

uOttawa

# Nanostructured DNA Barcodes: Paving the Way for Digital Data Storage with Solid-State Nanopore Readout

**Philipp Mensing**

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctor of Philosophy in Physics

Department of Physics
Faculty of Science
University of Ottawa

© Philipp Mensing, Ottawa, Canada, 2026

# Contents

Abstract ..... vi

Acknowledgements ..... viii

Statement of Originality ..... x

Statement of Contributions ..... xii

Publications ..... xii

Conference Presentations ..... xiii

List of Figures ..... xiv

List of Tables ..... xvii

1 Introduction ..... 1

1.1 Nanopore Fundamentals ..... 1

1.1.1 Nanopore Working Principle ..... 1

1.1.2 Types of Nanopores ..... 3

1.1.3 Nanopore Fabrication: Controlled Breakdown ..... 6

1.1.4 Instrumentation ..... 8

1.1.5 Nanopore Noise ..... 10

1.1.6 Nanopore Capture Process ..... 13

1.1.7 Nanopore Event Signatures ..... 18

1.1.8 Nanopore Translocation Control ..... 23

1.2 DNA Structure ..... 24

1.2.1 DNA as a Polymer ..... 24

1.2.2 DNA Origami ..... 27

1.2.3 DNA as a Carrier for Biomolecules ..... 30

1.2.4 DNA as a Data Storage Medium ..... 32

1.2.5 DNA Nanostructure Barcodes ..... 34

1.3 Thesis Overview ..... 36

2 DNA Dynamics and Capture ..... 38

2.1 Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-state Nanopores ..... 39

2.1.1 Abstract ..... 39

2.1.2 Introduction ..... 40

2.1.3 Results ..... 42

2.1.4 Conclusion ..... 56

2.1.5 Methods ..... 57

2.2 Supporting Information: Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-state Nanopores ..... 60

2.2.1 S1. Circular Double-stranded M13 – Random Cutting ..... 60

2.2.2 S2. Filtering Folded Fragmented LdsM13 Translocation Events ..... 61

2.2.3 S3. Folded State Survival Probability ..... 62

2.2.4 S4. Gel Images for Flexible Assemblies ..... 64

2.2.5 S5. 2D Heatmaps of Flexible Assemblies ..... 64

2.2.6 S6. Dependence of Heterostructure Radii on ssDNA Content ..... 65

2.2.7 S7. Effect of Polymer Rigidity on Monomer-Pore Distance ..... 65

2.2.8 S8. Triple Blockage ECD Plots ..... 66

2.2.9 S9. List of Oligo Sequences ..... 66

2.2.10 S10. Additional Event Traces for Flexible End Molecules ..... 72

3 DNA Barcoding Basics ..... 75

3.1 Nanostructured DNA Barcode Decoding with Solid-State Nanopores: Impact of Bit Spacing and Bit Shape ..... 76

3.1.1 Abstract ..... 76

3.1.2 Introduction ..... 77
3.1.3 Results and Discussion ..... 81
3.1.4 Conclusion ..... 97
3.1.5 Methods ..... 98
3.2 Supporting Information: Nanostructured DNA Barcode Decoding with Solid-state Nanopores: Impact of Bit Spacing and Bit Shape ..... 101
3.2.1 S1. Determining the Linear Carrier Level ..... 101
3.2.2 S2. Event Selection ..... 103
3.2.3 S3. Full Peak Separation Histograms ..... 105
3.2.4 S4. Determining Translocation Direction ..... 106
3.2.5 S5. Full Peak Distributions for Star Attachments ..... 108
3.2.6 S6. Additional Event Traces for Star Attachments ..... 109
3.2.7 S7. Full Peak Distributions for Bundle Attachments ..... 110
3.2.8 S8. Additional Event Filtering for Better Classification ..... 110
3.2.9 S9. List of Oligo Sequences ..... 111
3.2.10 S10. List of Nanopores and Experimental Settings ..... 120
4 DNA Barcoding – Challenges beyond the Nanopore ..... 121
4.1 Nanopore Barcoding ..... 121
4.1.1 Challenges with DNA Triangles ..... 123
4.1.2 Increasing Attachment Density ..... 125
4.1.3 DNA Origami Bundles for DNA Barcodes ..... 128
4.2 DNA Origami Assembly ..... 131
4.2.1 DNA Origami Purification ..... 131
4.2.2 DNA Origami Quality Control ..... 135
5 Conclusion and Outlook ..... 137

Appendix ..... 141
Appendix A: DNA Carrier Assembly ..... 141
Appendix B: Small Bundle Assembly ..... 143
References ..... 145

# Abstract

Humanity is creating information at speeds never seen before in human history. All this information needs to be stored and preserved for future generations. Modern electronic data storage is ideal for short-term storage and quick access but lacks longevity and requires huge amounts of energy. The production of storage capacity is also being outpaced by the rate of information generation. Alternatives for digital data storage solutions need to be found to provide a reliable, low-energy, and low-cost platform. Using DNA as a storage medium for data is one of the more promising technologies put forward to address this issue. DNA is stable for long times and requires minimal energy to keep the information contained within it. One of the ways proposed to store digital data on strands of DNA is the use of nanostructured DNA barcodes in the form of structured molecular attachments to a DNA carrier that encode sequences of binary pieces of information. An ideally suited readout mechanism for these nanostructured molecules is the solid-state nanopore. Nanopore blockage currents are proportional to the size of the molecule, and the presence of different sized attachments will generate distinguishable current spikes in the resulting nanopore current, which can then be interpreted as a binary sequence of bits, thus enabling readout of digital data stored in nanostructured molecules.

In this thesis I explore different strategies of building suitable nanostructured DNA barcode carriers that can be read using a solid-state nanopore. I first investigate the dynamics of DNA carrier translocations through solid-state nanopores to better understand the dynamics and kinetics involved with double-stranded DNA translocations to establish them as ideal carriers for these DNA data storage schemes.

Next, I investigate different designs of nanostructured DNA barcode carriers and bit attachments to find the optimal size, spacing, and attachment chemistry to generate molecules that can be read reliably by the solid-state nanopore. The key issues to solve to make DNA data storage with nanopore readout possible are ensuring each attachment generates a single peak, adjacent peaks are sufficiently far apart to not merge in the current trace, and different attachments generate distinguishable peaks that can be correctly identified based on their current signatures.

Finally, I show that nanostructured DNA barcoding with solid-state nanopore readout is possible and propose further steps to develop a full digital data storage scheme using this approach.

# Acknowledgements

This thesis was a long time coming and the journey to get here was by no means easy. Along the way I have had tremendous support by all the people (current and past) in the TCossaLab and I want to thank each and every one of them. My special thanks go out to Dr. Martin Charron, Dr. Kyle Briggs, Dr. Zachary Roelen, Matthew Waugh, Simon King, Dr. Liqun He, Mohamed Yassine Bouhamidi, and Nada Kerrouri for many fruitful discussions and fun distractions throughout my time in the lab. This thesis would not exist without them.

I also want to extend my special thanks to Prof. Vincent Tabard-Cossa for giving me the opportunity to do this research in his lab. His guidance and mentorship have time and again helped me keep this project on track.

I want to thank my family, especially my mother, father, and sister. They have given me nothing but support and despite being halfway across the globe, I never felt they were too far away.

The biggest thanks goes out to my wife Michelle Mensing. She is the most patient and loving wife I could have ever imagined and has certainly always been a reliable resource of energy and inspiration not just in my scientific career but in my personal life, as well. Our living situation has not been easy over the last 10 years but hopefully with the finishing of this work we can finally find a common place to live and raise our family.

And speaking of family, the list of people I need to thank would not be complete without the newest member in it. Our son Finn has brought an absolute spark of joy and love into my life, and I cannot wait to find out what our adventures together will bring in the future.

# Statement of Originality

The material presented in this thesis represents research carried out in the lab of Prof. Vincent Tabard-Cossa in the Physics Department at the University of Ottawa.

**Chapter 2** consists of an article that I am co-first author on together with Martin Charron. The article takes a deep dive into the translocation behavior of DNA strands with varying persistence lengths to investigate flexibility of a double strand of DNA with nicks in the backbone of one of the strands. I performed most of the experiments and assembled the required structures for this article while Martin Charron developed the analysis tools and provided the simulations used for interpretation of the experimental results.

**Chapter 3** consists of an article that is as of writing this thesis still in the peer review process. I am the first author of this article and designed and assembled most of the structures shown in the article myself. I also performed some of the experiments especially in the first sections of the paper. Additional experiments with barcoded designs were performed by Dr. Zachary Roelen, Nada Kerrouri, and Sofia Zaynullina. I also developed the analysis workflow used to interpret the results and wrote the initial code that has since been adapted by Nada Kerrouri to be made available through the Poriscope app developed in our lab.

**Chapter 4** is a collection of experiments and analyses that were performed over the course of my PhD studies, that informed the interpretation of the published results but have not been shown in articles and likely will not be published outside of this thesis. I performed all of those experiments and analyses myself with the help and input from other members in the lab.

**Chapter 5** summarizes and draws conclusions from the results presented in this thesis. It also gives a brief glimpse into the most recent advances in the lab and experiments that have yet to make it to publication in the future. It highlights the challenges still to overcome and steps to get to an application with the designs presented in this thesis. I helped in designing and assembling the assemblies in that chapter as well as performing and analyzing experiments. All these experiments and analyses were supported by Nada Kerrouri and Dr. Zachary Roelen.

# Statement of Contributions

## Publications

[1] L. He, M. Charron, **P. Mensing**, K. Briggs, J. Adams, H. de Haan & V. Tabard-Cossa, *DNA Origami characterized via a solid-state nanopore: insights into nanostructure dimensions, rigidity and yield*, Nanoscale 15 (2023), 14043-14054, DOI: 10.1039/d3nr01873c

[2] * **P. Mensing**, M. Charron, M. Y. Bouhamidi & V. Tabard-Cossa, *Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-State Nanopores*, Macromolecules 58 (2025), 4094-4102, DOI: 10.1021/acs.macromol.5c00119

[3] D. Wadhwa, **P. Mensing**, J. Harden, P. Branco, V. Tabard-Cossa & K. Briggs, *Nano Trees: nanopore signal processing and sublevel fitting using decision trees*, Digital Discovery 4 (2025), 1743-1750, DOI: 10.1039/D5DD00060B

[4] L. He, B. Elliott, **P. Mensing**, K. Briggs, M. Godin, J. Flax, J. McGrath & V. Tabard-Cossa, *Digital Immunoassays for Sensitive Quantification of Blood Biomarkers Using Solid-State Nanopores*, ACS Nano 20 (2026), 10308-10319, DOI: 10.1021/acsnano.5c16690

[5] * **P. Mensing**, N. Kerrouri, S. Zaynullina, Z. Roelen, P. Xie & V. Tabard-Cossa, *Nanostructured DNA Barcode Decoding with Solid-State Nanopores: Impact of Bit Spacing and Bit Shape*, ChemRxiv (2026), DOI: 10.26434/chemrxiv.15002416/v1 (Currently under review at ACS Applied Materials & Interfaces)

[X] *: Included in this thesis

## Conference Presentations

[6] **P. Mensing**, M. Charron, E. McConnell, K. Briggs & V. Tabard-Cossa, *Towards DNA data storage with solid-state nanopore readout*, Biophysical Society of Canada (BSC) annual meeting, Ottawa, May 2022

[7] **P. Mensing**, M. Charron & V. Tabard-Cossa, *Investigating Flexible and Semi-Flexible Polymers using Solid-State Nanopores*, Biophysical Society (BPS) annual meeting, San Diego, February 2023

[8] **P. Mensing**, M. Charron, S. Zaynullina & V. Tabard-Cossa, *Investigating static and dynamic structures of DNA origami using solid-state nanopores for barcoding applications*, Black Forest Nanopore Meeting, Freiburg, Germany, November 2023

[9] **P. Mensing**, M. Charron, N. Kerrouri, S. Zaynullina & V. Tabard-Cossa, *Toward Data Storage Using Nanostructured DNA with Nanopore Readout*, Ottawa-Carleton Institute for Physics (OCIP), December 2025

# List of Figures

**Figure 1.1: Nanopore sensing mechanism** ..... 2

**Figure 1.2: Overview of different nanopore variants** ..... 3

**Figure 1.3: Nanopore fabrication via controlled breakdown** ..... 7

**Figure 1.4: Nanopore instrumentation** ..... 9

**Figure 1.5: Nanopore noise** ..... 12

**Figure 1.6: Capture process of nanopores** ..... 17

**Figure 1.7: Nanopore event signatures** ..... 19

**Figure 1.8: Structural overview of DNA** ..... 26

**Figure 1.9: DNA origami assembly process** ..... 28

**Figure 1.10: Biomolecules attached to DNA** ..... 31

**Figure 1.11: Overview of DNA data storage schemes** ..... 33

**Figure 1.12: Nanopore barcoding schemes** ..... 35

**Figure 2.1: Semiflexible structures overview** ..... 42

**Figure 2.2: Comparison of nicked dsDNA to ligated dsDNA** ..... 45

**Figure 2.3: Nanopore analysis of flexible and semiflexible DNA** ..... 47

**Figure 2.4: Structural analysis of flexible DNA in nanopores** ..... 51

**Figure 2.5: Investigation of flexible ends on a dsDNA molecule in nanopores** ..... 53

**Figure 2.6: Scatter plots of dsM13 assembly using uncut M13mp18** ..... 61

**Figure 2.7: Filtering of folded fragmented LdsM13 events** ..... 62

**Figure 2.8: 1% Agarose gel performed in 1x TAE buffer of various flexible structures used in this study** ..... 64

**Figure 2.9: 2D heatmaps of all levels fitted to events for the various structures investigated in this study** ..... 64

**Figure 2.10: Average end-to-end distance calculated from 5000 heteropolymers**..... 65

**Figure 2.11: Histogram of pore-monomer distances calculated for simulated polymers of different rigidities anchored inside a nanopore** ..... 65

**Figure 2.12: Normalized ECD of the 3 level (x3/L) for events ending in 1-3-1 transitions**.. 66

**Figure 2.13: Event traces of LdsM13 molecule** ..... 72

**Figure 2.14: Event traces of -1 flexible end molecule** ..... 73

**Figure 2.15: Event traces of -3 flexible end molecule** ..... 73

**Figure 2.16: Event traces of -5 flexible end molecule** ..... 74

**Figure 3.1: Overview of peak finding in event traces** ..... 82

**Figure 3.2: Peak distance determination in nanopore current traces**..... 84

**Figure 3.3: Peak classification for 2-bit star assemblies**..... 89

**Figure 3.4: Peak classification of 2-bit bundle assemblies** ..... 93

**Figure 3.5: Determination of linear dsDNA blockage level**..... 102

**Figure 3.6: Example traces of common error modes during translocation of 2-bit assemblies** ..... 104

**Figure 3.7: Histograms of all peak distances measured with assemblies using the triangle-shaped bit design** ..... 106

**Figure 3.8: Determining the translocation direction for 2-bit assemblies** ..... 107

**Figure 3.9: All-points scatter plot for the peaks detected in the assembly using star-shaped bit attachments**..... 108

**Figure 3.10: Sample event traces for 2-bit assemblies**..... 109

**Figure 3.11: All-points scatter plot for the peaks detected in the assembly using bundle-shaped bit attachments**.....110

**Figure 3.12: Filtered peak classification**. .....111

**Figure 4.1: Overview of triangle structures for peak height measurements** ..... 124

**Figure 4.2: 40-bit assembly with star attachment**..... 126

**Figure 4.3: Schematic overview of all bundle attachment strategies investigated in this study** ..... 129

**Figure 4.4: Gel electrophoresis of the output of different purification methods for DNA barcodes** ..... 134

**Figure 4.5: DNA origami quality control** ..... 136

**Figure 5.1: Overview of most recent results** ..... 138

# List of Tables

**Table 2.1: List of all oligo sequences needed to fully assemble a dsDNA M13 scaffold.**..... 66

**Table 2.2: List of all assembled structures used in this study and the oligos that have been left out of the assembly mixture in the assembly procedure.** ..... 71

**Table 2.3: List of all pore parameters used in this study**..... 71

**Table 3.1: List of all oligo sequences needed to fully assemble a dsDNA M13 scaffold.**.....111

**Table 3.2: List of extra oligos needed to assemble 2-bit triangle assembly (design 1).** .....116

**Table 3.3: List of oligos used for assembly of 4-arm and 12-arm stars for design 2, including the oligos binding them to the carrier.** .....117

**Table 3.4: List of oligos used for assembly of short and long 3-helix bundle attachments, including the oligos binding them to the carrier.** .....117

**Table 3.5: List of all pore parameters used in this study**..... 120

# 1 Introduction

## 1.1 Nanopore Fundamentals

### 1.1.1 Nanopore Working Principle

The resistive pulse sensing idea, forming the basis for nanopore-based detection, begins in the 1940s with the Coulter counter, a device that was developed as a fast way to count blood cells.[1] It worked by suspending a single micrometer-sized aperture between two reservoirs of saline solution, applying a voltage, and monitoring the resulting ionic current flowing through the aperture. The magnitude of this ionic current is proportional to the size of the aperture, and blood cells flowing through the aperture would transiently block part of that current. The magnitude and frequency of these current blockages can then be measured by an ammeter and analyzed to infer the size and number of cells in the solution, respectively.[2] To measure smaller targets including single biological molecules like proteins or DNA the required measurement electronics had to improve. Tools developed for electrophysiology in the 1970s and 1980s enabled more sensitive current measurements in the pA to fA levels.[3–5] In parallel, molecular targets, being generally in the nanometer size range, require apertures with physical dimensions in the same range. These nanometer-sized apertures are called nanopores and were first shown to be useful for investigating single DNA molecules in 1996 by Kasianowicz, Deamer and colleagues.[6] While the working principle of these nanopores appears similar to that of the Coulter counter scheme, scaling down to nanometer sizes brings to light complex transport processes and polymer physics that this thesis delves into. The process is outlined in **Figure 1.1**.

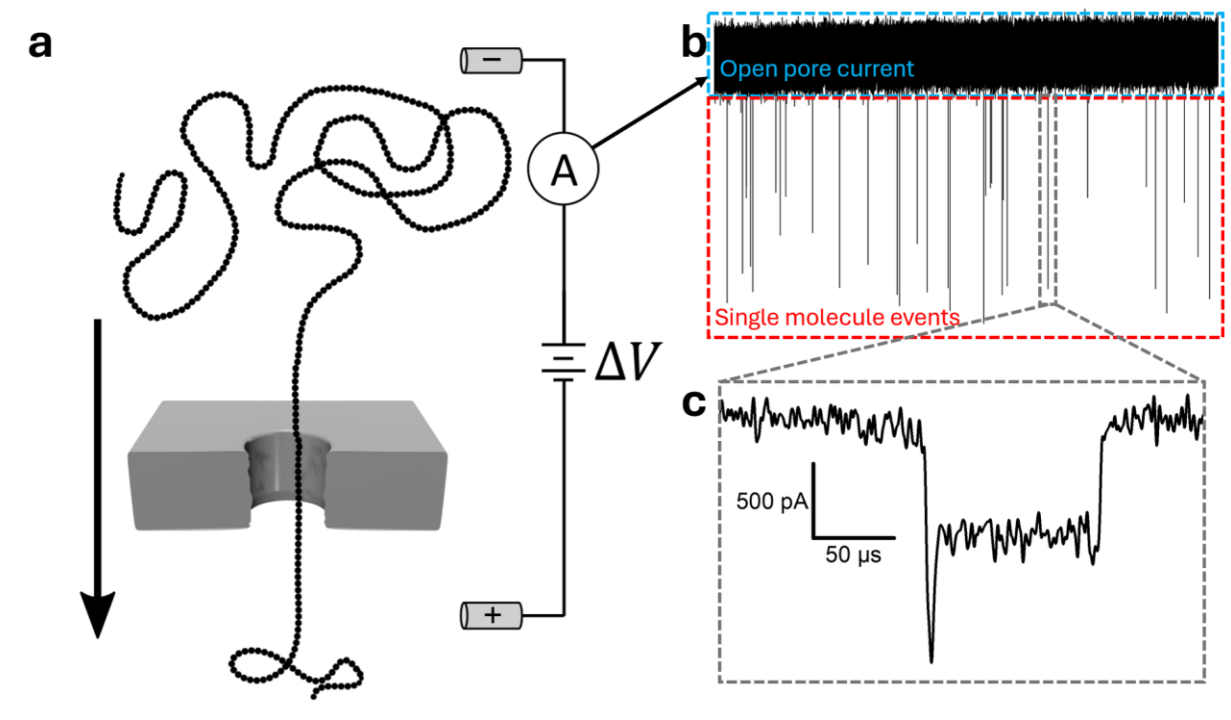



***Figure 1.1: Nanopore sensing mechanism.*** ***a)*** *Schematic of a polymer threading through a nanopore under an applied bias voltage, while the current is being measured.* ***b)*** *The resulting current trace recorded by the connected ammeter (typically a transimpedance amplifier). The open pore current is marked in blue while events caused by the translocation of biomolecules are marked in red.* ***c)*** *Zoom-in on one of the single-molecule events.*

Since the first proof of concept publication, there has been a wide range of applications for nanopore systems. These range from single molecule detection and counting,[7,8] structural analysis of polymers like DNA,[9–11] RNA,[12,13] proteins,[14,15] and other sequence-controlled polymers,[16–18] DNA sequencing,[19,20] protein profiling,[21–25] and more.[26–28]

### 1.1.2 Types of Nanopores

Nanopores can be roughly categorized into two types based on the material from which the pore is made. The main distinction is made between solid-state nanopores – artificially made pores in solid-state materials (silicon nitride, glass, graphene, etc.)[29] – and biological pores made from biological molecules, typically proteins[30] or DNA.[31,32] A variety of nanopore variants is shown in **Figure 1.2**.

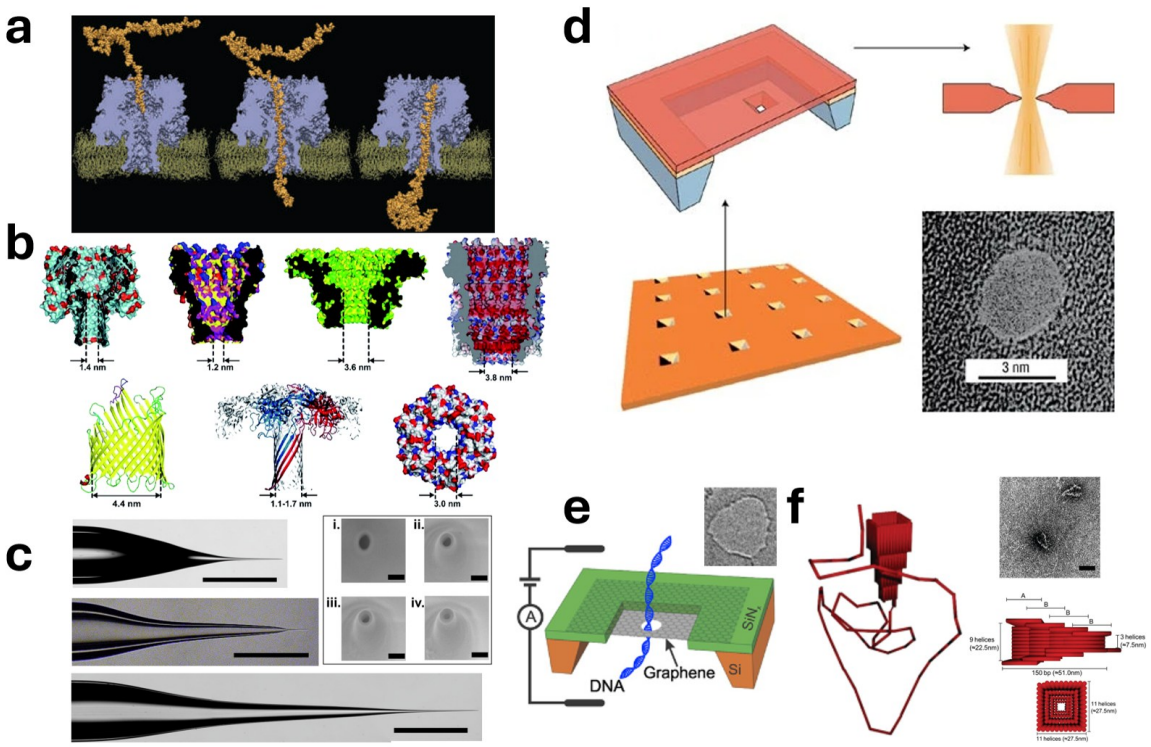



***Figure 1.2: Overview of different nanopore variants. a)*** *Schematic of the α-hemolysin protein forming a channel through a lipid membrane with a piece of ssDNA translocating.[33] Reprinted with permission from Springer Nature.* ***b)*** *Illustrations of commonly used protein nanopores.[34] Used with permission of Royal Society of Chemistry.* ***c)*** *TEM images of borosilicate nanocapillaries. Adapted and reprinted from Bafna[35] © 2016 Bafna, Soni. Licensed under CC BY 4.0.* ***d)*** *Schematic of a solid-state nanopore in a silicon nitride membrane fabricated through ion beam drilling with a TEM image of the resulting pore in the membrane.[33] Reprinted with permission from Springer Nature.* ***e)*** *A schematic and TEM image of a 2D nanopore in a suspended graphene sheet.[33] Reprinted with permission from Springer Nature.* ***f)*** *A schematic and TEM image of a DNA origami structure forming a nanopore on the inside.[31] Reprinted with permission. Copyright 2012 American Chemical Society.*

Biological pores are made from proteins found in living organisms where they act as passages through the insulating cell membranes surrounding cells and organelles, allowing ions and molecules to move across the membrane. As such their molecular structure is fixed and atomically well defined. Modifications and mutations can be applied to them in order to fine-tune the individual pore to their designed purpose making them useful for a variety of applications. The most important application for biological nanopores is sequencing of polymers, mainly DNA sequencing[36] but protein sequencing is actively being developed.[37] For this purpose a pore-forming molecule is carefully inserted into a suspended lipid membrane separating two liquid reservoirs. Performing this process while making sure that only a single protein gets inserted is tricky and the lipid membrane can be prone to breakage under stress when the membrane is large. This means that experimental conditions are limited and need to be carefully tuned to the type of protein used as well as the lipids used for forming the membrane. Applying high pressures or voltages risks breaking the membrane, while high salt concentrations, high or low pH values can denature the pore-forming protein and end the experiment. The fixed molecular structure also means that pore size cannot be controlled other than by changing the protein completely, limiting the range of analytes for each individual pore.

Artificially made pores in solid-state materials fix a lot of the issues with biological pores at the expense of giving up knowledge about the precise shape of the pore. The membranes can be made to a thickness much more suitable to the desired application (from single layer 2D pores[38] to nanopores in 40 nm thick membranes[39]), which allows for the application of pressure and high voltage without risking the breaking of the membrane. These materials are also, generally, chemically inert meaning changes in salt concentration, pH, and temperature do not degrade the

pore itself and it stays operational over a large range of experimental conditions. Additionally, the pore size can be fine-tuned to optimize the signal for each given analyte.

There are two different geometries for solid-state pores that are worth exploring, nanopipettes and membrane-based nanopores. Nanopipettes are small openings at the tip of a conical capillary of glass (see **Figure 1.2c**). They are made by locally heating a hollow glass cylinder with a laser while carefully pulling it apart until it snaps in half giving a nanometer sized aperture at the tip.[40] These pores are routinely formed to about 10 – 20 nm sizes and are stable for long periods of time.[41] Given the shape of the capillary the two sides of the nanopore are inherently asymmetric and the length of the sensing region (where the hole in the capillary is smallest) depends on the cone angle of the capillary. This asymmetry has effects on the capture and translocation dynamics of analytes.[42]

The other type of solid-state nanopores is formed in a flat membrane. Here a free-standing membrane of a dielectric material (usually silicon nitride, SiN) is the starting point. The membrane material is deposited on a wafer of silicon or glass and then a small section is etched away such that only the silicon nitride is left. In this section a pore can be drilled by any of the multitude of different methods of pore fabrication developed over the years.[43–46] The length of the nanopore is directly related to the thickness of the initially deposited film of silicon nitride, while the size can be adjusted either during fabrication or after suspending the pore in a liquid solution through a process called conditioning. These pores are symmetric as opposed to the asymmetry of the nanopipettes which means they can be operated either way around without the physics of

translocation being affected. The sensing region is smaller than the effective sensing region of nanopipettes, 10 nm for flat pores compared to ~100 nm for pipettes.

### 1.1.3 Nanopore Fabrication: Controlled Breakdown

Depending on the material and geometry of the nanopore, as well as the purpose for which it is being used, fabrication of nanopores can be done in several ways. In this thesis, solid-state nanopores are fabricated in silicon nitride (SiN) membranes using the controlled breakdown (CBD) process.[47,48] This process is described in detail in the following section. Interested readers are welcome to explore other nanopore fabrication methods in a review by Edel and colleagues[45] as well as a book chapter written by Briggs and colleagues.[43]

In the CBD process, the membrane is mounted in a fluidic cell and suspended in an electrolyte solution (usually KCl or LiCl). A linearly increasing voltage is applied, while the leakage current is monitored. As the electric field across the membrane approaches the dielectric strength of the material, it causes charge carriers to tunnel through the trap states from one side of the membrane to the other, generating defects that accumulate over time. A conductive weak point is created in the otherwise insulating membrane that eventually breaks down, creating a path for the ions to flow freely through the membrane.[49] This process is schematically shown in **Figure 1.3a**.

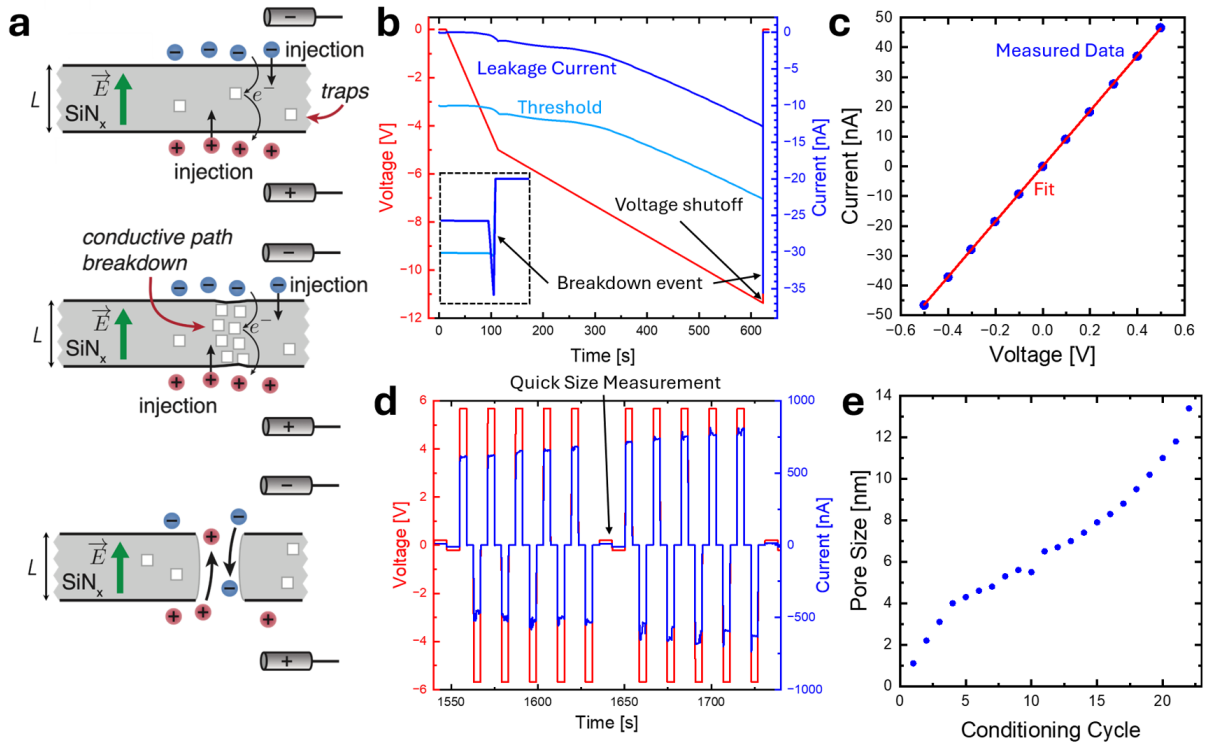



***Figure 1.3: Nanopore fabrication via controlled breakdown.*** ***a)*** *Schematic of the controlled breakdown process, where defects in the membrane accumulate until the membrane breaks down and forms a pore that creates an electrical path for ions to go through.*[50] *Reproduced with permission from Springer Nature.* ***b)*** *Voltage applied and current measured during nanopore fabrication via controlled breakdown. Once pore formation is detected by the current (blue) passing an offset moving average threshold (light blue), the voltage (red) is shut off. The inset shows a zoom-in on the breakdown event.* ***c)*** *I-V characteristic of a nanopore after fabrication is complete. The size of the nanopore can be estimated using the slope.* ***d)*** *The conditioning of a nanopore by applying short voltage pulses (red) in positive and negative direction. The current (blue) is monitored at the same time. Every 5 cycles a quick measurement is performed to roughly estimate the pore size and decide if conditioning is complete.* ***e)*** *Pore size estimates from quick measurements performed during conditioning showing the pore growing with each cycle.*

The size of the resulting hole can be estimated by measuring the current-voltage (I-V) behavior of the pore (see **Figure 1.3c**) using the equation

$$d = \frac{G}{2\sigma}\left(1 + \sqrt{\left(1 + \frac{16\sigma L}{\pi G}\right)}\right), \tag{1.1}$$

where $d$ is the diameter of the pore, $G$ is the conductance of the pore (measured current divided by the applied voltage), $\sigma$ is the conductivity of the ionic solution, and $L$ is the length of the pore.[51] The pore size can be adjusted through a process called conditioning by applying short voltage pulses,[52] usually 30-50% of the breakdown voltage, until the desired pore size is reached. Pore growth is proportional to the amount of current flowing through the pore. This means the fastest conditioning happens when higher voltages are applied. However, to not risk fabricating another pore at a different location in the membrane, the applied voltage is kept at or below 50% of the initial breakdown voltage. The voltages applied for sensing are generally less than 1 V and usually do not significantly grow the pore, due to the much smaller currents observed with those voltages. The above model for pore size is simplistic and assumes the pore to be a perfect cylinder of length equal to the thickness of the membrane in which the pore is formed. In reality, the shape of the pore is more complex. A more accurate model, using a more realistic pore shape, has been proposed recently to better estimate the size of the nanopore under more realistic assumptions.[53] Since this model relies on extra information from molecular translocations, it cannot be used at this stage of pore formation. It is discussed further down in **Section 1.1.7**.

### 1.1.4 Instrumentation

There are two major pieces of equipment needed for a nanopore experiment. A fluidic cell housing the membrane containing the pore, and a current amplifier capable of measuring the small, transient current changes observed in nanopore current recordings. The fluidic cells are fabricated in house using a 3D printer and have evolved over the course of this work to include a host of different arrangements, but the basic design has stayed roughly as shown in **Figure 1.4a**.

Based on its size, the current running through a nanopore ranges from roughly 100 nA down to <0.1 nA depending on the experimental setup, with the majority of work in this thesis taking place in the ~10 nA range. The transient current blockades generated by single molecules translocating through the pore generally range from ~$10^1$ µs up to >$10^5$ µs depending on the size of the molecule and pore. Because of the speed and range of translocation times, the current amplifiers used to measure these fast translocations need to be very sensitive and operate with high bandwidth to ensure proper resolution. In this study two current amplifiers were used: the VC100 and VC400, both manufactured by Chimera Instruments. The main difference is that the VC400 is a newer model with marginally increased sample rate of 5 MS/s (megasamples per second), while the VC100 operates at 4.167 MS/s, and a wider working range of currents up to ~100 nA (up from ~20 nA for the VC100). The VC400 also allows for simultaneous measurement of four channels. The amplifiers used in this work are shown in **Figure 1.4b**.

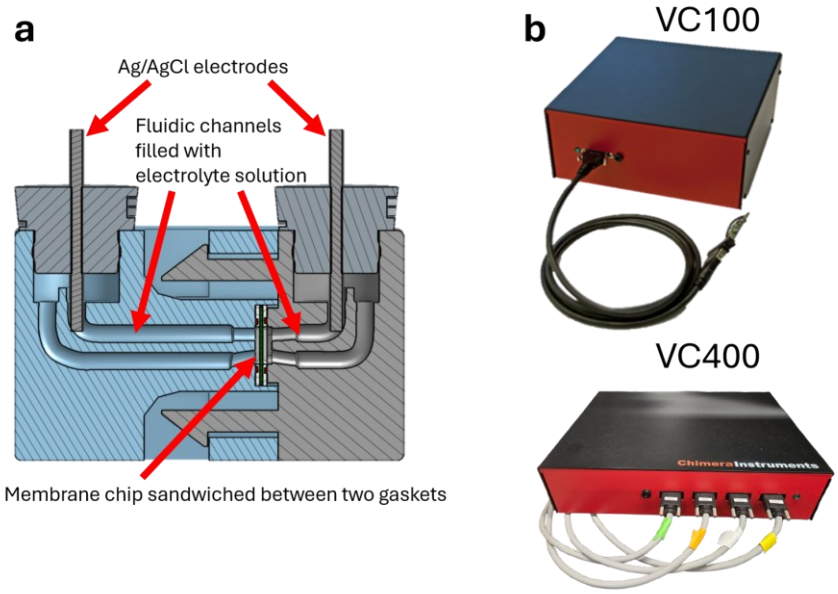



***Figure 1.4: Nanopore instrumentation. a)*** *Schematic of the nanopore fluidic cell.* ***b)*** *Current amplifiers used in this work. Both instruments are manufactured by Chimera Instruments (https://chimerainstruments.com/products/)*

### 1.1.5 Nanopore Noise

Every measurement of a physical quantity is subject to noise that may vary as a function of frequency. Nanopore current measurements are no exception. The noise in a nanopore setup is best understood by considering the equivalent circuit diagram in **Figure 1.5a**. Every component in that circuit, including the fluidics, pore, and membrane, acts as a source of a different type of noise that can be summarized in the power spectral density (PSD) of the resulting current trace. In general, the PSD of a nanopore current can be expressed in the following polynomial form[50]

$$S(f) = \frac{a_0}{f} + a_1 + a_2 f + a_3 f^2 \tag{1.2}$$

where $f$ is the frequency and $S$ is the spectral power at that frequency, estimated by taking the square magnitude of the Fourier transform of the steady-state nanopore current for a given voltage. The pre-factors $a_n$ represent the different contributions of the noise and are influenced by various different parameters of the measurement setup.

The $a_0$ contribution is also called *flicker noise* and usually dominates the low frequency end of the spectrum as shown in **Figure 1.5**. Its origin is still debated but it is thought to originate from the interactions of the ions in the electrolyte with the walls of the pore.[54–56] The presence of tiny air bubbles trapped on the surface in or near the pore also potentially contributes to this type of noise.[57] This type of noise scales quadratically with the applied voltage (and hence the current flowing through the pore).

There are two major contributions in the constant $a_1$ contribution to the noise spectrum, *thermal noise* and *shot noise*. These noise sources are present in all electrical circuits with thermal noise – as the name suggests – scaling with the temperature at the pore as well as the resistance of the pore, while shot noise arises from the fluctuations in the number of ions moving through the pore and therefore scales with the current flowing through the pore and the charge, type, and size of the ions carrying that current.

At higher frequencies (>10 kHz), the other electrical properties of the involved components start influencing the noise spectrum. The $a_2$ contribution is called dielectric noise and is a form of frequency dependent thermal noise due to the dissipation of energy in the capacitor formed by the suspended membrane. It is therefore not surprising that this noise contribution scales with the capacitance of the chip used.

At even higher frequencies, the PSD is eventually dominated by the $a_3$ contribution, which is called *capacitive noise*, especially at high bandwidths (>100 kHz). As the name would suggest the capacitive noise originates in the capacitance of the nanopore device and scales according to

$$S_{\text{capacitance}} \propto (C_{\text{total}} * f)^2, \tag{1.3}$$

with $C_{\text{total}}$ representing the total capacitance – comprising the capacitance of the membrane and supporting chip in which the nanopore is fabricated, as well as any stray capacitances arising from the interfaces in fluidic path and between the electronic components of the nanopore measurement system. This means that for high bandwidth measurements, minimizing the overall capacitance of the nanopore device is critical to achieving sufficient signal-to-noise ratio (SNR).

For a given bandwidth $B$ of the current measurement, a key metric to characterize the quality of the noise level is the root-mean-square (RMS) noise, which gives the standard deviation of the fluctuations in current at that bandwidth and is derived from the PSD according to

$$I_{\mathrm{RMS}}(B) = \sqrt{\int_{1}^{B} S(f)df} \tag{1.4}$$

The overall shape of the noise in nanopore current recording is shown in **Figure 1.5**.

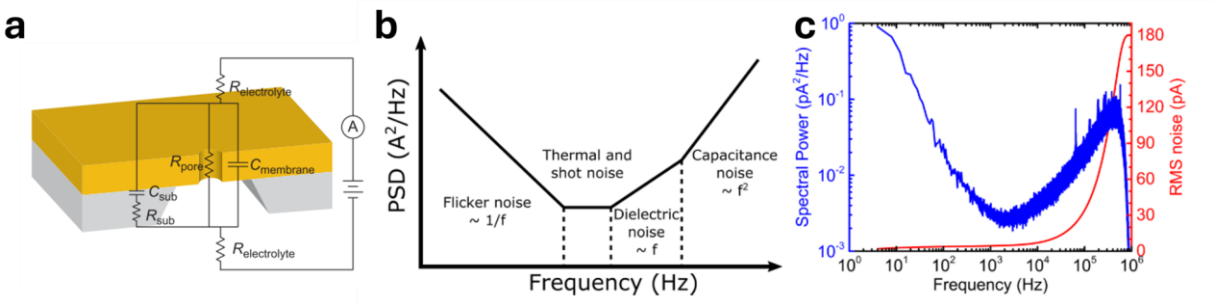



***Figure 1.5: Nanopore noise. a)*** *Equivalent circuit diagram of a nanopore fluidic cell. Reprinted from,[50] Copyright 2013, with permission from Elsevier.* ***b)*** *Schematic PSD for nanopore recording highlighting the different noise sources and their impact on the overall shape of the PSD.* ***c)*** *Example PSD (blue) and RMS noise (red) for a real 10 nm nanopore in a silicon nitride membrane suspended on a silicon chip under 200 mV bias voltage. The chip is painted with a layer of PDMS to reduce the capacitance of the chip.*

A simple way to reduce the RMS noise in a nanopore setup therefore lies in reducing the capacitance of the silicon chip e.g. by adding additional layers of dielectric materials underneath or on top of the membrane layer, or by minimizing the area of the chip exposed to the electrolyte.[58] This can be done, for example, by gaskets with a very small aperture to seal the nanopore chip, by manually painting a thin layer of polydimethylsiloxane (PDMS) on the chip, or by having a layer

of silicon oxide between the silicon of the chip and the membrane layer. All three methods were used extensively in this work to allow for high bandwidth (>500 kHz) measurements while ensuring an SNR of at least 6.

### 1.1.6 Nanopore Capture Process

The nanopore capture process for charged molecules is a complex process influenced by many factors inside and outside of the experimenters' control. The capture process is influenced by pore size, applied bias voltage, solution conductivity, molecular charge, size, shape, and length of the molecule. A charged molecule approaching a nanopore will feel the electric field generated inside the pore and be drawn in or repelled from it depending on the polarity of the charge. Due to the molecular nature of the setup, there are many parameters that affect translocations of polymers through a nanopore. A rough description of the main features is given here, but the reader is encouraged to read any of the more in-depth publications, especially the work of my colleague Martin Charron, for a more quantitative description of the processes and forces involved.[59–61]

Applying a voltage difference between the two reservoirs of a fluidic cell separated by a membrane containing a single nanopore creates an electric field inside the nanopore itself since this is by far the highest resistance of the circuit. The membrane is made of a dielectric material and therefore impermeable to electric charges, but the pore provides a path through which ions can translocate the membrane to the other side. The resistance of the pore is generally on the order of ~10–100 MΩ (when working in typically high molar salts like 3.6 M LiCl), which is much higher than the resistance along the fluidic paths connecting the pore to the electrodes (which are usually in the kΩ range). This means that almost the entire applied voltage drops across the nanopore, and

the resulting electric field is nonzero only in the immediate vicinity of the nanopore. The electric field is strongest inside the pore, where the resistance is largest, and drops off with the inverse square of the distance from the pore, analogous to Coulomb's law. From far away, the electric field lines of the pore appear like those of a point charge.[53] The exact shape and strength of the electric field close to the pore, however, depends on a variety of factors: the solution conductivity, ionic composition of the solution, the material and charge on the surface of the membrane, the shape and size of the pore itself,[51] and even the type and charge of the analyte in the solution[62,63] all contribute. Correctly modeling the electric field in and around the pore is challenging and has only recently been achieved using oblate spheroidal coordinates.[53,61] The localization of the electric field to the vicinity of the nanopore means that the bulk solution experiences a very weak electric field, so for charged molecules in the bulk solution, the electric field is not strong enough to meaningfully influence its movement. As a charged molecule comes closer to the pore through diffusion, there will be a point at which the electric field is strong enough to drive directed motion along the field lines. The interface where the motion of an analyte becomes dominated by the electric field as opposed to the diffusion is called the *capture radius* of the pore, as it is roughly hemispherical.[64,65]

Once a molecule has passed this point, it is very unlikely that it diffuses back out into the bulk solution and instead is more likely to be pulled in by the pore and translocate to the other side of the membrane. It should be emphasized that the concept of the capture radius is mostly a helpful analogy: there is no well-defined "radius" at which molecules are guaranteed to translocate the pore. Nevertheless, this is a useful conceptual framework to explain the two distinct regimes of molecular motion near a nanopore. There will be a range of distances from the pore where the

energy due to diffusion and the energy due to the electric field are similar in magnitude, and molecules in that region still have a chance to diffuse away from the pore instead of translocating.

A polymer in free solution has many possible conformations, while a polymer constrained by the pore during translocation has reduced conformational freedom. This leads to an entropic barrier for translocation of a large polymer through a nanopore.[66] The exact value of this barrier is dependent on the length and stiffness of the polymer, the size and shape of the pore, the temperature of the system, and the applied bias voltage. Furthermore, translocation behavior differs based on the height of the barrier relative to the energy provided by the electric field close to the pore.[67–69] If the barrier to entry is higher than the electric potential at the opening of the pore, the molecule is said to be in the barrier-limited regime, whereas if the opposite is true, the molecule is said to be in the diffusion-limited regime. In the barrier-limited regime, a molecule inside the capture radius of the pore is not guaranteed to successfully translocate through the pore and can sometimes attempt multiple times before translocation occurs or simply diffuse away. It is also possible for multiple molecules to occupy the capture radius at the same time, leading to crowding around the entrance of the pore. Experimentally, this regime can be identified from the capture statistics: the rate of translocation (the *capture rate*) remains proportional to the polymer concentration but depends exponentially on the bias voltage and increases with molecule length, unlike the diffusion-limited regime, where it is linear in voltage and largely independent of length (past a certain polymer size). This linear dependence on concentration breaks down at high concentrations, once multiple molecules crowd the region inside the capture radius. If the barrier to entry becomes too large, short and shallow blockages can be observed, which can be attributed to molecules "bumping" into the pore but failing to translocate. This happens when molecules approach the

mouth of the pore and block some of the electric field immediately outside the pore (the *access region*) and then move away again. The barrier-limited regime is usually observed in small pores, at low bias voltages and for shorter molecules, the reduced charge making it more difficult to pass on first attempt due to the lower forces acting on the molecule.

