

Nanopore Analysis of Nucleic Acids

Thomas Butler

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree: Physics

UMI Number: 3265304

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



UMI Microform 3265304

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

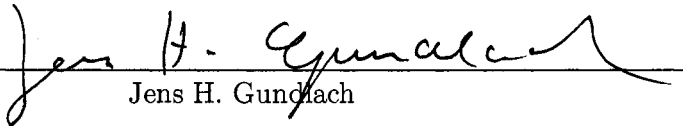
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

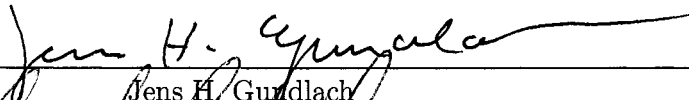
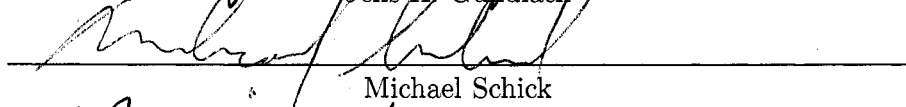
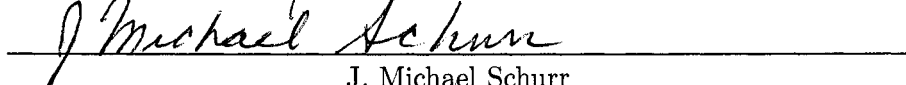
Thomas Butler

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:


Jens H. Gundlach

Reading Committee:


Jens H. Gundlach

Michael Schick

J. Michael Schurr

Date:

6-7-2007

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Proquest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Tom Butler

Date 6/8/07

University of Washington

Abstract

Nanopore Analysis of Nucleic Acids

Thomas Butler

Chair of the Supervisory Committee:
Professor Jens H. Gundlach
Physics

Nanopore analysis of nucleic acids is a novel technique to study the physical properties of nucleic acids at the single-molecule level. It also has the potential to be a central component of a fundamentally new DNA sequencing methodology. In this technique, individual single-stranded DNA or RNA molecules are detected and characterized by an ionic current blockade produced when they are electrophoretically driven through a nanometer-scale pore. The primary focus of this thesis is a detailed characterization and interpretation of the sub-level structure of the ionic current blockades produced during the interaction between single-stranded DNA and RNA molecules and the *Staphylococcus aureus* α -Hemolysin protein nanopore. These investigations reveal important mechanistic details of the interaction between nucleic acids and α -Hemolysin. The results presented in this thesis contribute significantly to the growing utility of nanopore analysis as a model system to study biologically relevant physical and chemical processes, and they may help with the development of novel, nanopore-based technological applications. A secondary focus of this thesis is the initial steps toward genetically engineering *Mycobacterium smegmatis* porin A as an alternative protein nanopore for nucleic acid analysis. Successful creation of an additional protein nanopore represents a substantial scientific and technological breakthrough.

TABLE OF CONTENTS

	Page
List of Figures	iv
List of Tables	vi
Chapter 1: Introduction	1
1.1 Principle of nanopore detection	1
1.2 Utility of nanopore-based nucleic acid analysis	3
1.2.1 Technological applications	5
1.2.2 Biological relevance	7
1.3 Thesis overview	7
Chapter 2: Structure and properties of α -hemolysin and single-stranded nucleic acids	9
2.1 <i>Staphylococcus aureus</i> α -Hemolysin: The prototypical nanopore	9
2.1.1 Biological role of α -HL	9
2.1.2 α -HL crystal structure	9
2.1.3 Conductance and gating behavior of α -HL	12
2.2 Nucleic acid structure and physical properties	13
2.2.1 Primary structure	13
2.2.2 Secondary structure	15
2.2.3 Nucleic acid charge distribution	16
2.2.4 The role of small ions	16
Chapter 3: Apparatus and procedures	18
3.1 Overview	18
3.1.1 Apparatus	18
3.1.2 Experimental procedure	18
3.2 Current amplification electronics	20

Chapter 4:	Structure of ionic current blockades	23
4.1	Introduction	23
4.2	Materials and methods	24
4.3	Results and discussion	25
4.3.1	Blockades demonstrate rapid transitions between well-defined conduc- tance levels	25
4.3.2	Molecular interpretation of blockade sub-states	27
4.3.3	Deep State statistics	30
4.3.4	Mid State statistics	35
4.3.5	Voltage dependence of the vestibule configuration	37
4.3.6	Polymer orientation and the vestibule configuration	41
4.3.7	Kinetic model of the vestibule configuration	43
4.3.8	Interpretation of the vestibule configuration model	45
4.3.9	Polymer capture and the vestibule configuration	47
4.4	Summary and Conclusions	50
Chapter 5:	Determination of RNA orientation during translocation	53
5.1	Introduction	54
5.2	Materials and methods	54
5.3	Results	55
5.3.1	$rA_{25}C_{50}$ and $rC_{50}A_{25}$ signals reflect diblock composition	55
5.3.2	Ambiguous signals	57
5.3.3	Analysis and classification of events	58
5.3.4	Deep State step signal depends strongly on RNA composition	62
5.3.5	Current levels and RNA orientation during translocation	62
5.3.6	Mid State characteristics	67
5.3.7	5' versus 3' capture asymmetry	69
5.3.8	rA_{50} Deep State step signals	69
5.4	Discussion	70
5.5	Summary and conclusions	73
Chapter 6:	Engineering <i>Mycobacterium smegmatis</i> porin A for nucleic acid analysis	74
6.1	Motivation for development of a new nanopore	75
6.1.1	Present day limitations of α -HL for DNA sequencing	75
6.1.2	Progress in solid-state nanopores	76
6.1.3	Scientific importance of additional pores	76

6.2	<i>Mycobacterium smegmatis</i> porin A is a promising candidate for nanopore analysis of nucleic acids	76
6.2.1	Biological role of MspA	76
6.2.2	MspA structure is promising for nanopore analysis	77
6.2.3	Additional advantageous characteristics of MspA	79
6.3	Experiments with Wild-Type MspA	81
6.4	Possible mechanisms prohibiting interaction between DNA and wild-type MspA	84
6.5	Engineering MspA to facilitate electrophoretic detection of DNA	86
6.6	Progress on experiments with MspA mutants	87
6.6.1	D90S/D91S shows reduced gating, but no interaction with DNA . . .	87
6.6.2	The MspA triple mutant D90S/D91S/D93N exhibits an increased rate of transient current blockades in the presence of ssDNA	88
6.7	Outlook	97
	Bibliography	98

LIST OF FIGURES

Figure Number	Page
1.1 Nanopore detection of DNA	2
1.2 A DNA molecule in the α -HL nanopore	4
1.3 Nanopore DNA sequencing concept	6
2.1 Ribbon diagram of α -HL crystal structure	10
2.2 Cross-section through a space-filling model of α -HL	11
2.3 DNA primary structure	14
3.1 Experimental apparatus	19
3.2 Electronics noise	22
4.1 Example ionic current traces illustrating the sub-state structure of polynucleotide induced current blockades	26
4.2 Molecular interpretation of current blockade sub-states	27
4.3 Statistical characteristics of the Deep State	31
4.4 Duration distribution of the extremely long dT_{50} Deep States	32
4.5 rC_{50} Deep State duration distribution	34
4.6 Statistical characteristics of the Mid State	38
4.7 Voltage dependence of vestibule configuration	40
4.8 M1 State duration distributions grouped by polymer orientation	42
4.9 Voltage dependant dA_{50} data fitted by a kinetic model	46
4.10 Polymer capture times	48
5.1 Examples of RNA homopolymer and copolymer blockades	56
5.2 Statistical characteristics of the Deep State for RNA homopolymers and diblock copolymers	57
5.3 Translocation event classification scheme	59
5.4 Example fit to a $rC_{50}A_{25}$ translocation event	60
5.5 Event classification results	63
5.6 Distributions of Deep State current levels	64
5.7 Diagram of relationship between RNA orientation and Deep State step signals	66
5.8 Distributions of Mid State durations	68

5.9	rA_{50} step signal characteristics	71
6.1	MspA structure	78
6.2	Comparison of MspA and α -HL	80
6.3	Conductance steps from insertion of individual wild-type MspA porins	82
6.4	Conductance and gating behavior of wild-type MspA	83
6.5	Distribution of conductance steps from insertion of SSN-MspA pores	89
6.6	Gating and conductance-voltage behavior of SSN-MspA pores	90
6.7	Example blockade events from an SSN-MspA pore in the presence of 40 μ M dA_{50}	92
6.8	SSN-MspA gating in the presence and absence of dA_{50}	94

LIST OF TABLES

Table Number	Page
4.1 Distributions of blockade types	26
4.2 Mid State characteristics	39

ACKNOWLEDGMENTS

I have had a number of excellent mentors during my graduate studies at the University of Washington. Jens Gundlach motivated the inception of our nanopore analysis effort and has been an excellent adviser, mentor, colleague and friend. Mark Troll has been a wonderful co-adviser, contributing much-needed biophysical and biochemical expertise to our research efforts. I am particularly grateful to Mark for his efforts to initiate our collaboration with Michael Niederweis. The support and advice of Bertil Hille has been invaluable to our project. He has loaned us two patch-clamp amplifiers, made us aware of the existence of MspA, and has provided very thorough and thoughtful comments on our manuscripts. J. Michael Schurr has also kindly supported us with expertise and thoughtful commentary on our manuscripts.

Jeff Sherman and Conor Buecheler have been wonderful friends and colleagues during my time at the University of Washington. Their companionship has been invaluable to my survival in graduate school. I also thank Ben Stottrup, Sarah Veatch and Kareem Kazkaz for being great friends and mentors. I have had the pleasure of working with Hannah Gelman, Lia Ball, and Risa Wong, three very talented undergraduates who have made significant contributions to our research effort. The recent exciting results with MspA are the product of hard work and persistence by Ian Derrington. His efforts have enabled us to make substantial progress on the MspA project while I have been occupied writing this thesis.

The IGERT program and center for nanotechnology provided the seed funding to get this project up and running. The University of Washington Royalty Research Fund and the University of Washington Physics Department provided the funding to keep us going over the past few years.

Most importantly, I thank my family. I am truly blessed to have such an amazing wife

and such an adorable son. I love them more than I ever imagined was possible and I cannot wait to see where the future leads us. The gratitude I feel toward my mother and father is beyond expression. It is my greatest hope in life that I can do as well for my children as they have done for my sister and I.

Chapter 1

INTRODUCTION

Nucleic acids are of fundamental scientific importance. Deoxyribonucleic acid (DNA) encodes and maintains the genetic information of living organisms. The ability to sequence DNA and investigate organisms at the genomic level has established a new paradigm for biological and biomedical research. Genomic data has the potential to enable significant advances in the diagnosis and treatment of disease, to give new insight into evolution, and to enable a much deeper understanding of how living systems function. Ribonucleic acid (RNA) is responsible for a variety of essential cellular functions. Genes encoded in DNA are transcribed into messenger RNA which carries the information to a ribosome for translation into a protein. The ribosome itself is a protein/RNA complex. Transfer RNA serves as an adapter that enables translation of the messenger RNA sequence into the appropriate amino acid sequence. The genetic information of some viruses is even carried as RNA. Understanding the physical and chemical details of cellular processes involving nucleic acids is essential to obtaining a deep, mechanistic understanding of cellular function. Over the past ten years, nanopore analysis of nucleic acids has emerged as a candidate for a revolutionary, fast, low-cost DNA sequencing technique and as novel method to study the physical properties of nucleic acids and pore proteins at the single-molecule level.

1.1 Principle of nanopore detection

Figure 1.1 illustrates the concept underlying nanopore analysis of nucleic acids. Two compartments filled with ionic solution are connected by a single pore with a diameter on the order of a nanometer. An externally applied voltage induces a small but measurable ionic current through the pore. DNA is added to the electrically negative compartment. Because it is a strong acid, DNA has net charge of $-1e$ per nucleotide in aqueous solution. Negatively

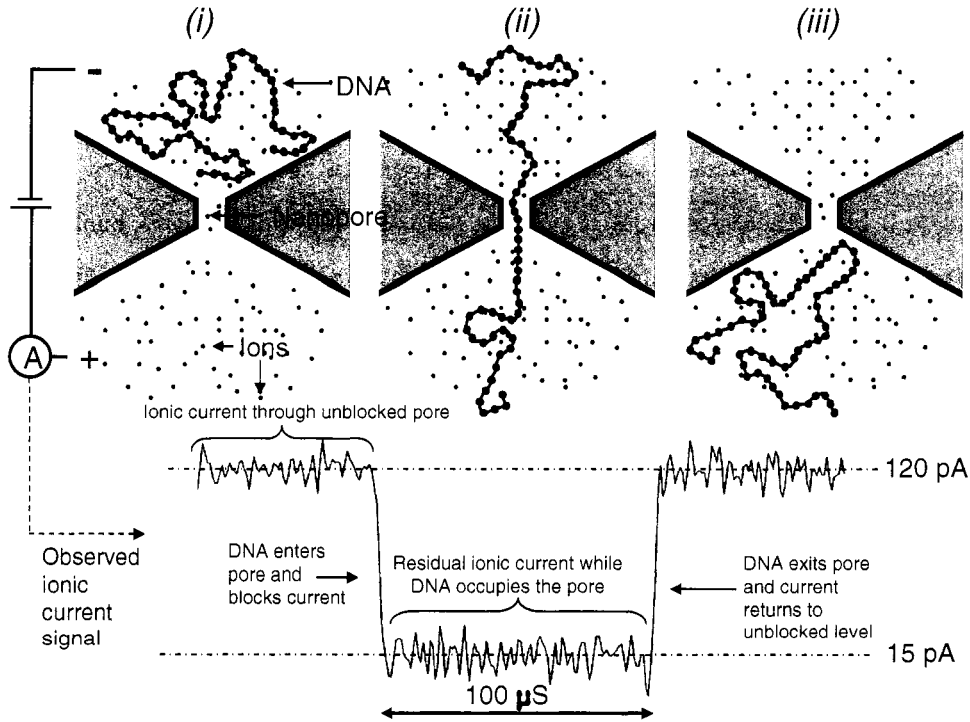


Figure 1.1: Principle of nanopore detection of individual DNA molecules. A nanopore connects two volumes of ionic solution and an externally applied voltage induces a small ionic current through the nanopore. (i) DNA is added to the electrically negative compartment. DNA acquires a net negative charge in solution which allows it to be driven into the nanopore by the voltage gradient. (ii) While a DNA molecule occupies the nanopore it obstructs the flow of ions and causes a blockade in the observed ionic current signal. (iii) After the DNA exits the pore into the electrically positive compartment, the ionic current returns to its initial, unobstructed level. Passage of individual DNA or RNA molecules through a nanopore are thus observed as transient blockades of the ionic current.

charged DNA molecules diffuse randomly about the electrically negative compartment until one molecule happens to pass into the region of high electric field that exists in and near the nanopore. The electric field exerts a force on the DNA and pulls the molecule into the pore. The presence of DNA in the pore impedes the flow of small ions and reduces the observed current. The current remains at a reduced level while the DNA is driven through the pore. Exit of the DNA into the electrically positive compartment is marked by a return of the ionic current to its initial level. Passage of individual DNA or RNA molecules through the nanopore are thus observed as transient blockades of the ionic current. Nanopore detection of nucleic acids was first demonstrated in 1996 by Kasianowicz, Branton, Brandin, and Deamer [34]. The nanopore used in this pioneering work was *Staphylococcus aureus* α -Hemolysin (α -HL), a heptameric, transmembrane protein pore. Figure 1.2 shows a snapshot from a molecular dynamics simulation [46] of a single-stranded DNA molecule being electrophoretically driven through α -HL. The focus of this thesis is the experimental characterization and understanding of the interaction between single-stranded DNA and RNA and the α -HL nanopore. This thesis also describes the initial steps of a project to develop an alternative nanopore for nucleic acid analysis through genetic engineering of *Mycobacterium smegmatis* Porin A.

1.2 Utility of nanopore-based nucleic acid analysis

Nanopore analysis of nucleic acids has many advantageous characteristics. It allows detailed observations of a dynamic, single-molecule interaction at rates of ~ 1 -10 interactions per second. The prototypical nanopore, α -HL, has a known atomic structure [73] and can be engineered at the molecular level [5, 30]. Recently, nanopore detection of nucleic acids has been demonstrated with pores that are drilled or etched in a variety of inorganic membranes [23, 25, 43, 75], the structure and composition of which can also be engineered. Temperature, buffer composition, and the applied voltage can be characterized and controlled. Finally, the instrumentation and infrastructure needed to carry out nanopore analysis of nucleic acids is far simpler and less expensive than that needed for other single-molecule techniques like optical tweezers or atomic force microscopy. These advantageous characteristics combined with the extremely detailed information it produces make nanopore-based DNA analysis

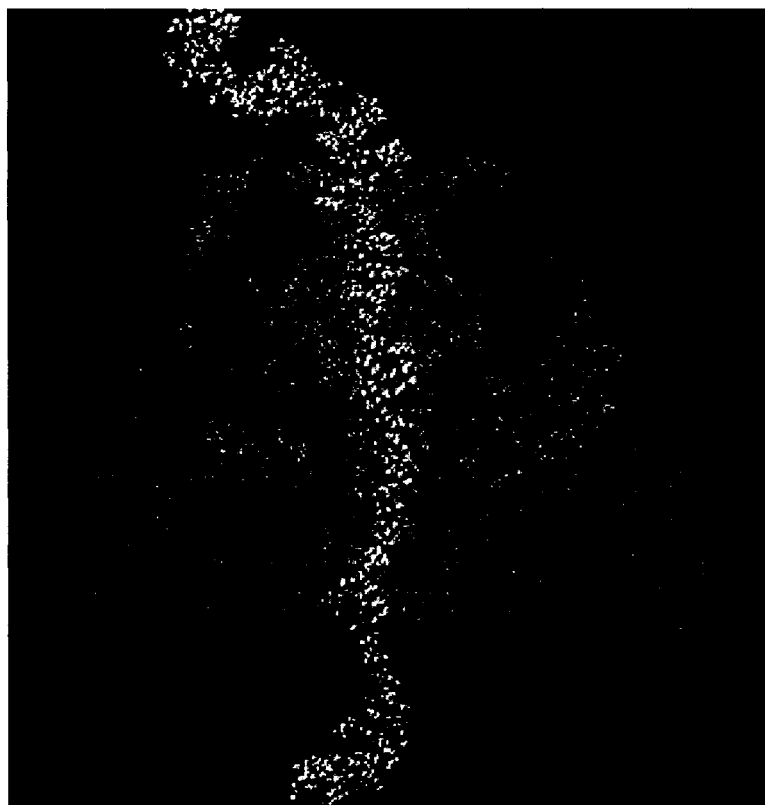


Figure 1.2: A DNA molecule (orange) in the α -HL nanopore (purple). The front half of the nanopore has been cut away to reveal the inner channel through which ssDNA or RNA can pass. The nanopore is sitting in a lipid bilayer (green). This system constitutes the first and most widespread implementation of the nanopore DNA detection process illustrated in Fig. 1.1. The image was kindly supplied by Dr. Aleksei Aksimentiev and was part of a molecular dynamics investigation of DNA translocation through α -HL [46].

a particularly attractive platform for the development of novel biosensors and an excellent model system for investigating biologically relevant molecular processes.

1.2.1 Technological applications

Single-molecule DNA sequencing

Over the past two decades, parallelization and other technological innovations have reduced the cost of DNA sequencing from tens of dollars per base in the 1980s to a few cents per base today. However, even at present-day prices, the cost of sequencing a mammalian-sized genome is still tens of millions of dollars [1]. These high costs significantly limit the number of different organisms for which entire genomes can be sequenced and prohibit the use comprehensive genomic information in individual health care. Novel approaches for sequencing DNA at low cost and high throughput are highly desirable, and nanopore analysis of DNA is one candidate for such a sequencing technology. Since DNA molecules traverse nanopores at a rate of ~ 1 base per μs or higher, nanopore sequencing can in principal provide a \sim seven orders of magnitude increase in sequencing rate over current state-of-the-art techniques [12]. Sequencing at very low copy number would be possible because nanopore detection efficiency is sensitive to DNA concentration and not the absolute amount of DNA. This could eliminate much of the DNA replication and amplification required by present-day methods, adding to the time and cost advantages of nanopore sequencing.

The potential ability of nanopores to sequence individual DNA molecules arises from the fact that the narrow diameter of the pore forces the DNA to translocate in a single-file, sequential manner. Any observable that is sensitive to identity of the short DNA segment localized in the nanopore could be used to derive the sequence. The residual ionic current passing through the pore alongside the DNA is the natural observable to use as a starting point for nanopore DNA sequencing. In this approach, base-specific modulation of the residual ionic current could be directly related to the DNA sequence (Fig. 1.3). As will be discussed later, significant technical obstacles must be addressed to use modulation of the co-passing ionic current as the sequencing observable. Another related approach involves integration of nanoscale electronics within, or next to, a solid state nanopore with the hope

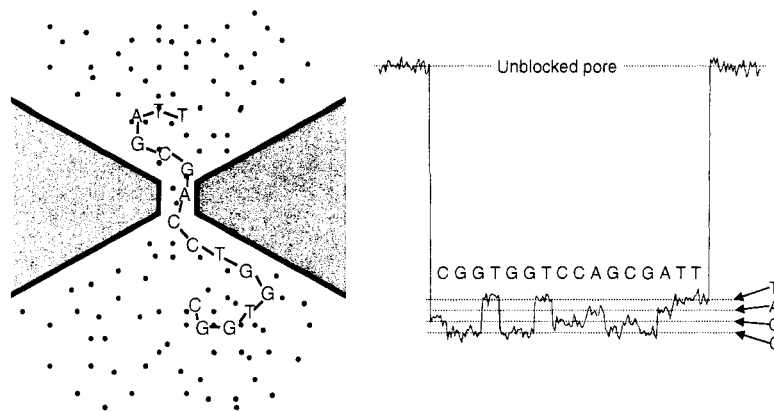


Figure 1.3: Hypothetical, idealized implementation of nanopore sequencing utilizing base-specific modulation of the co-passing ionic current to derive the DNA molecule's sequence.

of observing base-specific modulations in capacitance [21] or in a transverse current across the pore [43, 40]. Optical techniques are also being explored wherein the nanopore is used to both localize the DNA and activate different sequence-specific fluorophores the instant they pass through the pore [38]. All of these approaches are in an exploratory phase and proof-of-principle nanopore sequencing has yet to be demonstrated.

Stochastic sensing

The ability of nanopores to detect and characterize individual molecules via modulation of an ionic current has been termed “stochastic sensing” [5]. Proof-of-principle investigations demonstrated that nanopores can measure the concentration of DNA molecules in the vicinity of the pore [50] and distinguish between populations of very similar DNA molecules [30, 51, 79]. Sensitivity to concentration motivated the evaluation of nanopores as alternative detectors in conventional capillary electrophoresis [62]. Concentration measurements of several target molecules at the same time with a single α -HL pore enabled rapid, accurate assessment of the quality and purity of nucleic acid preparations [82] and was utilized in simultaneous detection of two different proteins [36]. Recently, Tropini *et al.* demonstrated a nanopore-based method for rapid detection of mutations in a DNA sequence [78], a technique which may find clinical application in genotyping and detection of single nucleotide

polymorphisms. Stochastic sensing of a variety of target molecules other than nucleic acids is also being pursued [5].

1.2.2 *Biological relevance*

Nanopore analysis of DNA is a versatile model system that allows detailed investigation of a variety of physical and chemical mechanisms that are important constituents of many biological processes. For example, infection of bacteria by phage viruses [41, 66], gene transfer during bacterial conjugation [11, 20], transfer of T-DNA from *Agrobacterium* to plant cells [86], and a variety of protein transport systems [70] involve the translocation of a long polymer across a membrane through a narrow channel. Confinement of part of the polymer within the pore represents an entropic barrier to translocation [4, 17, 49, 57, 58, 76] that is shared by all these processes. Nanopore analysis allows detailed investigation of this entropic barrier and may give insight into the importance of entropic effects for *in vivo* polymer translocation. Nanopore analysis has also been used to investigate the kinetics of DNA duplex dissociation [31, 48, 47, 69, 79], a central step in DNA replication and transcription. Experiments with the α -HL nanopore indicate that translocation dynamics are sensitive to RNA secondary structure [2] and to base-specific interactions between the DNA and the amino acid residues of the pore interior [50]. These observations suggest that nanopore analysis can also be used to study nucleic acid structure and DNA/protein interactions. Finally, nanopore analysis has provided a new system in which to study the ion atmosphere of DNA and its effects on DNA dynamics [9, 37, 85].

1.3 *Thesis overview*

This thesis presents a detailed analysis and interpretation of the sub-level structure of current blockades produced during interaction between single-stranded DNA and RNA molecules and the α -HL protein nanopore. Chapter 2 gives a brief review of the structure and properties of the α -HL nanopore and of single-stranded nucleic acids. Chapter 3 describes the nanopore apparatus and our experimental procedures. Chapter 4 characterizes the gross structure of the signals produced by DNA and RNA in α -HL and interprets

this structure with a model of the trajectories a polynucleotide may follow during its interaction with the pore. Chapter 5 presents a more detailed analysis of current blockade sub-structure, focusing on modulations of the ionic current that are related to the sequence of the translocating molecule. This analysis shows that polynucleotide orientation is an important factor to consider when interpreting nanopore data or modeling translocation dynamics. The data and analysis presented in Chapters 4 and 5 give insight into a number of interesting features observed in the signals produced during nanopore analysis and elucidate important mechanistic details of the translocation process. The results contribute significantly to the overall understanding of electrophoretic translocation of polynucleotide molecules through nanopores, benefiting both the development of technological applications of nanopore analysis and the growing utility of nanopore analysis as a model system for biologically relevant physical and chemical processes. Chapter 6 describes the initial steps toward developing an alternative protein nanopore for nucleic acid analysis. The project involves genetically engineering *Mycobacterium smegmatis* Porin A. If successful, the development of a second protein nanopore would be a significant advancement in the field of nanopore analysis.

Chapter 2

STRUCTURE AND PROPERTIES OF α -HEMOLYSIN AND SINGLE-STRANDED NUCLEIC ACIDS**2.1 *Staphylococcus aureus* α -Hemolysin: The prototypical nanopore***2.1.1 Biological role of α -HL*

Staphylococcus aureus is one of the most common causes of bacterial infections in humans. α -HL, also commonly referred to as α -toxin, is one of a number of toxic substances produced by *S. aureus*. It has been demonstrated that α -HL is relevant as a virulent factor in animal infections by *S. aureus*, and it is strongly suspected to play an important role in human infection [7]. α -HL is secreted by *S. aureus* as 33.2 kD water-soluble monomers that bind to the membranes of host cells and oligomerize to form relatively large, non-specific trans-membrane pores. Pore formation compromises the host cell's permeability barrier, leading to a variety of responses including osmotic swelling, lysis, and a complex array of secondary cellular reactions triggered by calcium influx [7]. At low concentrations, binding of α -HL to membranes occurs at specific binding sites that are only present in certain highly susceptible cell types including human monocytes and endothelial cells and rabbit erythrocytes. At high concentrations, non-specific absorptive binding to lipid bilayers enables α -HL to attack non-susceptible cells like human erythrocytes [27]. Translocation of DNA across lipid membranes has not been implicated as an *in vivo* function of α -HL, indicating that the ability of α -HL to allow electrophoretic translocation of DNA is a fortunate coincidence and not a result of evolutionary pressures.

