

**Genotyping by Nanopore Force Spectroscopy:
Method Development and Evaluation for Clinical
Diagnostics**

by

Matthew John Wiggin

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of

Doctor of Philosophy

in

The Faculty of Graduate Studies
(Biochemistry & Molecular Biology)

The University of British Columbia
(Vancouver)

November 2009

© Matthew John Wiggin, 2009

Abstract

Clinical diagnostic genotyping has the potential to predict an individual's response to a prescribed drug, and could thus dramatically improve drug efficacy and reduce adverse drug interactions. However, widespread implementation of clinical diagnostic genotyping is currently prevented by a lack of fast, simple clinical genotyping platforms. This thesis describes the development of a new genotyping technique based on nanopore force spectroscopy (NFS) which may fulfill this need, and serves as a feasibility study for further development towards a commercial instrument.

The thesis begins by describing NFS, which is a novel, general technique used to detect bio-molecules and characterize their physical interactions with one another. NFS is applied to base-calling by forming a duplex between an engineered single-stranded DNA probe and a DNA sample, and then measuring the dissociation rate under an applied force. The dissociation rate is shown to be extremely sensitive to duplex sequence homology: tests using purified synthetic DNA fourteen bases long demonstrate that even a single base mismatch can increase the dissociation rate over 100-fold.

This high specificity, combined with the sensitivity of nanopore detection, allows a base-call to be made from as few as 100 single molecule dissociation events involving the target. Based on these results, it is estimated that with further development, NFS genotyping could be possible from purified, unlabeled genomic DNA in less than 1 hour, without requiring PCR amplification. These characteristics would make NFS extremely attractive as a clinical diagnostic genotyping technology.

Further development is still required to produce an instrument capable of testing genomic DNA. However, based on the success of tests so far, this thesis concludes that further development of such an instrument is clearly warranted, especially given its potential impact on human health.

Table of Contents

Abstract.....	ii
Table of Contents.....	iv
List of Tables.....	viii
List of Figures	ix
List of Terms	xii
Acknowledgements	xvi
Dedication.....	xix
Chapter 1 - Introduction	1
1.1 The Importance of Genetic Testing to Medicine	1
1.2 The Need for New Clinical Diagnostic Genotyping Technology	4
Chapter 2 - Background	8
2.1 Genotyping by Allele Specific Hybridization.....	8
2.2 Genotyping by Allele Specific Enzymatic Reactions	12
2.3 Genotyping by DNA Sequencing.....	17
2.4 Common Technical Shortcomings of Existing Platforms.....	18
Chapter 3 - Principles of Nanopore Force Spectroscopy	21
3.1 Nanopore Detection & Analysis	21
3.1.1 Principles of Nanopore Detection	22
3.1.2 Classes of Nanopores	24
3.1.3 Seminal Nanopore Detection Experiments	26
3.2 Force Spectroscopy	29
3.2.1 Principles of Force Spectroscopy	30
3.2.2 Force Spectroscopy Methods.....	33

3.2.3 Seminal Force Spectroscopy Experiments.....	36
3.2.4 Limitations of Conventional Force Spectroscopy	
Methods as Diagnostic Tools	37
3.3 Nanopore Force Spectroscopy	39
3.4 Basic Concepts of Genotyping by Nanopore Force Spectroscopy	40
3.4.1 Proposed genotyping technology.....	43
3.4.2 From the Lab to the Real World – Challenges of Nanopore Force Spectroscopy Application	46
Chapter 4 - Probe Escape	49
4.1 Methods.....	51
4.2 Results & Data Analysis.....	54
4.2.1 Probe Escape Data Fitting.....	56
4.2.2 Probe Escape Kinetic Analysis	60
4.3 Discussion	68
4.3.1 Possible Causes of Energy Barrier Variation	68
4.3.2 Experimental Validation of the Characteristic Timescale as a Measure of the Energy Barrier	70
Chapter 5 - DNA Duplex Dissociation by Nanopore Force Spectroscopy	73
5.1 Methods.....	74
5.1.1 Single Pore Methods.....	76
5.1.2 Multi-Pore Methods.....	79
5.2 Results & Data Analysis.....	85
5.2.1 Nanopore Force Spectroscopy Data Fitting	86
5.2.2 Nanopore Force Spectroscopy Kinetic Analysis	89
5.3 Discussion	93
Chapter 6 Base-Calling by Nanopore Force Spectroscopy.....	100
6.1 Methods.....	100

6.2 Results & Data Analysis.....	101
6.3 Discussion	112
Chapter 7 - Future Studies.....	113
Chapter 8 - Conclusion.....	118
Bibliography	121
Appendix A - Characteristics of Common Force Spectroscopy Techniques	136
Appendix B Materials & Methods.....	137
B.1 Instrumentation	137
B.2 Software	140
B.2.1 Data Acquisition Software	140
B.2.2 Raw Data Analysis Software.....	141
B.3 Reagents	142
B.4 α -HL Nanopore Formation Methods	143
Appendix C - Measurement of DNA Length in α-HL	148
C.1 Methods & Raw Data Analysis	148
C.2 Results.....	149
C.3 Discussion	152
Appendix D - Calibration of Electrostatic Force	154
Appendix E Calculation of τ^* Error Bars	159
Appendix F Performance Evaluations	161
F.1 Sensitivity.....	161
F.1.1 Sampling Error	162
F.1.2 Probe-Target Hybridization Kinetics.....	164
F.2 Repeatability.....	168

F.3 Unrelated Sequence Rejection.....173

List of Tables

Table 5-1 Probe and target sequences used in nanopore force spectroscopy genotyping studies..	75
Table A-1 Characteristics of common force spectroscopy techniques ..	136

List of Figures

Figure 2-1 Genotyping by equilibrium hybridization	10
Figure 2-2 Genotyping by real-time PCR	14
Figure 2-3 Genotyping by structure specific cleavage.....	16
Figure 3-1 Nanopore detection of individual DNA molecules	23
Figure 3-2 Crystal structure of the α -hemolysin	25
Figure 3-3 Scanning electron micrograph of SiN _x solid state nanopores.....	26
Figure 3-4 Energy barrier perturbation by force spectroscopy	32
Figure 3-5 Atomic force microscopy	34
Figure 3-6 Magnetic tweezer force microscopy	35
Figure 3-7 Optical tweezer force microscopy	36
Figure 3-8 Genotyping by nanopore force spectroscopy	42
Figure 3-9 Proposed genotyping platform	45
Figure 4-1 A single molecule probe escape event	52
Figure 4-2 Event time distribution for a typical probe escape experiment	57
Figure 4-3 Calculated energy barrier probability density function, expressed as a function of τ , for a typical probe escape dataset.....	64
Figure 4-4 Probe escape characteristic timescales vs. applied potential	72

Figure 5-1 An unsuccessful single molecule nanopore force spectroscopy trial.....78

Figure 5-2 A typical multi-pore force spectroscopy trial82

Figure 5-3 Comparison of single pore and multi-pore force spectroscopy results87

Figure 5-4 Event time distribution for a typical single molecule force spectroscopy experiment88

Figure 5-5 Calculated energy barrier probability density functions, expressed as a function of τ , for a typical single molecule force spectroscopy dataset90

Figure 5-6 Single nanopore FS duplex dissociation characteristic timescale vs. applied potential for the Synth probe with two different targets.....94

Figure 5-7 Multi-pore force spectroscopy duplex dissociation characteristic timescale vs. applied potential for the Synth probe with six different targets97

Figure 5-8 Dissociation timescale extrapolated to zero force vs. melting energy for nanopore force spectroscopy of DNA duplexes98

Figure 6-1 Event time distributions for NFS base calling102

Figure 6-2 Proportion of events attributed to the perfect complement allele for three possible genotypes109

Figure 6-3 Characteristic timescale τ^*_{other} from equation (6-1) determined by NFS genotyping.110

Figure 6-4 Relative concentration of alleles calculated for three possible genotypes, as calculated by (6-12)111

Figure B-1 The nanopore apparatus, including the a-HL pore, PTFE cell, patch clamp amplifier, electrodes, and Faraday cage.....137

Figure C-1 Target hybridization probability as a function of probe length.....150

Figure C-2 Comparison of force spectroscopy event time distributions for probes containing 22-nt and 30-nt linkers at -60mV applied dissociation potential.....151

Figure D-1 Barrier height sensitivity to applied potential as a function of molecule length.....157

Figure F-1 Nanopore force spectroscopy characteristic timescale estimation error as a function of dataset size163

Figure F-2 Probability of probe-target hybridization vs. hold time166

Figure F-3 Repeatability of nanopore force spectroscopy kinetics between different trials170

Figure F-4 Variability in fraction of target capture in heterozygote genotyping experiments172

Figure F-5 Influence of unrelated DNA on NFS dissociation timescales.174

List of Terms

α	The stretch parameter of the stretched exponential decay function
α -HL	α -hemolysin, a pore-forming toxin protein produced by the bacterium <i>Staphylococcus aureus</i> , and the nanopore used in all experiments in this dissertation.
A_i	The weighting factor for a single characteristic timescale τ_i of the multi-exponential decay function
β	The control parameter of the Becquerel decay function
dsDNA	double stranded DNA
DPPC	1,2-diphytanoyl- <i>sn</i> -glycero-3-phosphocholine, the lipid used to form the bilayer in nanopore experiments
E	Electric field strength
e	The elementary charge, 1.602×10^{-19} Coulombs
f	Force
ϕ_B	The time parameter of the Becquerel decay function
ϕ_S	The time parameter of the stretched exponential decay function
FS	Force spectroscopy
FWHM	Full width at half maximum, used to characterize the spread of a distribution. It is calculated by determining the width of the distribution at the two points where the height is half of the maximum.

G_b	Gibbs' free energy
ΔG_b	Gibbs' free energy barrier to a reaction, including any contributions from applied force. Note that that throughout this thesis, ΔG_b is expressed in units of $k_B T$. At 300 K, $1 k_B T \approx 0.6$ kcal/M.
ΔG_b^0	The free energy barrier to a reaction, in the absence of applied force
i_b	The ionic current flowing through a single pore blocked by a probe molecule
i_o	The ionic current flowing through a single open pore
$I(t)$	The total ionic current, flowing through all pores, in a multi-pore experiment
k	The rate of a reaction
k_B	The Boltzmann constant, $1.38 \cdot 10^{-23}$ J/K
NFS	Nanopore force spectroscopy
$N_b(t)$	Number of blocked pores, as a function of time, in a multi-pore experiment
N_{events}	The total number of events in a force spectroscopy dataset
N_P	The total number of pores in a multi-pore experiment
$N_o(t)$	Number of open pores, as a function of time, in a multi-pore experiment
nt	Nucleotide
PCR	Polymerase chain reaction
PDF	Probability density function

$P_{survival}$	Survival probability, which measures the likelihood that a probe will still be present in the pore a time t after the dissociation (or escape) potential is reached
PTFE	Polytetrafluoroethylene
SNP	Single nucleotide polymorphism, a single base within the genome which varies between different individuals
ssDNA	Single stranded DNA
T	Temperature, measured in Kelvin
t	Time
τ	The characteristic timescale governing an ideal reaction with constant kinetics
τ^*	A single characteristic timescale representing the peak of the energy barrier PDF for a non-ideal reaction
τ_D	The diffusive relaxation time, i.e. the timescale at which a molecule or complex trapped in an energy well attempts to cross the energy barrier
V	The applied potential, measured in Volts. Defined as positive with respect to the trans-side of the pore, such that a probe molecule, captured from the <i>cis</i> -side is being pulled into the pore.
$w(\Delta G_b)$	The energy barrier PDF governing an energy barrier crossing process
$W(\tau)$	The energy barrier PDF governing an energy barrier crossing process, expressed as a function of τ
Δx_b	The distance, measured in units of distance, between an energy well, and the peak of the energy barrier in a structural transition reaction

z	The effective charge per nucleotide of a DNA molecule, used to account for the reduction in the effective force applied to a DNA molecule inside the pore by charge screening and electro-osmotic effects
-----	---

Acknowledgements

Many people contributed to the work in this thesis. I would like to thank my co-supervisors, Dr. Ivan Sadowski and especially, Dr. Andre Marziali, who supervised the day-to-day operations of my research. Andre was a constant source of guidance, insight, and support, and provided the optimism necessary to keep me going whenever I got stuck. As a scientist, I must concede that a better supervisor could, in theory, exist; however, I have no idea what one would improve in designing such a supervisor. Thanks also to my committee members, Dr. Lawrence McIntosh and Dr. Steve Plotkin, for helpful comments and feedback.

Thanks to the past and present members of the nanopore group in the Marziali lab, including Dr. Jon Nakane, who taught me to make nanopores, and performed the initial force spectroscopy studies that grew into this thesis. Thanks to Dhruti Trivedi and Dr. Jason Dwyer, for good times and shared ideas, and Nahid Jetha, collaborator and conference room-mate extraordinaire. Undergraduate students Chris Feehan, Sybil Drissler, Erika Chin, George Sterling, and Carolina Tropini all contributed to various nanopore studies over the years. Carolina deserves special thanks for initial development of the multi-pore force spectroscopy technique described in this thesis, and for

producing probe data as fast as I could analyze it. Special thanks also to Dr. Vincent Tabard-Cossa, for the most satisfying work partnership I have ever had.

Many thanks to all the other members of the Marziali lab, past and present including: Robin Coope, Peter Eugster, Kurtis Guggenheimer, Dylan Gunn, Miti Isbasecu, Joel Pel, Jaryn Perkins, Jared Slobodan, and Jason Thompson, for making work a fun place to be, and helping out whenever needed. Thanks in particular to our lab manager, David Broemeling, who made sure I kept getting paid, even when I forgot.

Thanks to everyone involved in Physics 253, including all of the teaching and support staff, and especially the students, whose enthusiasm, creativity, and ability to cause minor explosions using simple electrical components, was a constant source of inspiration. I am probably the only biochemistry graduate student in history to ever TA a robotics class, and for that, I am extremely lucky.

Lastly, thanks to my family. Thanks to my parents, John and Nancy Wiggin, for your unwavering love and support. You two instilled in me an enduring sense of curiosity by taking me out to the pond to collect tiny ecosystems in jars, and doing your best to answer the question “but why?” every one of the million-or-so times I asked it. Thanks to my sister, Laura Cain, and her family, Ryan, Molly, and

Anna, for being you guys. Vancouver isn't nearly as fun since you left, and I miss you all very much. Et, enfin, grâce à Karina Houle, la meilleure copine jamais, qui m'a gardé nourris, habillé, et plus ou moins sain d'esprit pendant l'écriture de cette thèse. Je t'aime plus que je sais comment le dire avec des mots.

Dedication

For Jack, who is missed.

Chapter 1 - Introduction

1.1 The Importance of Genetic Testing to Medicine

A relatively large amount of the information contained within the approximately 3 billion base pairs (bp) of the human genome varies between individuals, including thousands of regions with variable copy numbers [1, 2], and an estimated 10 million single nucleotide polymorphisms (SNPs) [3]. Subsets of these genetic differences affect an individual's predisposition to diseases, disorders, and clinical response to medication [4-7]. Rapid clinical tests for human genetic variation thus have the potential to dramatically improve health, by providing clinicians with information which could be used in diagnosis, and importantly, in tailoring drug treatment by prescribing an appropriate drug at the correct dose for that individual patient.

Studies linking genetics to healthcare have been revolutionized since the completion of the human genome project [8, 9], which provided the reference sequence required to identify millions of single nucleotide polymorphisms in thousands of individuals. Studies such as the international HapMap project [10], the Cancer Genome Atlas [11], and an increasing number of other genome wide association studies [5], are generating catalogues of common genetic variation in

the human population, many of which could be applied to improve clinical diagnostics and drug treatment.

In particular, pharmacogenetic and pharmacogenomic studies, which link polymorphisms to drug response, have found that 80-90% of variability in drug metabolism rates, and up to 60% of variability in pharmacodynamic response¹ are determined by a patient's genotype [4]. Therefore, clinical diagnostic genotyping could allow prescriptions to be tailored to an individual person, ensuring patients received the appropriate drug at a proper dose. This could revolutionize medical practice by increasing drug efficacy, which currently ranges from 25-80% for most major drugs [12], and reducing drug toxicity, which affects 2 million people per year in the U.S. including 100 000 deaths [7]. In addition, such tests could increase the number of drugs available, by allowing previously failed drugs to be released for a limited market along with an appropriate test [6, 13].

Though pharmacogenetics is still in its infancy, a few clinical diagnostic tests are now available [14-16] for a restricted set of disorders where the association between outcome and genetics is very

¹ Pharmacodynamic response is the effect of a drug on the body, and includes drug toxicity.

strong, the cost of treatment is high, and the effects of improper prescription are severe. Examples include tests for HER2 over-expression, which determines breast cancer response to the chemotherapeutic Herceptin [16-19], and genetic tests for activity in two genes of the cytochrome P450 family (CYP2D6 and CYP2C19) [15, 20], which are frequently implicated in drug failure, toxicity, and death [20].

Even for these examples, however, clinical genotyping is not yet commonplace, and tests do not exist for many cases where the potential benefit is enormous. One such case is sepsis / severe inflammatory response syndrome (SIRS), which is the 10th leading cause of death in the USA, with a mortality rate of 20-50%, and annual health care costs of \$5-10 billion in the USA [21]. Pharmacogenetic markers for sepsis are documented in the literature [22-26]; however no commercially available clinical genotyping technology can deliver results fast enough to make testing practical. Sepsis mortality increases by 7% with every hour of delay before appropriate treatment [21], while nearly all existing tests take at least six hours to perform [27].

According to clinicians, the lack of appropriate technology is a major reason clinical diagnostic genotyping has yet to be widely adopted [28]. Development of a technology to meet this urgent clinical need is the focus of this thesis.

1.2 The Need for New Clinical Diagnostic Genotyping

Technology

The utility of a clinical diagnostic genotyping platform rests on its ability to deliver sufficient genetic information about a patient in time to guide decisions made by medical practitioners. This goal defines a set of requirements including: accuracy, reproducibility, cost, capacity for multiplexing, speed, and the complexity of the process, all of which are briefly described here.

Accuracy and reproducibility are paramount, since life and death decisions are made based on test results. Though an ideal test would produce accurate results 100% of the time, accuracy and reproducibility of 99% are currently considered sufficiently robust for use in clinical diagnostics, with the proviso that a test should contain internal positive and negative controls in order to identify erroneous calls in as many cases as possible [29].

Ideally, the test should be inexpensive; however, according to clinical experts, the cost of the instrument and consumables are only moderately important since the final cost of a test is dominated by legal, marketing and clinical approval expenses [30, 31]. In addition, even a relatively expensive test can dramatically reduce the total cost of patient care by allowing targeted use of expensive drugs or reducing hospitalization times.

The test format and multiplexing capacity should be designed such that a single test, which can be performed for individual patients on an ad-hoc basis, gives medical practitioners sufficient information to make a treatment decision. Multiplexing requirements are moderate, with most available clinical pharmacogenetic tests examining 10-100 alleles in parallel [14-16, 32].

The speed of a test is critical, since many applications, such as sepsis / SIRS, require rapid clinical intervention. Even in less acute cases, a rapid test is still desirable, since it would allow patients to be diagnosed and treated in a single hospital visit. A reasonable goal is delivery of a genotype from a blood sample in one hour.

In order to meet this goal, the test must include a minimum number of steps, and it must be sufficiently simple that it can be performed by non-specialists in a hospital clinic. Complex, difficult tests must be outsourced to specialized testing facilities, which is costly, and represents a major barrier to widespread adoption of clinical genotyping [28], especially since it increases the time required to obtain results by hours to days.

As we shall see in the following chapter, no currently available clinical genotyping platform satisfies all of these requirements [33], with the most common problems being test complexity and speed.

These issues have led to low adoption rates among clinicians [28], motivating the need for new clinical diagnostic genotyping methods.

This thesis describes a novel genotyping method based on nanopore force spectroscopy [34, 35], which may address assay complexity and speed issues, in addition to fulfilling the other requirements listed above. NFS infers a sample's genotype by measuring dissociation kinetics between sample DNA and an engineered DNA probe, complementary to the sequence of interest.

The chapters that follow develop the basic technology for an NFS genotyping instrument, including experimental methods and data processing algorithms, and serve as a feasibility study for future development into a commercial instrument. The results of these studies demonstrate that base calls of nucleotide sequences can be made with single-nucleotide specificity from as few as 100 detected molecules.

It should be stressed that substantial further development is required before NFS can be applied to clinical genotyping. However, calculations in the latter portion of the thesis suggest that, with further development, NFS may be able to deliver results in less than 1 hour

from unamplified genomic DNA². NFS genotyping would also be extremely simple, requiring only that a clinician extract DNA from a blood sample, and load it into a disposable nanopore cartridge, which would test for multiple alleles simultaneously, once inserted into the instrument. Collectively, this work suggests that rapid clinical genotyping by nanopore force spectroscopy is feasible, and consequently, an exciting option for clinical application that warrants future development.

² See section 6.3 for calculations

Chapter 2 - Background

This chapter examines the strengths and weaknesses of currently available genotyping platforms, focusing primarily on those technologies that have been approved for use in hospital clinics by the FDA⁴. The first three sections describe a number of different genotyping techniques, and evaluate them from the standpoint of the end user. The final section considers common shortcomings and their sources in the detection technologies employed by these platforms. This analysis reveals the major challenges that must be surmounted in order to produce a widely useful clinical diagnostic genotyping instrument.

2.1 Genotyping by Allele Specific Hybridization

Most clinical diagnostic genotyping assays, including microarrays, are based on allele specific hybridization, which exploits the effect of sequence homology on the equilibrium binding constant of a DNA duplex. Sample DNA is hybridized to a set of engineered probes,

⁴ Many technologies used in research laboratories are inappropriate for clinical tests, due to issues with test complexity or multiplexing, and are therefore omitted from this review. Readers interested in a more inclusive list are referred to a number of excellent reviews on the subject: [5, 27, 36-38].

resulting in duplexes that are stable only if the probe and target DNA sequences are perfectly complementary (see Figure 2-1).

A variety of instruments based on this principle are described below, all of which follow the same general process. DNA is extracted and purified from a blood or tissue sample, and subjected to PCR to selectively amplify the loci of interest⁵. PCR products are labeled with either fluorescent tags, or with gold nano-beads, and then exposed to the surface-bound DNA probes for hybridization. Hybridization, which is performed at elevated temperature to maximize sequence specific differences in duplex stability, is allowed to come to equilibrium over several hours. Finally, the sample is imaged, indicating which alleles are present in the sample by the accumulation of signal at locations on the array.

⁵ One exception to this is the Verigene system, discussed below.

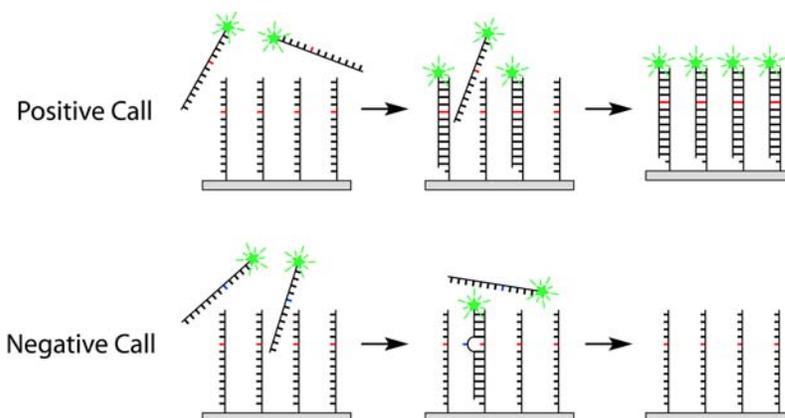


Figure 2-1 Genotyping by equilibrium hybridization. This detection scheme relies on the sensitivity of the equilibrium binding constant to mismatches in the DNA duplex. Labeled DNA samples are washed over a surface decorated with DNA probes complementary to the sequence of interest. Top - Perfectly complementary duplexes are thermodynamically stable, resulting in accumulation of signal (e.g. fluorescence). Bottom - Mismatched duplexes are unstable, resulting in little accumulation of signal.

The first allele specific hybridization-based system we will consider is the Roche Amplichip P450 test [15]⁶, which uses a fluorescence microarray to analyze 31 mutations and polymorphisms in the CYP2D6 and CYP2C19 genes of the cytochrome P450 family. Sample DNA is purified, amplified in a first PCR step, and labeled in a second PCR step. Fluorescent products are then hybridized to DNA probes on

⁶ Note that this was the first pharmacogenetic test to be cleared by the FDA [39].

a microarray slide, and detected by imaging. All told, the test requires between one and two full working days to complete [15], and the method is quite complex, requiring operator intervention for all steps, some of which require great care, such as the two PCR steps.

A related technology is the Luminex platform [40, 41], a fluorescence hybridization assay which multiplexes up to 100 samples per well in a standard 96-well plate. Multiplexing is accomplished by attaching the DNA probes to fluorescently labeled beads, each of which is internally labeled with a fluorescent dye corresponding to the associated probe [41, 42]. Sample DNA is purified and subjected to PCR for amplification and fluorescent labeling, and hybridized to the bead-ligated DNA probes. The beads are then passed through a flow-cytometry apparatus, which uses two separate lasers, one to identify the bead, and the other to quantify the amount of bound target DNA.

The scale of this setup is attractive to clinical labs, allowing the user to run anywhere from 1 to 10^4 samples, at a nearly linear increase in cost. However, the assay takes approximately 6 hours to complete, and suffers from similar problems of complexity as the Roche Amplichip P450 test. [27].

A third allele specific hybridization platform is the Nanosphere Verigene system [14, 43], which carries the advantage that it does not require PCR, due to two significant differences from other allele-specific hybridization platforms. The first difference is that labels are

not covalently attached to the sample DNA. Instead, the sample is sandwiched between two probes, the first of which is covalently attached to the surface of a slide. The second probe, which is labeled with a gold nano-particle, is mixed into solution with the sample. When the sample hybridizes to both probes, nano-particles accumulate at a specific spot on the array.

The second difference is the detection scheme, which is sufficiently sensitive that alleles can be detected directly from genomic DNA without PCR amplification. Following hybridization, the slide is stained with silver, which precipitates onto the nano-particles, forming large beads that can be detected by light scattering when illuminated with a laser.

The Verigene assay thus consists of DNA extraction, followed by hybridization and imaging, which is both simple and fast. The total time for the assay is approximately 2 hours, and nearly all steps are automated, making the Verigene a strong competing technology for clinical diagnostic genotyping.

2.2 Genotyping by Allele Specific Enzymatic Reactions

The next general technique we will consider uses enzymes to generate fluorescent products if and only if a specific allele is present.

The two methods considered here are real-time PCR and structure specific cutting by cleavase.

In real-time PCR assays, a molecular beacon probe containing a fluorescent dye-quencher pair is digested during the reaction if the sequence of interest is present (see Figure 2-2). Thus, as the reaction proceeds, dyes are released from their quenchers, generating a fluorescent signal. Though real-time PCR is extremely sensitive and specific, it is complicated, prone to contamination, labour-intensive, and difficult to multiplex [37].

One technology which overcomes many of these problems is IQuum's Lab-in-a-Tube platform (LIAT™) [44]. It uses a disposable, segmented tube pre-loaded with all the reagents and fluorescent DNA primers required for the assay. The tube is loaded with a blood sample, and inserted into a machine which automatically performs all steps from cell lysis to real-time PCR, giving results in approximately 1 hour [44]. Its speed and ease of use make it attractive for clinical genotyping; however, the multiplexing capability of this test is greatly limited, since each allele requires a separate test, making tests of ~100 polymorphisms for a patient unfeasible.

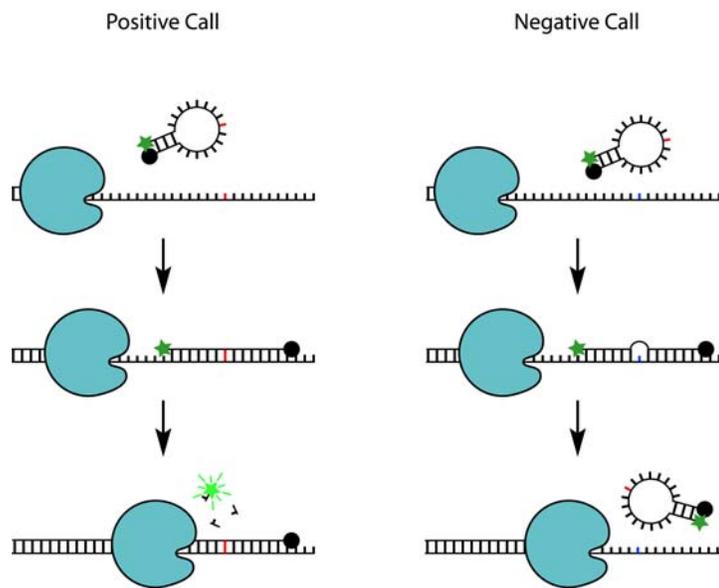


Figure 2-2 Genotyping by real-time PCR. A molecular beacon probe, coupled at its ends to a fluorescent dye / quencher pair, is designed to be complementary to the sequence of interest. During PCR amplification, the probe anneals to the DNA sample. Left - Exonuclease activity of DNA polymerase (blue) digests perfectly complementary probes, releasing the dye from the quencher, causing an increase in fluorescence. Right - DNA polymerase displaces mismatched probes without digesting them, resulting in no increase in fluorescence.