In the diffusion-limited regime (as the name suggests) the main restriction on molecules translocating the pore is the rate at which diffusion brings them into the capture radius. Here, the molecule is almost guaranteed to translocate through the pore once it enters the capture radius and it is unlikely to find more than one molecule inside the capture radius unless the concentration of the polymer is high enough that the mean distance between molecules is smaller than the capture radius. In this regime, the capture rate scales linearly with the applied bias voltage as well as the diffusion coefficient and concentration of the polymer, enabling the nanopore to quantitatively measure polymer concentration.[70,71] In most cases, performing experiments in the diffusion-limited regime is preferred, due to the generally higher capture rates, and extra strategies are often employed to increase the rate at which molecules arrive at the capture radius to decrease measurement times of counting molecules, by e.g. pre-concentration,[72] pre-confinement,[73] salt concentration gradients,[60,74,75] transverse flow,[76] or adding pressure gradients across the pore.[77]

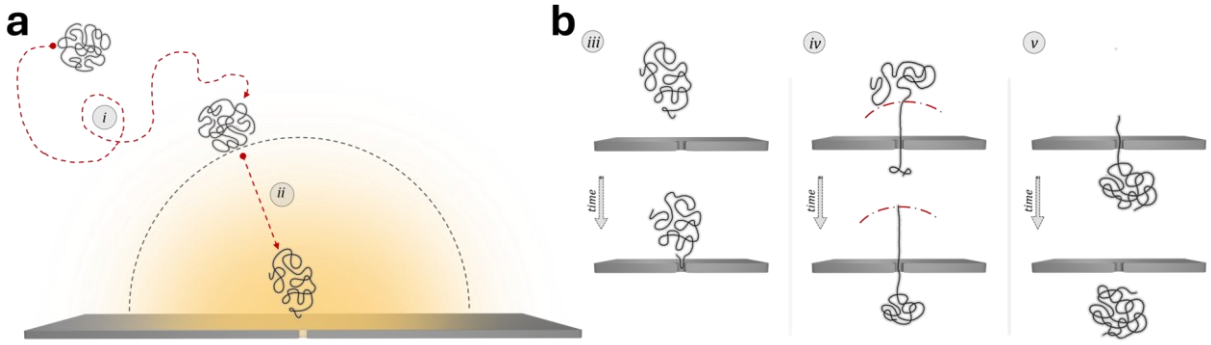



***Figure 1.6: Capture process of nanopores. a)*** *Schematic of the capture with the electric field shown by the yellow shadow and the capture radius as a dotted semicircle. The polymer is freely diffusing (i) outside the capture radius until it arrives at the capture radius and is driven towards the pore (ii).* ***b)*** *Schematic view of the translocation process. The initial threading of the polymer (iii), tension propagation while the polymer is translocating (iv) the end front of the tension is shown in red, the final phase of translocation (v) when the tension has reached the end and the last section of polymer is threaded through the pore. Reproduced with permission from Springer Nature.*[60]

The actual translocation process – once capture happened and the molecule starts threading through the nanopore – is similarly complex and depends heavily on the conditions of the pore and the molecule. For charged polymers like DNA, the impact of the electric field along its 3D extension as it approaches the pore is not uniform. The part of the molecule closer to the pore will experience a stronger electric field and therefore more directed motion towards the pore, while parts of the molecule further away from the pore are still more diffusive in their movement. A polymer in free solution is generally not in an extended rod-like state (see the DNA section further down for more in-depth description) but will be coiled up into a self-avoiding random walk. In free solution, a sufficiently long polymer (relative to its persistence length) takes up a spherical space on average, but any instantaneous conformation is likely quite extended.[78] If a coiled-up polymer like this approaches the electric field of a nanopore, the part of the molecule coming in contact with the electric field first is often not one of the ends of the molecule.[79] This leads to

folding as a section of the polymer is captured by the pore and the polymer buckles, pulling its two ends along, generating a very distinctive event signature with a deep blockage state followed by shallower blockage state with half the blocked current. The longer the polymer the less likely capture at one of the ends is and the more folding is observed.[11] As the polymer is unwinding into the pore, there is tension building up along the contour of the polymer. For semi-rigid polymers like DNA the tension front moves away from the pore and stretches the remaining parts of the polymer before reaching the end and the polymer translocates completely through the pore. This gives rise to a nonlinear velocity profile of the polymer where the initial velocity is high, slowing down as more and more parts of the molecule start moving and adding drag before speeding back up when the tension front reaches the end of the polymer and friction reduces due to fewer parts of the polymer being in motion.[60] The overall translocation time (also called *dwell time*) has been shown to be largely determined by the initial conformation of the polymer at the time of capture,[80,81] due to the internal relaxation times generally being longer than the translocation times observed in nanopore sensing. This leads to large spreads in translocation times, especially for long polymers.

### 1.1.7 Nanopore Event Signatures

Biomolecules translocating a nanopore transiently block the flow of ions. In this work, we refer to these current blockages as *events*. A lot of information about the translocating molecule can be deduced from the event *signature*, including duration (dwell time),[82,83] blockage depth,[7,84] number of sublevels,[85] and noise.[86] Nanopores have been used to investigate many different kinds of biomolecules, ranging from simple DNA strands[87] to more complex DNA origami assemblies,[9,88,89] proteins,[23,90] antibodies,[91,92] and more.[18,93] Each of these can generate different

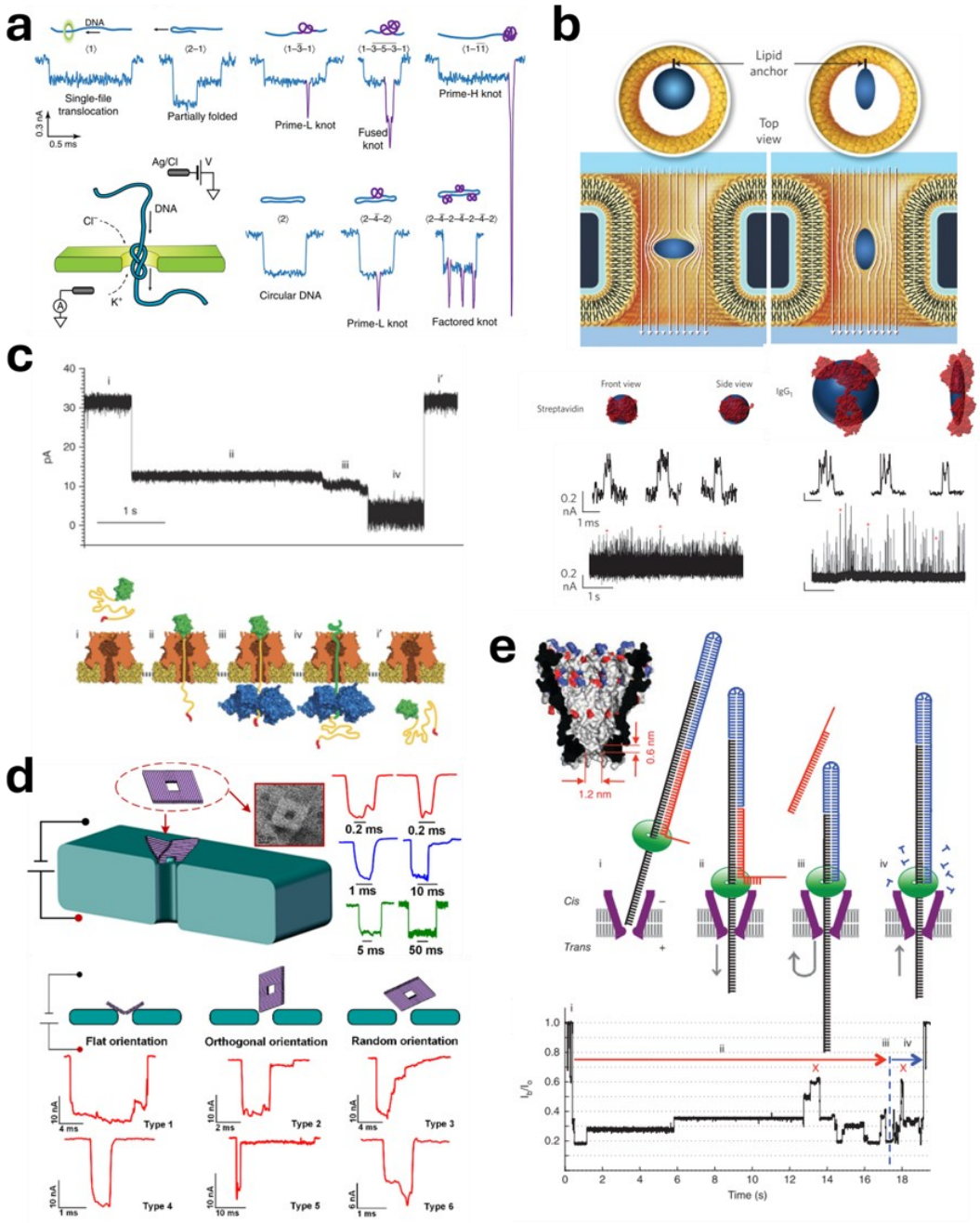



***Figure 1.7: Nanopore event signatures.*** ***a)*** *DNA secondary structure motifs and their impact on the nanopore signature.*[94] *Licensed under CC BY 4.0.* ***b)*** *Oblate shaped proteins rotating in a solid-state nanopore generate identifiable event signatures.*[21] *Reproduced with permission from Springer Nature.* ***c)*** *Proteins being unthreaded and pulled through a biological pore by a ClpX enzyme.*[95] *Reproduced with permission from Springer Nature.* ***d)*** *DNA origami plates translocating through a solid-state nanopore generating different event signatures depending on orientation. Reproduced with permission from* [96] *Copyright 2020 American Chemical Society.* ***e)*** *A DNA molecule being fed to a biopore by an enzyme (helicase) which generates a sequence of blockage levels that encodes the DNA sequence.*[97] *Reproduced with permission from Springer Nature.*

event signatures upon translocation that give information about the dimensions of the molecule,[98–100] its structure,[101] its charge,[102] its dynamical properties,[103] reaction pathways,[104–106] and potentially even its sequence, in the case of DNA,[20] peptides and proteins.[107] **Figure 1.7** shows a few examples of these event signatures.

On top of the molecular structure of the biomolecule, the shape of the event is also influenced by the location of the molecule relative to the pore (inside the pore or outside the pore in the access region), and by the noise inherent to the measurement system as rapid changes in nanopore current (relative to the bandwidth of the system) are distorted by the finite response time of the nanopore device (and fluidic cell) and the limited bandwidth of the electronics used in the experiment. Generally, smaller molecular features translocate faster and produce shallower blockages than larger molecular features, which means the ideal signal-to-noise ratio is seen when the pore is just barely larger than the molecule translocating. This means that finding the right balance between bandwidth and noise is crucially important in experimental design, including the choice of operating conditions (salt concentrations and applied voltage), and the pore dimensions for the molecular sizes studied.

A significant amount of modeling and simulation work has been done to correctly interpret and analyze current blockages generated by translocating molecules in various configurations and conformations that require precise modeling to understand and interpret. This has led to investigations of the structure of molecules like DNA,[94,108,109] proteins,[21,110] and RNA,[12,111,112] as well as DNA-DNA binding,[113] protein reactions,[14,114] and DNA-protein binding reactions.[115,116]

In this thesis, DNA is used as a molecular carrier for digital barcodes, and the interpretation of DNA translocation events is therefore the most important one to understand here. For double-stranded DNA, internal relaxation times are generally longer than the translocation times, meaning the conformation of the polymer is essentially frozen during the translocation process. Consequently, the main determinant of translocation dynamics generally, and the dwell time specifically, is the conformation of the polymer before and during translocation through the pore. These different conformations do not simply change the average speed and the dwell time of the resulting translocation – they also impact the instantaneous velocity of the segment of the molecule in the pore at each moment of the translocation event. This leads to tension propagation as described in the previous section and shown in **Figure 1.6** which dominates DNA translocation.[117–119]

On the scale of the nanopore (~10 nm), DNA behaves like a rigid rod (its persistence length is ~50 nm) and it is well approximated as a cylinder for the purposes of interpreting the blockage depth. In the simplest case, the translocating strand of DNA can be thought of as a single cylinder translocating a nanopore, and therefore, the blockage can be calculated as

$$\Delta G = G_0\left(d_{pore}\right) - G'^{cyl}\left(d_{pore}\right), \tag{1.5}$$

where $G_0$ is the open-pore conductance given by the conductance model underlying **Equation 1.1**, $d_{pore}$ is the size of the pore, and $d_{DNA}$ is the diameter of the DNA (which is usually taken to be ~2.2 nm). $G'^{cyl}$ is the blocked conductance by a cylindrical polymer given by

$$G'^{cyl} = \left( \frac{4L}{\pi\sigma\left(d_{pore}^2 - d_{DNA}^2\right)} + \frac{2K\left(\frac{d_{DNA}}{d_{pore}}\right)}{\pi\sigma d_{pore}} \right)^{-1}, \quad (1.6)$$

where $\sigma$ is the solution conductivity, $L$ is the length of the pore, and $K(d_{DNA}/d_{pore})$ denotes the complete elliptic integral of the first kind.[53,120] There are two major contributions to the nanopore blockage in **Equation 1.6**. The first term in the equation is the blockage arising from the molecule blocking current inside the pore, while the second term arises from the access region blockage of the parts of the molecule outside the pore. This model can be used to compare the observed blockages for DNA barcodes to the expected blockages based on the geometric shape of the molecules, which provides a crucial control on the correct interpretation of event signatures.

The duration of the translocation is a useful indicator about the length of the molecule, though the relationship is not linear.[33,121] Especially for longer strands of DNA (>1 kbp) there is often a more complex event structure than a simple head-to-tail translocation would indicate. Examples include DNA folding as described above,[79] but also knots,[94,122] secondary structure,[123] and supercoiling[124] (see **Figure 1.7a** for example traces). Even in these cases, however, the simple geometric model is often sufficient, as the blockage states usually fall on predictable multiples of the single dsDNA blockage level. **Chapter 2** of this thesis investigates this phenomenon for ssDNA-dsDNA hybrids and how persistence length influences the translocation behavior and the prevalence of folding.

### 1.1.8 Nanopore Translocation Control

Most factors that impact capture also impact other aspects of translocation behavior.[125,126] Generally, high capture rates and slow translocation are preferred, but experimentally this is difficult to achieve since most of the controlled aspects of the experiment cause these quantities to vary together. As such, it often requires the simultaneous use of different techniques to increase capture while slowing down the actual translocation process.

For smaller molecules, or when observing small features on larger molecules, capture rate is usually sacrificed to operate in a regime where passage kinetics are reduced, to resolve fast features in the current trace. The use of a processive enzyme to controllably feed the polymer in step slow enough to resolve each monomer, is what enabled biological nanopores to sequence DNA and peptides.[97] On solid-state nanopores, other physical and chemical approaches have been investigated to improve the sensing performance, including interfacing the nanopore with a high density medium like agarose gel,[127] varying the electric field,[128] balancing electrophoretic and electroosmotic flow,[129] applying a pressure gradient across the pore to induce liquid flow opposite the translocation direction,[77,130] blocking one side of the pore and trapping the analyte inside the sensing region of the pore,[14] and attaching a large bead to the polymer that gets stuck in the pore.[131,132]

## 1.2 DNA Structure

### 1.2.1 DNA as a Polymer

A polymer molecule is a chain of monomers that are chemically linked together to form a long strand. Deoxyribonucleic acid (DNA) is a naturally occurring polymer consisting of two polynucleotide chains that are intertwined to form a helix.

In the case of DNA, each of those chains is made up of a sequence of the four different nucleotides adenine (A), cytosine (C), guanine (G), and thymine (T) that are linked together covalently. The nucleotides consist of three main parts, a nucleobase that gives it its identity, a five-carbon sugar ring (deoxyribose), and a phosphate group connected to that sugar ring (see **Figure 1.8**). The phosphate of one nucleotide can chemically bond to the sugar ring of a different nucleotide to connect the two nucleotides together and these two parts therefore form the backbone of the resulting polynucleotide chain.

The nucleobases stick out from a single-stranded DNA molecule and can interact with other nucleobases through a process called *Watson-Crick-Franklin base pairing,* during which hydrogen bonds form between two complementary bases: A binds to T through two hydrogen bonds, while three hydrogen bonds bind C to G. Because of this specific interaction, base pairing is highly selective to the sequence of involved strands, and mismatches (while possible) are thermodynamically unfavorable and come at a high energetic penalty.

The stiffness of a polymer chain (or any chain for that matter) is measured by a metric called the *persistence length* $l_p$, which is defined as the length over which the movement of monomers remains correlated. The two complementary polynucleotide chains twist around each other to form a helical double strand of DNA. Due to the flatness of the nucleobases on the molecular level, adjacent basepairs in the double strand can stack on top of each other, further stabilizing this structure and giving the resulting double-stranded chain a persistence length nearly 50× larger than that of single-stranded DNA.[133] Polymers with large persistence lengths (like double-stranded DNA, dsDNA, $l_p \approx 50$ nm or ~150 basepairs (bp)) are stiffer than polymers with smaller persistence lengths (like single-stranded DNA, ssDNA, $l_p \approx 1 - 3$ nm or ~3–9 nt).[134–136] **Figure 1.8c** highlights the difference between the two persistence lengths in an AFM image where the ssDNA of the same length is much more coiled up than a comparable length dsDNA.

Each backbone group also carries a net negative charge of -1$e$, meaning the whole chain ends up with quite a large negative charge of -2e/bp when in the double-stranded form, though in the presence of an electrolyte in the solution the effective charge will be smaller due to a process called *Manning condensation*,[137] being generally on the order of -0.5 $e$/bp or smaller depending on the salt concentration and type.[131,138,139] The base pairing makes the resulting double strand of DNA quite chemically inert and this, together with the high electric charge of the backbone and the high rigidity, make dsDNA an ideal model polymer for nanopore experiments.

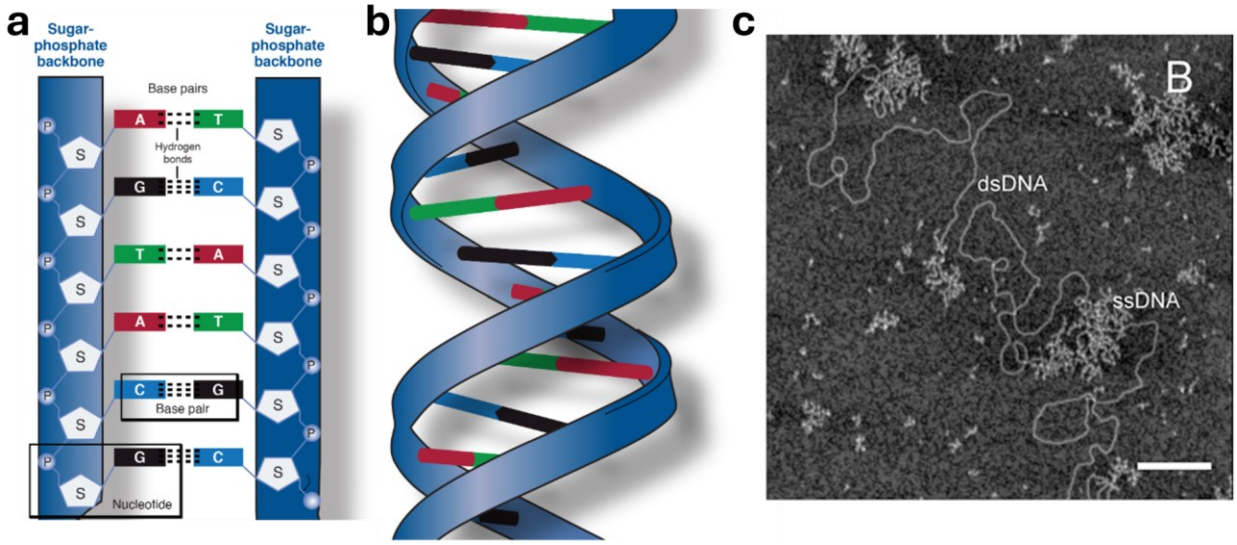



***Figure 1.8: Structural overview of DNA. a)*** *Schematic of the chemical structure of double-stranded DNA.* ***b)*** *Schematic of the 3D helical structure of double-stranded DNA. Images adapted from: https://www.genome.gov/genetics-glossary/Phosphate-Backbone* ***c)*** *AFM image of ssDNA and dsDNA deposited on mica sheets. The scale bar is 250 nm long.*[136] *© 2006 Federation of European Biochemical Societies.*

The behavior of a free-floating, isolated DNA molecule in solution largely depends on its length. Other factors like charge, chemical modifications, interactions with other DNA molecules or DNA binding proteins, as well as the presence of confining walls can affect the behavior of the DNA as well but are not considered here for now. The simplest model for a polymer in free solution is a self-avoiding random walk with step sizes on the order of the persistence length, whereas very short fragments (< ~200 bp) can be well approximated as rigid rods. Longer fragments of dsDNA (> ~200 bp but < ~ 400 kbp) need to be treated as *worm-like chains*,[140] meaning they are flexible enough to coil up but are still short enough that excluded volume effects do not need to be considered. The DNA fragments used in this study are all < 10 kbp and fall into this latter category. In free solution, these molecules exist in a random-coil conformation, and the strand is more likely to be captured by a pore somewhere along its contour than it is to be captured by either end.[79]

### 1.2.2 DNA Origami

The high sequence specificity of DNA base pairing makes it possible to use DNA oligonucleotides (short fragments of single-stranded DNA, often referred to as "oligos") as a form of "staple" to bend and fold a much larger polynucleotide "scaffold" into arbitrary shapes. This was first shown experimentally by Rothemund in 2006,[141] who termed the process "DNA origami". Rothemund showed the versatility and simplicity of this process by designing and assembling a variety of 2D shapes, but the technique has since been expanded to all kinds of 3D shapes[142] that are used for a range of applications, including for targeted drug delivery,[143] as a platform for controlled chemical reactions,[144] as building blocks for larger self-assembled structures,[145] as molds for microfabrication of solid-state materials,[146,147] and more.

The process of assembling a DNA origami structure is relatively simple. For scaffolded DNA origami (which refers to DNA origami that uses a single large "template" strand folded into shape by a large number of short "staple" strands), the process involves mixing a set of staple strands with the scaffold strand in salt solution.[148] This mixture is heated up to about 80 °C and then slowly cooled to room temperature.[149] Provided the oligo sequences are properly matched to the sequence of the scaffold strand, the final structure will self-assemble into the desired shape with high probability (the process is shown schematically in **Figure 1.9**).

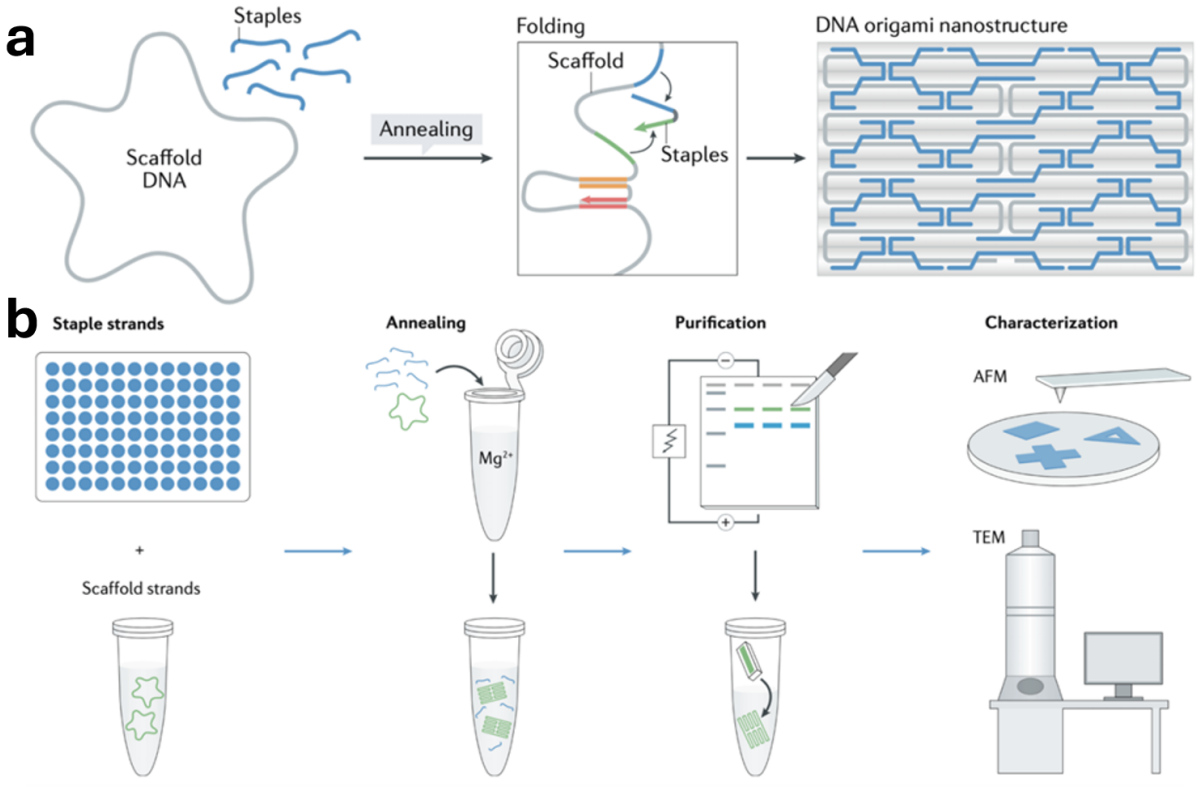



***Figure 1.9: DNA origami assembly process. a)*** *Schematic of the molecular assembly process, the short oligo staple strands fold the long scaffold strand into the desired shape as designed.* ***b)*** *The usual steps taken to assemble a DNA origami structure, the staple oligos are mixed with the scaffold strand in a buffer containing $Mg^{2+}$ ions. The mixture is purified (by e.g. gel extraction) to remove excess oligos and misfolded products and subsequently characterized with TEM or AFM to see if the structure is assembled correctly. Image reproduced from [150] under CC BY-NC-ND 4.0.*

The details of the optimal protocol with respect to the yield of the full target assembly (salt concentration, temperature profile, relative concentration of staples and templates) vary for different structures. More complex structures typically require longer assembly times, but the overall protocol remains largely the same.

There are two main ingredients needed for DNA origami. The first is a suitable scaffold strand. It is difficult to build long (> 5 knt) single-stranded DNA (ssDNA) strands with custom

sequences, and most naturally occurring DNA is double-stranded and requires additional digestion steps to be usable as a scaffold. There are a few viruses, whose genomes are naturally single-stranded, that are used in the DNA origami field as scaffolds. Most commonly used is the M13 bacteriophage, whose 7249 nt long single-stranded genome has become the default scaffold for most DNA origami applications. In recent years, efforts have been made to extend the range of suitable scaffolds both by introducing variants of the M13 genome with different lengths and by chemically synthesizing single-stranded DNA.[151,152]

The second ingredient needed for DNA origami is a set of staple oligos to correctly fold the scaffold strand into the desired shape. Oligos shorter than 200 nt with arbitrary custom sequences can readily be purchased off the shelf. The necessary sequence for those staples depends on the structure of the final assembly, and a whole suite of software tools exists to facilitate this process.[153,154] However, especially when attempting to build structures that can bind to other molecules, some manual work is still required to ensure that the design works as intended.[155] This includes double checking the proper orientation of binding sites as well as the orientation and design of the binding sequences to guarantee proper binding and minimizing steric interactions, which are usually not considered in a lot of the DNA origami design tools. For an overview of common pitfalls and best practices for the design of 3D DNA origami structures the group of Hendrik Dietz has an excellent review on their design and assembly process.[148] The assembly protocols for the structures used in this thesis are outlined in the **Appendix**.

In this thesis the M13 scaffold was used as the scaffold for all structures. The length of 7249 nt is a good compromise of being long enough to have plenty of room for DNA barcoding

but not long enough to cause problems during handling (long strands of DNA are prone to rupture during pipetting and handling) and nanopore translocation. It also is one of the few scaffold strands that is readily commercially available. The staple oligo sequences are generated by using the software tool caDNAno[156] and commercially synthesized by Integrated DNA Technologies (IDT).

### 1.2.3 DNA as a Carrier for Biomolecules

Evolution has provided a whole host of molecular machinery, such as proteins, to interact with DNA on various levels and to perform multiple functions necessary for life. Unlike DNA, the net charge and charge distributions of proteins vary with their sequence of amino acids. As a result, native proteins are often difficult to detect using a nanopore since their diffusion is often fast and their response to an externally applied field varies.[157] One approach to facilitate protein detection with nanopores, or more generally biomarkers, is to attach these proteins to a highly charged and well-understood DNA carrier that defines their passage dynamics and facilitate their detection.[115,158] In principle, the presence of a protein or other biomolecule on a DNA strand should be readily detected by a nanopore as a transient blockage that exceeds that of the DNA carrier strand alone[159] (see **Figure 1.10**).

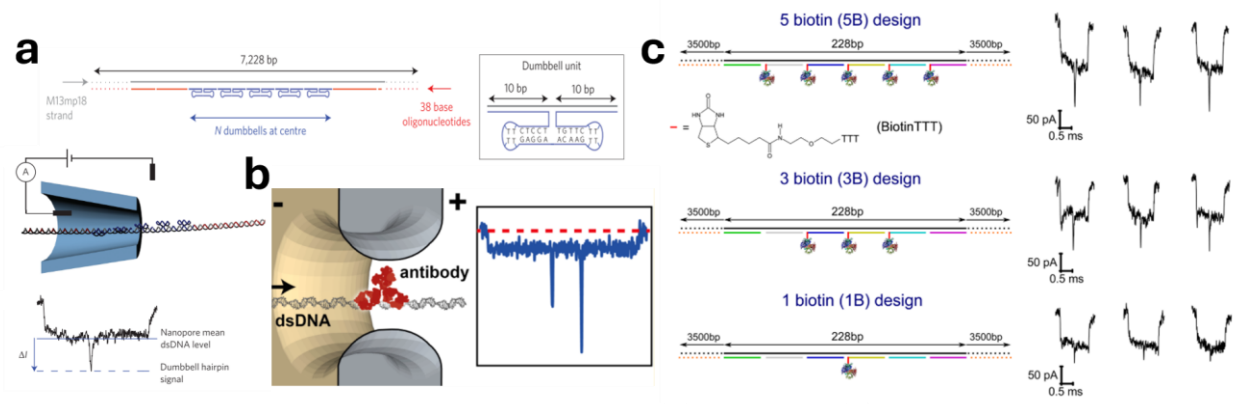



***Figure 1.10: Biomolecules attached to DNA. a)*** *DNA dumbbells formed by self-interaction of oligos on a DNA carrier and its impact on the current trace in a glass nanopipette nanopore.*[160] *Reproduced with permission from Springer Nature.* ***b)*** *An antibody attached to a DNA carrier via biotin-streptavidin interaction translocating through a thin solid-state nanopore and the resulting current trace. Reproduced with permission from*[91] *Copyright 2015 American Chemical Society.* ***c)*** *Proteins attached to various positions on a DNA carrier via biotin-streptavidin interactions, with sample events of translocation through a glass nanopipette.*[161] *Copyright © 2015 American Chemical Society. Licensed under CC BY 4.0.*

While this approach can be used to detect proteins that naturally bind to DNA,[113,162,163] it can also be extended to detect other proteins[161,164] and even other molecules (like DNA origami structures,[85,160] RNA,[12,165,166] antibodies,[91] heavy metal ions,[167] etc.). All that is required is that one can create a binding site on the DNA to which the target in question can attach.

Using DNA as a carrier for biomolecules like this enables many applications using nanopores that would not otherwise be possible. For example, this technique can be used to perform multiplexed detection, i.e. detect multiple different biomolecules at the same time, by binding them at different positions on the DNA carrier. Because the carrier is a linear piece of DNA, this gives the translocating molecules attached to the DNA carrier an intrinsic order during translocation, which allows one to infer the presence or absence of certain attached molecules.[168] This concept is known as DNA nanostructure barcoding and has been successfully demonstrated

by the group of Ulrich Keyser about a decade ago.[160] This should not be confused with DNA barcoding, which uses the DNA sequence to encode information. Here the information is encoded in the structure of the DNA and read by the nanopore sensing not by sequencing.

### 1.2.4 DNA as a Data Storage Medium

The idea of storing digital data in DNA is not new and early proposals of DNA data storage go back to the 1960s.[169] There are two main approaches to DNA data storage. The first one is to use the DNA chemical sequence as the storage medium, while the second approach is using the structure of the DNA to store information.[170,171] A schematic overview of the different strategies is shown in **Figure 1.11**.

DNA sequence-based data storage has the obvious advantage of high data density. Experimentally, a data density of up to $17\ \mathrm{EB/g}$ ($1\ \mathrm{EB} = 10^9\ \mathrm{GB}$) was achieved[172] while the theoretical density lies at about $455\ \mathrm{EB/g}$.[173] Through the addition of artificial nucleotides and modification of existing nucleotides the storage density can be improved even further.[174,175] However, while DNA sequencing has become dramatically faster and cheaper, writing data remains the bottleneck. Indeed, de novo synthesis of custom sequences is costly[176] and limited to relatively short strands,[177] so data must be fragmented across a large pool of oligonucleotides. Retrieving a specific file from such a pool typically requires selective amplification followed by sequencing, and once synthesized, the stored sequence cannot easily be modified.[178]

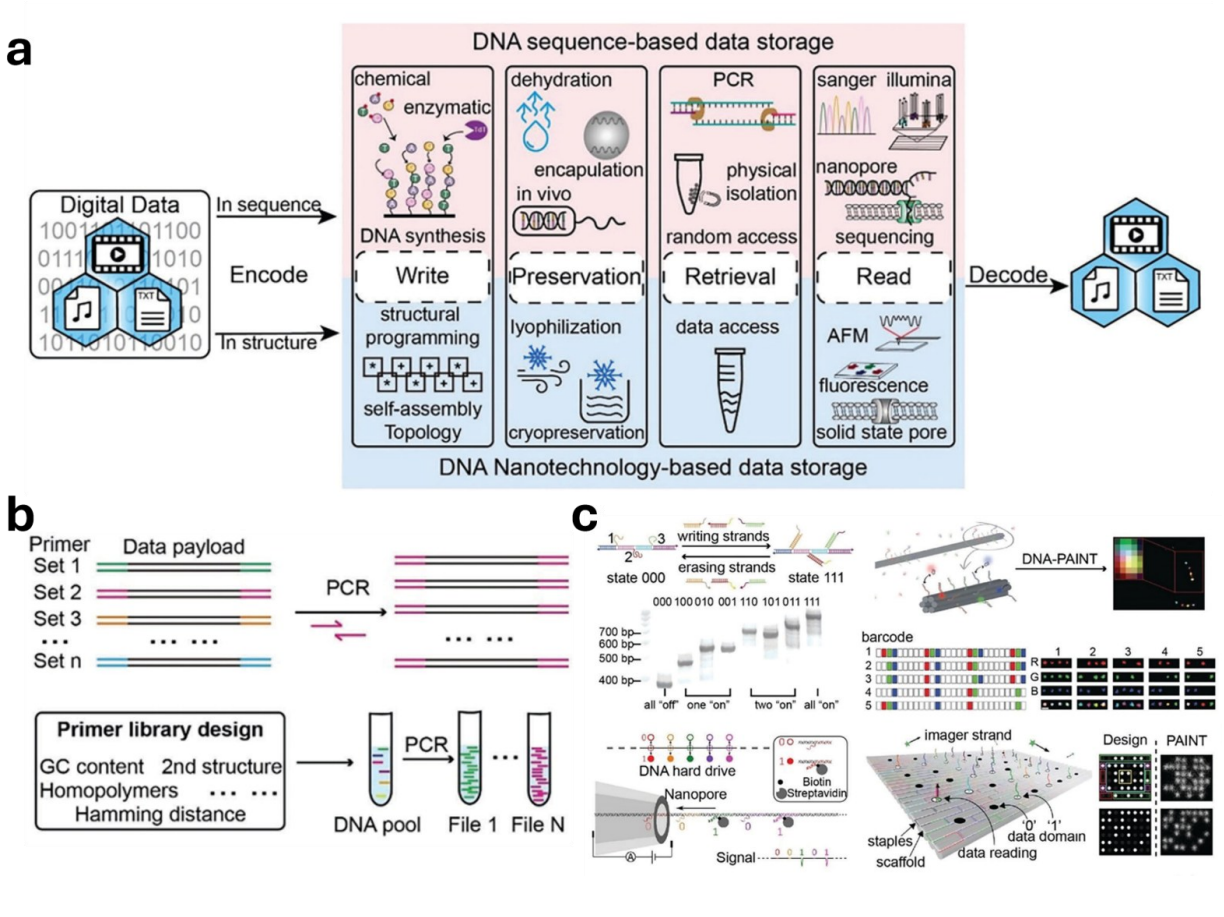



***Figure 1.11: Overview of DNA data storage schemes. a)*** *General workflow scheme for DNA data storage describing the steps required for sequence-based data storage and structure-based data storage.* ***b)*** *Simplified workflow for sequence-based data storage.* ***c)*** *Some examples of structure-based DNA data storage schemes with chemical, optical, and electrical readout mechanisms.*[170] *Copyright © 2023 Wiley-VCH GmbH.*

DNA structure-based data storage can take a variety of different forms with different readout mechanisms. Generally, this approach consists of a DNA scaffold structure (either a dsDNA strand or a DNA origami structure) with attachments at predefined positions that represent individual bits. Because data is written by assembling the scaffold with a reusable library of pre-synthesized components rather than by synthesizing a new sequence for each dataset, writing is simpler and less expensive. The readout of those bits can happen electrically[26] as ionic-current

signatures during nanopore translocation or optically, for example by AFM or super-resolution microscopy.[179,180] The advantage of adaptability to the specific readout mechanism comes at the cost of data density but can offer more robust storage, because each bit is a nanoscale feature formed by many base pairs, which offers more resilience to single point defects caused by e.g. radiation damage.[181,182]

### 1.2.5 DNA Nanostructure Barcodes

In a DNA nanostructure barcoding scheme, small biomolecules are attached to specific sites along the backbone of a long DNA carrier. The identity of a target molecule in a particular position then creates a "barcode" that is modified by the presence or absence of a given target molecule. In a nanopore translocation event, any molecules attached to the DNA carrier will generate a sequence of transient blockage "spikes" that can be detected either directly in the current trace,[183] in a parallel optical measurement,[184] or using a combination of the two.[185,186] The simplest interpretation of the barcode is to assign a numeric code to the sequence of spikes: 1 for the presence and 0 for the absence of a given molecule or structure on the carrier. Other encoding schemes involve binding multiple different molecules or DNA origami structures with different sizes or shapes that generate distinguishable spikes in the nanopore current trace. Experimentally, binary,[27] ternary,[187] quaternary,[188] and higher order[111] encodings have been demonstrated. The nature and size of the attachment can vary greatly and ranges from small intercalating molecules like dyes,[113] peptide nucleic acids (PNAs),[189,190] small proteins,[26,161,191,192] small DNA protrusions,[27,160,168,193] larger DNA structures,[187,188,194,195] DNA secondary structure motifs,[196] and DNA-RNA hybrid structures.[12,111] A variety of different attachments for the purposes of barcoding are highlighted in **Figure 1.12**.

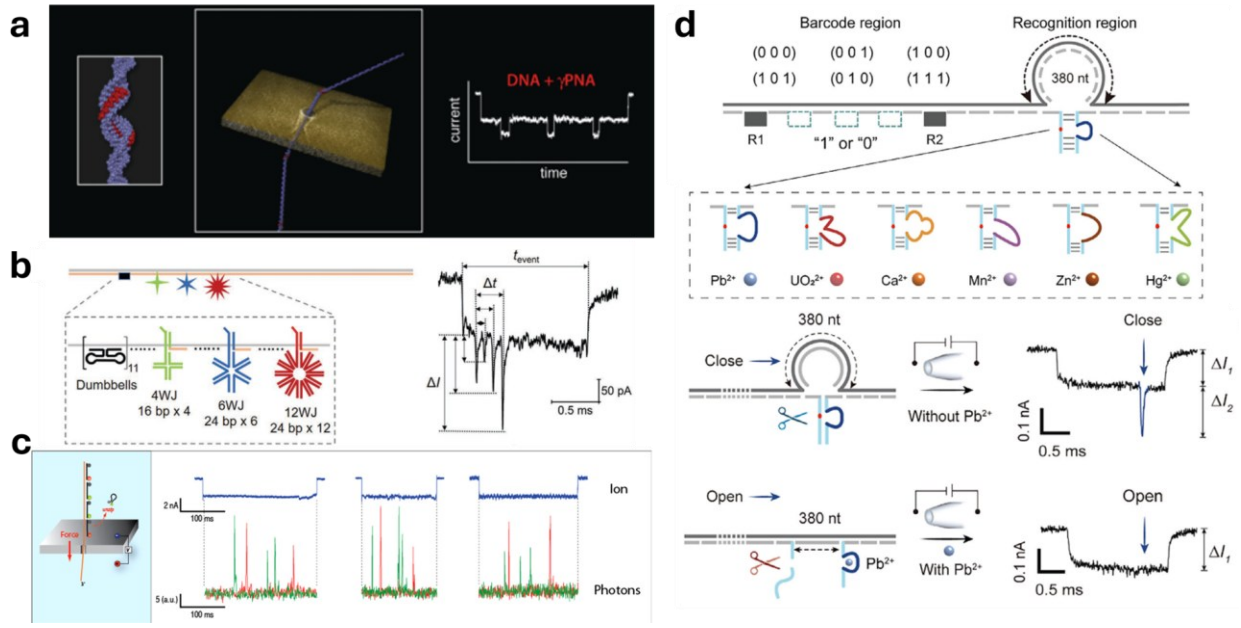



***Figure 1.12: Nanopore barcoding schemes. a)*** *Intercalating PNA used to generate extra blockage spikes in the ionic current traces. Reproduced with permission from [189] Copyright 2015 American Chemical Society.* ***b)*** *Small DNA origami stars used to generate different height peaks in the current trace for multi-level barcoding.[188] Licensed under CC BY 4.0.* ***c)*** *Fluorophores attached to a DNA backbone for barcoded detection of DNA fragments with different sequences, measured optically upon unzipping during passage and unquenching of the FRET pair.[184] Copyright © 2014 American Chemical Society.* ***d)*** *Small DNA protrusions used as a barcode for multiplexed detection of heavy metal ions via a circular detection mechanism.[167] © 20026 Wiley-VCH GmbH*

The application space for DNA barcoding is as diverse as the range of strategies used to build them, and the application dictates design of the optimal barcoding scheme. While most applications of DNA barcodes aim at multiplexed detection of a selection of target analytes,[183,197] others have been pursued, including digital data storage in DNA,[27,188] structural investigation of DNA polymers,[94,118,119,123,124] and the study of protein-DNA interactions.[157,163] Nanostructured barcoding schemes using carriers other than DNA have also been proposed.[16,18]

One of the more ambitious applications of these schemes is the development of digital data storage systems using these DNA barcodes which can then be read by a solid-state nanopore.

Modern digital data storage systems are quite energy intensive, expensive and short-lived.[176] The solution of using molecules to store information is especially applied for archival data applications. In those applications data is rarely accessed but must be migrated every few years to remain readable and correct accumulated errors.[198,199] Long-term storage of digital data is therefore often still done using older technologies like magnetic tape that are cheaper to store and do not require electric power just to stay in a readable format. These solutions tend to be very space intensive, however, which becomes problematic with today's ever-increasing amount of digital data produced. This is where DNA data storage schemes step in to provide a much cheaper and longer-lasting solution. DNA is stable in a wide variety of environments with minimal maintenance requirements[200] making it an attractive prospect for long-term digital data storage with a high potential data density of up to $4.5 \times 10^{11}\mathrm{GB/g}$.[173] While the field is exploring the sequence of synthetic DNA molecules to store information, the DNA nanostructured barcodes as described above are also being considered to harness DNA's data storage potential.[168,188]

## 1.3 Thesis Overview

For all these applications involving the passage of a DNA carrier through a nanopore, it is important to understand the dynamics of polymer capture and translocation and the interactions between the pore and the barcoded structure, in order to improve performance. In this thesis, I conduct a detailed investigation of the influence of molecular design choices and pitfalls to consider when building a nanostructured DNA barcoding scheme decoded by solid-state nanopore sensors. Other groups have recently characterized the performance of DNA carriers in solid-state pores with respect to resolution of their nanopore setup,[80,201] size and flexibility of the attachment

used for barcoding,[202] and simulations of molecular behavior inside the pore.[159] For a complete picture of how peaks are generated in the current traces, my aim is to design a fully resolvable and adaptable DNA origami-based design for DNA structural barcoding with the ultimate goal of digital data storage. This means designing a DNA structural barcode that not only resolves fully in the pore but ideally is adaptable on the fly by writing and rewriting individual bits, potentially also enabling DNA computing techniques to be readable using solid-state nanopore sensors. A robust design that can be fully resolved in a wide range of nanopore sizes enables classification of barcodes on a single molecule level and allows for easy transition into a mixture of barcodes without having to pre-sort events into different categories.

To that end, **Chapter 2** studies the dynamics of DNA origami scaffolds in solid-state nanopores and shows that these behave in principle like regular dsDNA fragments. It investigates the dynamics of scaffold strands with varying persistence lengths to show that nanopore events produced by dsDNA scaffolds are the most well behaved and easiest to interpret, making it an ideal candidate for structural DNA barcoding. **Chapter 3** investigates the requirements that barcoded designs need to fulfill to generate robust and easily distinguishable barcodes in solid-state nanopore current traces, while **Chapter 4** highlights the challenges in scaling these barcoded designs to larger numbers of bits and more robust classification. Finally, **Chapter 5** shows the most recent yet to be published results in the lab and highlights the next steps to push this DNA structural barcoding scheme to actual applications.

# 2 DNA Dynamics and Capture

This chapter is published in *Macromolecules* (58, 4094-4102, **2025**) as "Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-State Nanopores" authored by **Philipp Mensing**, Martin Charron, Mohamed Yassine Bouhamidi, and Vincent Tabard-Cossa.[*]

---

[*] Reproduced from Mensing, P.; Charron, M.; Bouhamidi, M. Y.; and Tabard-Cossa, V.; "*Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-State Nanopores*", **Macromolecules** (58, 4094-4102, **2025**) Copyright 2025 American Chemical Society.

## 2.1 Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-state Nanopores

*Philipp Mensing, Martin Charron, Mohamed Yassine Bouhamidi, and Vincent Tabard-Cossa*
*Department of Physics, University of Ottawa, Ottawa, Canada*

### 2.1.1 Abstract

Understanding the intricate details of the molecular mechanisms behind nanopore translocation is crucial to further the development of nanopores as single molecule sensors. In this study we assemble hybrid ss-dsDNA constructs with different effective persistence lengths to investigate their capture and folding behavior in solid-state nanopores. In general, we observe differences in event shapes for different flexible polymers with more flexible polymers leading to deeper, longer, and more complex events. We show that folding happens favorably in ssDNA segments along the heterogeneous polymer contour and can happen multiple times during the same event, blocking more current and generating more convoluted event shapes. By designing DNA heterostructures with various ss-dsDNA distributions, we further show that the propensity for folds at the ssDNA segments arise due to their higher flexibility allowing them to diffuse around the vicinity of a nanopore during the translocation process, as opposed to stiffer and non-relaxed dsDNA segments through which the tension propagates during passage. These findings highlight the different translocation regimes of polymers of various rigidity, improving our understanding of the folding of polymers into nanopores which will ultimately be useful for the development of applications requiring single-file translocations, for example, for the decoding of barcoded

nanostructured molecules for multiplexed biomarker detection or for molecular information storage schemes.