2.1.2 α -HL crystal structure

The structure of α -HL has been determined in atomic detail by X-ray crystallography [73]. This structural knowledge is central to the interpretation of data from DNA and RNA analysis experiments. The fully formed α -HL pore is a complex composed of seven identical

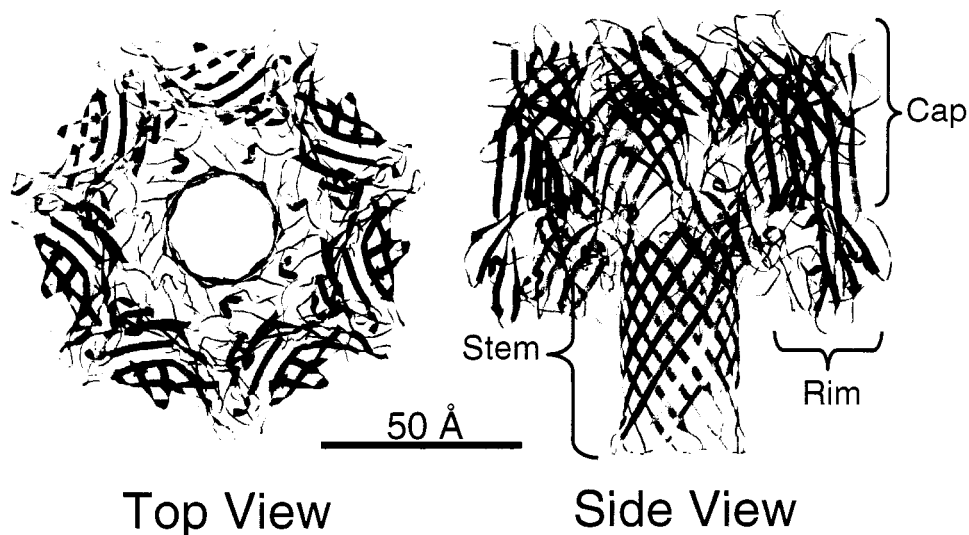


Figure 2.1: Ribbon diagram of α -HL crystal structure. The fully formed pore is a heptameric complex composed of seven identical monomers. The seven monomers are shown as different colors.

monomers, which are shown as different colors in the ribbon diagrams of Fig. 2.1. Figure 2.2 shows a cross-section through a space-filling model of α -HL that reveals the profile of the pore. The overall structure of the heptamer is generally divided into several domains (Fig. 2.1). The cap domain extends ~ 50 Å away from the lipid bilayer and is ~ 100 Å wide. The inside of the cap contains a cavity with a height of ~ 26 Å and a maximum diameter of ~ 46 Å. We refer to this cavity as the pore's "vestibule". The limiting constriction between the vestibule and the external solution has a diameter of ~ 28 Å, and we refer to this constriction as the "entrance" to the pore. The stem domain extends down from the bottom of the cap and spans the bilayer. It is composed of 14 anti parallel β strands in a cylindrical arrangement that is referred to as a " β -barrel". The orientations of the side chains along a given β strand alternate such that even numbered residues compose the exterior surface of the stem and the odd numbered residues compose the interior surface. The exterior residues are generally nonpolar and interact favorably with the hydrophobic core of the lipid bilayer to anchor the pore in the membrane. The residues lining the interior, solvent-accessible region of the stem are predominantly charged or polar. This β -barrel architecture

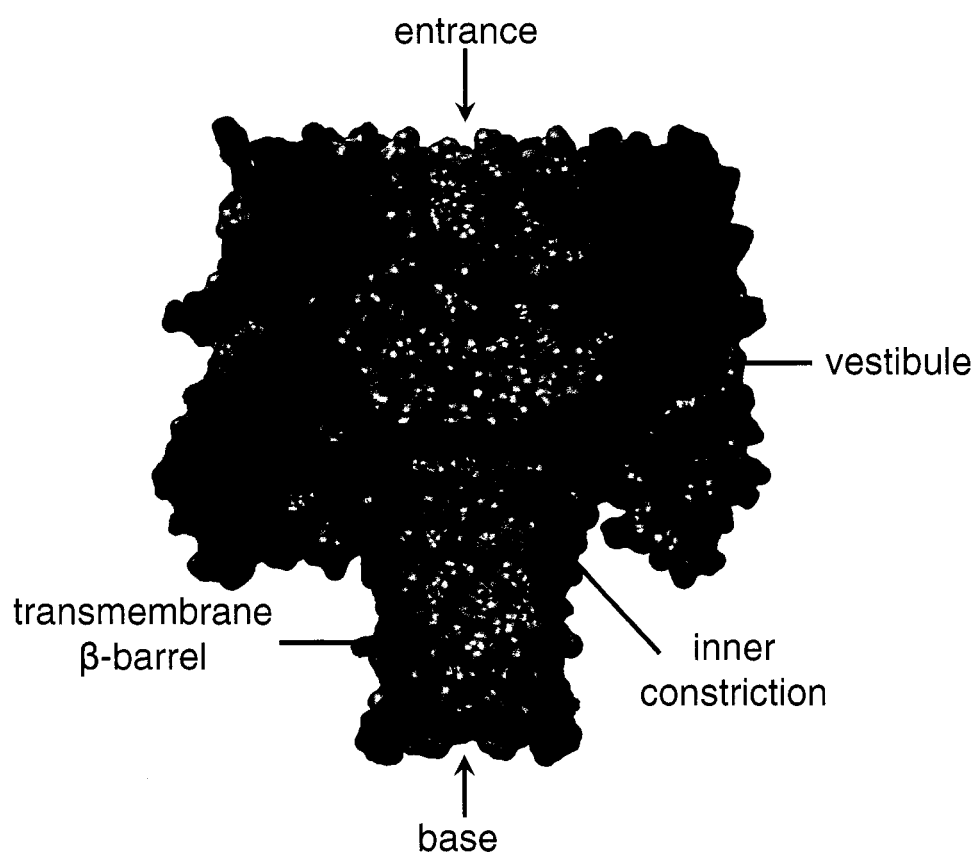


Figure 2.2: Cross-section through a space-filling model of α -HL. Positively charged basic residues are shown in blue and negatively charged acidic residues are in red. The surface of intersection between the protein and the clipping plane is colored dark gray.

allows the heptameric complex to take on the amphiphilic character required to span a lipid bilayer despite the fact that it is composed of hydrophilic monomers [73]. The limiting constriction of the pore is found at the junction between the vestibule and the interior of the stem. We refer to this constriction as the “inner constriction” of the pore. It has a diameter of ~ 14 Å and is lined by a ring of alternating glutamic acid and lysine residues, which are expected to be negatively and positively charged, respectively, under our typical experimental conditions. The C_α - C_α diameter of the stem is 26 Å, and the diameter of the stem’s interior channel ranges from ~ 14 Å at the inner constriction to ~ 24 Å, depending on the volume of the amino acid side chains. The stem “base”, which links the interior of the stem to the bulk solution on the opposite side of the bilayer, has an internal diameter of ~ 24 Å. The interior surface of the base is lined by alternating negatively charged aspartic acid and positively charged lysine residues, while the exterior surface presents another ring of negatively charged aspartic acid residues. Seven rim domains are located at the bottom of the cap. The rim domains are presumed to be in contact with the bilayer and are believed to play a key role in the initial binding of individual monomers to the bilayer membrane [81].

2.1.3 Conductance and gating behavior of α -HL

The conductance of α -HL is linearly dependent on the conductivity of the bathing solution over the concentration range $1 \text{ mM} < [\text{KCl}] < 1 \text{ M}$ [53]. The conductance depends asymmetrically on the polarity of the applied voltage with a conductance ratio of $G_+/G_- = 1.5$ in 1 M KCl at 100 mV (more current flows through the pore when positive voltage is applied to the side of the bilayer that contains the base of the pore) [54]. The pore is slightly anion selective, with an anion to cation transport ratio $t_a/t_c = 1.5$ at pH 7 [53]. In simple KCl solution at pH ~ 7 or greater, the conductance level of a single pore remains stable and relatively free of transient gating for periods of time on the order of hours. However, the pores undergo dose and voltage dependant inactivation in the presence of multivalent cations with the following inhibiting efficiencies: $\text{Zn}^{2+} > \text{Tb}^{3+} > \text{Ca}^{2+} > \text{Mg}^{2+} > \text{Ba}^{2+}$ [53]. Current fluctuations through the open pore peak at pH ~ 5.8 , decreasing with both increasing and

decreasing pH [6, 33]. This phenomenon was attributed to changes in current induced by protonation/deprotonation of select residues within the pore. Recently, Merzlyak *et al.* investigated the conductance, selectivity and current-voltage properties of 24 α -HL mutants, each having a single cystine substitution at a different site. The properties of each mutant were investigated before and after derivitization with positively or negatively charged reagents. Their results suggested that the net charge inside the pore is the primary determinant of its selectivity, and that the charge at the pore entrances is the main factor determining the non-linear, rectifying current-voltage behavior [54].

2.2 Nucleic acid structure and physical properties

A goal of this thesis is to relate the characteristics of the observed current blockades to the physical properties of DNA and RNA molecules. I restrict my discussion of nucleic acid physical properties to single-stranded molecules because α -HL is too narrow to allow translocation of the more well known and extensively studied double-stranded variety.

2.2.1 Primary structure

DNA and RNA are linear polymers with a repeating sugar-phosphate backbone. Attached to each sugar is one of several nitrogen-containing ring compounds called bases. The sequence of bases contains an organism's genetic information. Figure 2.3 shows the primary structure of a short DNA segment. The bases included in Fig. 2.3 constitute the four letter alphabet that is used by organisms to encode information in DNA. RNA has a similar primary structure, differing from DNA only in the presence of a hydroxyl group at the 2' position and the substitution of uracil for thymine in the four letter genetic alphabet. Replacement of the 2' hydrogen in DNA by a 2'-OH in RNA leads to significant differences in the physical properties of the two types of molecules and allows them to perform distinct yet complementary functions in cells [8]. Asymmetry of the primary structure of single-stranded DNA and RNA gives the polymers an intrinsic directionality. This directionality is generally expressed in terms of the 3' and 5' carbons of the ribose/deoxyribose sugar. The 3' end of the polymer is the end which first presents a 3' carbon, and likewise for the 5' end.

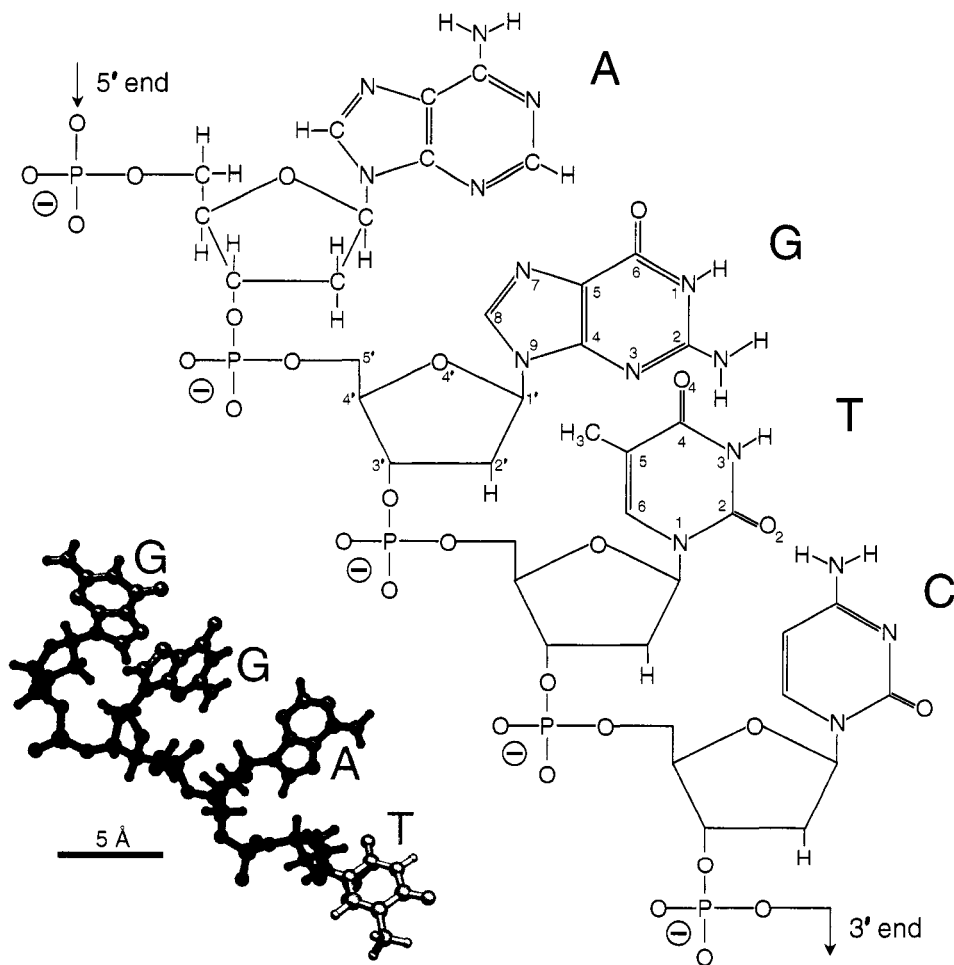


Figure 2.3: DNA is composed of a repeating sugar-phosphate backbone with one of four variable bases – Adenine (A), Guanine (G), Thymine (T) or Cytosine (C) – attached to the sugar. All hydrogen and carbon atoms are explicitly designated in the A nucleotide. The G nucleotide and the U base show the numbering scheme used to designate atoms. RNA differs from DNA in the replacement of the 2' hydrogen by a hydroxyl group and in the use of uracil instead of thymine in its four-letter genetic alphabet. Above pH 1 all phosphate groups are deprotonated and contribute $-1e$ of charge per nucleotide. Between pH 4.2 and pH 9.2 all of the bases are uncharged. Conformational flexibility results from rotation about any of the single bonds in the $P-O5'-C5'-C4'-C3'-O3'-P$ backbone, rotation about the five bonds of the ribose/deoxyribose ring and rotation about the C–N bond linking the base to the sugar. An example of a 3-dimensional conformation that can be assumed by a segment of single-stranded DNA is shown in the lower left. The sugar-phosphate backbone is blue and the bases are the indicated colors. Atomic coordinates are from the NMR structure of ssDNA complexed with Ribonucleoprotein D (PDB ID 1X0F) [18].

2.2.2 Secondary structure

Torsional rotation about the six single bonds that define the polynucleotide backbone, about the five bonds of the ribose/deoxyribose ring and about the C–N bond linking the base to the sugar give polynucleotide molecules significant conformational freedom (Fig. 2.3). Steric constraints and intra-molecular hydrogen bonding lead to sequence-dependant correlations amongst these torsional angles, translating into sequence-dependence of the local secondary structure [8, 67]. For homogeneous single-stranded polynucleotide molecules, the periodicity of the bases can lead to regularity in the local secondary structure along the length of the polymer. Specifically, poly(*rA*) and poly(*rC*) RNA molecules can adopt helical conformations, while poly(*rU*) does not seem to have any regular secondary structure [8]. The poly(*rC*) helix has six nucleotides per turn, a step height of 3.11 Å per nucleotide, and an outer diameter of ~13 Å while the poly(*rA*) helix has nine nucleotides per turn, a step height of 2.82 Å per nucleotide, and an outer diameter of ~22 Å [67]. The melting temperature of the poly(*rC*) is ~50°C, while that of poly(*rA*) is ~40°C [8]. The transition between helical and non-helical conformations occurs over a broad temperature range, suggesting that at 20°C, poly(C) and poly(A) will exist in partially helical, partially random coil conformations [8].

Differences in secondary structure lead to differences in the macroscopically observable behavior of polynucleotide molecules of various sequences. The flexibility of the backbone of a linear polymer is often characterized by the persistence length, ℓ_p , which is the distance along the backbone over which its orientational correlation falls by $1/e$. A value of $\ell_p \simeq 10$ Å was estimated from observations of the free diffusion of ssDNA in a denaturing solution of 8 M urea [77]. More recently, values of $\ell_p \simeq 7.5$ Å have been estimated from force-extension experiments with a Freely-Jointed Chain (FJC) model of flexibility that included a longitudinal stretching modulus [16, 72]. For poly(*rU*) RNA, $\ell_p \simeq 9$ Å was estimated from force-extension data and a Worm-Like Chain model (WLC) [71]. Distinctive elastic properties for DNA versus RNA are suggested by the observation that a WLC provided a poor fit to the ssDNA data [16] while the FJC provided a poor description of the RNA data [71].

2.2.3 Nucleic acid charge distribution

The charge distribution of nucleic acids play a fundamental role in nanopore analysis because the action of the external electric field on the nucleic acid charges provides the force that drives polymers through the pore. The phosphate groups of the polynucleotide backbone have one ionizable proton with a pK_a of about 1, so well above pH 1 each phosphate is deprotonated and has a charge of $-1e$. The five bases also have ionization sites. Under acidic conditions, Cytosine ($pK_a=4.2$), Adenine ($pK_a=3.5$), and Guanine ($pK_a=1.6$) become protonated, giving each base a net positive charge. In basic environments, Uracil ($pK_a=9.2$), Thymine ($pK_a=9.7$), and Guanine ($pK_a=9.2$) become deprotonated, giving each base a net negative charge [8]. All experiments reported in this thesis were conducted at pH 8, where all the bases are neutral and the ionized phosphate groups give the polynucleotides a net charge of $-1e$ per nucleotide. In addition to the net charge on the backbone, a nonuniform electron density enables many of the atoms of a nucleic acid to act as donors or acceptors of hydrogen bonds. Hydrogen bonding is central to DNA and RNA structure and function. For example, it enables the helical structure of double-stranded DNA, allows RNA to adopt functional secondary and tertiary structures, and is a central physical determinant in DNA-protein recognition and binding [8, 67]. The hydrogen bonding ability of DNA and RNA is relevant in nanopore analysis with α -HL because it plays a role in the intrinsic conformational properties of polynucleotide molecules and because hydrogen bonding between DNA and polar residues that line the α -HL pore may be important to translocation dynamics.

2.2.4 The role of small ions

Potassium and chloride ions provide the signal indicative of DNA interaction with α -HL and play an important role in determining the dynamics of the translocation process. All of our experiments were carried out at room temperature ($\sim 20^\circ \text{C}$) with a KCL concentration of 1 M. Under these conditions the ionic solution has a bulk conductivity of 10 Siemens/m and the Debye length is $\sim 3 \text{ \AA}$. This small Debye length implies that the DNA charges will be strongly screened by potassium counterions. This screening minimizes intra-molecular electrostatic repulsion of the DNA backbone, thus stabilizing secondary structure and min-

imizing the electrostatic component of the polymer's stiffness [8, 67]. Charge screening is also expected to reduce electrostatic interaction between the pore and the DNA as well as reducing the effect of the external voltage gradient on the DNA.

Chapter 3

APPARATUS AND PROCEDURES

3.1 Overview*3.1.1 Apparatus*

Figure 3.1 gives an overview of the experimental apparatus. The bilayer cuvette consists of a round Teflon puck containing two $\sim 175\ \mu\text{l}$ reservoirs that are connected by a U-shaped Teflon tube. One end of the tube is open and the other is occluded except for a $\sim 30\ \mu\text{m}$ diameter aperture across which the bilayers are formed. The reservoirs and U-tube are filled with a buffered solution containing 1 M KCl and 10 mM HEPES that is titrated to pH 8 with 2 M KOH. The bilayer forms an extremely high-resistance electrical seal ($> 100\ \text{G}\Omega$) and divides the buffer into two compartments; the *cis* compartment consists of the reservoir that contains the aperture and bilayer, and the *trans* compartment consists of the U-tube and the other reservoir. Electrical connection to the *cis* and *trans* reservoirs is made via direct contact between the buffer and Ag/AgCl electrodes. The *cis* compartment is grounded and voltage is applied to the *trans* compartment with a resistive-feedback current-to-voltage converter. The current-to-voltage converter sources or sinks whatever current is needed to maintain a fixed transmembrane voltage, and in doing so it outputs a voltage signal that is proportional to the current it supplies. After further amplification and filtering, the output voltage signal is digitized and recorded by a programmable data acquisition card.

3.1.2 Experimental procedure

The bilayer cuvette is cleaned by soaking in a self-heating solution of 30% hydrogen peroxide, 70% sulfuric acid or by boiling in 10% nitric acid followed by rinsing several times with HPLC-grade water. The aperture and inside of the U-tube are further rinsed by drawing $\sim 50\ \mu\text{l}$ of water, ethanol, and hexane through the aperture by applying suction to the *trans* side of the U-tube. The rinsed aperture is primed by dropping $\sim 1\ \mu\text{l}$ of a 2 mg/ml

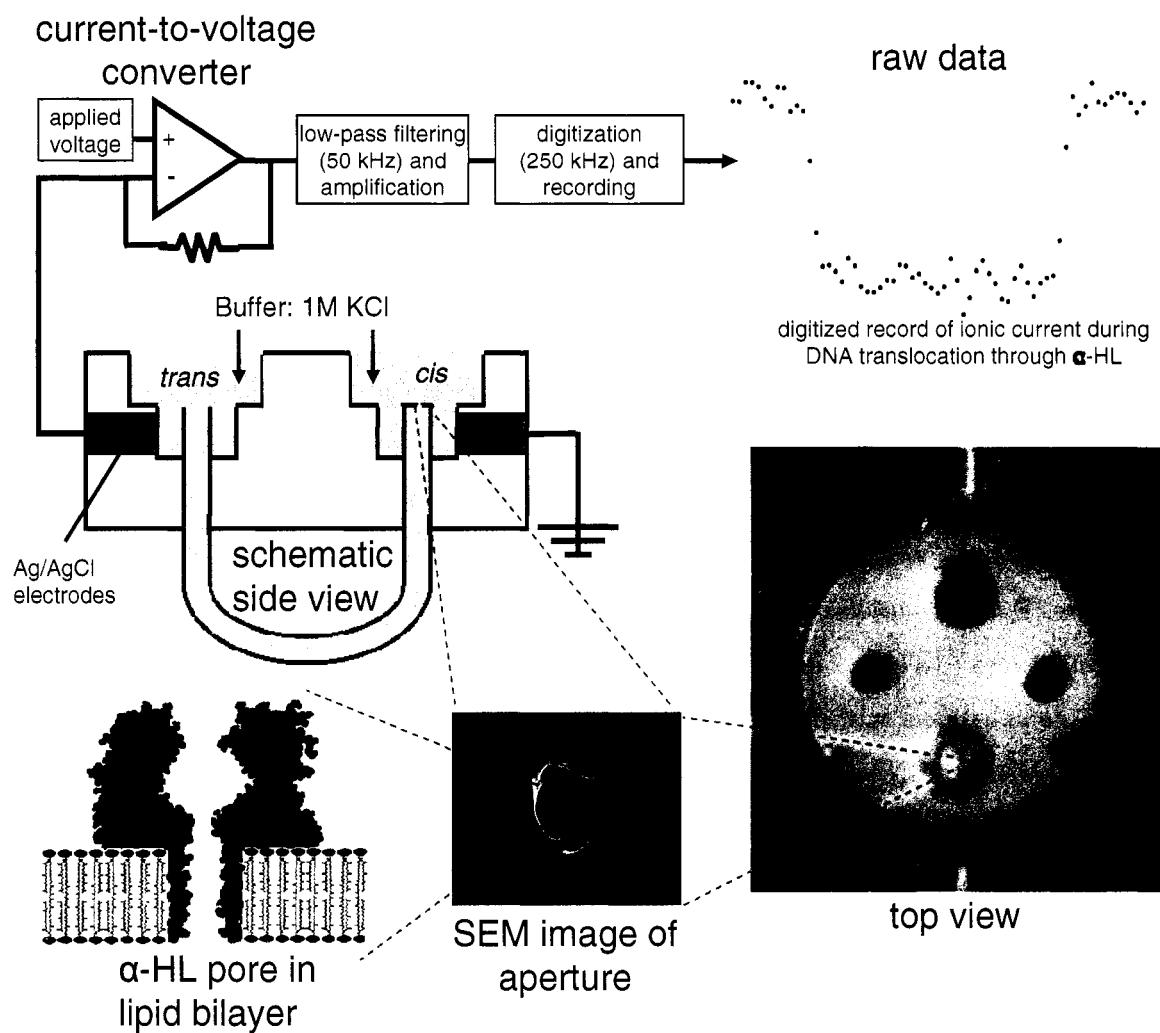


Figure 3.1: Diagram of experimental apparatus.

solution of lipid (Diphytanoyl PC) in hexane onto the aperture surface. After evaporation of hexane from the surface, residual lipid and hexane in the aperture are cleared by applying air pressure to the U-tube from the *trans* side. The primed cuvette is left under vacuum for ~ 15 minutes or at atmosphere for ~ 30 minutes to allow complete evaporation of the hexane. The cuvette is placed in an aluminum support structure, electrodes are inserted and the reservoirs and U-tube are filled with buffer. Electrical connection between the two reservoirs is indicated by saturation of the amplifier. A dried spot of lipid is mixed with hexadecene to a consistency that permits the lipid to be formed into small balls that can be rolled around on a glass surface without leaving a residual solvent trail. One of these lipid balls is attached to the end of a single-bristle brush and used to paint around the aperture surface. After painting, bilayers are formed by briefly occluding the aperture with a $\sim 5 \mu\text{l}$ air bubble blown from a micropipette. After formation of a stable bilayer has been verified by the absence of detectable current through the aperture, monomeric α -HL is added to the *cis* compartment at a concentration of $\sim 1 \mu\text{g/ml}$. Formation of an individual, heptameric transmembrane pore either occurs spontaneously or is induced by repeated bubbling of the aperture. Pore formation is indicated by an increase of the conductance of the bilayer from essentially zero to a characteristic value of 1 nS. The remaining α -HL is then perfused out of the *cis* compartment with fresh buffer to prevent the formation of additional pores. Single-stranded DNA or RNA is added to the *cis* compartment and is driven through the pore by application of positive voltage to the *trans* compartment. The ionic current flowing through the pore is low-pass filtered at a -3dB frequency of 50 kHz and continuously sampled and recorded at 4 μs intervals.

3.2 Current amplification electronics

We used an Axopatch 1B current-to-voltage converter. Figure 3.1 shows a simplified schematic of the current-to-voltage converter circuitry. The DC behavior of the circuit can be easily understood using the simplest model for an operational amplifier with negative feedback:

1. The inputs draw no current.

2. The output does whatever is necessary to keep the voltage of the two inputs equal.

If V_{IN} is applied to the inverting (-) input, then the output voltage, V_{OUT} , will sit at whatever voltage is necessary to keep the noninverting input (+) at V_{IN} . Because the noninverting input draws no current, any input current, I , must flow through the feedback resistor, R_f . V_{OUT} must sit at a voltage such that $V_{IN} - V_{OUT} = IR_f$. Thus, by measuring $V_{IN} - V_{OUT}$ one can deduce the input current $I = (V_{IN} - V_{OUT})/R_f$. Use of a $\sim G\Omega$ -sized feedback resistor allows conversion of picoamp input currents into millivolt output voltages. The time-dependant response of the circuit to changes in input current is determined by the characteristics of the operational amplifier, the size and stray parallel capacitance of the feedback resistor, and the total input capacitance [68]. When acquiring data we limit the bandwidth of our Axopatch 1B to 50 kHz with a four-pole, low-pass Bessel filter giving us a 10-90% rise time of $\sim 7 \mu s$.

Noise in the amplification electronics is an important consideration as it will mask any potential current fluctuations that are related to the interaction between nucleic acid molecules and the pore. Figure 3.2 shows the power spectrum of the signal coming out of our amplifier when it is in an open circuit configuration. In this configuration there are two primary mechanisms of noise generation. The first is Johnson noise in the feedback resistor with a spectral density given by $S_I(f) = 4kT/R_f$. The Axopatch 1B has a 500 M Ω feedback resistor, and at room temperature this gives $S_I(f) = 3.2 \times 10^{-5} \text{ pA}^2/\text{Hz}$, which is shown as the dashed horizontal line in Fig 3.2. Johnson noise does not vary with frequency and it is the leading noise source below 5 kHz. However, it is only a small fraction of the overall noise. The leading noise source in our recordings arises from coupling of voltage fluctuations at the input of the amplifier to the total capacitance seen at the input, C_{in} . C_{in} includes the intrinsic capacitance of the input FET itself and the capacitance of the leads connected to the FET. The capacitively-coupled noise has a spectral density of $S_I(f) = (2\pi f C_{in})^2 S_V(f)$, where $S_V(f)$ is the amplifier voltage noise spectral density, which is generally constant with frequency [22]. This noise increases with frequency as f^2 and is the reason we limit our recording bandwidth to 50 kHz. Figure 3.2 shows the f^2 component of the noise in our amplifier, along with a minor component proportional to f which is due to lossy dielectrics

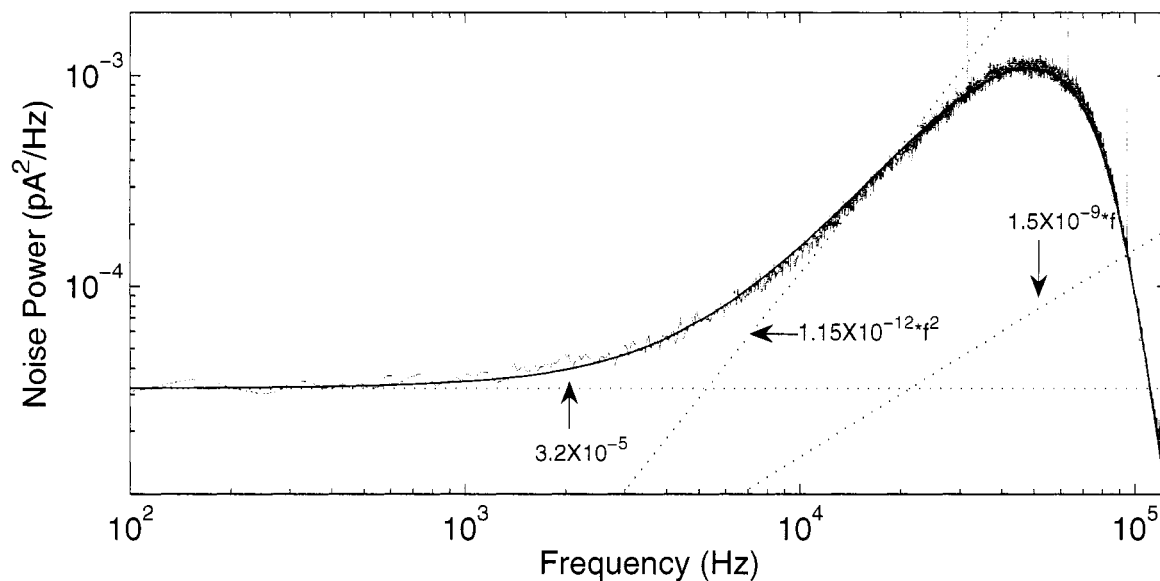


Figure 3.2: Power spectrum of the current noise in the Axopatch 1B when it is in an open circuit configuration. The gray trace shows the measured power spectrum with the 8-pole Bessel filter set to a -3dB frequency of 50 kHz. The dashed lines show the individual noise components as described in the text, and the solid line through the measured power spectrum shows the sum of the noise components convolved with the gain function of a 8-pole low-pass Bessel filter.

in the packaging and in capacitors [42]. The solid curve through the measured power spectrum in Fig. 3.2 is derived from the sum of the three indicated noise sources convolved with the transfer function of a 8-pole, low-pass Bessel filter. Connection of the headstage to the cuvette with a bilayer increases the overall noise by $\sim 15\%$, due primarily to the input capacitance added by the bilayer, and insertion of a pore causes a further moderate increase in noise, primarily at lower frequencies due to Johnson noise in the pore and the shot noise of the current.

Chapter 4

STRUCTURE OF IONIC CURRENT BLOCKADES

Abstract

We characterized the sub-state structure of current blockades produced when single-stranded polynucleotide molecules were electrophoretically driven into the α -HL protein pore. These blockades frequently contained sub-states where the ionic current was reduced by $\sim 50\%$. We associated most of these sub-states with a molecular configuration where a polymer occupies only the vestibule region of the pore, though a few appeared related to a polymer occupying only the transmembrane β -barrel region of the pore. The duration of the vestibule configuration depended on polymer composition and on which end of the polymer, 3' or 5', subsequently threaded into the narrowest constriction and initiated translocation. Below ~ 140 mV a polymer was more likely to escape from the vestibule configuration against the applied voltage gradient, returning to its initial side of the bilayer. At higher voltages a polymer was more likely to follow the voltage gradient by threading through the narrowest constriction and translocating through the pore. Increasing the applied voltage increased the residence time of a polymer in the vestibule configuration. A semi-quantitative model of these trends suggests that escape has a stronger voltage dependence than threading, and that threading is sensitive to polymer orientation while escape is not. These results emphasize the utility of α -hemolysin as a model system to study biologically relevant physical and chemical processes at the single-molecule level.

4.1 Introduction

Current blockades observed during the interaction between single-stranded polynucleotide molecules and α -HL often demonstrate well-resolved sub-states [2, 36]. A “Deep State” where the current is reduced to ~ 0 -30% of the unobstructed “Open State” is generally associated with translocation of a polynucleotide molecule through the pore [2, 10, 34,

52]. The dependence of the duration, depth, and frequency of occurrence of the Deep State on polymer length, applied voltage, temperature, and polymer composition have been experimentally characterized [26, 34, 50, 51, 52, 62]. A growing amount of theoretical work has sought to develop a model that can encompass as many of these observations as possible [4, 44, 49]. A “Mid State” where the current is reduced to 30-85% of the Open State is also often observed [2, 35, 82]. Previous studies with double-stranded DNA constructs [31, 79] indicate that a Mid State can be produced when a polynucleotide molecule is resident only in the vestibule portion of the pore. The dynamics of single-stranded polynucleotide molecules in the α -HL vestibule have been investigated in simulation studies [39, 60], but the detailed characteristics of Mid States produced by these polymers have not yet been investigated experimentally. We analyzed the sub-state structure of current blockades produced by single stranded polynucleotide molecules in the α -HL pore. We focused primarily on the vestibule-associated Mid State, exploring how its characteristics depend on polymer composition, polymer orientation, and applied voltage.