The other enzymatic method, employed by Third Wave Technologies' Invader assay [45], identifies alleles by structure-specific DNA cutting with the enzyme cleavase. The reaction involves two synthetic DNA sequences - a probe, and an "Invader oligo," which anneal adjacent to one another on the DNA template forming a Y-shaped complex, as shown in Figure 2-3. If the probe and the Invader

oligo are both capable of base-pairing with the template nucleotide at the boundary between them, they form an “overlapping structure” recognized by cleavase, which then cuts the probe.

In the Invader assay, two such reactions are performed simultaneously in a single tube. In the first reaction, sample DNA forms the template, with the single nucleotide polymorphism designed to lie at the boundary between the probe and Invader oligo, such that the reaction proceeds only if the desired allele is present. The cleaved product forms the template for the second reaction, in which the probe contains a fluorescent dye-quencher pair, generating a signal when it is cleaved. Note that each template can participate in repeated reactions, similar to PCR, resulting in up to 10 million-fold signal amplification.

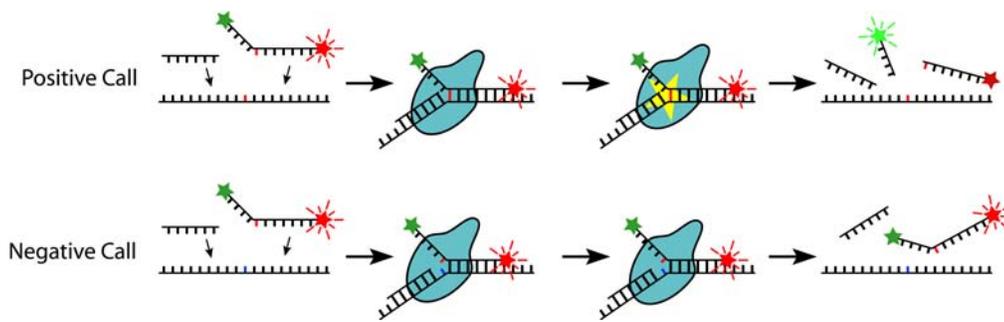


Figure 2-3 Genotyping by structure specific cleavage (simplified depiction). A DNA probe (dye-quencher tagged) and an Invader oligo anneal adjacent to one another on the sample DNA strand, forming a "Y" shaped structure. Top - If the probe and the Invader oligo are both capable of base-pairing with the template nucleotide at the boundary, a cleavase enzyme (blue) cuts the overhanging probe. Bottom - If there is no overlap, cleavase does not cut the probe, and no signal accumulates. Note that for simplicity, the figure shows only one of the two rounds of reactions involved in the Invader assay.

The Invader assay thus proceeds as follows. DNA is extracted from the sample, and PCR is performed, if required. The DNA is then added to a 96-well plate pre-loaded with reagents, incubated to allow the cleavase reaction to proceed. Finally, the plate is imaged to detect the fluorescent products. The assay is straightforward, sample handling is minimal, and in some cases, no PCR is required; however, the test takes at least 6 hours to perform from start to finish [45].

2.3 Genotyping by DNA Sequencing

The final technology we will consider is DNA sequencing. Rapid developments in sequencing technology will soon make it possible to sequence an individual's entire genome and maintain it in a databank⁷, giving health care practitioners instant access to genetic information. However, for reasons discussed below, this will remain less common than clinical diagnostic genotyping in the short term, and will likely never replace it completely.

In the short term, further development of both sequencing technology and the legal frameworks to deal with privacy of genetic information⁸ are required before personal genome sequencing can be commercialized. In the long term, widespread adoption will likely take decades, requiring other technologies to test patients whose genomes have not yet been sequenced. In addition, ad-hoc genotyping will be always required for disorders caused by somatic mutations, such as cancer.

⁷ Note that a simpler version of this, large scale personal single nucleotide polymorphism genotyping is already available from companies such as 23andMe [46], which analyzes 550 000 SNPs using microarrays, maintaining them in a databank.

⁸ The first American legislation protecting genetic privacy was adopted less than a year ago [47], and none has yet been drafted in Canada.

Targeted DNA sequencing of specific loci could be used for ad-hoc genotyping, however, this is a relatively complex, expensive, and slow technique. Even third generation sequencing machines, which achieve speed and cost improvements through massive parallelization, would not be appropriate for clinical genotyping.

Therefore, rapid clinical genotyping tools, which require a smaller leap both technologically and socially, will continue to fill an important need for the foreseeable future, as clinicians increasingly begin to rely on genetic data to make treatment decisions.

2.4 Common Technical Shortcomings of Existing Platforms

The clinical diagnostic genotyping technologies listed above show a trend of lower speed and higher complexity than what is desired. Most available tests take at least 6 hours to deliver results, and are so complex that they must be out-sourced to specialty labs. As we shall see below, the root of these problems lies in the detection mechanisms employed in these instruments, and low specificity and sensitivity of the detectors.

All of the assays listed above, with the exception of the Verigene test, are based on fluorescent detection of products, which is relatively

insensitive when used at the scale required for clinical genotyping⁹. Therefore, lengthy, and sometimes complex reactions are required in order for a detectable quantity of product to accumulate. For example, the Invader assay requires a reaction time of 5 hours [45], and the Amplichip P450 and Luminex tests require 1-3 hours for the hybridization step alone [15, 41].

Many of these assays also require PCR, both to label the products for fluorescent detection, and to greatly increase the product concentration to the point where it is detectable. PCR also compensates for low detector specificity, since prior to amplification, the detector faces a 1 in 10^9 signal-to-noise ratio, in the form of dilution of the nucleotide of interest against the entire complement of genomic sequence, which contains many partially homologous sequences which can interact with the probe.

PCR is highly undesirable, however, since it typically takes 1-3 hours to complete. It also requires a trained operator, as it is highly prone to contamination or failure [36] due to low target quantities and contamination from environmental DNA or blood components [48].

⁹ Note that confocal microscopy has the potential to detect individual fluorophores; however, this technique is inappropriate for clinical genotyping, since it requires extreme care, and cannot be readily parallelized to detect many different alleles simultaneously.

According to clinical experts, one major problem is the need to accurately pipette minute reagent volumes [30]. The need for expert operators is a major reason that clinical genotyping is not routinely performed in hospitals, and samples are instead sent out to specialized labs for testing [30].

Clinical genotyping would therefore benefit greatly from an instrument that did not require PCR amplification, sample labeling, or lengthy reaction times [36, 49]. In order to achieve this goal, a novel detection mechanism is required, capable of genotyping directly from unlabeled genomic DNA. Results presented in later chapters indicate that nanopore force spectroscopy has the potential to attain this goal.

NFS is capable of distinguishing a given target sequence from sequences with even a single base mismatch, by observing the dissociation of less than 100 duplexes, meaning that it may be possible to genotype DNA directly, without amplification or labeling. It achieves this by employing a novel detection method, based on measuring dissociation kinetics between the DNA sample and an engineered DNA probe under applied force. The basic concepts of nanopore force spectroscopy are the topic of the next chapter.

Chapter 3 - Principles of Nanopore Force Spectroscopy

Nanopore force spectroscopy was first developed and described in 2004 by Nakane, Wiggin, and Marziali [34]. NFS is a general technique for studying kinetics and thermodynamics of a wide variety of structural transitions at the molecular scale, including DNA duplex dissociation for the purpose of genotyping. Because NFS is a new technique, this chapter presents a detailed description of the basic underlying concepts, beginning with the techniques upon which it is based - nanopore detection and single molecule force spectroscopy. It then describes their integration to form nanopore force spectroscopy, and application of NFS to genotyping. The chapter concludes with a description of a proposed clinical diagnostic genotyping instrument.

3.1 Nanopore Detection & Analysis

A nanopore is perhaps the simplest conceivable nanotechnology - a nanometer-scale hole in an insulating membrane. Despite this simplicity, nanopores are employed in a surprisingly diverse set of single molecule techniques. Nanopores form excellent single molecule detectors due to their size, which is similar to the molecules they detect, facilitating molecular manipulation and detection. In addition, nanopores carry the benefit that both manipulation and detection can

be performed electrically, obviating the need for complex optics, lasers, or sample labeling.

3.1.1 Principles of Nanopore Detection

The basic concept of nanopore detection is illustrated in Figure 3-1, which shows a simple experiment in which individual DNA molecules are observed as they pass through the pore. The nanopore is placed in an insulating membrane separating two chambers filled with ionic solution, and an electrostatic potential is applied across the membrane, causing an ionic current to flow through the pore. When a charged polymer such as DNA is driven into the nanopore by the electrostatic potential, it is detected by an accompanying change in the ionic current.

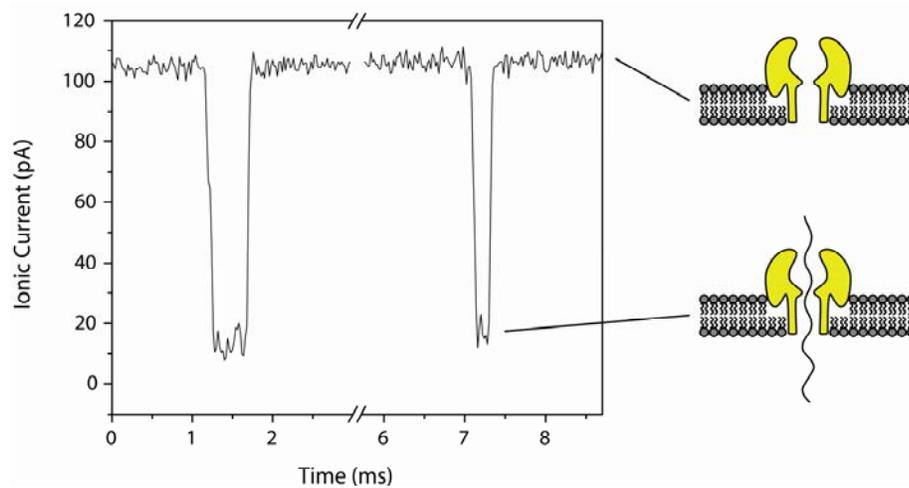


Figure 3-1 Nanopore detection of individual DNA molecules. DNA is driven into an α -hemolysin nanopore by a 115 mV applied potential, causing a measurable reduction in the ionic current through the pore.

The ionic current is affected by polymer entry through a combination of steric exclusion, which reduces the ionic current by decreasing the effective diameter of the pore, and ionic interactions, including co-ion repulsion and counter-ion attraction by the charged polymer [50, 51]. Co-ion repulsion tends to decrease the ionic current, while counter-ion attraction tends to increase it. The balance of these three effects depends on the salt concentration. In 1M KCl, the most commonly used ionic solution, steric exclusion and co-ion repulsion strongly dominate, such that ssDNA reduces the ionic current by $\sim 80\%$ when it enters the protein pore α -hemolysin (α -HL) [52-58] (see Figure 3-1).

3.1.2 Classes of Nanopores

Nanopores used in detection and manipulation of biomolecules can be divided into two main classes: protein pores and solid state pores. Protein pores used for nanopore detection are typically pore-forming toxins, the most prevalent of which is α -hemolysin from the bacterium *Staphylococcus aureus* [59, 60] (see Figure 3-2). This pore is nearly ideal for single molecule detection, due to its minimum internal diameter of 1.4 nm [60], which is well matched to the *ca.* 1.3 nm diameter of single stranded oligonucleotides. In addition, α -HL exhibits an unprecedented signal-to-noise ratio of *ca.* 50 for single molecule detection, which arises from a combination of the large ionic current reduction that accompanies oligonucleotide entry [52-58], and the pore's quiet electrical properties, including almost no propensity to gate¹⁰ [59].

¹⁰ Gating is reversible closure of a protein pore. Depending on the pore type, it can occur in response to applied potential, binding of small molecules or ions, or it can be spontaneous.

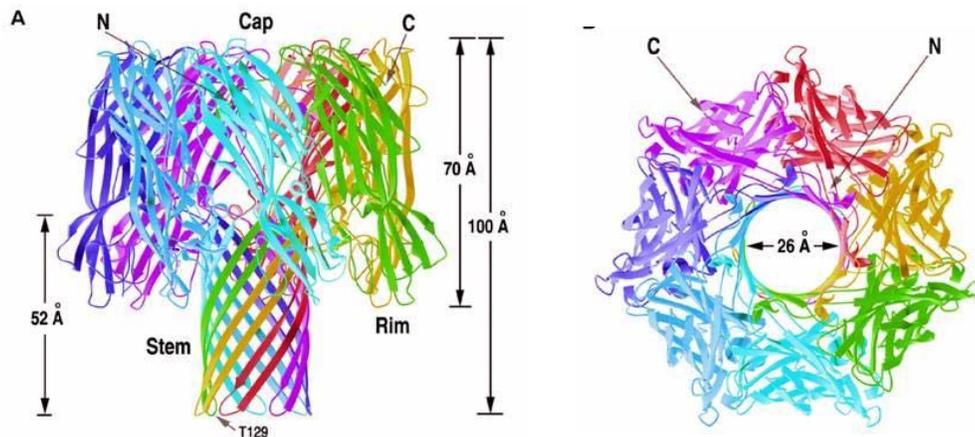


Figure 3-2 Crystal structure of the α -hemolysin, a heptameric pore-forming toxin produced by *Staphylococcus aureus* (PDB code: 7ahl]. The pore contains two major domains: the cap domain at the top, and the stem domain, comprising the β -barrel at the bottom. The cap encloses a large vestibule, 2.6 - 4.6 nm in diameter. Between the cap and the stem lies the limiting aperture (i.e. smallest constriction), which is 1.4 nm in diameter. The stem is approximately 2.6 nm diameter. Image source: [60].

Solid state nanopores are made by drilling a thin membrane of a suitable material, such as silicon nitride, with a high-energy electron beam and/or focused ion beam [61-63]. Solid state pores offer some advantages over α -HL including tunable size, ease of incorporation into instruments, stability over periods up to months [64], robustness under harsh conditions (*e.g.* low pH or high applied potentials), and potential for control over their geometric and surface properties. However, they also carry some significant disadvantages, including low signal-to-noise ratios, since they exhibit much higher noise than α -hemolysin [64]. In addition, DNA and protein tend to stick to the pore

surface, causing these solid state pores to block irreversibly. For these reasons, all experiments included in this dissertation use the more reliable α -HL pore.

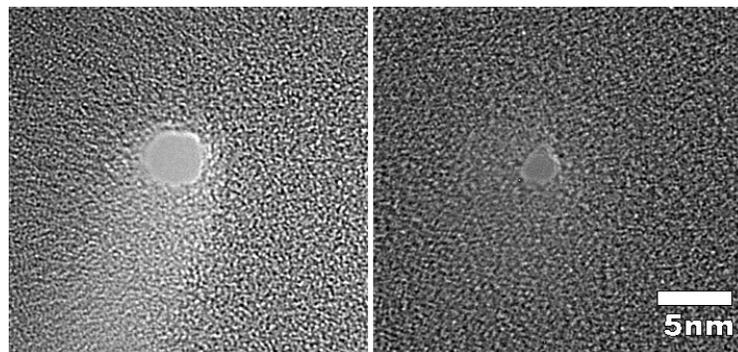


Figure 3-3 Scanning electron micrograph of SiN_x solid state nanopores. Pore diameters are estimated to be 4 nm (left), and 2 nm (right). Images and pores produced by Dhruvi Trivedi.

3.1.3 Seminal Nanopore Detection Experiments

The first demonstration of single molecule detection using nanopores was achieved in 1996 by Kasianowicz *et al.*, who observed individual single stranded RNA molecules as they passed through α -hemolysin [52]. A variety of studies detecting both DNA and RNA soon followed [53-58], and similar work has since been repeated using solid state pores [61, 65-67]. Later studies demonstrated that oligonucleotide properties such as length [52, 57, 68], orientation (*i.e.* 3'-first vs. 5'-first) [69] and most interestingly, sequence composition

[53, 56, 70] of block co-polymers, could be determined from current signatures during translocation, leading to speculation that nanopores could be used for rapid single-molecule DNA sequencing.

However, nanopore sequencing is not as straightforward as first hoped. The sequence-specific ionic current signals observed in early RNA homopolymer experiments were largely due to secondary structure¹¹, with only minor contributions from the bases themselves [53]. In addition, the current-sensitive region of the α -HL can accommodate 10-12 bases [57], each of which affects the ionic current. Each base resides in this region for only 1-10 μ s [56], meaning that the ionic current signal from an individual base is too low to be distinguished from shot noise¹² [71]. In short, more sophisticated methods will be required if nanopore sequencing is to be realized.

¹¹ Secondary structures of poly-U, poly-C, and poly-A RNA homo-polymers are: random coil, tight helices, and loose helices, respectively. Each of these structures causes a distinct ionic current blockage when the molecule passes through the pore [53].

¹² Shot noise arises from stochastic arrival times of the individual ions that make up the ionic current signal. In 1M KCl, a few hundred ions pass through the pore for each base that translocates, meaning shot noise is extremely high, relative to the difference in signal for different bases [71].

A number of variations on nanopore sequencing are being explored which may circumvent the above problems. The first approach plans to slow translocation with a processive enzyme such as DNA polymerase [72, 73]. This would improve the ionic current signal-to-noise ratio by slowing shot noise by increasing the integration time for each individual base. A second approach, pursued by Oxford Nanopore Technologies [74], plans to use a pore-coupled exonuclease to cleave nucleotides from the end of a DNA polymer. The cleaved nucleotides would then be driven, in sequence, through a chemically modified α -HL pore, with each nucleotide producing a distinctive ionic current signature [74, 75]. Finally, it has been proposed that electrodes embedded in solid state nanopores could specifically detect individual bases in a translocating DNA molecule, either by tunneling currents through the bases [76-78] or base-specific perturbations in capacitance [79]. These last schemes offer the fastest potential sequencing speeds, but also the greatest difficulties, particularly with respect to pore fabrication and the need to control DNA molecules with Angstrom resolution to ensure proper interaction with pore-embedded electronics during translocation [80].

It remains unclear whether nanopore sequencing will ever be realized at a scale that is competitive with existing sequencing technology. However, a number of other nanopore-based techniques have already demonstrated single-nucleotide specific DNA detection

[34, 35, 75, 81-84], including determination of DNA duplex sequence homology by nanopore force spectroscopy [34, 35, 83, 84].

In nanopore force spectroscopy, an electric field is used to apply a force which promotes duplex dissociation. The dissociation rate as a function of applied force is measured, yielding considerable information about the dissociation energy barrier. In the case of a genotyping assay, this can be used to determine sequence homology of a DNA duplex comprised of an engineered probe and an unknown sample with single nucleotide resolution [34, 35]. Before discussing this technique in more detail, we will provide a general overview of force spectroscopy.

3.2 Force Spectroscopy

Force spectroscopy techniques are uniquely suited to studying a wide range of biologically important, molecular scale structural transitions. Processes that can be studied by FS range from the action of molecular motors in processes such as cellular motility [85, 86], trafficking of proteins or organelles within the cell [87-89], and DNA or RNA synthesis [90, 91], to single molecule processes including protein folding [92-97], ligand-receptor interactions [98-100], and DNA duplex dissociation [34, 35, 83, 84, 101-104].

In a force spectroscopy experiment, the rate of a structural transition, such as DNA duplex dissociation, is enhanced (or occasionally, impeded) through application of a mechanical force. Simultaneously, an associated displacement, such as the separation of the two DNA strands in the duplex, is measured, allowing the rate of the structural transition to be observed. By examining the relationship between the rate and the applied force, it is possible to determine details of the energy barriers, forces, and mechanisms involved in the underlying process [105-110]. The physical principles of force spectroscopy analysis are described in the following section.

3.2.1 Principles of Force Spectroscopy

The theory of force spectroscopy stems from the physics governing reaction rates [105, 109-111]. Figure 3-4 shows the energy landscape for a theoretical reaction, in which a molecule or complex begins in an energy well (e.g., DNA duplex associated) and crosses over an energy barrier to a final state (e.g., DNA dissociated). If the barrier is sufficiently high¹³, the rate determining step is attainment of sufficient

¹³For barrier heights similar to the scale of random energy fluctuations, i.e. $< 5 k_B T \approx 3$ kcal/mol at 300 K, the rate limiting step becomes the diffusion time, and neither the Arrhenius relation, nor Kramers' rule (discussed below) apply [105, 109].

thermal energy, through random fluctuations, to overcome the potential energy barrier.

Because the reaction is driven by thermal energy fluctuations, reaction times for individual molecules are stochastic. Reaction kinetics are therefore characterized by the rate or characteristic timescale¹⁴, measured by observing many single molecule events, and analyzing the distribution of event times. The characteristic timescale, τ , is related to the energy barrier height through the Arrhenius equation [110]:

$$\tau = \tau_D e^{\Delta G_b / k_B T} \quad (3-1)$$

where τ_D , the diffusive relaxation time, is the timescale over which the molecule or complex attempts to cross the energy barrier, typically in the range of 10^{-9} – 10^{-10} s [112]. ΔG_b is the height of the free energy barrier. Note that throughout this thesis, ΔG_b is expressed in units of $k_B T$, where k_B is the Boltzmann constant, and T is the absolute temperature. $1 k_B T$ (0.6 kcal/M at 300 K), is the scale of thermal energy fluctuations, making it a convenient unit for measuring single molecule processes.

¹⁴ Note that the characteristic timescale is the inverse of the rate, i.e. $\tau = 1/k$.

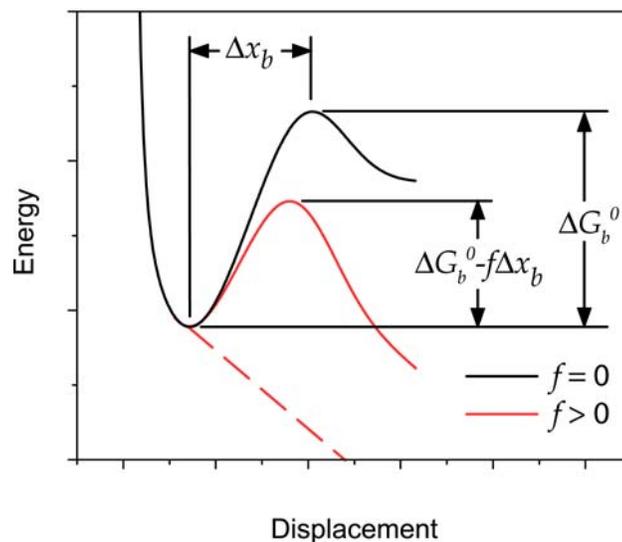


Figure 3-4 Energy barrier perturbation by force spectroscopy. Application of an assisting mechanical force (dashed line) reduces the energy barrier height, increasing the rate at which molecules or complexes cross the energy barrier. Note that the reaction coordinate (x-axis) is measured in space, e.g. the separation of two molecules in a complex.

In a force spectroscopy experiment, τ is modified by subjecting the molecule or complex to a mechanical force which assists, or occasionally, hinders the structural transition. As the reaction proceeds, this force does work, modifying the height of the energy barrier, which can be accounted for by including a force-dependent term in the Arrhenius rate law, as first proposed by Kramers [110-112]:

$$\tau = \tau_D e^{\Delta G_b^0 - f\Delta x_b} \quad (3-2)$$

where ΔG_b° is the free energy barrier at zero applied force, f is the component of the applied force (assumed here to be constant) acting along the direction of the reaction coordinate, and Δx_b is the distance between the ground state and the transition state. The term $f\Delta x_b$ therefore represents the work done by the applied force. Equation (3-2) is known as Kramers' rule, and forms the main theoretical underpinning of force spectroscopy.

3.2.2 Force Spectroscopy Methods

Forces in force spectroscopy experiments are typically applied by suspending the complex to be studied between a flexible mechanical probe and a surface, which are then pulled apart¹⁵. Structural transitions are determined from the separation between the probe and the surface, and forces are measured by monitoring the displacement of the probe from its equilibrium position.

This general method is employed by a wide array of force spectroscopy techniques. The three most common techniques - atomic force microscopy, magnetic tweezers, and optical tweezers, are described in this section. The details of these techniques are described

¹⁵ One important exception is nanopore force spectroscopy, discussed below.

in Figure 3-5, Figure 3-6, and Figure 3-7, and their characteristics are summarized in Table A-1.

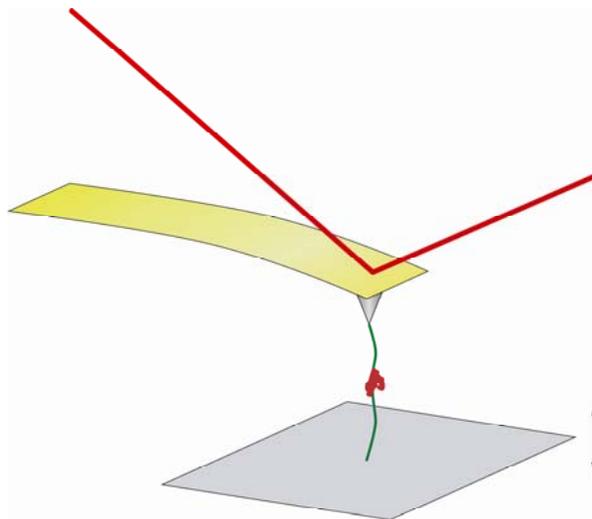


Figure 3-5 Atomic force microscopy [93, 99, 113-115]. The complex under study is suspended between the sharp tip at the end of a flexible gold cantilever and a glass slide attached to a piezoelectric manipulator. Force is applied by moving the glass slide away from the tip, and measured from the position of a reflected laser beam, which moves as the cantilever bends. Displacement is measured from the separation between the tip and the surface. Atomic force microscopy is popular because it is simple to set up, and instruments are commercially available [116, 117]. However, it is restricted to relatively large forces, and usually suffers from poor control over surface attachment [113].

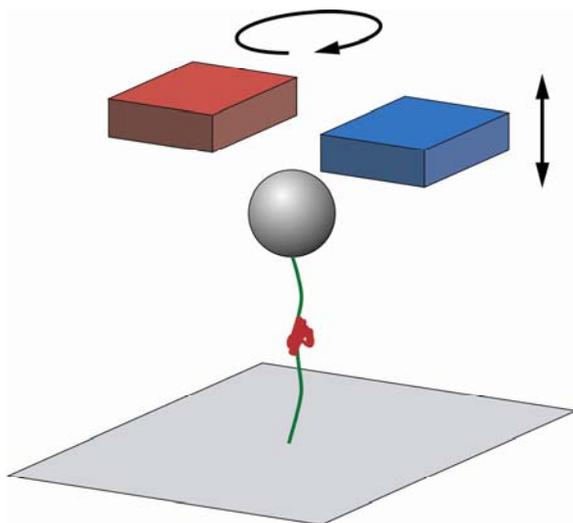


Figure 3-6 Magnetic tweezer force microscopy [113, 118-121]. The molecule is tethered between a glass slide and a paramagnetic bead, to which forces are applied by a magnet. Force is measured from diffusion of the bead in the plane parallel to the surface, and displacement is measured by the height of the bead above the surface. Magnetic tweezers offer some unique advantages, such as an ability to apply torque by rotating the magnet [120, 121], ease of constant-force application, and potential for parallelization; however, it is extremely difficult to apply time-varying forces by this technique.

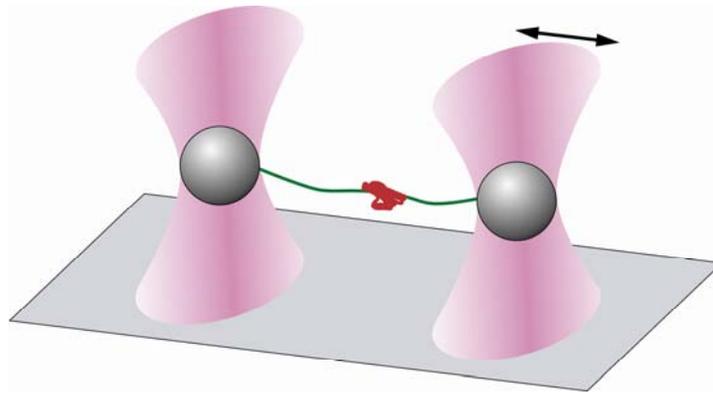


Figure 3-7 Optical tweezer force microscopy [90, 92, 113, 122-124]. A laser, focused to a diffraction-limited spot exerts radiation pressure on a small dielectric bead (or beads). Force is measured from the displacement of one bead from its equilibrium location (in the center of the laser focus), and displacement is measured from the separation of the beads. Optical tweezers offer tremendous control; molecules may be manipulated in three dimensions [113], and instruments using four or more beads can be used to study complex multi-molecular assemblies [125]. However, high intensity laser light can cause photo-damage or thermal-damage, and local heating can also affect kinetics [126].

3.2.3 Seminal Force Spectroscopy Experiments

A large body of force spectroscopy literature has been published over the last two decades; this section reviews a few of the most important results.

Early studies focused on measuring mechanical properties of biomolecules, including the elastic properties of individual DNA or RNA molecules subjected to force [119, 122, 127] and torque [120, 128]. However, the true power of force spectroscopy lies in studying

dynamic processes, such as folding of RNA [106-108, 129] protein molecules [92-97]. Similar studies have examined inter-molecular recognition in DNA duplexes [34, 35, 83, 84, 101-104], DNA binding proteins [130, 131], and receptor-ligand pairs [98-100]. In certain cases, it has even been possible to map out entire energy landscapes associated with these processes [106-108].