### 2.1.2 Introduction

Nanopore sensing is an active field of research owing to the simplicity of its working principles, and to the central role it played in achieving single-molecule DNA sequencing in a low-cost, portable format.[36] A recent promising sensing mode for the multiplexed detection of targets relies on DNA carrier-assisted sensing, wherein analytes of interest are bound at specific receptor sites along the contour of a long dsDNA backbone and electrically driven through a nanopore. Analyte identities and their binding position along the carrier can be inferred from nanopore ionic current signals, given that their presence inside the nanopore leads to sub-levels in the event signatures deeper than those produced by the single-file passage of the carrier. This technique has found applications for barcoded protein sensing,[158] DNA information storage,[188] and others.[119]

Many of these applications rely on polymers passing through nanopores in a single-file manner, wherein only one segment of the polymer is present inside the pore at any point during translocation, for accurate molecular identification. This is complicated by the fact that to have smooth unhindered translocations of the carrier, receptors and bound analytes, pore sizes need to be large, an experimental condition which unavoidably promotes folded carrier translocations, wherein at some point during its passage, distinct segments of a polymer are inside the pore at once, as recognized by different quantized blockage steps within a single blockage event (**Figure 2.1**). Moreover, longer polymers, which appealingly present carriers with more binding capacity, are also known to result in more folded translocations.[203] For most applications, folding of DNA

into nanopores is undesirable and efforts have been made to limit or remove folding through various means.[81,203] Folding is especially problematic when carriers designed with specific sequences of attachments, so-called barcoded designs,[202] are being used, as it adds extra complexity to the data analysis and can make classification of events on their linear barcodes challenging or even impossible.[204] Folding of DNA in solid-state nanopores has been investigated before both experimentally and theoretically,[79,94,122,205–208] yet the dynamics of polymer passage through a nanopore is complex, and a better understanding of the folding dynamics would help implement improved sensing schemes.

In this work, the passage of different linear DNA structures made up of interspaced single- (ss) and double-stranded (ds) DNA segments is studied, which allows us to infer the role of rigidity on polymer translocation and folding dynamics. The structures were assembled by binding 181 complementary 40 nucleotide (nt) long staple oligomers to a 7249 nt long scaffold based on the M13mp18 bacteriophage genome, where the periodic omission of oligomers during assembly resulted in ssDNA-dsDNA heterogeneous structures (**Figure 2.1a**). We first investigated the folding kinetics of the simplest of these structures, consisting of the 7249 nt scaffold bound to the entire set of 181 complementary staple oligonucleotides, i.e., forming a fully double-stranded polymer. Comparing the folding kinetics of this molecule to that of a plasmid dsDNA fragment of a similar length (enzymatically digested and chromatography-purified), and thus ligated, allowed us to establish the impact of the nicks along the assembly backbone, and to assess the quality of the assembly process. Second, we characterized the folding dynamics of DNA polymers of different flexibility by assembling linear structures in which some of the staple oligonucleotides were left out at defined and regular intervals, thereby introducing 40 nt long single stranded

sections of reduced rigidity along the DNA polymer. The spatial control of the flexibility of the polymer is used as a model system to provide insight into the folding characteristics of DNA carriers used in barcoded sensing schemes.

### 2.1.3 Results

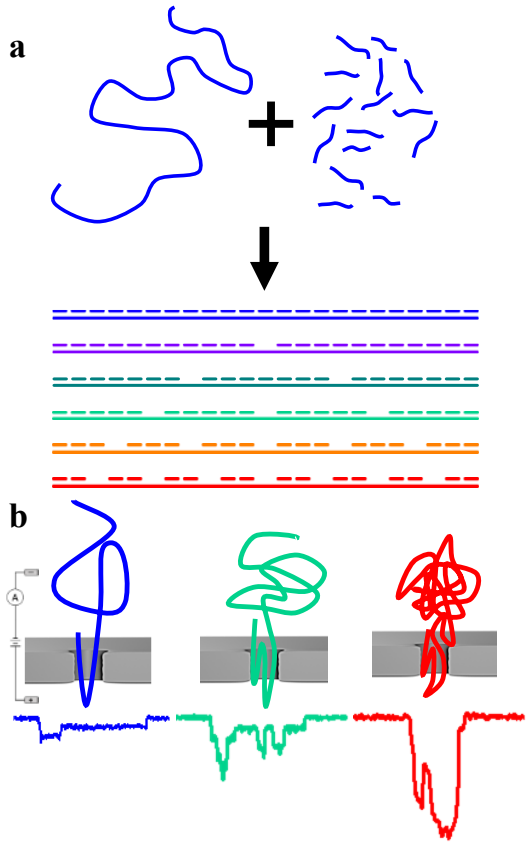



***Figure 2.1: Semiflexible structures overview. a)*** *Schematic of the assembly process generating linear DNA structures with incorporated flexible single stranded sections at different intervals. The scaffold strand is the 7249 nt long M13 genome, the added staple oligos are 40 nt long. ssDNA sections in the final assembly are generated by targeted omission of specific oligos in the mixture of staple strands at designated positions.* ***b)*** *Translocation of linear DNA assemblies through solid-state nanopores, with representative current signatures. Nanopore experiments are performed in 3.6 M LiCl at pH 8 with pore sizes ranging from 8 nm to 16 nm.*

#### 2.1.3.1 Influence of Nicks on DNA Translocation

Before investigating the effect of spatially varying the flexibility of the DNA carrier on the folding characteristics, we first performed control experiments to validate the quality of our assembly process and study the influence of nicks present along the assembly's backbone. For this purpose, we compared the translocation characteristics of commercially available, purified 7 kilobase-pair (kbp) enzymatically digested plasmid dsDNA fragments with our fully assembled linear structures. Our linear constructs were assembled by mixing a linearized M13 scaffold (7249 nt) with complementary oligonucleotides to form a complete dsDNA strand with nicks present along its contour due to the deliberate omission of a ligation step (see Methods). The comparison of signals from linear double-stranded M13 (LdsM13) and 7 kbp dsDNA translocations serves a two-fold purpose: Single-file translocations help assess the quality of the assembly process, whereas folded translocations help investigate the influence of nicks on folding dynamics.

**Figure 2.2a** displays the current traces of multiple 7 kbp (red) and LdsM13 (blue) translocation events obtained using an 8.4 nm pore in a 3.6 M LiCl solution, under a bias of 300 mV. Both single-file and folded events are observed, as identified by the single and dual blockage states in the event signatures, respectively. Individual events were fitted as a succession of blockage levels, and the total translocation duration $\tau$ and maximum blockage level $\Delta I_{max}$ were extracted, as plotted in **Figure 2.2b**. Both molecule types show quantized subpopulations centered around identical $\Delta I_{max}$ values, with the lowest blockage state ($\Delta I_{max} = 1.7$ nA) corresponding to single-file translocations and deeper blockages ($\Delta I_{max} = 3.4$ nA) corresponding to folded translocations.

**Figure 2.2c** plots the translocation time histograms for single-file events. The translocation time distributions of 7 kbp (red) and LdsM13 (blue) differ significantly: 7 kbp dsDNA events show tightly distributed translocation times with mean and standard deviation values of $\bar{\tau} = 264 \pm 37$ μs, whereas LdsM13 events show a similar peak value of $\bar{\tau} = 296 \pm 42$ μs in addition to a streak of shorter translocation times spanning down to the current amplifier's resolution limit of $\approx 5$ μs. Note that the slightly longer $\bar{\tau}$ value for LdsM13 can be attributed to its longer contour length (7249 bp vs 7000 bp). Given that the $\Delta I_{max}$ values of the shorter LdsM13 events are consistent with that of dsDNA, we suggest that they originate from the presence of shorter linear assemblies present in solution. This is further supported in Section S2 of the Supporting Information (SI) which demonstrates that a fraction of the M13 scaffold is randomly cut prior to the assembly process and that the addition of a localized cut by the restriction enzyme results in a wide distribution of partial LdsM13 lengths.

We next study how LdsM13 assemblies fold inside nanopores compared to 7 kbp dsDNA to assess whether nicks along the contour of DNA assemblies alter folded translocation kinetics. Using folded translocations only ($3\ \text{nA} < \Delta I_{max} < 5\ \text{nA}$), we first digitally filtered LdsM13 events to remove fragmented LdsM13 assemblies to provide a suitable comparison of similar-length linear polymers. The filtering method is illustrated in **Figure 2.2d** and makes use of the strong correlation between folded state duration $\tau_{fold}$ and total event duration $\tau$ [119]: Only events for which the condition $\tau > (200\ \mu\text{s} - \tau_{fold})$ holds true were considered as fully assembled LdsM13. See Section S2 of the SI for further details. **Figure 2.2e** displays an overlay of all the normalized selected events to show consistency of event shapes and highlight the folding fraction extraction method. The comparison of folding kinetics between LdsM13 and 7 kbp dsDNA was

achieved by calculating the folded state survival probability $S_{fold}(t)$, defined as the probability that an event is still in a folded state at normalized time $t \in [0,1]$. See Section S3 for details about $S_{fold}(t)$ measurements. **Figure 2.2f** plots the $S_{fold}(t)$ distributions calculated from folded 7 kbp dsDNA (red) and LdsM13 (blue) translocations, where both distributions are shown to be nearly identical. Given that the folded state duration is closely correlated to the folding location, i.e. the location along its contour at which a polymer folds inside a pore, **Figure 2.2f** indicates that the presence of nicks along a polymer does not alter folding kinetics. Note the inset of **Figure 2.2f** depicts the $S_{fold}(t)$ distribution from unfiltered LdsM13 events (**Figure 2.2d**), which differs from 7 kbp dsDNA and highlights the importance of filtering shorter fragments.

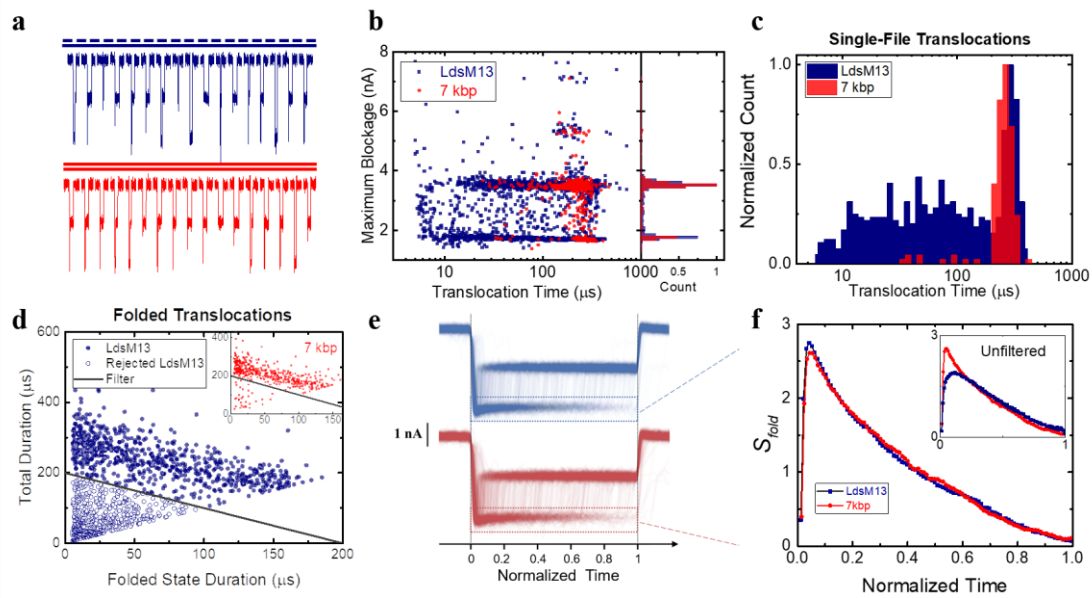



***Figure 2.2: Comparison of nicked dsDNA to ligated dsDNA. a)*** *Example traces of nicked LdsM13 (blue) and 7 kbp double-stranded DNA (red). Experiments performed in 3.6 M LiCl 10 mM HEPES pH 8 with an 8.4 nm pore at 300 mV.* ***b)*** *Maximum fitted blockage vs. translocation time for all events recorded for the two DNA polymer types.* ***c)*** *Translocation time distribution for only the single file events as identified in b).* ***d)*** *Definition of the filter used to compare folding statistics. Empty squares (below the $\tau > (200\ \mu s - \tau_{fold})$ condition) are filtered out of the analysis while solid squares are kept.* ***e)*** *Event overlay of all selected events from d).* ***f)*** *Survival probability of folding with event time for only the selected events. The inset shows the same plot for all events.*

### 2.1.3.2 Blockage Amplitudes of ssDNA-dsDNA Heterostructures

To investigate the effect of overall polymer rigidity on the capture, translocation and folding processes of linear polymer structures through solid-state nanopores, we now spatially vary the flexibility of the LdsM13 molecules by introducing 40-nucleotide long ssDNA segments at different locations along their contour (**Figure 2.3a**). This was achieved by simply omitting a selection of evenly distributed oligomers during the structure assembly. Naturally, structures with more ssDNA segments have lower overall rigidity, due to ssDNA having a significantly lower persistence length than dsDNA.[209,210] Throughout this work, we refer to the different dsDNA-ssDNA heterostructures by how many oligos are bound versus not bound. For example, the 5ds-1ss assembly has a pattern of 5 bound oligos and then 1 oligo missing as a unit repeated 30 times along the whole molecule. Since the oligos are 40 nt long, the 5ds-1ss structure is therefore made up of 200 bp dsDNA segments interspersed with 40 nt ssDNA segments.

**Figure 2.3b** shows current traces of translocations from eight heterostructures with different inherent rigidities ranging from that of dsDNA to that of ssDNA. Results were obtained in 3.6 M LiCl pH 8 under a bias of 100 mV to 400 mV in pore sizes of 8.3 nm to 15.7 nm (See **Table 2.3** in the SI for individual pore details). The data for ∞ds (purple traces) correspond to blockages from the LdsM13 structure discussed in the previous section, i.e. the nicked dsDNA fully assembled structure, whereas the ensuing six traces correspond to heterostructures with progressively more oligomers missing: 60ds-1ss (blue), 30ds-1ss (teal), 11ds-1ss (turquoise), 5ds-1ss (green), 3ds-1ss (yellow) and 2ds-1ss (orange). Finally, the data for ∞ss (red traces) correspond to translocations of linearized M13 ssDNA strands. Qualitatively, current traces in **Figure 2.3b**

show that heterostructures with higher ssDNA content (more missing oligos) lead to events with deeper and more complex blockage signatures with less observable quantized folded sub-levels.

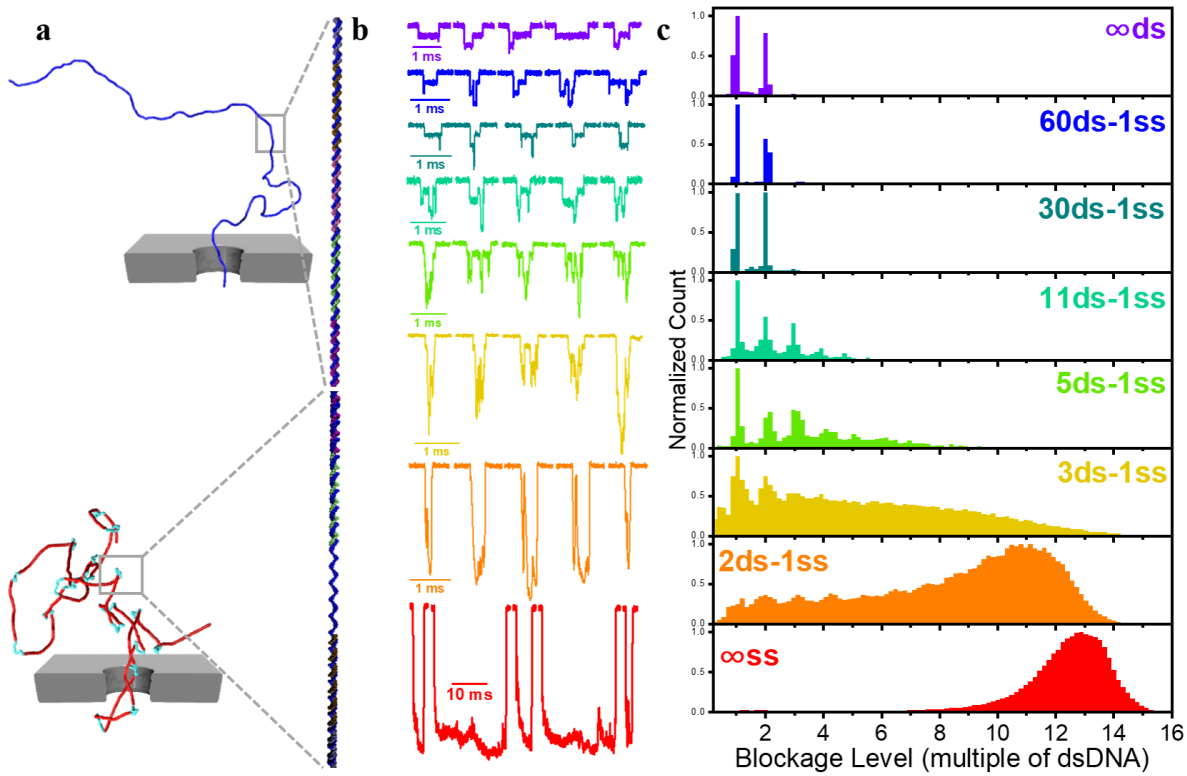



***Figure 2.3: Nanopore analysis of flexible and semiflexible DNA. a)*** *Schematic of the translocation process for fully double-stranded DNA (LdsM13, top) and DNA with single stranded sections (bottom). The more flexible polymers adopt a more compact blob shape for which each repeating unit is closer to the pore, which increases the probability of capture of multiple sections of the polymer at the same time.* ***b)*** *Example traces for the different tested flexible assemblies. The colors match the data in c). All experiments are performed in 3.6 M LiCl at pH 8 with voltages between 100 mV and 400 mV applied. Y-scale is normalized by the single level blockage to facilitate comparison of all events displayed.* ***c)*** *All-points histogram of levels fitted to events generated by different flexible assemblies. The flexibility of assemblies increases from top to bottom. Blockages are normalized by the blockage level of the single file blockage in each case (for the ∞ss, i.e. linearized single stranded M13, we added 2 kbp dsDNA during the experiment to normalize it).*

To quantify this observation, individual events were fitted to sequences of blockage states using a custom implementation of the CUSUM+ algorithm,[211] from which a histogram of fitted blockages was obtained and displayed in **Figure 2.3c**. Note that the fitted blockages were normalized by the single-file dsDNA blockage measured for each experiment to facilitate interpretation and calibrate for pore-to-pore variations. **Figure 2.3c** shows that for structures ranging from ∞ds (fully dsDNA) up to 5ds-1ss, increasing ssDNA content results in the presence of higher quantized blockage levels that line up perfectly with the expected blockages for three, four, and even five dsDNA strands in the pore at the same time. Interestingly, as the ssDNA content keeps increasing, as per the three most flexible structures studied here (3ds-1ss in yellow, 2ds-1ss in orange, fully ssDNA in red), translocation events generally show long and deep blockage levels without any clearly discernible internal structure, resulting in wide distributions of blockage states not tightly distributed around any multiples of single-file blockages. We note that no blockage states corresponding to single ssDNA strands stretched inside the pore (half-integer multiples of dsDNA blockages) are observed across all the heterostructures.

The results for flexible structures in **Figure 2.3c** agree with prior work[212] that reported significantly deeper blockages for M13 ssDNA translocations compared to dsDNA and reasoned that ssDNA blockades resulted not solely from strands inside the pore, but also from bundled ssDNA segments blocking current in the access region outside the channel. We suggest that **Figure 2.3c** supports this interpretation and iteratively shows the natural progression of the effect of rigidity on the blockage amplitude of linear polymers with identical contour lengths: For very rigid polymers such as dsDNA, only the strands inside pores contribute to the current blockage due to polymer segments outside the pore being far away. However, as ssDNA content and overall

flexibility increase, polymer radii reduce and as such segments outside the pore tend to be closer to pores and contribute to the current blockage amplitude. To support this, sections S6 and S7 demonstrate the drastic effect of increased ssDNA segments and flexibility on polymer radii and monomer distributions near nanopores through statistics acquired on simulated polymer conformations.[213]

Interestingly, the complex blockage states of more flexible polymers show no clear time-dependence which, following the above interpretation, indicates monomers are in the vicinity of pores throughout the entire translocation process. The constant proximity of polymer segments to pores suggests different translocation dynamics to those of more rigid polymers like dsDNA whose conformations are essentially fixed in space during the translocation process until its segments sequentially undergo tension and start moving towards the pore, as described by tension propagation principles.[60,118,214,215] **Figure 2.3** thus suggests that the translocation dynamics is significantly affected by a polymer's rigidity. As such, we now aim to obtain better insight into the origin of the more frequent deeper and complex observed blockades by analyzing the folding statistics of different heterostructures.

#### 2.1.3.3 Folding Kinetics of ssDNA-dsDNA Heterostructures

The intermediate range (5ds-1ss, 11ds-1ss, 30ds-1ss) of heterostructures studied in **Figure 2.3c** creates a variety of highly structured events that contain a high amount of information about the underlying physics of the translocation process. This is a direct consequence of the introduction of an increasing number of ssDNA sections along the contour of the molecule. We first note that all folded states with blockage amplitudes $\Delta I$ deeper than $2\Delta I_{dsDNA}$, i.e. two dsDNA strands inside

the pore, arise when a separate segment of the same molecule gets close enough to the pore to get captured alongside the already translocating strand(s). When the segment finding the pore is not a polymer end, it folds inside the pore and temporarily introduces two new dsDNA strands into the pore. **Figure 2.4a** shows the simplest folding case of such type consisting of one strand or segment already inside a pore, resulting in a temporary blockage level three times as deep as the level for one dsDNA strand. When the polymer is initially captured by an end, we identified those events as a 1-3-1 event (representing the sequence of quantized dsDNA blockades during the event). **Figure 2.4b** shows the current trace of a more complex 2-1-3-1-3-1 translocation event containing two such intra-event folding states. The $3\Delta I_{dsDNA}$ blockage states are more commonly featured in semi-rigid assembly events compared to fully dsDNA polymers, especially for the intermediate cases of 30ds-1ss, 11ds-1ss, and 5ds-1ss.

To better illustrate why flexible polymers result in more complex sequences of folded states, we now attempt to locate which segments along the polymer are more prone to be captured and fold inside the pore during the translocation process. To do this, we investigated all translocation events that ended with a 1-3-1 sequence of blockage levels. As discussed above and shown in **Figure 2.4a**, the ends of such events are easily physically interpreted and as such the mapping of the timeseries data to the molecular positions becomes relatively straightforward. To estimate the location of the last capture and folding point along the polymer $x$, we used the truncated electric charge deficit (TrECD)[204] measurements which are defined as the area under the event with respect to the open-pore current, i.e. $TrECD = \int_{t1}^{t2}(I(t) - I_o)dt$, and represent the effective ionic charge that molecules block from going through the pore (red area in **Figure 2.4b**) during a given time interval. Here, TrECD is used as a proxy measurement for the amount of DNA

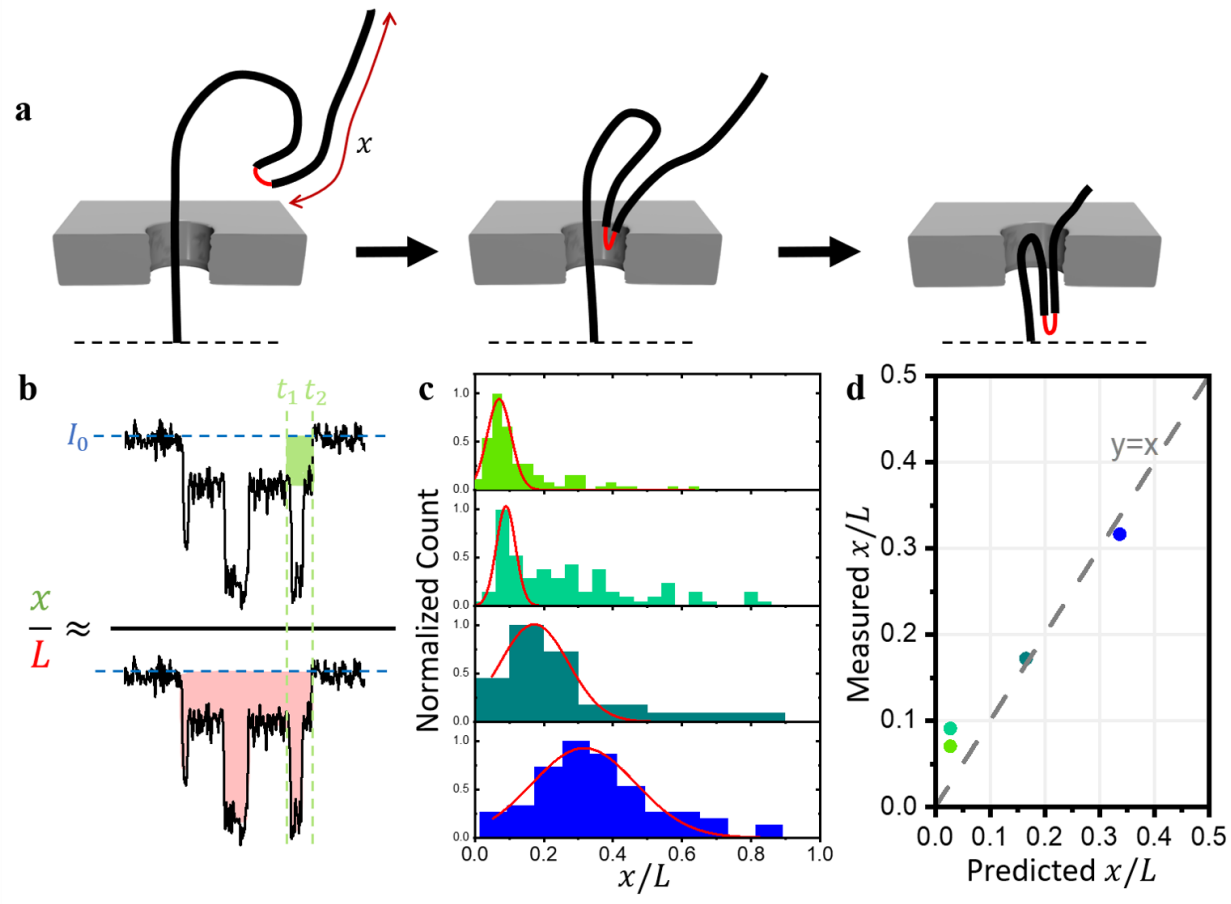



***Figure 2.4: Structural analysis of flexible DNA in nanopores. a)*** *Schematic progression of a translocation leading to a 1-3-1 strand folding pattern. The red line schematically represents a single stranded section along the double-stranded polymer in black.* ***b)*** *Current trace of a translocation event ending with 1-3-1 states. Experimental determination of the fractional distance $x/L$ from the events. The truncated electric charge deficit (TrECD) of the single level from when the secondary capture of a folded segment of the polymer enters (three strands) to the end of the event (green) over the total ECD of the whole event (red).* ***c)*** *Histograms of all fractional distances $x/L$ of the last instance of 3-1 transitions. The red curve is a single peak Gaussian fit.* ***d)*** *Measured fractional distance, extracted from the peaks of the fits in c), vs. theoretical fractional distance (as determined by the molecular position of the last ss section programmed during molecular assembly). Note that the position of the last ss section is the same for 5ds-1ss and 11ds-1ss. See* ***Table 2.2*** *in the Supplementary Information for details.*

contour length having passed through the pore[11,216] as this allows folded events to be taken into account. To get the TrECD of the final linear section of the polymer before the end of the event, we take the last blockage level of an event ending in a 1-3-1 and extend it to the beginning of the

3 strands level (green box in **Figure 2.4b**). Dividing this TrECD by the total ECD of the whole event gives us a relative ECD that scales with the fractional position of the last capture point along the polymer, $x/L$, for a polymer with a contour length of $L$. **Figure 2.4c** shows the distributions of $x/L$ values calculated from fractional TrECD values for intermediate assembly translocations (30ds-1ss, 11ds-1ss, and 5ds-1ss) ending in 1-3-1 blockage states, and the corresponding single peak Gaussian fits used to calculate the average $x/L$ value. Interestingly, the fractional distances calculated in **Figure 2.4c** correspond closely to the physical fractional position of the last ssDNA section on the polymer. This is illustrated in **Figure 2.4d** which plots the estimated $x/L$ against the predicted $x/L$ values expected from each heterostructure design and exhibits a linear dependence for the last capture point along the polymer with the position of the last ssDNA, indicating that ssDNA sections are more likely to be captured during translocation events than dsDNA sections. We note lastly that the duration of the $3\Delta I_{dsDNA}$ levels does not seem to correlate with the distance between two ssDNA sections along the contour indicating that the end of the triple folded level most likely is not located at a ssDNA section (See Supporting Information **Figure 2.12**).

Extrapolating from this observation for the 1-3-1 event case study, we further suggest that the complex folding and blockage sequences observed for heterostructures in **Figure 2.3** are caused by different ssDNA segments entering the pore throughout the translocation process. This is partly facilitated by the reduction in radius of gyration for the more flexible structures (See **Figure 2.10** and **Figure 2.11** of the SI) resulting in independent sections of the molecule being on average closer to the pore opening and hence more likely to be captured alongside the main strand translocation. Additionally, the breaking of the tension propagation in sections of ssDNA ensures that these sections of the polymer can move almost independently of the translocating main strand

and helps maintain this reduced radius of gyration during the translocation process. However, the size reduction of polymers does not explain the propensity of ssDNA segments to be captured as opposed to dsDNA segments, the origin of which is clarified in the next section.

### 2.1.3.4 Flexible Ends

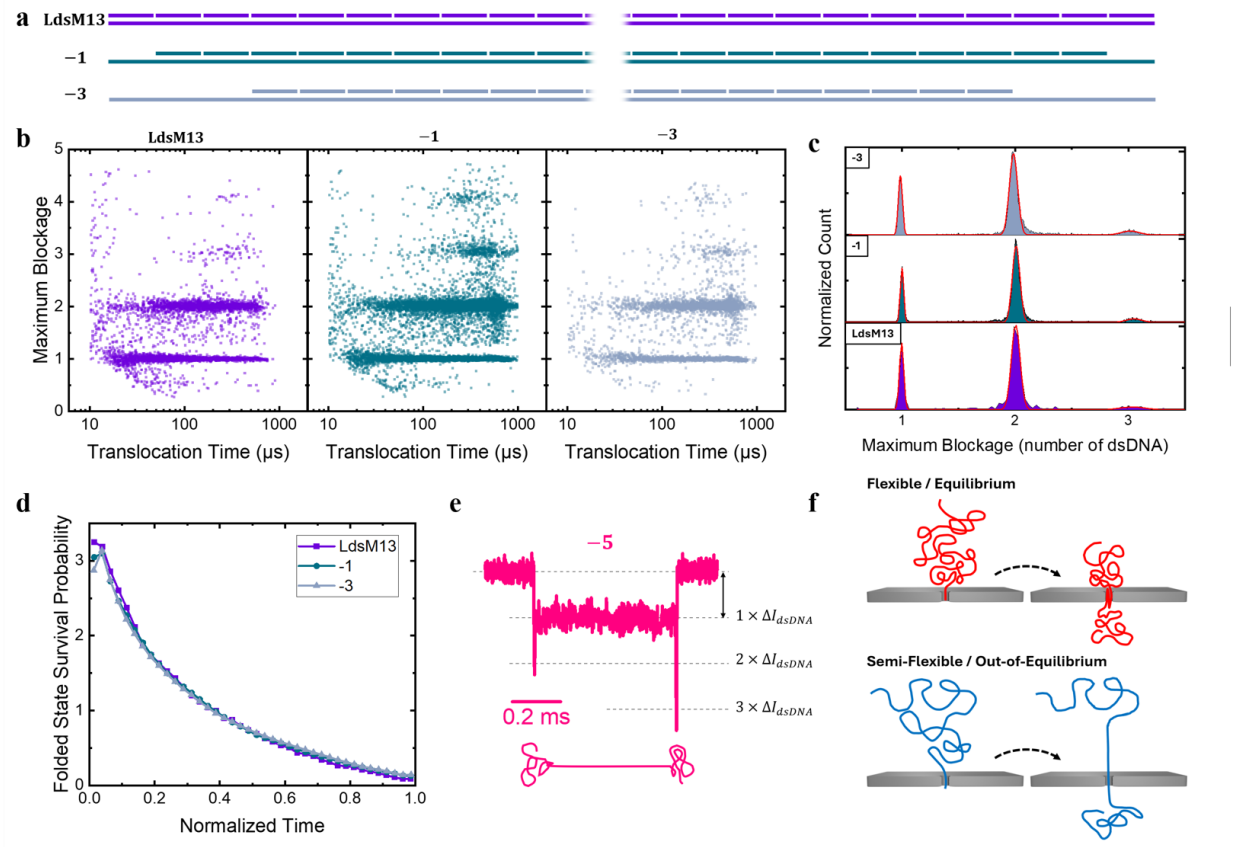



***Figure 2.5: Investigation of flexible ends on a dsDNA molecule in nanopores.*** ***a)*** *Design of LdsM13 (purple) -1(dark cyan) and -3 (grey) structures with flexible ends of varying lengths (0, 44 and 124 nt respectively).* ***b)*** *Maximum blockage in multiples of dsDNA vs translocation time scatterplots for LdsM13, -1 and -3 structures.* ***c)*** *Maximum fitted blockage histograms for LdsM13, -1 and -3 structures.* ***d)*** *Folded-state survival probability distribution for three structures.* ***e)*** *Current traces of -5 structure translocations.* ***f)*** *Schematic of the translocation process for flexible (top) and semi-flexible (bottom) polymers.*

Observations from **Figure 2.3** and **Figure 2.4** have led us to conclude that ssDNA segments are more readily captured throughout a translocation event compared to fully double-stranded sections. To test whether the presence of ssDNA segments also affects the kinetics of polymers approaching and entering nanopores, we studied the passage of LdsM13 structures with flexible ssDNA extremities. We hypothesized that if ssDNA segments were attracted to the pore more significantly during the capture process as they are throughout the translocation process, ssDNA extremities would promote single-file translocations, i.e. polymers entering the pore by an end, and alter the folding statistics introduced in **Figure 2.2f**. We note that since folding is a major hurdle for applications relying on the single file translocation of molecular nanostructured assemblies (e.g. DNA barcodes[160]), these experiments were also motivated by attempting to reduce folded translocations.

As per the heterostructures above, flexible ssDNA extremities were achieved by removing the 40-nucleotide-long ssDNA staple strands located at the polymer ends from the mix during LdsM13 assembly, as shown in **Figure 2.5a**. Structures used here are identified by how many oligos were omitted on each extremity: -1, -3 and -5, representing respectively 44, 124 and 204 nt long ssDNA segments at the end of the dsDNA polymer. **Figure 2.5b** shows the maximum blockage vs translocation times scatterplots measured for these structures, and **Figure 2.5c** shows the corresponding maximum blockage histograms. Note that only structures with full length scaffolds (i.e. events with ECD > 0.28 pC) were considered for analysis. Each quantized peak was fitted to a Gaussian function, and subpopulation amplitudes were extracted to calculate the probability of observing folded events ($\Delta I_{max} \geq 2\Delta I_{dsDNA}$) for each structure. Interestingly, the presence of flexible ssDNA extremities did not change the fraction of folded translocations, which

was consistently observed to be $75.7 \pm 0.2\%$ across the LdsM13, -1 and -3 polymers (as discussed below, the presence of additional blockage states for the -5 structure made accurate measurement of the folding fraction difficult), as calculated by the amplitudes of the Gaussian fits to $\Delta I_{max}$ distributions. Additionally, **Figure 2.5d** shows the folded state survival probability $S_{fold}$ of the three nanostructures, which appear to be nearly identical. These observations suggest that during the capture process these ssDNA segments are not more accessible to the pore than dsDNA segments, leading us to conclude that the overall dynamics are dominated by the behavior of the polymer as a whole and not by the dynamics of the extremities of the polymer.

Revealingly, traces of single-file translocations of the -5 structure (i.e. dsDNA polymer with 200 nt long ssDNA ends), shown in **Figure 2.5e**, wherein many events contain a blockage state at the start and end of the event, exhibit amplitudes that do not correspond to quantized dsDNA blockages. These initial and final non-dsDNA-integer states are barely observed for LdsM13, -1 and -3 structures, and as such are instead believed to correspond to extremity ssDNA segments passing through the nanopore not in an elongated fashion, like dsDNA, but as a blob instead. As such, even under the effects of the strong nanopore electric field, the conformations of ssDNA segments are relaxed. The combined observations from **Figure 2.3**, **Figure 2.4**, and **Figure 2.5** suggest that the ssDNA segments outside the pore, whilst also being attracted to the pore, are free to diffuse around and explore the space, unlike the more rigid dsDNA segments which are essentially immobile until the tension front reaches them. Throughout the translocation process, the relaxed ssDNA segments diffuse and find the pore entrance more readily than dsDNA segments which leads to more complex folded translocations (**Figure 2.3**) and deeper blockages (**Figure**

**2.3c**). This description is illustrated in **Figure 2.5f** which depicts how the overall rigidity of polymers affects their conformations throughout the translocation process.

### 2.1.4 Conclusion

In this study we presented ssDNA-dsDNA heterostructures as model systems to experimentally study the dynamics of folding of polymers translocating solid-state nanopores. We first established that our nanopore device is agnostic to the presence of nicks along the contour of the DNA and the nicked dsDNA behaves like fully ligated dsDNA with regards to blockage depth, passage time distribution, and folding statistics. This result was somewhat expected since it is well established that the main influence on the high persistence length of dsDNA is due to base stacking,[133,217] which is largely unaffected by the presence of a single nick along one of the strands.

To understand how the polymer's flexibility (persistence length, radius of gyration) affected the dynamics of polymer chains translocating through a nanopore, we built ss-dsDNA hybrid molecules with differing effective persistence lengths by removing specific short oligos from the mixture of oligos used to assemble these structures. Solid-state nanopore translocation events of these heterostructures revealed a difference in translocation complexity with more flexible structures generating deeper, longer and more complex current traces, compared to their more rigid counterparts. An in-depth look at the semi-flexible assemblies gives us a hint at the mechanism behind the translocation of these flexible assemblies by proposing parallel capture at the ssDNA segments. Heterostructure designs with ssDNA segments moved to the ends of the molecule showed no significant change in folding behavior yet clearly showed that ssDNA segments are relaxed throughout translocation. Combined, these observations show that more

flexible polymers, in addition to being more compact and closer to the pore mouth, are free to explore and diffuse in the surroundings of the nanopore throughout the translocation process, which leads to increased current blockages and increased folded translocations by having multiple segments of the same polymer being captured during the translocation process. Polymer folding in nanopores remains a big challenge for the broad application of solid-state nanopores, but we hope the work presented in this study will provide useful insights to researchers studying semi-flexible translocation through nanopores.

## 2.1.5 Methods

### 2.1.5.1 DNA Sample & Assembly

Linear dsDNA fragment controls were purchased from ThermoFisher (NoLimits 7000 bp DNA Fragment, SM1741). Nicked and flexible DNA assemblies were prepared by mixing an M13 scaffold (Tilibit Nanosystems M13mp18, type p7249, 7249 nt long) with five times excess of an equimolar mixture of the required oligos (Integrated DNA Technologies, see **Table 2.2** for a full list of the oligos used for each assembly) in 1x TE buffer with added 10 mM $MgCl_2$ (see **Table 2.1** in the SI for the full set of sequences). The mixture was heated to 70 °C for 10 minutes and then cooled down to room temperature (25 °C) over the course of 30 minutes. The scaffold was first linearized with the restriction enzyme HincII (New England Biolabs, R0103S). Excess oligos were removed with the help of Amicon 100 kDa molecular weight cutoff filters (Millipore Sigma, UFC510096).

#### 2.1.5.2 Nanopore Measurements

Nanopores were fabricated in 10–15 nm thick SiN membranes using the controlled breakdown (CBD) method in a custom design instrument as described in Waugh et al.[47,48] Nanopore fabrication was performed using a voltage ramp in 1 M KCl, 10 mM HEPES pH 8 until breakdown is detected. Nanopore enlargement was then performed in 3.6 M LiCl, 10 mM HEPES pH 8 until the desired size was achieved. Nanopore dimensions were extracted using the conductance model described in previous work.[51,53] All experiments were performed in 3.6 M LiCl 10 mM HEPES pH 8. For a list of all pores used in this study, refer to Supporting Information **Table 2.3**.

#### 2.1.5.3 Data Analysis

Translocation events in the recorded current traces were located and fitted using a custom implementation of the CUSUM+ algorithm.[211] A digital low-pass filter was applied prior to event detection and fitting. The value of the cutoff frequency was chosen on a per-experiment basis (usually either 500 kHz or 250 kHz), to maximize temporal resolution whilst maintaining good signal to noise ratio. Custom Python and R scripts were used for plotting and data filtering.

#### 2.1.5.4 Acknowledgements

All authors would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2021-04304]. M. C. acknowledges support of the Ontario Graduate Scholarship (OGS).

#### 2.1.5.5 Supporting Information

Circular Double-stranded M13 – Random Cutting; Filtering Folded Fragmented LdsM13 Translocation Events; Folded State Survival Probability; Gel Images for Flexible Assemblies; 2D Heatmaps of Flexible Assemblies; Dependence of Heterostructure Radii on ssDNA Content; Effect of Polymer Rigidity on Monomer-Pore Distance; Triple Blockage ECD Plots; List of Oligo Sequences; Additional Event Traces for Flexible End Molecules

## 2.2 Supporting Information: Folding Dynamics of Linear ssDNA-dsDNA Heterostructures through Solid-state Nanopores

### 2.2.1 S1. Circular Double-stranded M13 – Random Cutting

Commercial M13mp18 scaffolds come in circular form, but a portion of the molecules are randomly broken due to handling during the processing. To assess the quality of the M13mp18 scaffold and the origin of the wide distribution of translocation time of the LdsM13 results shown in **Figure 2.2** in the main text, the assembly of the LdsM13 was performed using uncut M13mp18 scaffold, for which a perfect yield would result instead in circular double-stranded M13 (CdsM13).

Upon assembly, we passed the CdsM13 molecules through a 7.9 nm pore in 3.6 M LiCl under a 300 mV bias. **Figure 2.6** shows the maximum blockage of the CdsM13 *versus* the corresponding translocation time. Two blockage sub-populations are fitted to have $\Delta G = 6.12$ nS and 12.40 nS. Having passed dsDNA through the same nanopore (not shown), we established that the lower blockade population corresponds to that of dsDNA and thus to passages of linearized CdsM13 which result from randomly broken scaffolds. The higher blockade population therefore either corresponds to passages of folded LdsM13 or CdsM13.

**Figure 2.6b** shows a scatterplot of the minimum blockage versus translocation time for the data of **Figure 2.6a**. Since the minimum blockage level of folded linear polymer translocations corresponds to the single-file state, **Figure 2.6b** better separates linearized and circular M13 translocations. By comparing subpopulation amplitudes, we estimate that roughly 55% of the scaffolds used for this experiment were broken and linearized.

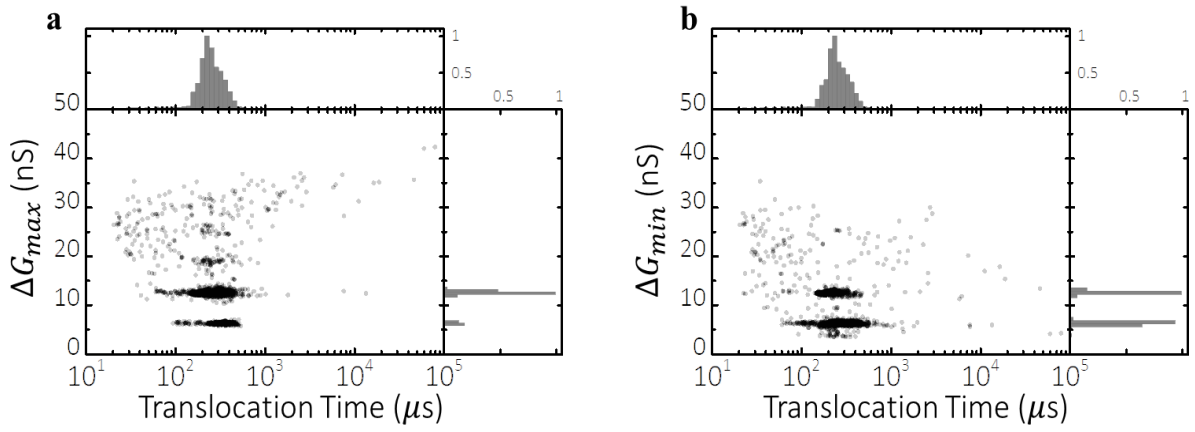



***Figure 2.6: Scatter plots of dsM13 assembly using uncut M13mp18. a)*** *Scatter plot of maximum blockage versus translocation time, with histograms of maximum blockage and translocation time.* ***b)*** *Scatter plot of minimum blockage versus translocation time, with histograms of minimum blockage and translocation time. The nanopore experiments were done in 3.6 M LiCl using a 7.9 nm pore with an applied bias of 300 mV.*

### 2.2.2 S2. Filtering Folded Fragmented LdsM13 Translocation Events

Supplementary **Figure 2.7a** shows the total duration of folded LdsM13 events plotted against the respective folded state duration of each individual LdsM13 folded translocations in an 8.4 nm pore, in 3.6 M LiCl, under a 300 mV bias. Supplementary **Figure 2.7b** shows the corresponding distribution of translocation times. Filtering misassembled LdsM13 fragments solely based on translocation times cannot be done precisely, since the translocation time distribution is a convolution of molecules folding at different locations along their contour, and of different lengths passing through the pore. To address this, a filter of the form $\tau = (200\ \mu s - \tau_{fold})$ was used instead, as displayed by the red dashed line in Supplementary **Figure**

**2.7a**. Supplementary **Figure 2.7c-d** justify the use of this filter by plotting the data from **Figure 2.7**a rotated by 45 degrees, and the corresponding x-axis ($\tau_{fold}\sin(-\pi/4) + \tau\cos(-\pi/4)$) distribution, respectively, showing a better separation between folded and misassembled LdsM13 events.