4.2 Materials and methods

An overview of our experimental apparatus and procedures is given in Chapter 2. Monomeric α -HL was purchased from EMD Biosciences (San Diego, CA), Diphytanoyl-PC was from Avanti Polar Lipids (Alabaster, AL) and all other reagents were from Sigma. After formation of a single pore, RNA or ssDNA was added to the *cis* compartment at a concentration of 1-2 μ M. In this configuration the α -HL pore was oriented such that polynucleotide molecules were driven into the pore through its extra-membranous vestibule region. All experiments were performed at $21 \pm 2^\circ\text{C}$ in 1 M KCl, 10 mM HEPES/KOH buffered at pH 8.

Homogeneous ssDNA oligonucleotides dA_{50} , dC_{50} , and dT_{50} were synthesized by Integrated DNA Technologies (Coralville, IA) and homogeneous RNA oligonucleotides rA_{50} , rC_{50} , and rU_{50} were synthesized by Dharmacon RNA Technologies (Lafayette, CO). None of these synthetic oligonucleotide constructs were phosphorylated at either the 5' or 3' end. All samples were PAGE purified by the suppliers. We re-suspended the oligonucleotide samples in the experimental buffer and stored them at -20°C until immediately before use.

4.3 Results and discussion

4.3.1 Blockades demonstrate rapid transitions between well-defined conductance levels

Entry of individual polynucleotide molecules into the α -HL nanopore caused current blockade events with two primary conductance states; a Mid State where the ionic current was between 30% and 85% of the unobstructed Open State, and a Deep State where the current was between 0% and 30% of the Open State (Fig. 4.1). The rise time of Open to Mid or Mid to Deep transitions was generally comparable to the $7\ \mu s$ limit imposed by our 50 kHz anti-aliasing filter. We thus modeled the structure of each blockade event as a sequence of Mid and Deep conductance states. Following typical threshold-based nanopore event detection algorithms [50, 51, 62, 82] we defined an “Open threshold” at $I_{OS} - 2\sigma_{OS}$ and a “Deep Threshold” at $0.3I_{OS}$, where I_{OS} is the average Open State current and σ_{OS} is the rms noise of the Open State (Fig. 4.1). Mid States were identified as periods when the ionic current was between the Open and Deep thresholds, and Deep States were identified as periods when the ionic current was below the Deep threshold. We parameterized each event by the durations and average currents of its constituent sub-states, and we required that a sub-state have a duration of $30\ \mu s$ or longer to be included in our description of the event.

We focused our analysis on the five types of events shown in Fig. 4.1. We defined a Mid State that occurred in the absence of a Deep State as an “ M_0 ” State, a Mid State that directly preceded a Deep State as an “ M_1 ” State, and a Mid State that directly followed a Deep State as an “ M_2 ” State. All Deep States were defined as “ D ”. We classified events according to their sub-state sequence. Table 4.1 lists the percentages of event types observed for different polymers in experiments with 120 mV applied across the bilayer. M_0 , M_1 - D and D were the most common types of events for all six polymers we investigated. Events with M_2 States occurred with small but appreciable frequency only for rA_{50} , dT_{50} and rU_{50} . Events that were not well described by any of the five event types shown in Fig. 4.1 were identified by one or more of the following characteristics: (1) the Mid State had an excursion into the Deep State that was longer than $22\ \mu s$ but shorter than $30\ \mu s$, too long to be attributed to electronic noise but too short for characterization as a Deep-level signal,

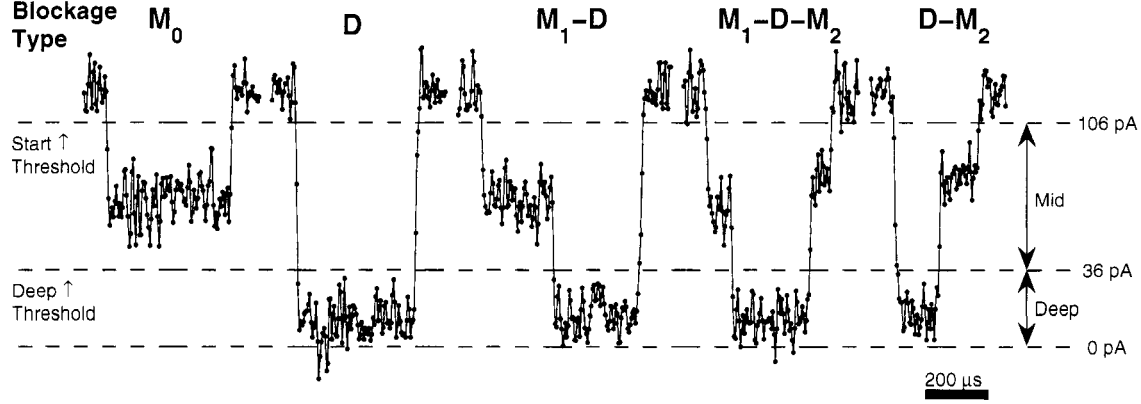


Figure 4.1: Example ionic current traces illustrating the sub-state structure of polynucleotide induced current blockades. These traces are from an experiment with dT_{50} at 120 mV. The horizontal lines show the zero current level and the thresholds used to subdivide blockades into constituent Mid and Deep States. We defined three kinds of Mid States: M_0 States occur in the absence of a Deep State, M_1 States directly precede a Deep State, and M_2 States directly follow a Deep State. All Deep States were abbreviated as “D”. We classified each event by its sub-state sequence. Our names for the various event types appear above the corresponding traces.

Table 4.1: Distributions of current blockade event types for the six polymers we investigated at +120 mV. The five explicit event types are illustrated in Fig. 4.1. “Other” type events are those that were not well described by any of the five explicit types. Values are given as percentages of the total number of events, averaged over repeated experiments. Uncertainties give the standard error of the mean, and the number of times each experiment was repeated is shown after the polymer name.

Polymer	M_0	D	M_1-D	M_1-D-M_2	$D-M_2$	Other
dA_{50} (6)	43 ± 4	28 ± 1	27 ± 4	< 1	< 1	2
dC_{50} (8)	19 ± 4	66 ± 5	7 ± 2	< 1	< 1	7 ± 2
dT_{50} (8)	39 ± 5	15 ± 1	38 ± 3	3	1	4
rA_{50} (7)	42 ± 2	14 ± 1	32 ± 2	2	1	9 ± 3
rC_{50} (5)	20 ± 6	62 ± 7	12 ± 1	< 1	< 1	6 ± 1
rU_{50} (3)	37 ± 8	31 ± 4	12 ± 3	3 ± 1	5 ± 2	12 ± 1

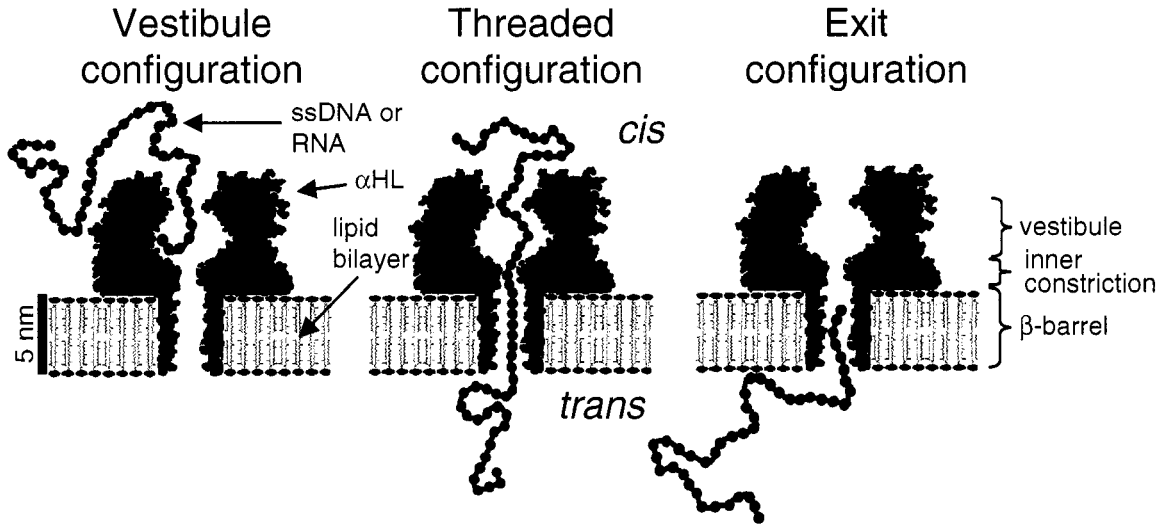


Figure 4.2: Molecular interpretation of conductance states and event structure shown in Fig. 4.1. We hypothesize that M_0 and M_1 states correspond to the vestibule configuration, D states correspond to the threaded configuration, and M_2 states correspond to the exit configuration.

- (2) the rms noise of one of the States was larger than twice the rms noise of the Open State,
- (3) the event demonstrated more than one Deep State.

4.3.2 Molecular interpretation of blockade sub-states

The known structure of the α -HL pore, described in detail in Chapter 1, is central to a physical interpretation of the different types of ionic current blockades. Figure 4.2 shows a schematic cross-sectional view of the pore. On the *cis* side of the bilayer the pore entrance narrows to a diameter of ~ 2.5 nm before opening into the internal “vestibule” region. The vestibule region widens to a maximum diameter of ~ 4.6 nm before narrowing down and joining a short ~ 1.4 nm diameter “inner constriction”. The inner constriction leads into a ~ 5 nm long, ~ 2 nm diameter β -barrel that spans the bilayer and opens into the *trans* compartment. All our experiments were configured such that ssDNA and RNA molecules were driven into α -HL from the *cis* side of the bilayer.

Knowledge of the structure of α -HL [73] combined with a crude volume exclusion model for the obstruction of current through α -HL [13] leads to the simple preliminary physical

interpretation of the various event types shown in Fig. 4.1. When a polymer occupies only the vestibule region of the pore, it reduces the current to the Mid State. This “vestibule configuration” leads to M_0 events if the polymer escapes back into the *cis* compartment. Alternatively, one end of the polymer may thread through the neck region of the pore and into the β -barrel, impeding the passage of ions enough to reduce the current to the Deep State. If the polymer continues threading through the β -barrel, it will eventually exit on the *trans* side of the bilayer. This translocation scenario accounts for M_1 - D type events if the vestibule configuration is longer than our 30 μ s temporal resolution and D -type events if it is shorter. It is also possible, in principle, for a polymer that has threaded into the inner constriction to retract back into the vestibule configuration. In the last two traces of Fig. 4.1, a well-resolved Mid State is observed after the Deep State. It is clear how retraction from the threaded configuration back into the vestibule configuration, followed by escape into the *cis* compartment could account for these M_2 States. However, if the Deep State corresponds to translocation, then the subsequent M_2 State could result if the polymer’s progress through the pore is arrested in an “exit configuration” where its trailing end occupies part of the β -barrel but is no longer threaded through the inner constriction.

The association of the vestibule configuration with reduction of the current to the Mid State and the threaded configuration with further reduction of the current to the Deep State is supported by results from previous investigations. Howorka et al. investigated the binding of free oligonucleotides to complementary strands that were tethered to the interior of the α -HL vestibule [29]. When the transient, double-stranded DNA constructs had sufficiently long single-stranded overhangs, they produced current blockades that fluctuated between Mid and Deep conductance states. Complexes with shorter overhangs produced blockades that only demonstrated Mid States. Similar data and interpretations have been given for experiments with hairpin DNA having variable length duplex regions [15, 80]. These results strongly support the association of the Mid State with the vestibule configuration and the association of the Deep State with the threaded configuration. Further, a recent simulation study that simultaneously computed the time evolution of a single-stranded polymer interacting with the α -HL pore and the corresponding modulation of the ionic current was able to reproduce this relationship between the molecular configurations and

degrees of current obstruction [60].

Homogeneous single-stranded DNA and RNA polymers are highly flexible, so it is more difficult to infer their particular arrangement in the α -HL vestibule than for the more rigid double-stranded DNA constructs. We suggest two distinct ways in which a polymer may enter the vestibule from the *cis* compartment. One possibility is that entrance is initiated by fluctuation of an end segment into the vestibule, and the other possibility is that an internal segment of the polymer forms a transient “loop” which fluctuates into the vestibule and initiates polymer entrance. The results of Vercoutere et al. [79], where ssDNA constructs with hairpins at both the 3' and 5' ends did not cause resolvable Mid State blockades, argue against loop-initiated entrance. We also note two potential arrangements of the polymer in the vestibule configuration. Either the entire 50mer is drawn inside the vestibule, or only part of the 50mer is in the vestibule while the rest of the polymer remains in the *cis* compartment. The latter possibility is illustrated in Fig. 4.2. In the duplex DNA experiments the α -HL vestibule accommodated up to ~ 24 nucleotides and still demonstrated Mid States [29], suggesting that at least half of a 50 nucleotide long ssDNA or RNA molecule could be drawn into the vestibule. Jung *et al.* estimated that the accessible volume of the α -HL vestibule, corrected for hydration of the vestibule surface, is $\sim 32,600 \text{ \AA}^3$ [32]. Deamer *et al.* estimated a molecular volume of $\sim 290 \text{ \AA}^3$ per dAMP [14]. Assuming that each nucleotide is accompanied by 4 water molecules leads to a rough estimate of $\sim 17,000 \text{ \AA}^3$ for the volume occupied by a hydrated dA_{50} molecule. These calculations indicate that the vestibule is in principle large enough to hold an entire ssDNA or RNA 50mer. While it is energetically favorable for the entire polymer to enter the vestibule, the entropic contribution to the free energy is quite complex. As the polymer enters the vestibule there are entropic costs associated with removing each monomer from the bulk *cis* volume [57] and from packing additional monomers into the confining vestibule [65]. However, there is also a potential entropic benefit associated with moving the polymer from a partially filled state, where a segment is restricted to the entrance, to a completely filled state where the polymer is free to reorient within the confining vestibule [58]. It is not *a priori* obvious which of these effects dominate in our experiments, and we will return to the question of how much of the polymer is drawn into the vestibule after presenting our data.

4.3.3 Deep State statistics

Each point in Fig. 4.3 gives the duration and average current of a Deep State. The panels show representative data from one experiment at 120 mV for each of the six polymers we investigated. We display the data on a log-linear scale to allow visualization of the Deep State duration distributions across many orders of magnitude. Clear grouping behavior is observed in all of the panels. Most of these groups correspond to local maxima in the two-dimensional distributions of Deep State durations and currents. However, the groups of long duration states in the dT_{50} and rU_{50} panels do not correspond to local maxima, but appear as “groups” because of the logarithmic scaling of the duration axis. Previous work strongly suggests that the signals in the maxima-associated groups arise from translocation of polymers through the pore [10, 34, 52]. The general characteristics we observed for these maxima-associated signals are in good agreement with previously reported results [2, 34, 50, 51, 82]. These references present extensive characterization and discussion of the Deep State. In this chapter I focus only on the aspects of the Deep State distributions that are important to our characterization and interpretation of blockade structure.

The durations of Deep States in the translocation-associated groups for the five homopolymers other than dC_{50} are appreciably longer than our 30 μs cutoff, suggesting that we were able to detect nearly all translocation events in experiments with these polymers at 120 mV. It is clear from the dC_{50} panel that many translocation-associated states were shorter than our 30 μs cutoff time. Events demonstrating these short translocation signals were classified as “Other”, and we hypothesize that fast translocation accounts for a significant fraction of dC_{50} “Other” type events listed in Table 4.1. In chapter 5 we show that the two groups in the rC_{50} distribution arise from translocation of polymers in either of two possible orientations, 3' first or 5' first. Orientation has also been shown to cause the two groups in the dA_{50} panel [46] and similar grouping in Deep State characteristics of a variety of heterogeneous ssDNA samples [82]. We will utilize this orientation information in our subsequent analysis of the Mid State.

The long-duration Deep States in the dT_{50} and rU_{50} panels of Fig. 4.3 have not been previously reported. For dT_{50} , Deep States with duration longer than 4 ms constitute $\sim 20\%$

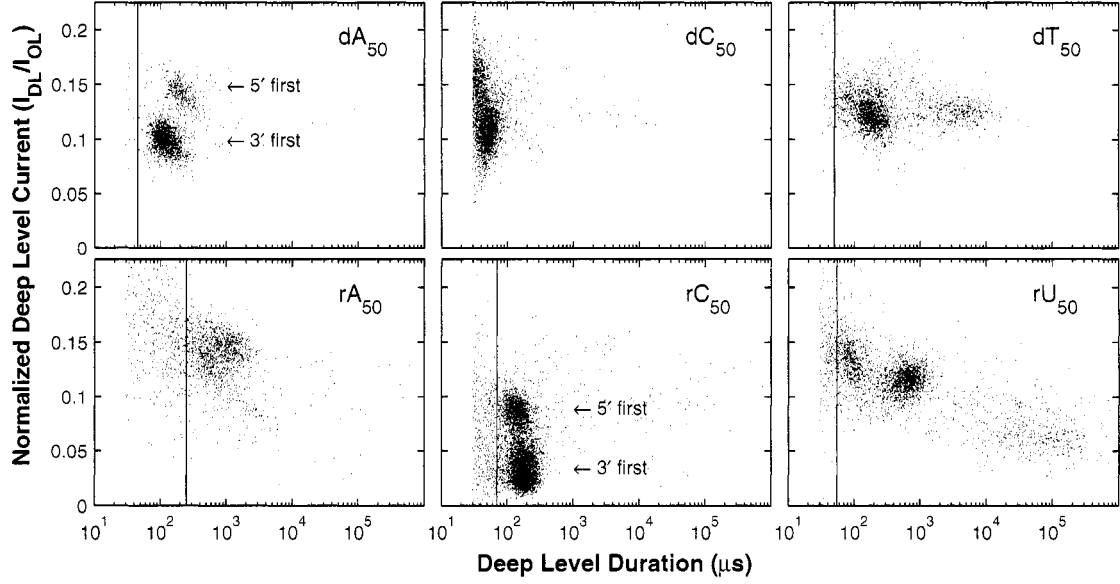


Figure 4.3: Distributions of Deep State durations and currents. Each point corresponds to the duration and normalized average current of one Deep State. Data were acquired with 120 mV applied across the bilayer. We display the data on a semi-logarithmic scale to allow visualization of the Deep State signal distributions across many orders of magnitude in duration. The vertical lines give the locations of minima in the duration distributions (see Fig. 4.5). We only associate States longer than these minima with translocation of polymers through the pore. The longest duration groups of signals in the dT_{50} and rU_{50} panels do not correspond to local maxima in the distributions. These long signals are well described by exponential distributions with time constants of ~ 3.7 ms for dT_{50} (Fig. 4.4) and ~ 94 ms for rU_{50} . They appear as “groups” because of the logarithmic scaling of the duration axis. Previous work found that the two groups in the dA_{50} distributions were related to 3' or 5' first translocation [46]. In chapter 5 we show that there is an analogous correlation between rC_{50} Deep State characteristics and translocation orientation.

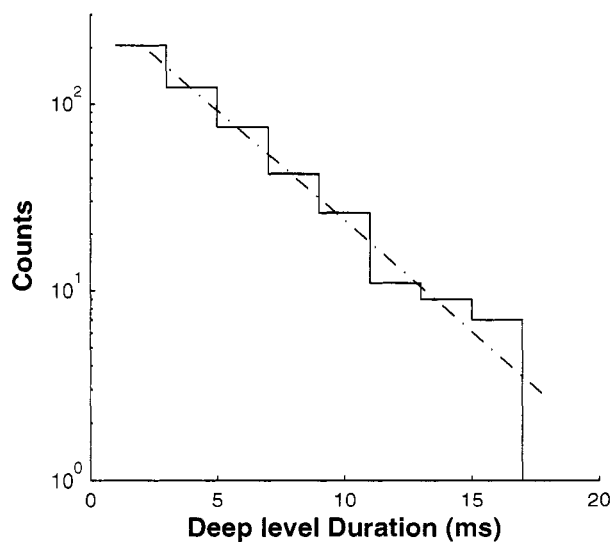


Figure 4.4: Duration distribution of the extremely long dT_{50} Deep States. These signals appear as the rightmost group in the dT_{50} panel of Fig. 4.3. As is clear from this plot, the group of long dT_{50} signals do not correspond to an additional maximum in the dT_{50} Deep State duration distribution. The long signals constitute $\sim 20\%$ of the total number of signals in the dT_{50} panel of Fig. 4.3 and are well described by an exponential distribution with a time constant of ~ 3.7 ms (dotted line).

of the total number of signals shown in the panel and their distribution is well described by a single exponential with a ~ 3.7 ms time constant (Fig. 4.4). Analogously for rU_{50} , Deep States with duration longer than 10 ms constitute $\sim 21\%$ of the total and their distribution is well described by a single exponential with a ~ 94 ms time constant. Previous analysis of Deep level data did not take into account the extremely long Deep States, but found a 10-100 μs exponential time constant for the falloff of the Deep level duration distribution in the vicinity of the maxima [51]. The characteristics of the Mid States preceding and following the long Deep States are similar to those of the translocation-associated signals, suggesting that the long signals also represent translocation. To explain these long Deep States, we hypothesize that during translocation a polymer's forward progress can be significantly inhibited or even stopped entirely for relatively long periods of time. Tangling of the polymer at the neck or binding of one or more bases to the residues lining the β -barrel are two possible causes of such a stoppage. The Deep State would then persist until untangling or unbinding occurred, allowing "normal" translocation to resume and the subsequent exit of the polymer from the pore into the *trans* compartment.

A final important feature in the Deep State distributions is that, with the exception of dC_{50} , we observed minima in the duration distributions between our 30 μs experimental cutoff time and the first maxima (Fig. 4.5). These minima are not readily apparent in Fig. 4.3 due to the logarithmic scaling, but their locations are given by the solid vertical lines. The rise in the distribution for durations shorter than these minima is not consistent with a simple picture of translocation, wherein one expects a monotonic decrease in probability for durations shorter than the most probable duration [44]. Similar minima in Deep State duration distributions have been previously observed [2, 34]. Signals shorter than the minima have been attributed to multiple possible effects such as retraction from the threaded configuration back into the vestibule configuration, very rapid translocation, or translocation of short polynucleotide fragments. Due to this ambiguity in the interpretation of the short Deep States, we only designated signals with durations longer than the minima as translocations (Fig. 4.5). If some of these short signals represent retraction, then we would expect to find events with multiple Deep States, as are often observed in experiments with short double-stranded DNA constructs in the α -HL vestibule [15, 29, 80]. We occasionally

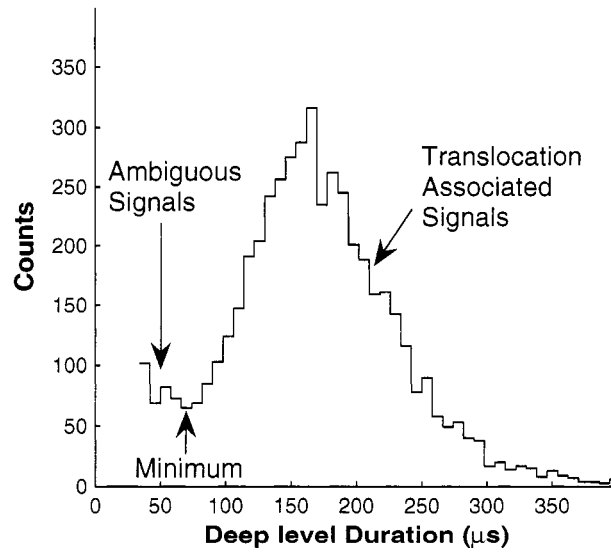


Figure 4.5: rC_{50} Deep State duration distribution. This distribution shows the same data as the rC_{50} panel of Fig. 4.3, but the signals have been binned in duration and are displayed on a linear scale. There is a clear minimum in this duration distribution between our 30 μs experimental cutoff time and the translocation-associated maxima. We observed similar minima in the Deep State duration distributions of other polymers we investigated (Fig. 4.3, solid vertical lines). We designated Deep States that were longer than the minima as “translocation-associated” signals and those shorter than the minima as “ambiguous” signals.

observed such events in experiments with dT_{50} and rA_{50} , but their frequency of occurrence was difficult to quantify because they were rare and difficult to identify with our simple data analysis algorithm. Multiple Deep State events in experiments with the other four polymers were almost never observed. The existence of events with multiple Deep States suggests that retraction can occur and be resolved in our experiments. However, this process seems quite rare and it remains unclear the degree to which it can account for the short, ambiguous Deep States.

4.3.4 Mid State statistics

Figure 4.6 shows the durations and currents of the three types of Mid States, M_0 , M_1 , and M_2 , from experiments at 120 mV with dT_{50} , rA_{50} , and rU_{50} . Each distribution displays a characteristic current, I_M , around which the majority of the signals are clustered. We quantified this characteristic current with a Gaussian fit to the largest peak in a weighted distribution of the currents of the Mid States. The duration of each state was used as the weighting factor, and the longest 5% of signals in each data set were excluded so that the peak location was not dominated by a minority of very long signals. I_M values for the distributions in Fig. 4.6 are shown as solid horizontal lines. We focused our analysis on Mid States of similar current by defining a “primary group” with boundaries at $I_M \pm (w + 2\sigma_{OS}/\sqrt{n_{obs}})$. The term w is an empirical width factor that we set to 7.5 pA. The term $2\sigma_{OS}/\sqrt{n_{obs}}$ accounts for the expected increased statistical variation in the estimated average current of shorter signals, where σ_{OS} is the rms noise observed in the Open State and n_{obs} is the number of samples of the ionic current time series used to calculate the average current of a given Mid State. The dashed curves in Fig. 4.6 show the boundaries defining the primary groups. We used the average duration of the Mid States in the primary group, t_M , to parameterize the Mid State duration distributions. The longest 5% of Mid States within the primary group were excluded from the calculation of t_M to again prevent the statistic from being dominated by a minority population of very long signals. The t_M values for each data set are shown as solid vertical lines in each panel of Fig. 4.6. M_0 and M_1 data for dA_{50} , dC_{50} , and rC_{50} followed the same general patterns as the data shown

in Fig. 4.6, while the occurrence of M_2 States was negligible in experiments with these polymers. Table 4.2 summarizes the average values of I_M and t_M at 120 mV for all six of the polymers we investigated.

In comparing the first two columns of Fig. 4.6, one sees that for a given polymer there is a strong similarity between the M_0 and M_1 distributions. This similarity is also reflected in the I_M and t_M values in Table 4.2 (except for the dT_{50} t_M values). These observations are consistent with the association of both M_0 and M_1 signals with the polymer in the vestibule configuration. Since we interpret M_0 States as escape from the vestibule back into the *cis* compartment, we define an “escape probability” as $N(M_0)/(N(M_0)+N(M_1))$, where $N(M_0)$ and $N(M_1)$ are the number of signals in the primary groups of the M_0 and M_1 distributions, respectively. Average escape probabilities for experiments at 120 mV are listed in Table 4.2 and range between ~ 0.5 and ~ 0.7 . This indicates that at 120 mV, a homogeneous, 50-nucleotide long DNA or RNA molecule residing in the vestibule configuration is slightly more likely to move against the applied voltage gradient and escape back to the electrically negative compartment than it is to follow the gradient and translocate through the pore.

The majority of the M_2 States demonstrate current levels that are clearly distinct from M_0 and M_1 current levels (Fig. 4.6, arrows). Assuming that the Deep States preceding these M_2 signals indicate translocation, we hypothesize that these Mid States occur when a polymer’s exit into the *trans* compartment is hindered while it still occupies part of the β -barrel but is no longer threaded through the inner constriction of the pore. The mechanism that hinders polymer exit could also occur at a previous stage in the translocation process, when the polymer was still threaded through the neck of the pore and blocking the current at the Deep State. In this case one might expect that polymers demonstrating longer, more frequent M_2 States would also demonstrate longer Deep States. This correlation is found in our experiments as dT_{50} and rU_{50} produce the longest Deep States and the most frequent M_2 signals. The M_2 distributions in Fig. 4.6 also show a few signals with current levels similar to the characteristic current levels in the M_0 and M_1 distributions. These M_2 signals are consistent with the hypothesis that a polymer can retract back into the vestibule state after threading into the inner constriction enough to produce a Deep State, but not enough to commit the polymer to translocation. We interpret such signals in the same way

we interpret the existence and rarity of events with multiple Deep States – retraction of a threaded polymer back into the vestibule configuration can be observed, but only very rarely.

Comparing Mid State data for different polymers at 120 mV shows that Mid State characteristics are moderately sensitive to polymer composition. M_0 and M_1 States for dA_{50} , dT_{50} , rC_{50} and rU_{50} all block the current at $I_M/I_{OS} \sim 0.59$, whereas dC_{50} signals show a slightly larger residual current and rA_{50} signals show a significantly smaller residual current. The characteristic durations of M_0 and M_1 States for rA_{50} and dT_{50} are significantly longer than those of the other four polymers. The largest escape probability was 0.69, demonstrated by rU_{50} , and the smallest was 0.48, demonstrated by dT_{50} , with the escape probabilities of the remaining polymers more or less evenly distributed between these two extreme values. Polymer-specific differences in Mid State characteristics could enhance the ability of α -HL to analyze and distinguish polynucleotide samples [36, 51]. However, further work is needed to understand the molecular origins of these differences. For example, the large difference between the Mid State characteristics of dA_{50} and rA_{50} suggests that specific interactions between bases and the residues lining the vestibule is not the primary factor determining the Mid State characteristics of a given polymer.

4.3.5 Voltage dependence of the vestibule configuration

We conducted experiments at a variety of voltages over the range of 100 mV to 200 mV. Data obtained at each voltage were analyzed as described in the previous section. Figure 4.7 shows the voltage dependencies of I_M/I_{OS} , t_M , and the escape probability for dA_{50} , rC_{50} , and rU_{50} . The I_M/I_{OS} and t_M values in Fig. 4.7 were derived from the M_1 distributions as opposed to the M_0 distributions because we observed a large number of M_1 signals for all polymers at all voltages. As we increased the voltage we saw a moderate but clear decrease in the normalized M_1 current level, a substantial increase in the typical M_1 duration, and a substantial decrease in the escape probability to near zero at 200 mV. Voltage curves for dC_{50} , dT_{50} and rA_{50} are not shown, though their trends are qualitatively the same as those shown in Fig. 4.7.