Force spectroscopy is increasingly being applied to determine the mechanisms, step sizes, and processivities of active components of the cell. For example, force spectroscopy studies observing individual nucleotide incorporation by RNA polymerase have yielded fundamental knowledge about the mechanism by which this enzyme moves along the DNA template [91]. A variety of other processive nucleic acid enzymes such as DNA polymerase [90], exonuclease [132], DNA helicases [133, 134], and topoisomerases [121, 135-137] have been studied by force spectroscopy, as well as molecular motors such as kinesin [85, 86], myosin [87-89], and the rotary motor F_0F_1 ATPase [138].

3.2.4 Limitations of Conventional Force Spectroscopy Methods as Diagnostic Tools

In principle, force spectroscopy could also be used for highly sensitive and specific molecular detection assays, such as genotyping, by forming a complex between the target molecule and a known probe, and measuring dissociation kinetics as they are pulled apart. Since the

dissociation timescale is extremely sensitive to the energy barrier height, τ would act as an indicator for the degree of complementarity between the probe and the target. While this is theoretically possible using the methods above, it is impractical due to their limited potential for parallelization, as discussed below.

A diagnostic assay requires approximately 10^6 probes for each target molecule of interest, since multiple dissociations are required to characterize kinetics, and only a small fraction of probes will typically bind to the target¹⁶. Since each probe in an optical trap or AFM requires an individually addressed laser, these techniques could not easily be parallelized to this level. Magnetic tweezers can simultaneously observe up to 10^3 beads [139], since they do not require lasers, but this number cannot reasonably be increased much further, since it would require high-resolution imaging of an unfeasibly large field of view.

Given that up to 100 such tests are required for a clinical genotyping assay (one for each allele), conventional force spectroscopy techniques simply cannot supply the throughput required to make them practical. However, as we shall see in the next section, nanopore force spectroscopy does not suffer from the limitations of these

¹⁶ See Appendix F for calculation.

techniques, while retaining a similar level of specificity, making it extremely attractive for rapid, sensitive molecular detection assays.

3.3 Nanopore Force Spectroscopy

Nanopore force spectroscopy marries the power of force spectroscopy to the simplicity of nanopore methods. Force is applied directly to a charged molecule by the electric field, with a balancing force applied by steric constraints between the small pore and the complex, such as a DNA duplex. Displacement in NFS is determined from the ionic current flowing through the pore, which is altered when the complex undergoes a structural transition. For two-state measurements, such as duplex dissociation, the signal is the increase in ionic current that accompanies the clearing of the pore after dissociation, typically giving a signal-to-noise ratio of 50 or greater.

Since both force application and detection are entirely electrical, nanopore force spectroscopy is fundamentally different from other force spectroscopy techniques, and accordingly, has different strengths and weaknesses. Its major strengths are: ease of parallelization, since no optics are required, and ease of automation, since molecules are manipulated directly by the electric field, rather than through careful control of beads or surfaces. However, molecular manipulation with the electric field means that only highly charged molecules can be studied by NFS. In addition, small structural changes are more

difficult to observe than in conventional FS methods, since the ionic current changes only if the charge or structure of the polymer in the pore is altered. For example, exit of a DNA molecule from the pore can easily be detected, but sliding of DNA within the pore cannot.

Thus, while nanopore force spectroscopy is inferior, in most cases, for observing subtle structural transitions, it is excellent for studying two-state transitions, such as DNA dissociation. Its simplicity and amenability to parallelization make NFS an ideal tool for ultra-sensitive detection of specific biomolecules, such as a genotyping assay, which we will discuss in the next section.

3.4 Basic Concepts of Genotyping by Nanopore Force Spectroscopy

Genotyping by nanopore force spectroscopy is based on measuring forced-dissociation kinetics for a DNA duplex comprised of an engineered single-stranded probe (complementary to the sequence of interest) and a target. Two versions of NFS genotyping are described in this thesis: single pore force spectroscopy, and multi-pore force spectroscopy, which uses many pores in parallel to permit rapid data collection, as required for a clinical diagnostic test.

We will first consider single pore force spectroscopy, depicted in Figure 3-8. The figure shows a schematic diagram of a typical single

molecule dissociation trial, along with corresponding voltage and current traces. The probe is driven into the nanopore by an electric field, allowed to hybridize to sample DNA, and then the field is reversed to perform a single force spectroscopy trial, in which the dissociation time is measured. This experiment is repeated many times, generating an event time distribution. The distribution is analyzed to determine the characteristic dissociation timescale, which is a sensitive indicator of sequence homology between the probe and the target.

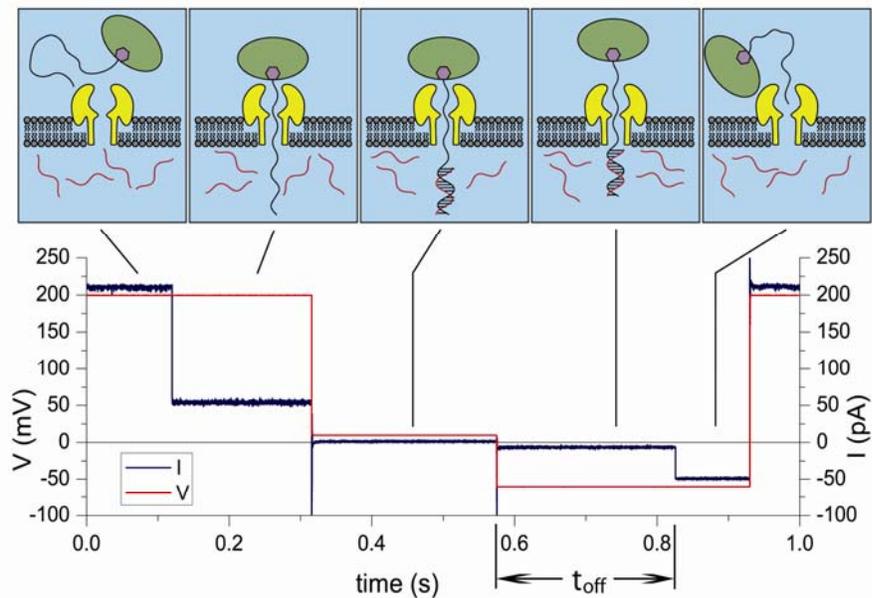


Figure 3-8 Genotyping by nanopore force spectroscopy. A schematic diagram of the process is shown, along with current and applied potential traces from a typical single molecule trial. An applied potential (red) of 200mV electrophoretically captures a probe molecule in the pore, reducing ionic current (blue) from 200pA to ~50pA. Next, a hybridization check is performed by reducing the potential to 10 mV. The duplex on the *trans*-side traps the probe in the pore, and the current remains in the blocked state. The potential is then changed to the dissociation potential (-70mV shown). After some time (t_{off}), the duplex dissociates and the probe escapes, causing a stepwise change in the ionic current. Note that changes in the applied potential are accompanied by transient current spikes, which are caused by capacitance of the lipid bilayer and electronic components of the patch clamp amplifier.

In a multi-pore assay, many pores are used in parallel, allowing multiple single-molecule dissociations to be observed in a single trial. As pores clear following dissociation, the ionic current increases from an initial value with most pores blocked, until it reaches a steady state value where all pores have cleared. The event time distribution is determined from the ionic current trace using a simple mathematical transformation, yielding results identical to a single pore experiment.

The characteristic timescale for the process is an extremely sensitive indicator of the dissociation energy, and therefore, the sequence homology between the probe and the target. Under appropriate conditions (discussed below), a single nucleotide mismatch reduces the dissociation timescale >100-fold, meaning that less than 100 single molecule dissociation events are required to detect a given target with single-base specificity. Thus, as demonstrated below, NFS may not require PCR, which would make it considerably simpler and faster than currently available genotyping assays.

3.4.1 Proposed genotyping technology

This section gives a conceptual description of a proposed commercial genotyping instrument, capable of rapidly analyzing a

DNA sample to determine a patient's genotype at up to 100 loci simultaneously by nanopore force spectroscopy¹⁷.

The heart of the instrument, shown in Figure 3-9 is a disposable chip with an array of 100 elements, each of which tests for a single allele. The chamber on the *cis*-side of each element is pre-loaded with a separate probe, specific to a single allele of interest, and contains a separate electrode for measuring the current through that element. The membrane within each array element contains 10^6 nanopores for rapid multi-pore force spectroscopy. The DNA sample is loaded into the *trans*-side sample chamber, which contacts all nanopore-membrane elements of the array.

¹⁷ Note, however, that this is not the instrument used for the proof-of-concept studies described in this dissertation.

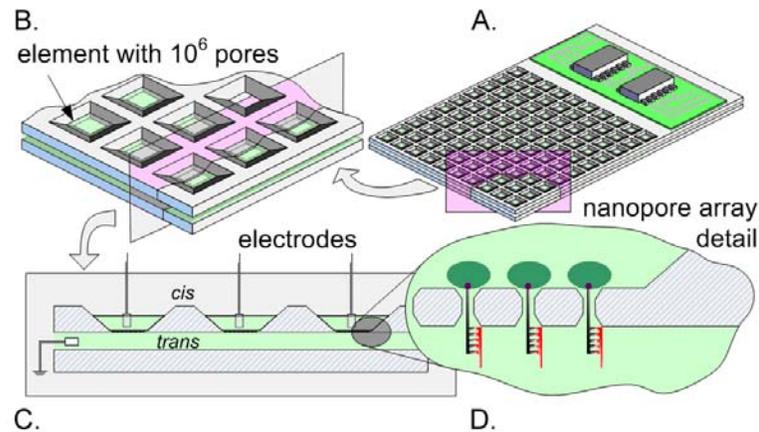


Figure 3-9 Proposed genotyping platform. (Anti-clockwise from upper right) A. Disposable nanopore chip containing 100 array elements. Each element tests for a single allele. B. Each array element contains 10^6 nanopores. C. The *trans*-chamber, which contains the sample, is common to all elements, and contains a ground electrode. The *cis*-chamber for each element contains a single allele-specific DNA probe, and an individually addressed electrode, allowing detection of a single DNA target in that element. D. Probes driven into the nanopores hybridize to the target DNA (red) on the *trans*-side of the array.

Ideally, this assay would not require PCR. The slowest step in a PCR-free assay would be probe-target hybridization, since the hybridization rate depends on target concentration. Since, at a conservative estimate, 250 dissociation events are required to make a base call, very low hybridization efficiencies (and short hybridization

times) could be tolerated¹⁸. Pores containing un-hybridized probes would be cleared prior to the FS experiment, creating a DC offset in the ionic current. However, this offset would not affect calculation of the dissociation timescale, which is based on the ionic current decay from the initial state, in which some pores are blocked, to the final state, in which all pores are open.

This proposed platform remains a work in progress. The next section introduces the major challenges that need to be addressed for development of a working prototype, including those challenges dealt with in this dissertation.

3.4.2 From the Lab to the Real World - Challenges of Nanopore Force Spectroscopy Application

This section describes the major challenges to development of a commercially viable genotyping technology based on nanopore force spectroscopy, many, but not all of which are solved through the work contained in the following chapters.

Since NFS is a novel technique, much of this work involves development and characterization of the instrument, and confirmation

¹⁸ This is discussed in detail in Appendix F.

that force spectroscopy is possible using nanopores. Chapter 4 describes probe escape studies, which are simple, well-controlled experiments used to characterize the sensor and develop the data processing algorithms for extraction of characteristic timescales from the event time distributions.

Chapter 5 develops and demonstrates nanopore force spectroscopy of DNA duplexes. These experiments are performed in both a single pore format, and a limited scale multi-pore format, which uses *ca.* 100 pores in parallel, the maximum allowable with the instruments currently available in the Marziali lab. This chapter focuses on the relationship between duplex sequence homology and the characteristic timescale for dissociation, demonstrating that NFS can distinguish between alleles of a single nucleotide polymorphism.

The next major milestone is base-calling of unknown samples, which is the focus of Chapter 6. This chapter develops the required data analysis algorithms, and demonstrates base-calling of simulated homozygous and heterozygous samples, composed of purified, synthetic DNA.

In order to move beyond the proof-of-concept studies presented in this thesis towards clinical diagnostic genotyping, an instrument containing millions of pores will be required, as described in Chapter 7 (future work). This instrument will be used to test the sensitivity of the proposed instrument by performing experiments with decreasing

concentrations of sample DNA¹⁹. Following these studies, it will be important to move towards genotyping on real-world samples, by examining the effects of contaminants, such as large amounts of genomic DNA, which will contain many sequences partially homologous to the probe.

Some preliminary studies examining hybridization kinetics and the rejection of unrelated DNA sequences are presented in Appendix E. Results of these studies suggest that extension to genotyping from genomic DNA should be feasible. However, these results should be taken with caution, as the experiments need to be expanded substantially, and performed using the next generation instrument.

Thus, considerable development is still required before clinical diagnostic genotyping by nanopore force spectroscopy can be achieved. However, as discussed in the following chapters, the fundamental principals of genotyping by nanopore force spectroscopy are sound, and development of such a commercial instrument appears to be possible, based on the results of this thesis.

¹⁹ The number of probes which hybridize to target molecules depends on the target concentration, meaning that many pores are required to form a sufficient number of duplexes for base-calling at low DNA concentrations.

Chapter 4 - Probe Escape

We now turn our attention to the experimental portion of this thesis; however, we are not yet ready to consider nanopore force spectroscopy of DNA duplexes. Instead, this chapter describes a series of control experiments, performed in the absence of target, which are used to develop data analysis methods for NFS [140].

The most important step in data analysis is determination of the characteristic timescale τ by fitting the distribution of single molecule event times. As discussed below, processes whose kinetics are determined by crossing a fixed energy barrier give rise to event times that follow a Poisson distribution. However, NFS event time distributions are substantially broader than this, requiring special data fitting and analysis methods in order to determine a characteristic timescale.

As discussed below, broadening of the event time distribution is caused by stochastic, non-covalent bonds between the probe and the α -HL pore. Since these bonds must be broken for the probe to exit the pore, they contribute to the energy barrier. Importantly, they are stochastic – *i.e.* probe-pore bond strengths vary from one single molecule event to the next. Therefore, the ensemble of events is governed by a broadened distribution of energy barriers, which leads to a broadened event time distribution.

In order to study probe-pore interactions in detail, we wish to remove the influence of the DNA duplex from the experiments. For this, we use probe escape experiments, in which no target molecule is present on the *trans*-side of the pore (see Figure 4-1). Instead, a *trans*-positive electrostatic potential is applied which tends to hold the probe in the pore, creating an energy barrier which is primarily electrostatic. This energy barrier is substantially easier to characterize and control than the barrier to DNA duplex dissociation. Therefore, probe escape experiments facilitate comparisons between theory and results, and allow the influence of probe-pore interactions on the energy barrier and event time distributions to be clearly observed.

Results from these experiments allow the stochasticity of probe-pore interactions to be measured²⁰. Through careful analysis of these results, methods permitting calculation of a characteristic timescale τ^* , from NFS data are developed. Since this analysis method has not previously been applied to force spectroscopy, experiments in this section are used to confirm that τ^* obeys Kramers' rule, and thus forms a useful metric of energy barrier height.

²⁰ Note that stochastic DNA-pore interactions are also important to nanopore sequencing efforts [72-79], since stochastic DNA-pore interactions can affect residence times for individual nucleotides passing through the pore. Results from this chapter are also applicable to this problem.

4.1 Methods

The probes used in these experiments are comprised entirely of adenosine, ranging in length 15-65 nucleotides (depending on the experiment), and coupled to Avidin at the 5' end by a biotin linker. Probes are added to the *cis*-side of the nanopore cell to a final concentration of 200 nM prior to the start of the experiment²¹.

For each molecule, at each potential, between 100 and 10 000 single molecule escape trials, similar to the trial depicted in Figure 4-1, were performed. Probes are captured in the pore by a 200 mV *trans*-positive potential, and held at that potential for 0.1-0.2 s, allowing them to thread through the pore completely [141]. The acquisition software then reduces the applied potential to the escape potential, and waits for the probe to thermally escape from the pore, which is detected as an increase in ionic current. Rare events, lasting longer than 10 s, are terminated by reversing the potential to -150 mV. Following escape, the potential is returned to the capture potential to await the next trial.

²¹ See Appendix A for details on pore formation protocols, instrumentation, and data acquisition software.

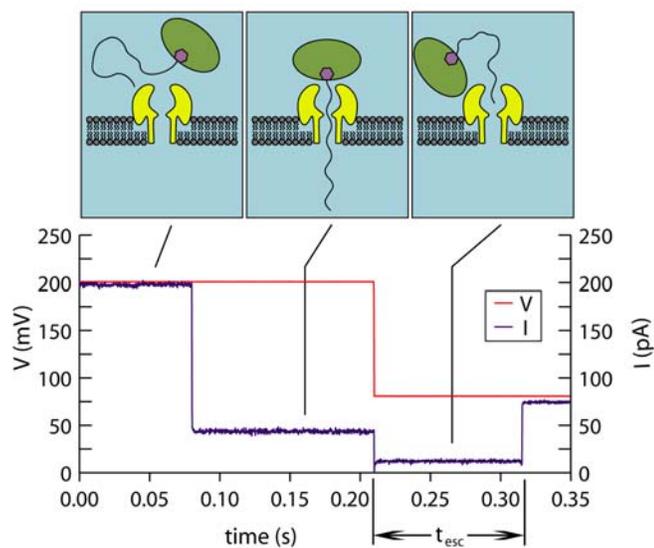


Figure 4-1 A single molecule probe escape event. A 200mV applied potential (red) electrophoretically drives a probe molecule into the pore. Probe capture reduces the ionic current through the pore from 200pA to ~50pA (blue). Following a 0.15 s hold, the applied potential is reduced (75mV shown) to begin the experiment. After some time (t_{esc}), the probe escapes, causing an increase in current to the open channel state.

After the completion of the experiment, custom-written data analysis software is used to determine single molecule escape times from voltage and current traces. Timing for each event begins when the escape potential is reached, and ends upon the increase in ionic current that accompanies probe escape. Long events (> 10 s), which are terminated by reversing the potential, and short events, in which the probe escapes too rapidly to be timed accurately (typically less than 300 μ s) are flagged as improperly timed, but still included in the dataset as either “long” or “short” events, respectively.

The analysis software rejects events where the probe is considered to be inserted into the pore with an unusual conformation. The software monitors the pore impedance, which has been shown to be a sensitive indicator of molecule conformation in α -HL nanopores [70, 81], and discards any events whose impedance differs by >10% from the expected value during either the 200mV capture phase. Unlike the long and short events mentioned above, these events are completely excluded from subsequent analysis. This filtering does not significantly change the characteristic timescale for escape calculated in subsequent analysis.

For further analysis, the event time distribution is expressed in terms of survival probability, $P_{survival}$, which measures the likelihood that the probe is still present in the pore at time t after the escape potential is reached²². This is calculated by sorting the event time distribution sorted in order of increasing duration, and assigning each individual event a survival probability as:

$$P_{survival} = 1 - \frac{i-1}{N_{events}} \quad (4-1)$$

²² Survival probability is equivalent to the reactant concentration in an ensemble experiment. It is normalized, such that $P_{survival} = 1$ at $t = 0$, and decays to 0 at $t = \infty$.

where i is the index of the event in the sorted dataset, and N_{events} is the total number of events. All subsequent analysis begins from plots of survival probability vs. time, as described in the following section.

4.2 Results & Data Analysis

As mentioned in the introduction to this chapter, analysis of probe escape and nanopore force spectroscopy data is complicated by the fact that the event time distribution is broader than expected. Before proceeding to the analysis, we will first consider the relationships between the event time distribution, the characteristic timescale, and the energy barrier in more detail. We begin with the ideal case, for which the energy barrier is fixed, *i.e.* identical for all events.

Processes whose kinetics are governed by crossing a fixed energy barrier are expected to follow the Poisson distribution. A Poisson process must satisfy two conditions. First, event times must be rare enough that a maximum of one event will occur in a sufficiently short time interval Δt . Second, probability that an event will occur in an interval Δt is independent of time, meaning that events occur at a constant rate.

As discussed in Chapter 3, the rate-limiting step in an energy barrier crossing process is attainment of sufficient thermal energy, to overcome the barrier ΔG_b . If ΔG_b is fixed, the probability that a

molecule will attain sufficient thermal energy, through random fluctuations, to overcome the barrier in a time interval Δt is constant. Further, no two molecules will cross the energy barrier at exactly the same time, satisfying the conditions for a Poisson process. The survival probability distribution for such a process takes the form of a simple exponential decay²³:

$$P_{\text{survival}}(t) = e^{-\frac{t}{\tau}} \quad (4-2)$$

where the characteristic timescale τ is related to the energy barrier and applied force through Kramers' rule (3-2).

Note, however, that if some phenomenon, such as stochastic probe-pore interactions, causes the energy barrier to vary from single molecule event to event, this will give rise to a distribution of characteristic timescales, all of which will be represented in the event time distribution. This can be accounted for by generalizing (4-2) to include a distribution of energy barriers [142]:

²³ Note that there are two exponential relationships governing constant rate kinetics. Kramers' rule states that the characteristic timescale is exponentially related to force. In addition, the distribution of event times follows an exponential decay due to Poisson statistics, with a time constant determined by the characteristic timescale for a given barrier. It is therefore extremely important to keep track of which exponential relationship is being considered during different stages of the analysis.

$$P_{\text{survival}}(t) = \int_0^{\infty} e^{-\frac{t}{\tau}} w(\Delta G_b) d\Delta G_b \quad (4-3)$$

where $w(\Delta G_b)$ is the energy barrier probability density function (PDF), which describes the probability of a probe molecule experiencing an energy barrier between ΔG_b and $\Delta G_b + d\Delta G_b$. As before, each specific value for τ is related to a specific energy barrier ΔG_b through Kramers' rule. Note that if $w(\Delta G_b)$ is a delta function, equation (4-3) reduces to equation (4-2).

Equation (4-3) indicates that all molecules experiencing a single, constant energy barrier obey Poisson kinetics; however, if a range of energy barriers is represented in the data, the ensemble of all event times does not follow Poisson kinetics.

The purpose of this section is therefore to extract the energy barrier distribution from the event time distribution, in order to characterize the interactions that give rise to non-Poisson behavior, and to determine an appropriate characteristic timescale for further analysis.

4.2.1 Probe Escape Data Fitting

Figure 4-2 shows the survival probability vs. time for a typical probe escape experiment. Probe escape data is fitted using four different functions, the first three of which are commonly used to model single molecule kinetics, and the fourth, the multi-exponential

decay, is used for further kinetic characterization, as discussed below.

All fits are performed by minimizing the linear least squares error

between the data and the fitted functions. Details of the fits are

described below; the results are shown in Figure 4-2.

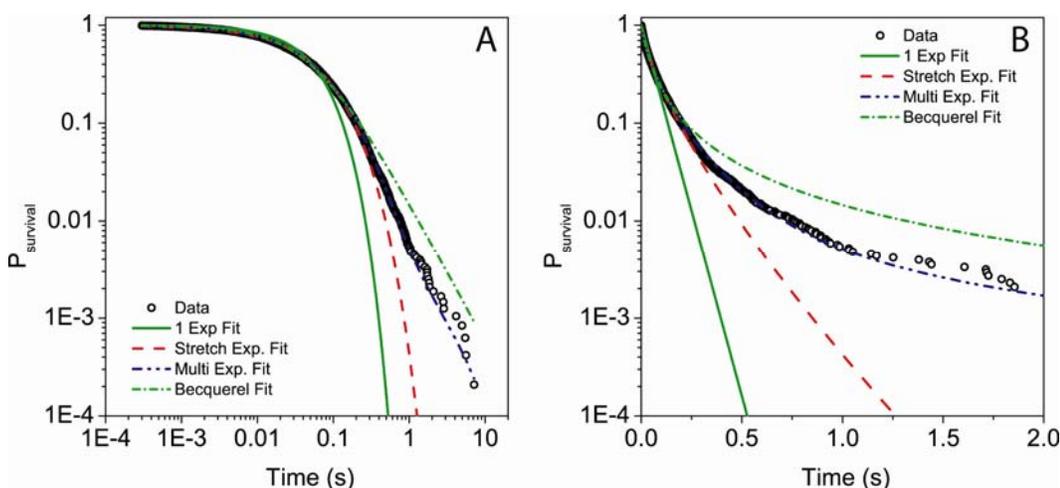


Figure 4-2 Event time distribution for a typical probe escape experiment. Data shown is for the dA_{27} trapped in the pore by an 80mV applied potential (4 783 points). Single exponential and stretched exponential functions both fail to model the power law region between 0.1 and 10 seconds. The Becquerel decay and multi-exponential functions are better, particularly at long times. A. A log-log plot of the data emphasizes the long power law region in the data. B. A semi-log plot of the same data emphasizes deviation from exponential behavior.

The first fitting function, the Poisson distribution defined by equation (4-2), is the distribution for an ideal process governed by a constant energy barrier.

The other two fits we will consider are commonly used to describe processes governed by a distribution of energy barriers [142-147]. The first of these is the Becquerel decay function, which has previously been used to describe processes such as myoglobin's carbon monoxide binding kinetics at low temperatures [143] and phosphor luminescence decay [144]:

$$P_{survival}(t) = \left(1 + \frac{t}{\phi_B}\right)^{-\beta} \quad (4-4)$$

where ϕ_B has dimensions of time, and $1 \leq \beta \leq \infty$ is a dimensionless control parameter, both of which are fitted to the data. Note that in the limit $\beta \rightarrow \infty$, (4-4) becomes an exponential decay [144].

The final fit is the stretched exponential decay, which is frequently used as phenomenological function to describe processes such as fluorescence decay [145] and enzymatic rate fluctuations at the single molecule scale [146, 147]. This takes the form:

$$P_{survival}(t) = e^{-\left(\frac{t}{\phi_S}\right)^\alpha} \quad (4-5)$$

where ϕ_S has dimensions of time, and $0 \leq \alpha \leq 1$ is a dimensionless stretch parameter, both of which are fitted to the data. Note that in the limit $\alpha \rightarrow 1$, (4-5) becomes an exponential decay.

In addition, a multi-exponential decay function, composed of a sum of many exponential decay terms, is shown in Figure 4-1. This function

contains too many parameters, and is consequently too computationally expensive to be viable for routine data fitting (*i.e.* in every experiment). As we shall see below, however, it is extremely useful for evaluating the quality of the fits described above, and for understanding the processes that cause event time distributions to deviate from Poisson kinetics. The multi-exponential decay function takes the form:

$$P_{survival}(t) = \sum_{i=1}^n A_i e^{-\frac{t}{\tau_i}} \quad (4-6)$$

where A_i is a weighting factor for the characteristic timescale τ_i , and $\sum_i A_i \equiv 1$. For the fit shown in Figure 4-2, 43 logarithmically spaced timescales were pre-defined ranging from 1×10^{-4} s, to *ca.* 100 s, and the associated A_i values were fitted to minimize least-squares error.

Examining the quality of the fits in Figure 4-2, it is clear that probe escape is a non-Poisson process. Equation (4-2) deviates substantially from the data at times greater than 0.1s in Figure 4-2, under-predicting the escape time by up to a factor of 50. In this region, the survival probability appears to obey a power law, following nearly a straight line on a log-log plot. Similar behavior at long times was observed for all probes, at all potentials.

The best of the three fits, the Becquerel decay function follows the data well at both short and long times²⁴. We will therefore use this fit for further analysis, beginning in the next section.

4.2.2 Probe Escape Kinetic Analysis

In this section, we determine the shape of the energy barrier PDF from Becquerel fit, and use this to determine both the spread in the energy barrier PDF, and an appropriate timescale τ^* which can be applied to Kramers' rule analysis.

Before proceeding to this analysis, it is important to clearly define some terms that will be referred to frequently in the derivation that follows. First, the derivation will consider ΔG_b , which is the total free barrier including contributions from the electrostatic force and any other interactions, such as non-covalent probe pore bonds. This is distinct from ΔG_b^0 , which we use to denote the free energy barrier in the absence of electrostatic force. Note also, that the derivation will

²⁴ Note that the multi-exponential decay also follows the data extremely well; however, with a wide enough range of timescales, this would describe almost any monotonically decaying function nearly exactly. The multi-exponential decay is used not as a fit, but rather as a method for extracting the energy barrier distribution directly from the data, as we shall see in the following section.

frequently refer to both τ and τ^* . τ is the general definition of a characteristic timescale, governing the exponential decay of a Poisson process (4-2). Any single value of τ is related to a specific energy barrier through Kramers' rule (3-2). As mentioned in the previous section, non-Poisson event time distributions are governed by a distribution of τ 's. For Kramers' rule analysis, we must choose a single τ from the distribution, which we denote by τ^* .

Calculation of τ^* and the shape of the energy barrier distribution (expressed as a function of τ) proceeds by taking the inverse Laplace transform of the event time distribution, as discussed below. This requires some derivation [140], beginning with the generalized event time distribution (4-3), which is closely related to the Laplace transform:

$$\mathcal{L}\{H(\tau)\} = \int_0^{\infty} e^{-\frac{t}{\tau}} H(\tau) d\tau \quad (4-7)$$

where $H(\tau)$ is any probability density function of τ . To convert (4-3) into this form, we must apply a change of variables, expressing ΔG_b in terms of τ . We begin by relating ΔG_b to τ using the Arrhenius relation:

$$\Delta G_b = \ln \tau - \ln \tau_D \quad (4-8)$$

Note that since the diffusive relaxation time, τ_D , is not measured in these experiments, we cannot solve this equation for ΔG_b . We can,

however define a new term, $\Delta G_b'$, which is solely dependent on τ , as follows:

$$\Delta G_b' = \Delta G_b + \ln \tau_D = \ln \tau \quad (4-9)$$

Inserting $\Delta G_b'$ into the energy barrier PDF as $w(\Delta G_b')$ shifts the distribution along the ΔG_b axis by $\ln \tau_D$ with respect to $w(\Delta G_b)$. Thus, $w(\Delta G_b')$ gives the shape of the energy barrier PDF, but not its location on the ΔG_b axis. Bearing in mind this limitation, we will continue to work in terms of $\Delta G_b'$ and τ for the remainder of the derivation.