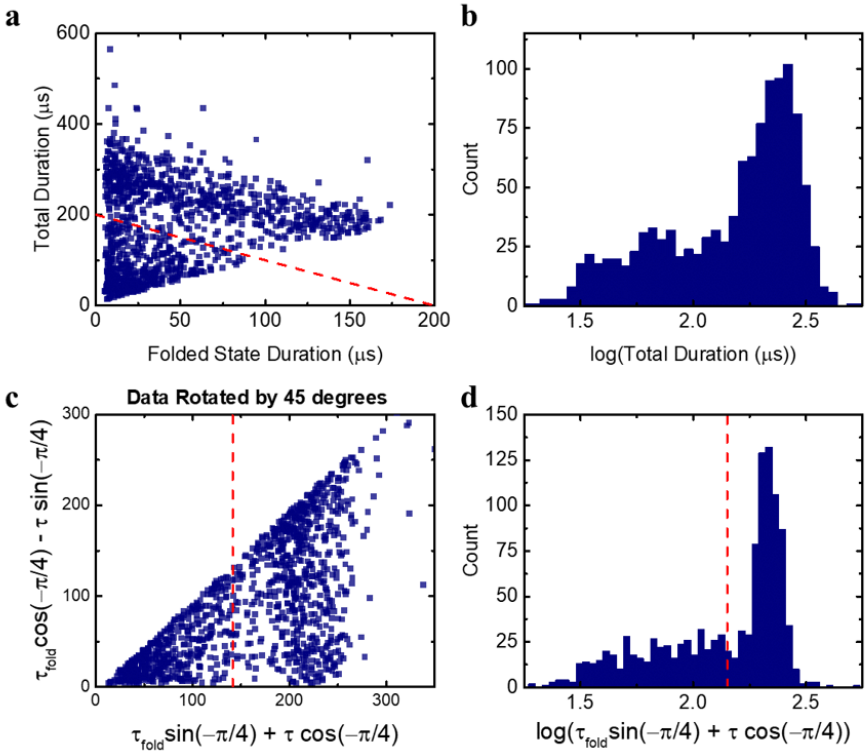



***Figure 2.7: Filtering of folded fragmented LdsM13 events. a)*** *Total translocation time versus folded portion of folded events.* ***b)*** *Histogram of the log of total translocation time of folded events.* ***c-d)*** *The data from a-b) rotated by 45 degrees. The experiment was performed using an 8.4 nm pore, in 3.6 M LiCl, with an applied bias of 300 mV.*

## 2.2.3 S3. Folded State Survival Probability

Folded-State Survival probability $S_{fold}(\tau)$ simply quantifies the probability that an event is still in a folded state at time $t$. Here $t$ is the normalized time such that 0 and 1 correspond

to the start and end of the event. Given $N$ translocation events, let each event contain $n_i$ datapoints between times $t = 0$ and 1, where each are characterized by a time and current $\left(t_{i,j}, I_{i,j}\right)$ with $i \in [1, N]$ and $j \in [1, n_i]$. With a folded state determined by upper and lower bounds $I_{min}$ and $I_{max}$, we can calculate the value of $S_{fold}$ at time $\tau$ with a resolution $\delta t$ using the following formula:

$$S_{fold}(t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n_i} w_{i,j}$$

Here, $w_{i,j}$ is the weight attributed to each datapoint, defined as follows:

$$w_{i,j} = \begin{cases} 0 & for\ I_{i,j} \notin [I_{min}, I_{max}]\ \text{or}\ t_{i,j} \notin \left[t - \frac{\delta t}{2}, t + \frac{\delta t}{2}\right] \\ (\delta t n_i)^{-1} & for\ I_{i,j} \in [I_{min}, I_{max}] \text{ and } t_{i,j} \in \left[t - \frac{\delta t}{2}, t + \frac{\delta t}{2}\right] \end{cases}$$

The above formulae simply count the number of current points in the folded state within time intervals of width $\delta t$. However, and importantly, to avoid overcounting events with long translocation times and undercounting shorter ones, each datapoint is weighted by the inverse of the corresponding event's length $n_i$. The above method allows using the raw data to compute folding statistics instead of relying on fits which handle short folded sub-levels poorly.

## 2.2.4 S4. Gel Images for Flexible Assemblies

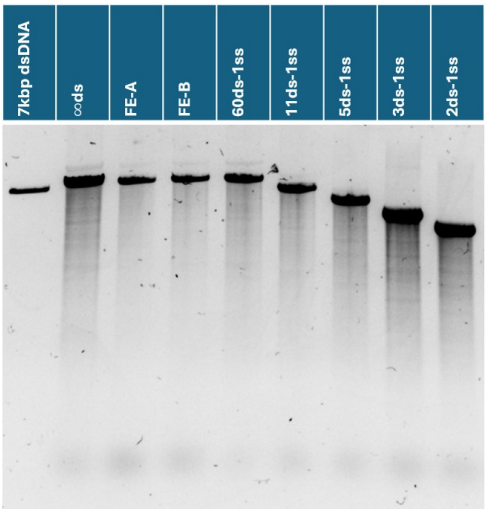



*Figure 2.8: 1% Agarose gel performed in 1x TAE buffer of various flexible structures used in this study.*

## 2.2.5 S5. 2D Heatmaps of Flexible Assemblies

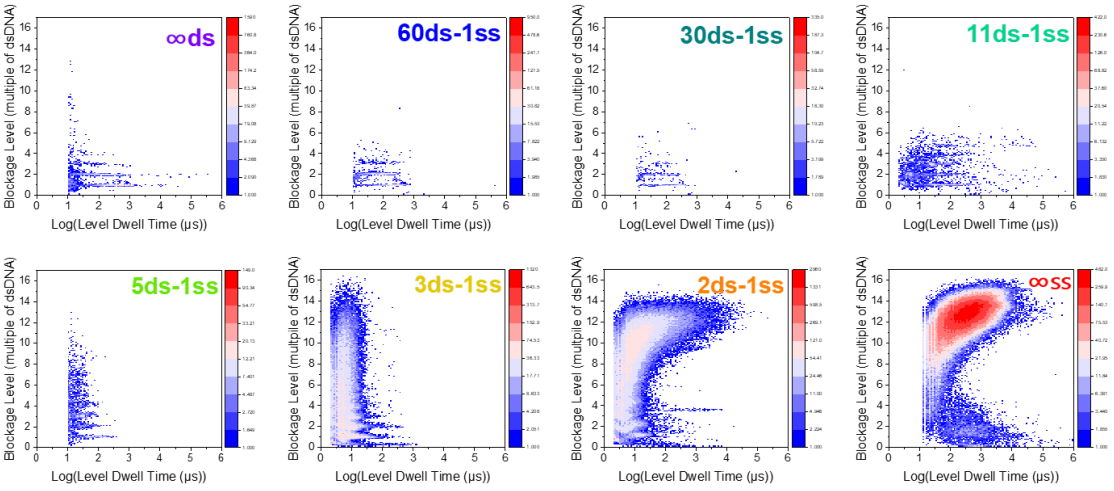



*Figure 2.9: 2D heatmaps of all levels fitted to events for the various structures investigated in this study.*

## 2.2.6 S6. Dependence of Heterostructure Radii on ssDNA Content

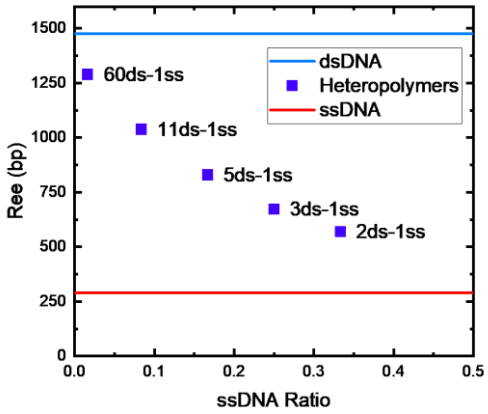



***Figure 2.10: Average end-to-end distance calculated from 5000 heteropolymers*** *simulated using discrete worm-like chain, as described in Tree et al.[213] Persistence lengths used for ssDNA and dsDNA were 2 nm and 50 nm, respectively.*

## 2.2.7 S7. Effect of Polymer Rigidity on Monomer-Pore Distance

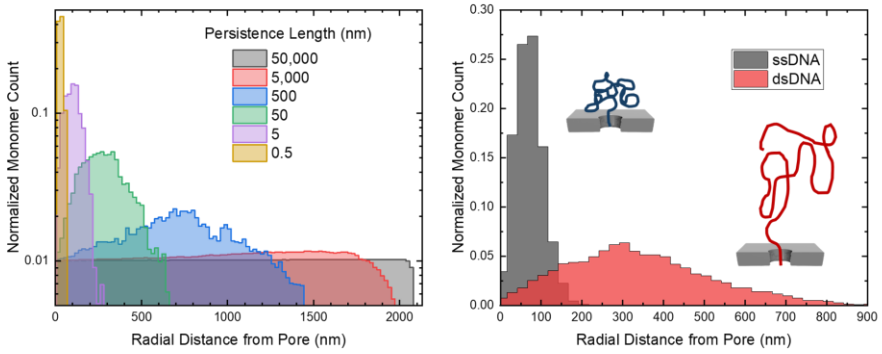



***Figure 2.11: Histogram of pore-monomer distances calculated for simulated polymers of different rigidities anchored inside a nanopore. a)*** *Persistence lengths of different orders of magnitudes;* ***b)*** *Persistence lengths of dsDNA (50 nm) and ssDNA (2 nm). Polymer conformations simulated using discrete worm-like chain (DWLC), as described in Tree et al.[213] Monomers here are defined as the end points of the biased randomized steps of the simulated DWLCs.*

### 2.2.8 S8. Triple Blockage ECD Plots

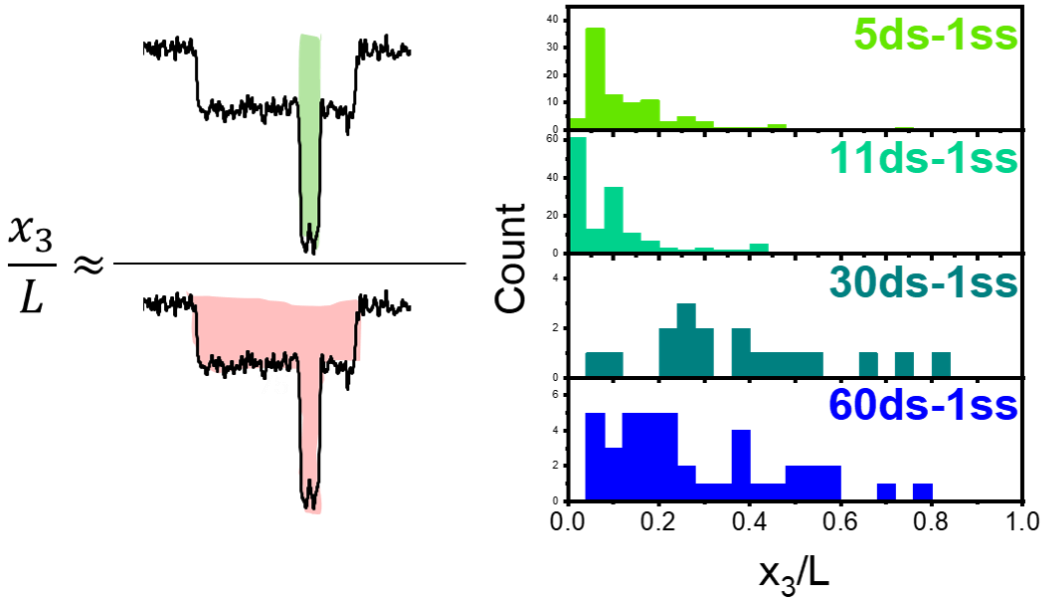



*Figure 2.12: Normalized ECD of the 3 level ($x_3/L$) for events ending in 1-3-1 transitions.*

### 2.2.9 S9. List of Oligo Sequences

*Table 2.1: List of all oligo sequences needed to fully assemble a dsDNA M13 scaffold.*

| Oligo ID | Sequence | Length (nt) |
|---|---|---|
| P001 | GACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATG | 45 |
| P002 | GTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA | 40 |
| P003 | ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG | 40 |
| P004 | CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGC | 40 |
| P005 | GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCG | 40 |
| P006 | TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAG | 40 |
| P007 | GCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCA | 40 |
| P008 | CCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTG | 40 |
| P009 | GCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGC | 40 |

| | | |
|---|---|---|
| P010 | CCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTCCGAAAT | 40 |
| P011 | CGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATA | 40 |
| P012 | GGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTAT | 40 |
| P013 | TAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGT | 40 |
| P014 | CTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAA | 40 |
| P015 | TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATC | 40 |
| P016 | GGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGG | 40 |
| P017 | AAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCG | 40 |
| P018 | AAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCA | 40 |
| P019 | CGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCC | 40 |
| P020 | GCTACAGGGCGCGTACTATGGTTGCTTTGACGAGCACGTA | 40 |
| P021 | TAACGTGCTTTCCTCGTTAGAATCAGAGCGGGAGCTAAAC | 40 |
| P022 | AGGAGGCCGATTAAAGGGATTTTAGACAGGAACGGTACGC | 40 |
| P023 | CAGAATCCTGAGAAGTGTTTTTATAATCAGTGAGGCCACC | 40 |
| P024 | GAGTAAAAGAGTCTGTCCATCACGCAAATTAACCGTTGTA | 40 |
| P025 | GCAATACTTCTTTGATTAGTAATAACATCACTTGCCTGAG | 40 |
| P026 | TAGAAGAACTCAAACTATCGGCCTTGCTGGTAATATCCAG | 40 |
| P027 | AACAATATTACCGCCAGCCATTGCAACAGGAAAAACGCTC | 40 |
| P028 | ATGGAAATACCTACATTTTGACGCTCAATCGTCTGAAATG | 40 |
| P029 | GATTATTTACATTGGCAGATTCACCAGTCACACGACCAGT | 40 |
| P030 | AATAAAAGGGACATTCTGGCCAACAGAGATAGAACCCTTC | 40 |
| P031 | TGACCTGAAAGCGTAAGAATACGTGGCACAGACAATATTT | 40 |
| P032 | TTGAATGGCTATTAGTCTTTAATGCGCGAACTGATAGCCC | 40 |
| P033 | TAAAACATCGCCATTAAAAATACCGAACGAACCACCAGCA | 40 |
| P034 | GAAGATAAAACAGAGGTGAGGCGGTCAGTATTAACACCGC | 40 |
| P035 | CTGCAACAGTGCCACGCTGAGAGCCAGCAGCAAATGAAAA | 40 |
| P036 | ATCTAAAGCATCACCTTGCTGAACCTCAAATATCAAACCC | 40 |
| P037 | TCAATCAATATCTGGTCAGTTGGCAAATCAACAGTTGAAA | 40 |
| P038 | GGAATTGAGGAAGGTTATCTAAAATATCTTTAGGAGCACT | 40 |
| P039 | AACAACTAATAGATTAGAGCCGTCAATAGATAATACATTT | 40 |
| P040 | GAGGATTTAGAAGTATTAGACTTTACAAACAATTCGACAA | 40 |
| P041 | CTCGTATTAAATCCTTTGCCCGAACGTTATTAATTTTAAA | 40 |
| P042 | AGTTTGAGTAACATTATCATTTTGCGGAACAAAGAAACCA | 40 |
| P043 | CCAGAAGGAGCGGAATTATCATCATATTCCTGATTATCAG | 40 |
| P044 | ATGATGGCAATTCATCAATATAATCCTGATTGTTTGGATT | 40 |
| P045 | ATACTTCTGAATAATGGAAGGGTTAGAACCTACCATATCA | 40 |
| P046 | AAATTATTTGCACGTAAAACAGAAATAAAGAAATTGCGTA | 40 |
| P047 | GATTTTCAGGTTTAACGTCAGATGAATATACAGTAACAGT | 40 |
| P048 | ACCTTTTACATCGGGAGAAACAATAACGGATTCGCCTGAT | 40 |
| P049 | TGCTTTGAATACCAAGTTACAAAATCGCGCAGAGGCGAAT | 40 |

| | | |
|---|---|---|
| P050 | TATTCATTTCAATTACCTGAGCAAAAGAAGATGATGAAAC | 40 |
| P051 | AAACATCAAGAAAACAAAATTAATTACATTTAACAATTTC | 40 |
| P052 | ATTTGAATTACCTTTTTTAATGGAAACAGTACATAAATCA | 40 |
| P053 | ATATATGTGAGTGAATAACCTTGCTTCTGTAAATCGTCGC | 40 |
| P054 | TATTAATTAATTTTCCCTTAGAATCCTTGAAAACATAGCG | 40 |
| P055 | ATAGCTTAGATTAAGACGCTGAGAAGAGTCAATAGTGAAT | 40 |
| P056 | TTATCAAAATCATAGGTCTGAGAGACTACCTTTTTAACCT | 40 |
| P057 | CCGGCTTAGGTTGGGTTATATAACTATATGTAAATGCTGA | 40 |
| P058 | TGCAAATCCAATCGCAAGACAAAGAACGCGAGAAAACTTT | 40 |
| P059 | TTCAAATATATTTTAGTTAATTTCATCTTCTGACCTAAAT | 40 |
| P060 | TTAATGGTTTGAAATACCGACCGTGTGATAAATAAGGCGT | 40 |
| P061 | TAAATAAGAATAAACACCGGAATCATAATTACTAGAAAAA | 40 |
| P062 | GCCTGTTTAGTATCATATGCGTTATACAAATTCTTACCAG | 40 |
| P063 | TATAAAGCCAACGCTCAACAGTAGGGCTTAATTGAGAATC | 40 |
| P064 | GCCATATTTAACAACGCCAACATGTAATTTAGGCAGAGGC | 40 |
| P065 | ATTTTCGAGCCAGTAATAAGAGAATATAAAGTACCGACAA | 40 |
| P066 | AAGGTAAAGTAATTCTGTCCAGACGACGACAATAAACAAC | 40 |
| P067 | ATGTTCAGCTAATGCAGAACGCGCCTGTTTATCAACAATA | 40 |
| P068 | GATAAGTCCTGAACAAGAAAAATAATATCCCATCCTAATT | 40 |
| P069 | TACGAGCATGTAGAAACCAATCAATAATCGGCTGTCTTTC | 40 |
| P070 | CTTATCATTCCAAGAACGGGTATTAAACCAAGTACCGCAC | 40 |
| P071 | TCATCGAGAACAAGCAAGCCGTTTTTATTTTCATCGTAGG | 40 |
| P072 | AATCATTACCGCGCCCAATAGCAAGCAAATCAGATATAGA | 40 |
| P073 | AGGCTTATCCGGTATTCTAAGAACGCGAGGCGTTTTAGCG | 40 |
| P074 | AACCTCCCGACTTGCGGGAGGTTTTGAAGCCTTAAATCAA | 40 |
| P075 | GATTAGTTGCTATTTTGCACCCAGCTACAATTTTATCCTG | 40 |
| P076 | AATCTTACCAACGCTAACGAGCGTCTTTCCAGAGCCTAAT | 40 |
| P077 | TTGCCAGTTACAAAATAAACAGCCATATTATTTATCCCAA | 40 |
| P078 | TCCAAATAAGAAACGATTTTTTGTTTAACGTCAAAAATGA | 40 |
| P079 | AAATAGCAGCCTTTACAGAGAGAATAACATAAAAACAGGG | 40 |
| P080 | AAGCGCATTAGACGGGAGAATTAACTGAACACCCTGAACA | 40 |
| P081 | AAGTCAGAGGGTAATTGAGCGCTAATATCAGAGAGATAAC | 40 |
| P082 | CCACAAGAATTGAGTTAAGCCCAATAATAAGAGCAAGAAA | 40 |
| P083 | CAATGAAATAGCAATAGCTATCTTACCGAAGCCCTTTTTA | 40 |
| P084 | AGAAAAGTAAGCAGATAGCCGAACAAAGTTACCAGAAGGA | 40 |
| P085 | AACCGAGGAAACGCAATAATAACGGAATACCCAAAAGAAC | 40 |
| P086 | TGGCATGATTAAGACTCCTTATTACGCAGTATGTTAGCAA | 40 |
| P087 | ACGTAGAAAATACATACATAAAGGTGGCAACATATAAAAG | 40 |
| P088 | AAACGCAAAGACACCACGGAATAAGTTTATTTTGTCACAA | 40 |
| P089 | TCAATAGAAAATTCATATGGTTTACCAGCGCCAAAGACAA | 40 |

| | | |
|---|---|---|
| P090 | AAGGGCGACATTCAACCGATTGAGGGAGGGAAGGTAAATA | 40 |
| P091 | TTGACGGAAATTATTCATTAAAGGTGAATTATCACCGTCA | 40 |
| P092 | CCGACTTGAGCCATTTGGGAATTAGAGCCAGCAAAATCAC | 40 |
| P093 | CAGTAGCACCATTACCATTAGCAAGGCCGGAAACGTCACC | 40 |
| P094 | AATGAAACCATCGATAGCAGCACCGTAATCAGTAGCGACA | 40 |
| P095 | GAATCAAGTTTGCCTTTAGCGTCAGACTGTAGCGCGTTTT | 40 |
| P096 | CATCGGCATTTTCGGTCATAGCCCCCTTATTAGCGTTTGC | 40 |
| P097 | CATCTTTTCATAATCAAAATCACCGGAACCAGAGCCACCA | 40 |
| P098 | CCGGAACCGCCTCCCTCAGAGCCGCCACCCTCAGAACCGC | 40 |
| P099 | CACCCTCAGAGCCACCACCCTCAGAGCCGCCACCAGAACC | 40 |
| P100 | ACCACCAGAGCCGCCGCCAGCATTGACAGGAGGTTGAGGC | 40 |
| P101 | AGGTCAGACGATTGGCCTTGATATTCACAAACAAATAAAT | 40 |
| P102 | CCTCATTAAAGCCAGAATGGAAAGCGCAGTCTCTGAATTT | 40 |
| P103 | ACCGTTCCAGTAAGCGTCATACATGGCTTTTGATGATACA | 40 |
| P104 | GGAGTGTACTGGTAATAAGTTTTAACGGGGTCAGTGCCTT | 40 |
| P105 | GAGTAACAGTGCCCGTATAAACAGTTAATGCCCCCTGCCT | 40 |
| P106 | ATTTCGGAACCTATTATTCTGAAACATGAAAGTATTAAGA | 40 |
| P107 | GGCTGAGACTCCTCAAGAGAAGGATTAGGATTAGCGGGGT | 40 |
| P108 | TTTGCTCAGTACCAGGCGGATAAGTGCCGTCGAGAGGGTT | 40 |
| P109 | GATATAAGTATAGCCCGGAATAGGTGTATCACCGTACTCA | 40 |
| P110 | GGAGGTTTAGTACCGCCACCCTCAGAACCGCCACCCTCAG | 40 |
| P111 | AACCGCCACCCTCAGAGCCACCACCCTCATTTTCAGGGAT | 40 |
| P112 | AGCAAGCCCAATAGGAACCCATGTACCGTAACACTGAGTT | 40 |
| P113 | TCGTCACCAGTACAAACTACAACGCCTGTAGCATTCCACA | 40 |
| P114 | GACAGCCCTCATAGTTAGCGTAACGATCTAAAGTTTTGTC | 40 |
| P115 | GTCTTTCCAGACGTTAGTAAATGAATTTTCTGTATGGGAT | 40 |
| P116 | TTTGCTAAACAACTTTCAACAGTTTCAGCGGAGTGAGAAT | 40 |
| P117 | AGAAAGGAACAACTAAAGGAATTGCGAATAATAATTTTTT | 40 |
| P118 | CACGTTGAAAATCTCCAAAAAAAAGGCTCCAAAAGGAGCC | 40 |
| P119 | TTTAATTGTATCGGTTTATCAGCTTGCTTTCGAGGTGAAT | 40 |
| P120 | TTCTTAAACAGCTTGATACCGATAGTTGCGCCGACAATGA | 40 |
| P121 | CAACAACCATCGCCCACGCATAACCGATATATTCGGTCGC | 40 |
| P122 | TGAGGCTTGCAGGGAGTTAAAGGCCGCTTTTGCGGGATCG | 40 |
| P123 | TCACCCTCAGCAGCGAAAGACAGCATCGGAACGAGGGTAG | 40 |
| P124 | CAACGGCTACAGAGGCTTTGAGGACTAAAGACTTTTTCAT | 40 |
| P125 | GAGGAAGTTTCCATTAAACGGGTAAAATACGTAATGCCAC | 40 |
| P126 | TACGAAGGCACCAACCTAAAACGAAAGAGGCAAAAGAATA | 40 |
| P127 | CACTAAAACACTCATCTTTGACCCCCAGCGATTATACCAA | 40 |
| P128 | GCGCGAAACAAAGTACAACGGAGATTTGTATCATCGCCTG | 40 |
| P129 | ATAAATTGTGTCGAAATCCGCGACCTGCTCCATGTTACTT | 40 |

| | | |
|---|---|---|
| P130 | AGCCGGAACGAGGCGCAGACGGTCAATCATAAGGGAACCG | 40 |
| P131 | AACTGACCAACTTTGAAAGAGGACAGATGAACGGTGTACA | 40 |
| P132 | GACCAGGCGCATAGGCTGGCTGACCTTCATCAAGAGTAAT | 40 |
| P133 | CTTGACAAGAACCGGATATTCATTACCCAAATCAACGTAA | 40 |
| P134 | CAAAGCTGCTCATTCAGTGAATAAGGCTTGCCCTGACGAG | 40 |
| P135 | AAACACCAGAACGAGTAGTAAATTGGGCTTGAGATGGTTT | 40 |
| P136 | AATTTCAACTTTAATCATTGTGAATTACCTTATGCGATTT | 40 |
| P137 | TAAGAACTGGCTCATTATACCAGTCAGGACGTTGGGAAGA | 40 |
| P138 | AAAATCTACGTTAATAAAACGAACTAACGGAACAACATTA | 40 |
| P139 | TTACAGGTAGAAAGATTCATCAGTTGAGATTTAGGAATAC | 40 |
| P140 | CACATTCAACTAATGCAGATACATAACGCCAAAAGGAATT | 40 |
| P141 | ACGAGGCATAGTAAGAGCAACACTATCATAACCCTCGTTT | 40 |
| P142 | ACCAGACGACGATAAAAACCAAAATAGCGAGAGGCTTTTG | 40 |
| P143 | CAAAAGAAGTTTTGCCAGAGGGGGTAATAGTAAAATGTTT | 40 |
| P144 | AGACTGGATAGCGTCCAATACTGCGGAATCGTCATAAATA | 40 |
| P145 | TTCATTGAATCCCCCTCAAATGCTTTAAACAGTTCAGAAA | 40 |
| P146 | ACGAGAATGACCATAAATCAAAAATCAGGTCTTTACCCTG | 40 |
| P147 | ACTATTATAGTCAGAAGCAAAGCGGATTGCATCAAAAAGA | 40 |
| P148 | TTAAGAGGAAGCCCGAAAGACTTCAAATATCGCGTTTTAA | 40 |
| P149 | TTCGAGCTTCAAAGCGAACCAGACCGGAAGCAAACTCCAA | 40 |
| P150 | CAGGTCAGGATTAGAGAGTACCTTTAATTGCTCCTTTTGA | 40 |
| P151 | TAAGAGGTCATTTTTGCGGATGGCTTAGAGCTTAATTGCT | 40 |
| P152 | GAATATAATGCTGTAGCTCAACATGTTTTAAATATGCAAC | 40 |
| P153 | TAAAGTACGGTGTCTGGAAGTTTCATTCCATATAACAGTT | 40 |
| P154 | GATTCCCAATTCTGCGAACGAGTAGATTTAGTTTGACCAT | 40 |
| P155 | TAGATACATTTCGCAAATGGTCAATAACCTGTTTAGCTAT | 40 |
| P156 | ATTTTCATTTGGGGCGCGAGCTGAAAAGGTGGCATCAATT | 40 |
| P157 | CTACTAATAGTAGTAGCATTAACATCCAATAAATCATACA | 40 |
| P158 | GGCAAGGCAAAGAATTAGCAAAATTAAGCAATAAAGCCTC | 40 |
| P159 | AGAGCATAAAGCTAAATCGGTTGTACCAAAAACATTATGA | 40 |
| P160 | CCCTGTAATACTTTTGCGGGAGAAGCCTTTATTTCAACGC | 40 |
| P161 | AAGGATAAAAATTTTTAGAACCCTCATATATTTTAAATGC | 40 |
| P162 | AATGCCTGAGTAATGTGTAGGTAAAGATTCAAAAGGGTGA | 40 |
| P163 | GAAAGGCCGGAGACAGTCAAATCACCATCAATATGATATT | 40 |
| P164 | CAACCGTTCTAGCTGATAAATTAATGCCGGAGAGGGTAGC | 40 |
| P165 | TATTTTTGAGAGATCTACAAAGGCTATCAGGTCATTGCCT | 40 |
| P166 | GAGAGTCTGGAGCAAACAAGAGAATCGATGAACGGTAATC | 40 |
| P167 | GTAAAACTAGCATGTCAATCATATGTACCCCGGTTGATAA | 40 |
| P168 | TCAGAAAAGCCCCAAAAACAGGAAGATTGTATAAGCAAAT | 40 |
| P169 | ATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCAT | 40 |

| | | |
|---|---|---|
| P170 | TAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGA | 40 |
| P171 | ACGCCATCAAAAATAATTCGCGTCTGGCCTTCCTGTAGCC | 40 |
| P172 | AGCTTTCATCAACATTAAATGTGAGCGAGTAACAACCCGT | 40 |
| P173 | CGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATG | 40 |
| P174 | GGATAGGTCACGTTGGTGTAGATGGGCGCATCGTAACCGT | 40 |
| P175 | GCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGCCTC | 40 |
| P176 | AGGAAGATCGCACTCCAGCCAGCTTTCCGGCACCGCTTCT | 40 |
| P177 | GGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGC | 40 |
| P178 | TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTC | 40 |
| P179 | GCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGG | 40 |
| P180 | CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGAC | 40 |
| P181 | GTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTC | 44 |

***Table 2.2: List of all assembled structures used in this study and the oligos that have been left out of the assembly mixture in the assembly procedure.***

| **Structure Name** | **Oligos missing** |
|---|---|
| ∞ds | - |
| 60ds-1ss | P061, P121 |
| 30ds-1ss | P031, P061, P091, P121, P151 |
| 11ds-1ss | P006, P018, P030, P042, P054, P066, P078, P090, P102, P114, P126, P138, P150, P162, P174 |
| 5ds-1ss | All oligos with numbers divisible by 6 |
| 3ds-1ss | All oligos with numbers divisible by 4 |
| 2ds-1ss | All oligos with numbers divisible by 3 |
| ∞ss | All |
| -1 | P001, P181 |
| -3 | P001, P002, P003, P179, P180, P181 |
| -5 | P001, P002, P003, P004, P005, P177, P178, P179, P180, P181 |

***Table 2.3: List of all pore parameters used in this study****. Different pore length indicates pores of different batches.*

| Pore ID | Diameter (nm) | Length (nm) | Sample | Voltage (mV) |
|---|---|---|---|---|
| 1 | 15.7 | 20.0 | ∞ds | 100 |
| 1 | 15.7 | 20.0 | 5ds-1ss | 100 |
| 2 | 13.6 | 20.0 | 3ds-1ss | 100 |
| 3 | 15.3 | 15.0 | 2ds-1ss | 200 |
| 4 | 8.3 | 15.0 | 11ds-1ss | 200 |
| 5 | 9.7 | 30.0 | 60ds-1ss | 400 |
| 5 | 9.7 | 30.0 | 30ds-1ss | 400 |
| 6 | 11.6 | 20.0 | ∞ss | 200 |

| | | | | |
|---|---|---|---|---|
| 7 | 8.4 | 12.0 | ∞ds | 300 |
| 7 | 8.4 | 12.0 | 7 kbp | 300 |
| 8 | 7.2 | 12.0 | -1 | 200 |
| 8 | 7.2 | 12.0 | -3 | 200 |
| 8 | 7.2 | 12.0 | -5 | 200 |

### 2.2.10 S10. Additional Event Traces for Flexible End Molecules

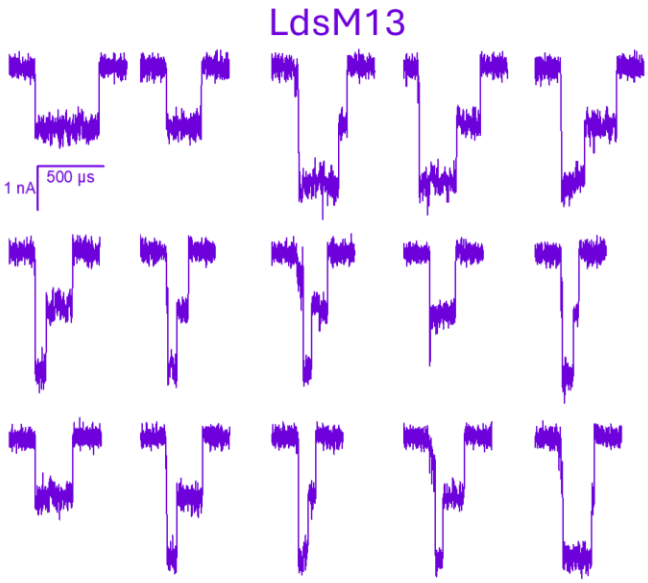



*Figure 2.13: Event traces of LdsM13 molecule.*

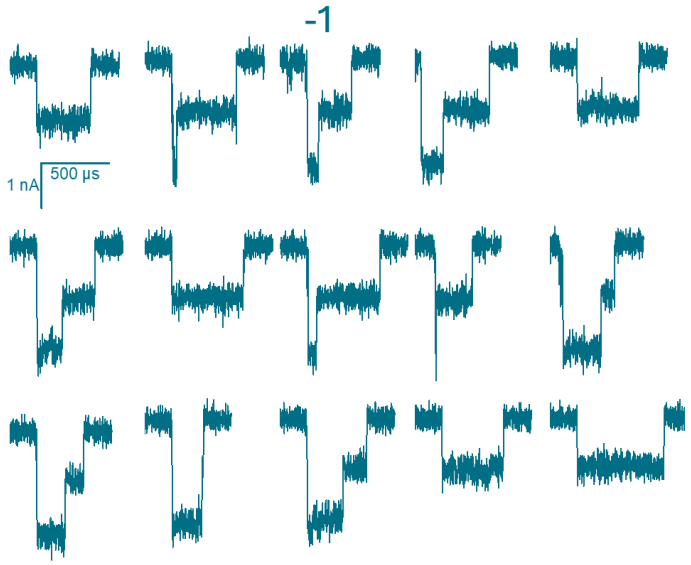



*Figure 2.14: Event traces of -1 flexible end molecule.*

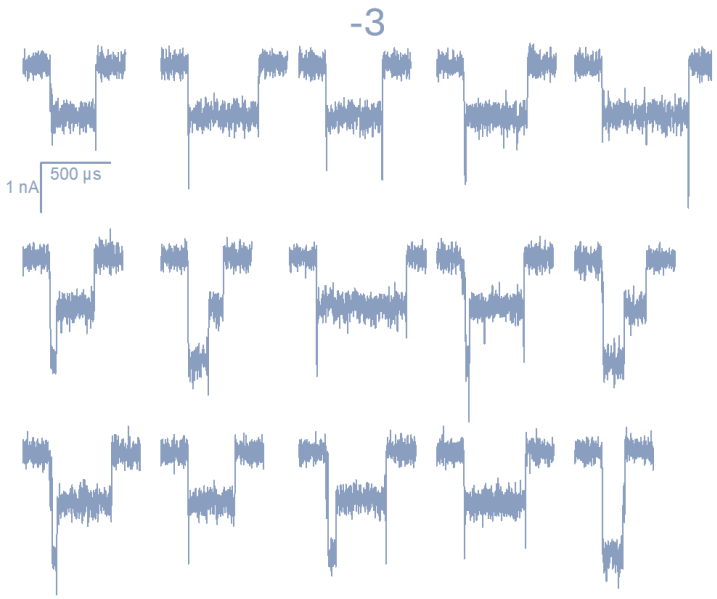



*Figure 2.15: Event traces of -3 flexible end molecule.*

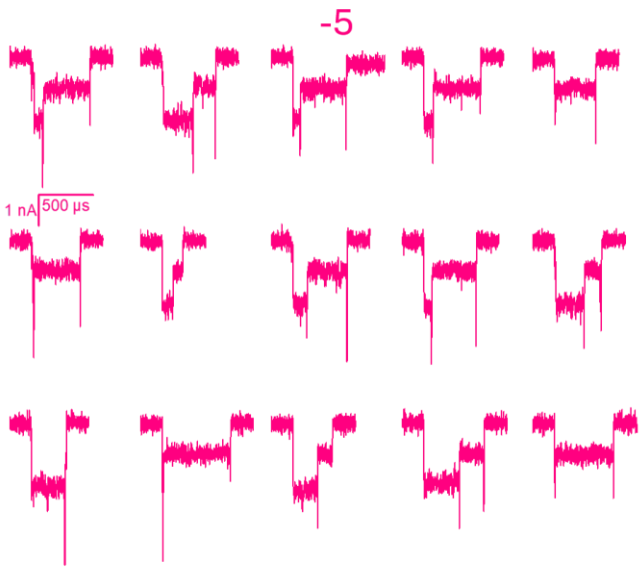



*Figure 2.16: Event traces of -5 flexible end molecule.*

# 3 DNA Barcoding Basics

This chapter is currently submitted and under peer review by *ACS Applied Materials & Interfaces* titled “Nanostructured DNA Barcode Decoding with Solid-State Nanopores: Impact of Bit Spacing and Bit Shape” authored by **Philipp Mensing**, Nada Kerrouri, Sofia Zaynullina, Zachary Roelen, Ping Xie, and Vincent Tabard-Cossa.

## 3.1 Nanostructured DNA Barcode Decoding with Solid-State Nanopores: Impact of Bit Spacing and Bit Shape

*Philipp Mensing[1], Nada Kerrouri[1], Sofia Zaynullina[1], Zachary Roelen[2], Ping Xie[3], and Vincent Tabard-Cossa[1]*

*[1]Department of Physics, University of Ottawa, Ottawa, ON, Canada*

*[2]Oxford Nanopore Technologies Canada, Ottawa, ON, Canada*

*[3]Oxford Nanopore Technologies, Cambridge, MA, USA*

### 3.1.1 Abstract

In recent years nanopores have emerged as a powerful tool for the detection of biomolecules at the single-molecule level. Yet simultaneous detection of many different molecules remains a challenge. Toward this goal, nanostructured DNA barcodes are a promising strategy for enabling the multiplexed detection of biomarkers. The general principle is based on DNA nanostructures assembled along the contour of a linear DNA carrier creating a binary sequence. Different molecules of interest are assigned a unique barcode. Here we address two key challenges in the design of nanostructured DNA barcodes when decoded using pores fabricated on thin SiN membranes (10–20 nm thick) and measured with state-of-the-art MHz-range bandwidth amplifiers. First, to maximize digital data density and multiplexing, we studied the minimal spacing required between nanostructured bits for clear signal resolution, testing this with compact DNA triangles attached to a carrier molecule. Second, to ensure the unique identification of

different bits in the current trace, we explored distinctively sized and shaped DNA nanostructures — specifically DNA stars and helix bundles — designed to generate distinguishable current signatures. Results reveal that for voltage-driven translocation, bits positioned at least 160 bp apart and with rigid shapes with fixed orientations optimize the decoding efficiency.

### 3.1.2 Introduction

Nanopore sensing is a powerful tool to measure the size and structure of individual molecules. A nanopore separating two reservoirs filled with electrolyte solution generates a current that can be blocked by the presence of a molecule with similar dimensions to those of the nanopore.[45] The length and depth of those transient blockages are to first order proportional to the size of the translocating object.[51,53,218] Beyond sequencing applications, this concept has been used to detect specific disease biomarkers.[7,8,45,219] Since many biomolecules of interest (e.g. proteins, miRNA) can generate similar nanopore blockage signals during free translocation or require different sensing conditions (salt concentration, pH, voltage, pore size etc.) to optimize the signal, it remains a challenge to detect and quantify specific molecules directly with a solid-state nanopore sensor. To tackle this problem, the field has pursued strategies involving the use of a proxy molecule to more uniquely label each individual molecule of interest in a sample.[7,8,85,160,195,197] Based on this strategy, we previously developed a digital immunoassay with nanopore readout that could quantify a protein biomarker from plasma or serum down to the high femtomolar sensitivity range with a single pore.[220] To enable multiplexed detection, however, these labels need to be highly differentiable, which quickly becomes challenging above 4-plex, based on blockage depth and passage time characteristics alone.[8,204] Keyser and colleagues.[161] first showed that a promising avenue to enable multiplexed detection of a panel of disease biomarkers with solid-state nanopores

involves associating targeted proteins of interest with barcoded DNA molecules. In this scheme, a carrier molecule, usually M13 ssDNA, is decorated with complementary ssDNA strands. A subset of these ssDNA "staples" is used to encode a digital barcode in the form of bound molecular structures as side chains, forming bits (e.g. a "0" and a "1") at known positions along the carrier. Some designs use the presence and absence of these structures as bits,[26] which requires additional positional information for the barcode to be interpreted correctly, while other designs use two or more structures aiming to generate distinct blockages, that can be interpreted as binary,[27] ternary,[187] or quaternary[188] bit codes – but this remains a challenge. The bit sequence, originating from the specific electrical blockage signatures of each individual bit type, serves to uniquely identify these carriers. This is an important first step for the detection of different target molecules in mixtures[160] with a single nanopore. Attachment geometry and chemistry can vary greatly for such a molecular barcoding approach, leading to a wide range of possible applications like DNA data storage,[188] multiplexed target detection,[183] structure determination,[94] and dynamics investigations.[119]

Previous studies have investigated the influence of attachment size and geometry on the resulting nanopore current signal. Bell *et al.*[160] studied DNA dumbbell structures to investigate the resulting peak size and detection reliability for bits of various lengths with a glass nanopipette system operated with a 50 kHz bandwidth. They determined that, with their measurement setup, a bit length of 220 bp (~75 nm) with a spacing of 312 bp (~106 nm) is required for reliable detection. This allowed for barcodes where the "0" is represented by the absence of an attachment and requires measuring the relative positions of peaks in order to correctly identify them, which runs into problems with long streaks of "0" bits in a given barcode due to the inherent spread of nanopore translocation times and the non-constant translocation velocity.[118] Gevers *et al.*[201] studied

the same system in SiN solid-state nanopores at 100 kHz bandwidth and concluded that they need a similar minimum bit length of about 350 bp (~119 nm), while the bit spacing was not investigated in their study. A design with much smaller attachments in the form of two differently sized DNA hairpins (8 and 16 bp length, respectively) and high-density packing (fitting 56 bits onto the M13 scaffold) was demonstrated by Chen *et al.*[27] In their study it was shown that for bit spacings less than 114 bp the chance of distinguishing adjacent attachments on the carriers dropped below 50%, while the blockage distributions for the two different attachments exhibited significant overlap, making barcode identification only possible as an ensemble measurement, which is not compatible with multiplexing. Finally, Zhu *et al.*[188] presented a design with multiple different attachments to build quaternary instead of binary barcodes using differently sized DNA origami stars. They showed robust classification from ensemble statistics after sorting different barcodes based on a binary address sequence in the front of the carrier molecule. While these studies provide promising proof-of-concept results, they rarely discuss in depth the failure modes of nanopore barcode decoding or the issues associated with molecular assembly, and the yields of correctly assembled and detected barcodes are often not reported. Partly as a result, barcodes are typically classified at the population level, from ensemble statistics, rather than event by event. Here, we classify each event individually.

In this study, we tackle two major challenges that remain to be investigated for pores fabricated on thin SiN membranes (10–20 nm thick) and measured with state-of-the-art MHz-range bandwidth amplifiers. These SiN nanopore devices represent an important class of solid-state nanopores, yet glass nanopipettes account for most nanostructured DNA barcode studies reported so far.[221] Here, the use of high bandwidth amplifiers (up to 1 MHz) allows for an improved

time resolution to aid in the detection of small and closely spaced attachments.[50,222–225] The first challenge lies in determining the minimal spacing between nanostructured bits along the carrier molecule so that adjacent peaks are clearly distinct from one another. Optimization of this distance will ultimately lead to an improvement in the density of bits that can be stored on the polymer and thus the multiplexing potential. To study this question, we designed a simple and compact DNA triangle structure that attaches to the side of the carrier. The second challenge relates to the unique identification of the two different bits in the event current trace, arising from differently sized and/or shaped attachments on the carrier. We explored multiple designs of DNA nanostructures (DNA stars and helix bundles) that attempt to solve this problem, based on our previous work on DNA structures that generate distinguishable current signatures in solid-state nanopores.[7,8] Although specific experimental settings (applied voltage, solution conductivity, pore size) affect the absolute values of our results, our broader conclusions should apply to most SiN nanopore devices. This is because we utilized standard operating parameters, state-of-the-art equipment, and pore sizes larger than the molecular structures. Using larger pores avoids a passage dynamics regime dominated by pore-wall friction, a phenomenon that is sometimes beneficial but difficult to control reproducibly. Varying these parameters is expected to yield only marginal improvements in the overall performance of the proposed barcoding schemes. Ultimately, these results provide a path toward designing optimized DNA nanostructured barcodes for multiplexed biosensing applications, which remains a significant challenge.

### 3.1.3 Results and Discussion

The concept of structured DNA barcodes translocating through nanopores is schematically outlined in **Figure 3.1a** for the case of a 2-bit polymer. Two-bit assemblies are sufficient to validate the performance of the nanopore system for distinguishing and correctly identifying the two different DNA nanostructures attached along a carrier that encode the binary signal. Important parameters needed to determine the decoding performance of a thin solid-state nanopore device – namely location of peaks, distinguishability of peaks, and determination of translocation direction – can be readily characterized with a 2-bit system. This approach is a first step toward multiplexed detection of biomarkers using nanostructured DNA barcodes as discussed above. A sample current trace of an experiment is shown in **Figure 3.1b** with a sample event trace of a typical 2-bit event shape shown in **Figure 3.1c**. To decode the binary information carried by the DNA carrier and tackle the challenges associated with optimizing bit spacing and bit differentiation, the task of the data analysis is to find the relative location of the current blockage peaks (generated by the bit nanostructures) within the event, as well as to extract their heights (i.e. blockage depths). For this purpose, we adapted a peak finding algorithm for use with nanopore current data (see Methods Section for more information). Briefly, for each set of experimental conditions, hundreds to thousands of single-molecule events are recorded, and from each event we identified the carrier blockage level by taking the maximum of the all-points histogram of its current trace values (see Section S1 in the Supporting Information). We then identified peaks within the event current traces using an algorithm that selects all points smaller in magnitude than their two neighbors and finds the separation of this point to the carrier level (its prominence) by looking for the left and right bases (the largest difference in current value in a window of 100 μs around the peak point). Only

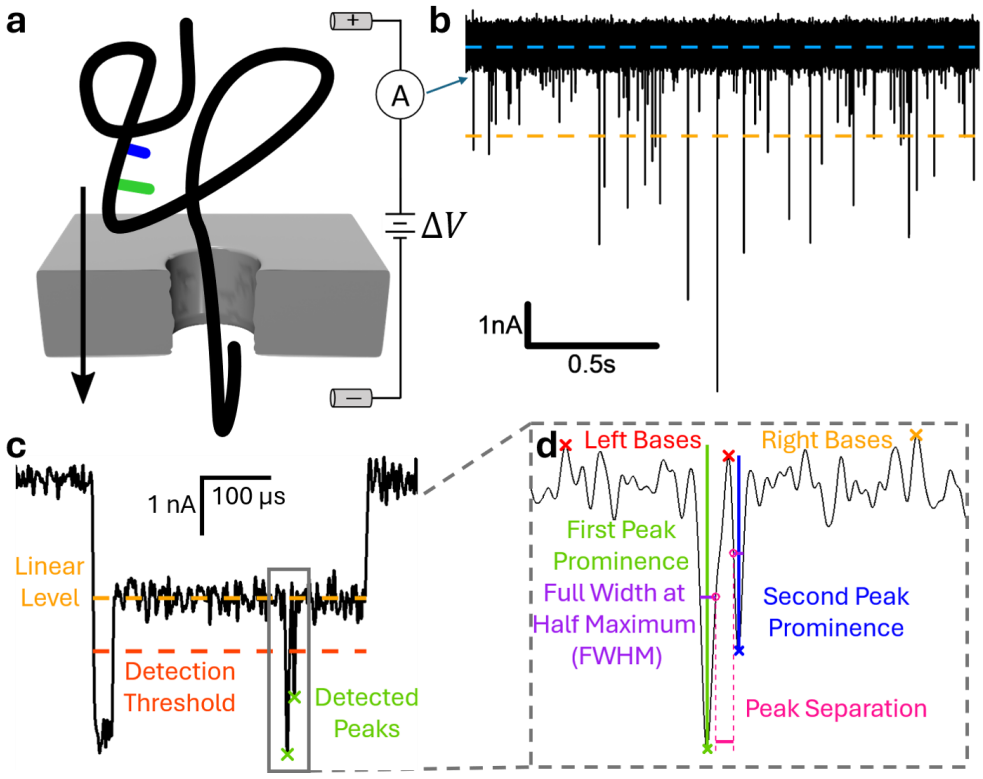



***Figure 3.1: Overview of peak finding in event traces. a)*** *Schematic of nanopore translocation of 2-bit DNA-assembly.* ***b)*** *A 2 s current trace of 2-bit DNA-assembly with triangle shaped attachments in 3.6 M LiCl at pH 8 in a 7.9 nm nanopore under a bias voltage of 300 mV. The open baseline current is shown with a blue dashed line, and the single-file level is marked with a yellow dashed line.* ***c)*** *Sample event trace of a 2-bit DNA-assembly, with the calculated unfolded level (yellow), the peak finding detection threshold (orange) and the identified peaks (green).* ***d)*** *Zoom-in of the region where peaks are detected for the event shown in c). Relevant peak finding parameters are highlighted: prominences (green and blue), peak bases (red and yellow), full width at half maximum (FWHM, purple, measured at half the prominence), and peak separation (pink).*

peaks with a prominence of at least six times the RMS noise (detection threshold in **Figure 3.1c**) of the baseline are considered. Once the peaks are identified, their widths are measured as the full width at half maximum (FWHM), specifically, the width at half the prominence value over the carrier level (see Section S2 of the Supporting Information for more details on the peak finding algorithm and peak/event selection steps). It is worth noting that the exact peak height is influenced not just by the size, shape, and flexibility of the molecular bit[202] on the carrier but also by the

velocity of translocation[53,118] (and therefore indirectly also by the dimensions of the nanopore), and the bandwidth of the measurement system. To calculate the separation between the peaks in the time domain of the current trace, we look for a rise of the current between the two peaks to a value within three times the RMS noise from the carrier level, ensuring that robust identification of both peaks is possible. The separation is then calculated as the time difference between the falling flank of the leading peak and the rising flank of the following peak as shown in pink in **Figure 3.1d**. The exact points of measurement are taken at the FWHM boundaries of the peaks (pink circles in **Figure 3.1d**). Measuring the separation between the flanks rather than between the peak centers acts as a conservative estimate of our ability to resolve adjacent peaks.