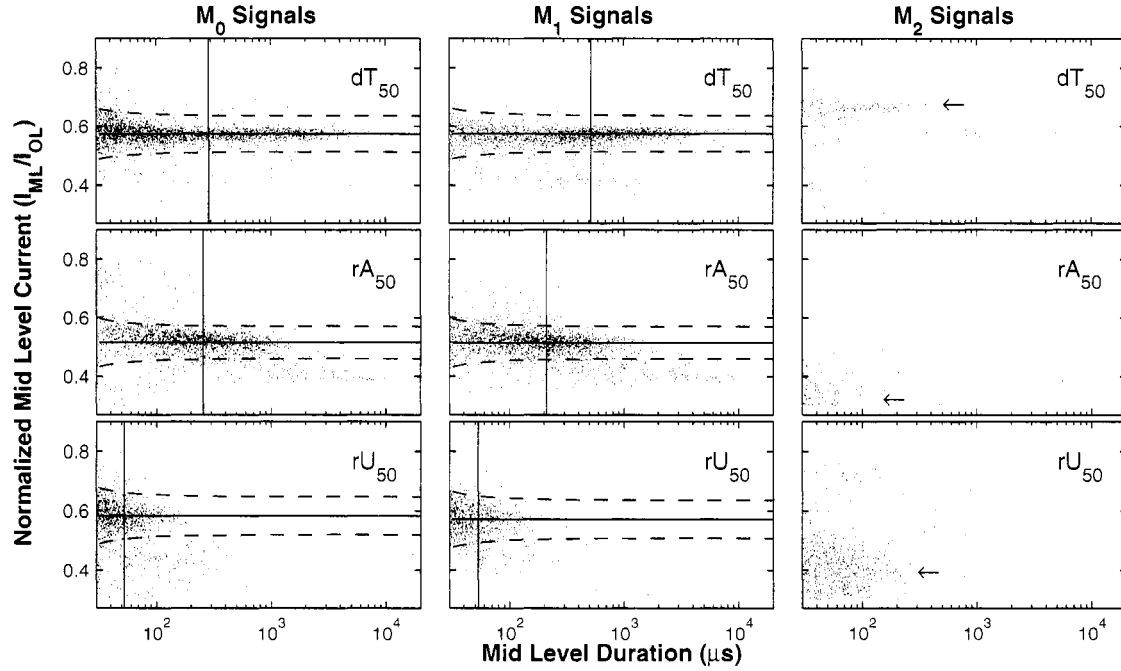


Figure 4.6: Distributions of Mid State durations and currents. Each point gives the duration and normalized average current of one Mid State. Rows show representative data from one experiment at 120 mV with the indicated polymer. Columns correspond to different types of Mid States (see Fig. 4.1). The solid horizontal lines show the characteristic Mid State current, I_M , and the solid vertical lines show t_M , the average duration of the Mid States contained within the dashed curves. We associate M_0 and M_1 States with the presence of a polymer in the α -HL vestibule. Arrows in the M_2 panels point out signals in the M_2 distributions that have different conductance than the M_0 or M_1 signals. We hypothesize that these differing M_2 States are produced when a polymer is held in a configuration where it still occupies part of the β -barrel but is no longer threaded through the inner constriction of the pore. Table 4.2 summarizes I_M and t_M values for the six polymers we investigated.

Table 4.2: Summary of Mid State characteristics for different polymers with 120 mV applied across the bilayer. Values are averages from repeated experiments, and uncertainties are the standard error of the mean. The number of repeated experiments for each polymer is given in Table 4.1. We do not quote an escape probability for dC_{50} because rapid translocation of these polymers made it difficult to accurately estimate the number of translocations. We do not parameterize the rA_{50} M_2 distributions because they do not demonstrate a well defined group of signals.

Polymer	Sub-state	I_M/I_{OS}	$t_M(\mu s)$	P_{escape}
dA_{50}	M_0	0.59 ± 0.01	75 ± 3	0.60 ± 0.05
	M_1	0.57 ± 0.01	69 ± 2	
dC_{50}	M_0	0.65 ± 0.01	54 ± 3	
	M_1	0.64 ± 0.01	57 ± 2	
dT_{50}	M_0	0.58 ± 0.01	288 ± 27	0.48 ± 0.05
	M_1	0.58 ± 0.01	65 ± 83	
	M_2	0.63 ± 0.01	122 ± 12	
rA_{50}	M_0	0.50 ± 0.01	171 ± 17	0.59 ± 0.03
	M_1	0.50 ± 0.01	154 ± 12	
rC_{50}	M_0	0.60 ± 0.02	61 ± 4	0.57 ± 0.07
	M_1	0.59 ± 0.01	59 ± 7	
rU_{50}	M_0	0.58 ± 0.02	53 ± 1	0.69 ± 0.09
	M_1	0.57 ± 0.02	55 ± 1	
	M_2	0.40 ± 0.01	67 ± 2	

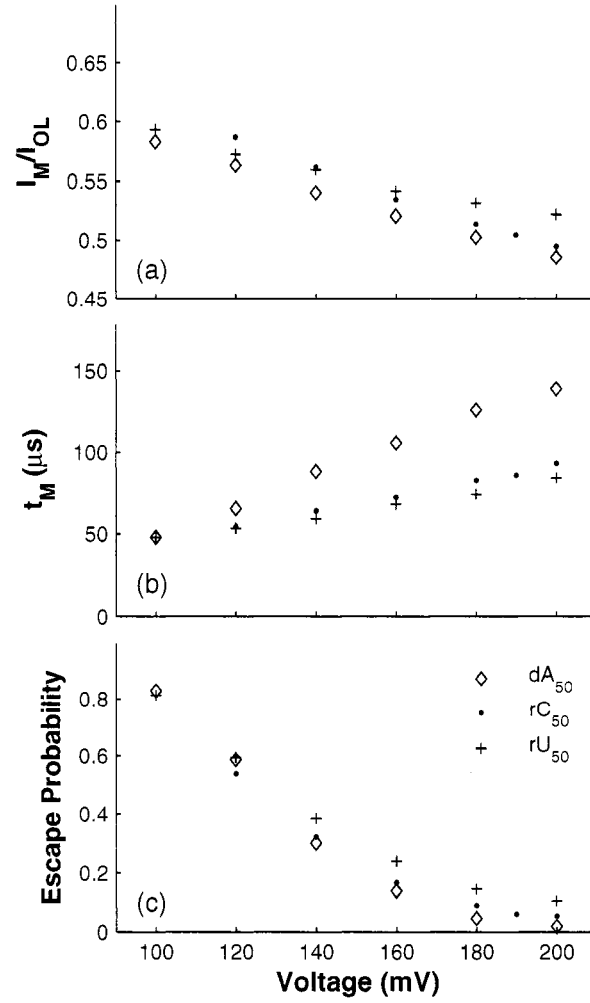


Figure 4.7: Voltage dependence of vestibule configuration. (a) Normalized M_1 current. (b) Mean duration of the primary group of M_1 signals. (c) Escape probability. This number reflects the likelihood that a polymer which has inserted into the vestibule will escape back into the electrically negative volume. It was calculated as the ratio of the number of M_0 signals to the sum of the number of M_0 and M_1 signals.

The general voltage-dependent trends shown in Fig. 4.7 provide further evidence supporting the hypothesis that M_0 and M_1 signals correspond to a polymer in the vestibule configuration. If a higher applied positive voltage on the *trans* side pulls more negatively charged bases from the *cis* side into the vestibule, then the volume exclusion mechanism of current obstruction accounts for the decrease in the normalized Mid State current. This simple mechanism can explain the general trend in the data, but a variety of additional factors may also contribute to the observed Mid State current [9, 85]. The trends in the durations and escape probabilities in Fig. 4.7 can be explained by a simple kinetic model wherein escape from the vestibule and threading of a polymer end from the vestibule into the pore neck are described as activated processes with associated rate constants β and α , respectively. If the vestibule configuration obeys such simple kinetics, its duration would be exponentially distributed with a time constant of $\tau_M = (\beta + \alpha)^{-1}$, and the escape probability would be given by $\beta/(\beta + \alpha)$. The trends in t_M and the escape probability shown in Fig. 4.7 could then result if (i) escape is favored at low voltages, while threading is favored at high voltages ($\beta > \alpha$ below ~ 140 mV, $\alpha > \beta$ above ~ 140 mV), (ii) the escape rate constant, β , decreases with increasing voltage, and (iii) escape has a stronger voltage dependence than threading (over the range 100-200 mV, $\partial\beta/\partial V > \partial\alpha/\partial V$).

4.3.6 Polymer orientation and the vestibule configuration

Before making a more detailed comparison between our data and this kinetic model of the vestibule state, we must consider the effects of polymer orientation. In chapter 5 we show that that translocation of an rC_{50} polymer in a $3' \rightarrow 5'$ orientation produces stronger obstruction of the ionic current than translocation in a $5' \rightarrow 3'$ orientation, and Mathe *et al.* found an analogous effect for poly-*dA* translocation [46]. This orientation-dependent Deep State conductance leads to the two well-resolved groups of signals in the rC_{50} and dA_{50} panels of Fig. 4.3. We utilized this effect to distinguish M_1 States that were followed by $3'$ -first translocation from those that were followed by $5'$ -first translocation. The resulting orientation-specific M_1 duration distributions are shown in Fig. 4.8. These distributions indicate that $5'$ -associated M_1 signals tend to be longer than $3'$ signals, and that this

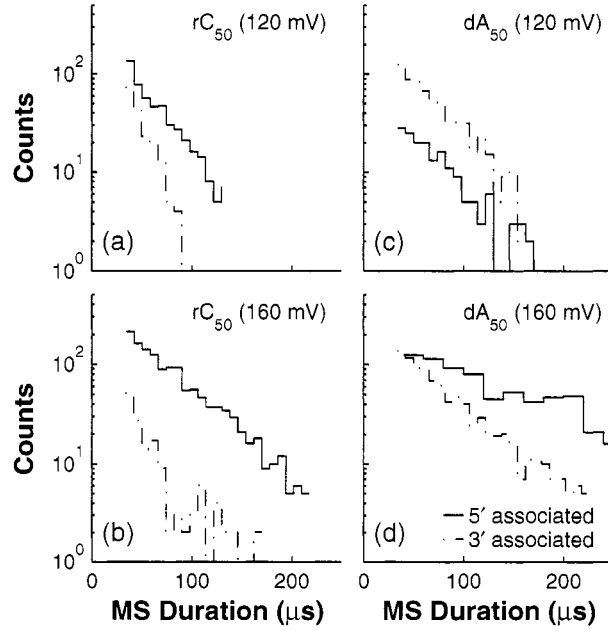


Figure 4.8: M1 State duration distributions grouped by polymer orientation. (a) and (b) show data for rC_{50} and (c) and (d) show data for dA_{50} . Data in top row were obtained at 120 mV while the data in the bottom row were obtained at 160 mV. The correlation between Deep State conductance and polymer orientation shown in Fig. 4.3 enabled us to group these M_1 signals by orientation.

difference is enhanced as the voltage increases. Application of the kinetic model we have proposed requires that the duration distributions of 3' and 5'-associated vestibule states be individually well-described by single exponentials. This requirement is met as single-exponential fits accurately modeled the dA_{50} data for both orientations and the rC_{50} data for 5'-associated Mid States. M_1 States followed by 3'-first translocation for rC_{50} were rare and short, making it difficult to assess the exponential character of these data. The existence of orientation-dependent differences in vestibule state durations for the other four polymers is suggested by the observation that, at voltages above ~ 140 mV, their overall M_1 duration distributions are better described using two exponentials rather than a single exponential.

We interpret the observation that vestibule configurations which are followed by 3' and 5'-first translocation demonstrate different lifetimes as indicating that there are two distinct

vestibule configurations, one for each orientation of the polymer. This could occur if only part of the polymer resides in the vestibule, with one end remaining in the *cis* compartment, or if the entire polymer resides in the vestibule but its conformational freedom is so restricted that it cannot reorient itself within the vestibule. While it remains unclear how much of the polymer resides in the vestibule, it nonetheless is appropriate to use separate sets of rate constants to model 3' and 5'-associated vestibule configuration kinetics.

4.3.7 Kinetic model of the vestibule configuration

We are able to compare our dA_{50} data to a kinetic model with the characteristics described above because its 3' and 5'-associated M_1 States can be distinguished and the orientation-dependant duration distributions are individually well described by single exponentials (Fig. 4.8). However, because we cannot distinguish between 3' and 5'-associated M_0 signals, we are unable to estimate orientation-specific 3' and 5'-associated escape probabilities and use them to derive estimates for the escape and threading rate constants directly from our data. We can estimate the overall escape probability, which is a weighted average of the orientation-specific escape probabilities. It is then possible to make a semi-quantitative analysis of the kinetics of the vestibule configuration by assuming a particular model for the voltage dependence of the rate constants and then assessing that model's ability to reproduce our voltage dependant lifetime and escape probability data (Fig. 4.9). We assume a very simple model where the 3' and 5'-associated escape rate constants, β_3 and β_5 , are identical and decrease exponentially with voltage (Eqn. 4.1), while the threading rate constants, γ_3 and γ_5 , are different but are both independent of voltage (Eqns. 4.2 and 4.3).

$$\beta_3 = \beta_5 = b \exp\left(-\frac{nze\delta V}{kT}\right) \quad (4.1)$$

$$\gamma_3 = g_3 \quad (4.2)$$

$$\gamma_5 = g_5 \quad (4.3)$$

In Eqn. 4.1, b is the escape rate constant with zero applied voltage, e is the elementary charge, z is the effective charge of a phosphate on the polynucleotide backbone after taking into account the effects of counter-ions, n is the number of phosphates resident in the

vestibule during the transition state for escape, δ is the fractional voltage drop in the vestibule, V is the applied voltage, k is the Boltzmann constant and T is the temperature. g_3 and g_5 are the voltage-independent rate constants for threading of a polymer end through the neck of the pore. This model is a simplified implementation of the Woodhull formalism [84], and a similar approach has previously been used to describe the kinetics of polypeptides in the α -HL pore [56]. To simplify our notation we define a characteristic escape voltage $V_E = [ezn\delta/kT]^{-1}$. This leads to the following expressions for the exponential time constants and the overall escape probability:

$$\tau_3 = \left[g_3 + b \exp\left(-\frac{V}{V_E}\right) \right]^{-1} \quad (4.4)$$

$$\tau_5 = \left[g_5 + b \exp\left(-\frac{V}{V_E}\right) \right]^{-1} \quad (4.5)$$

$$P_{escape} = \frac{1}{G_3 + G_5 + B} \left[G_3 \frac{b}{g_3} + G_5 \frac{b}{g_5} \right] \exp\left(-\frac{V}{V_E}\right) \quad (4.6)$$

G_3 and G_5 are the observed numbers of 3' and 5'-associated M_1 signals, and B is the total number of observed M_0 events. The parameter values $g_3 = 0.016\mu s^{-1}$, $g_5 = 0.006\mu s^{-1}$, $b = 18\mu s^{-1}$, and $V_E = 17\text{mV}$ give reasonable agreement between this model and our dA_{50} data (Fig. 4.9). These values were found by minimization of the total sum-squared deviation between the data in Fig. 9 and model in Eqns. 4.4–4.6. Weighting factors were adjusted such that the contributions to the total sum-squared deviation were approximately equal for the three curves.

Introducing voltage dependence into the threading rate constants as $\gamma = g \exp(V/V_T)$ required very weak voltage dependence ($V_T > 1000 \text{ mV}$) to give a reasonable description of our data. We found that a variety of other five-parameter models gave reasonable descriptions of our data, but none did so appreciably better than the four-parameter model of Eqns. 4.1–4.3. We also explored additional four-parameter models, one where the orientation dependence arose from differing voltage-independent factors in the escape rate constants ($\beta_3 = b_3 \exp(V/V_E)$, $\beta_5 = b_5 \exp(-V/V_E)$, and $\gamma_3 = \gamma_5 = g$), and another where the orientation dependence arose from differing 3' and 5'-associated escape voltages ($\beta_3 = b \exp(V/V_{E3})$,

$\beta_5 = b \exp(-V/V_{E5})$, and $\gamma_3 = \gamma_5 = g$). Neither of these four-parameter models gave acceptable descriptions of our data.

4.3.8 Interpretation of the vestibule configuration model

The ability of the model specified by Eqns. 4.1–4.3 to describe our data strongly supports the assertion that over the range ~ 100 – 200 mV, the escape process is much more sensitive to voltage than the threading process. The relatively weak voltage dependence of threading may be related to a small voltage gradient within the vestibule and the lack of negatively charged terminal phosphate groups at both the 3' and 5' ends of our synthetic polymers. The arrival of a polymer end at the neck may therefore be a diffusive process which is little biased by the applied voltage. Furthermore, insertion of the uncharged sugar/base portion of the terminal nucleotide far enough into the pore's inner constriction to allow the charged phosphate to experience the high electric field could also be a voltage-insensitive step in the threading process. Another aspect of the vestibule state suggested by our model is that the orientation dependence of the vestibule state lifetime (Fig. 4.8) can be primarily attributed to different rate constants for 3' versus 5' threading. Specifically, 3' threading is easier than 5' threading. This finding is consistent with the results of Mathe *et al.* [46] where experiments showed that poly(*dA*) segments passed more rapidly through the α -HL inner constriction when they moved in the 3' \rightarrow 5' direction as compared to the 5' \rightarrow 3' direction. In the same report, molecular dynamics simulations of a poly(*dA*) molecule threaded through the α -HL pore also found that 3' \rightarrow 5' passage was faster than 5' \rightarrow 3' passage. These simulations revealed a microscopic mechanism for the orientation asymmetry. When the polymer was confined in a tube with diameter < 15 Å, the DNA bases tilted toward the 5' end of the polymer. Because initial threading of a polymer and passage of a threaded polymer through the neck are very similar processes, we expect that the confinement-induced tilt is what makes 3' threading easier than 5' threading. Because the pore entrance that connects the vestibule to the *cis* compartment is ~ 25 Å in diameter, we do not expect an asymmetric tilting of the bases to be important to the escape process. Our finding that the voltage-independent factor in the escape rate constant (b in Eqn. 4.1) need

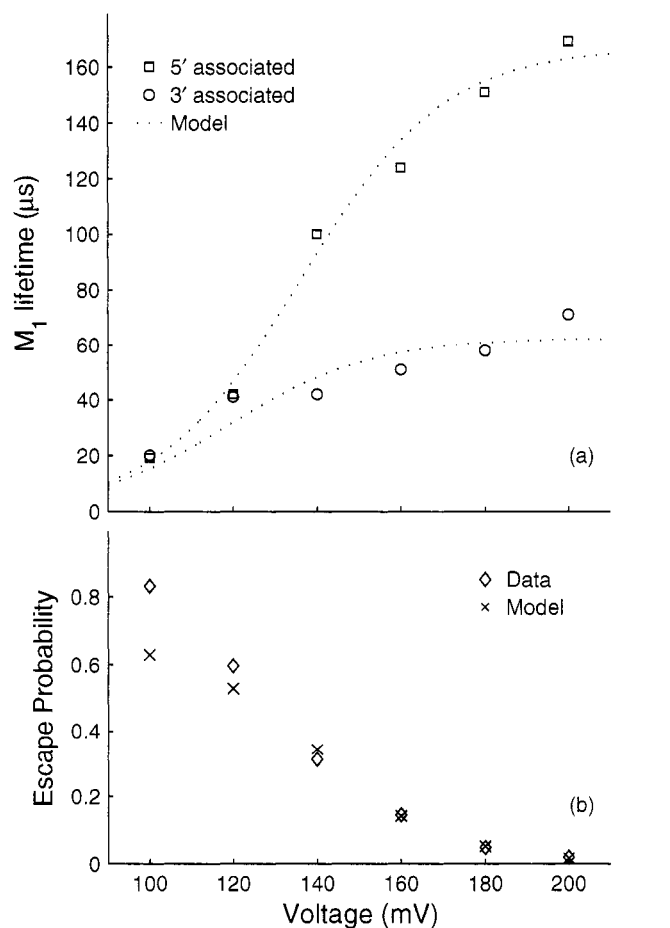


Figure 4.9: Voltage dependant dA_{50} data fitted by a kinetic model with voltage-dependant rate constants (Eqns. 4.1–4.3). M_1 lifetimes at each voltage were derived from exponential fits to the 3' and 5'-associated M_1 duration distributions. The model curves in the top panel were generated from Eqns. 4.4 and 4.5. The escape probability data in the bottom panel (\diamond) is the same as the dA_{50} escape probability data shown in Fig. 4.7, while the model values (\times) were derived from Eqn. 4.6. The model for the escape probability cannot be shown as a continuous curve because it requires experimental data (the observed numbers of 3' and 5'-associated M_1 signals). These model curves were generated with the parameters $g_3 = 0.016\mu s^{-1}$, $g_5 = 0.006\mu s^{-1}$, $b = 18\mu s^{-1}$, and $V_E = 17mV$.

not depend on orientation is consistent with this expectation. The orientation independence of escape also suggests that z , n , and δ in Eqn. 4.1, the microscopic factors that determine the voltage sensitivity of escape, do not depend strongly on polymer orientation. Finally, we note that the assumptions $z \sim 0.3$, $n \sim 20$ and $\delta \sim 0.2$ lead to the estimate $V_E \sim 20$ mV, indicating that our crude experimental value of 17 mV is physically plausible. The kinetic model we have presented is quite simple, but it gives a very reasonable description of our data and it suggests a number of interesting properties of the behavior of a polynucleotide molecule when it is confined in the α -HL vestibule.

4.3.9 Polymer capture and the vestibule configuration

The process of polymer capture by the pore has been investigated experimentally by measuring the dependencies of the capture rate on polymer concentration and applied voltage [26, 50, 62]. Exponential dependence of the capture rate on applied voltage was reported for voltages below ~ 120 mV [26]. For voltages greater than ~ 150 mV, an approximately linear [62] or a much weaker exponential dependence [50] was reported. The studies where the capture rate was measured for voltages greater than 120 mV both focused their analysis on the time intervals between successive translocations (t_{thread} in Fig. 4.10) as opposed to the interval between the end of one Mid State or Deep State and the beginning of the next Deep State (t_{enter} in Fig. 4.10). The work by Meller *et al.* employed an event detection threshold of $I_{block}/I_{OS} = 0.25$, and thus all of the M_0 events were excluded from this analysis. Nakane *et al.* employed an event detection threshold of $I_{block}/I_{OS} = 0.80$, a level that is sensitive to M_0 events. However, it is likely that they still missed some fraction of the shorter M_0 events due to inherent limitations in the response time of the current amplification electronics [61].

The existence of the vestibule configuration and the ability of a polymer to escape from it without translocation play an important role in the scaling change in the voltage dependence of the capture rate reported by Meller *et al.* and Nakane *et al.* Extending the model developed in the previous section, we assume one can account for the entry of polymers from the bulk into the vestibule with an effective unimolecular rate constant ζ . The probability density function of t_{thread} for such a model can be obtained by solving the

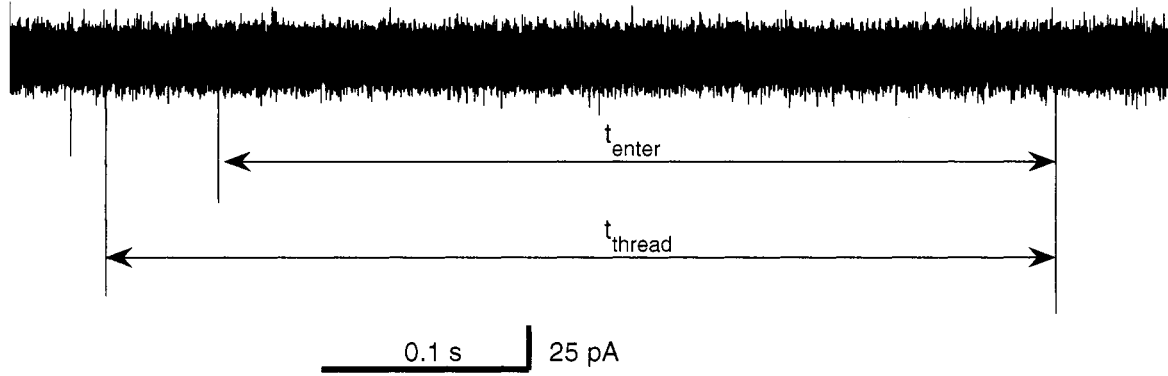


Figure 4.10: Continuous time series of the ionic current flowing through the pore demonstrating two translocation events and two escape events. There are two steps in *cis*-side capture by α -HL – entrance of a polymer into the vestibule and subsequent threading of the polymer from the vestibule into the inner constriction. We define the interval between the end of one event (either by translocation or escape) and the beginning of the next Mid State as “ t_{enter} ”, and we define the interval between the end of one Deep State and the beginning of the next Deep State as “ t_{thread} ”.

following system of equations [68]:

$$\frac{dP_M}{dt} = \zeta P_O(t) - (\beta + \gamma)P_M(t) \quad (4.7)$$

$$\frac{dP_O}{dt} = \beta P_M(t) - \zeta P_O(t) \quad (4.8)$$

$P_M(t)$ is the probability of finding the system in the vestibule configuration at time t , given that the pore was empty at time zero, and $P_O(t)$ is the probability of finding the pore empty at time t , given that it was empty at time zero. The probability density function of t_{thread} is given by $pdf(t_{thead}) = \gamma P_M(t_{thead})$. Elimination of P_O from Eqns. 4.7 and 4.8 leads to a second order differential equation for P_M . The solution to this equation, after applying appropriate initial and normalization conditions, is given by

$$P_M(t) = \frac{\zeta}{\lambda_+ - \lambda_-} \exp(-\lambda_- t) (1 - \exp(-\lambda_+ t)) \quad (4.9)$$

where

$$\lambda_{\pm} = \frac{1}{2} \{ (\zeta + \beta + \gamma) \pm \sqrt{(\zeta + \beta + \gamma)^2 - 4\zeta\gamma} \} \quad (4.10)$$

Under typical experimental conditions, the time between the end of one event and the beginning of the next is much longer than the characteristic duration of the events themselves (Fig. 4.10). In the model, this corresponds to $\zeta \ll \beta$ and $\zeta \ll \gamma$. Under these conditions, Eqn. 4.9 can be approximated as:

$$P_M(t) \simeq \frac{\zeta}{\beta + \gamma} \exp(-R_{slow}t)(1 - \exp(-R_{fast}t)) \quad (4.11)$$

where

$$R_{fast} \equiv \beta + \gamma \quad (4.12)$$

$$R_{slow} \equiv \frac{\zeta}{2} \left(1 - \frac{\beta - \gamma}{\beta + \gamma} \right) \quad (4.13)$$

The conditions on ζ , β , and γ imply that $R_{fast} \gg R_{slow}$. For t more than a few times greater than R_{fast}^{-1} , Eqn. 4.11 is well approximated by a single exponential with rate R_{slow} . Our data indicates that $\sim 100\mu\text{s}$ would be a typical value for R_{fast}^{-1} , suggesting that the deviation from single-exponential behavior in Eqn. 4.11 would show up in the experimental distribution for t_{thread} as a decrease in the number of intervals shorter than a few hundred μs . Such an effect would not be readily apparent in an analysis focused on time scales greater than 10 ms. Thus Eqn. 4.11 is consistent with the observation of Meller *et al.* that the distribution of t_{thread} was well fitted by a single exponential [50]. The capture rate measured by Meller *et al.* is then associated with the above rate R_{slow} . At lower voltages we find that escape is favored over translocation (Fig. 4.7). This corresponds to $\beta > \gamma$ in our model, and Eqn. 4.13 reduces to $R_{slow} \simeq \zeta(\gamma/\beta)$. At higher voltages we find that translocation is favored over escape. This corresponds to $\gamma > \beta$, and in this situation Eqn. 4.13 reduces to $R_{slow} \simeq \zeta(1 - \beta/\gamma)$. The transition between these two regimes occurs when $\beta \simeq \gamma$, which our data suggests is somewhere between 120-140 mV (Fig. 4.7). This is the same range over which the scaling change in the voltage dependence of the capture rate was observed by Meller *et al.* The capture rate measured by Nakane *et al.* cannot be related directly to R_{slow} . However, they did not account for the fraction of M_0 events that were

too short to trigger their threshold, and thus it is plausible that the scaling change they observed at ~ 150 mV is also attributable in part to the scaling change in R_{slow} .

Intuitively, this scaling change can be understood by reasoning that the capture rate is dependent upon both the rate of polymer entry into the vestibule and the probability of subsequent translocation. At lower voltage many of the polymers escape from the vestibule before they translocate, causing the observed capture rate to be much smaller than the entry rate. At higher voltage escape is unlikely and the capture rate becomes approximately equal to the entry rate (minus a small correction for the duration of the Mid State itself). A qualitatively similar explanation for the scaling change in the voltage dependence of the capture rate has been shown to emerge from detailed theoretical models of the capture process [4, 62]. These models used a simplified pore geometry that did not include a vestibule. Furthermore, they directly coupled polymer conformations in the bulk to polymer conformations within a narrow transmembrane channel, as opposed to linking the two states indirectly via an explicit intermediate state. In terms of our kinetic picture, these models would be related to the rate constant for entry into the vestibule, ζ . Further development of a theoretical model of a polynucleotide segment confined within the vestibule [58] could yield predictions of the rate constants for escape and translocation, β and γ . These predictions could be directly compared with the experimentally determined rate constants, providing a strong link between experiment and theory.

4.4 Summary and Conclusions

We conducted an analysis of current blockades produced by single-stranded polynucleotide molecules in the α -HL nanopore that took into account their sub-state structure. We modeled the structure of each blockade as a sequence of Mid (current at $\sim 50\%$ of unobstructed Open State) and Deep (current less than $\sim 30\%$ of Open State) conductance states. There appear to be two distinct configurations of the polymer in the pore that cause a Mid State. Our data are consistent with the interpretation that the most common Mid State occurs at the beginning of the polymer/pore interaction, when part of a polymer enters the α -HL vestibule from the *cis* compartment, but does not thread through the pore's inner constriction. The ionic current remains in the Mid State until the polymer either (i) escapes back

into the *cis* compartment, causing the current to return to the Open State, or (ii) threads through the inner constriction of the pore and reduces the current to the Deep State. For some polymers, up to $\sim 80\%$ of the observed events demonstrated resolvable signals indicative of a polymer in the vestibule configuration. Occasionally we observed Mid States that followed Deep States and demonstrated a different amount of residual current than the vestibule-associated Mid States. We attribute these Mid States to a configuration where the trailing end of the polymer occupies part of the β -barrel region of the pore, but is no longer threaded through the inner constriction. These β -barrel signals were much less frequent than the vestibule signals, occurring in 3% to 8% of the current blockade events for three of the polymers we investigated, and almost never occurring for the other three polymers. We observed clear statistical differences in the Mid State characteristics of different polymers, indicating that Mid State analysis could enhance the capability of α -HL as a single-molecule analytical device. Simulation studies of these polymer-specific differences could provide much needed insight into their molecular origins, and accurate prediction of these differences would provide very strong validation of the simulation methodology.