Applying a change of variables, we re-express (4-3) in terms of $\Delta G_b'$, obtaining:

$$P_{survival}(t) = \int_{\ln \tau_D}^{\infty} e^{-\frac{t}{\tau}} w(\Delta G_b') d\Delta G_b' \quad (4-10)$$

Next, we define a function $W(\tau)$, which expresses $w(\Delta G_b')$ in terms of τ :

$$W(\tau) \equiv w(\ln \tau) = w(\Delta G_b') \quad (4-11)$$

Using this equation, we apply a change of variables to (4-10) and obtain:

$$P_{survival}(t) = \int_{\tau_D}^{\infty} W(\tau) e^{-\frac{t}{\tau}} d \ln \tau \quad (4-12)$$

Since the integrand is approximately zero for $\tau < \tau_D$ [143] (also, see below), we can extend the lower bound of integration to zero:

$$P_{survival}(t) \approx \int_0^{\infty} \frac{W(\tau)}{\tau} e^{-\frac{t}{\tau}} d\tau \quad (4-13)$$

Equation (4-13) is the Laplace transform of the $W(\tau)/\tau$. Inverting the transform, we calculate $W(\tau)$:

$$W(\tau) = \tau \mathcal{L}^{-1} \{ P_{survival}(t) \} \quad (4-14)$$

where \mathcal{L}^{-1} is the operator for the inverse Laplace transform.

Substituting the Becquerel decay function (4-4) into the right-hand-side of (4-14) gives:

$$W(\tau) = \tau \mathcal{L}^{-1} \left\{ \left(1 + \frac{t}{\phi_B} \right)^{-\beta} \right\} \quad (4-15)$$

Equation (4-15) can be solved analytically [143], giving the final result:

$$W(\tau) = \frac{(\phi_B/\tau)^\beta e^{-\phi_B/\tau}}{\Gamma(\beta)} \quad (4-16)$$

This is the energy barrier probability density function, expressed in terms of τ .²⁵ Note that ϕ_B and β are the fitted parameters from the Becquerel decay function (4-4), and $\Gamma(\beta)$ is the Gamma function. Figure 4-3 shows plot of $W(\tau)$ vs. τ for the dA₂₇ probe at 80mV.

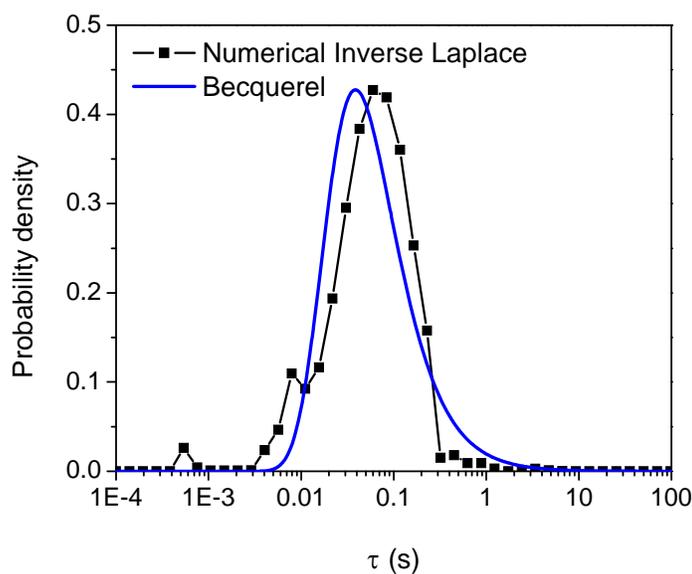


Figure 4-3 Calculated energy barrier probability density function, expressed as a function of τ , for a typical probe escape dataset. The PDF is calculated from data for the dA₂₇ probe escape at an 80mV potential (shown in Figure 4-2). Note that the shape of the distribution, when plotted on a logarithmic τ axis, has the same shape as the energy barrier distribution when plotted on a linear ΔG_b axis.

²⁵ Note that this is not the same as the timescale PDF, which is $H(\tau) = \underline{W}(\tau)/\tau$ (see equation (4-13)).

Before using this to proceed to further analysis, we will first check the accuracy of $W(\tau)$ in describing the energy barrier distribution, since Becquerel decay function upon which it is based was chosen without knowledge of the physical mechanism causing non-Poisson behavior. We do this by comparing $W(\tau)$ to a discrete, numerical inverse Laplace transform of event time distribution, derived from the multi-exponential fit. This makes fewer assumptions about the shape of the underlying energy barrier PDF than the Becquerel fit.

Comparing equations (4-6) and (4-3), it is apparent that the multi-exponential decay function is a discrete version of the generalized event time distribution. The multi-exponential fit generates a timescale histogram of (τ_i, A_i) values, each of which is the integral of the $w(\Delta G_b')$ over an interval $\Delta\Delta G_b'$.

Using equation (4-8), we can calculate the difference in energy barrier heights for two different timescales²⁶ as:

$$\Delta\Delta G_b' = \ln\left(\frac{\tau_{i+1}}{\tau_i}\right) \quad (4-17)$$

Since the multi-exponential decay (4-6) was fitted using logarithmically spaced timescales, the energy barrier difference

²⁶ Note that this assumes that τ_D is independent of energy barrier height [143].

between any two adjacent timescales is constant. We therefore relate A_i to $w(\Delta G_b')$ by integrating (4-6) over the interval $\Delta\Delta G_b'$:

$$A_i = \int_{\Delta G_{bi} - \Delta\Delta G_b/2}^{\Delta G_{bi} + \Delta\Delta G_b/2} w(\Delta G_{bi}') dG_b' \quad (4-18)$$

As before, we perform a change of variables, to express $\Delta G_b'$ in terms of τ :

$$\begin{aligned} A_i &= \int_{\ln \tau_i - (\ln \Delta\tau)/2}^{\ln \tau_i + (\ln \Delta\tau)/2} \frac{W(\tau)}{\tau} d\tau \\ &\approx \frac{W(\tau_i)}{\tau_i} \ln(\Delta\tau) \end{aligned} \quad (4-19)$$

Which can be solved for $W(\tau_i)$:

$$W(\tau_i) \approx \frac{A_i \tau_i}{\ln(\Delta\tau)} \quad (4-20)$$

Thus, the multi-exponential fit can be transformed into a discrete, numerical inverse Laplace transform of the data by re-scaling A_i values according to equation (4-20).

The resulting distribution, shown in Figure 4-3, is an extremely accurate representation of the true energy barrier PDF, since it is essentially calculated directly from the data. As seen in Figure 4-3, the energy barrier PDFs calculated from the Becquerel fit and the numerical inverse Laplace transform agree well, implying that the

Becquerel fit describes the energy barrier PDF governing the escape process.

We can now proceed to use these energy barrier PDFs to characterize the probe escape process, beginning with the spread in the distribution, i.e. the range of energy barriers experienced by different molecules escaping from the pore. This is calculated by determining the full width at half maximum (FWHM) of the energy barrier PDF, transforming τ to $\Delta G_b'$ using (4-17). For the PDF shown in Figure 4-3, the FWHM is $\sim 2.0 k_B T^{27}$, which is similar to the energy of 1-2 hydrogen bonds. Energy spreads for the full set of probe escape experiments (*i.e.* for all probes, at all potentials) ranged from $\sim 1 k_B T$ to $\sim 5 k_B T$, and tended to increase with the escape timescale. Possible causes for this variation are discussed in the next section.

The second piece of information we wish to extract from this distribution is a characteristic timescale, τ^* which can be used for Kramers' rule analysis of force spectroscopy results. A reasonable choice for τ^* is the timescale associated with the most heavily represented energy barrier, *i.e.* the timescale at the peak of the distribution. To calculate τ^* , we set the first derivative of equation (4-16) to zero and solve:

²⁷ $2 k_B T \approx 1.2 \text{ kcal/mol}$ at 293K

$$\tau^* = \frac{\phi_B}{\beta} \quad (4-21)$$

For the data in Figure 4-3, $\tau^*=0.037$ s. The timescale at the peak of the numerical inverse Laplace transform is 0.060 s. The difference between these two values represents a difference in energy barrier heights of $\sim 0.5 k_B T$, which is small. We will therefore continue to use (4-21) to calculate τ^* , noting that we will demonstrate the appropriateness of this choice in the following section by confirming that τ^* follows Kramers' rule.

4.3 Discussion

In this section, we use the results from the previous section to speculate about causes for non-Poisson behavior in probe escape experiments, and to perform a further check of τ^* calculation, by demonstrating that it follows Kramers' rule.

4.3.1 Possible Causes of Energy Barrier Variation

Though the mechanisms causing non-Poisson behavior were not studied directly, we can speculate about possible causes. Variation in the applied potential can be excluded, since feedback signals from the patch clamp amplifier indicate that the potential is stable to within 0.1mV, which would lead to energy barrier variation on the order of

$\sim 0.1 k_B T$ at maximum. Improper threading of the DNA molecule, such as capture of a DNA loop in the vestibule, could conceivably produce energy barrier variation, but is also unlikely to be the cause here. Such events are discarded from the analysis by excluding any events with non-characteristic ionic current signals (see section 4.1), which has been shown to be a highly sensitive indicator of DNA configuration [70, 81]. All remaining data can reasonably be assumed to consist of events where the probe is correctly threaded through the pore.

A compelling possibility is stochastic, non-covalent binding interactions between DNA and the α -HL pore. Lysine residues in the α -HL constriction could form salt bridges with phosphate groups in the DNA backbone [148], or amino acid residues within the pore could form hydrogen bonds with the bases of the DNA. If such binding interactions vary from event to event, it could conceivably produce an energy barrier distribution similar to what was seen above. Extensive experimental work would be required to prove this; however, this possibility is supported by the fact that probe escape kinetics are sequence-dependent [149], suggesting that base-specific interactions may be involved.

4.3.2 Experimental Validation of the Characteristic Timescale as a Measure of the Energy Barrier

Having developed a simple expression for calculating τ^* , we wish to validate our algorithm by demonstrating that τ^* obeys Kramers' rule before applying this analysis to genotyping experiments.

Kramer's rule (3-2) predicts that the characteristic timescale should increase exponentially with the energy barrier. The energy barrier in probe escape experiments is primarily electrostatic²⁸, and can thus be easily controlled by varying either the probe length, or the electric field. We model the electrostatic energy barrier as:

$$\Delta G_b(V) = zeV \sum_{i=1}^N \frac{\Delta V_i}{V} \quad (4-22)$$

where e is the elementary charge in Coulombs, N is the total number of nucleotides in the probe molecule, i is the index of a given nucleotide, numbered from the 5' end of the DNA molecule (which is tethered to avidin), $\Delta V_i / V$ is the fraction of the applied potential that the i^{th} nucleotide must cross to escape from the pore, and z is the effective charge per nucleotide. z accounts for charge screening by nearby K^+ ions which partially cancel negative charges on the DNA backbone,

²⁸ Note that for this analysis, we ignore potential probe-pore interactions described in the previous chapter, since energy of these interactions is much smaller than the electrostatic energy barrier.

and the effects of electro-osmotic flow, both of which reduce the effective force applied to the DNA molecule²⁹.

Equation (5-10) states that $\Delta G_b(V)$ is proportional to V , meaning that τ^* should increase exponentially with applied potential, according to Kramers' rule (3-2). We therefore confirm that τ^* forms a good metric of the energy barrier by demonstrating that τ^* scales as expected with applied potential and probe length.

This was accomplished by performing probe escape experiments using seven different probes, each composed exclusively of adenosine and ranging in length from 20-65 bases. Each probe was tested over a range of escape potentials. Data was fitted to the Becquerel decay function (4-4), and the dominant timescale τ^* was calculated from equation (4-21). Results are shown in Figure 4-4. Characteristic escape timescales are exponentially dependent on applied potential for each of the molecules tested, as predicted.

²⁹ z is calculated in Appendix D, which also calibrates the applied electrostatic force acting on the DNA molecule.

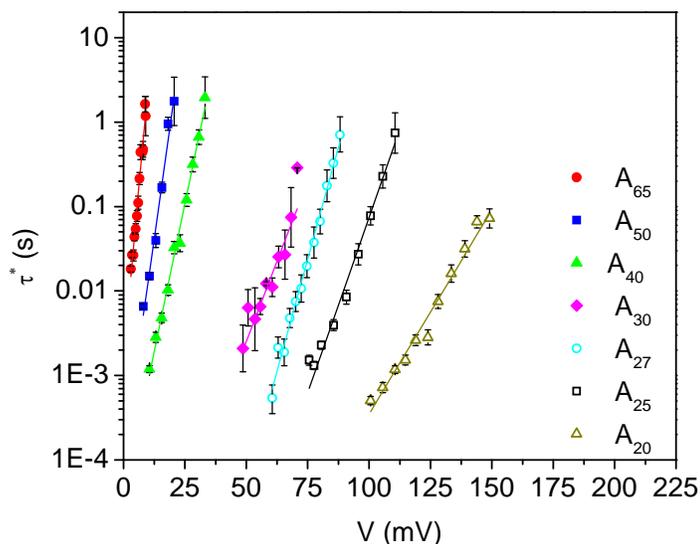


Figure 4-4 Probe escape characteristic timescales vs. applied potential. Lines are exponential fits to the data. Error bars indicate logarithmic sample standard deviation, estimated using a bootstrap algorithm, as described in Appendix E.

In addition, the relationship between probe length and the escape timescale is explored in Appendix D, confirming that the energy barrier behaves as predicted with probe length.

We can therefore state that τ^* obeys Kramers' rule, and is thus an appropriate parameter for force spectroscopy analysis. Since electrostatic forces are applied similarly in probe escape and duplex DNA force spectroscopy, this implies that τ^* can also be applied to analysis of nanopore force spectroscopy studies of DNA duplexes, which is the topic of the following chapter.

Chapter 5 – DNA Duplex Dissociation by Nanopore Force Spectroscopy

This chapter considers nanopore force spectroscopy of DNA duplexes, and lays the groundwork for base-calling in Chapter 6.

The first portion of the chapter repeats the analysis of event time and energy barrier distributions, following the general method developed in Chapter 4. NFS event time distributions are slightly different than those seen in probe escape experiments, requiring the use of different fitting functions. From this analysis, we characterize energy barrier variation and derive an expression for τ^* in duplex dissociation experiments. In the latter portion of the chapter, we confirm that τ^* forms a good indicator of the energy barrier height, by demonstrating that it scales as expected with force and energy barrier height.

In addition, results presented in this chapter act as standards for base calling, by providing known characteristic timescales for different target sequences.

5.1 Methods

This section describes the experimental and raw data analysis methods used for nanopore force spectroscopy of DNA duplexes. Single pore and multi-pore methods differ sufficiently that they are described separately, in sections 5.1.1 and 5.1.2 respectively. The probes and targets used in these experiments are common to both types of experiment, and are given in Table 5-1.

Probe Name	Probe Sequence	Target Name	Target Sequence	ΔG_{melt}	
				kcal/mol	k _B T
Probe Synth	5'-CCACCAACCAAACC-A ₃₀ -3'-biotin	PC	5'-GGTTTGGTTGGTGG	23.1	38.9
		5C	5'-GGTTTGGTT <u>C</u> GTGG	17.1	28.8
		8C	5'-GGTTT <u>G</u> CTTGGTGG	16.9	28.5
		8T	5'-GGTTT <u>G</u> T <u>T</u> TGGTGG	17.2	29.0
		3G12G	5'-GG <u>G</u> TGGTGGTGG <u>G</u> GG	17.0	28.6
		PC-rev	5'-GGT <u>GGTTGGTT</u> TGG	13.3	22.4
Probe rs7242-C	5'-ACGCATCTGAACTTTCTTCCA ₃₀ -3'-biotin	rs7242-G	5'-GGAAGAAAGGTCAGATCGCGT	28.4	49.6
		rs7242-T	5'-GGAAGAAA <u>T</u> GTCAGATCGCGT	24.7	43.1

Table 5-1 Probe and target sequences used in nanopore force spectroscopy genotyping studies. Mismatches are shown in red. The Synth probe - target set was designed to minimize secondary structure and out-of-register hybridization. Note that with the exception of PC (perfect complement) and PC-rev (reverse of the perfect complement), Synth target names reflect the position and identity of the mismatch, numbered from the 3' end, (nearest to the pore). rs7242 probe-target sets simulate genotyping of a single nucleotide polymorphism in the SERPINE1 gene, which has been implicated in sepsis response [150]; target names reflect SNP identities. All melting energies were determined using the DINAMelt web server's two-state hybridization energy calculator [151], which uses an empirical, nearest-neighbour algorithm; full details of the calculation are given in two related papers: [152, 153].

5.1.1 Single Pore Methods

All single pore force spectroscopy experiments use probe and target concentrations of 200 nM and 500 nM, respectively³⁰. For each experiment, between 250 and 2 500 single molecule dissociation trials, similar to the trial depicted in Figure 3-8, are performed.

The probe is driven into the nanopore by a 200 mV *trans*-positive potential, and detected by a stepwise decrease in the ionic current (see Figure 3-8) by the data acquisition software. Following capture, the probe is held at the capture potential for 0.5 s to allow it to hybridize with a target molecule.

Next, a binding check is performed, which allows analysis software to easily distinguish between un-hybridized probes and rapidly dissociating duplexes for proper short event counting (see raw data analysis procedures, below). In this step, the potential is reduced to 10 mV and held for 0.2 – 0.25 s. Hybridized probes remain trapped in the pore by the duplex, while un-hybridized probe escape³¹ during this step (compare Figure 3-8 and Figure 5-1).

³⁰ See Appendix A for details on pore formation protocols, instrumentation, and data acquisition software, and Appendix C for studies used to design probe length.

³¹ Over 99% of unhybridized probes escape during this step, as confirmed by probe escape data at 10 mV for the A₅₀ probe, which is longer, and therefore escapes more slowly than the force spectroscopy probe.

Next, the force spectroscopy trial is conducted by changing the applied potential to the dissociation potential. After some time, the probe thermally dissociates from the target and escapes from the pore, causing a stepwise change in the ionic current. Rare long events, lasting longer than 10 s, are terminated by changing the applied potential to -150 mV, which rapidly dissociates the duplex. Following dissociation, the potential is returned to the capture potential to await the next trial.

Note that the event time includes contributions from both duplex dissociation and probe exit, since the ionic current does not change until the probe exits the pore. However, under the conditions of these experiments, probe exit (*ca.* 10^{-4} s [141]) is typically much faster than dissociation (10^{-2} - 10 s), meaning that the rate-limiting step is the process we are interested in - probe-target dissociation.

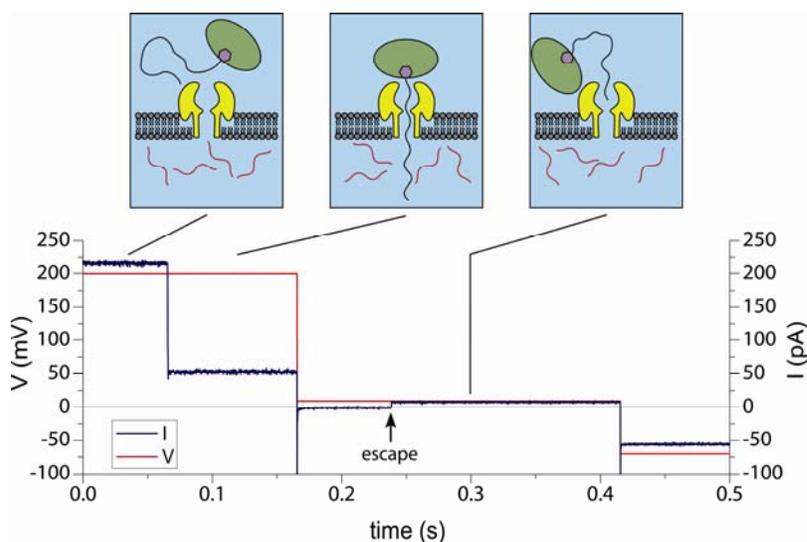


Figure 5-1 An unsuccessful single molecule nanopore force spectroscopy trial. Data shown is for the Synth probe with the PC target, and applied potentials are similar to those seen in Figure 3-8. In the trace shown, the un-hybridized probe escapes during the 10 mV hold step, causing an increase in current to the open-channel state. These events are discarded from the resulting dataset, as described below, and not used in further analysis.

After the completion of the experiment, single molecule event times are determined from voltage and current traces. Events involving un-hybridized probes, in which the ionic current increases to the open channel value during the hybridization check step (see Figure 5-1) are discarded. In addition, as with probe escape experiments, events where the probe is considered to be inserted into the pore with an unusual conformation are rejected by discarding any events whose impedance differs by >10% from the expected value during the 200mV capture phase.

The remaining events are timed for further analysis. Timing begins when the dissociation potential is reached, and ends upon the step-wise change in the ionic current that accompanies probe exit. Long events (> 10 s) which are terminated before dissociation occurs, and short events, where the probe escapes too rapidly to be timed accurately (typically less than $300 \mu\text{s}$) are flagged as improperly timed. These events are still counted for P_{survival} calculation as either “long” or “short” events, respectively. From the resulting set of event times, the survival probability as a function of time is calculated using equation (4-1).

5.1.2 Multi-Pore Methods

Multi-pore force spectroscopy was first developed and described by Tropini and Marziali [35]. Studies in this thesis modify and expand on this work, in order to demonstrate the feasibility of this method for genotyping.

Conceptually, multi-pore force spectroscopy experiments are similar to single pore experiments, in that many duplex dissociation events are observed, and used to calculate a survival probability distribution. However, many details of the procedure are different for multi-pore experiments, since data acquisition and analysis are performed for multiple single-molecule events in parallel.

In each multi-pore force spectroscopy experiment, the lipid bilayer contains between 50 and 150 pores³², and the probe and target concentrations are 200 nM and 500 nM, respectively. During data acquisition, 5 – 50 multi-pore trials are conducted, representing a total of 250 - 7500 single molecule events.

The applied potential during data acquisition follows a pre-determined waveform³³, similar to that shown in Figure 5-2. First, a 200 mV potential is applied for 10 s, to drive probes into the pore and allow them to hybridize with target molecules.

Next, the potential is reduced to 0 mV for 2 s, allowing un-hybridized probes to escape, while retaining probe-target duplexes within the pores. The potential is then briefly increased to 50 mV, so

³² See Appendix A for details on pore formation protocols, instrumentation, and data acquisition software, and Appendix C for studies used to design probe length.

³³ Recall that in single pore force spectroscopy experiments, data acquisition software changes the applied potential in response to single molecule events, such as probe capture or duplex dissociation. This is not possible here, since many single molecule events are observed in parallel; therefore, no feedback is used, e.g. the 200 mV capture potential is held for exactly 10 s in each trial, *etc.*

that the ionic current can be recorded for calculation of the number of blocked pores³⁴, used in $P_{survival}(t)$ calculations (see below).

The force spectroscopy trial is then conducted by changing the applied potential to the dissociation potential, causing probes to thermally dissociate from the target molecules under applied force. The potential is held at the dissociation potential for 10 - 100 s, giving time for nearly all probe-target duplexes to dissociate.

Following the force spectroscopy trial, any pores remaining blocked are cleared, and the open pore current is measured for $P_{survival}$ calculation. First, the applied potential is changed to -150mV, which rapidly dissociates any remaining duplexes, but also causes some pores to gate, i.e. spontaneously close. Gated pores are re-opened by brief application of a 10 mV potential, and finally, the dissociation potential is re-applied to measure the open pore current.

³⁴ Note that DNA capture rates have been measured to be *ca.* 0.005 s^{-1} at 50 mV [54], which is low enough that probe capture does not affect measurement of the number of hybridized probes in this step.

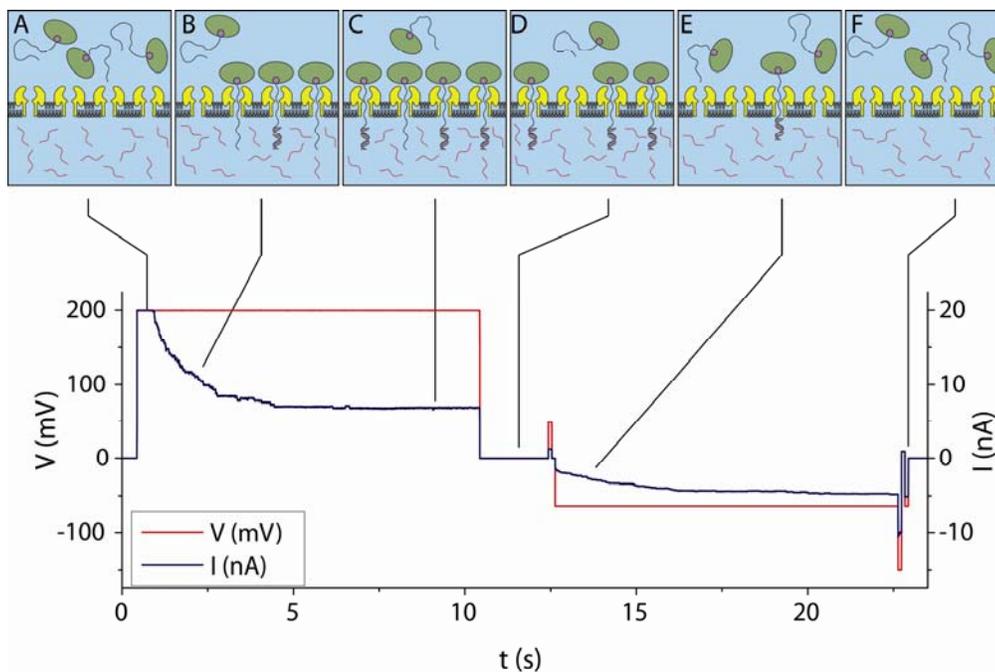


Figure 5-2 A typical multi-pore force spectroscopy trial. Data shown is for the rs7242-C probe and rs7242-G target. Probes are driven into the pores (A, B) by an applied potential (red) of 200 mV, causing the ionic current (blue) to decay as probes are captured. Note that at the start of this step, the ionic current briefly exceeds the measurement range of the patch clamp amplifier, and saturates at 20 nA. The capture potential is held for 10 s to allow probes to hybridize to targets (C). The potential is then reduced to 0 mV for 2.0 s, allowing any un-hybridized probes to escape from the pores (D), and then briefly increased to 50 mV. Next, the potential is changed to the dissociation potential (-65 mV shown), causing the current to decay as duplexes dissociate, and most pores clear (E). The applied potential is then changed to -150 mV, rapidly dissociating any remaining duplexes; however, this causes some pores to gate (i.e. spontaneously close), requiring that they be re-opened by briefly changing the potential to 10 mV. Finally, the potential is returned to the dissociation potential, allowing the open pore current to be measured (F).

Following experiments, $P_{survival}$ distributions are extracted from the ionic current decay during the dissociation step. The survival probability is the fraction of initially blocked pores still containing probes at time t during the dissociation step:

$$P_{survival}(t) = \frac{N_b(t)}{N_b(0)} \quad (5-1)$$

where $N_b(t)$ is the number of blocked pores, still containing duplexes at time t , and $N_b(0)$ is the number of blocked pores at the beginning of the dissociation step. We will now derive an expression for $P_{survival}(t)$ which is expressed entirely in terms of the ionic current (modified slightly from [35]).

We begin by noting that the ionic current through each pore during the dissociation step is either i_b (blocked) or i_o (open). These have been measured as a function of applied potential [35]:

$$i_o = \begin{cases} 0.97V & V > 0 \\ 0.74V & V < 0 \end{cases} \quad (5-2)$$

$$i_b = \begin{cases} 0.0006V^2 + 0.0953V & V > 0 \\ -0.0003V^2 + 0.0645V & V < 0 \end{cases} \quad (5-3)$$

The total ionic current at time t is a linear sum of the currents through all blocked and open pores:

$$\begin{aligned} I(t) &= N_b(t)i_b + N_o(t)i_o \\ I(t) &= N_b(t)i_b + [N_p - N_b(t)]i_o \end{aligned} \quad (5-4)$$

where $N_b(t)$ and $N_o(t)$ are the number of blocked and open pores, respectively, and $N_P = N_b(t) + N_o(t)$ is the total number of pores.

Rearranging (5-4) we can calculate the number of blocked pores from the ionic current:

$$N_b(t) = \frac{I(t) - N_P i_o}{i_b - i_o} \quad (5-5)$$

Note that $N_P i_o = I(\infty)$ is the total current when all pores are open, which is measured during the experiment. Substituting this into (5-5), we obtain:

$$N_b(t) = \frac{I(t) - I(\infty)}{i_b - i_o} \quad (5-6)$$

which expresses $N_b(t)$ entirely in terms of current.

Recall that the purpose of this derivation is to determine $N_b(t)/N_b(0)$ in terms of ionic current for calculation of the survival probability distribution. Some further derivation is required to calculate $N_b(0)$, since we must first determine $I(0)$. Capacitive currents and rapid duplex dissociations at the start of the dissociation step mean that we cannot measure $I(0)$ directly. We therefore calculate $I(0)$ using knowledge of pore impedance properties, and the number of blocked pores at the end of the hybridization check step.