#### 3.1.3.1 Spacing of Bits

To investigate the minimal spacing between bits at the frequency bandwidth of our system, we used the molecular nanostructure shown schematically in **Figure 3.2a**. We note that the results documented here will apply most relevantly to our particular operating conditions (pores in 10–20 nm thick SiN membranes, immersed in high molarity LiCl salts, recorded at 1 MHz bandwidth and an RMS noise in the range of 120–300 pA at full bandwidth), which represent standards of high performance in the field. The 2-bit carrying polymer is assembled from single-stranded, linear DNA from the M13mp18 bacteriophage hybridized with staple strands forming a double-stranded carrier 7249 bp long, and with the two bits sitting at roughly 2000 bp from one of its ends (see Methods Section for more details). The attachment here is a simple loop of dsDNA with a single oligo binding to two overhangs from the carrier and forming a triangular shape. We assembled a total of four different versions of this structure with different molecular spacings between the two attachment positions for the bits. Results shown in **Figure 3.2** were obtained using three 8 nm

diameter pores in nominally 12 ± 2 nm thick SiN membranes at 300 mV in 3.6 M LiCl at pH 8. Sample event traces are shown in **Figure 3.2b** for different inter-bit spacings of 80 bp (27 nm, yellow), 120 bp (41 nm, light green), 160 bp (54 nm, dark green), and 200 bp (68 nm, blue). The shortest distance was set to match the common estimate of the sensing length of the pore, defined here as the membrane thickness plus twice the pore diameter.[53] As expected, the current signatures show two peaks overlaid on an underlying event shape that is governed by the dsDNA carrier

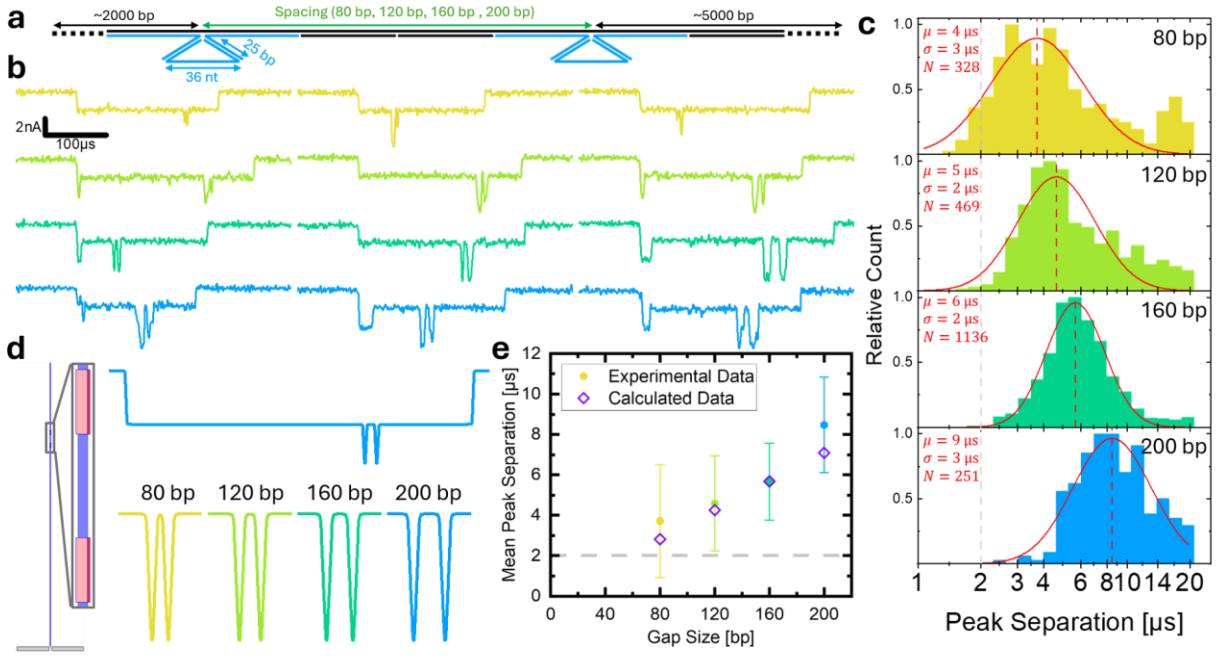



***Figure 3.2: Peak distance determination in nanopore current traces. a)*** *Schematic of the triangle 2-bit DNA design. The carrier is 7249 bp long with the two bits sitting at roughly 2000 bp along the contour.* ***b)*** *Sample event traces for 2-bit assemblies with different inter-attachment distances, color-matched to the distances shown in (c)-(e). These traces were recorded in 8 nm pores in 3.6 M LiCl pH 8 at 300 mV bias voltage, with a 500 kHz digital 8-pole low-pass Bessel filter and sampled at 4.167 MS/s.* ***c)*** *Histograms of peak separations for the different molecular spacings of the 2-bit triangle design (Number of events from top to bottom 328, 469, 1136, 251). Gaussian fits are shown in red with the means and standard deviations given in the inset; 2 µs resolution limit shown with a dashed grey line.* ***d)*** *Schematic and traces generated by the signal generator from a simple geometric model. Details of the geometric model are given in the Methods Section.* ***e)*** *Mean peak separations for all 2-bit triangle assemblies studied. The experimental data is extracted from the histograms in c). The error bars represent the standard deviation of the corresponding Gaussian fit. The purple diamonds represent the distances extracted from the signal generator traces. The time resolution limit of the system is again shown with a grey dashed line.*

(including potential folding at the beginning of the events). We find the blockage depth for these bit structures to vary significantly between events and between pores, making these structures unsuitable for determining the optimal blockage depth, but nonetheless useful for probing the peak spacing. We attribute the wide spread of blockage depths generated by these structures mainly to the flexibility of the single-stranded sections, which gives these structures a range of cross sections, and hence of blockage currents. These observations are quantified in **Figure 3.2c**, which shows the aggregated peak separations for all recorded events for the four different spacings (only separations <30 μs are plotted here; for the full distributions see Section S3 in the Supporting Information). The Gaussian fits for each spacing, shown in red, reveal a clear shift towards larger time differences for larger molecular spacings. The time difference shifts from 4 ± 3 μs for 80 bp spacing, 5 ± 2 μs for 120 bp spacing, 6 ± 2 μs for 160 bp spacing, to 9 ± 3 μs for 200 bp spacing. The time resolution limit of our system is estimated as approximately twice its rise time, i.e. ~2 μs here (grey dashed line in **Figure 3.2c**), due to a 500 kHz 8-pole low-pass Bessel filter being applied to the data before analysis, in order to maintain a signal-to-noise ratio above ~6 (in contrast to the higher 4.167 MS/s sampling rate and 1 MHz maximum bandwidth). This means that, at bit spacings of 160 bp and above, the most probable time separation is at least two standard deviations away from this limit and, thus, adjacent peaks can be differentiated in more than 95% of cases. It is worth noting that the dsDNA section in between adjacent bit attachments should behave like a semi-rigid rod and the translocation time of this section of the carrier is therefore only a function of the local velocity. Since the peaks are designed to lie sufficiently toward the middle of the carrier, where the velocity profile of a translocating DNA molecule is expected to be in its slowest, flattest section (for DNA lengths and pore sizes in this range, as determined by Charron *et al.*[118]), the inter-bit passage times should be optimally resolved and equivalent for both forward and

backward translocation directions. A simple linear relation is therefore expected between the molecular bit distance and the time separation measured in the current trace. It is worth reiterating that we are only considering events where the two peaks are distinguishable, based on the parameters we have set during peak finding analysis, which inherently biases the distributions toward larger separation times, due to the absence of lower values from the distributions (especially impacting the 80- and 120-bp spacing data). Our peak-finding algorithm considers two peaks separable if the current recovers to the carrier level between them. However, because peak separation is measured at FWHM, peaks separated by less than twice the rise time fail to recover to the carrier level and are misidentified as a single peak. The correlation between bit spacing and peak time separation is also supported by conductance calculations of the blockage produced by these nanostructures (see **Figure 3.2d**), which show similar trends. These generated signals are based on a purely geometric model and mimic event signatures by calculating the blocked state conductance of the system as a function of the distance from pore center to molecule center.[53] Assuming a uniform velocity profile, this distance information can easily be transformed into a time signal by fixing a total translocation time, here set at 400 μs, which lies in the middle of the dwell time distributions observed in the experiments. **Figure 3.2e** shows the means and standard deviations of the Gaussian fits from **Figure 3.2c** in comparison to the distances extracted from the theoretically calculated event signatures. For all four structures we find that the calculated values lie within the error interval of the measured data. The overall good agreement between the theoretically calculated event signatures and the experimental data shows the validity of using the simple geometric model to quickly validate the design on a rudimentary level and confirm that parameters like inter-attachment distance translate to proper separation of peaks in the resulting nanopore current traces. In particular, it easily illustrates the effect of the electrostatic length of the

pore on its ability to resolve closely spaced peaks for various pore dimensions and molecular geometries. Rather than capturing the full complexity of all electrokinetic phenomena, the modeled traces indicate the ideal resolution limit of the pore based on specific geometrical properties, access resistance, and changes in bulk conductance. This approximation is often sufficient to guide experimental design.

Based on the distributions of peak separations in **Figure 3.2c**, and the discussion above, we concluded that 160 bp is a safe minimal bit spacing for further bit design testing on this SiN nanopore platform, so as to ensure proper separation of peaks, as required for barcoding schemes using nanopore readout. Because translocation velocity decreases in larger pores,[118] this distance should yield separable peaks in a large range of pore sizes. It should be noted that, although this distance is specific to our pore geometry and measurement conditions, it lies well above a common estimate of the sensing length of the pore, approximated here as the membrane thickness plus twice the pore diameter. There should therefore be a point between the bits where the current, including the access-region contribution, is unaffected by either bit and can recover to the carrier level, as supported by the geometric modeling in **Figure 3.2d**. This can be reduced with the use of thinner membranes or of translocation speed reduction methods.[217,226,227] In fact, in a previous work using nanopipettes, a higher resolution was achieved, distinguishing smaller, tightly spaced attachments, down to 14 bp DNA dumbbells spaced at 18 bp intervals,[80] by translocating DNA from the inside of the pipette to the outside through a tightly sized pore. This special configuration resulted in slower translocation speeds enabling the higher resolution but also caused decreased capture rates and prolonged measurement times, making it impractical for most biomolecular barcoding schemes.

### 3.1.3.2 Bit Shape – DNA Stars

With the minimal spacing of bits on the molecular carriers experimentally validated for our particular nanopore devices and high-bandwidth measurement system, we now turn to address the challenge of distinguishing two different bit types in the event current trace, as generated by differently sized and shaped attachments on the carrier. Nanopore blockage is directly proportional to molecular size,[51] which means that the molecular structures to attach to the DNA carrier need to have specific shapes to generate two distinct current blockage levels on top of the signal of the carrier molecule. The particular triangle design of **Figure 3.2** exhibited a large spread of blockage depths even within the same nanopore experiment and is therefore unsuitable for this purpose. We have previously shown that dsDNA star-like structures generate distinguishable current blockage signals in solid-state nanopores[8] and therefore decided to use this as a basis for a two-bit system of attachments for DNA barcoding on dsDNA carriers. We chose a 4-arm 25 bp dsDNA star (encoding "0") and a 12-arm 25 bp dsDNA star (encoding "1") as our attachments to minimize overlap of the resulting nanopore blockage distributions. Similar structures have also been used as bits with glass nanopipette readout,[160] but performance on carriers translocating through thin pores in SiN membranes, which have shorter sensing regions and reduced electro-osmotic flow (EOF), remained unexplored.

The results of the DNA star 2-bit DNA carrier experiments are presented in **Figure 3.3**. A schematic view of the assembled structures is shown in **Figure 3.3a**. Similar to the triangle attachment shown in the previous section, the bits are positioned at roughly 2000 bp along the

contour of the 7249 bp carrier. The star attachments are assembled separately from the carrier molecule. To bind the stars to their assigned positions, each binding position DNA overhang has a unique sequence complementary to one of the two DNA stars. A linker strand (as highlighted in orange in **Figure 3.3a**) then ensures proper binding of stars to their assigned positions as demanded

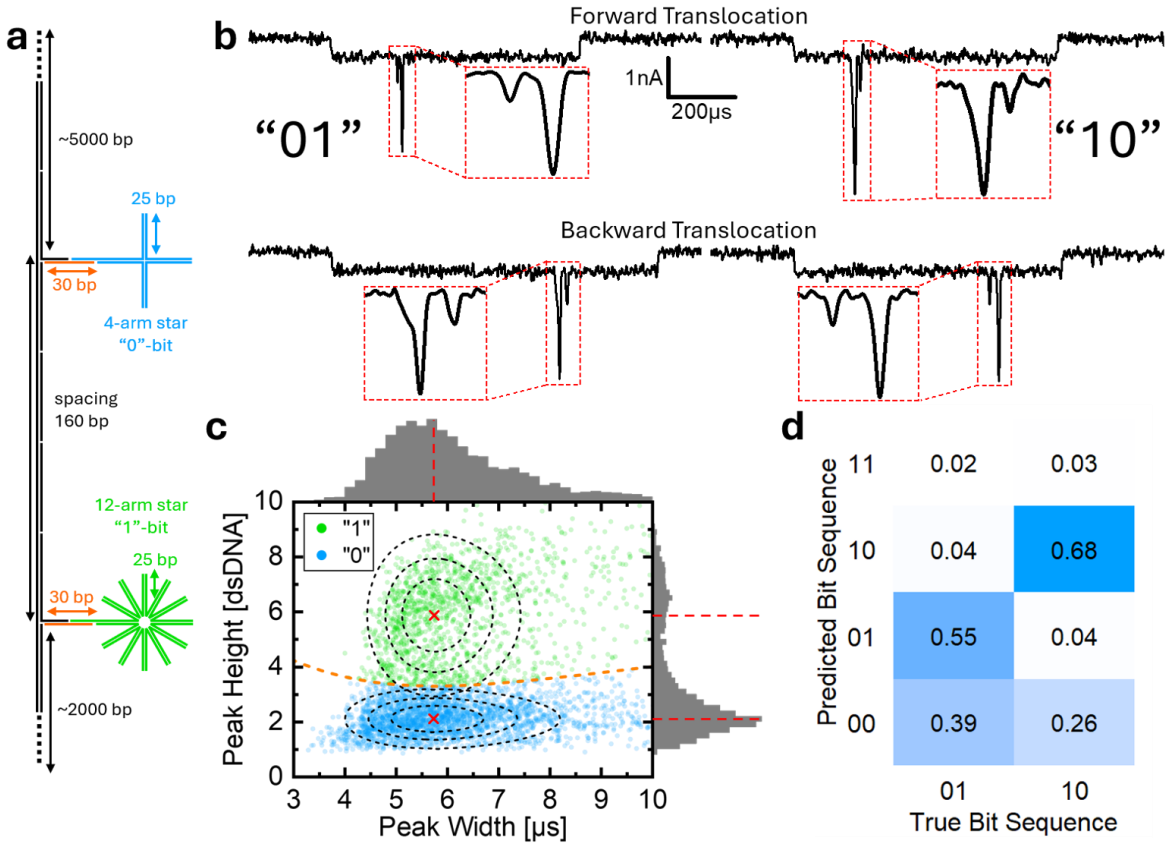



***Figure 3.3: Peak classification for 2-bit star assemblies. a)*** *Schematic of the 2-bit star assemblies with 4-arm (representing "0" bits) and 12-arm stars (representing "1" bits).* ***b)*** *Sample traces of correctly identified events for both 2-bit assemblies. Events are recorded in a 14 nm pore in 3.6 M LiCl at pH 8 and 150 mV bias voltage, with a 500 kHz digital 8-pole low-pass Bessel filter and sampled at 4.167 MS/s. The top row represents molecules translocating in the "forward" direction, and the bottom row represents molecules translocating in the "backward" direction.* ***c)*** *Scatter plot of all detected peak heights (normalized by the linear level for each event) and widths from a total of 2142 events (1159 events from "01" sample, 983 events from "10" sample) where two peaks were detected. The histograms along each axis are shown on the corresponding margin of the plot. The points are colored based on the assigned class in blue ("0") and green ("1"). The fitted 2D Gaussians are shown as black dashed contour lines (marking the areas where 25%, 50%, and 75% of peaks lie) with the peaks marked with a red cross at 2.1 ± 0.6 dsDNA blockages and 5.7 ± 1.2 µs for "0" and 5.9 ± 1.8 dsDNA blockages and 5.7 ± 1.2 µs for "1". Errors are the standard deviation along the corresponding axis. The orange dashed line marks the equal probability line between the two distributions.* ***d)*** *Confusion matrix of the classification of individual samples. The numbers show the fraction of events assigned to the corresponding class.*

by the specific barcode. We assembled two different versions of this structure with one “0” bit and one “1” bit in each of the positions; we denote the version with the “0” closer to the shorter end of the carrier as “01” and the one with the “1” bit closer to the shorter end of the carrier as “10” barcode. Categorized sample ionic current traces of these different molecular structures are shown in **Figure 3.3b** for a 14 nm pore in 3.6 M LiCl at 150 mV bias voltage. These selected event signatures reveal that this system is indeed able to produce differently sized peaks in the current traces as designed. The top row shows current traces identified as having translocated in a “forward” fashion while the bottom row shows events translocating “backward”. The translocation direction is determined by checking whether the peaks lie in the first half or the second half of the event trace (see Section S4 in the Supporting Information) since the molecular structure is asymmetric. To facilitate classification of peaks into the two designed categories, we plot the heights and widths of all peaks from events which show two distinct peaks as identified by our peak finding algorithm. **Figure 3.3c** shows the resulting scatter plot of the blockage depth of each peak, normalized by the linear carrier blockage level, versus the peak dwell time from a total of 2142 events, and displays two clear but overlapping populations resulting from the two different bit shapes. The “0” and “1” peaks here are mainly separated by different blockage depths rather than dwell times. The overall dataset is fit to a bimodal 2D Gaussian distribution (see Methods Section). Due to the relatively small sample sizes of the data sets, a simple 2D fit was deemed most appropriate over more complex clustering methods. The peak positions (red crosses) are at 2.1 ± 0.6 dsDNA blockages and 5.7 ± 1.2 μs for the “0” bits distribution and at 5.9 ± 1.8 dsDNA blockages and 5.7 ± 1.2 μs for the “1” bits distribution. The black contour plots mark the areas where 25%, 50%, and 75% of the events fall into. By selecting only events that show two peaks, we remove about 62% of all full-length carrier single-molecule events recorded. We calculated,

for each peak of both samples, the probability of belonging to either distribution and assigned them to the distribution that gave the higher probability therefore classifying them as either “0” or “1” as indicated by the colors in **Figure 3.3c** (see Section S5 of the Supporting Information for individual distributions for each of the samples). The line of equal probability between the two distributions is marked with a dashed orange line in the plot. Ultimately, this dividing line between 0 and 1 does not significantly alter the results shown here since the underlying mechanisms of the shortcomings are for the most part rooted in the misassembly of barcodes and the limited bandwidth of the nanopore setup. This is supported by Section S8 in the Supporting Information which shows that even removing edge cases for classification does not significantly impact overall performance. For events with either two “0” peaks or two “1” peaks the translocation direction is irrelevant, and classification is complete; for events with one “0” peak and one “1” peak, we also consider the translocation direction as described above. With the translocation direction and the peak classification, we can assign each event a binary barcode and the resulting confusion matrix of this assignment for experiments with each individual species of barcode is shown in **Figure 3.3d**. It emerges from the analysis of this confusion matrix that the “1” peaks (generated by 12-arm dsDNA star attachments) pose identification challenges (see SI Section S6 for additional event signatures highlighting the variability of the “1” bit peak height). This is supported by the fact that a significant fraction of events is classified as “00” even when one or two of the peaks are generated by 12-arm stars. We see a significant portion of peaks lying in the range where 4-arm star peaks should lie, leading to those peaks being misclassified as “0”. We speculate that the reasons why the 12-arm star structures do not generate larger blockage levels more reliably stem from a combination of factors including: (i) the disassembly of 12-arm dsDNA stars (or improper assembly in the first place) bound to the carrier leading to smaller structures being sensed; (ii) the

flexibility of the arms such that only a fraction of arms are inside the pore at any given time (as supported by the fact that the mean observed blockage depth is << 12× that of dsDNA); and (iii) the attenuation of the signals from fast translocations of the 12-arm stars moving alongside a carrier molecule through the pores (featuring larger diameters than in our previous study[8]). Recent work with nanopipettes has observed a similar trend, where flexible attachments generate shallower blockages than expected, which supports our findings.[202] A redesign of the bit structure offering longer shapes and better structural stability should in principle solve these issues. With the loss of detectability for the smaller 4-arm star structures in larger pores, it is deemed unfeasible to simply increase the size difference of these star structures by, for example, increasing the number or length of arms for the "1" bit structure to improve distinguishability of the different species of peaks. The fact that even the most distinguishable star structures from our previous study,[7,8] when measured individually, do not make for good separation on the carrier in these experiments, shows that a more fundamental redesign of the attachment structure is needed to design robust molecular bits.

#### 3.1.3.3 Bit Shape – DNA Helix Bundles

Inspired by a design proposed by Seeman and coworkers,[228] we devised a subsequent two-bit system design in an attempt to fix the shortcomings of the DNA star bit system. In this iteration of the 2-bit assembly, we attach a 63 bp (21 nm) 3-helix dsDNA bundle (3HB) onto the main carrier strand as the "0" bit, while a longer, 147 bp (50 nm) version of the 3-helix bundle functions as the "1" bit, as shown schematically in **Figure 3.4a**. The detection principle is based on the fact that the short 63 bp 3HB will translocate too fast to be fully resolved with the measurement bandwidth and response time of our nanopore device, thus providing a short and shallow peak labeled as "0" bit, while the long 147 bp 3HB is long enough to reach the expected, full blockage

level of 3× dsDNA. Moreover, further peak separation should be possible with this strategy based on the duration/width of the peaks (in addition to their height), giving potentially less overlap of the resulting distributions and improving peak classification. The analysis of these structures is performed in a similar fashion to the analysis for the star structures outlined above. For this design

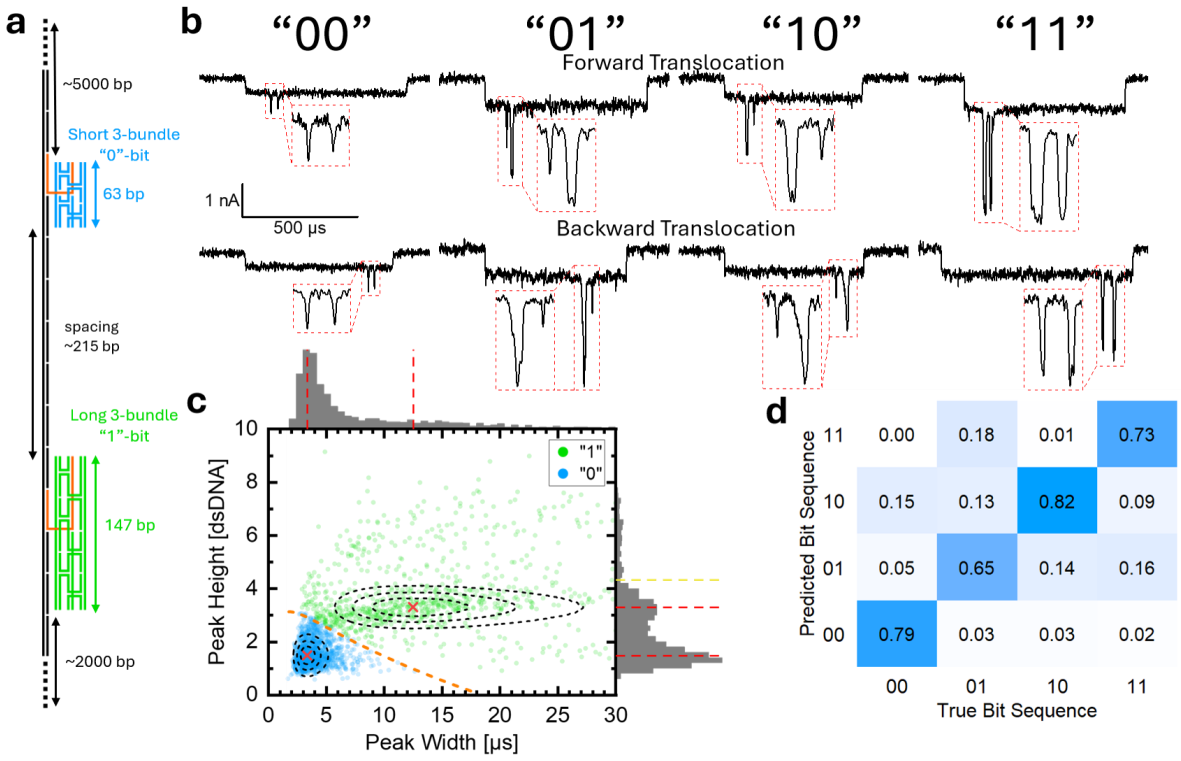



***Figure 3.4: Peak classification of 2-bit bundle assemblies. a)*** *Schematic of the 2-bit 3HB assemblies with short 3HB (blue, representing "0" bits) and long 3HB (green, representing "1" bits).* ***b)*** *Sample traces of correctly identified events for all four different 2-bit assemblies. Events are recorded in 4 pores ranging from 12 to 14 nm in size in 3.6 M LiCl at pH 8 and 200 mV bias voltage with a 500 kHz digital 8-pole low-pass Bessel filter and sampled at 5 MS/s. The top row represents molecules translocating in the forward direction and the bottom row represents molecules translocating in the backward direction.* ***c)*** *Scatter plot of all detected peak heights (normalized by the linear carrier level for each event) and widths from a total of 711 events (355 events from "00" sample, 126 events from "01" sample, 111 events from "10" sample, 119 events from "11" sample). The histograms along each axis are shown on the corresponding margin of the plot. The points are colored based on the assigned class in blue ("0") and green ("1"). The fitted 2D Gaussians are shown as black dashed contour lines (marking the areas where 25%, 50%, and 75% of peaks lie) with the peaks marked with a red cross at 1.5 ± 0.5 dsDNA blockages and 3.4 ± 1.3 µs for "0" and 3.3 ± 0.5 dsDNA blockages and 12.5 ± 1.6 µs for "1". Errors are the standard deviation along the corresponding axis. The orange dashed line marks the equal probability line between the two distributions.* ***d)*** *Confusion matrix of the classification of individual samples. The numbers show the fraction of events assigned to the corresponding class.*

we assembled all four possible 2-bit barcodes denoted “00”, “01”, “10”, and “11” in similar fashion to the star structures in the previous section.

**Figure 3.4b** shows a selection of correctly classified events for each of the samples and translocation directions, using data recorded in 4 different nanopores ranging from 12 to 14 nm in size in 3.6 M LiCl at pH 8 and 200 mV bias voltage. These pooled results highlight the two distinct peak types of this design, distinguishable by blockage depth (i.e. peak height). We also note that the deeper peaks representing “1” bits tend to be longer in duration, as expected, often reaching a brief, multi data point blockage plateau suggesting that these peaks are fully time-resolved with the settings used in these experiments (500 kHz 8-pole low-pass Bessel filter, 5 MS/s sampling frequency). As intended, the short 3HB “0” bits are faster, narrower in subevent width, not fully time-resolved, and therefore do not reach the same peak heights. The extracted peak heights and widths for all events recorded are plotted in **Figure 3.4c** together with the 2D Gaussian fit for both populations (see Section S7 in the Supporting Information for full distributions of all individual samples, the fit is only using points from the “01” and “10” samples). The red crosses mark the means of the distributions at 1.5 ± 0.5 dsDNA blockages and 3.4 ± 1.3 μs for “0” and 3.3 ± 0.5 dsDNA blockages and 12.5 ± 1.6 μs for “1”, with the errors indicating the standard deviations along the corresponding axes. The black contour lines mark areas where 25%, 50%, and 75% of peaks fall into. Beyond the two main populations for “0” and “1” peaks differing in peak height and duration, we also note the presence of much deeper peaks with blockage values larger than expected from the geometric shapes of the assemblies (marked with a yellow dashed line in **Figure 3.4c**). We hypothesize that these originate from the attachments rotating around their single staple strand attachment point to the carrier and translocating unaligned with the carrier axis. A similar

phenomenon has been simulated for small dumbbell structures.[159] Since this effect will be stronger for longer bundles, we assume that the vast majority of these long and deep peaks are more likely to be generated by the long bundle as opposed to the short bundle, which is only a couple of nanometers longer than the diameter of the pore and should therefore more readily align with the pore axis before or during translocation. This interpretation is supported by looking at the individual peak distributions in **Figure 3.11** in the Supporting Information, where we do not observe any deep and long peaks in that range for the "00" sample. Finally, we show the results of this classification in the confusion matrix in **Figure 3.4d**. Correct classification ranges from ~65% to ~82% across the 2-bit sequences, with an average true positive rate of ~75%. The values and variations observed in the off-diagonal elements in the confusion matrix can be explained by the differences in translocation kinetics of the "0" bits across the differently sized pores used with these samples, leading to different levels of resolvability (and therefore different rates of misclassification) as peaks shift to larger blockages which leads to a larger percentage of peaks landing on the other side of the equal probability line. This can to some extent be addressed by selecting events more rigorously and only keeping events within the 80% confidence region of either peak distribution (see Section S8). Another feature in this confusion matrix is the asymmetry of classification of "01" and "10" barcodes which most likely originates from folded events being removed during event selection. Since folding typically only happens at the beginning of the event,[10,11,81] events translocating in the forward direction (with the peaks closer to the beginning of the event) will have a higher chance of peaks being positioned within a folded level. In those cases, calling the correct ordering of the peaks becomes ambiguous and these events are therefore removed from downstream analysis. This creates an imbalance in the number of forward and backward translocations used in the classification of events, with roughly 1.5 times as many

backward as forward translocations across all experiments. Part of this discrepancy lies in the uncertainty in calling event direction simply based on position of peaks (as described in Section S4). Extra structure in the underlying event signature (e.g. folding, knots, etc.) can influence this determination without impacting the ability to call peaks correctly. The main hindrance in classification now seems to be in making sure that different molecular conformations of the same bit structure give the same blockage depth and dwell time. Differences in translocation times as a result of different orientation of the bit on the carrier might shift the resolvability of peaks originating from this design making it not just different from pore to pore but also from event to event on the same pore. This can cause, occasionally, even the short 3HB structures to become fully resolved, leading to an increased likelihood of misclassifying the "0" as a "1" bit. It should be noted that the reliance on the "0" peak not being fully resolved defines a range of experimentally usable pore sizes (> ~8 nm, < ~15 nm), as they need to be large enough for the barcodes to translocate, without too much friction with the pore wall, but not so large that the short 3HB structures are missed entirely, nor so small that they clog the pore. Restricting classification to events whose peaks are within the 80% confidence region of either distribution improves some of the classification metrics in the confusion matrix but does not completely solve these issues (see **Figure 3.12** in the Supporting Information) indicating that it is not only peaks lying close to the intersection of the two distributions that are responsible for misclassification. This implies that these issues most likely originate at a more fundamental level (e.g. nanostructure assembly, nanopore-analyte interactions, etc.) and are not purely artifacts of the analysis workflow or system bandwidth.

### 3.1.4 Conclusion

In this study we investigated how the design of DNA-based molecular barcodes affects the decoding performance for solid-state nanopores fabricated on thin SiN membranes (10–20 nm thick) and measured with state-of-the-art MHz-range bandwidth amplifiers. We used three different architectures of DNA nanostructured barcodes to investigate the minimally resolvable bit spacing and the effect of bit sizes and shapes on the ability to correctly classify "0" and "1" bits. We found the safe minimal molecular spacing to be 160 bp (54 nm) for an experimental setup with a ~2 μs response time (500 kHz low-pass filtered signal, sampled at 4.167 MS/s or 5 MS/s), operating outside of pore-wall friction dominated translocation regime, in order to balance density and resolution of peaks in the current traces. We showed that peak classification into the correct bit can be improved by changing to a 2D classifier that uses both peak duration and depth. However, current nanostructures are only able to call the correct barcode up to 82% of the time. This system can be improved with more robust designs that have higher assembly yield and more rigid structure to generate more consistent current traces in the nanopores, especially in pores of different sizes. This would also lead to cleaner distributions in peak height and width, which would further facilitate distinction using the fitting method presented here. This is another step on the path to building robust DNA barcodes that can be read using solid-state nanopores, toward DNA-based multiplexing designs, as well as DNA data storage systems.

### 3.1.5 Methods

#### 3.1.5.1 DNA Sample & Assembly

DNA assemblies were prepared by mixing an M13 scaffold (Tilibit Nanosystems M13mp18, type p7249, 7249 nt long) with five-fold excess of an equimolar mixture of the required oligos (Integrated DNA Technologies) in 1x TE buffer with added 10 mM $MgCl_2$ (see Section S9 in the Supporting Information for the full set of sequences of all the assemblies). The mixture was heated to 70 °C for 10 minutes and then cooled down to room temperature (25 °C) over the course of 30 minutes. Prior to assembly the scaffold was linearized with the restriction enzyme HincII (New England Biolabs, R0103S). Excess oligos were removed with the help of Amicon 100 kDa molecular weight cutoff filters (Millipore Sigma, UFC510096). For triangle assemblies, the attachment oligo was mixed in with the carrier assembly mix. Star and bundle attachment structures were assembled separately by mixing all involved oligos at equimolar ratio in 1x TE buffer with added 10 mM $MgCl_2$. The mixture was heated to 95 °C and slowly cooled to room temperature over the course of 48 hours. The final product was cleaned up using solid-phase reversible immobilization (SPRI) beads (HighPrep PCR Magbio, AC-60005).

#### 3.1.5.2 Nanopore Measurements

Nanopores were fabricated in 10–20 nm thick SiN membranes using the controlled breakdown (CBD) method in a custom design instrument as described in *Waugh et al*.[48] Nanopore fabrication was performed using a voltage ramp in 1 M KCl, 10 mM HEPES, pH 8 until breakdown was detected. Nanopore enlargement was then performed in 3.6 M LiCl, 10 mM HEPES, pH 8 until the desired size was achieved. Nanopore dimensions were extracted using the conductance

model described previously.[51] All experiments were performed in 3.6 M LiCl 10 mM HEPES pH 8. For a list of all pores and experimental parameters used in this study, refer to Section S10 in the Supporting Information. Nanopore current recordings were performed on the Chimera VC100 or VC400 amplifiers at 4.167 MS/s and 5 MS/s sampling frequency, respectively. Before analysis the data was low-pass filtered to 250 kHz or 500 kHz depending on signal-to-noise ratio using an 8-pole digital Bessel filter. The value of the cutoff frequency was chosen on a per-experiment basis, to maximize temporal resolution while maintaining a good signal-to-noise ratio.

#### 3.1.5.3 Data Analysis

Translocation events in the recorded current traces were located using a custom implementation of the CUSUM+ algorithm.[211] Custom Python and R scripts were used for plotting and data selection. Peak finding was performed using the find_peaks function from the SciPy library in Python with minimum prominence set to 6σ of the baseline noise. An implementation of this peak finding algorithm is available in Poriscope.[229] Peak distributions for DNA star- and bundle-shaped attachments were fitted using the curve_fit function of the SciPy library.

#### 3.1.5.4 Signal Generation

Synthetic nanostructured DNA barcode signals were generated using the nanopore signal generator on the TCossaLab website (https://www.tcossalab.net/signalGenerator).[53,218] The DNA carrier was modeled as a cylinder of radius 1.1 nm with lengths of 600 nm (~2000 bp, front) and 1500 nm (~5000 bp, back). The spacing between the bits was modeled as cylinders of radius 1.1 nm with length 27 nm, 40 nm, 54 nm, 68 nm (~80 bp, ~120 bp, ~160 bp, ~200 bp). The bits were modeled as a 90° wedge with R1 of 1.1 nm and R2 of 2.2 nm and length 15 nm (~50 bp). The pore

was taken as a cylindrical pore with radius 4.0 nm and length 12 nm with a resolution of 10000. To translate the distance trace to a time trace, the event length was taken as 400 μs (the length of a typical event for the 7249 bp scaffold), with a cutoff frequency of 500 kHz and noise of $0.01\sigma$.

**3.1.5.5 Acknowledgements**

We would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2021-04304], and of Oxford Nanopore Technologies. We would also like to thank Martin Charron for valuable discussions and help in writing this paper.

# 3.2 Supporting Information: Nanostructured DNA Barcode Decoding with Solid-state Nanopores: Impact of Bit Spacing and Bit Shape

*Philipp Mensing[1], Nada Kerrouri[1], Sofia Zaynullina[1], Zachary Roelen[2], Ping Xie[3], and Vincent Tabard-Cossa[1]*

*[1]Department of Physics, University of Ottawa, Ottawa, ON, Canada*

*[2]Oxford Nanopore Technologies Canada, Ottawa, ON, Canada*

*[3]Oxford Nanopore Technologies, Cambridge, MA, USA*

## 3.2.1 S1. Determining the Linear Carrier Level

One of the first steps in event analysis is the identification of the linear level, which corresponds to a single dsDNA strand's worth of blockage. This is used to normalize blockages across multiple pores and experiments, and as a baseline from which to measure the prominence of peaks. The presence of peaks along the current trace makes proper fitting of sublevels difficult, as the peaks themselves can sometimes be missed by fitting algorithms such as CUSUM,[211] which influences the determination of the average current for these sublevels. Here, we determined the linear carrier level with the following algorithm, illustrated on a sample event in **Figure 3.5a**. First, we trimmed the open-pore current baseline from the trace to obtain the current trace shown in **Figure 3.5b**. We then defined a moving window with width equal to the standard deviation of the open-pore baseline (1σ of the current noise). This window was then iterated through the span of

current values covered by the event, while counting the number of data points that fall in the window. The result of this process is plotted in **Figure 3.5c**. For folded events, such as the one used as an example, we obtain a bimodal distribution with two obvious peaks. Only events whose most prominent level is the unfolded level are considered for downstream analysis, so we can simply take the maximum of this distribution as the linear carrier level. Events for which the linear level determined here does not lie in the range of the unfolded level are discarded from the analysis, as described in Section S2.

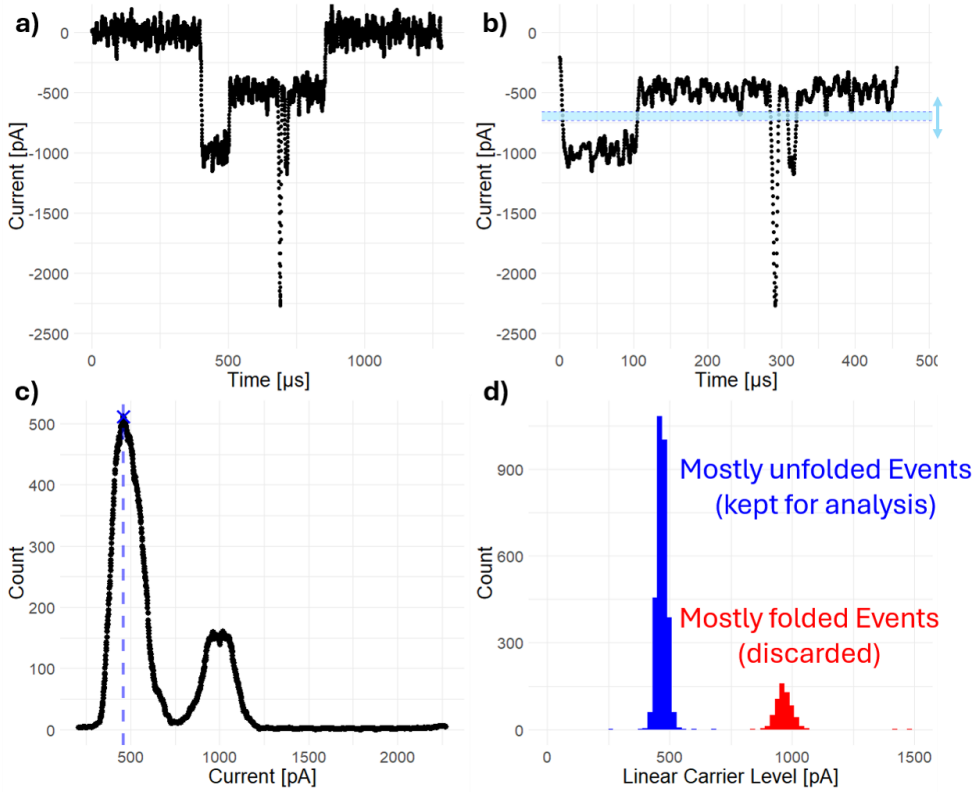



***Figure 3.5: Determination of linear dsDNA blockage level. a)*** *Sample event trace from the "01" star attachment structures.* ***b)*** *The same event trace with the baseline level trimmed away. The blue box marks a sample position of the moving window with the width of the standard deviation of the baseline level of the event.* ***c)*** *Count of the number of data points for each window position in b). The two peaks represent the two prominent levels in the event trace (folded and unfolded level). The blue line and cross mark the location of the maximum in this set, which is taken to be the value of the unfolded level for this particular event.* ***d)*** *Aggregated linear carrier levels for all events in the "01" star attachment experiment. The two peaks in this distribution represent mostly unfolded events (blue) and mostly folded events (red). This distribution is the basis for event selection later in the analysis.*

### 3.2.2 S2. Event Selection

Due to the presence of fragmented carrier molecules in the samples as well as leftover attachment molecules, correctly assembled, full-length translocation events need to be selected from the background. We have previously discussed how to select events of full-length M13 scaffold strands and reject those of smaller fragments.[10] Since the attachment molecules on their own are much shorter than the M13 carrier molecule, rejecting the short carrier fragments also removes all events from free-floating individual attachment molecules. This first selection step is done simply by setting lower bounds on the maximum blockage depth and electric charge deficit (ECD) of the events and discarding events that do not reach these values. The exact cutoff values depend on the pore parameters and change slightly for each pore used in this study.

A second selection step is then applied after running the peak finding on the events passing the first step. This step identifies peaks that are either located in the folded level or too large to have been caused by the designed attachments. Peaks whose bases do not fall within the noise band of the linear carrier level are discarded and not considered for downstream analysis. Identifying peaks in the folded level of the event is possible in principle, but determining the direction of the bit sequence is inhibited by the fact that it is not clear on which end of the carrier the peaks lie. To prevent the peak finding from triggering on short initial folded levels, peaks found within the first 20 μs of the event are also discarded. Additionally, any peak wider than 50 μs is removed from the analysis, as this could indicate misfolding or the detection of DNA folding/knotting as peaks.

All peaks passing the criteria above are, in principle, considered to be true peaks generated by the designed nanostructures. A final selection step then retains only events that show two peaks in the linear level and generally follow an easily interpretable translocation process (mostly linear translocation with minor folding not affecting the peaks). To achieve this, we selected events based on the linear carrier level and kept only events for which this level lies in the distribution identified as corresponding to the actual linear DNA level (see **Figure 3.5d**).

Of all events deemed full-length by the first selection step, on average 38% pass all subsequent selection steps across all experiments performed. A few example traces of events failing these selection criteria are given in **Figure 3.6**.