We focused particular attention on the signals indicative of the vestibule configuration because they are a very prominent but little studied component of the current blockades produced by polynucleotide molecules in the α -HL pore. To gain insight into the physics of the vestibule configuration, we analyzed vestibule configuration data at a variety of voltages between 100 and 200 mV. We observed a clear decrease in the normalized current with increasing voltage. This may be the result of crowding as more bases are pulled into the vestibule at higher voltages. We also found that the characteristic duration of the vestibule state increased with increasing voltage while the probability of polymer escape from the vestibule back into the electrically negative volume decreased. A simple kinetic picture where the rate constant for escape decreases strongly with voltage while the rate constant for threading has much weaker voltage dependence qualitatively explains these trends. This simple picture was slightly complicated by the observation that the duration of the vestibule configuration depended on which end of the polymer, 3' or 5', subsequently threaded into the inner constriction. We found that a kinetic model with the following characteristics gave a very nice description of our orientation and voltage-dependant dA_{50} data: (i) the escape

rate constant decreases exponentially with voltage but does not depend on orientation and, (ii) the threading rate constants are different for 3' versus 5' insertion, but do not depend on the applied voltage. From this semi-quantitative analysis we conclude that threading has much weaker voltage dependence than escape, threading is sensitive to polymer orientation, and escape is insensitive to polymer orientation. The intricate molecular details revealed by this analysis underscore the utility of α -HL as a model system to study physical and chemical mechanisms that are important constituents of many biological processes.

Chapter 5

DETERMINATION OF RNA ORIENTATION DURING
TRANSLOCATION***Abstract***

We conducted a comparative analysis of the ionic current blockades produced by the translocation of rA_{50} , rC_{50} , $rA_{25}C_{50}$, and $rC_{50}A_{25}$ RNA molecules through the α -HL nanopore. Translocation of rA_{50} reduced the ionic to $\sim 15\%$ of the Open State current level, while translocation of rC_{50} reduced the current to either $\sim 5\%$ or $\sim 10\%$ of the Open State. Translocation of $rA_{25}C_{50}$ and $rC_{50}A_{25}$ molecules often produced blockades showing a distinct bi-level signal within the Deep State. The current in the deeper portion of the bi-level signal was generally at $\sim 5\%$ or $\sim 10\%$ of the Open State, while the shallower portion was at $\sim 15\%$, suggesting that the bi-level signal corresponds to a transition between the poly- rC segment and the poly- rA segment threading through the pore. From the temporal ordering of the bi-level signals we inferred whether individual $rA_{25}C_{50}$ and $rC_{50}A_{25}$ molecules passed through the pore in a $3' \rightarrow 5'$ or $5' \rightarrow 3'$ orientation. Correlation between the levels of current obstruction and the inferred $rA_{25}C_{50}$ or $rC_{50}A_{25}$ orientation indicates that $3' \rightarrow 5'$ translocation of a poly- rC segment causes a significantly deeper current obstruction than $5' \rightarrow 3'$ translocation. This analysis also suggests that the $3'$ -ends of rC_{50} and $rA_{25}C_{50}$ RNA molecules are more likely to initiate translocation than the $5'$ -ends. Orientation dependent differences in the Mid States that immediately preceded many of the Deep States suggest that the Mid State also contains information about the orientation of an RNA molecule. These findings emphasize that the orientation of a polynucleotide molecule is an important factor in translocation and highlight how detailed analysis of the structural features within ionic current blockades can give new insights into the translocation process.

5.1 Introduction

The directionality of the sugar-phosphate backbone plays an important role in determining the biochemical properties of polynucleotide molecules. For example, during DNA replication, DNA-polymerase slides along both the leading and lagging template strands in a $3' \rightarrow 5'$ direction as it adds new nucleotides to the $3'$ -ends of the newly synthesized strands [3]. It is reasonable to expect that this directionality affects the electrophoretic translocation of ssDNA or RNA through the pore, and this expectation has motivated discussion of polynucleotide orientation in various theoretical treatments of translocation dynamics [44, 59].

When plotted in the two-dimensional space of translocation duration versus average translocation current, ssDNA and RNA translocation signals often group into two distinct, well defined groups (Fig. 4.3). It was hypothesized that the existence of these two groups was attributable to the directionality of the sugar-phosphate backbone and that they represent translocation in either the $3' \rightarrow 5'$ or $5' \rightarrow 3'$ orientation [2, 34, 59]. Recent experimental work supported this hypothesis for ssDNA [46, 82] but correlation between grouping and orientation had not yet been investigated for poly-*rC* RNA.

It was previously demonstrated that RNA diblock copolymers comprised of a poly-*rA* and a poly-*rC* segment produce current blockades that show a pronounced bi-level signal within the Deep State, and that this signal most likely reflected the diblock composition of the molecule [2]. We built upon these observations and utilized the fact that the time-ordering of the bi-level signals is indicative of polymer orientation to make a detailed experimental test of the orientation hypothesis for poly-*rC* RNA.

5.2 Materials and methods

An overview of our apparatus and procedures are given in Chapter 3, and the detailed Materials and Methods described in Chapter 4 also apply to these experiments. For all data presented in this report, 120 mV (*trans*-side positive) was applied across the bilayer. The average of the mean Open State current observed in the four data sets presented in this chapter is 128 pA. RNA concentrations of $\sim 1 \mu\text{M}$ in the $\sim 160 \mu\text{l}$ *cis* compartment during

data collection resulted in event rates of ~ 0.5 -2 Hz.

The four RNA samples used in these experiments, rA_{50} , rC_{50} , $rA_{25}C_{50}$ and $rC_{50}A_{25}$, were synthesized and PAGE purified by Dharmacon RNA Technologies (Lafayette, CO). RNA samples were re-suspended to a concentration of 20 μ M in the experimental buffer and stored at -20°C until immediately before use. After completion of a ten month experimental period, sample aliquots of each of the four RNA constructs were 5' end-labeled with Polynucleotide Nucleotide Kinase and $\gamma^{32}\text{P}$ ATP and run on a 10% denaturing polyacrylamide gel. The dried gel was exposed and scanned on a PhosphorImager (Molecular Dynamics) and gel images were analyzed with ImageQuant (Molecular Dynamics). Image analysis indicated that 60% of the rA_{50} sample, 81% of the rC_{50} sample, 74% of the $rA_{25}rC_{50}$ sample and 67% of the $rC_{50}A_{25}$ sample were full length. The rA_{50} sample also demonstrated 8% N-1 and 6% N-2, which we interpret as incomplete sample purification after synthesis. Shorter RNA molecules in each sample are most likely the result of degradation during storage or handling.

We present data from one experiment with each of the four RNA samples. These experiments were conducted with 120 mV applied across the bilayer. We repeated these experiments over a period of ten months with 120, 140 or 160 mV applied across the bilayer and the trends described in this report were consistently observed in all experiments. The rA_{50} data set detailed here is representative of 16 data sets from 7 different pores, rC_{50} is representative of 12 data sets from 5 pores, $rA_{25}C_{50}$ is representative of 6 data sets from 3 pores and $rC_{50}A_{25}$ is representative of 12 data sets from 5 pores.

5.3 Results

5.3.1 $rA_{25}C_{50}$ and $rC_{50}A_{25}$ signals reflect diblock composition

Typical traces from translocation of $rA_{25}C_{50}$ and $rC_{50}A_{25}$ are shown in Fig. 5.1. As with the homopolymer events described in Chapter 4, the diblock events had two primary conductance states; a Mid State and a Deep State. In contrast to the homopolymer events, the diblock events often demonstrated clear bi-level modulations within the Deep State ("step signals"). Fig. 5.1 also shows an rA_{50} and rC_{50} event that have been superposed to illus-

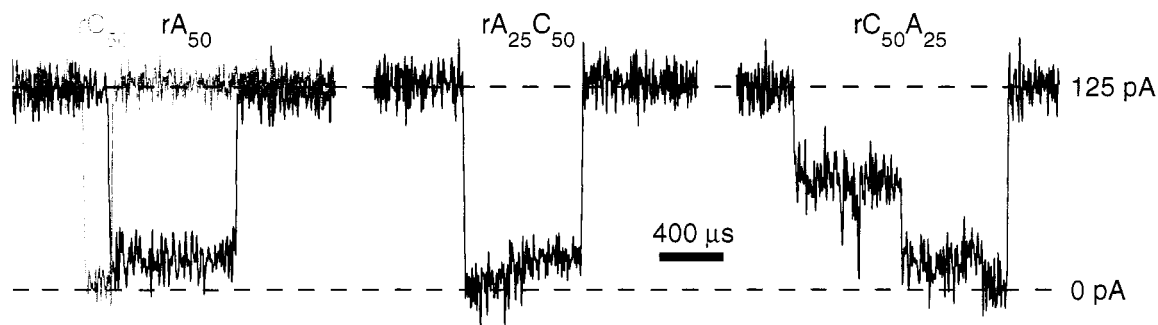


Figure 5.1: Examples of RNA homopolymer and copolymer blockades.

trate the plausibility of associating the step signals in the copolymer Deep States with the transition between a poly-*rA* segment and a poly-*rC* segment in the inner constriction of the α -HL pore.

The statistical behavior of the Deep States from translocation of these four constructs at +120 mV is shown in Fig. 5.2. Each point gives the duration and average current of one Deep State. The current values represent the average current over the entire Deep State, and thus step signals are represented by a current that is intermediate between the two constituent levels. The location of each group was estimated as the centroid of the region bounded by a contour denoting 50% of the highest density of Deep States. In Fig. 5.2 most of the *rA*₅₀ events cluster into one main group that is located at a duration of $\sim 600 \mu s$ and $I_{DS}/I_{OS} \simeq 0.13$. Most of the *rC*₅₀ events cluster into two well resolved groups centered at $\sim 150 \mu s$, $I_{DS}/I_{OS} \simeq 0.025$ and $\sim 120 \mu s$, $I_{DS}/I_{OS} \simeq 0.08$. From these group locations we make the rough approximations that adenine segments take $\sim 12 \mu s/nt$ to move through the pore and obstruct the current at $I_{DS}/I_{OS} \simeq 0.13$, while cytosine segments take $\sim 3 \mu s/nt$ and obstruct the current at $I_{DS}/I_{OS} \simeq 0.05$. These approximations lead to an estimate of $\sim 450 \mu s$ and $I_{DS}/I_{OS} \simeq 0.095$ for the average duration and current obstruction levels expected for *rA*₂₅*rC*₅₀ and *rC*₅₀*A*₂₅ translocation. In Fig 5.2, the main *rA*₂₅*rC*₅₀ group is centered at $\sim 490 \mu s$, $I_{DS}/I_{OS} \simeq 0.08$ and the main *rC*₅₀*A*₂₅ is centered at $\sim 330 \mu s$, $I_{DS}/I_{OS} \simeq 0.10$. The durations and current levels exhibited by events in the main *rA*₂₅*rC*₅₀ and *rC*₅₀*A*₂₅ groups are thus reasonably intermediate between the *rA*₅₀ and *rC*₅₀

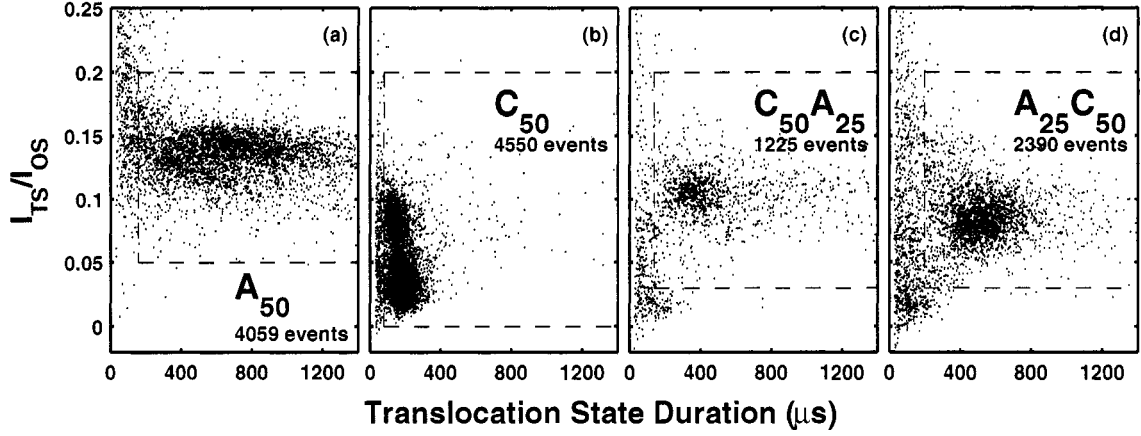


Figure 5.2: Distribution of Deep State durations and currents for rA_{50} , rC_{50} , $rA_{25}C_{50}$ and $rC_{50}A_{25}$ RNA. The coordinates of the point give the duration and average current of the Deep State of an event. Deep State current is normalized to the Open State current. Deep States falling outside of the dotted rectangles were excluded from our analysis. The number of events given in each panel corresponds to the number of points falling within the dashed rectangles. These four data sets were obtained under identical experimental conditions with 120 mV (*trans* side positive) applied across the bilayer.

homopolymer translocation characteristics.

5.3.2 Ambiguous signals

While the majority of the Deep States we observed were in or near the main groups, we did observe signals that were well separated from the centers of the main groups. In Fig. 5.2a there are a number of rA_{50} Deep States with very short durations, and many of these shorter signals have greater conductance than the signals near the center of the main rA_{50} group. As discussed in Chapter 4, these short Deep States could be the result of a polymer threading into the pore's inner constriction and then retracting back into the vestibule. They could also be the result of translocation of short RNA fragments produced by degradation of our sample [82], or very rapid translocation of full length molecules. Figs. 5.2c and 5.2d also demonstrate secondary populations of Deep States that are separated from the main groups by regions of low event density. Signals in these secondary groups have shorter durations and many of them tend to cause a greater obstruction of the ionic current than signals in

the main groups. Similar to the short duration rA_{50} events, events in the secondary groups in Figs. 5.2c and 5.2d could be interpreted as either incomplete translocation, translocation of short fragments, or complete, rapid translocation of full length molecules. Since PAGE analysis of our RNA samples suggested that $\sim 30\%$ of our rA_{50} , $rA_{25}rC_{50}$, and $rC_{50}A_{25}$ molecules were shorter than full length, we believe translocation of short RNA fragments to be the most likely explanation for the presence of these secondary populations of Deep States. However, we cannot rule out the other two possibilities. Lacking an unambiguous interpretation of these secondary populations of Deep States, we chose to focus our analysis on signals falling near the centers of the main groups. The dotted rectangles in Fig. 5.2 enclose the Deep States included in the results presented in this chapter. The trends we observe and our overall conclusions are insensitive to the values assigned to these current and duration cutoffs, but we feel that the exclusion of ambiguous signals allows for a clearer characterization of the majority of the Deep States.

5.3.3 Analysis and classification of events

Classification scheme

We classified each current blockade as one of six event “types”. Schematic diagrams of these six types are shown in Fig. 5.3. The six event types are the combination of three signal classes characterizing the Deep State and the presence or absence of a Mid State. The Deep State classes are NoStep, indicating that the Deep State does not exhibit a significant step, HiLo, indicating a “high to low” current step within the Deep State, and LoHi, indicating a “low to high” current step. Events demonstrating HiLo or LoHi signals will be collectively referred to as “step” events. The Mid State signal classes, Mid and NoMid, indicate the presence or absence of a resolvable Mid State preceding the Deep State of an event. In our nomenclature, the name of each event type is formed by concatenating the Mid State and Deep State class names. For example, a MidNoStep event has a Mid State but no step signal within the Deep State, and a NoMidLoHi does not have a Mid State, but does show a “low-to-high” step within the Deep State.

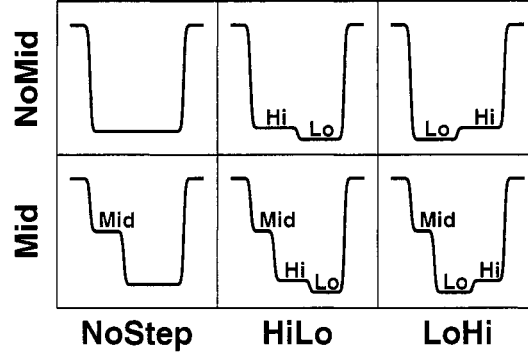


Figure 5.3: Translocation event classification scheme. An algorithm was used to classify each event as one of these six “types”. The six types arise from the presence or absence of a Mid State combined with three possible Deep State signatures. For example, a MidLoHi event begins with a resolvable Mid State, transitions to a very low $\sim 0\text{-}10\%$ “Lo” conductance level and finally to a moderately low $\sim 10\text{-}20\%$ “Hi” conductance level before returning to the initial Open State.

Algorithm for event parameterization and classification

Because the Hi and Lo Deep State conductance levels were not resolved from one another above the noise, we could not use a simple threshold algorithm to subdivide and parameterize the step events. Instead, we fitted each event by six phenomenological functional forms that represent the six event types described in the previous sub-section. These “type-functions” are sums of logistic functions. Each transition between conductance states corresponds to one term in the sum. The type-function used to represent MidHiLo and MidLoHi events is the following:

$$\begin{aligned}
 I(t) = I_0 &+ \frac{\Delta I_1}{1 + \exp\left[-\frac{(t-t_1)}{\tau_1}\right]} + \frac{\Delta I_2}{1 + \exp\left[-\frac{(t-t_2)}{\tau_2}\right]} \\
 &+ \frac{\Delta I_3}{1 + \exp\left[-\frac{(t-t_3)}{\tau_3}\right]} - \frac{\Delta I_1 + \Delta I_2 \Delta I_3}{1 + \exp\left[-\frac{(t-t_4)}{\tau_4}\right]}
 \end{aligned} \tag{5.1}$$

I_0 was fixed at the Open State current level. ΔI_i is the difference in current between the (i+1)th and ith levels, t_i is the time at which the ith transition occurs, and τ_i governs the steepness of the ith transition. The other event types were represented by analogous sums of logistic functions, and distinctions between event types were maintained by constraining the

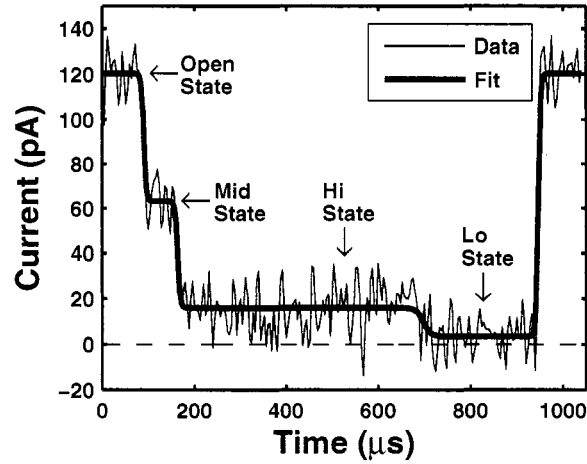


Figure 5.4: Example fit to a $rC_{50}A_{25}$ translocation event. The dark line shows the best fit to the data. The functional form of this fit is given in Eqn. 5.1 and the event is classified as MidHiLo. Comparison of this fit to the MidNoStep fit with the F-test yielded a p-value of $\sim 10^{-14}$.

bounds of the type-function fit parameters. The I_i and t_i parameters of each type-function were adjusted to optimize the fit to the event data by the Trust-region algorithm available in the Matlab Curve Fitting Toolbox. All τ_i parameters describing Open \leftrightarrow Mid \leftrightarrow Deep State transitions were fixed to give a 10%-90% rise-time of $7 \mu s$ in order to mimic the response characteristics of our recording electronics. We chose to fix the τ_i parameter describing the Lo \leftrightarrow Hi Deep State transition to give a 10%-90% rise time of $20 \mu s$ instead of $7 \mu s$ in order to better fit the actual Lo \leftrightarrow Hi transitions observed in the data. Our results were relatively insensitive to the specific values at which the τ parameters were fixed. Data and the fit for a MidHiLo event are shown in Fig. 5.4.

We then used the F-test [55] to determine the type-function that provided the best fit to a given event. We first checked for the presence of a Mid State, and we then checked for a step signal within the Deep State. To check for the presence of a Mid State, we compared the MidNoStep fit to the NoMidNoStep fit. If the p -value was less than 0.05 the event was classified as Mid, otherwise it was classified as NoMid. To check for the presence of a step signal in the Deep State, we compared the HiLo and LoHi fits to the NoStep fit. The

event was classified as HiLo or LoHi if either comparison yielded a p -value less than 0.05. If neither comparison resulted in a p -value less than 0.05, then the event was classified as NoStep. The p -value obtained in comparing the MidHiLo fit to the MidNoStep fit for the event shown in Fig. 5.4 is $\sim 10^{-14}$, reflecting the very strong HiLo step signal in the Deep State. The parameters of the best fitting type-function were used to estimate the Mid State, Deep State, Hi, and Lo current levels and durations.

Test of data analysis algorithm

To test our data analysis we created numerical simulations of translocation event signals and applied our analysis algorithm to these simulated signals. A range of idealized “noiseless” signals were generated by specifying parameters for the above described type-functions. To generate noise in our simulated signals we randomly selected segments of inter-event Open State time series data and subtracted off the mean value. We then added the resulting time series data as noise on the simulated noiseless signals. The rms of the noise added to simulated signals was 9.5 pA and the mean Open State current level for the simulated signals was 130 pA, reflecting the experimental conditions we observed with 120 mV applied across the bilayer. These simulations revealed the ranges of Deep State step sizes and Deep State sub-level durations in which our algorithm is effective and the ranges in which it is limited. We found that simulated signals were correctly classified more than 90% of the time if the following criteria were met: The step size, $\Delta I/I_{OS}$, had to be larger than 0.08 and the durations of both the Hi and Lo sub-levels had to be longer than 60 μ s. As Hi or Lo state durations were made shorter or step sizes smaller, it became increasingly likely that steps that were present in the noiseless signals were not resolved in the signals with noise. Our analysis algorithm classified such events as NoStep, and thus NoStep events can best be interpreted as events that do not meet the above described criteria. In a later section we elaborate on the effects of these known detection inefficiencies in the context of experimentally observed translocation signals.

5.3.4 Deep State step signal depends strongly on RNA composition

Figure 5.5 shows the distribution of event types for the four RNA samples we investigated. Our classification algorithm designated 88% of the rC_{50} events as NoStep. The rC_{50} distribution shows that the presence of strong, frequent steps in the Deep State is not a general phenomenon common to translocation of all RNA sequences. rA_{50} translocation produced 23% HiLo events and 35% LoHi events, indicating that rA_{50} does have a tendency to produce step signals. $rA_{25}C_{50}$ translocation produced a very high percentage of LoHi events. This result is consistent with a simple model wherein the different current blockade levels observed in rA_{50} and rC_{50} translocation are reflected as step signals in copolymer Deep States. $rC_{50}A_{25}$ translocation produced a high percentage of step signals, but it also produced a significant number of NoStep events. This result is not consistent with the simple model of copolymer translocation signals. In subsequent sections we compare the detailed characteristics of the step signals observed in rA_{50} , $rC_{50}A_{25}$ and $rA_{25}C_{50}$ translocation. This comparison leads to a clear interpretation of the presence of NoStep signals in copolymer translocation and shows that there are significant differences between rA_{50} step signals and copolymer step signals.

5.3.5 Current levels and RNA orientation during translocation

The distributions of the ionic current levels observed in rA_{50} and rC_{50} NoStep events are shown in Figs. 5.6a and 5.6b. The rA_{50} current distribution has one peak centered at $I/I_{0S} \simeq 0.14$, while the rC_{50} current distribution has two peaks centered at $I/I_{0S} \simeq 0.08$ and $I/I_{0S} \simeq 0.03$. HiLo and LoHi events are those in which we detect a step signal within the Deep State. The distributions of the current levels for the “Lo” and “Hi” levels observed in both HiLo and LoHi events are shown in Figs. 5.6c–5.6f. The current of the Lo levels in $rA_{25}C_{50}$ LoHi events (Fig. 5.6c) are very similar to the Deep State currents of the lower rC_{50} group (Fig. 5.6b) and the currents of the Hi levels in these events are very similar to the currents of rA_{50} (Fig. 5.6a). These similarities suggest that $rA_{25}C_{50}$ LoHi step signals represent poly(rC) translocation followed by poly(rA) translocation and that these events result from $3' \rightarrow 5'$ passage of polymers through the pore. The Hi level current in $rC_{50}A_{25}$

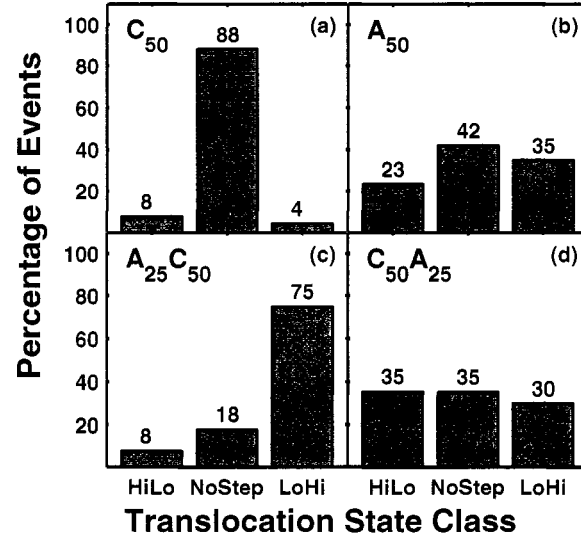


Figure 5.5: Event classification results. The bars in each panel correspond to the frequency of occurrence of the three Deep State classes: HiLo, NoStep, and LoHi. (a) shows that the large majority of rC_{50} translocation events do not exhibit strong step signals within the Deep State. (c) shows that a significant majority of $rA_{25}C_{50}$ events exhibit a strong LoHi step signal. We observe moderate differences in the event type distributions of rA_{50} and $rC_{50}A_{25}$.

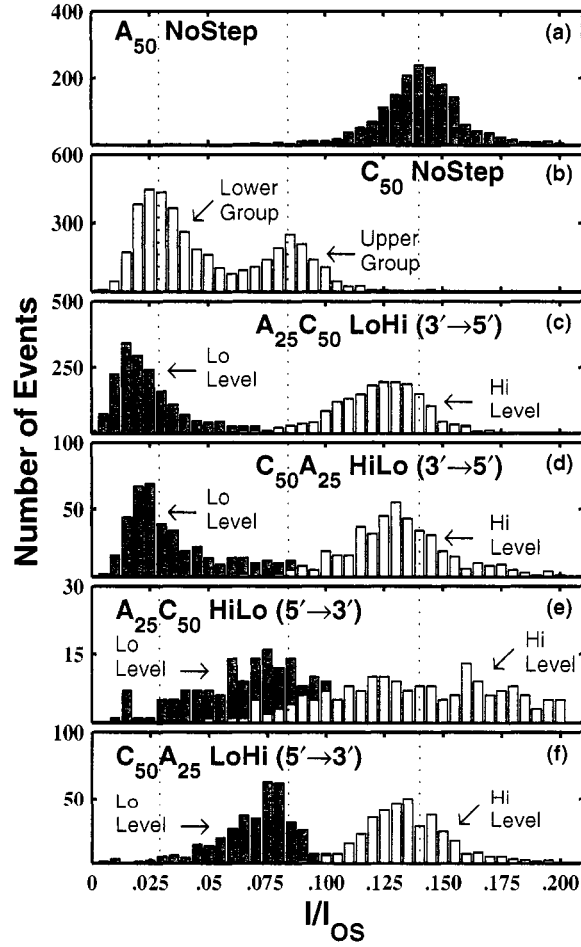


Figure 5.6: Distributions of Deep State current levels. Panels (a) and (b) show the distribution of Deep State current levels for all NoStep events observed in the rA_{50} and rC_{50} data, respectively. The two peaks in the rC_{50} distribution correspond to the two rC_{50} groups shown in Fig. 5.2. The dashed vertical lines correspond to mean current levels determined by Gaussian fits to the rA_{50} distribution and to the lower and upper rC_{50} distributions. Panels (c), (d), (e) and (f) show the distributions of the Hi and Lo Deep State current sub-levels observed in the indicated event classes. From the time ordering of the step signal we inferred the order in which the segments moved through the pore. Thus, we interpret HiLo events as $rA_{25} \rightarrow rC_{50}$ and LoHi events as $rC_{50} \rightarrow rA_{25}$. Comparing this information with the RNA sequence gives the orientation, $3' \rightarrow 5'$ or $5' \rightarrow 3'$, of the molecule during translocation. The inferred translocation orientations for each group of copolymer events are indicated in the corresponding panel.

HiLo events (Fig. 5.6d) is very similar to rA_{50} current, while the Lo level is similar to the current in the lower rC_{50} group. We interpret these similarities as indicative of poly(rA) translocation followed by poly(rC) translocation, suggesting that $rC_{50}A_{25}$ HiLo events also represent $3' \rightarrow 5'$ translocation. $rA_{25}C_{50}$ LoHi and $rC_{50}A_{25}$ HiLo events both appear to represent translocation in the $3' \rightarrow 5'$ orientation and they both demonstrate Lo current levels that are similar to the Deep State current levels of the lower of the two rC_{50} groups. From this we deduce that $3' \rightarrow 5'$ translocation of a poly- rC segment results in current levels of $I/I_{OS} \simeq 0.03$. The Hi current levels for $rA_{25}C_{50}$ HiLo (Fig. 5.6e) and $rC_{50}A_{25}$ LoHi (Fig. 5.6f) events are similar to the rA_{50} Deep State currents, whereas the Lo level for these events are similar to the current of the upper rC_{50} group. This indicates that $rA_{25}C_{50}$ HiLo and $rC_{50}A_{25}$ LoHi both represent $5' \rightarrow 3'$ translocation, and that $5' \rightarrow 3'$ poly- rC translocation results in a current of $I/I_{OS} \simeq 0.08$. Furthermore, signals suggestive of $5' \rightarrow 3'$ copolymer translocation rarely demonstrated a Lo level current of $I/I_{OS} \simeq 0.03$, while signals suggestive of $3' \rightarrow 5'$ copolymer translocation rarely demonstrated a Lo level current of $I/I_{OS} \simeq 0.08$. These trends indicate a strong correlation between poly- rC orientation and the degree of obstruction of the ionic current. A schematic diagram illustrating the orientation dependent step signals is shown in Fig. 5.7. We further interpret the strong similarities between the step signal Lo levels and the two rC_{50} current levels as evidence that the lower group of rC_{50} events in Fig. 5.2b is representative of $3' \rightarrow 5'$ translocation of rC_{50} molecules while the upper group is representative of $5' \rightarrow 3'$ translocation.