First, we calculate $N_b(0)$ from the ionic current during the 50 mV check step using (5-5), noting that i_b and i_o in this calculation are for an

applied potential of 50 mV. We then use this to calculate $I(0)$ at the dissociation potential using (5-4).

Finally, using (5-6) to express $N_b(t)$ and $N_b(0)$ in (5-1) in terms of the ionic current at the dissociation potential, we obtain:

$$P_{survival}(t) = \frac{(I(t) - I(\infty)) / (i_b - i_o)}{(I(0) - I(\infty)) / (i_b - i_o)} \quad (5-7)$$
$$P_{survival}(t) = \frac{I(t) - I(\infty)}{I(0) - I(\infty)}$$

which is the desired expression.

Equation (5-7) is used to calculate the survival probability from multi-pore force spectroscopy current traces by sampling $I(t)$ over a logarithmically increasing time interval. This preserves temporal resolution at short times, while preventing over-fitting to the data at long times (where few dissociations occur). Note that multiple trials (*i.e.* capture / dissociation cycles) are performed at each potential; prior to further analysis, all $P_{survival}(t)$ curves at a given dissociation potential are averaged.

5.2 Results & Data Analysis

We now turn our attention to characteristic timescale determination from nanopore force spectroscopy data. This follows the same general procedure developed in section 4.2. However, the event time

distribution takes on a slightly different shape than in probe escape, requiring the use of different fitting functions.

5.2.1 Nanopore Force Spectroscopy Data Fitting

In this section, we consider the event time distribution for nanopore force spectroscopy of DNA duplexes, using examples from a single-pore experiment involving the dissociation of Synth probe from the PC target. It should be noted that the results from multi-pore experiments were similar (see Figure 5-3), meaning that the analysis developed below applies to both single pore and multi-pore experiments.

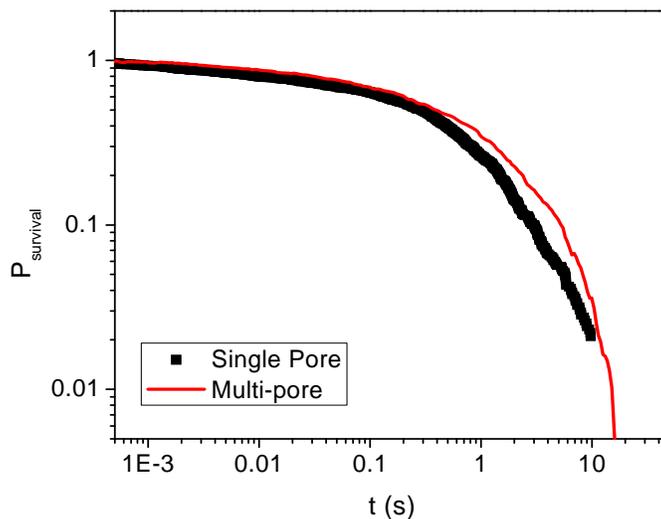


Figure 5-3 Comparison of single pore and multi-pore force spectroscopy results. Event time distributions for force spectroscopy dissociation of the PC target from Synth probe (see Table 5-1) at -50 mV are shown. 933 and *ca.* 450 events are included in the single pore and multi-pore datasets, respectively.

Figure 5-4 shows the event time distribution for a typical nanopore force spectroscopy experiment, along with fits to the exponential decay (4-2), Becquerel decay (4-4) and stretched exponential decay (4-5) functions, as well as the multi-exponential decay function, (4-6), used to determine the energy barrier distribution from the numerical inverse Laplace transform.

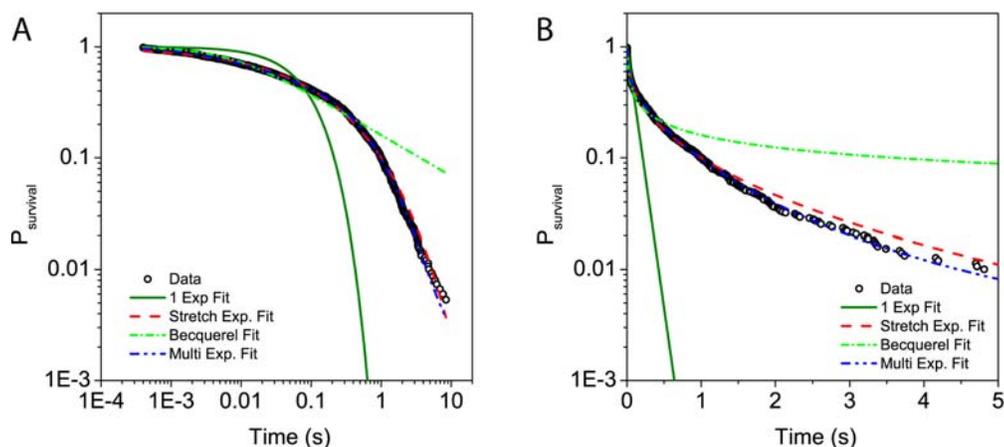


Figure 5-4 Event time distribution for a typical single molecule force spectroscopy experiment. Data shown is for dissociation of the Synth probe from the PC target at -60 mV (1391 points). Fits are described in section 4.2.1. Note that events longer than 10 s are manually terminated, and do not appear on the plot. A. Log-log plot of the data. B. Semi-log plot of the same data as in A, emphasizing deviation from exponential behavior at long times.

Examining the fits, it is clear that the event time distribution is best described by a stretched exponential decay, rather than the Becquerel decay function used in probe escape studies (see Figure 5-4). This is not surprising, given that the process is fundamentally different – i.e. duplex dissociation, rather than escape against an electrostatic potential. Regardless of the mechanism leading to stretched exponential behavior, we need to repeat the energy barrier distribution analysis developed in the previous chapter.

5.2.2 Nanopore Force Spectroscopy Kinetic Analysis

As before, we begin by calculating the energy barrier PDF for the best fit, which in this case, is the stretched exponential decay (4-5). The inverse Laplace transform of the stretched exponential decay function cannot, in general, be calculated analytically; however, a good approximation does exist [145]:

$$W(\tau) \approx \frac{B}{(\phi_s/\tau)^{\alpha/2(1-\alpha)}} \exp\left[-\frac{(1-\alpha)\alpha^{\alpha/(1-\alpha)}}{(\phi_s/\tau)^{\alpha/2(1-\alpha)}}\right] m(\tau) \quad (5-8)$$

where:

$$m(\tau) = \begin{cases} 1/\left[1 + C\left((\phi_s/\tau)^{\alpha(0.5-\alpha)/(1-\alpha)}\right)\right] & \alpha \leq 0.5 \\ 1 + C\left((\phi_s/\tau)^{\alpha(\alpha-0.5)/(1-\alpha)}\right) & \alpha > 0.5 \end{cases}$$

and the empirical parameters B and C are functions of α [145]. Recall that a plot of $W(\tau)$ vs. $\ln(\tau)$, has the same shape as $w(\Delta G_b)$ vs. ΔG_b ; however we cannot relate any specific τ to the corresponding energy barrier, since τ_D is unknown (see section 4.2.2).

Figure 5-5 shows a plot of $W(\tau)$, as calculated from (5-8), for the dataset shown in Figure 3-8. As before, we wish to compare this to the energy barrier PDF calculated from the numerical inverse Laplace transform (4-20), which gives a more accurate representation of the

energy barrier PDF. Before comparing the two distributions, we will first examine some of the characteristics of the PDF calculated from the numerical inverse Laplace transform.

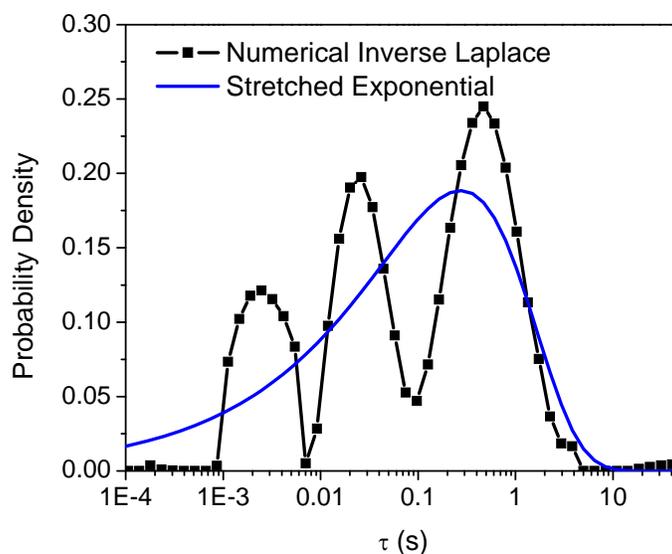


Figure 5-5 Calculated energy barrier probability density functions, expressed as a function of τ , for a typical single molecule force spectroscopy dataset. The PDF is calculated from Synth probe - PC target dissociation data at -60 mV applied potential (event time distribution shown in Figure 5-4).

The numerically calculated energy barrier PDF contains three well-separated peaks. Qualitatively, the weight and energy of the peaks are correlated, - *i.e.* higher energy peaks in the distribution are more heavily weighted. Energy barrier PDFs from other from other NFS experiments contained 1-4 peaks, and showed a similar correlation between peak height and energy. It is possible that PDFs where few

peaks were observed contained additional peaks that could not be resolved, due to the fact that events faster than $\sim 300 \mu\text{s}$ could not be timed accurately.

Peaks were typically 1-3 $k_B T$ wide at half maximum, and similar in shape to those seen in probe escape experiments, suggesting that peak broadening is caused by the same mechanism in both cases. As discussed in section 4.3.1, this mechanism most likely involves interactions between the probe DNA and the pore. The mechanism leading to peak splitting is unclear. The duplex is somehow involved, since peak splitting was never observed in probe escape experiments; however further experiments would be required to determine the detailed mechanism.

Comparing the two distributions in Figure 5-5, we see that the stretched exponential-derived energy barrier PDF appears to be a smoothed representation of the numerically calculated energy barrier PDF. Though the stretched exponential-derived energy barrier contains only a single peak, other characteristics of the two distributions are similar. Both are broad and skewed towards higher energy. In addition, their dominant peaks (*i.e.* highest points) occur at

nearly the same location³⁵. Since the stretched exponential captures the most important characteristics of the energy barrier PDF – in particular, the location of the dominant peak, we will continue to use this for further analysis.

As with probe escape experiments, we will take the timescale associated with the most heavily represented energy barrier for τ^* , noting that we will confirm that this forms a good metric of the dissociation energy in the next section. We therefore proceed to calculate τ^* from the peak of $W(\tau)$.

An empirical function was used to calculate τ^* , since setting derivative of (5-8) to zero gave an equation that could not easily be solved. First, equation (5-8) was solved numerically for τ^* for a large number of values of τ and α . The resulting distribution was found to be fitted by the following empirical function:

$$\tau^* = \varphi \phi_s \alpha^\chi e^{-\frac{\alpha}{\theta}} \quad (5-9)$$

where $\varphi = 18.24$, $\chi = 2.190$, and $\theta = 0.3610$.

³⁵ The separation between the two peaks was found to be less than 0.5 k_BT for all datasets. As we shall see below, this is small compared to the change in dissociation energy when mismatches are introduced into the duplex.

Thus, we have a straightforward method for calculating τ^* from nanopore force spectroscopy event time distributions. We are therefore ready to proceed to Kramers' rule analysis, as described in the next section.

5.3 Discussion

In this section, we establish that τ^* is a good measure of the energy barrier by confirming that it obeys Kramers' rule. More importantly, we demonstrate that τ^* is sufficiently sequence specific to permit base-calling with single nucleotide resolution.

Figure 5-6 shows timescale versus applied potential results for the dissociation of Synth probe from 2 different targets: PC, which is the perfect complement, and 8T, which contains a single nucleotide mismatch in the middle of the duplex (see also Figure 5-7, below). The most striking feature of Figure 5-6 is the separation between the two datasets. Though the targets differ by only a single nucleotide, their characteristic timescales differ by up to a factor of 500, clearly demonstrating that nanopore force spectroscopy can distinguish between DNA sequences with single nucleotide specificity.

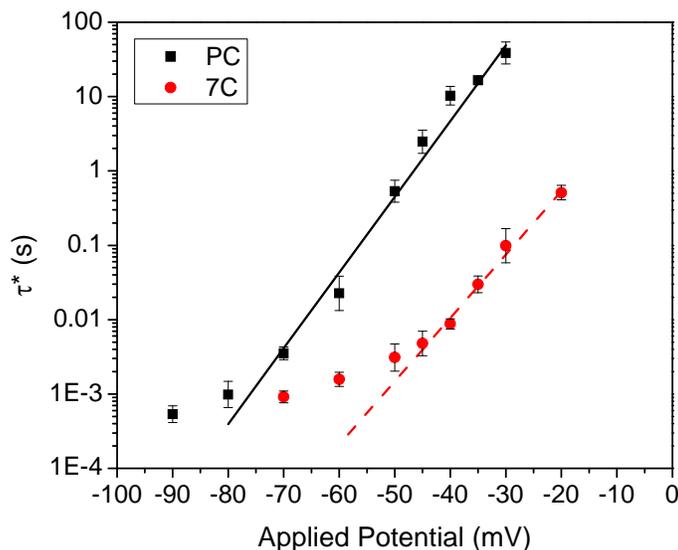


Figure 5-6 Single nanopore FS duplex dissociation characteristic timescale vs. applied potential for the Synth probe with two different targets. Targets tested are PC (perfect complement) and 8C (single mismatch, see Table 5-1). Note that each point is an average, taken over 3-6 different trials. Each trial contains *ca.* 250 - 2500 single molecule events. Exponential fits (5-10) to the data are performed by minimizing logarithmic least squares error, and consider only points in the regime where Kramers' rule is valid, as discussed below. Error bars show logarithmic standard error of measurement, calculated by (D-2) from the set of τ^* values from multiple trials.

Having demonstrated single-nucleotide specificity in timescales, we now wish to explore the behavior of the data with respect to Kramers' rule in more detail. We begin by re-expressing Kramers' rule (3-2) to explicitly include the electrophoretic force (C-3), noting that

this substitution involves a sign change in the force term, since pulling force is positive when the applied potential is negative:

$$\begin{aligned}\tau^* &= \tau_D \exp\left[\Delta G_b^0 - f\Delta x_b\right] \\ \tau^* &= \tau_D \exp\left[\Delta G_b^0 + \frac{zeV}{l}\Delta x_b\right]\end{aligned}\tag{5-10}$$

where $z \approx 0.4$ is the effective charge per nucleotide (see Appendix D), e is the elementary charge, $l = 0.45\text{nm}$ is the length per nucleotide (see Appendix C), and Δx_b is the width of the energy barrier (measured as displacement of the probe). Recall, also that τ_D is the diffusive relaxation time and ΔG_b^0 is the free energy barrier in the absence of applied force.

To a first approximation, all of the terms in equation (5-10) are constant, except τ^* , ΔG_b^0 , and V . We are therefore interested in exploring these relationships, in order to confirm that duplex dissociation by nanopore force spectroscopy obeys Kramers' rule.

Examining the relationship between τ^* and V in Figure 5-6, it is apparent that, so long as the applied potential remains within reasonable limits, τ^* obeys Kramers' rule, decreasing exponentially as the applied potential becomes more negative. Note, however, that for large forces, *i.e.* $V < -70\text{ mV}$ for the PC target, τ^* deviates from Kramers' rule, converging to a value near $500\ \mu\text{s}$ for both targets. This is not surprising, since $500\ \mu\text{s}$ is close to the expected timescale for exit of single stranded DNA probe in this range of potentials [141]. It is

therefore likely that at large negative potentials, the energy barrier is too low for Kramers' rule to apply [112], and probe exit, rather than duplex dissociation is the rate limiting step.

Kramers' rule also predicts that τ^* will increase exponentially with ΔG_b° , the dissociation energy in the absence of force. To test this, multi-pore force spectroscopy experiments were performed with the Synth probe against a six targets with differing sequences and melting energies (see Table 5-1). Characteristic timescales vs. applied potential for these experiments are shown in the Figure 5-7.

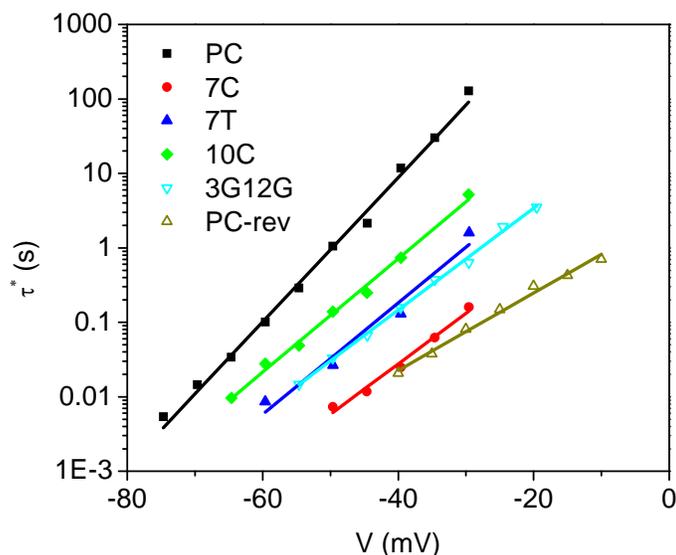


Figure 5-7 Multi-pore force spectroscopy duplex dissociation characteristic timescale vs. applied potential for the Synth probe with six different targets (see Table 5-1). Each point represents a single experiment containing *ca.* 500 single molecule dissociation events. For clarity, only data well within the regime that follows Kramers' rule ($\tau^* > 0.005$ s) is shown. Note characteristic timescales were re-calculated from data originally published by Tropini & Marziali in [35] using analysis methods described above, which differ from those published in the paper. Exponential fits (5-10) to the data were calculated by minimizing logarithmic least squares error.

Results in Figure 5-7 show a qualitative correlation between melting energy (see Table 5-1) and characteristic timescale; however, it is not immediately clear whether timescales increase exponentially with ΔG_b° . To confirm this relation, we extrapolate fits in Figure 5-7 to 0 mV to estimate the characteristic timescale at zero force, and plot this

vs. ΔG_{melt} , the melting energy (see Table 5-1). ΔG_{melt} is taken as an approximation of ΔG_b° .³⁶ The results, shown in Figure 5-8, follow the predicted exponential relationship surprisingly well, given that both the characteristic timescales and the energies plotted are approximations.

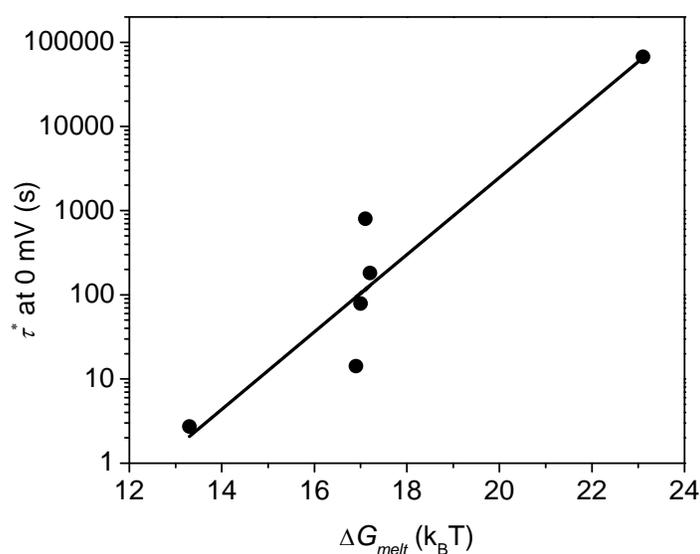


Figure 5-8 Dissociation timescale extrapolated to zero force vs. melting energy for nanopore force spectroscopy of DNA duplexes. Note that dissociation timescales were extracted from data in Figure 5-7.

³⁶ Note that ΔG_{melt} is not necessarily equal to ΔG_b° . Many reaction pathways are available for melting (e.g. target strand beginning to dissociate at the 5' end or 3' end). In a force spectroscopy experiment, the force selectively lowers only the energy barriers for only a subset of these pathways, which contribute to ΔG_b° (e.g. target strand beginning to dissociate at the 3' end, but not the 5' end).

Thus, these studies yield two important results. First, characteristic timescales follow predictions made from kinetic theory, confirming that DNA duplex dissociation can be studied by nanopore force spectroscopy. Second, and more importantly, the characteristic timescale forms an extremely sensitive indicator of probe-target sequence homology. This suggests that, under appropriate conditions, base calling with single nucleotide accuracy should be straightforward.

Note, however, that in order to maximize differences in τ^* for different alleles in base calling, two conditions must be satisfied. First, data should be collected at relatively low potentials, and second, the mismatch should be located near the middle of the duplex, as demonstrated by single pore force spectroscopy experiments in which mismatches located at the ends of the duplex were found to have little effect on the characteristic timescale [34]. Neither of these conditions are particularly difficult to satisfy. We therefore move on to apply these results to base-calling by nanopore force spectroscopy.

Chapter 6 Base-Calling by Nanopore Force Spectroscopy

We now turn our attention to the main goal of this thesis: base calling. In this chapter, methods for determining whether a target sequence of interest is present in the sample are developed and demonstrated using synthetic DNA samples simulating both homozygous and heterozygous genotypes. Note that while this work lays the groundwork for clinical diagnostic genotyping, some additional development will be required before the feasibility of genotyping complex samples containing genomic DNA can be demonstrated. This left for future studies, as described in Chapter 7.

Regardless, the results presented in this chapter clearly show that base-calling is possible for simple samples, and are extremely encouraging for further development.

6.1 Methods

For the studies in this chapter, the Synth probe was used to base-call a simulated polymorphism, for which the two ‘alleles’ corresponded to PC (perfect complement), and 8C (single nucleotide mismatch, see Table 5-1). Note that the polymorphism was located in the middle of the duplex, to maximize the difference in kinetics between the two targets.

Simulated genotypes were created by mixing PC and 8C targets at ratios 1:0, 0:1, and 1:1, corresponding to G/G, C/C, and G/C genotypes, respectively. The total target DNA concentration was fixed at 500 nM for all studies, and no genomic DNA was present³⁷. Note that for genotyping of real-world samples, these tests would be duplicated using a probe complementary to the 8C target for unambiguous genotype assignment; however, this was not performed for the proof-of concept studies described here.

Experimental procedures, including initial setup, data acquisition, and raw data analysis followed protocols described in section 5.1. The dissociation potential for these experiments was - 50 mV, which was predicted to give large differences in kinetics between the two targets, based on results from Chapter 5.

6.2 Results & Data Analysis

The event time distributions for these three samples are shown in Figure 6-1, and are all easily distinguishable from one another. The two homozygote datasets are very well separated, with characteristic timescales matching previous results, as discussed below. The

³⁷ Note, however, that an experiment demonstrating that limited quantities of unrelated DNA sequences have no effect on kinetics is described in Appendix F.

heterozygote dataset displays two distinct timescales, corresponding to characteristics for the 8C (fast) and PC (slow) targets.

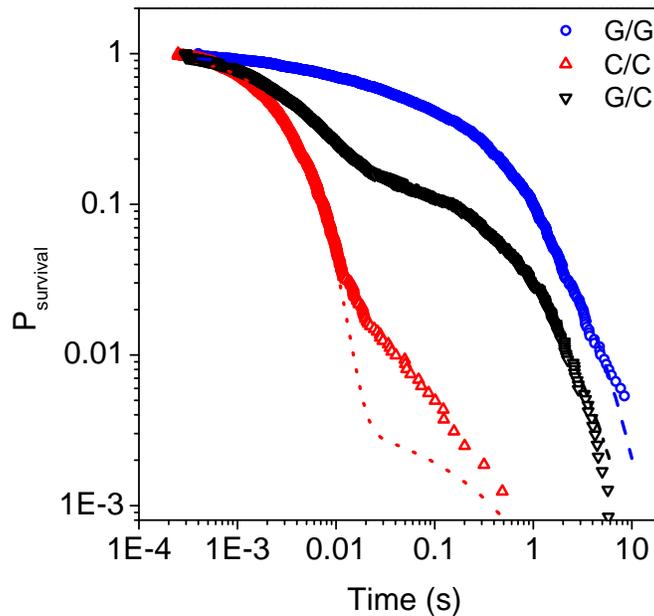


Figure 6-1 Event time distributions for NFS base calling. The 3 single pore force spectroscopy data sets are for the Synth probe with PC (genotype G/G), 8C (C/C) and a 1:1 mix of PC:8C (G/C) at a -50 mV dissociation potential. Note that the C/C dataset deviates from expected kinetics at times greater than 0.01s. The reason for this deviation is unknown. Fits shown are to a 2-term stretched exponential (6-1), as described below.

In this section, we develop and test two related base calling methods. The first, and simplest of these, determines the fraction of dissociation events attributable to the desired target. The second method extends this analysis using knowledge of hybridization

kinetics to improve base-calling results, and could be further developed to measure target concentrations, as discussed below.

Both analyses begin by analyzing the event time distribution, in order to determine the fraction of events attributable to the desired target sequence. The first step is fitting the event time distribution to a two-term exponential function³⁸, of the form:

$$P_{survival} = A_T \exp\left[(-t/\tau_T)^{\alpha_T}\right] + A_M \exp\left[(-t/\tau_M)^{\alpha_M}\right] \quad (6-1)$$

The weighting factor A_T represents the fraction of events attributable to the desired target. Kinetic parameters τ_T and α_T are fixed to average values from previous experiments, noting that data for base-calling experiments were excluded when calculating these averages. All other parameters are fitted to the data, making no assumptions about partially complementary (mismatched) DNA sequences contributing to A_M , τ_M , and α_M . Note, however, that the constraint $A_T + A_M = 1$ was applied to ensure that survival probability at the start time was 1, as predicted from kinetic theory.

³⁸ Data acquisition and analysis methods for this case may require modifications to account for the presence of large amounts of partially complementary DNA. Data analysis could be modified, either by discarding very fast dissociation events prior to fitting, or by including additional terms in (6-1). Alternatively, data acquisition could be modified by inclusion of a low-force dissociation step, which would selectively reject partially complementary sequences prior to the force spectroscopy trial.

Base-calling in this case consists of determining whether A_T exceeds some threshold, indicating that the target is present in the sample. Results of this analysis are presented below. However, we first derive an extension of this analysis method to which improves base calling, and may be capable of measuring target concentrations.

This method uses knowledge of hybridization kinetics to determine target concentrations from the weighting factors A_T and A_M . The simplest case of this analysis assumes that target concentrations are low, making hybridization a non-competitive process. This is not the case for the experiments above, where target concentrations are high enough that hybridization is competitive³⁹. Therefore, we begin by deriving an expression for target concentration in the non-competitive case, and then extend this analysis to the competitive case, which can be used to analyze our data.

Assuming that hybridization is a first order non-competitive reaction, the hybridization rate will be:

$$\frac{dP_{PT}}{dt} = k_{PT}[T](1 - P_{PT}) \quad (6-2)$$

³⁹ Note, however, that hybridization kinetics are expected to be non-competitive in a mature instrument, where millions of pores will be used to analyze DNA at low concentrations.

where P_{PT} is the fraction of all captured probes which form probe duplexes with the perfectly complementary target, k_{PT} is the hybridization rate, and $[T]$ is the target concentration. Integrating, we obtain:

$$P_{PT} = 1 - e^{-k_{PT}[T]t} \quad (6-3)$$

where t is the hybridization time. This can be rearranged to calculate the target concentration⁴⁰:

$$[T] = -\frac{\ln(1 - P_{PT})}{k_{PT}t} \quad (6-4)$$

Tests of this analysis method will require a large pore array, which is not currently available, and measurements of hybridization kinetics over a range of target concentrations, which were not performed here⁴¹. This is therefore left for future work. We can, however, use this to develop a related analysis technique which is applicable to current studies. This method uses knowledge of hybridization kinetics to improve base-calling results by determining the target concentration relative to other molecules in the sample.

⁴⁰ Note that (6-4) could potentially be used to measure target concentrations, which is useful for detecting genomic copy number variations.

⁴¹ Note, however, that hybridization kinetics were measured for both targets at 500 nM concentration. These results are described in F.1.2.