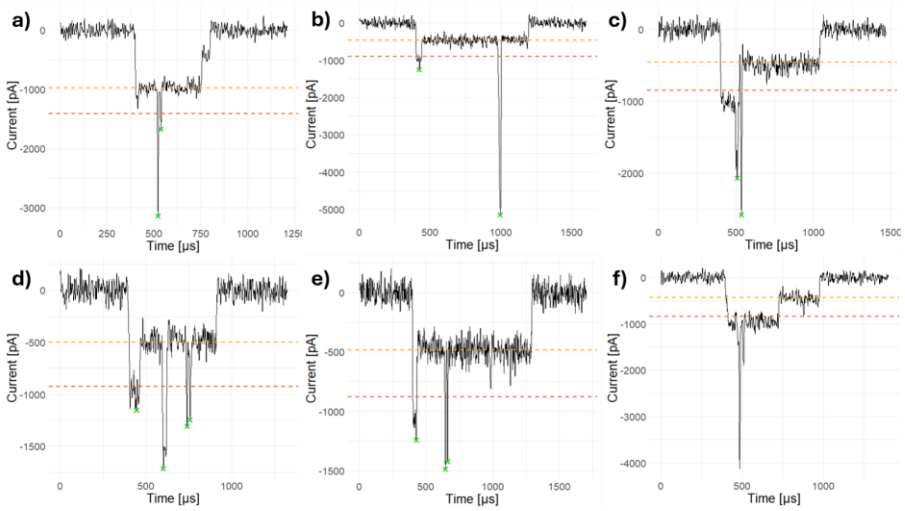



***Figure 3.6: Example traces of common error modes during translocation of 2-bit assemblies.*** *All traces shown here are recorded in an 8 nm pore in 3.6 M LiCl at pH 8 under 300 mV bias voltage. The determined linear carrier level is shown as an orange dashed line for each event, the peak detection threshold is shown as a red dashed line, and identified peaks are marked with green crosses.* ***a)*** *The event is mostly folded, meaning that all peaks are found in the folded level.* ***b)*** *The two main peaks are either merged or one of them is missing, but the peak finding triggered on the folded level at the beginning of the event. Events of this type can pass all the defined selection criteria, but in this case the two peaks are too far apart.* ***c)*** *The linear level is correctly identified but the peaks are partially in the folded level leading to potential confusion about the translocation direction.* ***d)*** *Additional peaks identified due to the presence of DNA knots or similar secondary DNA structures.* ***e)*** *Additional peak identified in the folded level on top of two correctly identified attachment peaks.* ***f)*** *Peaks located in the folded level are discarded during peak selection, leading to an apparent no-peak event.*

### 3.2.3 S3. Full Peak Separation Histograms

**Figure 3.7** below shows the full histograms of the peak separations measured using the triangle-shaped attachment. For the fitting of peak separations as discussed in the main text, only peak separations less than 30 μs are considered. The long tail of the distribution shown in the full histogram is mostly caused by the peak finding triggering on an initial short, folded level in the current trace. If one of the actual peaks is subsequently missed by the peak finding (e.g., if the peaks are merged because they are too close together), the event will pass the selection step that keeps only events with two peaks. This is supported by the fact that this secondary, longer distribution spans almost the entire range of durations up to the average event length. Additionally, this longer distribution appears more prominently for the smaller spacings because the peaks are more often merged in those cases. This indirectly corroborates the conclusion in the main text that peaks are more often merged for the 80 bp and 120 bp spacings, while this happens much less frequently for the 160 bp spacing and almost never for the 200 bp spacing.

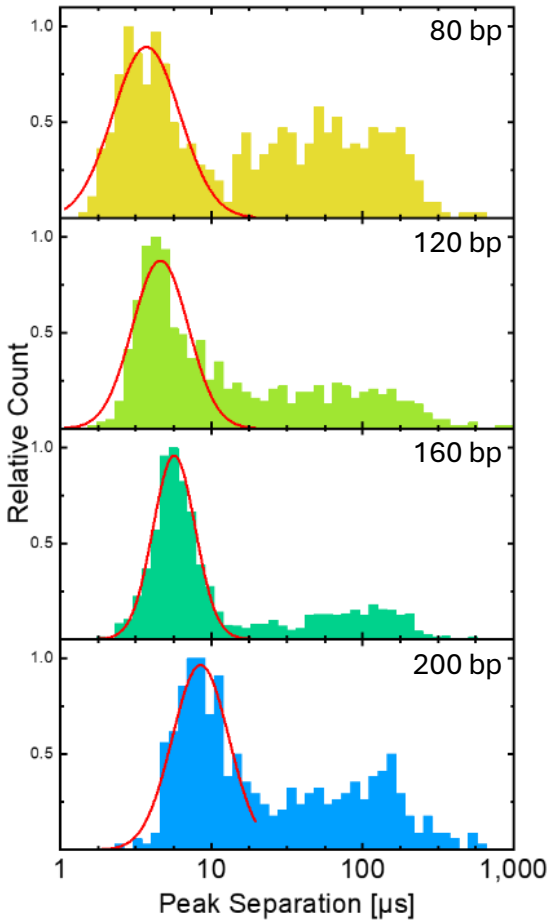



***Figure 3.7: Histograms of all peak distances measured with assemblies using the triangle-shaped bit design.*** *Peaks separated by more than ~30 µs are considered to originate from the peak finding triggering on the initial folded level of the event and missing one of the main peaks of the structures.*

### 3.2.4 S4. Determining Translocation Direction

All attachments used in this study are positioned about 25% along the contour. This facilitates determining the translocation direction of events. To that end, we calculate the partial electric charge deficit (pECD) of the event current trace before the first peak (green area in **Figure 3.8**) and after the last peak of the event (blue area in **Figure 3.8**). If the pECD before the first peak

is larger than the pECD after the last peak, the event is considered to be translocating in the forward direction; otherwise, it is considered to be translocating in the backward direction.

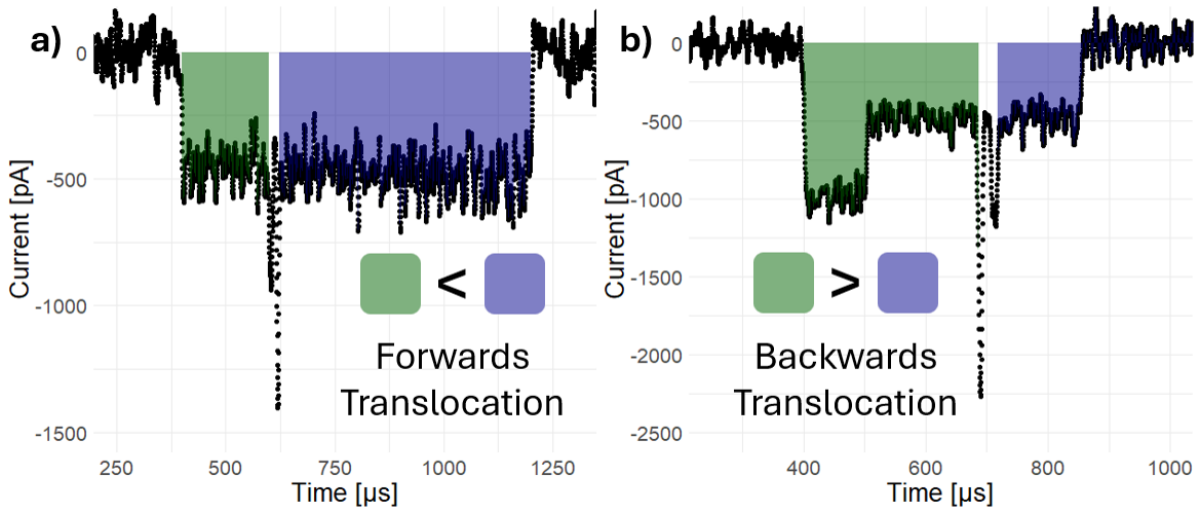



***Figure 3.8: Determining the translocation direction for 2-bit assemblies. a)*** *Sample event determined to have translocated in the forward direction; the green and blue areas mark the pECD of the corresponding sections of the event.* ***b)*** *Sample event determined to have translocated in the backward direction.*

### 3.2.5 S5. Full Peak Distributions for Star Attachments

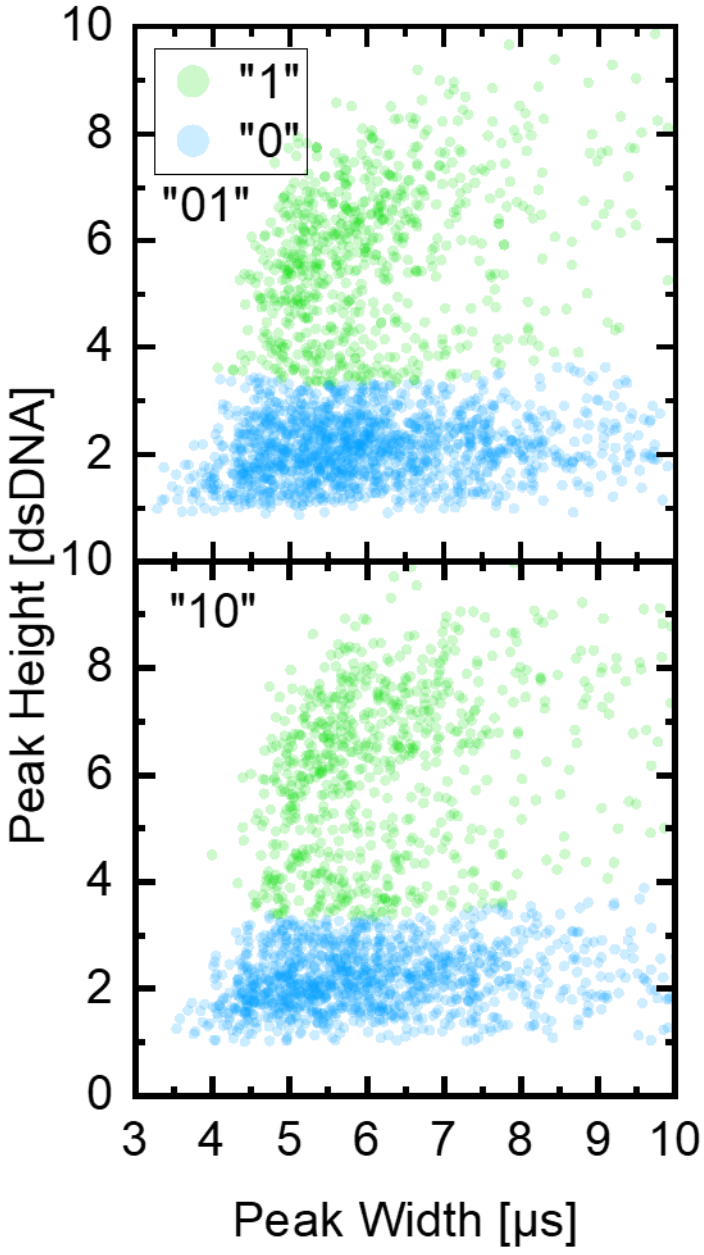



***Figure 3.9: All-points scatter plot for the peaks detected in the assembly using star-shaped bit attachments.*** *Peaks classified as "0" bits are marked in blue while peaks classified as "1" are marked in green. Classification of the peaks is based on the 2D Gaussian distribution obtained from fitting the combination of "01" and "10" samples as described in the main text.*

### 3.2.6 S6. Additional Event Traces for Star Attachments

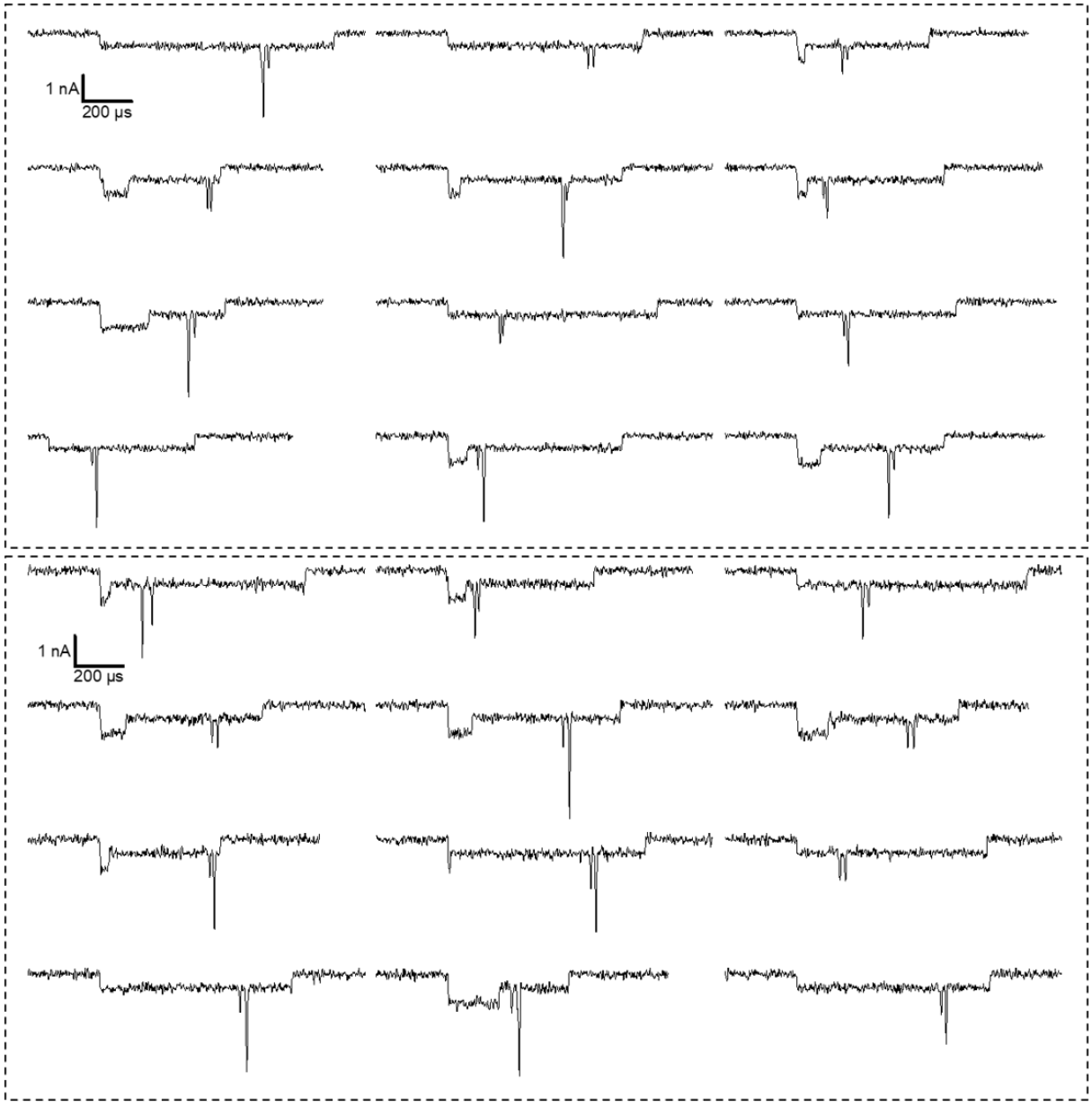



***Figure 3.10: Sample event traces for 2-bit assemblies.*** *Selection of events for the "01" (top) and "10" (bottom) samples with star-shaped attachments in an 13.7 nm pore with 150 mV bias voltage in 3.6 M LiCl at pH 8. This selection highlights the difference in blockage depth for the 12-arm star attachment and the difficulty in classification resulting from that as described in the main text.*

### 3.2.7 S7. Full Peak Distributions for Bundle Attachments

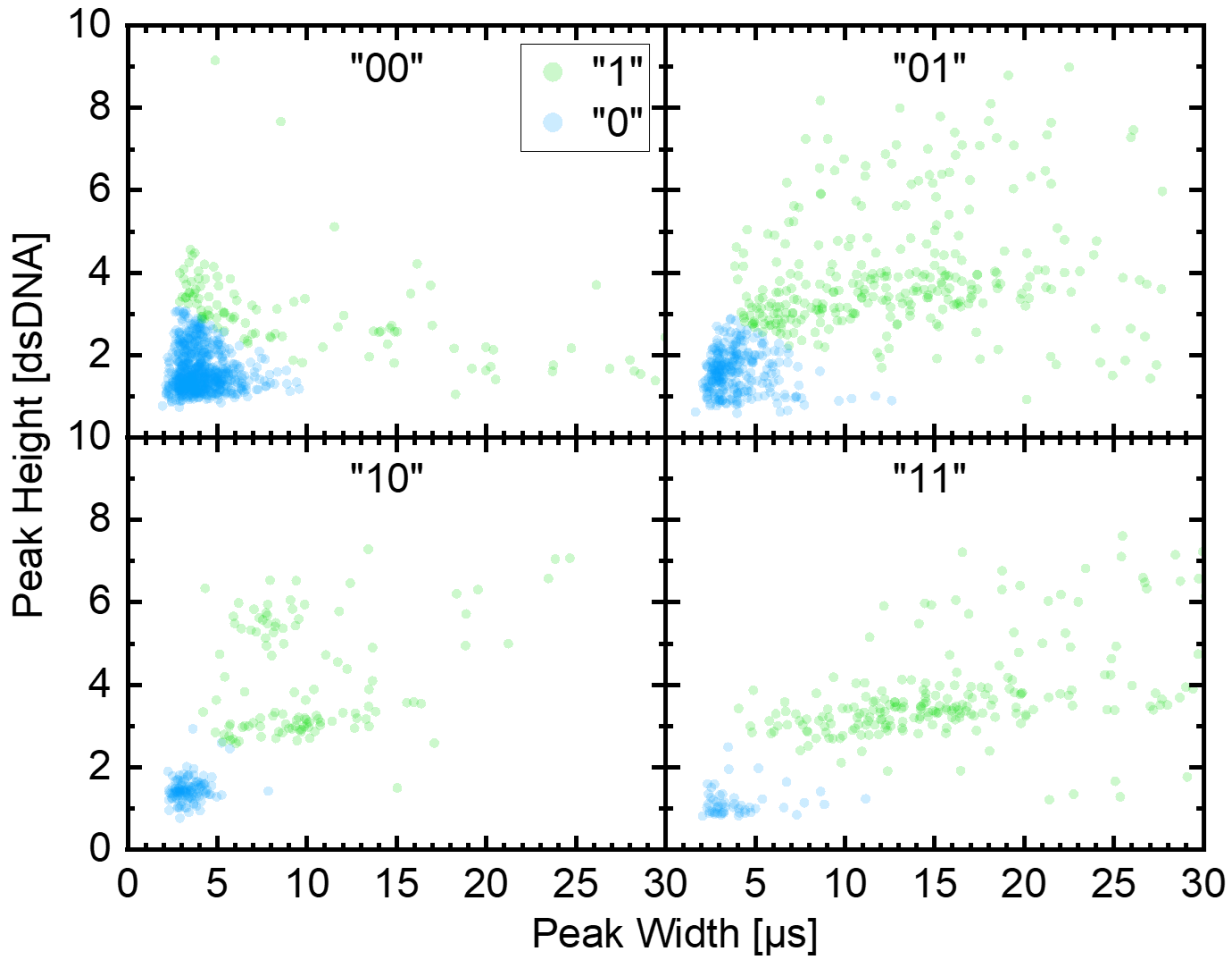



***Figure 3.11: All-points scatter plot for the peaks detected in the assembly using bundle-shaped bit attachments.*** *Peaks classified as "0" bits are marked in blue while peaks classified as "1" are marked in green. Classification of the peaks is based on the 2D Gaussian distribution obtained from fitting the combination of "01" and "10" samples combined as described in the main text.*

### 3.2.8 S8. Additional Event Filtering for Better Classification

To investigate if classification performance can be improved with stricter rejection criteria, we applied additional filtering after peak distribution fitting. We removed all events in which at

least one peak lay outside the 80% confidence region of either peak in the 2D Gaussian distribution (this removed about 58% of the remaining events). The resulting confusion matrix for the classification of these events is shown in **Figure 3.12**. While the diagonal elements do improve for the "00" and "01" samples, the impact on the classification for the "10" and "11" samples is minimal.

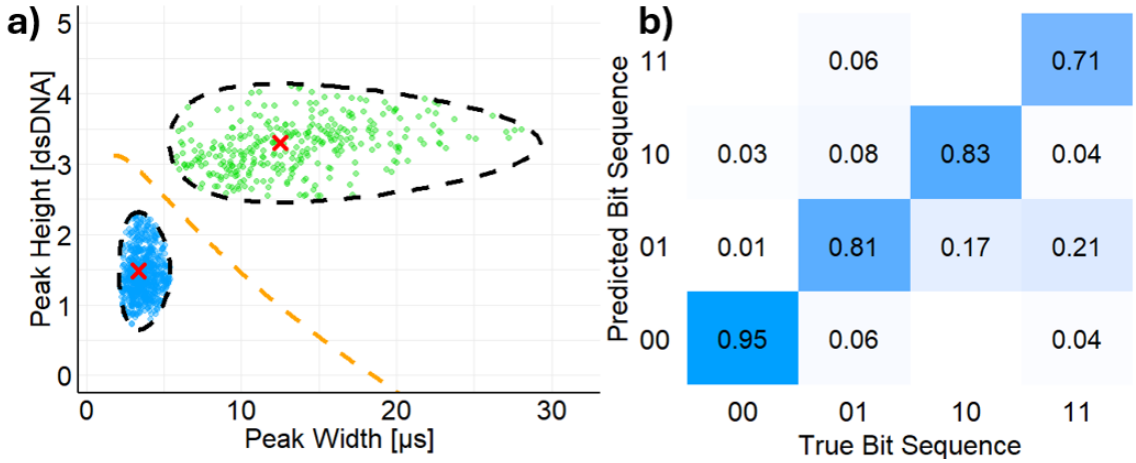



***Figure 3.12: Filtered peak classification. a)*** *Filtered all-points scatter plot for the peaks detected in the assembly using bundle-shaped bit attachments for all samples. Included are only peaks that fall into the 80% confidence region of the two Gaussian distributions. Peaks classified as "0" bits are marked in blue while peaks classified as "1" are marked in green.* ***b)*** *Confusion matrix of classification only considering events with two peaks in the filtered set shown in a), numbers represent fractions of the total event count per column (Total event numbers 144 for "00", 52 for "01", 53 for "10", 52 for "11").*

### 3.2.9 S9. List of Oligo Sequences

***Table 3.1: List of all oligo sequences needed to fully assemble a dsDNA M13 scaffold.***

| Oligo ID | Sequence | Length (nt) |
|---|---|---|
| P001 | GACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATG | 45 |
| P002 | GTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA | 40 |
| P003 | ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG | 40 |

| | | |
|---|---|---|
| P004 | CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGC | 40 |
| P005 | GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCG | 40 |
| P006 | TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAG | 40 |
| P007 | GCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCA | 40 |
| P008 | CCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTG | 40 |
| P009 | GCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGC | 40 |
| P010 | CCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTCCGAAAT | 40 |
| P011 | CGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATA | 40 |
| P012 | GGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTAT | 40 |
| P013 | TAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGT | 40 |
| P014 | CTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAA | 40 |
| P015 | TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATC | 40 |
| P016 | GGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGG | 40 |
| P017 | AAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCG | 40 |
| P018 | AAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCA | 40 |
| P019 | CGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCC | 40 |
| P020 | GCTACAGGGCGCGTACTATGGTTGCTTTGACGAGCACGTA | 40 |
| P021 | TAACGTGCTTTCCTCGTTAGAATCAGAGCGGGAGCTAAAC | 40 |
| P022 | AGGAGGCCGATTAAAGGGATTTTAGACAGGAACGGTACGC | 40 |
| P023 | CAGAATCCTGAGAAGTGTTTTTATAATCAGTGAGGCCACC | 40 |
| P024 | GAGTAAAAGAGTCTGTCCATCACGCAAATTAACCGTTGTA | 40 |
| P025 | GCAATACTTCTTTGATTAGTAATAACATCACTTGCCTGAG | 40 |
| P026 | TAGAAGAACTCAAACTATCGGCCTTGCTGGTAATATCCAG | 40 |
| P027 | AACAATATTACCGCCAGCCATTGCAACAGGAAAAACGCTC | 40 |
| P028 | ATGGAAATACCTACATTTTGACGCTCAATCGTCTGAAATG | 40 |
| P029 | GATTATTTACATTGGCAGATTCACCAGTCACACGACCAGT | 40 |
| P030 | AATAAAAGGGACATTCTGGCCAACAGAGATAGAACCCTTC | 40 |
| P031 | TGACCTGAAAGCGTAAGAATACGTGGCACAGACAATATTT | 40 |
| P032 | TTGAATGGCTATTAGTCTTTAATGCGCGAACTGATAGCCC | 40 |
| P033 | TAAAACATCGCCATTAAAAATACCGAACGAACCACCAGCA | 40 |
| P034 | GAAGATAAAACAGAGGTGAGGCGGTCAGTATTAACACCGC | 40 |
| P035 | CTGCAACAGTGCCACGCTGAGAGCCAGCAGCAAATGAAAA | 40 |
| P036 | ATCTAAAGCATCACCTTGCTGAACCTCAAATATCAAACCC | 40 |
| P037 | TCAATCAATATCTGGTCAGTTGGCAAATCAACAGTTGAAA | 40 |
| P038 | GGAATTGAGGAAGGTTATCTAAAATATCTTTAGGAGCACT | 40 |
| P039 | AACAACTAATAGATTAGAGCCGTCAATAGATAATACATTT | 40 |
| P040 | GAGGATTTAGAAGTATTAGACTTTACAAACAATTCGACAA | 40 |
| P041 | CTCGTATTAAATCCTTTGCCCGAACGTTATTAATTTTAAA | 40 |
| P042 | AGTTTGAGTAACATTATCATTTTGCGGAACAAAGAAACCA | 40 |
| P043 | CCAGAAGGAGCGGAATTATCATCATATTCCTGATTATCAG | 40 |

| | | |
|---|---|---|
| P044 | ATGATGGCAATTCATCAATATAATCCTGATTGTTTGGATT | 40 |
| P045 | ATACTTCTGAATAATGGAAGGGTTAGAACCTACCATATCA | 40 |
| P046 | AAATTATTTGCACGTAAAACAGAAATAAAGAAATTGCGTA | 40 |
| P047 | GATTTTCAGGTTTAACGTCAGATGAATATACAGTAACAGT | 40 |
| P048 | ACCTTTTACATCGGGAGAAACAATAACGGATTCGCCTGAT | 40 |
| P049 | TGCTTTGAATACCAAGTTACAAAATCGCGCAGAGGCGAAT | 40 |
| P050 | TATTCATTTCAATTACCTGAGCAAAAGAAGATGATGAAAC | 40 |
| P051 | AAACATCAAGAAAACAAAATTAATTACATTTAACAATTTC | 40 |
| P052 | ATTTGAATTACCTTTTTTAATGGAAACAGTACATAAATCA | 40 |
| P053 | ATATATGTGAGTGAATAACCTTGCTTCTGTAAATCGTCGC | 40 |
| P054 | TATTAATTAATTTTCCCTTAGAATCCTTGAAAACATAGCG | 40 |
| P055 | ATAGCTTAGATTAAGACGCTGAGAAGAGTCAATAGTGAAT | 40 |
| P056 | TTATCAAAATCATAGGTCTGAGAGACTACCTTTTTAACCT | 40 |
| P057 | CCGGCTTAGGTTGGGTTATATAACTATATGTAAATGCTGA | 40 |
| P058 | TGCAAATCCAATCGCAAGACAAAGAACGCGAGAAAACTTT | 40 |
| P059 | TTCAAATATATTTTAGTTAATTTCATCTTCTGACCTAAAT | 40 |
| P060 | TTAATGGTTTGAAATACCGACCGTGTGATAAATAAGGCGT | 40 |
| P061 | TAAATAAGAATAAACACCGGAATCATAATTACTAGAAAAA | 40 |
| P062 | GCCTGTTTAGTATCATATGCGTTATACAAATTCTTACCAG | 40 |
| P063 | TATAAAGCCAACGCTCAACAGTAGGGCTTAATTGAGAATC | 40 |
| P064 | GCCATATTTAACAACGCCAACATGTAATTTAGGCAGAGGC | 40 |
| P065 | ATTTTCGAGCCAGTAATAAGAGAATATAAAGTACCGACAA | 40 |
| P066 | AAGGTAAAGTAATTCTGTCCAGACGACGACAATAAACAAC | 40 |
| P067 | ATGTTCAGCTAATGCAGAACGCGCCTGTTTATCAACAATA | 40 |
| P068 | GATAAGTCCTGAACAAGAAAAATAATATCCCATCCTAATT | 40 |
| P069 | TACGAGCATGTAGAAACCAATCAATAATCGGCTGTCTTTC | 40 |
| P070 | CTTATCATTCCAAGAACGGGTATTAAACCAAGTACCGCAC | 40 |
| P071 | TCATCGAGAACAAGCAAGCCGTTTTTATTTTCATCGTAGG | 40 |
| P072 | AATCATTACCGCGCCCAATAGCAAGCAAATCAGATATAGA | 40 |
| P073 | AGGCTTATCCGGTATTCTAAGAACGCGAGGCGTTTTAGCG | 40 |
| P074 | AACCTCCCGACTTGCGGGAGGTTTTGAAGCCTTAAATCAA | 40 |
| P075 | GATTAGTTGCTATTTTGCACCCAGCTACAATTTTATCCTG | 40 |
| P076 | AATCTTACCAACGCTAACGAGCGTCTTTCCAGAGCCTAAT | 40 |
| P077 | TTGCCAGTTACAAAATAAACAGCCATATTATTTATCCCAA | 40 |
| P078 | TCCAAATAAGAAACGATTTTTTGTTTAACGTCAAAAATGA | 40 |
| P079 | AAATAGCAGCCTTTACAGAGAGAATAACATAAAAACAGGG | 40 |
| P080 | AAGCGCATTAGACGGGAGAATTAACTGAACACCCTGAACA | 40 |
| P081 | AAGTCAGAGGGTAATTGAGCGCTAATATCAGAGAGATAAC | 40 |
| P082 | CCACAAGAATTGAGTTAAGCCCAATAATAAGAGCAAGAAA | 40 |
| P083 | CAATGAAATAGCAATAGCTATCTTACCGAAGCCCTTTTTA | 40 |

| | | |
|---|---|---|
| P084 | AGAAAAGTAAGCAGATAGCCGAACAAAGTTACCAGAAGGA | 40 |
| P085 | AACCGAGGAAACGCAATAATAACGGAATACCCAAAAGAAC | 40 |
| P086 | TGGCATGATTAAGACTCCTTATTACGCAGTATGTTAGCAA | 40 |
| P087 | ACGTAGAAAATACATACATAAAGGTGGCAACATATAAAAG | 40 |
| P088 | AAACGCAAAGACACCACGGAATAAGTTTATTTTGTCACAA | 40 |
| P089 | TCAATAGAAAATTCATATGGTTTACCAGCGCCAAAGACAA | 40 |
| P090 | AAGGGCGACATTCAACCGATTGAGGGAGGGAAGGTAAATA | 40 |
| P091 | TTGACGGAAATTATTCATTAAAGGTGAATTATCACCGTCA | 40 |
| P092 | CCGACTTGAGCCATTTGGGAATTAGAGCCAGCAAAATCAC | 40 |
| P093 | CAGTAGCACCATTACCATTAGCAAGGCCGGAAACGTCACC | 40 |
| P094 | AATGAAACCATCGATAGCAGCACCGTAATCAGTAGCGACA | 40 |
| P095 | GAATCAAGTTTGCCTTTAGCGTCAGACTGTAGCGCGTTTT | 40 |
| P096 | CATCGGCATTTTCGGTCATAGCCCCCTTATTAGCGTTTGC | 40 |
| P097 | CATCTTTTCATAATCAAAATCACCGGAACCAGAGCCACCA | 40 |
| P098 | CCGGAACCGCCTCCCTCAGAGCCGCCACCCTCAGAACCGC | 40 |
| P099 | CACCCTCAGAGCCACCACCCTCAGAGCCGCCACCAGAACC | 40 |
| P100 | ACCACCAGAGCCGCCGCCAGCATTGACAGGAGGTTGAGGC | 40 |
| P101 | AGGTCAGACGATTGGCCTTGATATTCACAAACAAATAAAT | 40 |
| P102 | CCTCATTAAAGCCAGAATGGAAAGCGCAGTCTCTGAATTT | 40 |
| P103 | ACCGTTCCAGTAAGCGTCATACATGGCTTTTGATGATACA | 40 |
| P104 | GGAGTGTACTGGTAATAAGTTTTAACGGGGTCAGTGCCTT | 40 |
| P105 | GAGTAACAGTGCCCGTATAAACAGTTAATGCCCCCTGCCT | 40 |
| P106 | ATTTCGGAACCTATTATTCTGAAACATGAAAGTATTAAGA | 40 |
| P107 | GGCTGAGACTCCTCAAGAGAAGGATTAGGATTAGCGGGGT | 40 |
| P108 | TTTGCTCAGTACCAGGCGGATAAGTGCCGTCGAGAGGGTT | 40 |
| P109 | GATATAAGTATAGCCCGGAATAGGTGTATCACCGTACTCA | 40 |
| P110 | GGAGGTTTAGTACCGCCACCCTCAGAACCGCCACCCTCAG | 40 |
| P111 | AACCGCCACCCTCAGAGCCACCACCCTCATTTTCAGGGAT | 40 |
| P112 | AGCAAGCCCAATAGGAACCCATGTACCGTAACACTGAGTT | 40 |
| P113 | TCGTCACCAGTACAAACTACAACGCCTGTAGCATTCCACA | 40 |
| P114 | GACAGCCCTCATAGTTAGCGTAACGATCTAAAGTTTTGTC | 40 |
| P115 | GTCTTTCCAGACGTTAGTAAATGAATTTTCTGTATGGGAT | 40 |
| P116 | TTTGCTAAACAACTTTCAACAGTTTCAGCGGAGTGAGAAT | 40 |
| P117 | AGAAAGGAACAACTAAAGGAATTGCGAATAATAATTTTTT | 40 |
| P118 | CACGTTGAAAATCTCCAAAAAAAAGGCTCCAAAAGGAGCC | 40 |
| P119 | TTTAATTGTATCGGTTTATCAGCTTGCTTTCGAGGTGAAT | 40 |
| P120 | TTCTTAAACAGCTTGATACCGATAGTTGCGCCGACAATGA | 40 |
| P121 | CAACAACCATCGCCCACGCATAACCGATATATTCGGTCGC | 40 |
| P122 | TGAGGCTTGCAGGGAGTTAAAGGCCGCTTTTGCGGGATCG | 40 |
| P123 | TCACCCTCAGCAGCGAAAGACAGCATCGGAACGAGGGTAG | 40 |

| | | |
|---|---|---|
| P124 | CAACGGCTACAGAGGCTTTGAGGACTAAAGACTTTTTCAT | 40 |
| P125 | GAGGAAGTTTCCATTAAACGGGTAAAATACGTAATGCCAC | 40 |
| P126 | TACGAAGGCACCAACCTAAAACGAAAGAGGCAAAAGAATA | 40 |
| P127 | CACTAAAACACTCATCTTTGACCCCCAGCGATTATACCAA | 40 |
| P128 | GCGCGAAACAAAGTACAACGGAGATTTGTATCATCGCCTG | 40 |
| P129 | ATAAATTGTGTCGAAATCCGCGACCTGCTCCATGTTACTT | 40 |
| P130 | AGCCGGAACGAGGCGCAGACGGTCAATCATAAGGGAACCG | 40 |
| P131 | AACTGACCAACTTTGAAAGAGGACAGATGAACGGTGTACA | 40 |
| P132 | GACCAGGCGCATAGGCTGGCTGACCTTCATCAAGAGTAAT | 40 |
| P133 | CTTGACAAGAACCGGATATTCATTACCCAAATCAACGTAA | 40 |
| P134 | CAAAGCTGCTCATTCAGTGAATAAGGCTTGCCCTGACGAG | 40 |
| P135 | AAACACCAGAACGAGTAGTAAATTGGGCTTGAGATGGTTT | 40 |
| P136 | AATTTCAACTTTAATCATTGTGAATTACCTTATGCGATTT | 40 |
| P137 | TAAGAACTGGCTCATTATACCAGTCAGGACGTTGGGAAGA | 40 |
| P138 | AAAATCTACGTTAATAAAACGAACTAACGGAACAACATTA | 40 |
| P139 | TTACAGGTAGAAAGATTCATCAGTTGAGATTTAGGAATAC | 40 |
| P140 | CACATTCAACTAATGCAGATACATAACGCCAAAAGGAATT | 40 |
| P141 | ACGAGGCATAGTAAGAGCAACACTATCATAACCCTCGTTT | 40 |
| P142 | ACCAGACGACGATAAAAACCAAAATAGCGAGAGGCTTTTG | 40 |
| P143 | CAAAAGAAGTTTTGCCAGAGGGGGTAATAGTAAAATGTTT | 40 |
| P144 | AGACTGGATAGCGTCCAATACTGCGGAATCGTCATAAATA | 40 |
| P145 | TTCATTGAATCCCCCTCAAATGCTTTAAACAGTTCAGAAA | 40 |
| P146 | ACGAGAATGACCATAAATCAAAAATCAGGTCTTTACCCTG | 40 |
| P147 | ACTATTATAGTCAGAAGCAAAGCGGATTGCATCAAAAAGA | 40 |
| P148 | TTAAGAGGAAGCCCGAAAGACTTCAAATATCGCGTTTTAA | 40 |
| P149 | TTCGAGCTTCAAAGCGAACCAGACCGGAAGCAAACTCCAA | 40 |
| P150 | CAGGTCAGGATTAGAGAGTACCTTTAATTGCTCCTTTTGA | 40 |
| P151 | TAAGAGGTCATTTTTGCGGATGGCTTAGAGCTTAATTGCT | 40 |
| P152 | GAATATAATGCTGTAGCTCAACATGTTTTAAATATGCAAC | 40 |
| P153 | TAAAGTACGGTGTCTGGAAGTTTCATTCCATATAACAGTT | 40 |
| P154 | GATTCCCAATTCTGCGAACGAGTAGATTTAGTTTGACCAT | 40 |
| P155 | TAGATACATTTCGCAAATGGTCAATAACCTGTTTAGCTAT | 40 |
| P156 | ATTTTCATTTGGGGCGCGAGCTGAAAAGGTGGCATCAATT | 40 |
| P157 | CTACTAATAGTAGTAGCATTAACATCCAATAAATCATACA | 40 |
| P158 | GGCAAGGCAAAGAATTAGCAAAATTAAGCAATAAAGCCTC | 40 |
| P159 | AGAGCATAAAGCTAAATCGGTTGTACCAAAAACATTATGA | 40 |
| P160 | CCCTGTAATACTTTTGCGGGAGAAGCCTTTATTTCAACGC | 40 |
| P161 | AAGGATAAAAATTTTTAGAACCCTCATATATTTTAAATGC | 40 |
| P162 | AATGCCTGAGTAATGTGTAGGTAAAGATTCAAAAGGGTGA | 40 |
| P163 | GAAAGGCCGGAGACAGTCAAATCACCATCAATATGATATT | 40 |

| | | |
|---|---|---|
| P164 | CAACCGTTCTAGCTGATAAATTAATGCCGGAGAGGGTAGC | 40 |
| P165 | TATTTTTGAGAGATCTACAAAGGCTATCAGGTCATTGCCT | 40 |
| P166 | GAGAGTCTGGAGCAAACAAGAGAATCGATGAACGGTAATC | 40 |
| P167 | GTAAAACTAGCATGTCAATCATATGTACCCCGGTTGATAA | 40 |
| P168 | TCAGAAAAGCCCCAAAAACAGGAAGATTGTATAAGCAAAT | 40 |
| P169 | ATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCAT | 40 |
| P170 | TAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGA | 40 |
| P171 | ACGCCATCAAAAATAATTCGCGTCTGGCCTTCCTGTAGCC | 40 |
| P172 | AGCTTTCATCAACATTAAATGTGAGCGAGTAACAACCCGT | 40 |
| P173 | CGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATG | 40 |
| P174 | GGATAGGTCACGTTGGTGTAGATGGGCGCATCGTAACCGT | 40 |
| P175 | GCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGCCTC | 40 |
| P176 | AGGAAGATCGCACTCCAGCCAGCTTTCCGGCACCGCTTCT | 40 |
| P177 | GGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGC | 40 |
| P178 | TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTC | 40 |
| P179 | GCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGG | 40 |
| P180 | CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGAC | 40 |
| P181 | GTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTC | 44 |

*Table 3.2: List of extra oligos needed to assemble 2-bit triangle assembly (design 1).*

| **Oligo ID** | **Sequence** | **Length (nt)** |
|---|---|---|
| SP050_L | TATTCATTTCAATTACCTGAGTGTCTTGAAGAAGGACCAGGGAAC | 45 |
| SP050_R | GGAAACGAGGTGCATTTTGATACCGGCAAAAGAAGATGATGAAAC | 45 |
| SP052_L | ATTTGAATTACCTTTTTTAAGTGTCTTGAAGAAGGACCAGGGAAC | 45 |
| SP052_R | GGAAACGAGGTGCATTTTGATACCGTGGAAACAGTACATAAATCA | 45 |
| SP053_L | ATATATGTGAGTGAATAACCGTGTCTTGAAGAAGGACCAGGGAAC | 45 |
| SP053_R | GGAAACGAGGTGCATTTTGATACCGTTGCTTCTGTAAATCGTCGC | 45 |
| SP054_L | TATTAATTAATTTTCCCTTAGTGTCTTGAAGAAGGACCAGGGAAC | 45 |
| SP054_R | GGAAACGAGGTGCATTTTGATACCGGAATCCTTGAAAACATAGCG | 45 |
| SP055_L | ATAGCTTAGATTAAGACGCTGTGTCTTGAAGAAGGACCAGGGAAC | 45 |
| SP055_R | GGAAACGAGGTGCATTTTGATACCGGAGAAGAGTCAATAGTGAAT | 45 |
| A1_86 | GTTCCCTGGTCCTTCTTCAAGACACAATTAGTGACCACTTCAGAATCCT TTCAACCGCGTACGGTATCAAAATGCACCTCGTTTCC | 86 |

*Table 3.3: List of oligos used for assembly of 4-arm and 12-arm stars for design 2, including the oligos binding them to the carrier.*

| Oligo ID | Sequence | Length (nt) |
|---|---|---|
| J12_01 | CCTGCAAATTTTACTGATGCTGATCTGCTATGGCACTGCCGTCTGCGATT | 50 |
| J12_02 | AATCGCAGACGGCAGTGCCATAGCATGTTATTGGGTCCATACATGAAGAC | 50 |
| J12_03 | GTCTTCATGTATGGACCCAATAACAGAAAAACCGGGGTAAGGTAGCCGCC | 50 |
| J12_04 | GGCGGCTACCTTACCCCGGTTTTTCCGGTAAAGTTTAAGCCTGTAATGAG | 50 |
| J12_05 | CTCATTACAGGCTTAAACTTTACCGTCGAAGTTTTCGTGATTTTCAGCTC | 50 |
| J12_06 | GAGCTGAAAATCACGAAAACTTCGATGTGTTAGACTAACCTGTGTCCGTA | 50 |
| J12_07 | TACGGACACAGGTTAGTCTAACACAATGCAGCAAAGGGTCTACAGTCGGG | 50 |
| J12_08 | CCCGACTGTAGACCCTTTGCTGCATTCTGCAGGACTTATGTCATTTGCAG | 50 |
| J12_09 | CTGCAAATGACATAAGTCCTGCAGACGAGAGCCGCCTCAGTGCAATACAT | 50 |
| J12_10 | ATGTATTGCACTGAGGCGGCTCTCGAGCGTGTCTTTTCTATCTTTACGGC | 50 |
| J12_11 | GCCGTAAAGATAGAAAAGACACGCTAGAAAACGATCCGCAGCGGGATTCG | 50 |
| J12_12A | GTTTCCGAATGACACCGAATCCCGCTGCGGATCGTTTTCTGATCAGCATC AGTAAAATTTGCAGG | 65 |
| J12_04A | AAGATCCCAGCGCTCGGCGGCTACCTTACCCCGGTTTTTCCGGTAAAGTT TAAGCCTGTAATGAG | 65 |
| SP050 | TATTCATTTCAATTACCTGAGCAAAAGAAGATGATGAAACCACACTTGGA CCTTC | 55 |
| SP054 | TATTAATTAATTTTCCCTTAGAATCCTTGAAAACATAGCGACATCCCACA GAGGG | 55 |
| J12P050 | GAAGGTCCAAGTGTGGTGTCATTCGGAAAC | 30 |
| J4P050 | GAAGGTCCAAGTGTGGAGCGCTGGGATCTT | 30 |
| J12P054 | CCCTCTGTGGGATGTGTGTCATTCGGAAAC | 30 |
| J4_P054 | CCCTCTGTGGGATGTGAGCGCTGGGATCTT | 30 |

*Table 3.4: List of oligos used for assembly of short and long 3-helix bundle attachments, including the oligos binding them to the carrier.*

| Oligo ID | Sequence | Length (nt) |
|---|---|---|
| C0_1 | TTTTTCATCTAGCCGGTGTTTTTT | 24 |
| C0_2 | GCAGCGCTATGTGCGACACTGTTTTT | 26 |
| C0_3 | TTATGCTTCGAGGTCAGTCGT | 21 |
| C0_4 | TTTTTTTCTAACACTCTCAAACCGCA | 26 |
| C0_5 | TTTTTTGACTGTCGTCCAGTTTTT | 24 |
| U0_1_1 | TGAAGCAAAAGTGAAACAACGCAACAGCTTCGGCG | 35 |
| U0_1_2 | AAAGCGGGTTCTTCGCAGAGATATATGCTCGTCGT | 35 |
| U0_3_1 | CCTCACTGTGCGAGTGACTGGCATTCTGGACCGCT | 35 |
| U0_3_2 | AACGACCAAGGCTGTGGCAGAGAAAATCACGGATA | 35 |

| | | |
|---|---|---|
| S1_1 | ACACCGGCGCACCAGTTTTCACGCCGAAAGCATAAACGA CGAGCACATATGAGAGTGTTAGAA | 63 |
| S21_1 | CTGGACGTGCGGTTTCTCTGCGAAGAACACCT | 32 |
| S22_1 | GATTATTTACATTGGCAGATTCACCAGTCACACGACCAGT CGAGCTGTTGCGTTGTTGCGCTGCCTAGATG | 71 |
| S3_1 | CAGTGTCGCACATATCACTTTTGCTTCAACGACTGCCGCT TTCGCTGTATCCCTAGACAGTCA | 63 |
| S4 | ACAGTAAGCGCTTTTTGTAGGCAGGGGCTCCCCCTGTGAC TAACTGCGCCAAAACATCTTCGG | 63 |
| S5_1 | ATCCCCTTGTCCAATCTAACTCACCGAATTCTTACATTTT AGACCCTAATATCACATCATTAG | 63 |
| S6 | AGATTAATTGCCACTGCCAAAATTCTGTCCACAAGCGTTT TAGTTCGCCCCAGTAAAGTTGTC | 63 |
| S1_3 | ACACCGGCCTCTAAGTAAGAGAGCGGTCAGCATAATATCC GTGATTTTCTGAGAGTGTTAGAA | 63 |
| S21_3 | CTGGACGTGCGGTTTCTGCCACAGCCTTACCT | 32 |
| S22_3 | TCAATCAATATCTGGTCAGTTGGCAAATCAACAGTTGAAA CGACAGAATGCCAGTCAGCGCTGCCTAGATG | 71 |
| S3_3 | CAGTGTCGCACATACTCGCACAGTGAGGACGACTGGGTCG TTCGGCTCTACCGACAACAGTCA | 63 |
| S5_3 | ATCCCCTTGTCCAACTTTCAGAATCATGTTCTTACCGGCA AAGTGACGCTATCACATCATTAG | 63 |
| 0E1 | TTTTTAGCGTGTCTAGCGT | 19 |
| 0E2 | GCAGTCAACTGCGATTTTT | 19 |
| 0CA | TTTTTCTTGATGGAGCAGACCTATCGTCACGTATTTTT | 38 |
| 0CB | TTTTTCCGAGATTGTCGAACGTATCTGCCTTACTTTTT | 38 |
| 0C1 | GTTACTCGTTGGACGTTGGCACTCGACA | 28 |
| 0C2 | TTCACGTTAGACTGCACGATACACCTGTGAACCTG | 35 |
| 0C3 | TGTCATCACCGAGTCAGCGACAACTATG | 28 |
| 0C4 | GATAACGGTTGCAGCAGGAGATGTCAGC | 28 |
| 0C5 | AGCAGAAGCGTTGAGTGCTGTGCCATAAGCTCACC | 35 |
| 0C6 | CTGACAACGTATCGCTGTCAGTCGGAGA | 28 |
| 0C7 | GTCGAATACTGTACCAACGGAATTCCTC | 28 |
| 0C8 | CTACACTAGGACGTCTCTAGTGTCCTAGCTGAACT | 35 |
| 0C9 | AGGTAGTCTGATAGCATGCAGACTTAGC | 28 |
| 0L1 | GATGCGTGAACTATGACATCTGATGTGTGCTACTTGTCACCA | 42 |
| 0L2 | TCGAGTACGCTCTTGGTTACTGTGACCTGTGCTCACGAGGAC | 42 |
| 0B1 | TCGCAGTTGACTGCTAGTACCTGAGCACCTATCAGACTACCT AGTTCAGGTACAGTATTCGACTCTCCGATCAACGCTTCTGCT GCTGACAGATGACACAGGTTCACAGGTGGAGTAACGATGTAG AAGTAGCTCTGCTCCATCAAG | 147 |
| 0B2 | GCACCTGACAGCGATACGTCCGTTGCTAGGACACTAGAGCTG | 74 |

| | | |
|---|---|---|
| | CATGAGGTCACAGTAACCTTCGACAATCTCGG | |
| 0B3 | GTAAGGCAGATACGAAGAGCGAGGCTATGCTAAGTACGTCCT AGTGTAGGAGGAATTTGTCAGGGTGAGCTTATGGCCGTTATC CATAGTTGTCGCTGACGTGAATGTCGAGTGCCAACATAGTTC TCGCACGACGCTAGACACGCT | 147 |
| 0S1 | TACGTGACGATAGGACACATCAGATGTCGTCCAACTATCGTG CAGTCTAACTCGGTTCTCCTGCTGCAACACAGATTATTTACA TTGGCAGATTCACCAGTCACACGACCAGT | 113 |
| 0S1_2 | TACGTGACGATAGGACACATCAGATGTCGTCCAACTATCGTG CAGTCTAACTCGGTTCTCCTGCTGCAACACATCAATCAATAT CTGGTCAGTTGGCAAATCAACAGTTGAAA | 113 |
| 1E1 | CTGTACCGTTGGCGTTTTT | 19 |
| 1E2 | TTTTTCCACGTTGAGTTGG | 19 |
| 1CA | TTTTTGGCTGTCCGTGCAACCGATCAATCCGCATTTTT | 38 |
| 1CB | TTTTTCTTGGCAATGTCCACCATTGGATCGTGGTTTTT | 38 |
| 1C1 | CACGTCTAGGTAGACTCGAGATCCAATC | 28 |
| 1C2 | TCAAGGAACGCATGTGACGTTATCTCAGCTAACGC | 35 |
| 1C3 | CTGAGTCATGTGGCAATCGGTGTTAACG | 28 |
| 1C4 | ACTCTTCGACTCGTTGGACTTGTCGTGA | 28 |
| 1C5 | CTCCTATTCGACTCGTTCCGAGCGTTCCATCGTCA | 35 |
| 1C6 | GTCAGATCAATCTGCCAACGCAGGCTTC | 28 |
| 1C7 | CCAAGATTCGCTTCAACGCATGTGTTCT | 28 |
| 1C8 | ATATGGCTAACTCATACTCGCAGCTCTCTCCACCT | 35 |
| 1C9 | TACAGTGCGACTTAGGTTCGTGTGAACG | 28 |
| 1L1 | CGTGCGACTGGCATGTGATACATACACTGGTTGGACTACATC | 42 |
| 1L2 | ATAGCCTGGCTCATGCAATACGGTCGTTGCGTTATGGTACTA | 42 |
| 1B1 | TGCGGATTGATCGGATGCCAGACGCATCGATTGGACATGCGT TCCTTGACGTTAACGAAGAGTTGACGATGGAACGCATCTGAC AGAACACATGCGTTGCCATATCGTTCACACGAACCATGAGCC TACTCGACCAACTCAACGTGG | 147 |
| 1B2 | CCACGATCCAATGGAACGACCGTATTGCTAAGTCGGCGAGTA TGAGTTAGAAGCGAGCGTTGGCAGATTGTCGGAACAAGTCCA ACGAGTCACCGATTCTGAGATAACGTCATCTCGAGAGTGTAT GTATCACTTGCACGGACAGCC | 147 |
| 1B3 | CGCCAACGGTACAGTGGTGACTCCAACCTCTACCTAGACGTG GCGTTAGGCCACATGACTCAGTCACGACGAGTCGAATAGGAG GAAGCCTATCTTGGAGGTGGAGAGAGCTCACTGTAGTCCTCG ATAACGCTGGACATTGCCAAG | 147 |

### 3.2.10 S10. List of Nanopores and Experimental Settings

***Table 3.5: List of all pore parameters used in this study.*** *Different pore length indicates pores of different batches.*

| Pore ID | Nominal Membrane Thickness (nm) | Diameter (nm) | Samples (bit spacing) | Voltage (mV) |
|---|---|---|---|---|
| 1 | 12 ± 2 | 7.9 | Triangle (160 bp) | 300 |
| 2 | 12 ± 2 | 7.9 | Triangle (80 bp, 120 bp, 160 bp, 200 bp) | 300 |
| 3 | 12 ± 2 | 8.0 | Triangle (200 bp) | 300 |
| 4 | 12 ± 2 | 13.7 | Stars (160 bp, “01”, “10”) | 150 |
| 5 | 17 ± 2 | 12.6 | Bundles (320 bp, “00”) | 200 |
| 6 | 17 ± 2 | 13.6 | Bundles (320 bp, “01”) | 200 |
| 7 | 17 ± 2 | 13.2 | Bundles (320 bp, “10”, “11”) | 200 |
| 8 | 17 ± 2 | 13.0 | Bundles (320 bp, “01”) | 200 |

# 4 DNA Barcoding – Challenges beyond the Nanopore

## 4.1 Nanopore Barcoding

The challenges with molecular barcoding from the perspective of the nanopore sensors themselves are already highlighted in **Chapter 3**. However, there is another class of challenges that must be overcome before it will be possible to generate proper barcoded signals in the nanopore current traces, namely, those arising from the molecular design choices made with respect to the DNA assembly itself. Many iterations need to occur to arrive at a feasible design. This includes experiments to investigate assembly yields and purification of the samples from their unreacted starting products. In this chapter, I present the additional experimental work and analyses performed, and highlight lessons learned not included in the preceding chapters. This includes extra analysis for nanopore translocations of failed intermediate DNA nanostructure designs, as well as investigations using gel electrophoresis to examine the yield of assembly for these structures, and the effect of different purification strategies. These intermediate process steps are often glossed over in the literature but represent an important contribution to progress in the field.