Our interpretation of the data in Fig. 5.6 implies that $3' \rightarrow 5'$ translocation of $rA_{25}C_{50}$ and $rC_{50}A_{25}$ produces a strong $\Delta I/I_{OS} \sim 0.11$ step in the Deep State while $5' \rightarrow 3'$ translocation produces a smaller $\Delta I/I_{OS} \sim 0.06$ step. As described earlier, we determined that our data analysis algorithm efficiently detects Deep State step signals that are larger than $\Delta I/I_{OS} \sim 0.08$, and so it is unlikely that many $3' \rightarrow 5'$ copolymer translocation events are classified as NoStep. However, our detection efficiency and accuracy declines with decreasing Deep State step signal size below $\Delta I/I_{OS} \sim 0.08$. Thus it is reasonable to expect that many $5' \rightarrow 3'$ copolymer translocation events are classified as NoStep events.

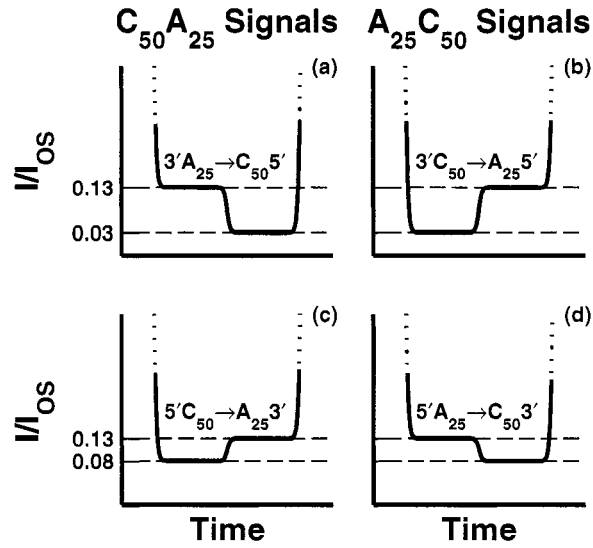


Figure 5.7: Diagram of the relationships between RNA orientation during translocation and observed Deep State step signals for $rA_{25}C_{50}$ and $rC_{50}A_{25}$ copolymers. These relationships are inferred from the time ordering and current obstruction levels of the Deep State step signals shown in Fig. 5.6. We conclude that poly- rA translocation results in $I/I_{OS} \simeq 0.13$, regardless of orientation. However, poly- rC translocation results in $I/I_{OS} \simeq 0.03$ for $3' \rightarrow 5'$ translocation and $I/I_{OS} \simeq 0.08$ for $5' \rightarrow 3'$ translocation.

5.3.6 Mid State characteristics

As detailed in Chapter 4, we observed significant differences in the characteristics of the Mid States produced immediately before rA_{50} and rC_{50} translocation. However, unlike the Deep State distributions shown in Figs. 5.2a and 5.2b, the 2-dimensional rA_{50} and rC_{50} Mid State distributions have a significant region of overlap. Consequently, rC_{50} and rA_{50} Mid States are not distinguishable at the level of individual events. However, the statistical behavior of rC_{50} and rA_{50} Mid States is readily distinguishable. Figs. 5.8a and 5.8e show the rA_{50} and rC_{50} Mid State duration distributions. We parameterized the distributions with an exponential fit and defined the exponential time constant as a phenomenological “characteristic Mid State duration.” As shown in Figs. 5.8a and 5.8e, the characteristic rA_{50} Mid State duration is significantly longer than the characteristic rC_{50} Mid State duration. Figs. 5.8b and 5.8c show the Mid State duration distributions of $rC_{50}A_{25}$ and $rA_{25}C_{50}$ MidHiLo events. As described in the previous section, we interpret these events as arising from poly rA translocation followed by poly(rC) translocation. Figs. 5.8f and 5.8g show the Mid State duration distributions of $rC_{50}A_{25}$ and $rA_{25}C_{50}$ MidLoHi events, both of which we interpret as arising from poly(rC) translocation followed by poly(rA) translocation. We find that the copolymer Mid States identified with rA -first entry have significantly longer characteristic durations than copolymer Mid States that are identified with rC -first entry. This trend suggests that the characteristics observed in rA_{50} and rC_{50} homopolymer Mid States are reflected in $rA_{25}C_{50}$ and $rC_{50}A_{25}$ Mid States and it strongly supports the conclusion of Chapter 4 that Mid State characteristics are indicative of the end of the molecule that goes on to initiate translocation. Furthermore, the $rA_{25}C_{50}$ MidNoStep Mid State distribution in Fig. 5.8d is very similar to the $rA_{25}C_{50}$ MidHiLo distribution and quite distinct from the $rA_{25}C_{50}$ MidLoHi distribution. Analogously, the $rC_{50}A_{25}$ MidNoStep distribution in Fig 5.8h is similar to the $rC_{50}A_{25}$ MidLoHi distribution and distinct from the $rC_{50}A_{25}$ MidHiLo distribution. These observations support our conjecture that most NoStep copolymer Deep States arise from $5' \rightarrow 3'$ translocation.

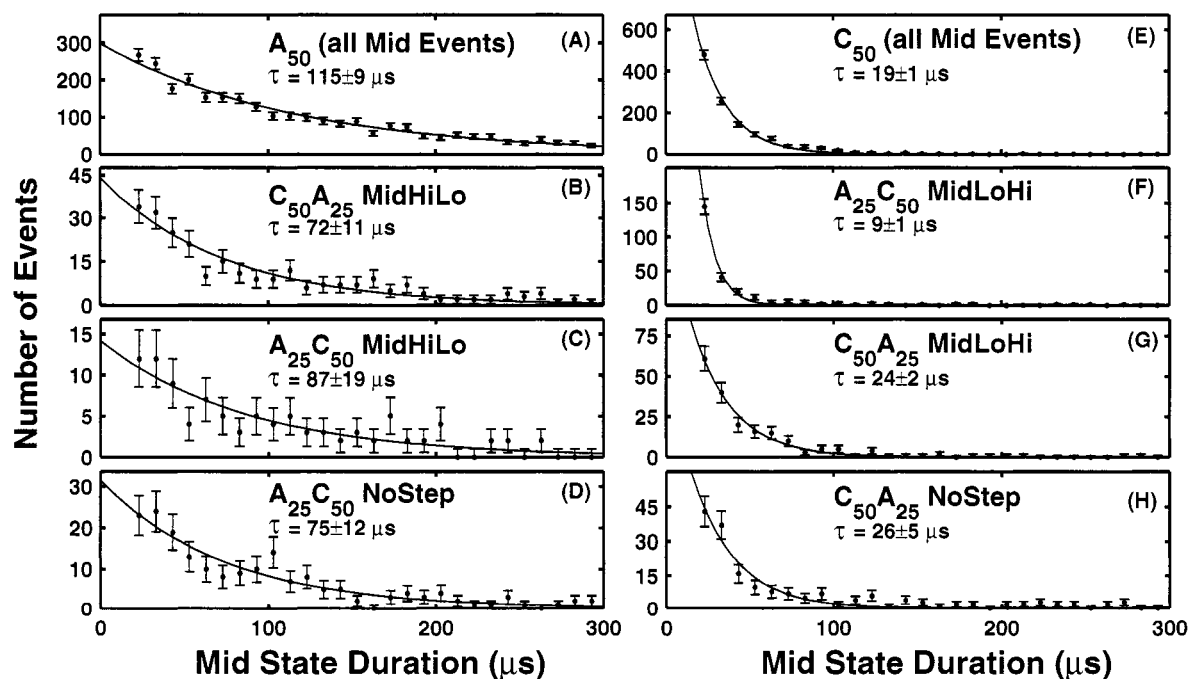


Figure 5.8: Distributions of Mid State durations. Error bars on the data points represent the statistical uncertainty in binning the data. The solid lines show exponential fits to the data. The exponential time constant and 95% confidence intervals derived from the fit are given for each distribution. Panels (a) and (e) show that Mid States produced by rA_{50} tend to be significantly longer than Mid States produced by rC_{50} . The distributions shown in panels (b) and (c) arise from events where we inferred that the molecules entered the pore with the poly(rA) segment first, while the distributions in panels (f) and (g) correspond to rC -first entry.

5.3.7 5' versus 3' capture asymmetry

The correlation between the inferred $\text{poly}(rC)$ orientation and Deep State current level suggests that translocation events in the upper rC_{50} group in Fig. 5.2a represent 5'-first entry, while events in the lower group represent 3'-first entry. By fitting the distribution shown in Fig. 5.6b to the sum of two Gaussian curves and then comparing the relative areas under each curve, we estimate that 64% of the time rC_{50} translocation begins with the 3'-end (lower group) whereas 36% of the time it begins with the 5'-end (upper group). Our classification algorithm designates 75% of the $rA_{25}C_{50}$ events as LoHi, indicative of 3'-first entry for this molecule. Following our conjecture about $5' \rightarrow 3'$ classification and evidence in the Mid State data, we make the crude but reasonable assumption that all NoStep events represent 5'-first entry of polymers into the pore. With this assumption, we estimate a lower bound of 75% on the likelihood that an $rA_{25}C_{50}$ molecule will enter the pore 3'-first and a corresponding upper bound of 25% on 5'-first entry. Thus we deduce a significant asymmetry in the capture probabilities for the two ends of $rA_{25}C_{50}$ molecules favoring the 3', $\text{poly-}rC$ end. For $rC_{50}A_{25}$, our algorithm designates 35% of the events as HiLo, which we interpret as 3'-first entry. Again assuming that all NoStep events represent $5' \rightarrow 3'$ translocation, we estimate an upper bound of 65% and a lower bound of 35% for the likelihood that a $rC_{50}A_{25}$ molecule will enter the pore 5' first or 3' first, respectively. The $rC_{50}A_{25}$ data suggests a possible capture asymmetry favoring the 5', $\text{poly-}rC$ end, but this asymmetry is dependent upon the accuracy of the assumption about NoStep event orientations. Preferred $\text{poly-}rC$ first entry of $rC_{50}A_{25}$ molecules, corresponding to $5' \rightarrow 3'$ translocation and the smaller step signal, does provide a plausible explanation for the observation that $rC_{50}A_{25}$ translocation produces a much higher percentage of NoStep events than $rA_{25}C_{50}$ translocation.

5.3.8 rA_{50} Deep State step signals

During visual inspection of individual rA_{50} translocation events, we observed frequent, sizable fluctuations in the level of ionic current obstruction. We analyzed the rms noise observed in the Open State and the Deep State of five rA_{50} experiments at 120 mV. The average variance of the Open State current for these five experiments was $\sim 55 \text{ pA}^2$, while

the average Deep State variance was $\sim 88 \text{ pA}^2$. In three rC_{50} experiments at 120 mV we also observed an average Open State variance of $\sim 55 \text{ pA}^2$, but the average Deep State variance was only $\sim 72 \text{ pA}^2$. These observations are consistent with visual inspection of translocation signals and suggest that rA_{50} translocation results in greater fluctuation of the Deep State ionic current than rC_{50} translocation. These $\text{poly}(rA)$ fluctuations can be interpreted by our data analysis as step signals, producing the rA_{50} step events shown in Fig. 5.5b. Fig. 5.9 indicates that there are significant differences in step signals produced by $\text{poly}(rA)$ fluctuations and step signals produced by transitions between $\text{poly}(rA)$ and $\text{poly}(rC)$ segments within one copolymer molecule. Specifically, the $\text{poly}(rA)$ induced Lo level does not obstruct the current as much as either of the $\text{poly}(rC)$ induced Lo levels. The data presented in Fig. 5.9, the strong agreement of $rA_{25}C_{50}$ and $rC_{50}A_{25}$ Hi and Lo current levels with rA_{50} and rC_{50} current levels, and the consistency of Mid State and Deep State orientation information all suggest that it is unlikely that $\text{poly}(rA)$ fluctuations are responsible for a large fraction of the $rA_{25}C_{50}$ or $rC_{50}A_{25}$ step events. However, the presence of $\text{poly}(rA)$ fluctuations in copolymer translocation signals likely complicates detection and characterization of $\text{poly}(rA) \rightarrow \text{poly}(rC)$ step signals. In particular, $\text{poly}(rA)$ fluctuations are likely to further hinder detection and characterization of the already small Deep State step signal arising from $5' \rightarrow 3'$ copolymer translocation.

5.4 Discussion

The characteristics we observe in rA_{50} and rC_{50} homopolymer translocation are in good agreement with those previously reported for translocation of $\text{poly}(rA)$ and $\text{poly}(rC)$ homopolymers, and the step signal characteristics we observe in $rA_{25}C_{50}$ translocation are consistent with those reported for the translocation of $rA_{30}C_{70}Gp$ RNA molecules [2]. The partitioning of translocation events into two fairly distinct groups appears to be a rather general phenomenon [2, 34, 51, 82]. Our investigation gives strong experimental support to the hypothesis that the two distinct groups observed in $\text{poly}(rC)$ homopolymer translocation represent $3' \rightarrow 5'$ and $5' \rightarrow 3'$ translocation. Recent investigations involving chemically modified ssDNA molecules [82] and hairpin ssDNA molecules with homogeneous, 50-nucleotide long adenine overhangs [46] both found a correlation between DNA orientation and the

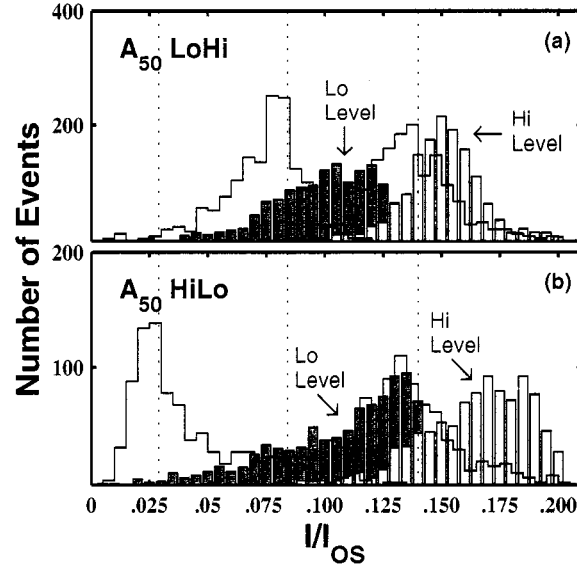


Figure 5.9: We observed frequent step signals in rA_{50} translocation. Distributions of Hi and Lo current levels for rA_{50} LoHi and rA_{50} HiLo step events are shown in (a) and (b), respectively. As a comparison, the solid black lines show the corresponding $rC_{50}A_{25}$ current level distributions. The dashed vertical lines correspond to mean current levels determined by Gaussian fits to the rA_{50} distribution and to the lower and upper rC_{50} distributions in Fig. 5.6. The Lo level in rA_{50} step events causes a smaller ionic current obstruction than either poly(rC) related Lo level. This provides evidence that step events observed in $rC_{50}A_{25}$ and $rA_{25}C_{50}$ translocation are caused by differences between poly(rA) and poly(rC) ionic current obstruction, and not by fluctuations in poly(rA) current obstruction levels.

partitioning of translocation events into two groups. In our work with rC_{50} RNA and both investigations with ssDNA, events that caused a larger obstruction in the ionic current were more frequent and were identified with 3'-first entry of polynucleotide molecules into the α -HL pore. In addition to orientation dependent differences in the level of ionic current obstruction, we also saw evidence of orientation-dependent differences in translocation durations. Continued refinement of our data analysis algorithm, the use of more sophisticated feature extraction and classification algorithms [83], improvements in the signal to noise in our measurements and more precise control of experimental parameters should allow an accurate characterization of the Deep State sub-level duration distributions. Such measurements may enable the exploration of the orientation dependence of translocation durations and may lead to meaningful comparison between experimental observations and theoretical models of translocation dynamics [44, 59].

Consistent with the investigations described in Chapter 4, these experiments provided additional evidence that the Mid State is indicative of the segment of the RNA molecule that goes on to initiate translocation. While the Mid State analysis presented in this work cannot be used to determine the orientations of individual molecules during translocation, statistical information contained in the Mid State provides useful qualitative support for our conjecture that copolymer translocation events lacking a detectable Deep State step were produced predominately by $5' \rightarrow 3'$ translocation.

Further development of experimental methods and data analysis techniques should enable the determination of $5'$ vs. $3'$ entry likelihoods with a high degree of precision. Accurate characterization of these orientation-dependent entry likelihoods combined with a more thorough investigation and characterization of the Mid State should give further insight into the capture phase of the translocation process [26, 62, 50]. Increased understanding of this capture phase is particularly important for development of detector applications of the α -HL system [36, 62].

A number of molecular level mechanisms underlying various translocation phenomena have been proposed [2, 26, 46, 50]. It was suggested that differing secondary structure is the principal cause of the current and duration differences observed in the translocation of rA_{50} and rC_{50} molecules [2] and a confinement-induced asymmetric tilt of the bases toward the $5'$

end of the polymer has been implicated as the mechanism underlying observed orientation-dependent differences in the ionic current obstruction and diffusional dynamics of ssDNA molecules [46].

5.5 Summary and conclusions

We have conducted a systematic analysis of the prominent features observed in the ionic current blockades produced by translocation of four different RNA molecules. We inferred the orientation of individual $rA_{25}C_{50}$ and $rC_{50}A_{25}$ RNA molecules as they were electrophoretically driven through the α -HL nanopore pore by monitoring the time ordering of the step signals they produced during translocation. Our analysis provided strong evidence that the orientation of a poly(rC) segment as it moves through the pore significantly affects the level at which it obstructs the ionic current flowing through the pore. Our analysis also indicated a difference in the likelihood of 5' versus 3' capture by the electric field in the pore for rC_{50} and $rA_{25}C_{50}$ molecules. Based on the analysis presented in Chapter 4, we interpret these asymmetries as indicating that 3' threading of rC_{50} into the inner constriction is easier than 5' rC_{50} threading, and that threading of either end of an rC_{50} molecule is easier than threading of either rA_{50} end. We also found clear orientation dependent trends in the Mid State, suggesting that it also contains useful information about translocation orientation. These results emphasize that polynucleotide orientation is an important factor that should be taken into consideration when interpreting translocation data, developing models of the translocation process, or developing technological applications for the *alpha*-HL system. This work also illustrates how analysis of structure within translocation events can give insight into subtle details of the translocation process. Models of the molecular mechanisms underlying these findings will lead to a better understanding of the dynamics of confined polymers and the interactions between polynucleotide molecules and the protein pore. Such understanding is vital to the development of the α -HL system as a useful model for biological processes and as a practical single molecule analytical device.

Chapter 6

ENGINEERING *MYCOBACTERIUM SMEGMATIS* PORIN A FOR NUCLEIC ACID ANALYSIS***Abstract***

Both the technological development and biophysical utility of nanopore analysis of nucleic acids would benefit greatly from the development of additional pores. We have initiated a collaborative research effort with the microbiology group of Dr. Michael Niederweis at the University of Alabama Birmingham to engineer a porin (“MspA”) found in the outer membrane of *Mycobacterium smegmatis* for application in nanopore analysis of nucleic acids. This porin has many advantageous characteristics for nucleic acid analysis including a short, narrow inner constriction, remarkable robustness, ease of use, and the retention of pore-forming activity despite the introduction of multiple amino-acid substitutions. Initial experiments did not show evidence of interaction between wild-type MspA and single-stranded DNA molecules. We identified electrostatic repulsion and steric hindrance as potential factors prohibiting ssDNA translocation through MspA. Further, voltage-dependant gating of MspA complicated the observation of potential ssDNA-induced signals and limited the level of applied voltage at which experiments were conducted. We developed a variety of mutation strategies to address these issues. Initial mutants wherein 2/3 of the negative charge of the inner constriction had been removed did not show evidence of interaction with ssDNA. However, a mutant with all of the excess negative charge removed has recently demonstrated frequent transient current blockades in the presence of ssDNA. We are presently working to verify the exciting possibility that these blockades are a result of interaction between ssDNA and the MspA mutant.

6.1 Motivation for development of a new nanopore

6.1.1 Present day limitations of α -HL for DNA sequencing

As shown in previous chapters and in the literature [30, 36, 51, 79, 82], α -HL has remarkable sensitivity to DNA composition. However, it remains unclear if α -HL possesses the degree of sensitivity required to achieve the ultimate goal of single-molecule DNA sequencing. As is shown in Fig. 4.3, the distributions of Deep State current levels for the nucleic acid homopolymers we have investigated show considerable overlap. Because they are derived from the average currents of the entire Deep States of homopolymer translocation events, the differences in these distributions suggest upper limits for base-specific differences in ionic current obstruction. The ~ 5 to 15 pA differences between the blockade levels of rC_{50} and the other homopolymers cannot be factored into an estimate of base-specific current differences because they are most likely a result of the helical secondary structure of rC_{50} . Averaging the most probable Deep State currents over repeated experiments at 120 mV, we estimate characteristic Deep State current levels (I_{DS}/I_{OS}) of 0.118 ± 0.002 , 0.119 ± 0.003 , 0.120 ± 0.006 for dC_{50} (N=6), dT_{50} (N=8), and rU_{50} (N=3), respectively (uncertainties give the standard error of the mean, n gives the number of repeated experiments). For 3' and 5'-first dA_{50} Deep States we find I_{DS}/I_{OS} of 0.104 ± 0.005 and 0.152 ± 0.005 (N=6), while for rA_{50} we find $I_{DS}/I_{OS} = 0.138 \pm 0.004$ (N=6). The average I_{OS} for all of these experiments is 126 pA. These numbers indicate that the differences between dC_{50} , dT_{50} , and rU_{50} Deep State current levels are less than 1 pA, dA_{50} Deep States differ from these three by ~ 1.9 or ~ 4.2 pA, and rA_{50} Deep States differ by ~ 2.4 pA. In nanopore sequencing applications at single nucleotide resolution, we expect that base-specific current differences are likely to be smaller than current differences found in homopolymers. Further, the trans-membrane β -barrel of α -HL is long enough to hold ~ 12 bases at any given time, and the macroscopically observed ionic current level will reflect blockade contributions from all 12 bases. A different protein pore with more optimal geometric and chemical properties may give an enhanced base-specific signal.

6.1.2 *Progress in solid-state nanopores*

Various research groups are developing nanopores that have been drilled or etched in solid-state membranes for use in nanopore analysis [23, 25, 43, 75]. Solid-state pores have a number of advantages over protein pores including robustness to a wider variety of chemical and physical environments and the possibility of integration directly with nanoscale electronic sensors. Despite these advantages and considerable activity in this field, there has been no experimental demonstration of nucleic acid analysis with a solid-state pore that has comparable sensitivity to α -HL. Further, it is unlikely that techniques to produce these solid-state pores will, in the near future, achieve the truly sub-nanometer scale engineerability and reproducibility that is intrinsic to protein pores. Continued exploration of protein pores thus remains very important for the technological development of nanopore analysis of nucleic acids.

6.1.3 *Scientific importance of additional pores*

The utility of nanopore analysis of nucleic acids as a model system for biological processes is strongly dependant on the generality of the insights it reveals. Experimental results with α -HL have motivated a large number of theoretical investigations seeking to elucidate the underlying physics of the system. Without additional pores it is very difficult to determine which aspects of the α -HL data are attributable to the general physics underlying translocation of a confined polymer, and which are due to the particular structure of α -HL and specific interactions between nucleic acids and the residues lining the interior of the α -HL pore. Development of an additional pore will show which aspects of the α -HL data are pore-specific and which aspects are potentially of broad generality.

6.2 *Mycobacterium smegmatis* porin A is a promising candidate for nanopore analysis of nucleic acids

6.2.1 *Biological role of MspA*

Mycobacterium smegmatis (*M. smegmatis*) is a species of rapidly growing mycobacterium normally found in soil, water and dust. Unlike its well known, slowly growing cousins *M.*

tuberculosis and *M. leprae*, *M. smegmatis* is not generally considered a human pathogen [28]. One distinguishing characteristic of mycobacteria is their $\sim 90\text{-}100$ Å thick outer membrane that is rich in exceptionally long fatty acids called mycolic acids. It is hypothesized that the permeability barrier created by this thick outer membrane is an important factor in the intrinsic resistance of mycobacteria to most common antibiotics, chemotherapeutic agents and chemical disinfectants [63]. MspA has been identified as the major diffusion pathway for hydrophilic solutes across the outer membrane of *M. smegmatis* [74]. MspA has been purified, its gene, *mshA*, has been cloned [64] and its structure has been determined by X-ray crystallography [19]. As with α -HL, translocation of polynucleotide molecules is not an *in vivo* function of MspA.

6.2.2 *MspA structure is promising for nanopore analysis*

MspA is an octomeric complex composed of identical monomers. Each monomer contributes a 134 residue globular domain to the rim of the pore and a 50 residue loop. The loop participates in two β -barrels, the wider, longer of which constitutes the stem region of the pore and the shorter, narrower of which constitutes the base and contains the inner constriction (Fig. 6.1). In the outer membrane of *M. smegmatis*, the MspA rim is exposed to the environment while the base is pointed in toward the periplasm [19]. The pore complex is ~ 96 Å tall and has outer diameters of ~ 88 Å and ~ 37 Å at the rim and base, respectively. The outer surface of the rim presents hydrophilic residues that are presumed to interact with water while the outer surfaces of the stem and base regions are hydrophobic and presumed to be embedded in the mycolic-acid rich outer membrane of *M. smegmatis*. It is interesting to note that the width of the hydrophobic band on the stem/base outer-surface is only ~ 37 Å, in stark contrast to the proposed ~ 90 Å thickness of present models of the mycobacterial outer membrane. A recent study using cystine scanning mutagenesis to determine the accessibility of various residues to membrane-impermeant reagents suggests that an appreciable portion of the hydrophilic rim region is embedded in the outer membrane of *M. smegmatis* [45].

The short inner constriction of MspA is of particular importance to nucleic acid analysis. In contrast to α -HL, where the inner constriction is followed by a ~ 50 Å long, ~ 20

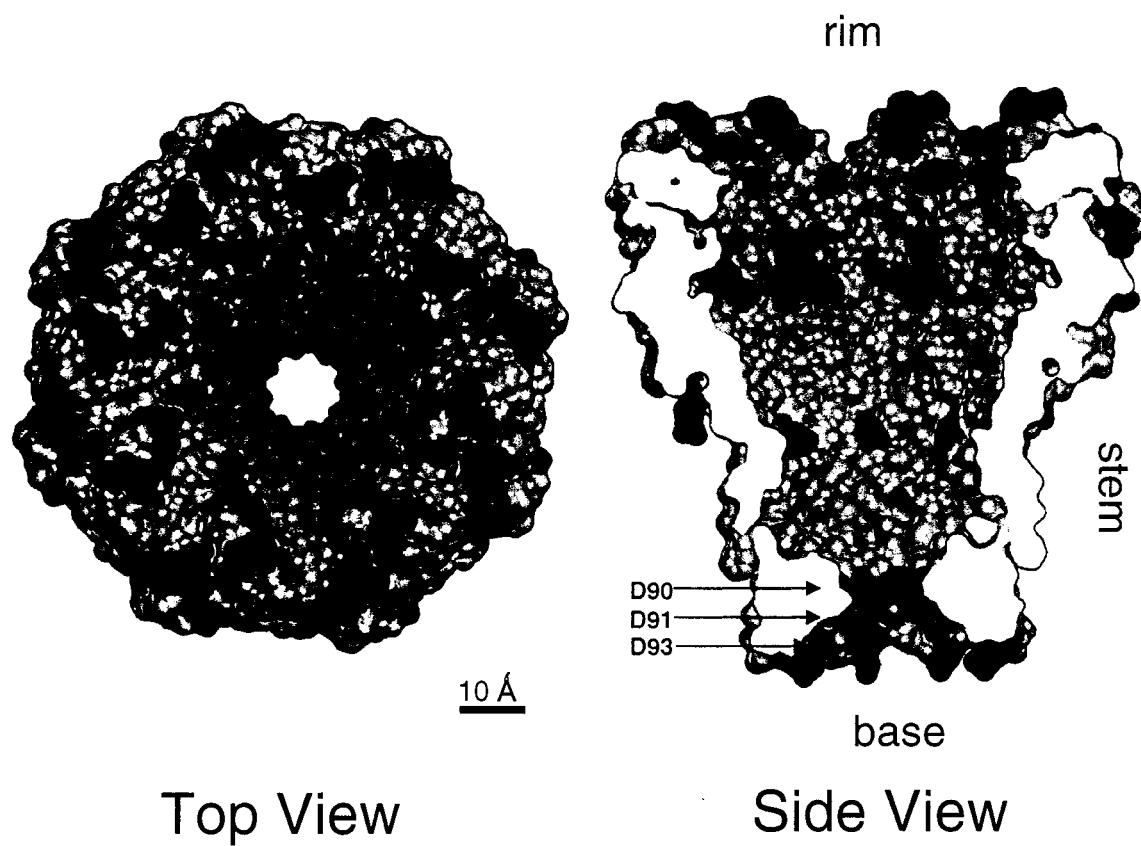


Figure 6.1: Top view and cross-sectional side view of a space-filling model of MspA. Positively-charged basic residues are shown in blue and negatively-charged acidic residues are in red. The surface of intersection between the protein and the clipping plane is colored light gray. The locations of the aspartic acid residues that we have thus far mutated (D90, D91, and D93) are indicated in the side view. The 10 Å scale bar applies to both views.

\AA diameter β -barrel, the MspA inner constriction is flanked on both sides by much wider channels. Figure 6.2 shows a side-by-side view of the two pores. The shorter MspA constriction is expected to improve the spatial resolution of the ionic current signal by decreasing the length of the blocking polymer that contributes significantly to the observed resistance. The bottom two images of Fig. 6.2 show expanded views of only the residues that define the inner constrictions of the two pores. Both inner constriction models are shown on the same scale. It is clear that the diameters of the two inner constrictions, as determined from their crystal structures, do not differ by more than 1 or 2 \AA . It is interesting to note that in the α -HL structure report [73] and in all subsequent reports on nanopore analysis with α -HL, the inner constriction diameter is quoted as ~ 14 or ~ 15 \AA . The MspA structure report quotes ~ 10 \AA for the diameter of the MspA inner constriction. Perhaps the α -HL estimate gives atomic center-to-center distances across the constriction, while the MspA estimate gives atomic surface-to-surface distances. Despite the fact that different inner constriction diameters are given in the literature, we conclude from our own analysis of the coordinate files that the inner constrictions of MspA and α -HL have nominally the same diameters, which can be taken as ~ 10 – 11 \AA (atomic surface-to-surface distance).

6.2.3 Additional advantageous characteristics of MspA

The fully-formed MspA porin is extremely robust [24]. Heating to 92°C is required to dissociate the octamer and this dissociation temperature is not appreciably reduced by the presence of detergent (2% SDS) or strong denaturant (7.6 M Urea). MspA is also resistant to dissociation despite incubation at pH 1 or at pH 14. In addition to its demonstrated robustness, we have found MspA pore formation in bilayer experiments to be extremely reliable, requiring less time to obtain a stable pore and demonstrating longer pore lifetimes than α -HL. Perhaps the most advantageous property of MspA is its engineerability. Our collaborators have developed an efficient expression and purification system for MspA mutants in *M. smegmatis* [45]. They have produced 33 different mutants with substitutions in the inner constriction, all of which were well expressed and most of which retain channel-forming ability.