Under the conditions of the experiments performed here, target concentrations are relatively high, making hybridization competitive. Hybridization rates for the molecules are sequence specific, as demonstrated in Appendix F, which affects the proportion of events involving desired target and mismatched molecules. However, by measuring binding kinetics for the targets, we can recover the mole-fraction of each molecule in the sample. We therefore begin by describing hybridization kinetics with a set of coupled differential equations:

$$\frac{dP_{Pu}}{dt} = -k_{PT}[T]P_{Pu} - k_{PM}[M]P_{Pu} \quad (6-5)$$

$$\frac{dP_{PT}}{dt} = k_{PT}[T]P_{Pu} \quad (6-6)$$

$$\frac{dP_{PM}}{dt} = k_{PM}[M]P_{Pu} \quad (6-7)$$

where the subscripts PT , PM , and Pu denote probe-target duplexes, probe-mismatch duplexes, and unhybridized probes, respectively. Note that $[T]$ and $[M]$ measure DNA concentrations, while P_{PT} , P_{PM} , and P_{Pu} denote fractions of the total probes present. Hybridization probabilities are thus subject to the constraint:

$$P_{PT} + P_{PM} + P_{Pu} = 1 \quad (6-8)$$

To obtain the relative concentration of the perfectly complementary target in the sample requires further derivation. We calculate the

fraction of unhybridized probes as a function of time by integrating (6-5) and applying the boundary condition that $P_{Pu} = 1$ at $t = 0$, obtaining:

$$P_{Pu} = e^{-k_{PT}[T]t - k_{PM}[M]t} \quad (6-9)$$

Next, we perform similar a similar integration for probe-target duplexes, noting that the P_{Pu} is substituted from (6-9):

$$\begin{aligned} \frac{dP_{PT}}{P_{PT}} &= \int k_{PT}[T] e^{-k_{PT}[T]t - k_{PM}[M]t} dt \\ P_{PT} &= \frac{k_{PT}[T]}{k_{PT}[T] + k_{PM}[M]} \left(1 - e^{-k_{PT}[T]t - k_{PM}[M]t}\right) \end{aligned}$$

Simplifying by substituting P_{Pu} from (6-9) we obtain:

$$P_{PT} = \frac{k_{PT}[T]}{k_{PT}[T] + k_{PM}[M]} (1 - P_{Pu}) \quad (6-10)$$

Note, however, that $1 - P_{Pu} = P_{PM} + P_{PT}$, from (6-8). Applying this to (6-10) gives:

$$P_{PT} = \frac{k_{PT}[T]}{k_{PT}[T] + k_{PM}[M]} (P_{PT} + P_{PM}) \quad (6-11)$$

We now wish to solve this equation, in order to determine the relative allele concentration from the fraction of events attributable to perfect match, and mismatch targets. This is done by performing some algebra on (6-11) as follows:

$$P_{PT} = \frac{k_{PT}[T]}{k_{PT}[T] + k_{PM}[M]} (P_{PT} + P_{PM})$$

$$k_{PM}[M]P_{PT} = k_{PT}[T]P_{PM}$$

$$k_{PM}[M]P_{PT} + k_{PM}[T]P_{PT} = k_{PT}[T]P_{PM} + k_{PM}[T]P_{PT}$$

$$\frac{[T]}{[T] + [M]} = \frac{k_{PM}P_{PT}}{k_{PM}P_{PT} + k_{PT}P_{PM}}$$

Noting that P_{PT} and P_{PM} are proportional to A_T and A_M (the fraction of dissociation events attributable to each of the two target molecules), we can make a substitution to obtain the final form:

$$\frac{[T]}{[T] + [M]} = \frac{k_{PM}A_M}{k_{PM}A_T + k_{PT}A_M} \quad (6-12)$$

The relative concentration of the mismatched allele follows a similar derivation.

We now proceed to apply both base-calling algorithms developed above to the data in Figure 6-1, beginning with determination of the fraction of dissociation events attributable to the various target molecules in the samples.

Figure 6-1 shows fits to the two-term stretched exponential (6-1), performed by minimizing the least squares error between the data and the fit. From the fits, we extract the fraction of events attributed to the desired target, with the results shown in Figure 6-2. In the G/G (PC) and C/C (8C) the algorithm assigns over 95% and less than 1% of

observed events to PC, respectively. In the G/C heterozygote, *ca.* 30% of events are attributed to the target, which less than the expected 50%, but sufficient to unequivocally state that the G allele is present in the sample, and that the sample is likely a heterozygote.

It should be noted that fraction events attributable to the perfect complement varied considerably (by up to $\pm 15\%$) from experiment to experiment, the causes of which variation are unclear (see F.2). The data shown in Figure 6-2 is a typical dataset, since the average proportion of events assigned to the perfectly complementary target was 30%.

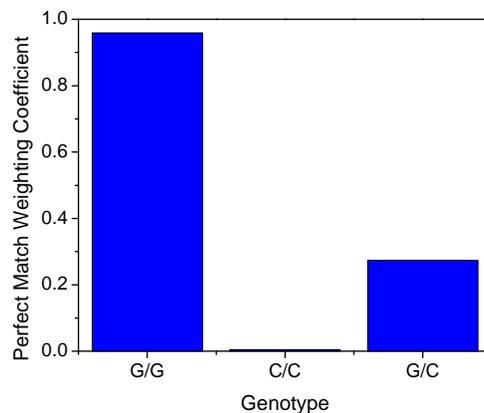


Figure 6-2 Proportion of events attributed to the perfect complement allele for three possible genotypes. The G/G genotype is homozygous for the desired allele.

Figure 6-3 shows that the values for τ_M^* in the C/C and G/C samples extracted from the fits agree well with the Synth2-8C

characteristic timescale, as expected. τ^*_M for the G/G sample differs from the expected value by a factor of approximately 3; however this is not expected to agree with the characteristic timescale for 8C dissociation, since this is an artifact arising from imperfect fitting of the stretched exponential decay to the data. Note that in the G/G case, less than 5% of the events are attributed to mismatched targets.

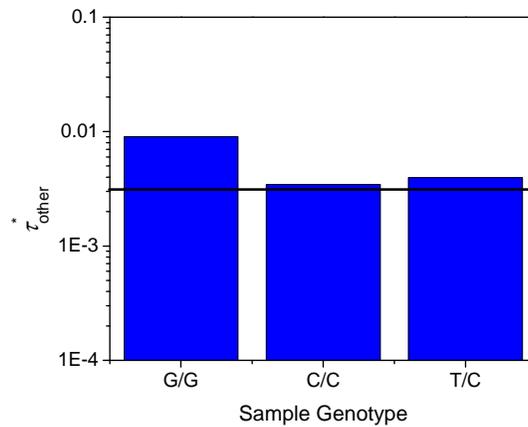


Figure 6-3 Characteristic timescale τ^*_{other} from equation (6-1) determined by NFS genotyping. In the C/C and G/C datasets, τ^*_{other} is predicted to match the known characteristic timescale for the 8C DNA sequence, denoted by a black line.

We now move on to calculate the mole fraction of each target in the sample, noting before we do, that in the proposed instrument, a separate test will be performed for each allele. Since one would expect a positive call for both alleles in a heterozygote, this information could be used to improve base-calling results by taking advantage of

sequence-specific hybridization rates. We can mimic that analysis here, where tests were performed for only one of the two alleles, by assuming that we have identified the other sequence. Note that this is a reasonable assumption, given that the characteristic timescales in Figure 6-3 can be used to identify the mismatch in these samples. We therefore analyze the data with (6-12), using sequence-specific hybridization information presented in F.1.2, to determine the mole fraction of each target in the sample. Results of this analysis are shown in Figure 6-4.

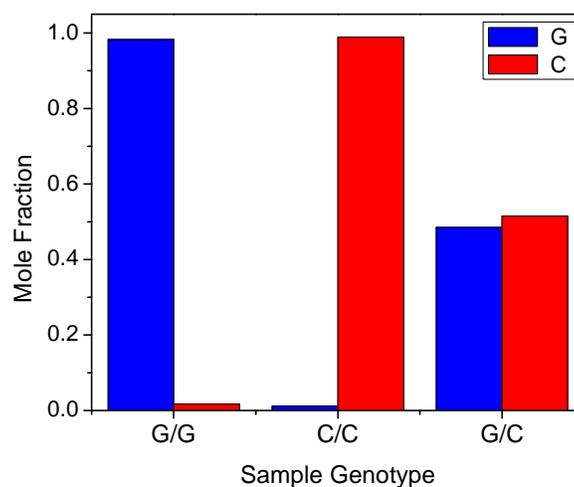


Figure 6-4 Relative concentration of alleles calculated for three possible genotypes, as calculated by (6-12). Data is presented as calculated mole fractions of PC (G) and 8C (C) targets.

This analysis dramatically improves the agreement to expected values – the fraction of each target deviates from the expected values

by less than 2% for the data shown. It should be noted that mole fractions for each target in heterozygous samples were observed to vary by up to 15% from experiment to experiment, similar to the previous analysis. However, calculation of relative concentrations did improve results considerably over the previous analysis, since the average mole calculated fraction for the perfect complement was 49.9%, in contrast to the previous analysis, which found an average of 30% of events attributable to the perfect complement.

6.3 Discussion

The main conclusion of this section is that base calling is feasible by nanopore force spectroscopy, under the conditions tested. It should be noted that these experiments represent the simplest case, since only two unique sequences were present in the sample, and the DNA concentration was high. Further tests are still required to demonstrate genotyping in complex samples, such as genomic DNA. However, even in the presence of genomic DNA, the two alleles of a single nucleotide polymorphism should be the most difficult to distinguish from one another. Given the results above, and the possibility of rejecting distantly related DNA sequences from the analysis, extension of this process to a real-world DNA sample may be possible, following the studies outlined in Chapter 7.

Chapter 7 - Future Studies

As noted throughout this thesis, further technological improvements and tests are required to demonstrate genotyping from genomic DNA samples, prior to the development of a commercial instrument.

The first of these studies, already under way in the Marziali lab, is evaluation of solid state nanopores, which carry the advantage that they could easily be incorporated into a disposable chip for clinical genotyping. Preliminary probe escape and duplex dissociation results using solid state pores are similar to results presented in this dissertation, including non-ideal event time distributions, and single base specificity of dissociation timescales. However, results are currently difficult to reproduce in these experiments, since solid state pores frequently enter a stable, partially blocked state in which it is nearly impossible to capture targets. This is presumably due to adsorption of DNA or protein to the pore surface, which could be solved by chemically modifying the pore surface.

Regardless of whether solid state or α -HL pore based instruments are chosen for further development, the studies that follow are similar. Some of these studies, which could be performed using single pores, would improve data analysis through understanding kinetics and mechanisms involved in nanopore force spectroscopy.

As discussed in Chapters 4 and 5, the functions used to fit event time distributions were chosen phenomenologically, since the underlying mechanisms leading to non-ideal behavior remain poorly understood. As a result, the fitting functions do not model duplex dissociation (or probe escape) kinetics exactly, leading to errors in data analysis. It is expected that these errors will become increasingly important in tests containing many DNA sequences partially complementary to the probe. Therefore, further research into the causes of non-ideal behavior is very important. For example, studies attempting to modify or suppress these interactions by chemically modifying the pore could lead to improved fitting algorithms, or recovery of ideal behavior in nanopore force spectroscopy experiments.

The bulk of future studies will require an instrument containing millions of pores, in order to test samples containing low DNA concentrations. This may not be overly difficult to achieve, since large arrays of both α -HL or solid state pores are currently being developed commercially. An instrument containing large arrays of α -HL pores is currently being developed by Oxford Nanopore Technologies [74] for the purpose of sequencing, which could conceivably be modified to permit genotyping. On the other hand, silicon membranes containing millions of nanopores are already commercially available [154], which could be incorporated into a disposable chip for nanopore force spectroscopy.

Once an instrument containing millions of pores is created, its performance will need to be evaluated under increasingly stringent conditions. Early tests will focus on the sensitivity, specificity, and speed of the sensor, by performing base-calling on samples containing decreasing concentrations of chemically synthesized DNA, with the goal of demonstrating base-calling at concentrations simulating unamplified genomic DNA. Calculations suggesting that this may be feasible are presented here.

Ultimately, the sensitivity of the genotyping instrument will depend on its ability to observe ionic current changes accompanying DNA dissociation and probe exit from a small fraction of the pores. Though not measured experimentally, it can be argued that the signal to noise ratio of the sensor will be sufficient to allow genotyping from unamplified genomic DNA. Assuming a dissociation potential of -50 mV in 1 M KCl, the current change in each pore will be $\Delta I \approx 20\text{pA}$, as estimated from current changes observed in α -HL. The total current change is therefore expected to be:

$$\Delta I_{Total} = P_{PT} N_p \Delta I \quad (6-13)$$

where the product of the hybridization probability for the probe, P_{PT} , and the total number of pores, N_p , gives the number of pores containing target-hybridized probes. Assuming 10^6 pores, and $P_{PT} = 2.5 \times 10^{-4}$, (estimated in Appendix F), the total signal will be $\Delta I_{Total} \approx 10\mu\text{A}$.

The ionic current noise is expected to be dominated by thermal noise, which arises from random motion of current carriers, in this case, ions within the pore. It is calculated as [155]:

$$I_{RMS} = \sqrt{\frac{4k_B TBN}{R}} \quad (6-14)$$

where I_{RMS} is the root-mean square of the current (i.e. noise), B is the bandwidth, assumed to be 100 Hz. R is the measured resistance of a single pore, taken as 1G Ω . For these values, $I_{RMS} \approx 100$ pA. Thus, the total signal-to-noise ratio is:

$$S.N.R = \frac{\Delta I_{Total}}{I_{RMS}} \quad (6-15)$$

For the parameters listed above, the signal-to-noise ratio is approximately 250. While this clearly needs to be tested, it is extremely encouraging, suggesting that an array of 10^6 pores could easily detect targets at femto-molar concentration.

Following verification of the above result, it will be important to examine the effects of partially complementary DNA in genomic samples. These could initially be simulated using large amounts of mixed DNA sequences capable of interacting with the probe. Results of this work will be used to optimize experimental protocols - *e.g.* suppressing binding of non-target sequences using elevated temperature, or using a low-force dissociation step prior to the force

spectroscopy trial. This work will also determine whether genotyping in less than 1 hour from unamplified DNA is feasible.

This work will culminate in genotyping of DNA extracted from biological tissue. Ideally, no PCR amplification will be required, which would represent a dramatic improvement over currently available technologies. However, even an instrument that requires a few rounds of PCR could still be faster, less sensitive to error, and therefore, more appealing to clinicians than available clinical genotyping platforms.

Chapter 8 - Conclusion

The nanopore force spectroscopy technique developed in this dissertation shows great potential as a rapid clinical genotyping technique. From the proof-of-concept studies presented in this thesis, we can make a number of conclusions regarding its basic principles, and potential for clinical diagnostic applications.

First, the basic genotyping scheme for nanopore force spectroscopy carries some significant advantages over currently available genotyping methods. Detection is entirely electrical, meaning that sample DNA need not be labeled. In addition, NFS is able to actively discriminate between target molecules which bind to the probes, giving it an exceptional degree of sequence specificity.

Relatively sophisticated data analysis methods are required to interpret nanopore force spectroscopy data, owing to the fact that DNA dissociation kinetics in nanopore force spectroscopy do not obey Poisson kinetics. Despite this non-ideal behavior, it is possible to extract timescales which appropriately characterize the data and obey Kramers' rule, using the data processing algorithms developed in Chapter 5.

The success of the experimental work presented in the latter portions of this thesis suggests that clinical diagnostic genotyping by nanopore force spectroscopy is possible. Tests using a variety of DNA

probes and targets show that even single nucleotide mismatches can have a profound effect on NFS dissociation timescales. Characteristic timescales can differ by over a factor of 100 for different alleles of a single nucleotide polymorphism. With this degree of sequence specificity, it is straightforward to unambiguously determine whether a given allele is present in a chemically synthesized DNA sample. NFS base calling is also extremely sensitive. At a conservative estimate, a target can be identified from as few as 100 successful probe-target dissociation events.

If similar sensitivity can be retained in the presence of genomic DNA, and preliminary calculations of probe-target hybridization rates turn out to be correct, purified genomic DNA could be genotyped in less than one hour, without the need for PCR, by nanopore force spectroscopy. However, considerable further development is still required to develop a working prototype of a commercial NFS genotyping instrument. In particular, an instrument employing millions of pores in parallel must be developed and tested, particularly with regards to the detection limit and specificity in the presence of genomic DNA.

In conclusion, the successful results presented in this thesis indicate that further development of a commercial NFS genotyping instrument is clearly warranted. This instrument would be rapid, potentially delivering results in less than one hour. In addition, it would be extremely simple, requiring only that DNA be extracted from a sample

and loaded into a disposable cartridge for testing. These advantages would make NFS genotyping ideal for rapid, point-of-care use in a hospital. Given that one of the greatest barriers to widespread adoption of clinical diagnostic genotyping is the lack of suitable instruments, genotyping by nanopore force spectroscopy has potential to dramatically improve human health.

Bibliography

1. Redon, R., et al., *Global variation in copy number in the human genome*. *Nature*, 2006. **444**(7118): p. 444-454.
2. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome*. *Science*, 2004. **305**(5683): p. 525-528.
3. Kruglyak, L. and D.A. Nickerson, *Variation is the spice of life*. *Nature Genetics*, 2001. **27**(3): p. 234-236.
4. Grossman, I., *Routine pharmacogenetic testing in clinical practice: dream or reality?* *Pharmacogenomics*, 2007. **8**(10): p. 1449-1459.
5. Seng, K.C. and C.K. Seng, *The success of the genome-wide association approach: a brief story of a long struggle*. *European Journal of Human Genetics*, 2008. **16**(5): p. 554-564.
6. Weinshilboum, R., *Genomic medicine - Inheritance and drug response*. *New England Journal of Medicine*, 2003. **348**(6): p. 529-537.
7. Wilkinson, G.R., *Drug therapy - Drug metabolism and variability among patients in drug response*. *New England Journal of Medicine*, 2005. **352**(21): p. 2211-2221.
8. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. *Nature*, 2001. **409**(6822): p. 860-921.
9. Venter, J.C., et al., *The sequence of the human genome*. *Science*, 2001. **291**(5507): p. 1304-+.
10. Altshuler, D., et al., *A haplotype map of the human genome*. *Nature*, 2005. **437**(7063): p. 1299-1320.
11. *NIH Launches Comprehensive Effort to Explore Cancer Genomics (Press Release)*. [web page] 2005 [cited 2008 Aug 1]; Available from:
http://cancergenome.nih.gov/media/news_12_13_2005.asp.
12. Spear, B.B., M. Heath-Chiozzi, and J. Huff, *Clinical Applications of Pharmacogenetics*. *Trends in Molecular Medicine*, 2001. **7**(5): p. 201-204.

13. Evans, W.E. and H.L. McLeod, *Drug therapy - Pharmacogenomics - Drug disposition, drug targets, and side effects*. New England Journal of Medicine, 2003. **348**(6): p. 538-549.
14. *F.D.A. device and test approval for Verigene system and Verigene warfarin metabolism nucleic acid test*, U.S. F.D.A, Editor. 2007.
15. *AmpliChip CYP450 Test*. 2008 2007-10 [cited 2008 Sept. 17]; Available from:
http://www.amplichip.us/documents/CYP450_P.I._US-IVD.pdf.
16. *Pathvysion HER-2 Fluorescence In Situ Hybridization Test*. [web page] 2008 [cited 2008 October 21]; Available from:
<http://www.pathvysion.com>.
17. Pegram, M.D., et al., *Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185(HER2/neu) monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment*. Journal of Clinical Oncology, 1998. **16**(8): p. 2659-2671.
18. Ross, J.S. and J.A. Fletcher, *The HER-2/neu oncogene in breast cancer: Prognostic factor, predictive factor, and target for therapy*. Stem Cells, 1998. **16**(6): p. 413-428.
19. Slamon, D.J., et al., *Human-Breast Cancer - Correlation of Relapse and Survival with Amplification of the Her-2 Neu Oncogene*. Science, 1987. **235**(4785): p. 177-182.
20. Ingelman-Sundberg, M., M. Oscarson, and R.A. McLellan, *Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment*. Trends in Pharmacological Sciences, 1999. **20**(8): p. 342-349.
21. Kumar, A., et al., *Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Critical Care Medicine, 2006. **34**(6): p. 1589-1596.
22. Winning, J., et al., *Molecular biology on the ICU - From understanding to treating sepsis*. Minerva Anestesiologica, 2006. **72**(5): p. 255-267.

23. Majetschak, M., et al., *Relation of a TNF gene polymorphism to severe sepsis in trauma patients*. *Annals of Surgery*, 1999. **230**(2): p. 207-214.
24. Mira, J.P., et al., *Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality - A multicenter study*. *Jama-Journal of the American Medical Association*, 1999. **282**(6): p. 561-568.
25. Russell, J.A., H. Wellman, and K.R. Walley, *Protein C rs2069912 C allele is associated with increased mortality from severe sepsis in North Americans of East Asian ancestry*. *Human Genetics*, 2008. **123**(6): p. 661-663.
26. Walley, K.R. and J.A. Russell, *Protein C-1641 AA is associated with decreased survival and more organ dysfunction in severe sepsis*. *Critical Care Medicine*, 2007. **35**(1): p. 12-17.
27. Johnson, M.A., M.J. Yoshitomi, and C.S. Richards, *A comparative study of five technologically diverse CFTR testing platforms*. *Journal of Molecular Diagnostics*, 2007. **9**(3): p. 401-407.
28. Collins, F., *Personal Communication to Dr. Andre Marziali at International Conference on Genomics*. 2006: Hangzhou, China.
29. Isler, J.A., O.E. Vesterqvist, and M.E. Burczynski, *Analytical validation of genotyping assays in the biomarker laboratory*. *Pharmacogenomics*, 2007. **8**(4): p. 353-368.
30. Walley, K.R., MD, Professor, University of British Columbia, Vancouver. *Personal communication to Dr. Andre Marziali*. 2006.
31. Willis, T., PhD Former Chief Scientific Officer and co-founder of ParAllele. *Personal communication to Dr. Andre Marziali*. 2006.
32. *F.D.A. test approval for Tag-it Cystic Fibrosis test kit*, U.S. F.D.A., Editor. 2007 June 7
33. Pare, P.D., Walley, K.R., and Tebbutt, S.J., *Personal communication to Dr. Andre Marziali*, A. Marziali, Editor. 2007: Vancouver.
34. Nakane, J., M. Wiggin, and A. Marziali, *A nanosensor for transmembrane capture and identification of single nucleic acid molecules*. *Biophysical Journal*, 2004. **87**(1): p. 615-621.

35. Tropini, C. and A. Marziali, *Multi-nanopore force Spectroscopy for DNA analysis*. Biophysical Journal, 2007. **92**(5): p. 1632-1637.
36. Kim, S. and A. Misra, *SNP genotyping: Technologies and biomedical applications*. Annual Review of Biomedical Engineering, 2007. **9**: p. 289-320.
37. Kwok, P.Y., *Methods for genotyping single nucleotide polymorphisms*. Annual Review of Genomics and Human Genetics, 2001. **2**: p. 235-258.
38. Tsongalis, G.J. and W.B. Coleman, *Clinical genotyping: The need for interrogation of single nucleotide polymorphisms and mutations in the clinical laboratory*. Clinica Chimica Acta, 2006. **363**(1-2): p. 127-137.
39. *FDA Clears First of Kind Genetic Lab Test*, U.S. F.D.A, Editor. 2004.
40. *F.D.A. device approval for Luminex Lx 100/200 instrument*, U.S. F.D.A., Editor. 2008 March 7.
41. *Luminex XMap Technology*. [web page] [cited 2008 October 3]; Available from:
<http://www.luminexcorp.com/technology/index.html>.
42. Iannone, M.A., et al., *Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry*. Cytometry, 2000. **39**: p. 131-140.
43. *Nanosphere Verigene technology*. [web page] [cited 2008 October 3]; Available from:
http://www.nanosphere.us/DirectDetectionofNucleicAcids_4401.aspx.
44. *IQuum's Liat Analyzer*. [web page] 2006 [cited 2007 December 5]; Available from:
<http://www.iquum.com/products/analyzer.shtml>.
45. *Third Wave Technologies' Invader Technology*. [web page] 2008 [cited 2008 October 21]; Available from:
<http://www.twt.com/invader/invader.html>.
46. *23andMe*. [web page] 2009 [cited 2009 March 27]; Available from: <https://www.23andme.com/>.

47. Wadman, M., *Genetics bill cruises through Senate*. *Nature*, 2008. **453**(7191): p. 9-9.
48. Radstrom, P., et al., *Pre-PCR processing - Strategies to generate PCR-compatible samples*. *Molecular Biotechnology*, 2004. **26**(2): p. 133-146.
49. Chen, X. and P.F. Sullivan, *Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput*. *Pharmacogenomics Journal*, 2003. **3**(2): p. 77-96.
50. Smeets, R.M.M., et al., *Salt dependence of ion transport and DNA translocation through solid-state nanopores*. *Nano Letters*, 2006. **6**(1): p. 89-95.
51. Chang, H., et al., *DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels*. *Nano Letters*, 2004. **4**(8): p. 1551-1556.
52. Kasianowicz, J.J., et al., *Characterization of individual polynucleotide molecules using a membrane channel*. *Proceedings of the National Academy of Sciences of the United States of America*, 1996. **93**(24): p. 13770-13773.
53. Akeson, M., et al., *Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules*. *Biophysical Journal*, 1999. **77**(6): p. 3227-3233.
54. Henrickson, S.E., et al., *Driven DNA transport into an asymmetric nanometer-scale pore*. *Physical Review Letters*, 2000. **85**(14): p. 3057-3060.
55. Meller, A. and D. Branton, *Single molecule measurements of DNA transport through a nanopore*. *Electrophoresis*, 2002. **23**(16): p. 2583-2591.
56. Meller, A., et al., *Rapid nanopore discrimination between single polynucleotide molecules*. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(3): p. 1079-1084.

57. Meller, A., L. Nivon, and D. Branton, *Voltage-driven DNA translocations through a nanopore*. *Physical Review Letters*, 2001. **86**(15): p. 3435-3438.
58. Kasianowicz, J.J., et al., *Simultaneous multianalyte detection with a nanometer-scale pore*. *Analytical Chemistry*, 2001. **73**(10): p. 2268-2272.
59. Menestrina, G., *Ionic Channels Formed by Staphylococcus-Aureus Alpha-Toxin - Voltage-Dependent Inhibition by Divalent and Trivalent Cations*. *Journal of Membrane Biology*, 1986. **90**(2): p. 177-190.
60. Song, L.Z., et al., *Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore*. *Science*, 1996. **274**(5294): p. 1859-1866.
61. Li, J., et al., *Ion-beam sculpting at nanometre length scales*. *Nature*, 2001. **412**(6843): p. 166-169.
62. Storm, A.J., et al., *Fabrication of solid-state nanopores with single-nanometre precision*. *Nature Materials*, 2003. **2**(8): p. 537-540.
63. Chen, L.M., et al., *Fast fabrication of large-area nanopore arrays by FIB*. *Acta Physica Sinica*, 2005. **54**(2): p. 582-586.
64. Tabard-Cossa, V., et al., *Noise analysis and reduction in solid-state nanopores*. *Nanotechnology*, 2007. **18**(30).
65. Li, J.L., et al., *DNA molecules and configurations in a solid-state nanopore microscope*. *Nature Materials*, 2003. **2**(9): p. 611-615.
66. Tabard-Cossa, V., et al., *Noise analysis and reduction in solid-state nanopores*. *Nanotechnology*, 2007. **18**(30): p. 305505.
67. Storm, A.J., et al., *Fast DNA translocation through a solid-state nanopore*. *Nano Letters*, 2005. **5**(7): p. 1193-1197.
68. Muthukumar, M., *Polymer translocation through a hole*. *Journal of Chemical Physics*, 1999. **111**(22): p. 10371-10374.
69. Butler, T.Z., J.H. Gundlach, and M.A. Troll, *Determination of RNA orientation during translocation through a biological nanopore*. *Biophysical Journal*, 2006. **90**(1): p. 190-199.

70. Butler, T.Z., J.H. Gundlach, and M. Troll, *Ionic current blockades from DNA and RNA molecules in the alpha-hemolysin nanopore*. *Biophysical Journal*, 2007. **93**(9): p. 3229-3240.
71. Deamer, D.W. and M. Akeson, *Nanopores and nucleic acids: prospects for ultrarapid sequencing*. *Trends in Biotechnology*, 2000. **18**(4): p. 147-151.
72. Cockroft, S.L., et al., *A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution*. *Journal of the American Chemical Society*, 2008. **130**(3): p. 818-+.
73. Benner, S., et al., *Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore*. *Nature Nanotechnology*, 2007. **2**(11): p. 718-724.
74. *Oxford Nanopore - Nanopore Sequencing*. [web page] 2008 [cited 2008 October 21]; Available from: <http://www.nanoporetech.com/sequences>.
75. Astier, Y., O. Braha, and H. Bayley, *Toward single molecule DNA sequencing: Direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter*. *Journal of the American Chemical Society*, 2006. **128**(5): p. 1705-1710.
76. Zwolak, M. and M. Di Ventra, *Electronic signature of DNA nucleotides via transverse transport*. *Nano Letters*, 2005. **5**(3): p. 421-424.
77. Branton, D., et al., *The potential and challenges of nanopore sequencing*. *Nature Biotechnology*, 2008. **26**(10): p. 1146-1153.
78. Lagerqvist, J., M. Zwolak, and M. Di Ventra, *Fast DNA sequencing via transverse electronic transport*. *Nano Letters*, 2006. **6**(4): p. 779-782.
79. Gracheva, M.E., A. Aksimentiev, and J.P. Leburton, *Electrical signatures of single-stranded DNA with single base mutations in a nanopore capacitor*. *Nanotechnology*, 2006. **17**(13): p. 3160-3165.
80. Zwolak, M. and M. Di Ventra, *Colloquium: Physical approaches to DNA sequencing and detection*. *Reviews of Modern Physics*, 2008. **80**(1): p. 141-165.