Some of the challenges discussed in this chapter arose from the interactions between barcode molecules and the nanopore, while others arose from the principles that were employed in designing the DNA structures. From the lessons learned in the process, two guiding principles for barcode design emerge to inform the semi-successful designs presented in the previous chapters.

The first such principle is that each barcoded event should be interpretable without prior knowledge of the encoded bit sequence. In our early encoding schemes, we attempted to use the presence and absence of a particular target to indicate 1 and 0 respectively, but it quickly became clear that it is difficult to be confident in our inference with respect to the absence of a target. Practically, this means that each bit must be encoded such that each event has the same number of peaks regardless of encoded sequence, i.e. we need an explicit "0" bit (that is distinguishable from both the "1" bit and the "null" state that is the absence of a bit).

This requirement is a consequence of the fact that nanopore translocation velocities vary, both from event to event and within a single event,[118] which can make mapping positions of peaks in the current trace to molecular positions difficult. This is especially problematic when "0" bits are at the ends of the sequence or when multiple "0" bits appear in a row. Having a dedicated "0" bit solves these issues at the cost of increased assembly complexity. This extra complexity arises not only due to the need for two different attachments, but also from the fact that these attachments must produce sufficiently distinct signals when translocating the nanopore. One way to achieve this is to make the "0" and "1" bit attachments different sizes such that the resulting current peaks have different blockage depths. But, as discussed in **Chapter 3**, the spread in blockage depth of each bit, due to fluctuations in its molecular orientation on the carrier, can render this task more challenging than one may initially anticipate.

The second guiding principle is that it should be possible to rewrite and update the sequence of a barcode after it has been written. This would open the door to manipulating the stored data after the fact, providing both reading and writing capability, for DNA-based computing

applications. On the molecular level, this means that the binding of the attachments should be reversible and controllable. A key advantage of this principle is that each barcode can be assembled by combining standard carrier strands and specific attachments. This eliminates the need to rebuild the entire carrier structure for every new barcode, significantly improving the technique's scalability.

### 4.1.1 Challenges with DNA Triangles

The very first design tested for this project is the triangle design introduced in **Chapter 3**. In addition to the spacing between bits, this design was also intended to generate peaks of different height based on the presence or absence of a single additional 86 nucleotide long oligo. The "0" bit version of this design consisted of two 25 nt ssDNA overhangs along the backbone of the carrier. For the "1" bit, a single 86 nt oligo binds to both overhangs and leaves an exposed 36 nt sequence (initially designed for easy access for a potential strand displacement reaction to remove the extra oligo after writing data on it). An alternative version of the "1" bit includes an extra 36 nt oligo to cover the exposed sequence and add stability to the attachment in the hope of increasing distinguishability. The three structures together with sample event traces are shown in **Figure 4.1a**.

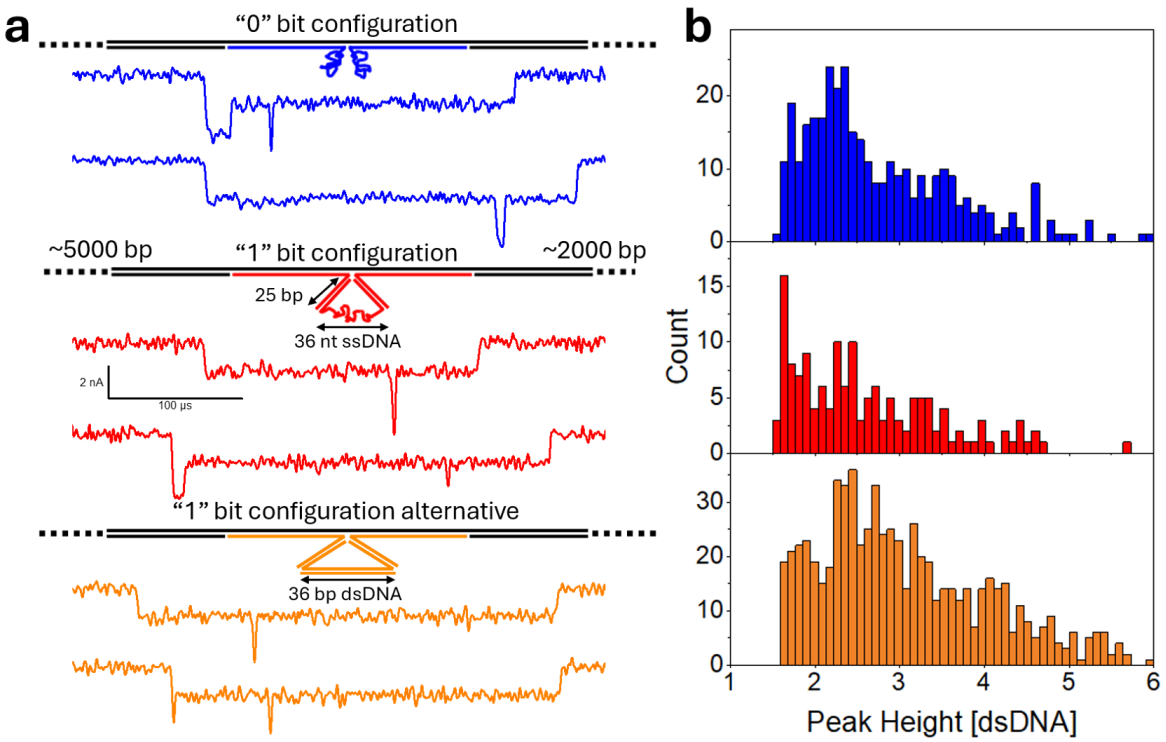



***Figure 4.1: Overview of triangle structures for peak height measurements. a)*** *Schematic structure of the different triangle attachments studied with sample events. From top to bottom: the "0" bit version (blue) of the triangle attachment consisting of only two short (25 nt) ssDNA overhangs on the main carrier; the "1" bit version (red) of the triangle attachment including the added 86 nt long oligo; an alternate version of the "1" bit (orange) with an extra addition of a 36 nt oligo to make the triangle completely double-stranded and add structural integrity.* ***b)*** *Peak height histograms for all three versions measured in an 8.0 nm pore in 3.6 M LiCl at pH 8 under 300 mV bias voltage. The peak height is normalized by the unfolded level of each event.*

From the event traces for all these triangular structures, the peaks were found through a simple thresholding algorithm. Peaks are identified as any number of data points that deviate from the baseline by more than 1.5 times the deviation caused by a single dsDNA strand. The peak height is taken to be the single data point that is furthest away from the baseline within those areas identified as peaks. **Figure 4.1b** shows the aggregated peak heights normalized by the single file level blockage for each event. From this plot it becomes clear that the distributions of peak heights

for the different states of this design are very similar, making distinguishing the different states all but impossible. Additionally, all three distributions span a wide range of blockages and even the small attachment of two ssDNA sequences is capable of generating peaks much deeper than expected based on purely the volume of the sidechains, indicating that other factors impact the observed blockage depths.[202] With those two observations, a new design using previously studied DNA origami stars[7] was investigated.

### 4.1.2 Increasing Attachment Density

While the main findings and challenges with the star barcoding scheme are highlighted in **Chapter 3,** this design was also used to test the attachment of up to 40 bits to a single carrier. Recall that this design (**Figure 3.3**) has a 4-arm star as a "0" bit and a 12-arm star as a "1" bit positioned every 160 bp along the dsDNA carrier. This was one of the first experiments attempted before more detailed studies to understand the origins of the challenges encountered were performed. Over the course of 4 early experiments trying to measure a 40-bit version of the star barcodes, two main problems occurred. The first was that a majority of the events recorded showed significant secondary structure in the current traces: deep, long blockages not observed with the 2-bit structures (see **Figure 4.2a** and **b** for example traces). The second is that even for events that did not show this anomalous behavior, the detectable number of peaks did not ever reach 40. The largest number of detected peaks in a single event was 30 (see **Figure 4.2c**). This indicated problems not just on the nanopore detection side, but assembly issues with these structures as well.

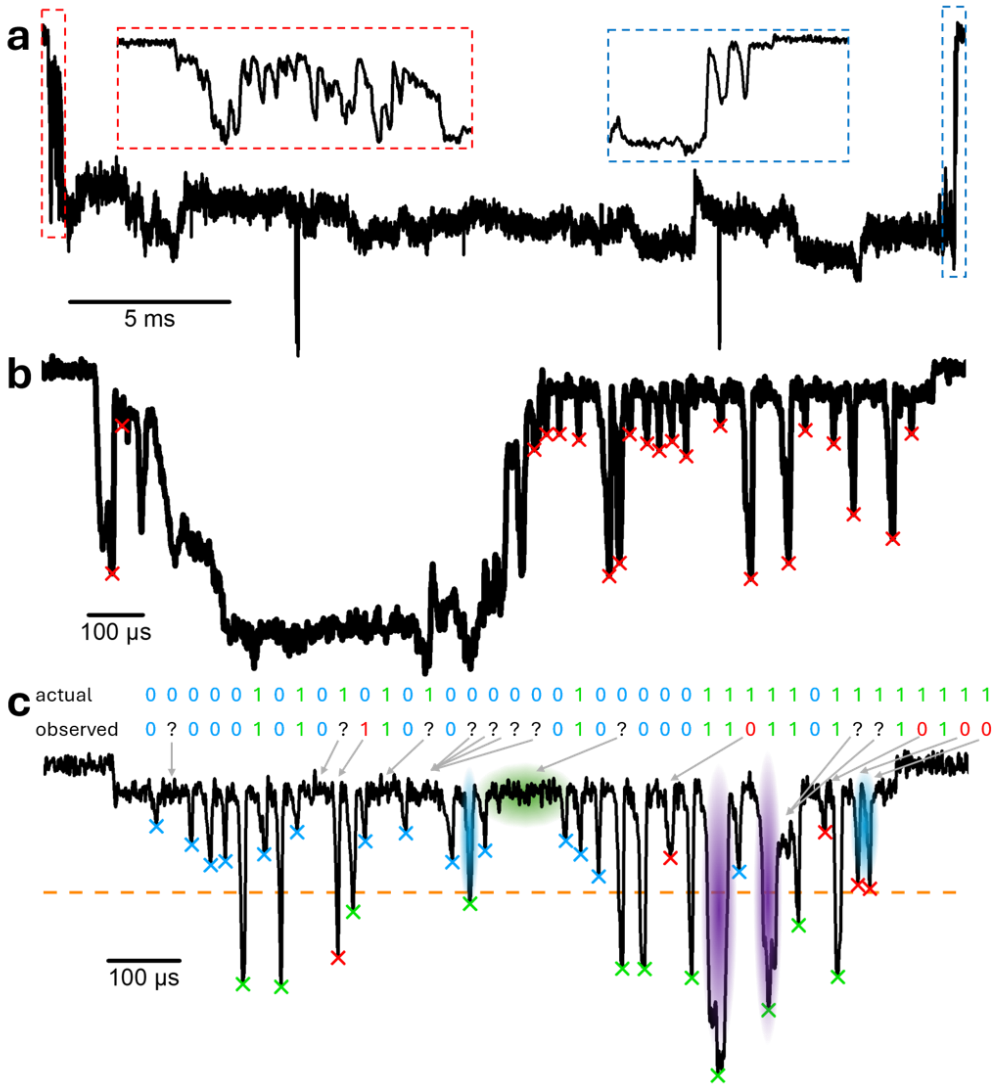



***Figure 4.2: 40-bit assembly with star attachment. a), b)*** *Example trace of events with extended blockage state in the middle, the insets in a) show a zoom-in of the beginning (red box) and end (blue box) of the event. The red crosses in b) mark peaks detected by our peak finding algorithm.* ***c)*** *Trace of an event with 30 detected peaks (marked by blue and green crosses, red crosses mark peaks that are wrongly categorized, the orange line marks the threshold for classification). Colored areas highlight common problems that are explained in the main text. The table on top represents the best attempt at mapping the observed sequence to the actual sequence of bits with question marks indicating spots of presumed missing peaks. These events were recorded in a 12.6 nm pore in 3.6 M LiCl at pH 8 with 100 mV bias voltage applied and Bessel low-pass filtered to 500 kHz.*

When zooming in on the beginning (red inset **Figure 4.2a**) or the end (blue inset) of the event, we see the expected event shape, with peaks deviating from the expected carrier level.

However, the middle section blocks more current than expected, and seems to get stuck in the pore, in some cases leading to orders of magnitude increase in overall event duration.

Events that behave as expected are exemplified by the trace shown in **Figure 4.2c** which highlights some of the issues seen with these high-density star structured barcodes. Even though the stars are supposed to attach in 120 bp intervals, observed peak distances are irregular and some positions are missing peaks entirely (green highlight). Some of the peaks observed are significantly longer in duration (and sometimes deeper in blockages as well) than expected (purple highlights) potentially indicating adjacent peaks merging into a single larger peak in the current traces. As described in **Chapter 3**, the main problem with this design is the difficulty distinguishing the two different attachments. Peaks with ambiguous height (blue highlights) make confident mapping of the observed peaks to the designed bit sequence very challenging if not impossible on an event-by-event basis.

These observations indicate an additional design consideration not discussed in the analysis above: bit spacing. Beyond the nanopore restricting bit density via the finite spatial resolution of the nanopore itself and the finite temporal resolution of the instruments used in the experiment, there is an additional steric constraint when attempting to attach many structures onto a single carrier molecule. Steric hindrance, electrostatic repulsion, or both, may play a role in reducing the binding efficiency for positions along the contour length of the molecule where attachments bind at a later point in time. Whether this phenomenon is unique to the DNA stars studied here or if there is a more general effect requires further study, for example using monovalent streptavidin or

polyethylene glycol (PEG) molecules, which would reduce the electrostatic repulsion. This insight was central to informing the final design presented in this thesis.[118]

### 4.1.3 DNA Origami Bundles for DNA Barcodes

With the shortcomings of the star design better understood, the next iteration of the barcoding design attempted to fix these by making the structure more rigid, both to decrease the width of the blockage depth distribution and to enable more accurate peak classification. The logical next idea was to use different numbers of parallel DNA strands to distinguish bits of different sizes. For normal dsDNA translocation, different blockage levels are readily detected and distinguished by the number of strands in the pore at the same time.[94,122] With this fact in mind, the hypothesis was that a barcoding scheme using DNA origami helix bundles would offer simple and predictable blockage levels that would be more easily interpreted. The main challenge with this approach is ensuring that the attachment sections are long enough to be fully resolved in the current trace. For this to work properly, the DNA origami attachments need to be aligned with the axis of the main strand during translocation. A total of 3 different strategies were designed and tested to make this idea work and they are schematically laid out in **Figure 4.3**.

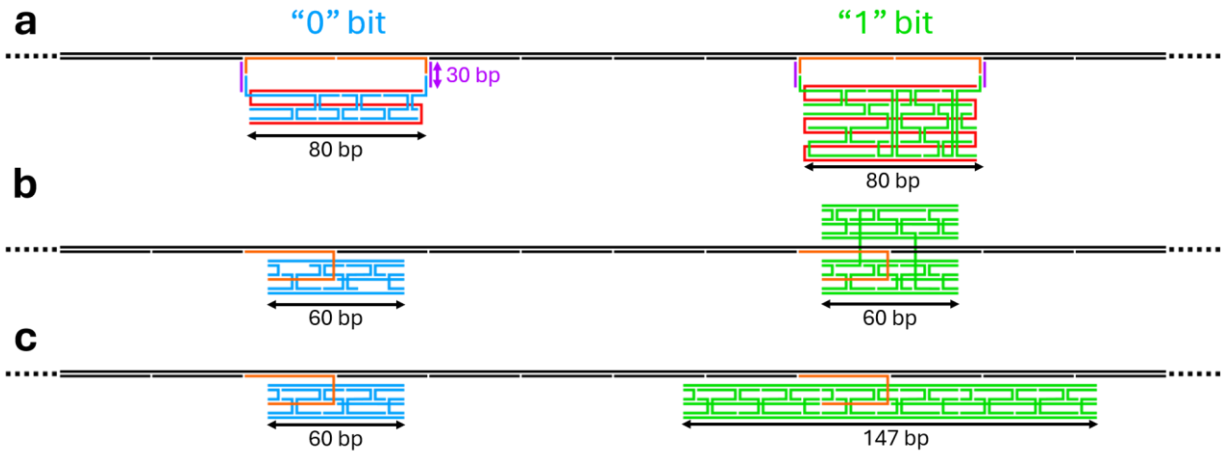



***Figure 4.3: Schematic overview of all bundle attachment strategies investigated in this study.*** *"0" bit designs are shown on the left while the corresponding "1" bit design is shown on the right. Oligos involved in forming the attachment structure are colored green (for "0") and blue (for "1"), oligos binding the attachment to the carrier are highlighted in orange.* ***a)*** *In this design the attachment is a small 3-helix bundle (3HB, "0" bit) and a 6-helix bundle (6HB, "1" bit) using scaffolded DNA origami with a long scaffold strand highlighted in red, and oligos encoding which attachment goes to which position marked in purple. The "0" and "1" bits are assembled separately and then bound to their respective positions.* ***b)*** *In this design the attachment is a small 3HB origami structure without a main scaffold strand. The "0" bit acts as the default state with empty ssDNA recognition sites to which an extra extension in the form of another 3HB origami binds to form the "1" bit, which is designed to wrap around the main carrier strand.* ***c)*** *This design is a variation of the previous one using the same structure for the "0" bit with the "1" bit varying the length of the attachment instead of the width.*

In the initial design of n-strand DNA helix bundle (nHB) structures, we used two binding sites per structure (one at the beginning and one at the end) to favor alignment of the attachment with the carrier axis. In this design the "0" consists of an 80 bp long 3HB structure while the "1" bit is a 6HB structure of the same length. Both attachments were synthesized using a standard DNA origami protocol with a custom-made scaffold strand of the proper length following a protocol (see **Appendix**) proposed by Wu *et al.*[151] This added extra steps to the assembly of the attachment bundles which meant the yield suffered (especially for the 6HB structure), but the scaffolded design should have made the structure more robust and kept its shape even when oligos are missing in the assembly. Additionally, the two-step binding of the attachment to the carrier to

align the structure along the axis of the carrier strand meant that there is the potential for 2 of the attachments attaching to either binding site at the same bit position, depending on the concentration of the attachment. On the nanopore side, the strategy of having the bundles off to the side of the main carrier strand with a 30 bp gap in between means that there are still degrees of freedom in the system – the attachment is able to move with respect to the carrier – widening the peak height distribution leading to increased rates of misclassification of peaks. This, together with the added complexity introduced by the scaffolded design of the attached bundles, led to a revamped design based on a paper by Seeman and colleagues.[228]

This next iteration of the bundle design is shown in **Figure 4.3b,** with a variant shown in **Figure 4.3c**. The main changes to the previous design include moving to a non-scaffolded design for the attachment to simplify the assembly protocol, only having a single attachment point to the carrier sequence to limit binding to the carrier to one structure at one position, and having the attachment bundles wrap around the carrier strand (for the 6HB version) to reduce attachment orientation flexibility on the carrier. The variant of this design changes the length of the attachment instead of the width (shown and analyzed in **Chapter 3)**. While more effective than the first iteration of the bundle, this design suffers from some of the same assembly issues, especially when trying to change "0" bits into "1" bits by binding an extra 3HB structure on the "0" bit. This, with the explained rotational motion of the attachment arising from the single attachment point acting as a hinge, necessitated yet another redesign. It should be noted that these failure modes are hard to detect in our DNA origami quality control. Since the overall structural behavior is dominated by the long carrier strand, the presence or absence of attachments usually does not significantly impact the diffusion behavior and is therefore not detectable by gel electrophoresis. Other than

direct imaging with AFM or TEM microscopes, the nanopore experiments are therefore the only option to observe these error modes.

## 4.2 DNA Origami Assembly

### 4.2.1 DNA Origami Purification

When assembling DNA origami (see **Section 1.2.2** and **Appendix**), especially when using a scaffold, the usual protocol is to add the staple oligos in excess of the scaffold strand. Most DNA origami studies have settled on a ratio of 3–5 times excess of staple strands to the scaffold strand. This solves several issues that would otherwise arise during DNA origami assembly. First, it guarantees that most scaffolds are folded into the desired shape as there are more than enough oligos in the mixture. Second, it counteracts slight variations in concentration of the staple oligos that might arise from differences in oligo synthesis and subsequent pipetting errors that might accumulate. Finally, it controls the reaction speed by reducing the time it takes for each structure to find all the necessary oligos. If the oligos were at the same concentration as the scaffold, the staple concentration would drop over the course of the assembly as oligos are bound by the forming structures and removed from free solution. This also minimizes the probability of forming multimers (structures where multiple scaffolds are bound together by the staple oligos). By the nature of a single staple oligo binding to multiple non-consecutive sections on the scaffold, it can also bind two scaffolds together. The rate at which multimers form depends on the relative concentration of the scaffold strand and staples and can be countered by increasing the staple oligo concentration in relation to the scaffold concentration. In this work, all origami assemblies are performed with a staple concentration at least 5× the scaffold concentration. In my work I therefore

work at 80 nM scaffold concentration and a 500 nM staple concentration giving a ratio of more than 5×.

After the assembly of the DNA origami, in addition to the main assembly, the final solution contains extra oligos (and any misfolded product, like half-formed structures or multimers, etc.). This requires cleanup, and potentially even an increase in the concentration of the final product, depending on the intended application. Importantly these steps need to be compatible with the structure of the origami and must not harm the staples holding the 3D shape together. Various methods have been shown and proven to be compatible with DNA origami,[148,230–232] and over the course of this study, multiple methods (gel purification, molecular weight cutoff filters, solid-phase reversible immobilization beads, ethanol precipitation, PEG precipitation) have been tried in service of maximizing yield and efficiency.

These purification methods make use of the different dynamics of DNA strands with different lengths. Molecular weight cutoff (MWCO) filters work by loading a mixture into a tube with porous membranes with certain sized holes and spinning it in a centrifuge.[233] Smaller molecules will pass right through the membrane while larger molecules are held back and can be retrieved by spinning the membrane tubes upside down. For this thesis Amicon® 100 kDa filters (Millipore Sigma, UFC510096) are used in all steps where MWCO are employed. Solid-phase reversible immobilization (SPRI) beads work by binding DNA based on the concentration of salt in the solution.[230] Larger molecules have a higher affinity of binding to the beads than shorter fragments and the exact cutoff point can be tuned by changing the bead-to-sample volume ratio.[234] Since the beads are paramagnetic, they can be pulled down through the use of a magnet and the

DNA bound to the beads can be released by resuspending them in a salt-free buffer solution. Gel purification makes use of the different migration speed of differently sized DNA fragments by electrophoresis through a gel medium (usually agarose or polyacrylamide) under a constant voltage.[235] Different migration speeds of DNA fragments lead to discrete bands of DNA that can be excised from the gel and dissolved to retrieve only the DNA in a specific band therefore removing all fragments of different sizes. PEG and ethanol precipitation work by mixing the DNA sample into a solution containing either PEG molecules or ethanol and centrifuging the mixture.[236,237] Long strands of DNA tend to aggregate in these solutions and form a pellet at the bottom of the tube, which can be resuspended after removing the supernatant in a clean buffer solution. The cutoff for the length at which the DNA precipitates can be tuned by changing the concentration of the PEG or ethanol solution.

Initially, MWCO filters were used for barcode purification and concentration due to their ease of use. More recently, SPRI beads have been employed due to their higher observed yield for these assemblies and easier scalability to cleaning up the smaller helix bundle (HB) structures used for attachment to the carrier. With the newest iteration of bundle structures (**Figure 5.1a**), this has led to problems of aggregation of barcoded structures which has prompted an investigation into purification methods that show minimal aggregation of the sample. **Figure 4.4** shows the result of four different purification methods of the DNA barcode using the carrier-integrated bit design. SPRI bead purification and PEG precipitation show significant aggregation of the barcode, identified by a large smear at much higher molecular weight. Ethanol precipitation does not show the same level of aggregation, but also barely removes any of the excess oligos. The MWCO filters show the lowest yield of all purification methods tested. This highlights the importance of properly

selecting and testing the purification method used. For purification steps of fully assembled DNA origami structures, when only excess oligos are removed, the Amicon filters perform well enough especially when yield is not an issue. For purification of smaller structures or when yield is important SPRI beads are the preferred option due to their tunable size cutoff and higher observed yield in recovery.

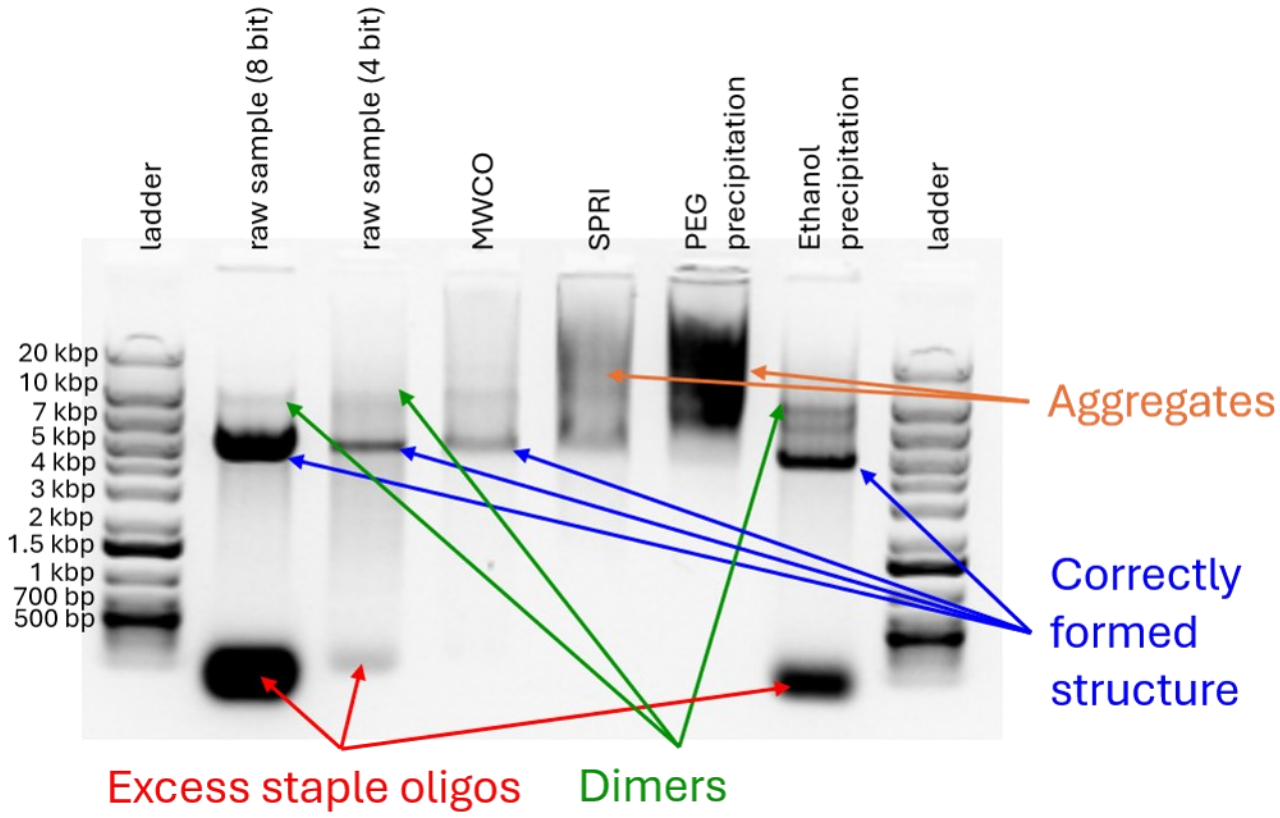



***Figure 4.4: Gel electrophoresis of the output of different purification methods for DNA barcodes.*** *The first and last lanes contain a GeneRuler 1 kb Plus DNA ladder for control. Lanes 2 and 3 contain an unpurified sample with slight variation (a 4-bit design and an 8-bit design) using the carrier-integrated HB structure. Lanes 4–7 contain the purified 8-bit version using the indicated purification method.*

### 4.2.2 DNA Origami Quality Control

Since the structure of DNA origami dictates its function, it is important to be able to confirm the correct assembly of the structure. In the field, this is done, in most cases, with a two-pronged approach: using gel electrophoresis in addition to an imaging technique like atomic force microscopy (AFM)[238,239] or cryo-EM.[240,241] While the gel electrophoresis is relatively quick and easy to do, it only gives information about the rough size of the resulting product without resolving the actual structure or resolving defects in the assembly beyond aggregation – a few missing staples are not typically visible in the gel. It is also a good first check to see if there are any multimers and to quickly estimate yield. Directly imaging the final origami structures gives direct evidence of the correct spatial assembly of the structure. This works particularly well with large 2D structures and 3D structures that have easily identifiable large-scale features.[210,242,243]

The barcoding structures studied in this work consist mostly of linear dsDNA with small attachments. While the presence or absence of these attachments could be resolvable by these imaging techniques, I preferred to perform secondary control of how these structures assembled by running them on a nanopore. Indeed, we have shown that solid-state nanopores can be very sensitive sensors to probe the assembly yield and the mechanical properties of DNA origami.[9] One example of this is shown in **Figure 4.5**.

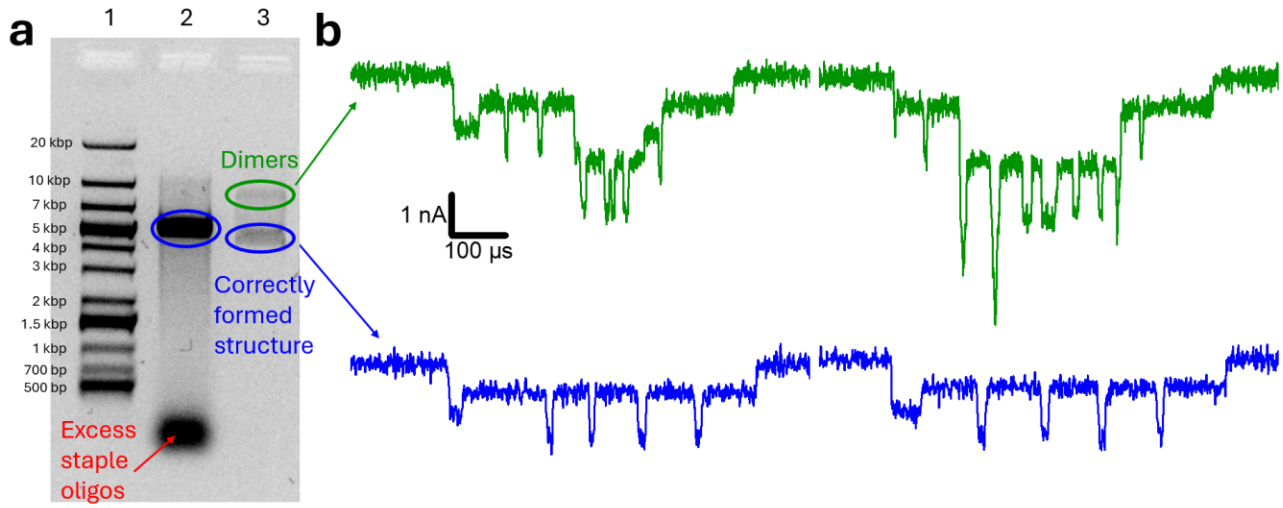



***Figure 4.5: DNA origami quality control*** *example using the 4-bit folded scaffold structure as shown in* ***Figure 5.1a****.* ***a)*** *Gel electrophoresis of a sample of unpurified (Lane 2) and purified (Lane 3) carrier with the folded backbone. Lane 1 contains the GeneRuler 1 kb Plus DNA ladder with the sizes of the bands indicated. The excess oligos are visible as a diffuse band in Lane 2.* ***b)*** *Example trace of the dimeric structure (green) and the correctly assembled barcode (blue) measured in a 10.8 nm nanopore in 3.6 M LiCl at 200 mV bias voltage.*

The main two quality control measurements in this work are therefore gel electrophoresis, to assess yield and discover obvious misfolding like half-formed products and multimers, and nanopore experiments for smaller molecular sized errors in assembly, like bits binding in the wrong position along the carrier and intramolecular secondary structures forming, that are not picked up by the gel electrophoresis.

# 5 Conclusion and Outlook

In this thesis I have presented a nanostructured DNA barcoding scheme that enables digital information to be encoded in the structure of molecules and decoded using solid-state nanopores. I have investigated the necessary requirements that a structure like this must satisfy to work reliably and explored the dynamics of passage of these molecules. I studied different designs of nanostructured DNA carriers with the goal of generating distinguishable signatures in the resulting current traces. While further studies will be needed to scale these results into a practical digital information storage method, the results shown in this thesis clearly show that molecular information storage using solid-state nanopores as a readout mechanism is feasible in principle.

With the insights gleaned in the course of the research that led to this thesis, a practical, reliable, and adaptable DNA nanostructured barcode for multiplexed target detection and digital data storage that is optimized for membrane-based solid-state nanopores of the dimensions used here appears closer than ever. In the final moments leading up to the submission of this thesis, further experiments were performed, in collaboration with Nada Kerrouri (PhD student in the group, who will continue this type of work toward multiplexed biomarker detection assays) and Dr. Zachary Roelen (Nanopore Scientist at Oxford Nanopore Technologies Canada), with the latest design of the DNA nanostructures, building on the knowledge developed in this thesis. **Figure 5.1** shows the preliminary results, which look highly promising. This version combines all the lessons learned with the previous barcoded designs and is based on a proven structure that has been successful in measuring the DNA velocity profiles during nanopore passage.[118] One advantage of this design is that the “0” bit is formed by folding of the main carrier strand, meaning it is directly

aligned with the axis of the contour of the main carrier, removing the issue with the flexibility of that attached bit entirely. It also means that each event should always have the correct number of peaks (given a full-length carrier) even if binding of the "1" bit extension fails. Experiments with this design are still underway, but preliminary results suggest that this design allows for robust classification with the "0" bit producing reliable, tightly distributed peak heights. The issue of encoding bits into structures with sufficiently different size to generate distinguishable peaks in the nanopore current traces also appears to have found a solution with this design: events with fully readable binary barcodes of "0" and "1" bits are now more frequently observed.

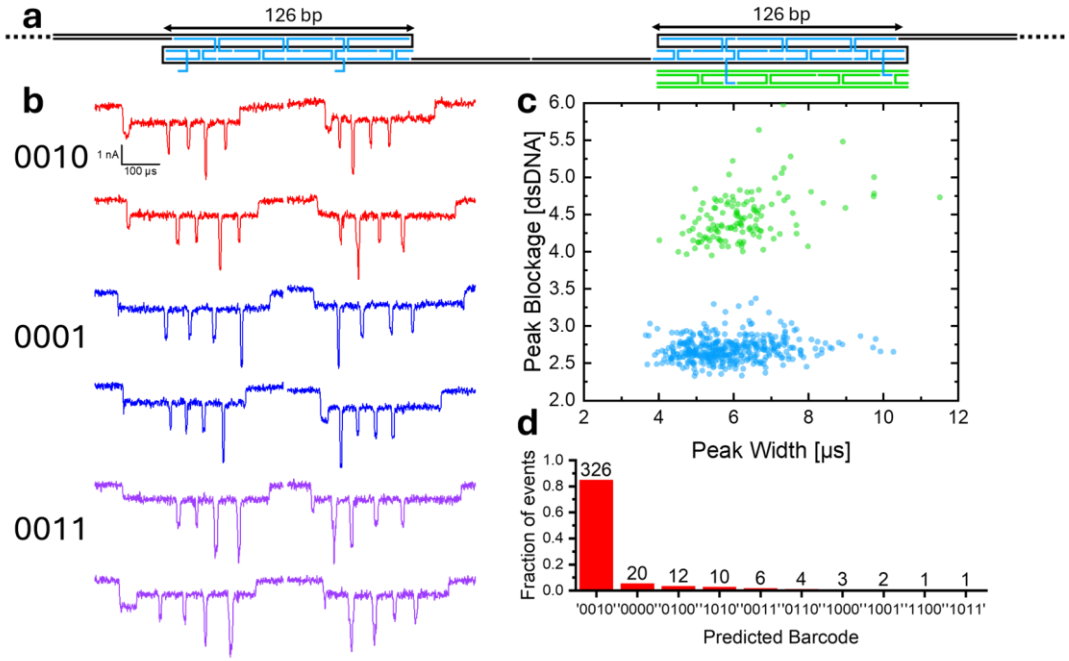



***Figure 5.1: Overview of most recent results. a)*** *Schematic of the design integrating the bundle into the carrier strand by folding it using DNA origami staple strands. The "0" bit version on the left consists of a 3HB with two overhangs, the "1" bit binds an extra 2HB structure.* ***b)*** *Sample traces for the three barcodes tested so far. Traces on the left are identified as having translocated in the forward direction, traces on the right are identified to have translocated in the backward direction determined by the ECD before the first detected peak and after the last detected peak.* ***c)*** *Scatter plot of blockage depth (in multiples of dsDNA blockage) and width of all detected peaks for events with 4 identified peaks for the "0010" sample. Green points are classified as "1", while blue points are classified as "0".* ***d)*** *Full classification of all events with 4 peaks detected for the "0010" sample. Numbers in the plot show the total number of events in each class.*

Despite the success of these experiments, there remain unanswered questions that will need to be addressed before this DNA nanostructured barcode can be used for molecular information storage or for multiplexed detection of biomarkers. First and foremost, the misclassification rate of complete barcodes needs to be investigated by expanding the analysis presented in **Chapter 3** to classify barcodes with up to 8 binary bits instead of just 2. Second, it will be important to perform statistical analysis of multiple barcodes in a mixture at the same time to confirm that these structures remain distinguishable even when the barcode for each individual event is unknown. Third, the scope of the barcodes should be expanded by adding additional binary bits to increase multiplexing capabilities, laying the groundwork for nanostructured DNA data storage.

While the next steps should focus on ensuring that the majority of events are readable barcodes, there will always be a subset of events that do not give readable barcodes due to the flexible nature of these polymer structures and the inefficiencies in the biochemical interactions that mediate their assembly. This can be addressed in the data analysis stage to an extent, and there is still work to be done to reliably filter unreadable barcodes while guaranteeing that the maximal number of readable barcodes are kept in the analysis pipeline. A fraction of structures, however, are also not forming correctly during the assembly process for a variety of reasons and making sure that the biochemical processes happen as efficiently as possible (or misfolded structures are removed as efficiently as possible from the sample mixture) is still an outstanding problem once a design has been finalized. Secondary structures and dimers are formed when off-target binding of staple oligos creates stable bindings that are not desired and can significantly affect yield of assembly, reducing the number of correctly identifiable events by the nanopore.

Once these issues are resolved, these structures should provide a reliable and scalable platform for multiplexed single-molecule detection schemes using solid-state nanopores as well as nanostructured DNA data storage schemes. Being able to tag, for example, different biomarkers in a blood sample, with these structures in the same sample gives each biomarker a unique barcode allowing simultaneous detection, which cuts down on experiment time as well as the required amount of sample volume, which especially in medical applications can be in limited quantities. This can be done either by directly binding the biomarker to these barcodes[160] or by translating the biomarker through the use of a bioassay into a molecule or structure that then binds to the barcode.[220,244]

For barcoded DNA data storage applications the scalability of the structures presented here would be of great use and being able to reliably read and write nanostructured DNA bit sequences enables the application of a whole host of strategies from information theory that can be applied to the reading and writing of binary data to turn it into a robust data storage scheme.[26,168] Being able to write individual bits allows, for example, error-correction algorithms to detect and fix errors in the binary sequence on the spot without having to reassemble the entire structure from scratch. As our knowledge in this area continues to expand, the foundation laid here offers a highly promising frontier for future doctoral scholars to explore.

# Appendix

## Appendix A: DNA Carrier Assembly

The assembly of the carrier for each of the assemblies shown in this thesis follows a standard DNA origami protocol as outlined by this review paper from the Dietz lab.[148] Due to the carrier mostly consisting of dsDNA the structural complexity of this carrier is much lower than most conventional DNA origami structures that are usually designed to take on a certain 2D or 3D shape which require complex folding patterns and staples that interact with very disparate parts of the scaffold strand. As such, the staple oligo sequences are generated in a straightforward way from the reverse complement of the scaffold sequence that is split into oligos of similar lengths (for the purposes of this thesis that is 40 nt or 60 nt long oligos). For the purposes of attaching structures onto the carrier, overhang sequences are generated randomly and manually checked for off-target interactions with other overhang sequences and their complements through the use of tools like the oligo analyzer offered on the IDT website (https://www.idtdna.com/calc/analyzer).

The scaffold used for all structures in this thesis is based on the M13 bacteriophage genome and is commercially available for purchase through many vendors (e.g. Tilibit Nanosystems https://www.tilibit.com/products/single-stranded-scaffold-dna-type-p7249). This genome is a circular 7249 nt long single stranded DNA fragment. To act as a carrier for the attachments it needs to be linearized which is done through the use of a restriction enzyme. Many choices for restriction enzymes acting on the M13 sequence exist but, in this thesis, HincII (https://www.neb.com/en-ca/products/r0103-hincii) was used for all structures. The M13 sequence contains exactly one

restriction site for HincII making it ideal for our purposes. Because it only cuts dsDNA, a short 20 nt oligo is added to the mixture when performing the restriction reaction with HincII (the left-over pieces are later removed through strand displacement reactions and do not interfere with downstream processes, as HincII leaves both ends blunt). The full linearization protocol is adapted from the instructions for the use of HincII that come with the commercially available version of the protein (see link above). The full reaction mixture contains these components:

- 10 µl M13 stock (400 nM as per Tilibit Nanosystems)
- 5 µl HincII oligo (10 µM)
- 10 µl HincII
- 5 µl NEB CutSmart Buffer
- 20 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

This mixture is incubated at 37 °C for at least 1 h and then heated to 65 °C for 20 minutes to denature the HincII protein and stop the HincII from cutting at off-target sites. This cutting step introduces a couple of problems that are described in detail in **Chapter 2**. The final cut M13 mixture can be used as is without cleanup in the next step which involves adding the necessary staple oligo to build the carrier strand.

