuration. The "Heal" step fractures the pattern into small trapezoids, which are recognized by the EBL machine during the writing process and the "PEC" step, as name suggests, adjusts the dosage of e-beam intensity according to the density of the patterns. Finally, add an export step to export files in the GPF format. At the end of this fracture process, two GPF files are generated for each GDS file. Review the GPF file using "Cview" to ensure that the files have been fractured correctly. An incorrect fracture will result in overlapping trapezoids. This error is generally caused by high density of patterns and can be removed by iterating the design to achieve optimal pattern density.

4) Create final job file using "Cjob" software. This software allows to choose the electric flux density, current and aperture size of the e-beam for the writing process. Set electric flux density or dosage to 450 $\mu c/cm^2$ for the nanochannel GPF file. Then, chose the 1 nano amp, 300 μm beam. It corresponds to 17 nm spot size of the beam, which should always be at least two times smaller than the smallest pattern size. This is critical for getting patterns of precise dimensions. For global alignment marks, chose dosage of 700 $\mu c/cm^2$ and 10 nano amp, 300

μm beam. Export the final job file after reviewing the write times. The write time for 9 replications of the nanochannel pattern described above was 9 minutes.

5) Acid piranha clean a 4 inch wide, 500 μm thick fused silica wafer. Acid piranha is freshly prepared from 3 parts sulfuric acid and 1 part hydrogen peroxide solution. Leave a fresh wafer from University Wafers in this solution for 20 minutes. Rinse thoroughly three times using the wet bench dump rinser and the rinser on the spin dryer. Ensure the wafer is dry by heating on a hot plate at 180°C for 180 sec.

6) Spin coat e-beam resist. This should give a thin, uniform layer on the wafer. Coat 950-PMMA C4 using spin coater at 4000 RPM with 1000 RPM/sec acceleration for 30 sec. It gives an approximately 500 nm thick layer of the e-beam resist, which is sufficient for the final etch depth of fused silica up to 300 nm.

7) Soft bake. Wait for 1 minute and bake the wafer at 180°C for 5 minutes. This step removes the solvent from the e-beam resist.

8) Coat gold discharge layer. Due to the electrical insulating properties of fused silica, static charges build up inside resist during EBL, compromising the resolution of the nanoscale features. To get around the charging problem, a thin gold discharge layer (~ 20 nm) is coated. Use AJA-II metal sputter system in the PAN cleanroom for 36 sec to coat a thin and uniform gold layer on top of the e-beam resist (~ 18 nm). Even though AJA-II gives a uniform layer crucial for obtaining consistent patterns on the entire wafer, it leaves an uncoated 5 mm wide rim. To cover the edge area, use one cycle of low quality Cressington metal sputter system for 15 sec (~ 5 nm).

9) Expose. Securely place the wafer on the 4 inch wafer holder of the EBL

machine. Before loading the wafer holder into the EBL machine, make sure a) the tilt of the wafer is smaller than 1 μm/mm b) the gold layer on the wafer is within $\pm$50 μm distance of the e-beam focus location and c) the copper clips on the wafer holder are in firm contact with the gold layer to provide a grounding pathway for the residual charges. Both a) and b) checks can be performed on the laser microscope located next to the EBL machine. Check c) has to be done manually. Using the job file created in **step 4** on the EBL machine, transfer the pattern to the e-beam resist on the wafer. After the exposure to the e-beam, the polymer layer undergo chain scission allowing those chains to be dissolved selectively in the developer and produce very high resolution patterns (~ 20 nm).

10) Strip gold layer. Dissolve the exposed wafer in one part of gold etchant, GE6, and one part of water mixture for 60 sec. Rinse in water thoroughly three times, and dry the wafer using a N_2 gun blow dry. The PMMA resist layer breaks down in a basic medium. Therefore, the etchant should be either acidic or neutral.

11) Develop. Isopropanol (IPA) diluted with water dissolves PMMA resist. Prepare a three parts IPA and one part water mixture to develop the wafer for 30 sec. After developing, immediately dissolve the wafer in water to stop developing for 30 sec and use the N_2 gun to blow dry. The total developing time can be best determined by checking the post-array using dark-field microscopy. The spacing between any two pillars should be clear. If required, additional developing can be done with 10 sec long steps until clear features are visible under the microscope.

12) Hard bake at 90°C for 120 sec in a oven. Cool the wafer before etching.

13) Plasma O_2 clean. Use AV etcher 'O2-clean' recipe for 10 sec. It uses O_2 plasma to clean any residual resist left on the developed areas, often called a resist

scum layer.

14) Reactive Ion Etch. Use the 'pjs-oxide' recipe on the AV etcher to etch the nanochannel region. It uses fluoride anions (F^-) to selectively etch the fused silica wafer. This etching method can be transport limited. In this case, large areas with no resist will etch faster than small holes or narrow lines. For example, the global marks region which are wider than 10 μm, etched at the rate of 19.5±1 nm/minute where as the nanochannel region etched at the rate of 15.5±0.5 nm/minute. Measure depth using P16 profilometer or an AFM.

To increase the anisotropy of the etch, use a cyclic process where etching is done for a maximum of 1 minute at a time, interleaved with N_2 flushing steps of 90 sec (Wüest et al., 2005). This method is based on the principle that the PMMA etches slowly at the beginning, but its etch rate increases with increase in the etch duration, if etched continuously. However, the fused silica etches at a constant rate irrespective of the etch cycle duration. This etching strategy led to relatively straight walls, when compared to the slanted walls observed in some previous nanochannel studies (Gupta et al., 2015; Iarko et al., 2015).