81. Vercoutere, W.A., et al., *Discrimination among individual Watson-Crick base pairs at the termini of single DNA hairpin molecules*. Nucleic Acids Research, 2003. **31**(4): p. 1311-1318.
82. DeGuzman, V.S., et al., *Sequence-dependent gating of an ion channel by DNA hairpin molecules*. Nucleic Acids Research, 2006. **34**(22): p. 6425-6437.
83. Mathe, J., et al., *Equilibrium and irreversible unzipping of DNA in a nanopore*. Europhysics Letters, 2006. **73**(1): p. 128-134.
84. Mathe, J., et al., *Nanopore unzipping of individual DNA hairpin molecules*. Biophysical Journal, 2004. **87**(5): p. 3205-3212.
85. Svoboda, K. and S.M. Block, *Force and Velocity Measured for Single Kinesin Molecules*. Cell, 1994. **77**(5): p. 773-784.
86. Block, S.M., L.S.B. Goldstein, and B.J. Schnapp, *Bead Movement by Single Kinesin Molecules Studied with Optical Tweezers*. Nature, 1990. **348**(6299): p. 348-352.
87. Guo, B. and W.H. Guilford, *Mechanics of actomyosin bonds in different nucleotide states are tuned to muscle contraction*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(26): p. 9844-9849.
88. Tanaka, H., et al., *Orientation dependence of displacements by a single one-headed myosin relative to the actin filament*. Biophysical Journal, 1998. **75**(4): p. 1886-1894.
89. Nishizaka, T., et al., *Unbinding Force of a Single Motor Molecule of Muscle Measured Using Optical Tweezers*. Nature, 1995. **377**(6546): p. 251-254.
90. Wuite, G.J.L., et al., *Single-molecule studies of the effect of template tension on T7 DNA polymerase activity*. Nature, 2000. **404**(6773): p. 103-106.
91. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase*. Nature, 2005. **438**(7067): p. 460-465.
92. Kellermayer, M.S.Z., et al., *Folding-unfolding transitions in single titin molecules characterized with laser tweezers*. Science, 1997. **276**(5315): p. 1112-1116.

93. Rief, M., et al., *Reversible unfolding of individual titin immunoglobulin domains by AFM*. *Science*, 1997. **276**(5315): p. 1109-1112.
94. Tskhovrebova, L., et al., *Elasticity and unfolding of single molecules of the giant muscle protein titin*. *Nature*, 1997. **387**(6630): p. 308-312.
95. Fernandez, J.M. and H.B. Li, *Force-clamp spectroscopy monitors the folding trajectory of a single protein*. *Science*, 2004. **303**(5664): p. 1674-1678.
96. Dietz, H., et al., *Anisotropic deformation response of single protein molecules*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(34): p. 12724-12728.
97. Wright, C.F., et al., *Parallel protein-unfolding pathways revealed and mapped*. *Nature Structural Biology*, 2003. **10**(8): p. 658-662.
98. Merkel, R., et al., *Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy*. *Nature*, 1999. **397**(6714): p. 50-53.
99. Florin, E.L., V.T. Moy, and H.E. Gaub, *Adhesion Forces between Individual Ligand-Receptor Pairs*. *Science*, 1994. **264**(5157): p. 415-417.
100. Grubmuller, H., B. Heymann, and P. Tavan, *Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force*. *Science*, 1996. **271**(5251): p. 997-999.
101. EssevazRoulet, B., U. Bockelmann, and F. Heslot, *Mechanical separation of the complementary strands of DNA*. *Proceedings of the National Academy of Sciences of the United States of America*, 1997. **94**(22): p. 11935-11940.
102. Lee, G.U., L.A. Chrisey, and R.J. Colton, *Direct Measurement of the Forces between Complementary Strands of DNA*. *Science*, 1994. **266**(5186): p. 771-773.
103. Rief, M., H. Clausen-Schaumann, and H.E. Gaub, *Sequence-dependent mechanics of single DNA molecules*. *Nature Structural Biology*, 1999. **6**(4): p. 346-349.

-
104. Strunz, T., et al., *Dynamic force spectroscopy of single DNA molecules*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(20): p. 11277-11282.
 105. Evans, E., *Probing the relation between force - Lifetime - and chemistry in single molecular bonds*. Annual Review of Biophysics and Biomolecular Structure, 2001. **30**: p. 105-128.
 106. Woodside, M.T., et al., *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(16): p. 6190-6195.
 107. Liphardt, J., et al., *Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality*. Science, 2002. **296**(5574): p. 1832-1835.
 108. Woodside, M.T., et al., *Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid*. Science, 2006. **314**(5801): p. 1001-1004.
 109. Hanggi, P., P. Talkner, and M. Borkovec, *Reaction-Rate Theory - 50 Years after Kramers*. Reviews of Modern Physics, 1990. **62**(2): p. 251-341.
 110. Pollak, E. and P. Talkner, *Reaction rate theory: What it was, where is it today, and where is it going?* Chaos, 2005. **15**(2).
 111. Kramers, H.A., *Brownian motion in a field of force and the diffusion model of chemical reactions*. Physical Review E, 1940. **7**: p. 284-304.
 112. Evans, E. and K. Ritchie, *Dynamic strength of molecular adhesion bonds*. Biophysical Journal, 1997. **72**(4): p. 1541-1555.
 113. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*. Nature Methods, 2008. **5**(6): p. 491-505.
 114. Binnig, G., C.F. Quate, and C. Gerber, *Atomic Force Microscope*. Physical Review Letters, 1986. **56**(9): p. 930-933.
 115. Rief, M., et al., *Single molecule force spectroscopy on polysaccharides by atomic force microscopy*. Science, 1997. **275**(5304): p. 1295-1297.

-
116. *Ambios Technology, Inc.* [web page] 2008 [cited 2008 November 6]; Available from:
http://www.ambiosotech.com/atomic_force_microscope.htm.
117. *Novascan Technologies, Inc.* . [web page] 2008 [cited 2008 November 6]; Available from:
<http://www.novascan.com/index.php>.
118. Gosse, C. and V. Croquette, *Magnetic tweezers: Micromanipulation and force measurement at the molecular level*. *Biophysical Journal*, 2002. **82**(6): p. 3314-3329.
119. Smith, S.B., L. Finzi, and C. Bustamante, *Direct Mechanical Measurements of the Elasticity of Single DNA-Molecules by Using Magnetic Beads*. *Science*, 1992. **258**(5085): p. 1122-1126.
120. Strick, T., et al., *Twisting and stretching single DNA molecules*. *Progress in Biophysics & Molecular Biology*, 2000. **74**(1-2): p. 115-140.
121. Strick, T.R., V. Croquette, and D. Bensimon, *Single-molecule analysis of DNA uncoiling by a type II topoisomerase*. *Nature*, 2000. **404**(6780): p. 901-904.
122. Smith, S.B., Y. Cui, and C. Bustamante, *Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules*. *Science*, 1996. **271**: p. 795-799.
123. Ashkin, A., et al., *Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles*. *Optics Letters*, 1986. **11**(5): p. 288-290.
124. Davenport, R.J., et al., *Single-molecule study of transcriptional pausing and arrest by E-coli RNA polymerase*. *Science*, 2000. **287**(5462): p. 2497-2500.
125. Dame, R.T., M.C. Noom, and G.J.L. Wuite, *Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation*. *Nature*, 2006. **444**(7117): p. 387-390.
126. Peterman, E.J.G., F. Gittes, and C.F. Schmidt, *Laser-induced heating in optical traps*. *Biophysical Journal*, 2003. **84**(2): p. 1308-1316.

127. Cluzel, P., et al., *DNA: An extensible molecule*. *Science*, 1996. **271**(5250): p. 792-794.
128. Strick, T.R., et al., *The elasticity of a single supercoiled DNA molecule*. *Science*, 1996. **271**(5257): p. 1835-1837.
129. Liphardt, J., et al., *Reversible unfolding of single RNA molecules by mechanical force*. *Science*, 2001. **292**(5517): p. 733-737.
130. Hornblower, B., et al., *Single-molecule analysis of DNA-protein complexes using nanopores*. *Nature Methods*, 2007. **4**(4): p. 315-317.
131. Koch, S.J., et al., *Probing protein-DNA interactions by unzipping a single DNA double helix*. *Biophysical Journal*, 2002. **83**(2): p. 1098-1105.
132. Perkins, T.T., et al., *Sequence-dependent pausing of single lambda exonuclease molecules*. *Science*, 2003. **301**(5641): p. 1914-1918.
133. Dumont, S., et al., *RNA translocation and unwinding mechanism of HCVNS3 helicase and its coordination by ATP*. *Nature*, 2006. **439**(7072): p. 105-108.
134. Perkins, T.T., et al., *Forward and reverse motion of single RecBCD molecules on DNA*. *Biophysical Journal*, 2004. **86**(3): p. 1640-1648.
135. Gore, J., et al., *Mechanochemical analysis of DNA gyrase using rotor bead tracking*. *Nature*, 2006. **439**(7072): p. 100-104.
136. Koster, D.A., et al., *Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB*. *Nature*, 2005. **434**(7033): p. 671-674.
137. Koster, D.A., et al., *Antitumour drugs impede DNA uncoiling by topoisomerase I*. *Nature*, 2007. **448**(7150): p. 213-217.
138. Itoh, H., et al., *Mechanically driven ATP synthesis by F-1-ATPase*. *Nature*, 2004. **427**(6973): p. 465-468.
139. Danilowicz, C., D. Greenfield, and M. Prentiss, *Dissociation of ligand-receptor complexes using magnetic tweezers*. *Analytical Chemistry*, 2005. **77**(10): p. 3023-3028.

-
140. Wiggin, M., et al., *Nonexponential Kinetics of DNA Escape from alpha-Hemolysin Nanopores*. *Biophysical Journal*, 2008. **95**(11): p. 5317-5323.
141. Bates, M., M. Burns, and A. Meller, *Dynamics of DNA molecules in a membrane channel probed by active control techniques*. *Biophysical Journal*, 2003. **84**(4): p. 2366-2372.
142. Xie, X.S., *Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic kinetics*. *Journal of Chemical Physics*, 2002. **117**(24): p. 11024-11032.
143. Austin, R.H., et al., *Dynamics of Ligand-Binding to Myoglobin*. *Biochemistry*, 1975. **14**(24): p. 5355-5373.
144. Berberan-Santos, M.N., E.N. Bodunov, and B. Valeur, *Mathematical functions for the analysis of luminescence decays with underlying distributions: 2. Becquerel (compressed hyperbola) and related decay functions*. *Chemical Physics*, 2005. **317**(1): p. 57-62.
145. Berberan-Santos, M.N., E.N. Bodunov, and B. Valeur, *Mathematical functions for the analysis of luminescence decays with underlying distributions 1. Kohlrausch decay function (stretched exponential)*. *Chemical Physics*, 2005. **315**(1-2): p. 171-182.
146. van Oijen, A.M., et al., *Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder*. *Science*, 2003. **301**(5637): p. 1235-1238.
147. Yang, H., et al., *Protein conformational dynamics probed by single-molecule electron transfer*. *Science*, 2003. **302**(5643): p. 262-266.
148. Nadassy, K., S.J. Wodak, and J. Janin, *Structural features of protein-nucleic acid recognition sites*. *Biochemistry*, 1999. **38**(7): p. 1999-2017.
149. Jetha, N.N., et al., *Sequence-dependent DNA escape kinetics from a-HL nanopores*. (in preparation).
150. Mooder, K.P., et al. *Polymorphisms in coagulation and fibrinolytic pathway genes mark the evolution of host-defense response*. 2007 [cited 2008 Nov 21]; Available from: <http://www.ashg.org/genetics/ashg07s/f21404.htm>.

151. Markham, N.R. and M. Zuker. *DINAMelt Two-state hybridization server*. [Web page] 2005 Jan 18 2005 [cited 2009 Jan 4]; Available from: <http://dinamelt.bioinfo.rpi.edu/twostate.php>.
152. Markham, N.R. and M. Zuker, *DINAMelt web server for nucleic acid melting prediction*. *Nucleic Acids Research*, 2005. **33**: p. W577-W581.
153. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. *Nucleic Acids Research*, 2003. **31**(13): p. 3406-3415.
154. *SimPore Membrane Technology*. [web page] 2008 [cited 2008 December 10]; Available from: <http://www.simpore.com/technology.html>.
155. Horowitz, P. and W. Hill, *The Art of Electronics*. Second ed. 1989, New York: Cambridge University Press. 1125.
156. Greenleaf, W.J., M.T. Woodside, and S.M. Block, *High-resolution, single-molecule measurements of biomolecular motion*. *Annual Review of Biophysics and Biomolecular Structure*, 2007. **36**: p. 171-190.
157. Jetha, N.N., M. Wiggin, and A. Marziali, *Forming an alpha-hemolysin nanopore for single molecule analysis*, in *Microfluids, Nanotechnologies, and Physical Chemistry in Separation, Detection, and Analysis of Biomolecules*, J.W. Lee and R.S. Foote, Editors. 2008, Humana: Totowa, NJ.
158. Wonderlin, W.F., A. Finkel, and R.J. French, *Optimizing Planar Lipid Bilayer Single-Channel Recordings for High-Resolution with Rapid Voltage Steps*. *Biophysical Journal*, 1990. **58**(2): p. 289-297.
159. Luan, B.Q. and A. Aksimentiev, *Electro-osmotic screening of the DNA charge in a nanopore*. *Physical Review E*, 2008. **78**(2).
160. Keyser, U.F., et al., *Direct force measurements on DNA in a solid-state nanopore*. *Nature Physics*, 2006. **2**(7): p. 473-477.
161. Sauer-Budge, A.F., et al., *Unzipping kinetics of double-stranded DNA in a nanopore*. *Physical Review Letters*, 2003. **90**(23): p. 238101.

162. Cugnet, P., *Confidence Interval Estimation for Distribution Systems Power Consumption by Using the Bootstrap Method*, in *Electrical Engineering*. 1997, Virginia Tech.
163. Stuve, E. *Logarithmic Error Bars*. 2008 [cited 2008 August 21]; Available from:
http://faculty.washington.edu/stuve/uwess/log_error.pdf.

Appendix A – Characteristics of Common Force Spectroscopy Techniques

	AFM	Magnetic Tweezers	Optical Tweezers	Nanopore Force Spectroscopy
Force range (pN)	5 - 10 ⁴ pN	10 ⁻³ - 10 ⁴ pN	0.1 - 100 pN	1-10 ³ pN
Displacement range (nm)	0.5 - 10 ⁴ nm	5 - 10 ⁴ nm	0.1 - 10 ³ nm	0.1-10 nm
Measurement timescale (s)	10 ⁻³ - 10 ² s	10 ⁻⁴ - 10 ⁵ s	10 ⁻⁴ - 10 ³ s	10 ⁻⁴ - 10 ³ s
Spatial resolution (nm)	0.5 - 1 nm	2 - 10 nm	0.1 - 2 nm	0.1-10 nm
Advantages	Easy, rapid setup	Applies force and torque Constant force Easy to parallelize	3D manipulation	Easy to parallelize Often label-free Direct force application
Limitations	Large, stiff probe Nonspecific attachment	Difficult to vary applied force	Photo-damage Sample heating	Low spatial resolution (except in special cases) Requires charged analyte

Table A-1 Characteristics of common force spectroscopy techniques. Data for AFM, magnetic tweezers, and optical tweezers taken from [113, 156].

Appendix B Materials & Methods

This Appendix describes instruments, software, reagents, and protocols used to form α -HL nanopores⁴². Experimental methods specific to individual experiments are described in the appropriate sections of the main text.

B.1 Instrumentation

A schematic view of the apparatus used in nanopore experiments is shown in Figure B-1.

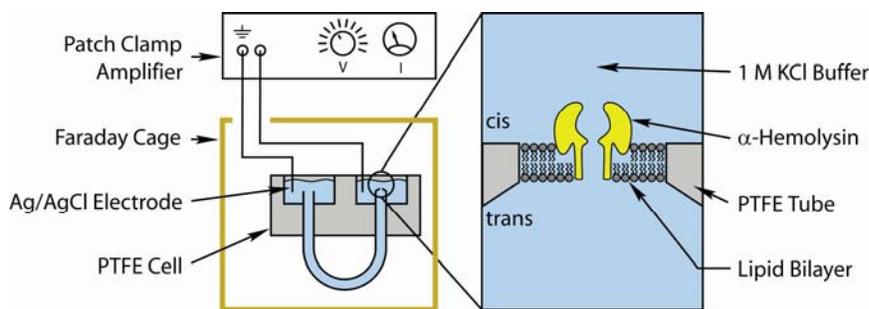


Figure B-1 The nanopore apparatus, including the α -HL pore, PTFE cell, patch clamp amplifier, electrodes, and Faraday cage.

⁴² Readers interested in more detailed protocols, suitable for setting up a nanopore laboratory, are referred to a recently published book chapter on the subject [157].

Experiments are conducted in a custom-built polytetrafluoroethylene (PTFE) chamber containing two 250- μ L chambers connected by a U-shaped PTFE shrink tube (Small Parts Inc. Miramar FL.) The tube has a 9 mm internal diameter, except at the end which forms the support for the lipid bilayer / nanopore assembly, where the diameter of the tube is reduced to a 25 μ m aperture⁴³, as follows.

The aperture is formed by inserting a 25 μ m diameter Constantine wire (Omega Engineering Stamford CT) into the tube, and heating it with a heat gun to close the shrink tube around the wire. The tip is then cut from the end of the tube with a razor blade, and the wire is removed, leaving a 25 μ m aperture at one end.

Ag/AgCl electrodes, used to control the applied potential across the bilayer and measure the ionic current flowing through the nanopore during experiments, are made by soldering 1 mm diameter, 99.9% pure silver wire (Alfa Aesar, Ward Hill MA) to 1 mm gold pins, which form connections to the patch clamp amplifier. The silver wire is insulated by covering it with PTFE shrink tube, leaving 1 cm bare at the end opposite the gold pin. The uncoated end is then sanded to remove impurities, immersed in concentrated bleach (NaClO) for at

⁴³ Note that this minimized the size of the lipid bilayer used during experiments, minimizing dielectric noise [158] and prolonging the life of the bilayer.

least 1 hour, forming AgCl on the surface, and finally, rinsed thoroughly with dH₂O. Electrodes are periodically re-surfaced by this procedure to maintain the AgCl coating.

The electrodes are connected to an Axopatch 200B patch clamp amplifier (MDS Analytical Devices, Sunnyvale, CA; formerly Axon Instruments, Union City, CA). The amplifier has a dynamic range of *ca.* 1 pA - 18 nA, creating an upper limit of ~180 α -HL pores when used for multi-pore experiments. The Axopatch contains an internal, 4-pole, low pass Bessel filter with an adjustable cut-off frequency, which is typically set to 10 kHz, resulting in 1-2 pA current standard deviation during experiments. In addition, the Axopatch has controls for subtraction of offset potentials between the electrodes, and capacitance compensation, which are adjusted prior to the start of experiments, as described below.

The voltage-control input, and both voltage and ionic current outputs of the patch clamp amplifier are connected to a LabView PCI-MIO-16E-4 data acquisition card (National Instruments Canada, Vaudreuil-Dorion, QC) in a computer. Data acquisition software is described below.

A custom-built copper Faraday cage is used to enclose the nanopore cell and Axopatch headstage (which detects and amplifies ionic current signals), in order to reduce electrical noise due to stray electromagnetic fields. This is mounted on an air suspension table (LW

series, Newport Corporation, Irvine CA), in order to reduce mechanical noise due to vibration.

Four additional pieces of equipment are used in these experiments: a hot plate (VWR, Mississauga, ON), a vacuum chamber (VWR, Mississauga, ON), and an oil-free diaphragm vacuum pump (Gast, Cole Parmer Canada, Montreal QC), all of which were used in cell cleaning procedures, and a dissecting microscope (Leica S6E, VWR International Mississauga ON), used in lipid bilayer formation.

B.2 Software

All data acquisition and analysis software was created using the LabView graphical programming language (National Instruments Canada, Vaudreuil-Dorion, QC). A conceptual description of the software is given here; details of operation are given where appropriate in the main body of the thesis.

B.2.1 Data Acquisition Software

Data acquisition software is used to record ionic current and applied potential signals to a PC hard drive for later analysis, and to control the applied potential during experiments in response to events such as DNA entering or exiting the pore. To accomplish this, the data acquisition software is a user-configurable state machine, where the

state is defined by the ionic current and time spent in the present state. State switching occurs in response to the current crossing a threshold, or when the time in the present state exceeds a time-out value. For example, during probe escape experiments, probe capture causes the ionic current to drop below a threshold value, and the data acquisition software responds by reducing the applied potential to the escape potential.

B.2.2 Raw Data Analysis Software

After an experiment is completed, raw data analysis software is used to determine event times from the recorded current and voltage data. Similar to the data acquisition software, analysis software is a state machine, with the state determined by applied potential, ionic current, and time spent in the present state. The software is configured by the user to call and time events, as described in the appropriate sections of the main text.

A second piece of software, which allows the user to observe voltage, current traces, and event timing (as called by analysis software) is used to confirm that events are called and timed correctly by the raw data analysis software.

B.3 Reagents

The buffer used in all experiments is 1 M KCl, 10 mM HEPES, pH 8.0 (1X nanopore buffer). Following initial mixing of KCl and HEPES in ultra-filtered, de-ionized water, the pH is adjusted by adding KOH until the pH reached 8.0. Finally, the volume is adjusted by addition of dH₂O to give a final KCl concentration of 1.00 M. All buffers are filtered through a 100 nm Millipore Express PLUS Membrane (VWR International Mississauga ON) prior to use.

Working DNA or probe solutions are made by diluting the stock solution, along with an equal volume of a 2X nanopore buffer, into 1X nanopore buffer to the desired final DNA concentration. This minimizes the perturbation of the ionic strength when adding DNA or probe solutions to the nanopore cell.

Black lipid membranes are formed from 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPPC), purchased as lyophilized powder from Avanti Polar Lipids, Inc. (Alabaster, AL). Upon receipt, the lipid is dissolved at 10mg/mL in chloroform, and stored in a -20 °C freezer for up to 3 months before replacement.

α -Hemolysin from *Staphylococcus aureus* is purchased from either CalBiochem (San Diego CA), or Sigma Aldrich (Oakville ON) as a lyophilized powder. Upon receipt, the α -HL is dissolved to a final concentration of 1mg/mL in 1:1 glycerol:dH₂O, and stored in a -20 °C freezer for up to 3 months before replacement.

Single stranded DNA used for probes and targets is chemically synthesized (MWG Biotech, High Point, NC or IDT DNA, Coralville, IA). Probes are biotinylated at either the 3' or 5' end for coupling to Avidin (see below). Probe sequences used for probe escape experiments are described in section 4.1, and probe and target sequences used in DNA force spectroscopy experiments are given in Table 5-1 of the main text. DNA is used without further purification; however, a subset of molecules were verified by mass spectrometry, and not found to contain detectable quantities of truncated DNA.

Avidin (Sigma-Aldrich, Oakville, ON) is coupled to the biotin moiety of all probes to keep them escaping to the *trans*-side of the pore after capture. Avidin and DNA are mixed to a final concentration of 10 μ M each in 1M KCl 10 mM HEPES, pH 8.0, vortexed for 10 seconds, and incubated for at least 30 min at room temperature before use. Working target solutions are diluted to a final concentration of 500 nM in 1X nanopore buffer.

B.4 α -HL Nanopore Formation Methods

The day before the experiment, the lipid is prepared aliquoting 50 μ L of DPPC / chloroform solution into each of 4 borosilicate test tubes. Tubes are left uncovered overnight, allowing the chloroform to evaporate, leaving a small spot of concentrated lipid in the bottom of the tube.

Immediately prior to the experiment, the PTFE nanopore cell is cleaned to remove DNA or other contaminants by boiling in 15% nitric acid (v/v) in dH₂O. The cell is then thoroughly rinsed by boiling twice in dH₂O, followed by forcing 500 μL each of dH₂O, ethanol, and HPLC grade hexane through the U-shaped PTFE tube with a syringe. After cleaning, the cell is dried for ~5 minutes under vacuum to remove any remaining hexane.

Next, a thin layer of DPPC is formed on the surface of the PTFE aperture, to assist in bilayer formation. The lipid in one test tube is dissolved in 250 μL of HPLC grade hexane, and 1 μL of this lipid solution is transferred onto the 25 μm aperture. The tube is allowed to dry in air, excess solution is forced out of the aperture by a syringe, and any remaining hexane solvent is removed by placing the cell under vacuum for 5 minutes.

The nanopore cell is then assembled in its housing with the Ag/AgCl electrodes in place. Electrodes are then connected to the head stage of the patch clamp amplifier. Next, the cell is filled with solution, which differs according to the experiment. For probe escape experiments, the cell is filled with 1X nanopore buffer, while for duplex dissociation experiments, it is filled with 1X nanopore buffer containing 500 nM target DNA.

Next, lipid is prepared for bilayer formation by adding 5 μL of 1-hexadecene (99% pure, Sigma-Aldrich, Oakville, ON) to a test tube

containing dry lipid, and allowing it to stand for ~45 s. Excess hexadecene is removed by inverting the tube and tapping it gently. Approximately 0.5 μL of hexadecene-dissolved lipid is then transferred to the U-shaped PTFE tube, adjacent to the 25 μm aperture, and mixed with a single bristle paint brush until it has the consistency of toothpaste-spit.

The lipid bilayer is then formed by using a pipette to gently blow an air bubble over the surface of the hexadecene mixture, across the aperture, and then withdrawing it back into the pipette. Bilayer formation is detected by monitoring the electrical characteristics of the system as follows. Using the patch clamp amplifier, a 5 mV square wave is applied, and the current was monitored. Bilayer formation causes the ionic current signal to change from a large magnitude square wave to a series of rapidly decaying capacitive spikes.

Bilayer strength is then tested by applying a large DC potential to burst the bilayer⁴⁴. Bilayers that break at a potential between 450 mV and 550 mV give good results for experiments, i.e. they are long-lived, and easy to form pores in. Bilayers breaking outside this range are not used in experiments.

⁴⁴ Note that the bilayers in any given experiment are highly reproducible, breaking at nearly the same potential each time. This means that the strength of a bilayer can be inferred from the breaking strength of previous bilayers.

Once a good bilayer has been formed, nanopore formation is straightforward. α -HL is diluted either to 0.05 mg/mL (single pore) or 0.2 mg/mL (multi-pore) in the same solution used to fill the nanopore cell. Approximately 2 μ L of diluted α -HL solution is added to the *cis*-side of the nanopore cell (i.e. the side containing the 25 μ m aperture) and stirred by refluxing.

Self-assembly of α -HL pores in the lipid bilayer is monitored by applying a 100 mV potential and observing the ionic current. A single pore typically conducts 97 ± 4 pA at +100 mV (trans-positive) and -66 ± 4 pA at -100 mV. Once the desired number of pores has incorporated, the cell is flushed with 3 mL of 1X nanopore buffer to remove excess α -HL, preventing further pore formation.

At this point, two adjustments are made to the patch clamp amplifier. First, the pipette offset is adjusted to cancel out any offset potential between the two electrodes⁴⁵, such that 0 mV applied potential results in 0 pA current through the pore(s). Second, the capacitance compensation is adjusted minimize observed transient currents resulting from capacitance of the lipid bilayer and patch

⁴⁵ Note that the offset potential is caused by differences in the AgCl coating between the two electrodes. The offset potential must be removed prior to experiments, to ensure that the potential applied across the bilayer is equal to the potential observed (and controlled) by the patch clamp amplifier.

clamp components when the applied potential was changed⁴⁶. With proper capacitance compensation adjustment, it is possible to observe probe exit events as fast as 200 μ s following potential changes during experiments.

At this point, set-up is complete, and experiments are started, using protocols described in the appropriate parts of the main text.

⁴⁶ Note that this does not affect the actual ionic current inside the nanopore cell; rather, the patch clamp amplifier subtracts an exponentially decaying component from the output signal, meaning that the capacitive current is not observed.

Appendix C - Measurement of DNA Length in α -HL

The experiments described in this section are designed to measure the length of the DNA probe with respect to the nanopore. Note that the length of the nanopore is already known to be 100 Å from X-ray crystallography [60] studies; however, length of the DNA molecule in the pore is required for two reasons. First, it is necessary for designing the length of the probe. In order to ensure that probes could hybridize to target molecules on the *trans*-side of the bilayer, the DNA linker between Avidin and the DNA binding region must to be long enough to extend all the way through the pore. Second, it is required in order to determine l , the length per monomer of the DNA probe in the pore, in order to calculate the applied force on the DNA molecule (see Appendix D).

Therefore, the length of the DNA molecule inside the pore was measured experimentally, by making probes of varying length, and testing their ability to target molecules on the *trans*-side of the pore.

C.1 Methods & Raw Data Analysis

Experimental methods follow those used in single molecule force experiments (see section 5.1 for details). Note, however, that the probe is held in the pore at 200 mV for 3.0 s in these experiments, in order

allow probe-target hybridization to reach equilibrium prior to the hybridization check step. In addition, two separate data analyses are performed, in order to measure both hybridization kinetics and dissociation kinetics.

The main goal of data analysis is determination of the fraction of probes that bound target during the 200 mV hybridization step. This is done by using the raw data analysis software to determine the fraction of probes which remain in the blocked, low current state during the 10 mV hybridization check step. The hybridization probability is the ratio the number of successful hybridizations, N_{hyb} to the total number of probe captures, N_{total} :

$$P_{hybridization} = \frac{N_{hyb}}{N_{total}} \quad (A-1)$$

where $P_{hybridization}$ is the probability of hybridization.

In addition, the dissociation kinetics for probe-target duplexes are measured, using analysis methods described in section 5.1.1.

C.2 Results

A plot of hybridization probability vs. probe length is shown in Figure C-1. The data shows a dramatic transition in binding effectiveness at a linker length of 20-22 nucleotides. Probes with linkers 20-nt long or shorter all show a hybridization probability near zero,

indicating that the DNA binding region of these probes is at least partially inside the nanopore. Molecules with linkers 22-nt long or longer all bind target with approximately equal effectiveness, suggesting that these molecules extend far enough through the pore to give the DNA binding region access to the *trans*-side.