The carrier assembly is an adapted version of standard DNA origami protocols and involves mixing of the staples with the carrier mixture from the linearization step and adding $MgCl_2$ to stabilize the DNA origami. The origami mixture contains these components:

- 10 µl M13 linearized (~80 nM estimated from mixture above)
- 10 µl oligo mixture (~500 nM for each individual oligo, up to 190 oligos in the mixture)
- 30 µl TEMg buffer (10 mM Tris-HCl, 1 mM EDTA, 20 mM $MgCl_2$, pH 8)

This mixture is heated to 70 °C for 10 minutes then cooled quickly to 60 °C and slowly ramped down to 40 °C over the course of 5 hours and subsequently further cooled to room temperature as per a recent study.[149] Afterwards the assembled products are purified using Amicon molecular weight cutoff filters or SPRI beads as described in **Chapter 4** to remove excess oligos as well as the remainder of the denatured protein. This carrier mixture is then used as the basis to be mixed with the corresponding attachments and incubated at room temperature.

# Appendix B: Small Bundle Assembly

The first iteration of bundle design used small DNA origami bundles as attachments for structural barcoding. The design of these small bundles is the same as the design for larger DNA origami structures using a scaffold strand and a set of staple oligos that fold the scaffold strand into the desired shape. The scaffolds for the 3HB (235 nt) and 6HB (473 nt) bundle structures were randomly generated sequences that were synthesized according to the protocol proposed by Wu *et al*.[151] This protocol is a two-step reaction that takes a set of 10 oligos (for 3HB) and 20 oligos (for 6HB) to synthesize a full-length double-stranded scaffold strand. The first step is a polymerase cycling assembly (PCA) in which the scaffold strand is built from the component oligos into one full length strand. The reaction mixture contains these components:

- 1 µl oligo mixture (1.2 µM each oligo)
- 10 µl 2x Polymerase mix (Taq Polymerase master mix, ready-to-use https://www.thermofisher.com/order/catalog/product/14966005)
- 9 µl deionized water

This mixture is then put into the thermocycler (94 °C for 2 min followed by 55 cycles of 94 °C for 15 s, 60 °C for 15 s, 68 °C for 15 s). The second step is a polymerase chain reaction that uses two

modified primers to amplify the full-length scaffold only without amplifying the intermediate products that might have been synthesized in the PCA step. The modifications on the primers consist of five phosphorothioated bases on the 5' end of the scaffold strand and a phosphorylated base at the end of the complement strand. These modifications are important for the digestion step to turn the double-stranded scaffold into a single-stranded scaffold strand ready to use for DNA origami. The PCR mixture contains the following components:

- 2 µl PCA reaction mixture
- 10 µl Primer mixture (5 µM each, working concentration 1 µM)
- 25 µl 2x Polymerase mix (same as above)
- 13 µl deionized water

This mixture is the put in the thermocycler using the 70 cycles with the same temperatures as described above. The final mixture is purified using SPRI beads as described in Chau *et al*.[230] In a final step the scaffold strands are digested using a λ exonuclease to selectively digest the complement strand leaving behind a single-stranded scaffold strand of the desired length and sequence. The reaction mixture for this digestion contains the following components:

- 20 µl PCR mixture
- 5 µl λ exonuclease buffer (10x)
- 1 µl λ exonuclease (5 units, https://www.neb.com/en-ca/products/m0262-lambda-exonuclease)
- 24 µl deionized water

This mixture is then incubated at 37 °C for 3 h before being heated to 75 °C for 10 minutes to denature the protein. The resulting solution is used as is for DNA origami assembly as described in the previous section. After full assembly of the small bundle structures another purification step is performed using SPRI beads.

# References


1. Bull, B. S., Schneiderman, M. A. & Brecher, G. Platelet Counts with the Coulter Counter. *The American Journal of Clinical Pathology* **44**, 678–688 (1965).
2. Hurley, J. Sizing Particles with a Coulter Counter. *Biophys. J.* **10**, 74–79 (1970).
3. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved Patch-Clamp Techniques for High-Resolution Current Recording from Cells and Cell-Free Membrane Patches. *European Journal of Physiology* **391**, 85–100 (1981).
4. Neher, E., Sakmann, B. & Henry Steinbach, J. The Extracellular Patch Clamp: A Method for Resolving Currents through Individual Open Channels in Biological Membranes. *European Journal of Physiology* **375**, 219–228 (1978).
5. Sakmann, B. & Neher, E. Patch Clamp Techniques for Studying Ionic Channels in Excitable Membranes. *Annu. Rev. Physiol.* **46**, 455–472 (1984).
6. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *PNAS* **93**, 13770–13773 (1996).
7. Karau, P. & Tabard-Cossa, V. Capture and Translocation Characteristics of Short Branched DNA Labels in Solid-State Nanopores. *ACS Sens.* **3**, 1308–1315 (2018).
8. He, L., Karau, P. & Tabard-Cossa, V. Fast capture and multiplexed detection of short multi-arm DNA stars in solid-state nanopores. *Nanoscale* **11**, 16342–16350 (2019).
9. He, L. *et al.* DNA origami characterized via a solid-state nanopore: insights into nanostructure dimensions, rigidity and yield. *Nanoscale* **15**, 14043–14054 (2023).
10. Mensing, P., Charron, M., Bouhamidi, M. Y. & Tabard-Cossa, V. Folding Dynamics of Linear ssDNA–dsDNA Heterostructures through Solid-State Nanopores. *Macromolecules* **58**, 4094–4102 (2025).
11. Storm, A. J., Chen, J. H., Zandbergen, H. W. & Dekker, C. Translocation of double-strand DNA through a silicon oxide nanopore. *Phys. Rev. E* **71**, 051903 (2005).
12. Platnich, C. M., Earle, M. K. & Keyser, U. F. Chemical Annealing Restructures RNA for Nanopore Detection. *J. Am. Chem. Soc.* **146**, 12919–12924 (2024).
13. Tripathi, P. *et al.* Discrimination of RNA fiber structures using solid-state nanopores. *Nanoscale* **14**, 6866–6875 (2022).
14. Schmid, S., Stömmer, P., Dietz, H. & Dekker, C. Nanopore electro-osmotic trap for the label-free study of single proteins and their conformations. *Nat. Nanotechnol.* **16**, 1244–1250 (2021).
15. Sauciuc, A. *et al.* Blobs form during the single-file transport of proteins across nanopores. *PNAS* **121**, e2405018121 (2024).
16. Boukhet, M. *et al.* Translocation of Precision Polymers through Biological Nanopores. *Macromol. Rapid Commun.* **38**, 1–6 (2017).

17. Ren, M. *et al.* Single-molecule resolution of the conformation of polymers and dendrimers with solid-state nanopores. *Talanta* **286**, 127544 (2025).

18. Cao, C. *et al.* Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. *Sci. Adv.* **6**, eabc2661 (2020).

19. Bayley, H. Nanopore sequencing: From imagination to reality. *Clin. Chem.* **61**, 25–31 (2015).

20. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).

21. Yusko, E. C. *et al.* Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* **12**, 360–367 (2017).

22. Meng, F., Li, X., Zou, N. & Wang, X. Protein Profiling by Nanopore-Based Technology. *Anal. Chem.* **97**, 10110–10125 (2025).

23. Afshar Bakshloo, M. *et al.* Nanopore-Based Protein Identification. *J. Am. Chem. Soc.* **144**, 2716–2725 (2022).

24. Lucas, F. L. R., Versloot, R. C. A., Yakovlieva, L., Walvoort, M. T. C. & Maglia, G. Protein identification by nanopore peptide profiling. *Nat. Commun.* **12**, 1–9 (2021).

25. Soni, N. *et al.* Full-length protein classification via cysteine fingerprinting in solid-state nanopores. *Nat. Nanotechnol.* **20**, 1482–1490 (2025).

26. Chen, K., Zhu, J., Bošković, F. & Keyser, U. F. Nanopore-Based DNA Hard Drives for Rewritable and Secure Data Storage. *Nano Lett.* **20**, 3754–3760 (2020).

27. Chen, K. *et al.* Digital Data Storage Using DNA Nanostructures and Solid-State Nanopores. *Nano Lett.* **19**, 1210–1215 (2019).

28. Shi, X. *et al.* Sustained unidirectional rotation of a self-organized DNA rotor on a nanopore. *Nat. Phys.* **18**, 1105–1111 (2022).

29. Meyer, N. *et al.* Solid-state and polymer nanopores for protein sensing: A review. *Adv. Colloid Interface Sci.* **298**, 102561 (2021).

30. Zuo, J. *et al.* Review—Single-Molecule Sensors Based on Protein Nanopores. *J. Electrochem. Soc.* **168**, 126502 (2021).

31. Bell, N. A. W. *et al.* DNA Origami Nanopores. *Nano Lett.* **12**, 512–517 (2012).

32. Bell, N. A. W. & Keyser, U. F. Nanopores formed by DNA origami: A review. *FEBS Lett.* **588**, 3564–3570 (2014).

33. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2**, 209–215 (2007).

34. Ying, Y.-L., Cao, C. & Long, Y.-T. Single molecule analysis by biological nanopore sensors. *Analyst* **139**, 3826–3835 (2014).

35. Bafna, J. A. & Soni, G. V. Fabrication of Low Noise Borosilicate Glass Nanopores for Single Molecule Sensing. *PLoS One* **11**, e0157399 (2016).

36. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).

37. Wei, X. *et al.* Engineering Biological Nanopore Approaches toward Protein Sequencing. *ACS Nano* **17**, 16369–16395 (2023).

38. Arjmandi-Tash, H., Belyaeva, L. A. & Schneider, G. F. Single molecule detection with graphene and other two-dimensional materials: nanopores and beyond. *Chem. Soc. Rev.* **45**, 476–493 (2015).

39. Wu, M.-Y., Krapf, D., Zandbergen, M., Zandbergen, H. & Batson, P. E. Formation of nanopores in a SiN/SiO2 membrane with an electron beam. *Appl. Phys. Lett.* **87**, 113106 (2005).

40. Stanley, J. & Pourmand, N. Nanopipettes—The past and the present. *APL Mater.* **8**, 100902 (2020).

41. Alawami, M. F. *et al.* Lifetime of glass nanopores in a PDMS chip for single-molecule sensing. *iScience* **25**, 104191 (2022).

42. Bell, N. A. W., Chen, K., Ghosal, S., Ricci, M. & Keyser, U. F. Asymmetric dynamics of DNA entering and exiting a strongly confining nanopore. *Nat. Commun.* **8**, 380 (2017).

43. Briggs, K., Waugh, M. & Tabard-Cossa, V. Fabricating Solid-State Nanopores for Single-Molecule Sensing. in *Solid State Nanopores: From Fabrication to Biosensing* 1–33 (Springer, 2023). doi:10.1007/978-3-031-42336-9_1.

44. Lin, K. *et al.* Fabrication of solid-state nanopores. *Nanotechnology* **33**, 272003 (2022).

45. Xue, L. *et al.* Solid-state nanopore sensors. *Nat. Rev. Mater.* **5**, 931–951 (2020).

46. Fried, J. P. *et al.* In situ solid-state nanopore fabrication. *Chem. Soc. Rev.* **50**, 4974–4992 (2021).

47. Kwok, H., Briggs, K. & Tabard-Cossa, V. Nanopore fabrication by controlled dielectric breakdown. *PLoS One* **9**, e92880 (2014).

48. Waugh, M. *et al.* Solid-state nanopore fabrication by automated controlled breakdown. *Nat. Protoc.* **15**, 122–143 (2020).

49. Briggs, K. *et al.* Kinetics of nanopore fabrication during controlled breakdown of dielectric membranes in solution. *Nanotechnology* **26**, 084004 (2015).

50. Tabard-Cossa, V. Instrumentation for Low-Noise High-Bandwidth Nanopore Recording. in *Engineered Nanopores for Bioanalytical Applications* (eds. Edel, J. B. & Albrecht, T.) 59–93 (Elsevier, New York City, 2013). doi:10.1016/B978-1-4377-3473-7.00003-0.

51. Kowalczyk, S. W., Grosberg, A. Y., Rabin, Y. & Dekker, C. Modeling the conductance and DNA blockade of solid-state nanopores. *Nanotechnology* **22**, 315101 (2011).

52. Beamish, E., Kwok, H., Tabard-Cossa, V. & Godin, M. Fine-tuning the Size and Minimizing the Noise of Solid-state Nanopores. *Journal of Visualized Experiments* e51081 (2013) doi:10.3791/51081.

53. Charron, M., Roelen, Z., Wadhwa, D. & Tabard-Cossa, V. Improved Conductance Blockage Modeling of Cylindrical Nanopores, from 2D to Thick Membranes. *Nano Lett.* **24**, 10527–10533 (2024).

54. Fragasso, A., Pud, S. & Dekker, C. 1/ *f* noise in solid-state nanopores is governed by access and surface regions. *Nanotechnology* **30**, 395202 (2019).

55. Saharia, J., Bandara, Y. M. N. D. Y., Karawdeniya, B. I., Alexandrakis, G. & Kim, M. J. Assessment of 1/ *f* noise associated with nanopores fabricated through chemically tuned controlled dielectric breakdown. *Electrophoresis* **42**, 899–909 (2021).

56. Su, S. *et al.* Origin of nonequilibrium 1/f noise in solid-state nanopores. *Nanoscale* **12**, 8975–8981 (2020).

57. Smeets, R. M. M., Keyser, U. F., Wu, M. Y., Dekker, N. H. & Dekker, C. Nanobubbles in solid-state nanopores. *Phys. Rev. Lett.* **97**, 1–4 (2006).

58. Tabard-Cossa, V., Trivedi, D., Wiggin, M., Jetha, N. N. & Marziali, A. Noise analysis and reduction in solid-state nanopores. *Nanotechnology* **18**, 1–6 (2007).

59. Brauburger, S. & Keyser, U. F. A biophysicist's guide to translocation of double-stranded nucleic acids through solid-state nanopores. *Phys. Life Rev.* **57**, 85–131 (2026).

60. Charron, M., Philipp, L., He, L. & Tabard-Cossa, V. Elucidating the dynamics of polymer transport through nanopores using asymmetric salt concentrations. *Nano Res.* **15**, 9943–9953 (2022).

61. Charron, M. Nanopore Resistive Sensing: Bridging Theory and Experiments. (University of Ottawa, Ottawa, 2025).

62. Shi, L. *et al.* The Influence of Electric Field Intensity and Particle Length on the Electrokinetic Transport of Cylindrical Particles Passing through Nanopore. *Micromachines (Basel).* **11**, 722 (2020).

63. Ai, Y., Liu, J., Zhang, B. & Qian, S. Field effect regulation of DNA translocation through a nanopore. *Anal. Chem.* **82**, 8217–8225 (2010).

64. Vollmer, S. C. & de Haan, H. W. Translocation is a nonequilibrium process at all stages: Simulating the capture and translocation of a polymer by a nanopore. *J. Chem. Phys.* **145**, 154902 (2016).

65. Qiao, L., Ignacio, M. & Slater, G. W. Voltage-driven translocation: Defining a capture radius. *Journal of Chemical Physics* **151**, 244902 (2019).

66. Rowghanian, P. & Grosberg, A. Y. Electrophoretic capture of a DNA chain into a nanopore. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **87**, 042722 (2013).

67. Grosberg, A. Y. & Rabin, Y. DNA capture into a nanopore: Interplay of diffusion and electrohydrodynamics. *J. Chem. Phys.* **133**, 165102 (2010).

68. Muthukumar, M. Theory of capture rate in polymer translocation. *Journal of Chemical Physics* **132**, 195101 (2010).

69. Qiao, L. & Slater, G. W. Capture of rod-like molecules by a nanopore: Defining an "orientational capture radius". *J. Chem. Phys.* **152**, 144902 (2020).

70. Charron, M., Briggs, K., King, S., Waugh, M. & Tabard-Cossa, V. Precise DNA Concentration Measurements with Nanopores by Controlled Counting. *Anal. Chem.* **91**, 12228–12237 (2019).

71. Zhang, Y. *et al.* Concentration effects on capture rate and translocation configuration of nanopore-based DNA detection. *Electrophoresis* **41**, 1523–1528 (2020).

72. Freedman, K. J. *et al.* Nanopore sensing at ultra-low concentrations using single-molecule dielectrophoretic trapping. *Nat. Commun.* **7**, 10217 (2016).

73. Lam, M. H. *et al.* Entropic Trapping of DNA with a Nanofiltered Nanopore. *ACS Appl. Nano Mater.* **2**, 4773–4781 (2019).

74. Jeon, B. & Muthukumar, M. Polymer capture by α-hemolysin pore upon salt concentration gradient. *J. Chem. Phys.* **140**, 015101 (2014).

75. Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y. & Meller, A. Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient. *Nat. Nanotechnol.* **5**, 160–165 (2010).

76. Sohi, A. N., Beamish, E., Tabard-Cossa, V. & Godin, M. DNA Capture by Nanopore Sensors under Flow. *Anal. Chem.* **92**, 8108–8116 (2020).

77. Sharma, V., Farajpour, N., Lastra, L. S. & Freedman, K. J. DNA Coil Dynamics and Hydrodynamic Gating of Pressure-Biased Nanopores. *Small* **18**, 2106803 (2022).

78. Biswas, P., Paramekanti, A. & Cherayil, B. J. Shapes of generalized random walks. *Journal of Chemical Physics* **104**, 3360–3365 (1996).

79. Mihovilovic, M., Hagerty, N. & Stein, D. Statistics of DNA Capture by a Solid-State Nanopore. *Phys. Rev. Lett.* **110**, 028102 (2013).

80. Chen, K. *et al.* Super-Resolution Detection of DNA Nanostructures Using a Nanopore. *Adv. Mater.* **35**, 2207434 (2023).

81. Briggs, K. *et al.* DNA Translocations through Nanopores under Nanoscale Preconfinement. *Nano Lett.* **18**, 660–668 (2018).

82. Fleming, A. M., Mathewson, N. J., Howpay Manage, S. A. & Burrows, C. J. Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. *ACS Cent. Sci.* **7**, 1707–1717 (2021).

83. Jetha, N. N., Feehan, C., Wiggin, M., Tabard-Cossa, V. & Marziali, A. Long dwell-time passage of DNA through nanometer-scale pores: Kinetics and sequence dependence of motion. *Biophys. J.* **100**, 2974–2980 (2011).

84. Butler, T. Z., Gundlach, J. H. & Troll, M. Ionic current blockades from DNA and RNA molecules in the α-hemolysin nanopore. *Biophys. J.* **93**, 3229–3240 (2007).

85. Roelen, Z. & Tabard-Cossa, V. Synthesis of length-tunable DNA carriers for nanopore sensing. *PLoS One* **18**, e0290559 (2023).

86. Das, N., Samanta, N., Joseph, A. M., Ghosh, H. & RoyChaudhuri, C. Noise Spectroscopy Analysis for Estimation of a Specific Biomolecule in a Complex Mixture Using Solid-State Nanopore. *IEEE Trans. Nanotechnol.* **17**, 1138–1145 (2018).

87. Healy, K., Schiedt, B. & Morrison, A. P. Solid-State Nanopore Technologies for Nanopore-Based DNA Analysis. *Nanomedicine* **2**, 875–897 (2007).

88. Plesa, C. *et al.* Ionic permeability and mechanical properties of DNA origami nanoplates on solid-state nanopores. *ACS Nano* **8**, 35–43 (2014).

89. Wang, J. *et al.* Rational design of a 3D DNA origami cube as an ideal signal carrier for glass nanopore-based biosensors. *Anal. Chim. Acta* **1343**, 343660 (2025).

90. Chuah, K. *et al.* Nanopore blockade sensors for ultrasensitive detection of proteins in complex biological samples. *Nat. Commun.* **10**, 2109 (2019).

91. Plesa, C., Ruitenberg, J. W., Witteveen, M. J. & Dekker, C. Detection of individual proteins bound along DNA using solid-state nanopores. *Nano Lett.* **15**, 3153–3158 (2015).

92. Freedman, K. J., Bastian, A. R., Chaiken, I. & Kim, M. J. Solid-State Nanopore Detection of Protein Complexes: Applications in Healthcare and Protein Kinetics. *Small* **9**, 750–759 (2013).

93. Baaken, G. *et al.* High-Resolution Size-Discrimination of Single Nonionic Synthetic Polymers with a Highly Charged Biological Nanopore. *ACS Nano* **9**, 6443–6449 (2015).

94. Kumar Sharma, R., Agrawal, I., Dai, L., Doyle, P. S. & Garaj, S. Complex DNA knots detected with a nanopore sensor. *Nat. Commun.* **10**, 4473 (2019).

95. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).

96. Zhu, L., Zhang, Z. & Liu, Q. Deformation-Mediated Translocation of DNA Origami Nanoplates through a Narrow Solid-State Nanopore. *Anal. Chem.* **92**, 13238–13245 (2020).

97. Manrao, E. A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).

98. Maugi, R. *et al.* A methodology for characterising nanoparticle size and shape using nanopores. *Nanoscale* **12**, 262–270 (2020).

99. Heng, J. B. *et al.* Sizing DNA using a nanometer-diameter pore. *Biophys. J.* **87**, 2905–2911 (2004).

100. Straathof, S. *et al.* Protein Sizing with 15 nm Conical Biological Nanopore YaxAB. *ACS Nano* **17**, 13685–13699 (2023).

101. Sutherland, T. C. *et al.* Structure of peptides investigated by nanopore analysis. *Nano Lett.* **4**, 1273–1277 (2004).

102. Li, S., Cao, C., Yang, J. & Long, Y. T. Detection of Peptides with Different Charges and Lengths by Using the Aerolysin Nanopore. *ChemElectroChem* **6**, 126–129 (2019).

103. Waduge, P. *et al.* Nanopore-Based Measurements of Protein Size, Fluctuations, and Conformational Changes. *ACS Nano* **11**, 5706–5716 (2017).

104. Liu, W., Yang, Z.-L., Yang, C.-N., Ying, Y.-L. & Long, Y.-T. Profiling single-molecule reaction kinetics under nanopore confinement. *Chem. Sci.* **13**, 4109–4114 (2022).

105. Ying, Y.-L., Yang, C.-N., Liu, W., Wu, X.-Y. & Long, Y.-T. Understanding single-molecule reactions using nanopore-based techniques. *Nat. Chem.* **17**, 1450–1461 (2025).

106. Chau, C. C. C., Weckman, N. E., Thomson, E. E. & Actis, P. Solid-State Nanopore Real-Time Assay for Monitoring Cas9 Endonuclease Reactivity. *ACS Nano* **19**, 3839–3851 (2025).

107. Nova, I. C. *et al.* Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2023).

108. Zheng, F. & Han, Q. Distinct DNA conformations during forward and backward translocations through a conical nanopore. *Analyst* **149**, 5131–5138 (2024).

109. Zheng, H. *et al.* Nanopore Label-Free Monitoring of B-DNA and Z-DNA Conformational Transition in Real Time. *ACS Nano* **20**, 5208–5216 (2026).

110. Ratinho, L., Meyer, N., Greive, S., Cressiot, B. & Pelta, J. Nanopore sensing of protein and peptide conformation for point-of-care applications. *Nat. Commun.* **16**, 3211 (2025).

111. Bošković, F. & Keyser, U. F. Nanopore microscope identifies RNA isoforms with structural colours. *Nat. Chem.* **14**, 1258–1264 (2022).

112. Schmidt, T. T. *et al.* Supramolecular Interactions Modulate RNA:DNA Folding Observed via Nanopore Sensing. *Angew. Chem. Int. Ed.* **64**, e202508917 (2025).

113. Wanunu, M., Sutin, J. & Meller, A. DNA Profiling Using Solid-State Nanopores: Detection of DNA-Binding Molecules. *Nano Lett.* **9**, 3498–3502 (2009).

114. Thakur, A. K. & Movileanu, L. Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore. *Nat. Biotechnol.* **37**, 96–104 (2019).

115. Marshall, M. M. *et al.* Nanopore analysis of single-stranded binding protein interactions with DNA. *Langmuir* **31**, 4582–4588 (2015).

116. Hurt, N., Wang, H., Akeson, M. & Lieberman, K. R. Specific nucleotide binding and rebinding to individual DNA polymerase complexes captured on a Nanopore. *J. Am. Chem. Soc.* **131**, 3772–3778 (2009).

117. Lu, B., Albertorio, F., Hoogerheide, D. P. & Golovchenko, J. A. Origins and consequences of velocity fluctuations during DNA passage through a nanopore. *Biophys. J.* **101**, 70–79 (2011).

118. Charron, M., Elliott, B., Kerrouri, N., He, L. & Tabard-Cossa, V. Velocity Fluctuation and Force Scaling During Driven Polymer Transport through a Nanopore. *ACS Nano* **19**, 30950–30960 (2025).

119. Plesa, C., van Loo, N., Ketterer, P., Dietz, H. & Dekker, C. Velocity of DNA during Translocation through a Solid-State Nanopore. *Nano Lett.* **15**, 732–737 (2015).

120. Charron, M. & Tabard-Cossa, V. Near-exact resistance of an orifice in a 2D membrane blocked by a cylindrical obstruction. *Appl. Phys. Lett.* **129**, 163701 (2026).

121. Tian, P. & Smith, G. D. Translocation of a polymer chain across a nanopore: A Brownian dynamics simulation study. *Journal of Chemical Physics* **119**, 11475–11483 (2003).

122. Plesa, C. *et al.* Direct observation of DNA knots using a solid-state nanopore. *Nat. Nanotechnol.* **11**, 1093–1097 (2016).

123. Maheshwaram, S. K., Sreenivasa, K. & Soni, G. V. Fingerprinting branches on supercoiled plasmid DNA using quartz nanocapillaries. *Nanoscale* **13**, 320–331 (2021).

124. Zheng, F. *et al.* Torsion-Driven Plectoneme Formation During Nanopore Translocation of DNA Polymers. *Phys. Rev. X* **15**, 031041 (2025).

125. Lu, B. *et al.* Pressure-controlled motion of single polymers through solid-state nanopores. *Nano Lett.* **13**, 3048–3052 (2013).

126. Hoogerheide, D. P., Lu, B. & Golovchenko, J. A. Pressure–Voltage Trap for DNA near a Solid-State Nanopore. *ACS Nano* **8**, 7384–7391 (2014).

127. Waugh, M. *et al.* Interfacing solid-state nanopores with gel media to slow DNA translocations. *Electrophoresis* **36**, 1759–1767 (2015).

128. Noakes, M. T. *et al.* Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nat. Biotechnol.* **37**, 651–656 (2019).

129. Asandei, A. *et al.* Electroosmotic Trap Against the Electrophoretic Force Near a Protein Nanopore Reveals Peptide Dynamics During Capture and Translocation. *ACS Appl. Mater. Interfaces* **8**, 13166–13179 (2016).

130. Hsu, W.-L. & Daiguji, H. Manipulation of Protein Translocation through Nanopores by Flow Field Control and Application to Nanopore Sensors. *Anal. Chem.* **88**, 9251–9258 (2016).

131. Keyser, U. F. *et al.* Direct force measurements on DNA in a solid-state nanopore. *Nat. Phys.* **2**, 473–477 (2006).

132. Tabard-Cossa, V. *et al.* Single-Molecule Bonds Characterized by Solid-State Nanopore Force Spectroscopy. *ACS Nano* **3**, 3009–3014 (2009).

133. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574 (2006).

134. Lu, Y., Weers, B. & Stellwagen, N. C. DNA persistence length revisited. *Biopolymers* **61**, 261–275 (2001).

135. Tinland, B., Pluen, A., Sturm, J. & Weill, G. Persistence Length of Single-Stranded DNA. *Macromolecules* 5763–5765 (1997).

136. Adamcik, J., Klinov, D. V., Witz, G., Sekatskii, S. K. & Dietler, G. Observation of single-stranded DNA on mica and highly oriented pyrolytic graphite by atomic force microscopy. *FEBS Lett.* **580**, 5671–5675 (2006).

137. Manning, G. S. Counterion condensation theory constructed from different models. *Physica A* **231**, 236–253 (1996).

138. Schellman, J. A. & Stigter, D. Electrical double layer, zeta potential, and electrophoretic charge of double-stranded DNA. *Biopolymers* **16**, 1415–1434 (1977).

139. Colchero, A., Pastor, I. & Ritort, F. Determining the Effective DNA Charge Density from Nanopore Translocation Dynamics. *Nano Lett.* **26**, 3365–3372 (2026).

140. Rubinstein, M. & Colby, R. H. *Polymer Physics*. (Oxford University Press, 2003).

141. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).

142. Hong, F., Zhang, F., Liu, Y. & Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **117**, 12584–12640 (2017).

143. Zhang, Q. *et al.* DNA Origami as an *In Vivo* Drug Delivery Vehicle for Cancer Therapy. *ACS Nano* **8**, 6633–6643 (2014).

144. Fan, S. *et al.* Spatiotemporal Control of Molecular Cascade Reactions by a Reconfigurable DNA Origami Domino Array. *Angew. Chem. Int. Ed.* **61**, e202116324 (2022).

145. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science (1979).* **338**, 1177–1183 (2012).

146. Diagne, C. T., Brun, C., Gasparutto, D., Baillin, X. & Tiron, R. DNA Origami Mask for Sub-Ten-Nanometer Lithography. *ACS Nano* **10**, 6458–6463 (2016).

147. Ye, J., Helmi, S., Teske, J. & Seidel, R. Fabrication of Metal Nanostructures with Programmable Length and Patterns Using a Modular DNA Platform. *Nano Lett.* **19**, 2707–2714 (2019).

148. Wagenbauer, K. F. *et al.* How We Make DNA Origami. *ChemBioChem* **18**, 1873–1885 (2017).

149. Dey, A. *et al.* Effect of Temperature Ramp in Rapid Folding of 3D DNA Origami Structures. *Small Struct.* **7**, e202500682 (2026).

150. Takiguchi, L. P. *et al.* NanoNERF: A nanoscale NERF blaster replica made of DNA. *bioRxiv* Preprint at https://doi.org/10.1101/2023.10.02.560388 (2023).

151. Wu, H. *et al.* Expanding DNA Origami Design Freedom with De Novo Synthesized Scaffolds. *J. Am. Chem. Soc.* **146**, 16076–16084 (2024).

152. Bush, J. *et al.* Synthesis of DNA Origami Scaffolds: Current and Emerging Strategies. *Molecules* **25**, 3386 (2020).

153. Glaser, M. *et al.* The Art of Designing DNA Nanostructures with CAD Software. *Molecules* **26**, 2287 (2021).

154. Dey, S. *et al.* DNA origami. *Nature Reviews Methods Primers* **1**, 13 (2021).

155. Majikes, J. M. & Liddle, J. A. DNA Origami Design: A How-To Tutorial. *J. Res. Natl. Inst. Stand. Technol.* **126**, 126001 (2021).

156. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001–5006 (2009).

157. Carlsen, A. T., Zahid, O. K., Ruzicka, J. A., Taylor, E. W. & Hall, A. R. Selective detection and quantification of modified DNA with solid-state nanopores. *Nano Lett.* **14**, 5488–5492 (2014).

158. Sandler, S. E. *et al.* Sensing the DNA-mismatch tolerance of catalytically inactive Cas9 via barcoded DNA nanostructures in solid-state nanopores. *Nat. Biomed. Eng.* **8**, 325–334 (2023).

159. Voorspoels, A. *et al.* Design Principles of DNA-Barcodes for Nanopore-FET Readout, Based on Molecular Dynamics and TCAD Simulations. *J. Phys. Chem. A* **128**, 3926–3933 (2024).

160. Bell, N. A. W. & Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat. Nanotechnol.* **11**, 645–651 (2016).

161. Bell, N. A. W. & Keyser, U. F. Specific Protein Detection Using Designed DNA Carriers and Nanopores. *J. Am. Chem. Soc.* **137**, 2035–2041 (2015).

162. Morin, T. J. *et al.* Nanopore-Based Target Sequence Detection. *PLoS One* **11**, e0154426 (2016).

163. Kowalczyk, S. W., Hall, A. R. & Dekker, C. Detection of local protein structures along DNA using solid-state nanopores. *Nano Lett.* **10**, 324–328 (2010).

164. Kong, J., Bell, N. A. W. & Keyser, U. F. Quantifying Nanomolar Protein Concentrations Using Designed DNA Carriers and Solid-State Nanopores. *Nano Lett.* **16**, 3557–3562 (2016).

165. Patiño-Guillén, G. *et al.* Single-molecule RNA sizing enables quantitative analysis of alternative transcription termination. *Nat. Commun.* **15**, 1699 (2024).

166. Li, Y. *et al.* DNA Nanoarray for Multiplexed RNA Detection with Single-Molecule Readout. *Advanced Science* **13**, e75319 (2026).

167. Wang, Z. *et al.* Digital DNA Nanostructures for Ultrasensitive Multiplexed Heavy Metal Sensing via Nanopore Platforms. *Adv. Funct. Mater.* **36**, e75756 (2026).

168. Bošković, F., Ohmann, A., Keyser, U. F. & Chen, K. DNA Structural Barcode Copying and Random Access. *Small Struct.* **2**, 2000144 (2021).

169. Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **20**, 456–466 (2019).

170. Wang, S., Mao, X., Wang, F., Zuo, X. & Fan, C. Data Storage Using DNA. *Advanced Materials* **36**, 2307499 (2024).

171. Doricchi, A. *et al.* Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* **16**, 17552–17571 (2022).

172. Organick, L. *et al.* Probing the physical limits of reliable DNA data retrieval. *Nat. Commun.* **11**, 616 (2020).

173. Church, G. M., Gao, Y. & Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science (1979).* **337**, 1628–1628 (2012).

174. Thomas, C. A. *et al.* Sequencing a DNA analog composed of artificial bases. *Nat. Commun.* **16**, 7240 (2025).

175. Hoshika, S. *et al.* Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science (1979).* **363**, 884–887 (2019).

176. Sensintaffar, A., Wei, Y., Ou, L., Du, D. & Li, B. Advancing Archival Data Storage: The Promises and Challenges of DNA Storage System. *ACM Transactions on Storage* **21**, 1–34 (2025).

177. Song, L.-F., Deng, Z.-H., Gong, Z.-Y., Li, L.-L. & Li, B.-Z. Large-Scale de novo Oligonucleotide Synthesis for Whole-Genome Synthesis and Data Storage: Challenges and Opportunities. *Front. Bioeng. Biotechnol.* **9**, 689797 (2021).

178. Sadremomtaz, A. *et al.* Digital data storage on DNA tape using CRISPR base editors. *Nat. Commun.* **14**, 6472 (2023).

179. Wisna, G. B. M. *et al.* High-speed 3D DNA PAINT and unsupervised clustering for unlocking 3D DNA origami cryptography. *Nat. Commun.* **16**, 11514 (2025).

180. Shin, J.-S. & Pierce, N. A. Rewritable Memory by Controllable Nanopatterning of DNA. *Nano Lett.* **4**, 905–909 (2004).

181. Ameixa, J., Sala, L., Kocišek, J. & Bald, I. Radiation and DNA Origami Nanotechnology: Probing Structural Integrity at the Nanoscale. *ChemPhysChem* **26**, e202400863 (2025).

182. Sala, L. *et al.* Folding DNA into origami nanostructures enhances resistance to ionizing radiation. *Nanoscale* **13**, 11197–11203 (2021).

183. Ren, R. *et al.* Multiplexed detection of viral antigen and RNA using nanopore sensing and encoded molecular probes. *Nat. Commun.* **14**, 7362 (2023).

184. Assad, O. N., Di Fiori, N., Squires, A. H. & Meller, A. Two Color DNA Barcode Detection in Photoluminescence Suppressed Silicon Nitride Nanopores. *Nano Lett.* **15**, 745–752 (2015).

185. Cai, S. *et al.* Single-molecule amplification-free multiplexed detection of circulating microRNA cancer biomarkers from serum. *Nat. Commun.* **12**, 3515 (2021).

186. McNally, B. *et al.* Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett.* **10**, 2237–2244 (2010).

187. Xia, P. *et al.* DNA Helix Bundle-Encoded Multi-Bit Information Readout by Sapphire-Supported Nanopores. *Adv. Funct. Mater.* **36**, e23998 (2026).

188. Zhu, J., Ermann, N., Chen, K. & Keyser, U. F. Image Encoding Using Multi-Level DNA Barcodes with Nanopore Readout. *Small* **17**, 2100711 (2021).

189. Singer, A., Rapireddy, S., Ly, D. H. & Meller, A. Electronic Barcoding of a Viral Gene at the Single-Molecule Level. *Nano Lett.* **12**, 1722–1728 (2012).

190. Singer, A. *et al.* Nanopore based sequence specific detection of duplex DNA for genomic profiling. *Nano Lett.* **10**, 738–742 (2010).

191. Weckman, N. E. *et al.* Multiplexed DNA Identification Using Site Specific dCas9 Barcodes and Nanopore Sensing. *ACS Sens.* **4**, 2065–2072 (2019).

192. Zhao, L., Wang, J., Wu, L.-S. & Zhao, X. Solid-State Nanopore Readout of Programmable DNA and Peptide Nanostructures for Scalable Digital Data Storage. *Biosensors (Basel).* **15**, 287 (2025).

193. Zhu, Z. *et al.* Adaption of a Solid-State Nanopore to Homogeneous DNA Organization Verification and Label-Free Molecular Analysis without Covalent Modification. *Anal. Chem.* **90**, 814–820 (2018).

194. Chandrasekaran, A. R., Levchenko, O., Patel, D. S., Macisaac, M. & Halvorsen, K. Addressable configurations of DNA nanostructures for rewritable memory. *Nucleic Acids Res.* **45**, 11459–11465 (2017).

195. Beamish, E., Tabard-Cossa, V. & Godin, M. Programmable DNA Nanoswitch Sensing with Solid-State Nanopores. *ACS Sens.* **4**, 2458–2464 (2019).

196. Liu, S. C., Li, Q., Ying, Y. L. & Long, Y. T. Detection of structured single-strand DNA via solid-state nanopore. *Electrophoresis* **40**, 2112–2116 (2019).

197. Zhu, J. *et al.* Multiplexed Nanopore-Based Nucleic Acid Sensing and Bacterial Identification Using DNA Dumbbell Nanoswitches. *J. Am. Chem. Soc.* **145**, 12115–12123 (2023).

198. Lunt, B. M. How Long Is Long-Term Data Storage? *Archiving Conference* (2011).

199. Lunt, B. M., Kemp, D., Linford, M. R. & Chiang, W. How Long Is Long-Term? - An Update. *Archiving Conference* **21**, 56–59 (2024).

200. Matange, K., Tuck, J. M. & Keung, A. J. DNA stability: a central design consideration for DNA data storage systems. *Nat. Commun.* **12**, 1358 (2021).

201. Gevers, J. *et al.* Impact of Translocation Dynamics and Bandwidth on the Readout of DNA Structural Barcodes with Membrane-Based Solid-State Nanopores. *ACS Nano* **19**, 6058–6068 (2025).

202. Li, Y., Sandler, S. E., Keyser, U. F. & Zhu, J. DNA Volume, Topology, and Flexibility Dictate Nanopore Current Signals. *Nano Lett.* **23**, 7054–7061 (2023).

203. Ermann, N. *et al.* Promoting single-file DNA translocations through nanopores using electro-osmotic flow. *J. Chem. Phys.* **149**, 163311 (2018).

204. Roelen, Z., Briggs, K. & Tabard-Cossa, V. Analysis of Nanopore Data: Classification Strategies for an Unbiased Curation of Single-Molecule Events from DNA Nanostructures. *ACS Sens.* **8**, 2809–2823 (2023).

205. Chen, P. *et al.* Probing single DNA molecule transport using fabricated nanopores. *Nano Lett.* **4**, 2293–2298 (2004).

206. Plesa, C., Cornelissen, L., Tuijtel, M. W. & Dekker, C. Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore. *Nanotechnology* **24**, 475101 (2013).

207. Seth, S. & Bhattacharya, A. How capture affects polymer translocation in a solitary nanopore. *J. Chem. Phys.* **244902**, 244902 (2022).

208. Li, J., Gershow, M., Stein, D., Brandin, E. & Golovchenko, J. A. DNA molecules and configurations in a solid-state nanopore microscope. *Nat. Mater.* **2**, 611–615 (2003).

209. Sim, A. Y. L., Lipfert, J., Herschlag, D. & Doniach, S. Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**, 1–5 (2012).

210. Roth, E., Glick Azaria, A., Girshevitz, O., Bitler, A. & Garini, Y. Measuring the Conformation and Persistence Length of Single-Stranded DNA Using a DNA Origami Structure. *Nano Lett.* **18**, 6703–6709 (2018).

211. Forstater, J. H. *et al.* MOSAIC: A Modular Single Molecule Analysis Interface for Decoding Multi-state Nanopore Data. *Anal. Chem.* **88**, 11900–11907 (2016).

212. Kowalczyk, S. W., Tuijtel, M. W., Donkers, S. P. & Dekker, C. Unraveling single-stranded DNA in a solid-state nanopore. *Nano Lett.* **10**, 1414–1420 (2010).

213. Tree, D. R., Muralidhar, A., Doyle, P. S. & Dorfman, K. D. Is DNA a good model polymer? *Macromolecules* **46**, 8369–8382 (2013).

214. Rowghanian, P. & Grosberg, A. Y. Force-driven polymer translocation through a nanopore: An old problem revisited. *Journal of Physical Chemistry B* **115**, 14127–14135 (2011).

215. Sakaue, T. Nonequilibrium dynamics of polymer translocation and straightening. *Phys. Rev. E* **76**, 021803 (2007).

216. Fologea, D., Brandin, E., Uplinger, J., Branton, D. & Li, J. DNA conformation and base number simultaneously determined in a nanopore. *Electrophoresis* **28**, 3186–3192 (2007).

217. Leitao, S. M. *et al.* Spatially multiplexed single-molecule translocations through a nanopore at controlled speeds. *Nat. Nanotechnol.* **18**, 1078–1084 (2023).

218. Charron, M., Roelen, Z., Wadhwa, D. & Tabard-Cossa, V. Predicting Resistive Pulse Signatures in Nanopores by Accurately Modeling Access Regions. *ArXiv* https://doi.org/arXiv:2411.05589v2 (2025)

219. Ying, Y.-L. *et al.* Nanopore-based technologies beyond DNA sequencing. *Nat. Nanotechnol.* **17**, 1136–1146 (2022).

220. He, L. *et al.* Digital immunoassay for biomarker concentration quantification using solid-state nanopores. *Nat. Commun.* **12**, 5348 (2021).

221. Zheng, F. *et al.* Comparative Benchmarking of Glass and Silicon Nitride Nanopores for Single-Molecule Detection. *ACS Nano* **20**, 14256–14267 (2026).

222. Briggs, K. Solid-State Nanopores: Fabrication, Application, and Analysis. (PhD Thesis, University of Ottawa, 2018).

223. Balan, A. *et al.* Improving signal-to-noise performance for DNA translocation in solid-state nanopores at MHz bandwidths. *Nano Lett.* **14**, 7215–7220 (2014).

224. Shekar, S. *et al.* Measurement of DNA translocation dynamics in a solid-state nanopore at 100-ns temporal resolution. *Nano Lett.* **16**, 4483–4489 (2016).

225. Lin, C.-Y. *et al.* Ultrafast Polymer Dynamics through a Nanopore. *Nano Lett.* **22**, 8719–8727 (2022).

226. Yuan, Z., Liu, Y., Dai, M., Yi, X. & Wang, C. Controlling DNA Translocation Through Solid-state Nanopores. *Nanoscale Res. Lett.* **15**, 80 (2020).

227. Fologea, D., Uplinger, J., Thomas, B., McNabb, D. S. & Li, J. Slowing DNA translocation in a solid-state nanopore. *Nano Lett.* **5**, 1734–1737 (2005).

228. Wang, R., Liu, W. & Seeman, N. C. Prototyping Nanorod Control: A DNA Double Helix Sheathed within a DNA Six-Helix Bundle. *Chem. Biol.* **16**, 862–867 (2009).

229. González González, A. C., Kerrouri, N., Wadhwa, D., Tabard-Cossa, V. & Briggs, K. Poriscope: A Configurable Pipeline for Nanopore Data Analysis. *J. Open Res. Softw.* **14**, 32 (2026).

230. Chau, C., Mohanan, G., Macaulay, I., Actis, P. & Wälti, C. Automated Purification of DNA Origami with SPRI Beads. *Small* **20**, 2308776 (2024).

231. Shaw, A., Benson, E. & Högberg, B. Purification of Functionalized DNA Origami Nanostructures. *ACS Nano* **9**, 4968–4975 (2015).

232. Lin, C., Perrault, S. D., Kwak, M., Graf, F. & Shih, W. M. Purification of DNA-origami nanostructures by rate-zonal centrifugation. *Nucleic Acids Res.* **41**, e40 (2013).

233. Fuhrman, J. A., Comeau, D. E., Hagstrom, A. & Chan, A. M. Extraction from Natural Planktonic Microorganisms of DNA Suitable for Molecular Biological Studies. *Appl. Environ. Microbiol.* **54**, 1426–1429 (1988).

234. Stortchevoi, A., Kamelamela, N. & Levine, S. S. SPRI beads-based size selection in the range of 2–10kb. *Journal of Biomolecular Techniques* **31**, 7–10 (2020).

235. Sowersby, D. S. & Lewis, L. K. SURE gel electrophoresis: A method for improved detection and purification of dilute nucleic acid samples. *Anal. Biochem.* **684**, 115373 (2024).

236. Humphreys, G. O., Willshaw, G. A. & Anderson, E. S. A Simple Method for the Preparation of Large Quantities of Pure Plasmid DNA. *Biochim. Biophys. Acta* **383**, 457–463 (1975).

237. Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H. Facile and scalable preparation of pure and dense DNA origami solutions. *Angew. Chem. Int. Ed.* **53**, 12735–12740 (2014).

238. Thamm, S., Slesiona, N., Dathe, A., Csáki, A. & Fritzsche, W. AFM-Based Probing of the Flexibility and Surface Attachment of Immobilized DNA Origami. *Langmuir* **34**, 15093–15098 (2018).

239. Xin, Y., Zargariantabrizi, A. A., Grundmeier, G. & Keller, A. Magnesium-Free Immobilization of DNA Origami Nanostructures at Mica Surfaces for Atomic Force Microscopy. *Molecules* **26**, 4798 (2021).

240. Kube, M. *et al.* Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution. *Nat. Commun.* **11**, 6229 (2020).

241. Ni, H. *et al.* Direct visualization of floppy two-dimensional DNA origami using cryogenic electron microscopy. *iScience* **25**, 104373 (2022).

242. Kuzuya, A., Sakai, Y., Yamazaki, T., Xu, Y. & Komiyama, M. Nanomechanical DNA origami 'single-molecule beacons' directly imaged by atomic force microscopy. *Nat. Commun.* **2**, 449 (2011).

243. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

244. He, L. *et al.* Digital Immunoassays for Sensitive Quantification of Blood Biomarkers Using Solid-State Nanopores. *ACS Nano* **20**, 10308–10319 (2026).