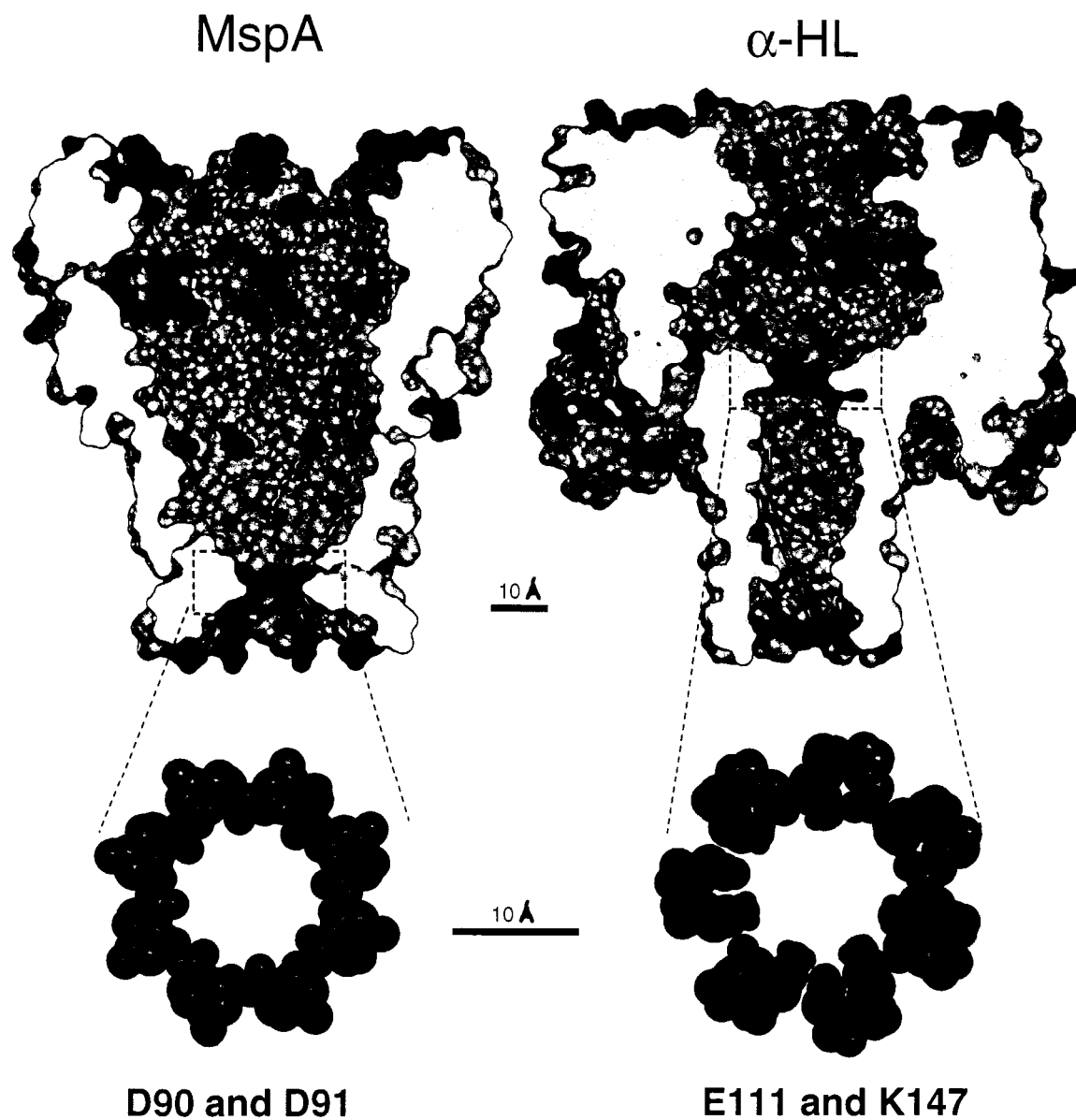


Figure 6.2: Top images show cross-sectional side views through space-filling models of MspA and α -HL. Positively charged basic residues are shown in blue and negatively-charged acidic residues are in red. The surface of intersection between the protein and the clipping-plane is colored gray. Bottom images show expanded views of only the residues that define the inner constrictions of the two pores. Note the different size scales for the top and bottom images.

6.3 Experiments with Wild-Type MspA

Conductance and gating characteristics of Wild-type MspA

We followed our α -HL experimental procedure and found that wild-type MspA (WT-MspA) readily formed pores in our bilayers. Figure 6.3a shows part of the time series of the current flowing across a bilayer with an applied voltage of nominally 10 mV after adding ~ 4 ng/ml WT-MspA to the *cis* compartment. Each step in the current corresponds to the insertion of one pore. Figure 6.3b shows the sizes of all of the current steps observed in this particular experiment. There is a clear primary current step size, but there are also minority populations of larger steps. The average primary current step is 54.6 pA while the average step sizes of the secondary populations are 109 pA and 164 pA. The observation that the step sizes of the secondary populations are exactly two and three times primary step size strongly suggests that these secondary events represent multiple pore insertions. We are unable to resolve brief sub-states within the presumed double and triple insertion events, indicating that the multiple insertions cannot be more than ~ 20 μ s apart. Because the average time interval between typical insertion events is ~ 10 s, we conclude that there must be some mechanism that occasionally acts to synchronize pore insertion. In two identical repeated experiments we observed average primary step sizes of 43 pA and 47 pA. We attribute these differences in single pore conductances to different junction potential offsets for the electrodes contacting the *cis* and *trans* buffer. This electrode offset was eliminated only in the experiment yielding a single channel current of 47 pA, and so we use this value to estimate a single-channel conductance of 4.7 nS (at +10 mV) for WT-MspA. This value is in excellent agreement with the value of 4.6 nS found in the literature [64].

Figure 6.4 shows 70 second snapshots of the ionic current through the one WT-MspA pore at 60 and 100 mV. The uppermost levels in the traces show the unblocked, Open State signal with positive voltage applied to the *trans* side (our typical arrangement for DNA translocation experiments). Gating of the pore is observed as downward deflections of the ionic current out of the Open State. We identify two types of gating events. In “transient” gating events the current remains at a reduced level for only a brief time before returning spontaneously to the Open State. In “indefinite” events the current remains at a reduced

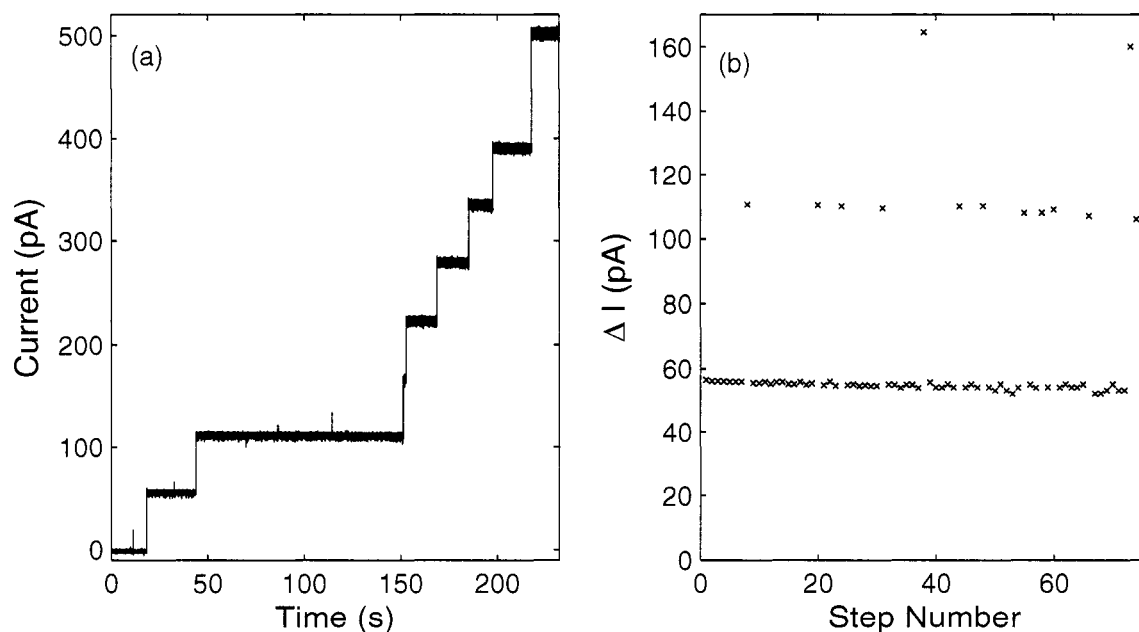


Figure 6.3: Conductance steps from insertion of individual wild-type MspA porins. (a) Time series of the ionic current flowing across the bilayer with an applied voltage of 10 mV and ~ 4 ng/ml WT-MspA in the *cis* compartment. Steps in the ionic current correspond to insertion of either individual pores or nearly simultaneous insertion of 2 or three pores. (b) Ionic current step sizes for all the insertion events of the experiment shown in (a). The minority populations of large current steps are almost exactly two and three times as large as the primary current step. The small linear decrease in ΔI with increasing step number is likely caused by drift in the junction potentials for the electrodes, and the increased scatter with increasing step number results from the fact that the pores already present in the bilayer increase the noise in the baseline current level.

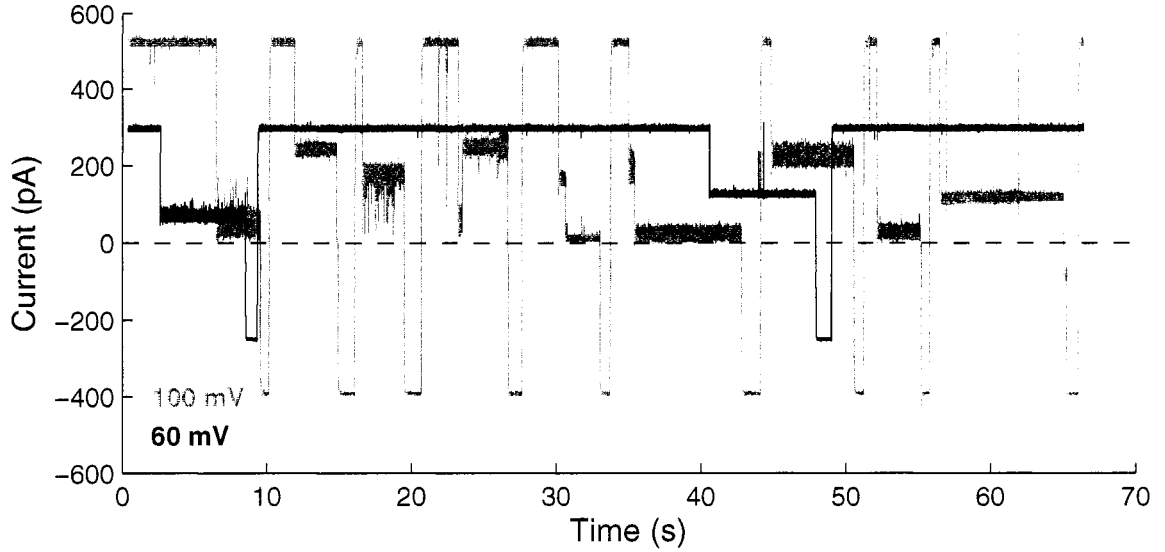


Figure 6.4: Snapshots of the ionic current through a WT-MspA pore at 60 and 100 mV. The uppermost levels in the traces show the unblocked, Open State signal with positive voltage applied to the *trans* side. Gating of the pore is observed as downward deflections of the ionic current. Often the current remains at a reduced level for several seconds and requires brief flipping of the polarity of the applied voltage in order to return to the Open State. Voltage flips are observed as brief intervals of negative current flow.

level for several seconds and requires flipping of the polarity of the applied voltage in order to return to the Open State. Voltage flips are observed in Fig. 6.4 as brief intervals of negative current flow.

The traces in Fig. 6.4 illustrate a number of characteristics we have observed in individual WT-MspA pores. To investigate the current-voltage properties of the pores we looked at the incremental conductance, $g(V) \equiv (I(V_2) - I(V_1))/(V_2 - V_1)$ where V is the voltage mid-way between the two test voltages V_1 and V_2 . From the data in Fig. 6.4 we estimate $g(80 \text{ mV}) \sim 5.6 \text{ nS}$ and $g(-80 \text{ mV}) \sim 3.6 \text{ nS}$ (Open State current at negative voltage was obtained from the periods of voltage polarity reversal). Using additional data for this pore at +20 mV and -20 mV we estimate $g(0 \text{ mV}) \sim 4.5 \text{ nS}$. These estimates show that, for our experimental conditions, the I-V curve for WT-MspA is non-linear and asymmetric with greater current flow for *trans*-side positive application of a given voltage than for *trans*-side negative. The traces in Fig. 6.4 also illustrate the strong voltage dependence of the gating

behavior of WT-MspA. Below ~ 60 mV gating is rarely observed, while above ~ 100 mV the pore only remains in the Open State for a few seconds before the occurrence of an indefinite gating event. Further, the gating behavior of the pores is extremely complex as evidenced by the presence of at least 6 clearly distinct sub-states in the 100 mV trace of Fig. 6.4. These characteristics are important to our translocation experiments as the asymmetry of the I-V curve reveals the orientation of a pore in the bilayer. Because the gating signals look like our expected nucleic acid translocation signals, the intrinsic gating behavior will complicate the implementation and interpretation of nucleic acid translocation experiments.

The addition of ssDNA or RNA to the *cis* compartment caused no obvious change in the gating pattern of WT-MspA pores. This result indicates that nucleic acids are either unable to be electrophoretically driven into WT-MspA or that their presence in the pore produces a very weak signal. Below we discuss several mechanisms that potentially prevent or mask the signal of DNA interaction with WT-MspA.

6.4 Possible mechanisms prohibiting interaction between DNA and wild-type MspA

The MspA inner constriction contains 24 negatively-charged residues

Another look at the MspA structure reveals a clear barrier to DNA entry – each monomer contains negatively charged aspartic acid residues at positions 90 and 91. These residues define the MspA inner constriction and result in a total charge of $-16e$ ringing the constriction. Position 93, which is directly adjacent to the constriction toward the base of the pore, is also an aspartic acid residue. These charges present a significant electrostatic barrier to entry of negatively charged DNA into the inner constriction of MspA and we postulate that the barrier is large enough to prevent translocation.

Voltage-dependant gating of MspA

As described earlier, WT-MspA has strongly voltage-dependent gating behavior. DNA translocation experiments at applied voltages of ~ 120 mV or higher (*trans*-side positive) are impractical due to rapid pore closure. In experiments with α -HL, the rate of DNA

capture by the pore is strongly voltage dependant [26, 50, 62] and it is extremely hard to observe translocation at voltages below 80 mV. We expect that DNA capture by MspA is also strongly voltage dependant, and the minimum practical voltage for observing translocation in MspA may be considerably higher than the ~ 80 mV value of α -HL. The upper limit of ~ 120 mV imposed on experiments with WT-MspA by pore gating is thus a significant obstacle to observing DNA translocation.

The cis-side entrance of MspA has an excess of negative charge and is far away from the inner constriction

Analogous to the α -HL M_0 events from DNA in the vestibule configuration (Chapter 4), it is plausible that entry of a polynucleotide molecule into the stem region of a WT-MspA pore (Fig. 6.1) would cause a shallow blockade of the ionic current. The electrostatic and steric barriers presented by the inner constriction would not preclude DNA entry into the stem region. However, we observed no shallow blockade events in our experiments with WT-MspA, suggesting that an inability of DNA to penetrate the *cis* entrance, and not the barriers of the inner constriction, may be the immediate obstacle preventing translocation. The surface of the rim region surrounding the entrance to the WT-MspA pore (Fig. 6.1) has 56 negatively charged residues (D13, D56, E57, E59, D134 and E139, D171) and only 24 positively charged residues (R14, R45 and K47), leaving it with a net surface charge of $-32e$. This is in stark contrast to α -HL, where the surface surrounding the *cis*-side entrance has a net charge of $+21e$. The excess negative charge around the WT-MspA entrance makes it energetically unfavorable for nucleic acids to approach and enter the pore. Despite considerable counterion screening from the high concentration of K^+ ions in our experiments (1 M), this electrostatic barrier may be sufficient to prevent nucleic acid entry into the WT-MspA stem. Further, the gradient in the externally applied voltage is expected to be quite small at the *cis*-side entrance. Most of the voltage will drop across the inner constriction, which is ~ 85 Å away from the entrance. The net negative charge of the surface surrounding the entrance and the longer distance from the entrance to the inner constriction suggest that *cis*-side entry of DNA into WT-MspA may be much more difficult than for α -HL.

These barriers are largely absent for *trans*-side entry of DNA into WT-MspA. However, our present experimental geometry (Fig. 3.1) prevents us from accessing the *trans* side of the bilayer. We are in the process of modifying our apparatus to allow *trans*-side access so that we can test for DNA translocation through MspA in both directions.

6.5 Engineering MspA to facilitate electrophoretic detection of DNA

Removal of negative charge from the inner constriction

Re-engineering of the WT-MspA inner constriction will likely be required due to the large electrostatic barrier created by the negatively charged residues at positions 90, 91 and 93 (Fig. 6.1). In addition to eliminating the negative charge, it will be desirable to use residues with small side-chains in order to minimize any possible steric barriers. We have opted to use serine residues at positions 90 and 91 and asparagine at position 93 because these residues are small, neutral, and hydrophilic. An alternative strategy to overcome the negative charge of the inner constriction is the introduction of a positive charge at either position 90 or 91 while leaving the aspartic acid at the other inner constriction position and introducing a neutral residue at position 93. It is interesting to note that a ring of alternating positive and negative residues define the inner constriction of α -HL. This motif may be a critical component of the nucleic acid analysis capabilities of α -HL.

Suppression of voltage-dependant gating

The bottom of the base of MspA faces the periplasm when the porin is in the outer membrane of *M. smegmatis* and its constituent residues (positions 91-103) form loops that do not participate in either of the β -barrels of the pore. We refer to these structures as the “periplasmic loops” and we hypothesize that fluctuation of the periplasmic loops into the inner constriction is a possible mechanism for the voltage-dependant gating of WT-MspA. Since loops usually have no structural function in proteins, our first approach will be to delete larger and larger portions of the periplasmic loops. Each successive deletion mutant will be tested for reduction of voltage dependant gating. If deletion is unsuccessful, a second approach will be to fix the periplasmic loops in an open conformation by creating

disulfide bridges between cysteines in adjacent loops. This will be achieved by exchanging the residues threonine 95 and leucine 101 to cysteines.

Removal of negative charge from the rim domain

As with the inner constriction, the electrostatic barrier created by the excess negative surface charge of the rim domain can be reduced by introducing residues with either neutral or positive side chains. The negative charge of D13 is counterbalanced by the adjacent arginine R14 and will be left unchanged. To avoid early potential problems with untested mutation sites we will construct the mutants E139Q (substitution of the glutamate residue, E, at position 139 with a glutamine residue, Q) and D134N, which were already shown not to affect the function or stability of MspA [45], before making mutations D56R and E57Q. The rim of the final mutant will have a net surface charge of $+8e$.

6.6 Progress on experiments with MspA mutants

6.6.1 D90S/D91S shows reduced gating, but no interaction with DNA

We have conducted a number of experiments with the D90S/D91S mutant (“SS-MspA”). We observe two distinct “types” of pores that are differentiated by their conductance properties. The conductance properties of both types of pores are asymmetric with respect to the polarity of the applied voltage; for *trans*-side negative voltage the current-voltage relationship appears to be linear, while for *trans*-side positive voltage it exhibits a small but clear nonlinearity. In the linear direction, “low conductance” pores demonstrate a conductance of 1.40 ± 0.13 nS ($N=7$, uncertainties give standard error in the mean for repeated experiments), while “high conductance type” pores exhibit a conductance of 3.01 ± 0.07 nS ($N=4$). In the nonlinear direction, the I-V curves were well fit by a quadratic functional form: $I(V) = GV + QV^2$. In this direction, the low conductance pores were characterized by $G = 1.42 \pm 0.06$ nS and $Q = (3.11 \pm 0.14) \times 10^{-3}$ nS², while the high conductance pores demonstrate $G = 3.19 \pm 0.07$ nS and $Q = (3.49 \pm 0.10) \times 10^{-3}$ nS². Because the quadratic coefficient of the I-V curves of large conductance pores is not a factor of two greater than the quadratic coefficient for small conductance pores, we believe that these the two types of

SS-MspA pores represent distinct populations. This is in contrast to the “double-insertion” explanation for the large conductance steps observed in experiments with WT-MspA (Fig. 6.3). A surprisingly convenient property of SS-MspA pores is that they remain largely free of gating behavior up to and above applied voltages of 200 mV. This result refutes our hypothesis that modification of the periplasmic loops will be needed to suppress the gating behavior and raises the possibility that we will not need to modify the loops in order to observe translocation.

The addition of ssDNA or RNA to the *cis* compartment did not result in any noticeable change in the characteristics of the ionic current flowing through an SS-MspA pore. These results suggest that either the excess negative charge on the surface of the rim domain or the remaining negative charge of D93 near the inner constriction is the immediate barrier preventing observable interaction between ssDNA and MspA.

6.6.2 The MspA triple mutant D90S/D91S/D93N exhibits an increased rate of transient current blockades in the presence of ssDNA

Distribution of single-channel conductances for D90S/D91S/D93N pores

We next conducted experiments with the MspA triple mutant D90S/D91S/D93N (“SSN-MspA”). SSN-MspA does not have negatively charged residues in or directly adjacent to the inner constriction. We first ran experiments where many pores insert into a bilayer in order to assess the heterogeneity of single-channel conductances. As shown in Fig. 6.5 the conductance distribution shows a peak at ~ 25 pA. At the present time we do not know the origin of the broadness observed in the SSN-MspA single-channel conductance distribution.

SSN-MspA gating and current-voltage characteristics

We also analyzed the voltage-dependant gating characteristics and current-voltage behavior of 12 individual SSN-MspA pores. Figure 6.6a shows snapshots of the ionic current through an SSN-MspA pore at three different voltages. These traces show how SSN-MspA does not exhibit gating behavior for positive voltages up to and above 200 mV, while transient and indefinite gating is observed for voltages more negative than ~ -80 mV. Like WT-MspA

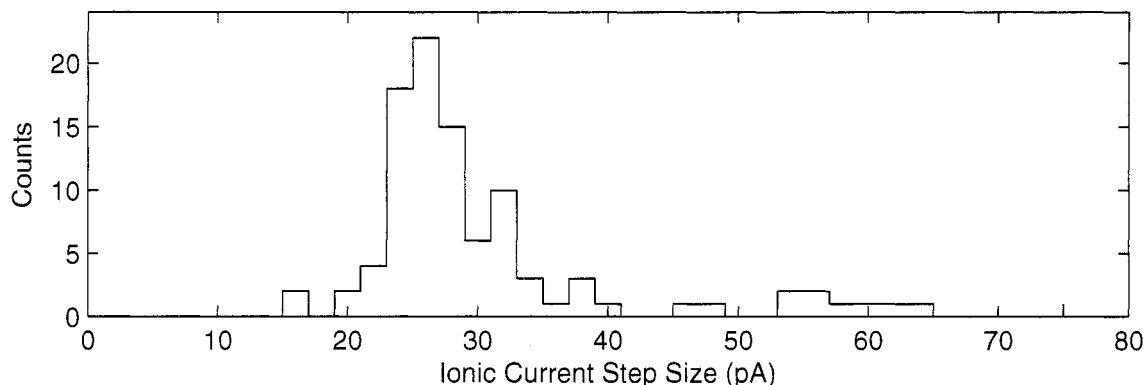


Figure 6.5: We added ~ 1 ng/ml of D90S/D91S/D93N MspA to the *cis* compartment while applying 10 mV across the bilayer. Analogous to the experiments with WT-MspA shown in Fig. 6.3, pore insertion was observed as a step in the ionic current flowing across the bilayer. SSN-MspA insertions from six repeated experiments were combined in this figure, and the resulting distribution shows a peak at a step sizes of ~ 25 pA and a possible second peak at ~ 31 pA.

and SS-MspA, SSN-MspA exhibits nonlinear, asymmetric current voltage characteristics. However, unlike WT-MspA and SS-MspA, SSN-MspA passes more current for *trans*-side negative application of a given voltage than for *trans*-side positive. We have noticed the existence of multiple types of pores in our SSN-MspA sample which are differentiated by their conductance-voltage properties. We have reproducibly observed two of these types in our experiments. We define these reproducible types as Type 1 and Type 2 pores, and their conductance-voltage relationships are shown in Fig. 6.6b. Of the 12 SSN-MspA pores we have analyzed, 5 had Type 1 conductance-voltage behavior, 4 had Type 2 behavior, and three were distinct from both of these types and from one another. The Type 1 and Type 2 curves have similar shape, both demonstrating minima in the conductance at moderate positive voltage, but they differ considerably in the magnitude of the conductance for positive voltage. Both of the pores shown in Fig. 6.6b would fall beneath the main peak in the distribution of Fig. 6.5. While we do not know the cause of the heterogeneity of SSN-MspA pore characteristics, we are nonetheless able to reproducibly obtain both Type 1 and Type 2 pores. Because these pore types remain largely free of gating and excessive noise at voltages above +200 mV, we will test both types for interaction with nucleic acids.

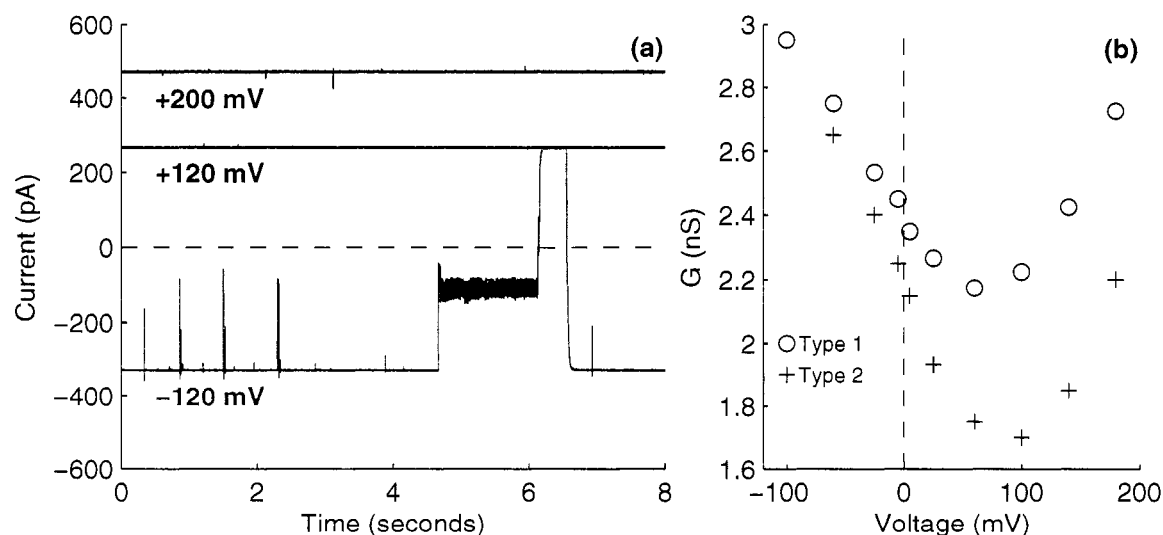


Figure 6.6: Gating and conductance-voltage behavior of SSN-MspA pores. (a) Time series of the ionic current signal through an SSN-MspA pore at three different voltages. We do not observe gating behavior for positive voltages up to and above 200 mV, while transient and indefinite gating is observed for voltages more negative than ~ -80 mV. (b) Incremental conductance, $g(V)$, of two different SSN-MspA pores. $g(V) \equiv (I(V_2) - I(V_1))/(V_2 - V_1)$ where V is the voltage mid-way between the two test voltages V_1 and V_2 . Of the 12 SSN-MspA pores we have analyzed, 5 had Type 1 conductance-voltage behavior, 4 had Type 2 behavior, and three were distinct from both of these types and from one another.

Transient blockades in the presence and absence of single-stranded DNA

We have conducted a detailed analysis of the blockade characteristics of SSN-MspA in the presence and absence of ssDNA. We detect the beginning of a blockade event when the ionic current drops below $I_{OS} - 3\sigma_{OS}$ and the end of the event when the ionic current subsequently rises above $I_{OS} - \sigma_{OS}$, and we exclude events that are shorter than 10 μ s. Figure 6.7 shows 12 sample blockade events from data collected at +200 mV with 4 μ M dA_{50} in the *cis* compartment. We parameterized each event by its duration and by the minimum current value observed within the blockade. We used the minimum current instead of the average current because most of the blockades were too brief to be fully resolved by our amplification electronics, and they thus appear as “spikes” and do not demonstrate well-defined conductance states. The majority of the blockades we observed were shorter than 50 μ s and had a minimum current that was above $I_{OS}/2$. We defined these blockades as “Short-Mid” events. Especially in the presence of DNA, we observed many “Short-Deep” events with duration less than 50 μ s and minimum current below $I_{OS}/2$. Blockades longer than 50 μ s also tended to demonstrate a minimum current that was either well above or well below $I_{OS}/2$, and we defined these signals as “Long-Mid” and “Long-Deep” events, respectively. The classification of each event is listed above the trace in Fig. 6.7.