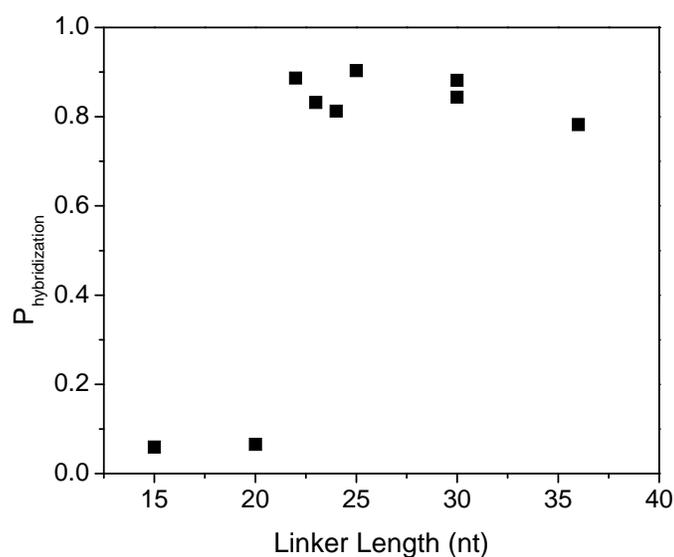


Figure C-1 Target hybridization probability as a function of probe length.

To confirm that all probes form complete duplexes, i.e. that probes with linkers close to 22 nucleotides in length are not prevented from forming complete duplexes by steric hindrance from the pore, we measure dissociation kinetics as a function of probe length. Probes forming incomplete duplexes would be expected to dissociate faster

than probes forming complete duplexes. A subset of these results is presented in Figure C-2.

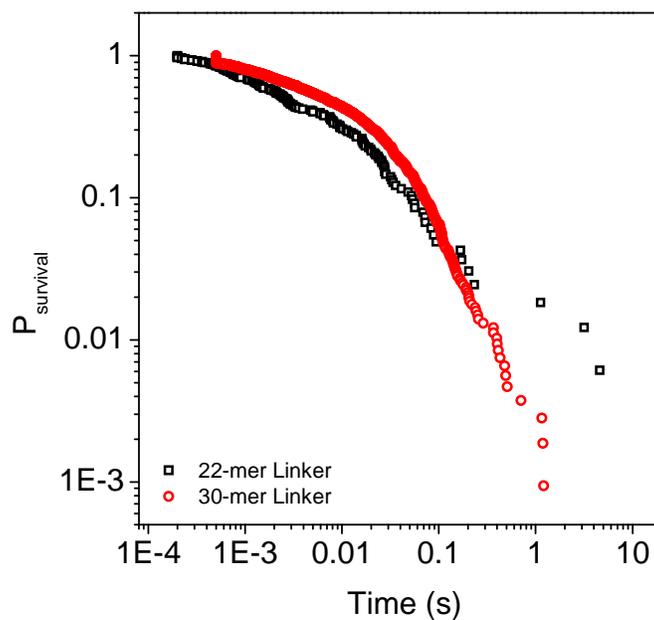


Figure C-2 Comparison of force spectroscopy event time distributions for probes containing 22-nt and 30-nt linkers at -60mV applied dissociation potential.

The event time distributions shown in Figure C-2 are nearly identical⁴⁷, implying that the dissociation energy is nearly identical for these molecules. It is therefore highly likely that the DNA binding regions of probes with linkers 22-nt long or longer are capable of forming complete duplexes on the *trans*-side of the pore.

C.3 Discussion

The probe with a 22-nt linker is the shortest probe capable of forming a complete duplex on the *trans*-side of the pore, making the effective length of the pore 22 nucleotides. Since the α -HL pore is 100 Å long [60], the monomer unit length for a DNA molecule in the pore is $\ell \approx 4.5$ Å / nucleotide. This agrees well with optical trap force-extension curves of ssDNA, which predict $\ell \approx 4.8$ Å / nucleotide [122] at forces similar to those applied in these experiments⁴⁸.

⁴⁷ The minor variation between the curves is well within the limits of repeatability of single molecule nanopore force spectroscopy experiments. Minor variations in dissociation time distributions appear to be caused by the differences in the α -HL pore and lipid bilayer from experiment to experiment. A more detailed analysis of this issue is described in Appendix F.2.

⁴⁸ See Appendix D for force calculation.

Somewhat surprisingly, no other direct measurement of the length of ssDNA molecules in the entire α -HL pore has been reported in the literature. These results can, however, be compared to an indirect measurement of the length of DNA in the β -barrel, made by Meller and Branton [57]. They measured ionic current reduction as a function of DNA length, and found that current reduction was length-dependent for short molecules, but independent of length for molecules greater than or equal to 12 nucleotides long. Based on this, they estimated that 12 nucleotides fit into the β -barrel, which is half the length of the entire pore. Thus, they predict that the whole pore could accommodate 24 nucleotides, which agrees well with the result of 22 nucleotides given here.

Appendix D - Calibration of Electrostatic Force

Nanopore force spectroscopy could be applied to many problems beyond genotyping, including single-molecule measurement of energy barriers governing biologically important structural transitions. In order to measure the height of the free energy barrier in a force spectroscopy experiment, it is essential that the applied force be known⁴⁹. Therefore, in the interest of stimulating such studies, this section calibrates force with respect to the electrostatic potential. As discussed below, the work performed here gives only a rough estimate of force; however, this is still useful as a starting point for further work.

The force in nanopore force spectroscopy arises from the pull of the electric field on the charged DNA backbone. If the DNA was completely unshielded (see below), and experienced no other forces, each nucleotide within the field would experience a force:

$$f = eE \tag{C-1}$$

where E is the electric field, which can be approximated as $\Delta V/l$, where ΔV is the potential drop at occurring across the nucleotide, and l

⁴⁹ Note that the genotyping method described in this thesis does not require force calibration, since it is semi-quantitative. The main goal is determining sequence homology between the probe and the target, rather than measuring the exact energy barrier to duplex dissociation.

is the nucleotide spacing, previously determined to be 0.45 nm (see Appendix C).

Note, however that (C-1) over-estimates the effective force experienced by the nucleotide. The effective force is lowered by counter-ion condensation, i.e. K^+ ions which associate with the DNA molecule, partially canceling the negative charges in the backbone. In addition, electro-osmotic flow as fast-moving K^+ ions pull water past the DNA molecule creates a frictional force, which opposes the electrostatic force [159, 160]. These effects can be accounted for by inclusion of a scaling term z , which reduces the effective charge on each nucleotide:

$$f_E = ze \frac{\Delta V}{l} \quad (C-2)$$

where f_E is the effective force acting on the nucleotide.

The total force on the DNA molecule will be the sum of the effective forces applied to all nucleotides residing within the field:

$$\begin{aligned} f_{total} &= \sum_i ze \frac{\Delta V}{l} \\ f_{Total} &= \frac{zeV}{l} \end{aligned} \quad (C-3)$$

Therefore, we can calibrate the force on the DNA molecule, if z is known. We determine z using data from probe escape experiments,

beginning by considering the relationship between the energy barrier, applied force, and probe length.

The energy barrier to probe escape can be modeled as:

$$\Delta G = \Delta G_b^0 + \frac{zeV}{l} L_{eff} \quad (C-4)$$

where ΔG_b^0 is the energy barrier in the absence of applied force, arising from any probe-pore interactions, and the second term represents the energy barrier arising from the effective force. Note that this is assumed to act through a distance L_{eff} , which is less than the length of the molecule, accounting for the fact that not all nucleotides must cross the entire potential for the probe to escape.

Equation (C-4) can be simplified, by making the substitution $L_{eff} = N_{eff}l$, where N_{eff} is the effective number of nucleotides which would have to cross the potential to create the energy barrier. Making this substitution, we obtain:

$$\Delta G = \Delta G_b^0 + zeVN_{eff} \quad (C-5)$$

Inserting this into the Arrhenius relation (3-1) gives the force-lifetime relationship for probe escape:

$$\ln(\tau^*) = \ln(\tau_D) + \Delta G_b^0 + zeVN_{eff} \quad (C-6)$$

Therefore, the slope of $\ln(\tau^*)$ vs. V is zeN_{eff} . This will be referred to as the barrier height sensitivity to applied potential, and is shown in Figure D-1.

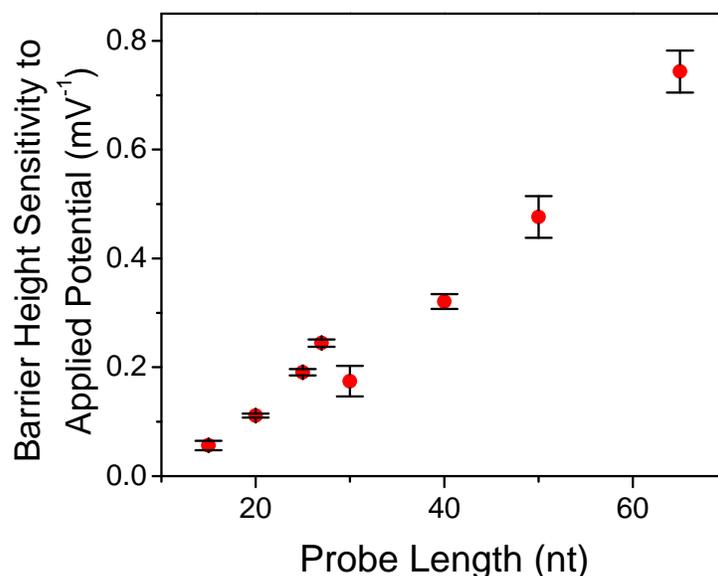


Figure D-1 Barrier height sensitivity to applied potential as a function of molecule length. Data points are slopes of $\ln(\tau^*)$ vs. V probe escape data in Figure 4-4. Error bars indicate standard error of measurement, calculated by error propagation. The data behaves unexpectedly for molecules 25-30 nucleotides long, and actually shows a decrease in sensitivity to the applied potential as length increases from 27nt to 30nt. The reason for this is unknown; however, repetition of the experiment gave similar results. Originally published in [140].

To estimate z , we consider only probes longer than the pore, since a 1-nt increase in length for these molecules causes a 1-nt increase in N_{eff} .

The slope of barrier height sensitivity vs. length in this region is ze . Taking the slope of the data from the linear region from 30-65 nucleotides long in Figure D-1, we find that $z \approx 0.4$ [140]. A number of other groups have estimated z using a variety of experimental methods [160, 161] and molecular dynamics simulations [159], and have found values ranging from $z \approx 0.1$ [161] to $z \approx 0.3$ [159, 160]. Given that estimates of z vary so widely depending on how it is measured, further studies in this area are certainly warranted.

Having estimated z , we can now solve (C-3) for the applied force. Taking $z \approx 0.4$, as calculated here, we find that $F_{Total} \approx 1.5$ pN/mV. Note, however, that given the uncertainty in z , this should be considered a rough estimate. Taking the range of estimates $0.1 \leq z \leq 0.4$, we find that 0.4 pN/mV $\leq F_{Total} \leq 1.5$ pN/mV.

Appendix E Calculation of τ^* Error Bars

This Appendix describes the calculation of error bars for τ^* in both probe escape and duplex force spectroscopy experiments. This is done by bootstrapping [162], which is a convenient method for estimating the error in a sampled dataset.

Bootstrapping consists of creating many datasets of equal size to the original dataset by re-sampling the original dataset, with replacement, and then calculating the desired parameter (τ^*) for each re-sampled dataset. The resulting set of calculated values is used to estimate the error [162].

For each dataset to be analyzed, 10 000 re-sampled data-sets, each containing the same number of points as the original dataset are created. First, a set of random numbers, $0 \leq R_i \leq 1$, is generated. Each R_i is assigned an event time by choosing the event whose survival probability is closest to R_i . The resulting set of event times is treated as an experimental dataset. $P_{survival}$ vs. t is calculated, and used to determine τ^* , using methods described in Chapters 4 and 5.

Error in τ^* is calculated logarithmically [163], since τ^* is exponentially related to the applied potential. Therefore, the variance is calculated for the log of τ^* :

$$\ln(s_{\tau^*}) = \frac{\sqrt{\sum (\ln \tau_i^* - \langle \ln \tau^* \rangle)^2}}{n-1} \quad (\text{D-1})$$

where $\ln(s_{\tau^*})$ is the logarithmic sample variance, τ^*_i is the characteristic timescale calculated from re-sampled dataset i , and n is the number of re-sampled datasets.

Error bars are then calculated as:

$$\tau^* \pm \delta\tau^* = \exp\left[\ln \tau^* \pm \ln s_{\tau^*}\right] \quad (\text{D-2})$$

Appendix F Performance Evaluations

This Appendix describes additional studies related to instrument performance, including repeatability, and preliminary studies of sensitivity and unrelated sequence rejection. It should be noted that all of these studies were performed within the limitations of the current instrument, and as such, any statements regarding the envisioned instrument, i.e. a sensor based on millions of pores, in tests involving human genomic DNA, should be taken as estimates.

F.1 Sensitivity

Aside from sequence selectivity, sensitivity is perhaps the most important measure of sensor performance in genotyping. Sensitivity, i.e. the minimum amount of target required to make a base-call, affects assay complexity and speed. It determines whether PCR is required, and the length of the hybridization step, which is most likely to be the slowest step in a clinical diagnostic genotyping assay. A number of issues related to instrument sensitivity are explored here. These include sampling error in τ^* , due to the limited dataset size, and hybridization kinetics, which determines the number of dissociation events which can be observed in a reasonable length of time. Note that a third factor relating to sensitivity, the signal-to-noise ratio of current measurements, is discussed in Chapter 6.

F.1.1 Sampling Error

Sampling error in τ^* arises from the fact that τ^* is measured from an event time distribution of finite size. This section attempts to measure the effect of dataset size on the error in τ^* , using a Monte Carlo algorithm⁵⁰.

To simulate sampling, a large experimental single pore dataset, containing 1497 events was used re-sampled to create datasets ranging in size from 10 to 3000 points. For each sample size, 10 000 re-sampled datasets were created and analyzed to determine the characteristic timescale. The error in the $\log(\tau^*)$ was measured from the resulting timescale distribution, as described in Appendix E. Results of the analysis are given in Figure F-1. As expected, the error decreases proportional to the square root of dataset size.

⁵⁰ See Appendix D for a description of Monte Carlo methods, and error bar calculation.

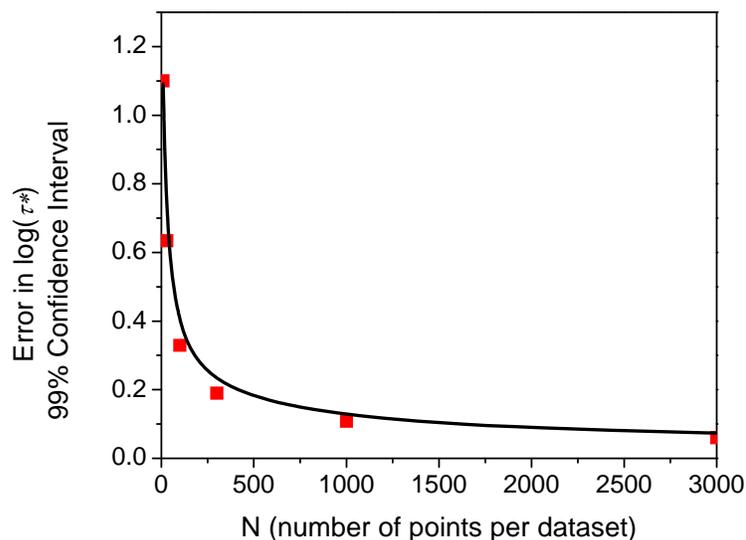


Figure F-1 Nanopore force spectroscopy characteristic timescale estimation error as a function of dataset size. Each point represents the 99% confidence interval for the distribution of $\log(\tau^*)$ for 10^4 Monte Carlo simulated sample sets. The trend line is proportional to $1/\sqrt{n}$.

From results in Figure F-1, it is apparent that for a dataset containing 30 dissociations, the 99% confidence interval of the error in $\log(\tau^*)$ is 0.67, meaning that we can state, with 99% confidence, that the true characteristic timescale is within a factor of 4.5 of the measured value⁵¹. For a well-designed probe, FS timescales differ by at least a factor of 50 for a single nucleotide substitution. This suggests that the

⁵¹ Note that an error in $\log(\tau^*)$ of 0.67 corresponds to an interval of $10^{0.67} \tau^*$ to $\tau^*/10^{0.67}$, *i.e.* the true timescale should lie within a factor of $10^{0.67} = 4.5$.

genotype of a sample could be unambiguously determined from a dataset containing as few as 30 single molecule dissociations. Note, however, that this error estimate assumes that each dissociation event is timed accurately, which may not be possible for large, multi-pore arrays, whose signal-to-noise ratios have not yet been measured experimentally. Therefore, for further analysis, we will take a more conservative estimate of 100 - 250 events required to identify an allele during base calling.

F.1.2 Probe-Target Hybridization Kinetics

Having come up with an estimate of the number of events required to detect a given sequence, we can now estimate the hybridization time for an NFS genotyping experiment. This estimate requires that we first measure probe-target hybridization kinetics.

Hybridization kinetics are measured using methods similar to those described in Appendix C, at a single target concentration of 500 nM. The Synth probe is captured in a single pore at 200 mV, and held for varying amounts of time ranging from 0.01 s to 5 s, where it can potentially hybridize to a target molecule (PC or 8C). The potential is then reduced to 10 mV for 0.2 s, to test for hybridization, by determining whether the probe remains trapped in the pore during this step.

Results of these experiments are not perfectly reproducible, showing a variation in hybridization rates of up to a factor of two between different trials. Therefore, each experiment is repeated at least three times, and the results were averaged.

Results of these experiments are shown in Figure F-2. To extract probe-target hybridization rates, data in Figure F-2 are fitted by assuming target hybridization to be a first order reaction, with a rate of $k_{on}[C]$, where k_{on} is the rate constant and $[C]$ is the target concentration. This gives hybridization kinetics that follow:

$$P_{hybridization}(t) = 1 - Ae^{-k_{PT}[T]t} \quad (\text{E-1})$$

where the pre-factor A describes the maximum hybridization probability, observed at long times, k_{PT} is the probe-target hybridization rate, and $[T]$ is the target concentration. The fits in Figure F-2 follow the data very closely; however, two important features, considered below, are worth discussing.

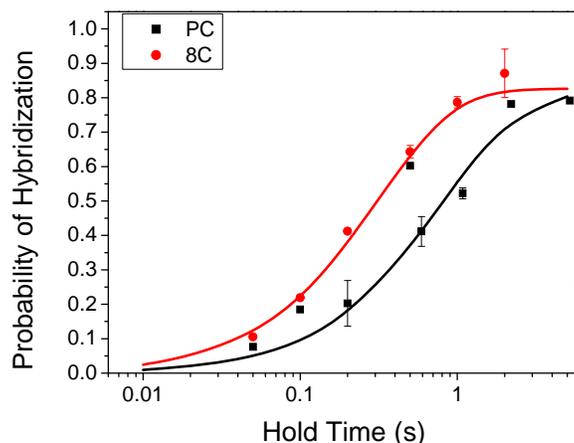


Figure F-2 Probability of probe-target hybridization vs. hold time. Experiments were performed using a single pore with the Synth probe, and either PC (perfect complement) or 8C (single base mismatch) targets, each at 500 nM concentration. Each data-set is an average of four (PC) and three (8C) trials. Data are fitted to equation (E-1), a single exponential, as described in the text.

The first of these features is the fact that hybridization kinetics are sequence specific. Rate constants extracted from the fits for PC and 8C targets are $2.4 \times 10^6 \text{ s}^{-1} \text{M}^{-1}$ and $6 \times 10^6 \text{ s}^{-1} \text{M}^{-1}$, respectively. This is surprising, especially given that the mismatched target binds the probe faster than the perfect complement. Similar results (not shown) were obtained in a single experiment, in which the probe was swapped for a probe complementary to the 8C target, i.e. the mismatched target was still found to hybridize faster than the perfect complement. The reasons for this are unclear, and certainly warrant further investigation, especially since despite an extensive search, no references to such

behavior were found in the literature. However, these further studies lie beyond the scope of this thesis.

The second important feature of Figure F-2 is the fact that hybridization probability for both molecules saturates at less than the expected value of 1. This suggests that either hybridization is either reversible under these conditions, or some probes become inaccessible to target molecules. The latter of these possibilities is more likely, since if hybridization were reversible, the perfect complement would be expected saturate closer to 100% hybridization than the mismatched molecule. Probe inaccessibility could be explained by DNA binding to the lipid bilayer.

This is an extremely important issue to further development of nanopore force spectroscopy genotyping. Regardless of the cause, it is possible that the saturating hybridization probability is dependent on the target concentration, meaning that low target concentrations could prove very difficult to detect. Further investigation of this phenomenon, over an extensive range of target concentrations, is therefore required.

It should be noted, however, that it may be possible to increase the saturation hybridization probability, for example by using a different lipid (or appropriate surface coatings in the case of solid state nanopores) to suppress probe-membrane interactions. For now, we

shall assume this issue could be dealt with by future improvements, and note that the calculations below should be taken with caution.

Based on these results, we can estimate hybridization times for a sensor incorporating millions of nanopores, for the purpose of calculating an overall signal-to-noise ratio, as presented in Chapter 6. To demonstrate this, we take a very conservative estimate of 250 dissociations required to accurately determine a FS timescale (see Figure F-1). We solve equation (E-1) assuming a solid state membrane containing 10^6 pores, a rate constant of $2.5 \times 10^6 \text{ s}^{-1} \text{ M}^{-1}$, a 50 fM target concentration, and a required hybridization efficiency of 0.025% (250 duplexes). This gives a total hybridization time of ~30 minutes, without amplification by PCR.

F.2 Repeatability

Consistent characteristic timescale determination from experiment to experiment is essential for a clinical genotyping assay. This was tested within the limitations of currently available instruments using purified, chemically synthesized DNA. NFS experiments were repeated under identical conditions, and consistency of survival probability curves, extracted timescales, and for heterozygote genotyping, the relative proportion of each target measured in the sample, were examined.

Survival probabilities for repeated single pore and multi-pore assays are shown in Figure F-3. Single pore kinetics vary substantially from experiment to experiment; the population standard deviation of characteristic timescales is typically a factor of ~ 2 ; the fastest and slowest datasets shown in Figure F-3 vary by a factor of ~ 5 . While this is clearly undesirable, it is possible to identify the perfectly complementary target unambiguously in laboratory tests, since the characteristic timescales for PC and 8C differ by a factor of >100 under these conditions.

Note, however, that the multi-pore results are extremely consistent, with almost no variation in characteristic timescales. This is an encouraging result, since the proposed instrument would contain many pores, suggesting that timescale variability would be low from experiment to experiment.

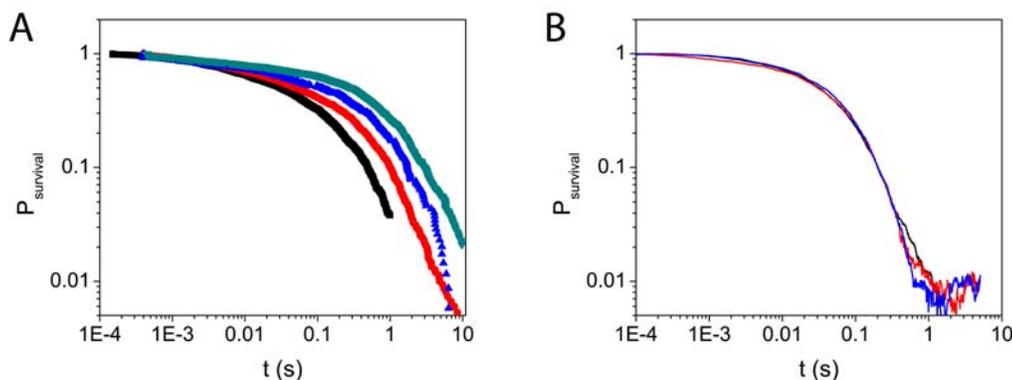


Figure F-3 Repeatability of nanopore force spectroscopy kinetics between different trials. Both trials are for Synth probe dissociation from the PC target. A) Single pore results. Timescales vary by a factor ~ 5 . B) Multi-pore results. Timescales do not vary detectably between trials. Note that noise in multi-pore $P_{survival}$ data at long times is due to gating (i.e. spontaneous closure) of some α -HL pores at negative potential.

These results suggest that the most likely cause for kinetic variation is inconsistency in the α -HL nanopore between experiments. As seen in Chapters 4 and 5, α -HL interacts with the probe, and possibly the duplex. If such interactions vary from pore to pore, it could be expected to cause variability in results where only a single pore is used, while when hundreds of pores are used, such variability would be averaged out.

Variation in target hybridization rates was also observed, and proved much more difficult to control. Great care was taken in these experiments to ensure consistent results. Target stock solution concentrations were measured by absorbance at 260 and 280 nm. A

single working solution was made, mixed thoroughly, and aliquoted for replicate experiments of a given type. In addition, during setup for each experiment, the entire cell was filled with target solution while the lipid membrane and nanopores were being created, to avoid possible diffusion effects.

Despite these attempts to control the hybridization rate, the population standard deviation is still approximately a factor of 2. Surprisingly, sequence-specific binding rate variation is also observed in mixed target FS experiments. This can clearly be seen in the multi-pore dissociation kinetics curves presented in Figure F-4. Each dataset in the heterozygote genotyping tests contains at least 500 single molecule dissociations. The expected population standard deviation is 2% or less, compared to the value of 15% measured over multiple trials.

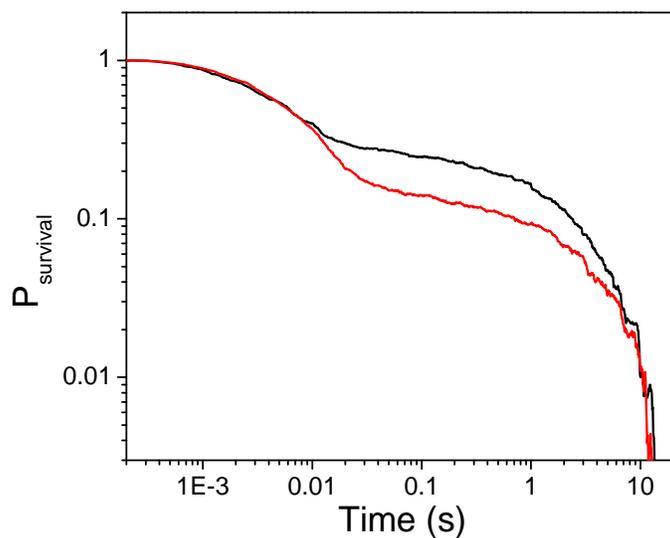


Figure F-4 Variability in fraction of target capture in heterozygote genotyping experiments. Results shown are for a multi-pore assay.

The cause of hybridization rate variation is unknown. It is conceivable that variability in measured binding rates (section F.1.2) stems from target adsorption to the surfaces of the lipid membrane and PTFE nanopore cell. However, this would not explain the sequence-specific hybridization rate differences in Figure F-4. It is unclear what mechanism could cause such sequence specific variation. Further optimization of hybridization rate variation is certainly warranted; however the success of a commercial instrument does not depend on such improvements, since results are sufficiently repeatable at present for binary tests such as genotyping.

F.3 Unrelated Sequence Rejection

All experiments described thus far used chemically synthesized, purified target DNA samples. Genomic DNA samples will contain extensive amounts of partially complementary and unrelated DNA. A preliminary experiment was conducted to simulate the effects of unrelated DNA sequences on sensor operation.

A 500 nM sample of PC DNA was spiked with 5 μ g/mL each of single-stranded M13 and double stranded λ DNA, and then tested by single pore force spectroscopy. Characteristic timescales were not detectably affected by the addition of unrelated DNA, as shown in Figure F-5.

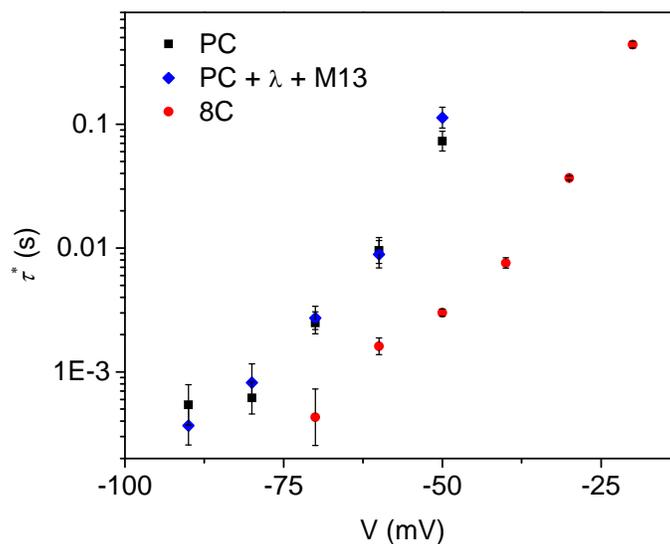


Figure F-5 Influence of unrelated DNA on NFS dissociation timescales. The addition of 5 μ g/mL each of M13 and λ phage DNA does not significantly affect characteristic dissociation timescales, since the PC and PC+ λ +M13 data are within error of one another. Each point represents *ca.* 500-1500 single molecule events, from a single experiment. Error bars were calculated as described in Appendix E.

It should be noted that the genomes for both M13 (6.7 kb, single stranded) and λ (48 kb, double stranded) are small compared to a human genome, and therefore contain little sequence homology to the probe. Therefore, although these studies demonstrate that unrelated sequences do not affect sensor operation, studies involving human genomic DNA are still required. This is left for future studies.