To obtain multiple channel sizes of varying aspect ratio, etch different devices to different depths. For example, I have made 9 identical devices on a single wafer with nanofunnels of varying width (100 nm - 240 nm). To achieve different depths, a) allow all the 9 devices to be etched down to approximately 100 nm, b) cover two devices using glass slides, and kapton tape, c) etch the remaining devices for additional 50 nm, and d) repeat step b) and c).

15) Strip resist. Dissolve the wafer in N-Methyl-2-pyrrolidone (NMP) solution for 15 min at 70°C followed by acid piranha clean described in **step 5**.

C.2 Microchannel fabrication

The mask design used for microchannel fabrication is illustrated in Figure 3.2.

1) HMDS vapor treatment. This vapor increases photoresist adhesion. Use it for 3 to 5 minutes.

2) Spin coat photoresist. This should give a thin, uniform layer on the wafer. Choose the photoresist type and spin speed based on the thickness of the resist you want. For example, S1818 at 2000 RPM with acceleration rate of 2000 RPM/sec for 30 sec gives approximately 2 μm thick layer.

3) Soft bake at 115°C for 60 sec on a hot plate for Shipley series photoresists. This step removes the solvent from the photoresist. Different photoresists require different soft bake parameters.

4) Expose. Align the wafer so that the alignment marks from the nanochannel design and the microchannel design overlap. Transfer the microchannel pattern from a mask to the wafer by exposing the photoresist to change its solubility in photoresist developer. The best pattern transfer comes from using the exposure mode 'Soft Contact' on the MA6 or MABA6 aligner with alignment gap of 60 μm. Additionally, placing a silicon wafer underneath a glass wafer helps in improving the visibility of very shallow alignment marks and reduces the required exposure time. To achieve a good contact between silicon and glass wafers, sandwich a small drop of water between them. The exposure time must be tuned to account for the photoresist thickness, the substrate surface reflectivity, and the intensity of the bulb in the mask aligner. For example, I have used 15 sec exposure time for S1818 when a silicon wafer is placed underneath the fused silica wafer. Under-exposed

or over-exposed wafers are only apparent after developing.

5) Develop. Microposit 351 developer dissolves exposed Shipley series photoresists. This developer can be diluted with water to increase the developing time. The wafer is developed once all exposed photoresist is dissolved. This is best determined by watching any large exposed areas on the wafer. Developing time must be tuned based on the photoresist thickness, exposure dose, and developer reactivity. After developing, immediately rinse the wafer with DI water to stop developing.

6) Review. Check the wafer under microscope to ensure that the microchannels are aligned to the nanchannel location.

7) Hard Bake at 120°C for 90 sec on a hot plate for Shipley series photoresists.

8) Reactive Ion Etch. Use 'pjs-oxide' recipe on the AV etcher to etch nanochannel region. It uses fluoride anions (F^-) to selectively etch the fused silica wafer. Do not etch for more than 20 minutes at a time for Shipley series photoresists. They tend to melt with heat generated during etch process and spread onto exposed region.

9) Wafer Dicing. Cover the wafer with a 2 μm thick Shipley series photoresists to cover the exposed region of patterned glass. Prepare the wafer for dicing by sticking the blue saw tape on the non-patterned side only. The adhesive from the tape can be very hard to clean and may adversely impact the fusion bonding step, which requires sticking super-clean surfaces together. Wafers can be precisely diced with the MNC wafer saw.

10) AFM characterization. Characterize depth of individual devices before bonding which is an irreversible process.

11) SEM characterization. Characterize width using a device which will not be bonded. SEM characterization of a glass surface requires a layer of metal to avoid charge build up, which renders the device unusable for DNA stretching. It is difficult to clean metal residues in nanochannel region which will cause DNA to stick when inserted inside such a nanochannel. Coat a layer of gold using Cressington for 45 sec before SEM (~ 15 nm).

C.3 Access ports and fusion bonding

1. Take a device coated with photoresist for protection and mark the microchannel reservoirs using a sharpie.

2. Cover both sides of the device with scotch tape. Cut out 2 mm, square holes in scotch tape using blade on both sides near sharpie mark for microchannel reservoirs.

3. Use a dental sand blaster to create through holes in the device. The sand gun should be held at a 90° angle to the surface and about one inch away to avoid damaging the surface and create holes in an efficient manner. For example, if the gun is loaded with sand, it takes 20 sec to create one hole. Repeat the same procedure for other three holes.

4. Blow dry away all the sand particles as much as you can. Dissolve the tape coated device in acetone solution and sonicate it, if needed. All the tape and the photoresist should come out very easily.

5. Take a fused silica, 170 μm thick cover slip. Cover slips thicker than 170 μm will not work on high numerical aperture, oil objectives which typically have focal distance close to 170 μm.

6. Clean both the patterned glass piece and cover slip in acid piranha followed by base piranha. Acid piranha is freshly prepared 3 parts sulfuric acid and 1 part hydrogen peroxide solution. Leave the glass pieces in the mixture for 20 minutes using a coverslip holder and later wash thoroughly at least three times.

7. For base piranha, heat 5 parts of water to 80° C. Add 1 part of ammonium hydroxide and heat the mixture to 80° C. Then, add 1 part of hydrogen peroxide and heat the mixture to 80° C. Leave the glass pieces in the mixture for 20 minutes using the coverslip holder while heating it continuously to 80° C and later wash thoroughly at least three times.

8. Bond the two glass pieces together by applying pressure in a bonding well and removing all the bubbles near the edges, through-holes and the patterned regions. This will create a temporary bond which can be reversed by placing a small blade between the two surfaces.

9. To permanently bond the glass pieces, heat them at 1000° C for six hours in blue lindberg furnace. Use a 4° C/minute ramp rate while heating and cooling.