Figure 6.8 shows the blockade characteristics of one SSN-MspA pore before the addition of DNA, in the presence of ~ 4 μ M dA_{50} , and after the dA_{50} has been removed from the *cis* compartment by perfusion with DNA-free buffer. With +200 mV applied across the bilayer, SSN-MspA exhibits 0.1 to 1 transient blockades per second even in the absence of DNA. These “control” events look qualitatively the same as the events shown in Fig. 6.7. When 4 μ M dA_{50} is added to the *cis* compartment, the blockade rate increases by a factor of 10 to 100, and when the dA_{50} is removed by perfusion with DNA-free buffer, the blockade rate is reduced almost to its initial level. The dashed vertical and horizontal lines are at the 50 μ s and $I_{OS}/2$ boundaries that we used to classify the events, and the four numbers in the upper right of each panel give the number of blockades per second for each of the four types of events. The blockade characteristics shown in Fig. 6.8 were qualitatively reproduced in four repeated experiments. Short Mid events were always the most common, followed

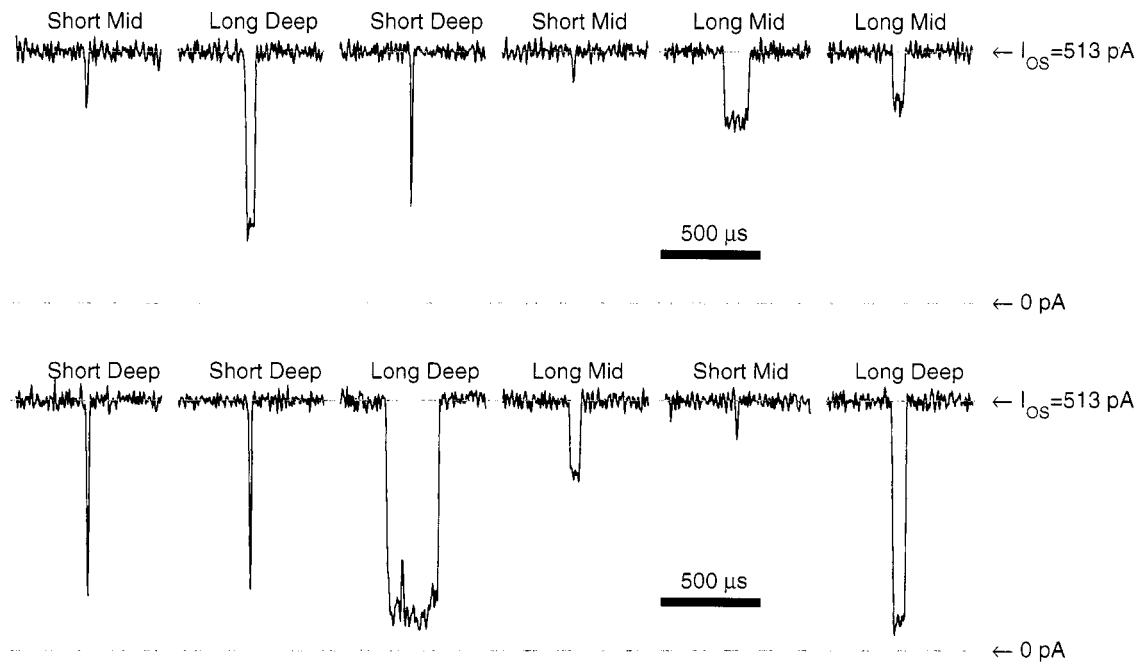


Figure 6.7: Example blockade events from an SSN-MspA pore in the presence of $4 \mu\text{M}$ dA_{50} . We classified each event as one of four types; “Short-Mid” events are shorter than $50 \mu\text{s}$ and have a minimum current greater than $I_{OS}/2$, “Long-Mid” events are longer than $50 \mu\text{s}$ and have a minimum current greater than $I_{OS}/2$, “Short-Deep” events are shorter than $50 \mu\text{s}$ and have a minimum current less than $I_{OS}/2$, and “Long-Deep” events are longer than $50 \mu\text{s}$ and have a minimum current less than $I_{OS}/2$.

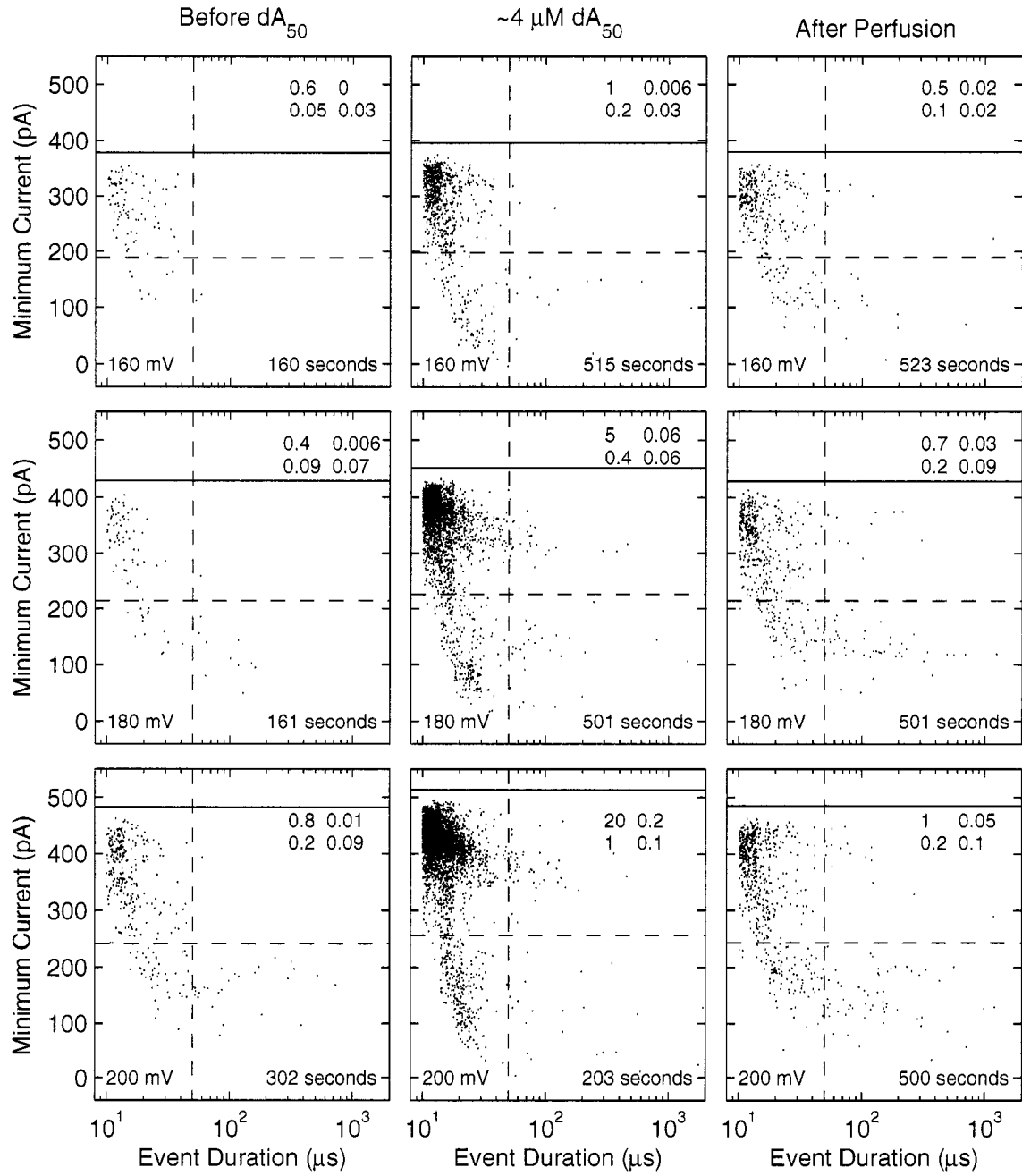
by Short Deep events. The rates of Short Mid, Short Deep and Long Mid events always increase by at least a factor of 10 in the presence of dA_{50} , while the rate of Long Deep events seems independent of the presence of dA_{50} . The magnitude of the dA_{50} -correlated blockade rate increases varied by more than a factor of 10 in different experiments. We hypothesize that this large variability in blockade rate increase reflects differences in the accessibility of SSN-MspA to the *cis* compartment buffer in different experiments. It is interesting to note in Fig. 6.8 that the Open State current level increases by $\sim 5\%$ when dA_{50} is present in the *cis* compartment. We have observed this effect in most but not all of our experiments with dA_{50} and we do not yet have a good explanation for this phenomena.

Potential mechanisms for the increased blockade rate in SSN-MspA

The data in Fig. 6.8 are consistent with the scenario where dA_{50} molecules are electrophoretically driven into the SSN-MspA pore and cause transient blockades of the ionic current. Perhaps the increase in the rate of Short Mid signals results from escape events analogous to the α -HL M_0 signals, while the increase in the rate of Short Deep events represents translocation of dA_{50} through SSN-MspA. Our inability to resolve conductance levels within most of the events would result if dA_{50} escape and translocation were much faster for SSN-MspA than for α -HL.

While this dA_{50} -induced blockade explanation is both plausible and encouraging, we can identify at least two other candidate mechanisms for the observed blockade rate increase. First, it is possible and even likely that MspA gating can be induced by molecules other than single-stranded nucleic acids. For example, gating in α -HL can be induced by a variety of divalent and trivalent cations [53]. Such molecules would be contaminants in our DNA experiments. If these hypothetical contaminant molecules are present in the dA_{50} aliquots that we add to the *cis* compartment, then we would observe the correlation between increased blockade rate and the presence of dA_{50} . A second possibility is that the moderate rate of transient blockades observed in the DNA-free control experiments results from intrinsic conformational fluctuations of the SSN-MspA structure. As described earlier, fluctuation of a periplasmic loop into the inner constriction is one possible gating

Figure 6.8: Each point gives the duration and minimum current observed within one transient blockade. All data in this figure were collected with the same pore. The solid horizontal lines show the average Open State current levels. The left row illustrates blockade behavior before the addition of dA_{50} to the *cis* compartment, the center row shows blockade behavior with $\sim 4 \mu\text{M}$ dA_{50} in the *cis* compartment, and the right column shows blockade behavior after perfusion of the $\sim 200 \mu\text{l}$ *cis* compartment with 6 ml of DNA-free buffer. Rows correspond to data taken with either 160 mV, 180 mV or 200 mV applied across the bilayer (*trans*-side positive). The rate of transient blockades is significantly higher when dA_{50} is present, and the effect becomes more pronounced as the voltage increases. The dashed vertical and horizontal lines are at the $50 \mu\text{s}$ and $I_{OS}/2$ boundaries that we used to classify the events, and the four numbers in the upper right of each panel give the rates of each of the four types of events in units of events per second. The upper left number gives the rate of Short Mid blockades, the upper right number gives the rate of Long Mid events, and the two lower numbers give the rates of Short Deep and Long Deep events.



mechanism. Perhaps dA_{50} molecules bind to the outer surface of SSN-MspA and change the conformational free-energy landscape in a manner that favors partially closed conformations. In this scenario, the increase in the blockade rate is caused by interaction between SSN-MspA and dA_{50} , but the blockade mechanism is not physical obstruction of the ionic current flow by the presence of dA_{50} inside of the SSN-MspA pore.

Determination of blockade mechanism

Experiments with single-stranded DNA hairpin constructs may provide a convenient way to test if the dA_{50} -correlated blockades shown in Fig. 6.8 are caused by DNA entry into the SSN-MspA pore. If the duplex region of the hairpin construct is too wide to pass through the inner constriction of SSN-MspA, then the molecule will remain trapped until the duplex dissociates and allows completion of translocation. We thus hope to observe a substantial increase in the rate of Long Deep events in experiments with hairpin constructs. We will run experiments with two hairpin constructs, one containing a 5 base-pair duplex region (“hp-5”) and the other containing a 10 base pair duplex region (“hp-10”). We expect the hp-10 experiments to show Long Deep events with significantly greater duration than those of hp-5 experiments. Analogous effects have been reported for experiments involving DNA duplex dissociation in α -HL [29, 46, 79].

Positive results with the hairpin experiments described above would constitute very strong, though still indirect, evidence for translocation. It will be necessary to obtain direct proof of translocation by directly detecting single-stranded DNA molecules on the *trans* side of the bilayer after an experiment [25, 34]. If the blockade characteristics for the hairpin constructs do not differ substantially from those of dA_{50} , then either the SSN-MspA inner constriction is wide enough to allow passage of the duplex region without dissociation, or the blockades are not caused by the presence of DNA inside of the SSN-MspA pore. Experiments involving ssDNA constructs linked to a “blocking” particle that is clearly too wide to pass through MspA [26] could resolve this ambiguity. If these experiments indicate that DNA is unable to pass through the SSN-MspA pore, then we will continue developing and testing additional mutants for the ability to translocate ssDNA molecules.

6.7 Outlook

The first critical milestone for this project will be verification of ssDNA translocation through a mutant MspA pore. We are hopeful that this can be achieved with the SSN-MspA mutant. Even if the blockades shown in Fig. 6.8 are not caused by translocation of dA_{50} , there are still a number of reasonable mutation strategies to pursue in our effort to engineer another protein pore that allows DNA translocation. If we achieve this first milestone, then we will begin a series of comparative experiments with a variety of mutants with the dual goals of optimizing MspA for nucleic acid analysis and understanding the nanoscale mechanisms that govern electrophoretic translocation of ssDNA and RNA through MspA. Results from these experiments will give significant insight into the physics underlying nanopore analysis and will hopefully lead to new MspA-based nanopore biosensors with improved analytical capabilities.

BIBLIOGRAPHY

- [1] National Human Genome Research Institute RFA-HG-05-004. Revolutionary Genome Sequencing Technologies The \$1000 Genome, 2005. <http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-05-004.html>.
- [2] M. Akeson, D. Branton, J.J. Kasianowicz, E. Brandin, and D.W. Deamer. Microsecond Time-Scale Discrimination Among Polycytidylic Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments Within Single RNA Molecules. *Biophysical Journal*, 77(6):3227–3233, 1999.
- [3] B. Alberts. DNA replication and recombination. *Nature*, 421:431–435, 2003.
- [4] T. Ambjornsson, SP Apell, Z. Konkoli, EA Di Marzio, and JJ Kasianowicz. Charged polymer membrane translocation. *Journal of Chemical Physics*, 117(8):4063–4073, 2002.
- [5] H. Bayley and P.S. Cremer. Stochastic sensors inspired by biology. *Nature*, 413:226–230, 2001.
- [6] S.M. Bezrukov and J.J. Kasianowicz. Current noise reveals protonation kinetics and number of ionizable sites in an open protein ion channel. *Physical Review Letters*, 70(15):2352–2355, 1993.
- [7] S. Bhakdi and J. Tranum-Jensen. Alpha-toxin of Staphylococcus aureus. *Microbiology and Molecular Biology Reviews*, 55(4):733–751, 1991.
- [8] V.A. Bloomfield, D.M. Crothers, and I. Tinoco. *Nucleic Acids: Structures Properties and Functions*. University Science Books, sausalito, CA, 2000.
- [9] D.J. Bonthuis, J.S. Zhang, B. Hornblower, J. Mathe, B.I. Shklovskii, and A Meller. Self-energy-limited ion transport in subnanometer channels. *Physical Review Letters*, 97(12):128104, 2006.
- [10] T.Z. Butler, J.H. Gundlach, and M.A. Troll. Determination of RNA Orientation during Translocation through a Biological Nanopore. *Biophysical Journal*, 90(1):190–199, 2006.
- [11] I. Chen, P.J. Christie, and D. Dubnau. The Ins and Outs of DNA Transfer in Bacteria. *Science*, 310(5753):1456–1460, 2005.

- [12] D.W. Deamer and M. Akeson. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol*, 18(4):147–151, 2000.
- [13] D.W. Deamer and D. Branton. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res*, 35(10):817–825, 2002.
- [14] D.W. Deamer, H. Olsen, M.A. Akeson, and J.J. Kasianowicz. Mechanism of ionic current blockades during polymer transport through pores of nanometer dimensions. In *Structure and dynamics of confined polymers*, pages 165–176. Kluwer Academic Publishers, 2002.
- [15] V.S. DeGuzman, C.C. Lee, D.W. Deamer, and W.A. Vercountere. Sequence-dependent gating of an ion channel by DNA hairpin molecules. *Nucleic Acids Research*, 34(22):6425–6437, 2006.
- [16] M.N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette. Stretching Single Stranded DNA, a Model Polyelectrolyte. *Physical Review Letters*, 89(24):248102, 2002.
- [17] E.A. Di Marzio and A.J. Mandell. Phase transition behavior of a linear macromolecule threading a membrane. *The Journal of Chemical Physics*, 107:5510, 1997.
- [18] Y. Enokizono, Y. Konishi, K. Nagata, K. Ouhashi, S. Uesugi, F. Ishikawa, and M. Katahira. Structure of hnRNP D Complexed with Single-stranded Telomere DNA and Unfolding of the Quadruplex by Heterogeneous Nuclear Ribonucleoprotein D. *Journal of Biological Chemistry*, 280(19):18862–18870, 2005.
- [19] M. Faller, M. Niederweis, and G.E. Schulz. The Structure of a Mycobacterial Outer-Membrane Channel. *Science*, 303(2):451–464, 2004.
- [20] F. X. Gomis-Ruth and M. Coll. Cut and move: protein machinery for DNA processing in bacterial conjugation. *Curr Opin Struct Biol*, 16(6):744–752, 2006.
- [21] M.E. Gracheva, A. Aksimentiev, and J.P. Leburton. Electrical signatures of single-stranded DNA with single base mutations in a nanopore capacitor. *Nanotechnology*, 17(13):3160–3165, 2006.
- [22] OP Hamill, A. Marty, E. Neher, B. Sakmann, and FJ Sigworth. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflugers Archiv European Journal of Physiology*, 391(2):85–100, 1981.
- [23] C.C. Harrell, Y. Choi, L.P. Horne, L.A. Baker, Z.S. Siwy, and C.R. Martin. Resistive-Pulse DNA Detection with a Conical Nanopore Sensor. *Langmuir*, 22(25):10837–10843, 2006.

- [24] C. Heinz, H. Engelhardt, and M. Niederweis. The Core of the Tetrameric Mycobacterial Porin MspA Is an Extremely Stable β -Sheet Domain. *Journal of Biological Chemistry*, 278(10):8678–8685, 2003.
- [25] J.B. Heng, C. Ho, T. Kim, R. Timp, A. Aksimentiev, Y.V. Grinkova, S. Sligar, K. Schulten, and G. Timp. Sizing DNA Using a Nanometer-Diameter Pore. *Biophysical Journal*, 87(4):2905–2911, 2004.
- [26] S.E. Henrickson, M. Misakian, B. Robertson, and J.J. Kasianowicz. Driven DNA Transport into an Asymmetric Nanometer-Scale Pore. *Physical Review Letters*, 85(14):3057–3060, 2000.
- [27] A. Hildebrand, M. Pohl, and S. Bhakdi. Staphylococcus aureus alpha-toxin: Dual mechanism of binding to target cells. *Journal of Biological Chemistry*, 266(26):17195–17200, 1991.
- [28] ST Howard and TF Byrd. The rapidly growing mycobacteria: saprophytes and parasites. *Microbes Infect*, 2(15):1845–53, 2000.
- [29] S. Howorka and H. Bayley. Probing Distance and Electrical Potential within a Protein Pore with Tethered DNA. *Biophysical Journal*, 83(6):3202–3210, 2002.
- [30] S. Howorka, S. Cheley, and H. Bayley. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature Biotechnology*, 19:636–639, 2001.
- [31] S. Howorka, L. Movileanu, O. Braha, and H. Bayley. Kinetics of duplex formation for individual DNA strands within a single protein nanopore. *Proceedings of the National Academy of Sciences*, 98(23):12996–13001, 2001.
- [32] Y. Jung, S. Cheley, O. Braha, and H. Bayley. The Internal Cavity of the Staphylococcal R-Hemolysin Pore Accommodates 175 Exogenous Amino Acid Residues. *Biochemistry*, 44(25):8919–8929, 2005.
- [33] JJ Kasianowicz and SM Bezrukov. Protonation dynamics of the alpha-toxin ion channel from spectral analysis of pH-dependent current fluctuations. *Biophysical Journal*, 69(1):94–105, 1995.
- [34] J.J. Kasianowicz, E. Brandin, D. Branton, and D.W. Deamer. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences*, 93(24):13770–13773, 1996.
- [35] J.J. Kasianowicz, S.E. Henrickson, M. Misakian, H.H. Weetall, B. Robertson, and V. Stanford. Physics of DNA threading through a nanometer pore and applications to simultaneous multianalyte sensing. In *Structure and Dynamics of Confined Polymers*, pages 141–163. Kluwer Academic Publishers, 2002.

- [36] J.J. Kasianowicz, S.E. Henrickson, H.H. Weetall, and B. Robertson. Simultaneous multianalyte detection with a nanometer-scale pore. *Anal. Chem.*, 73(10):2268–2272, 2001.
- [37] U.F. Keyser, B.N. Koeleman, S. Van Dorp, D. Krapf, R.M.M. Smeets, S.G. Lemay, N.H. Dekker, and C. Dekker. Direct force measurements on DNA in a solid-state nanopore. *Nature Physics*, 2(7):473–477, 2006.
- [38] M.J. Kim, M. Wanunu, D.C. Bell, and A. Meller. Rapid Fabrication of Uniformly Sized Nanopores and Nanopore Arrays for Parallel DNA Analysis. *Adv. Materials*, 18:3149–3153, 2006.
- [39] C.Y. Kong and M. Muthukumar. Modeling of polynucleotide translocation through protein pores and nanotubes. *Electrophoresis*, 23(16):2697–2703, 2002.
- [40] J. Lagerqvist, M. Zwolak, and M. Di Ventra. Fast DNA Sequencing via transverse electronic transport. *Nano letters(Print)*, 6(4):779–782, 2006.
- [41] L. Letellier, P. Boulanger, M. de Frutos, and P. Jacquot. Channeling phage DNA through membranes: from in vivo to in vitro. *Res. Microbiol.*, 154:283–287, 2003.
- [42] RA Levis and JL Rae. Low-noise patch-clamp techniques. *Methods Enzymol.*, 293:218–66, 1998.
- [43] J. Li, M. Gershow, D. Stein, E. Brandin, and JA Golovchenko. DNA molecules and configurations in a solid-state nanopore microscope. *Nature Materials*, 2(9):611–615, 2003.
- [44] D.K. Lubensky and D.R. Nelson. Driven Polymer Translocation Through a Narrow Pore. *Biophysical Journal*, 77(4):1824–1838, 1999.
- [45] M. Mahfoud, S. Sukumaran, P. Hulsman, K. Grieger, and M. Niederweis. Topology of the Porin MspA in the Outer Membrane of *Mycobacterium smegmatis*. *Journal of Biological Chemistry*, 281(9):5908, 2006.
- [46] J. Mathe, A. Aksimentiev, D.R. Nelson, K. Schulten, and A. Meller. Orientation discrimination of single-stranded DNA inside the α -hemolysin membrane channel. *Proceedings of the National Academy of Sciences*, 102(35):12377–12382, 2005.
- [47] J. Mathe, A. Arinstein, Y. Rabin, and A. Meller. Equilibrium and irreversible unzipping of DNA in a nanopore. *Europhysics Letters*, 73(1):128–134, 2006.
- [48] J. Mathe, H. Visram, V. Viasnoff, Y. Rabin, and A. Meller. Nanopore Unzipping of Individual DNA Hairpin Molecules. *Biophysical Journal*, 87(5):3205–3212, 2004.

- [49] S. Matysiak, A. Montesi, M. Pasquali, A.B. Kolomeisky, and C. Clementi. Dynamics of Polymer Translocation through Nanopores: Theory Meets Experiment. *Physical Review Letters*, 96(11):118103, 2006.
- [50] A. Meller and D. Branton. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis*, 23(16):2583–2591, 2002.
- [51] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton. Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences*, 97(3):1079–1084, 2000.
- [52] A. Meller, L. Nivon, and D. Branton. Voltage-Driven DNA Translocations through a Nanopore. *Physical Review Letters*, 86(15):3435–3438, 2001.
- [53] G. Menestrina. Ionic channels formed by *Staphylococcus aureus* alpha-toxin: Voltage-dependent inhibition by divalent and trivalent cations. *Journal of Membrane Biology*, 90(2):177–190, 1986.
- [54] P.G. Merzlyak, M.F.P. Capistrano, A. Valeva, J.J. Kasianowicz, and O.V. Krasilnikov. Conductance and Ion Selectivity of a Mesoscopic Protein Nanopore Probed with Cysteine Scanning Mutagenesis. *Biophysical Journal*, 89(5):3059–3070, 2005.
- [55] H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression*. Oxford University Press New York, 2004.
- [56] L. Mervilleau, J.P. Schmittschmitt, J.M. Scholtz, and H. Bayley. Interactions of Peptides with a Protein Pore. *Biophysical Journal*, 89(2):1030–1045, 2005.
- [57] M. Muthukumar. Polymer translocation through a hole. *The Journal of Chemical Physics*, 111:10371, 1999.
- [58] M. Muthukumar. Translocation of a Confined Polymer through a Hole. *Physical Review Letters*, 86(14):3188–3191, 2001.
- [59] M. Muthukumar. Theory of sequence effects on DNA translocation through proteins and nanopores. *Electrophoresis*, 23(10):1417–1420, 2002.
- [60] M. Muthukumar and CY Kong. Simulation of polymer translocation through protein channels. *Proceedings of the National Academy of Sciences*, 103(14):5273–5278, 2006.
- [61] J. Nakane. Personal Communication, 2006.
- [62] J. Nakane, M. Akeson, and A. Marziali. Evaluation of nanopores as candidates for electronic analyte detection. *Electrophoresis*, 23(16):2592–601, 2002.

- [63] M. Niederweis. Mycobacterial porins- new channel proteins in unique outer membranes. *Molecular Microbiology*, 49(5):1167–1177, 2003.
- [64] M. Niederweis, S. Ehrt, C. Heinz, U. Kloecker, S. Karosi, K.M. Swiderek, L.W. Riley, and R. Benz. Cloning of the mspA gene encoding a porin from Mycobacterium smegmatis. *Molecular Microbiology*, 33(5):933–945, 1999.
- [65] A.S. Petrov and S.C. Harvey. Structural and Thermodynamic Principles of Viral Packaging. *Structure*, 15:21–27, 2007.
- [66] L. Ponchon, S. Mangenot, P. Boulanger, and L. Letellier. Encapsulation and transfer of phage DNA into host cells: From in vivo to single particles studies. *Biochimica et biophysica acta(G). General subjects*, 1724(3):255–261, 2005.
- [67] W. Saenger. *Principles of nucleic acid structure*. Springer, 1984.
- [68] B. Sakmann and E. Neher, editors. *Single-Channel Recording*. Plenum Press, 1983.
- [69] A.F. Sauer-Budge, J.A. Nyamwanda, D.K. Lubensky, and D. Branton. Unzipping Kinetics of Double-Stranded DNA in a Nanopore. *Physical Review Letters*, 90(23):238101, 2003.
- [70] G. Schatz and B. Dobberstein. Common principles of protein translocation across membranes. *Science*, 271(5255):1519–1526, 1996.
- [71] Y. Seol, G.M. Skinner, and K. Visscher. Elastic Properties of a Single-Stranded Charged Homopolymeric Ribonucleotide. *Physical Review Letters*, 93(11):118102, 2004.
- [72] S.B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795–799, 1996.
- [73] L. Song, M.R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J.E. Gouaux. Structure of Staphylococcal alpha-Hemolysin, a Heptameric Transmembrane Pore. *Science*, 274(5294):1859, 1996.
- [74] C. Stahl, S. Kubetzko, I. Kaps, S. Seeber, H. Engelhardt, and M. Niederweis. MspA provides the main hydrophilic pathway through the cell wall of Mycobacterium smegmatis. *Molecular Microbiology*, 40(2):451–464, 2001.
- [75] AJ Storm, JH Chen, HW Zandbergen, and C. Dekker. Translocation of double-strand DNA through a silicon oxide nanopore. *Physical Review E*, 71(5):51903, 2005.

- [76] W. Sung and P.J. Park. Polymer Translocation through a Pore in a Membrane. *Physical Review Letters*, 77(4):783–786, 1996.
- [77] B. Tinland, A. Pluen, J. Sturm, and G. Weill. Persistence length of single-stranded DNA. *Macromolecules*, 30(19):5763–5765, 1997.
- [78] C. Tropini and A. Marziali. Multi-Nanopore Force Spectroscopy for DNA Analysis. *Biophysical Journal*, 92(5):1632, 2007.
- [79] W. Vercoutere, S. Winters-Hilt, H. Olsen, D. Deamer, D. Haussler, and M. Akeson. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nature Biotechnology*, 19:248–252, 2001.
- [80] W.A. Vercoutere, S. Winters-Hilt, V.S. DeGuzman, D. Deamer, S.E. Ridino, J.T. Rodgers, H.E. Olsen, A. Marziali, and M. Akeson. Discrimination among individual Watson–Crick base pairs at the termini of single DNA hairpin molecules. *Nucleic Acids Research*, 31(4):1311–1318, 2003.
- [81] B. Walker and H. Bayley. Key Residues for Membrane Binding, Oligomerization, and Pore Forming Activity of Staphylococcal alpha-Hemolysin Identified by Cysteine Scanning Mutagenesis and Targeted Chemical Modification. *Journal of Biological Chemistry*, 270(39):23065, 1995.
- [82] H. Wang, J.E. Dunning, A.P.H. Huang, J.A. Nyamwanda, and D. Branton. DNA heterogeneity and phosphorylation unveiled by single-molecule electrophoresis. *Proceedings of the National Academy of Sciences*, 101(37):13472–13477, 2004.
- [83] S. Winters-Hilt, W. Vercoutere, V.S. DeGuzman, D. Deamer, M. Akeson, and D. Haussler. Highly Accurate Classification of Watson-Crick Basepairs on Termini of Single DNA Molecules. *Biophysical Journal*, 84(2):967–976, 2003.
- [84] A.M. Woodhull. Ionic Blockage of Sodium Channels in Nerve. *The Journal of General Physiology*, 61(6):687–708, 1973.
- [85] J. Zhang and B.I. Shklovskii. Effective charge and free energy of DNA inside an ion channel. *Physical Review E*, 75(2):21906, 2007.
- [86] J.R. Zupan and P. Zambryski. Transfer of T-DNA from *Agrobacterium* to the plant cell. *Plant Physiology*, 107(4):1041–1047, 1995.

VITA

Thomas Butler

Education

University of Washington

Seattle, WA

PhD in Physics

May 2007

Thesis Advisor: Professor Jens H. Gundlach

Thesis Title: Nanopore Analysis of Nucleic Acids

MS in Physics

September 2002

Northern Arizona University

Flagstaff, AZ

Merged BS in Physics and Mathematics

May 2001

Honors and Awards

Weiss Travel Prize, University of Washington Physics Department

May, 2006

Nanotechnology Fellow, NSF IGERT Program

2002-2004

Karrer Prize, University of Washington Physics Department

May 2002

Outstanding Graduating Senior, Northern Arizona University

May 2001

Presentations

Poster. 51st Biophysical Society Annual Meeting.

March 7, 2007

Talk. Biosensing with channels conference. Bremen, Germany

August 3, 2006

Talk. University of British Columbia. Vancouver, Canada

February 3, 2006

Talk. Simon Fraser University. Vancouver, Canada

March 27, 2006

Poster. 50th Biophysical Society Annual Meeting. February 22, 2005

Poster. 49th Biophysical Society Annual Meeting. February 15, 2004

Peer Reviewed Publications

T.Z. Butler, J.H. Gundlach and M.A. Troll. Determination of RNA orientation during translocation through a biological nanopore. *Biophysical Journal*, 90(1):190-199

T.Z. Butler, J.H. Gundlach and M.A. Troll. Ionic current blockades from DNA and RNA molecules in the alpha-hemolysin nanopore. Submitted to *Biophysical Journal*