UCLA
UCLA Electronic Theses and Dissertations

Title
Improved Detection of Native Conformation Proteins with Hydrogel Interfaced Nanopores

Permalink
https://escholarship.org/uc/item/9rx7h0fk

Author
Acharya, Shiv Jagat

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Improved Detection of Native Conformation Proteins with Hydrogel Interfaced Nanopores

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Bioengineering

by

Shiv Jagat Acharya

2019

© Copyright by

Shiv Jagat Acharya

2019

ABSTRACT OF THE DISSERTATION

Improved Detection of Native Conformation

Proteins with Hydrogel Interfaced Nanopores

by

Shiv Jagat Acharya

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2019

Professor Jacob J. Schmidt, Chair

Nanopores are powerful sensors capable of directly measuring the physical characteristics of single molecules through electrical measurement. Native proteins are particularly interesting targets because of their relevance in medicine and biology. However, theoretical and experimental work has shown that proteins translocate through pores too rapidly to be resolved accurately, resulting in misleading attenuated blockade amplitudes, reduced event frequencies, and incomplete dwell time distributions. Here, it is shown that a hydrogel placed on the distal side of a nanopore can sufficiently increase the residence time of proteins to be resolved accurately by typical measurement apparatus. Measurements of proteins with hydrogel backed nanopores are shown to have event frequencies that exceeded theoretical predictions at concentrations multiple orders of magnitude below those used in typical protein studies, dwell time distributions that extend to the millisecond time scale, and event amplitudes that correctly predict the volume of proteins. It is

demonstrated that IgG and BSA can be measured simultaneously in mixture and current blockades can inform on the shape of a protein as a result of the improved detection. Analytical and computational models are used in conjunction with experimental data to investigate the mechanism of capture, indicating that IgG and BSA do not penetrate the hydrogel, but instead escape by diffusion back into the *cis* chamber. The residence time is also shown to be much more sensitive to the diffusivity and electrophoretic mobility of the protein as a result of this capture mechanism. Lastly, a method to analyze nanopore measurements is presented using hidden Markov models and a modified Canny edge detection algorithm. The method addresses difficulty in analyzing data that shows significant flicker or capacitive noise or has long duration events, such as data produced by hydrogel facilitated detection of protein. The methods shown here serve to introduce interfacing hydrogels with solid state nanopores as an effective, amenable, and robust approach to addressing issues in protein detection related to short dwell times.

The dissertation of Shiv Jagat Acharya is approved.

Gerard C.L. Wong

Andrea M. Kasko

Harold G. Monbouquette

Jacob J. Schmidt, Committee Chair

University of California, Los Angeles

2019

Table of Contents

Introduction 1

1.1 Nanopores as single-molecule sensors 1

1.1.1 Principles of nanopore sensing 1

1.1.2 Biological nanopores and solid state nanopores 4

1.1.3 Single molecule measurements with nanopores 5

1.2 Measurement of native conformation proteins with nanopores 6

1.2.1 Value in measurement of native conformation proteins 7

1.2.2 Measurements of native proteins with solid state nanopores 8

1.3 Inaccurate measurements of native proteins due to limited residence time 9

1.3.1 Inadequate residence time of proteins in nanopores 9

1.3.2 Slowing down nucleic acid translocation through nanopores 13

1.4 Addressing the rapid translocation speed of native proteins 14

1.4.1 Tethering proteins to lipid bilayer coatings on nanopores 15

1.4.2 High bandwidth measurement of proteins 16

1.4.3 Drawbacks to previous methods and use of a hydrogel to increase residence time 17

Improved measurement of proteins using a solid state nanopore coupled with a hydrogel 30

2.1 Introduction 30

2.2 Results and Discussion 32

2.2.1 Interfacing hydrogels with nanopores and potential effect on protein detection 32

2.2.2 Hydrogels on the distal side of the nanopore greatly improve protein detection 34

2.2.3 Measurement of a two protein mixture 42

2.2.4 Shape information from hydrogel enhanced measurements 47

2.3 Conclusion 52

2.4 Materials and Methods 54

2.4.1 Nanopore fabrication 55

2.4.2 Hydrogel preparation 56

2.4.3 Measurement of nanopores 57

2.4.4 Data analysis 57

The mechanism of capture and study of dwell time for hydrogel facilitated detection of proteins 62

3.1 Introduction 62

3.2 Results and Discussion 64

3.2.1 Analytical models for dwell time distribution 64

3.2.2 COMSOL models for dwell time distribution 67

3.2.3 Relationship between voltage and dwell time 74

3.2.4 Increased dwell time sensitivity to diffusivity and electrophoretic mobility 77

3.2.5 Protein-pore interaction and effect on dwell time 82

3.3 Conclusion 84

3.4 Materials and Methods 85

3.4.1 COMSOL model 85

3.4.2 Formation of Nanopores 87

3.4.3 Measurement of Protein 88

Nanopore measurement data analysis method based on Hidden Markov models and Canny edge detection 93

4.1 Introduction 93

4.2 Results and Discussion 98

4.2.1 Common methods of nanopore data analysis 98

4.2.2 Hidden Markov modeling 102

4.2.3 Canny edge detection and baseline reconstruction 106

4.2.4 Hidden Markov modeling with edge detected baseline correction 113

4.3 Conclusion 120

4.4 Materials and Methods 121

4.4.1 Data analysis 121

4.4.2 Data collection 122

Future Directions 126

5.1 Summary 126

5.2 Future Work 127

5.2.1 Exploring limits of hydrogel facilitated detection of proteins 127

5.2.2 Novel directions in protein detection with hydrogel backed nanopores 131

5.2.3 Development of data analysis methods 134

5.3 Concluding remarks 135

Preface

This work discussed in this dissertation focuses on interfacing hydrogels with solid-state nanopores to address the poor detection of proteins in conventional, solid-state nanopore measurement. This project was my primary research pursuit during the graduate program. However, I would like to bring notice to other work that is not discussed in this dissertation. This additional work consists primarily of artificial lipid bilayer research, including published research on the reconstitution of mammalian ion channels in artificial lipid bilayers, the study of lipid oxidation on membrane permeability, and the use of a high-throughput artificial bilayer system to study cell membrane-nanoparticle interactions. These works have been omitted here to keep the contents of this dissertation concise and cohesive.

Acknowledgements

Above all, I would like to thank my advisor, Prof. Jacob Schmidt, for his invaluable guidance and support. His insight, suggestions, and mentorship drove this project forward and more importantly, made me a better scientist and person. I would also like to thank my committee members, Prof. Gerard C.L. Wong, Prof. Andrea M. Kasko, and Porf. Harold G. Monbouquette, for their useful suggestions and recommendations. I would also like to thank members of the Schmidt Research Group who contributed to my research–Reyhaneh Nazarian, Eric Lee, Tyler Smith, Bin Lu, Shayson Edwards, Chenyu (Ann) Jiang, Chance Kuo, Anthony Ma, Peter Lymberopoulos, Katherine Li, Christopher Toh, Andre Petersen, Brian Siegel, and Wilson Xie. I would also like to thank Ahmad El-Arabi, Carl Salazar, Rob Tan, and Shiva Portonovo, who first trained me in the Schmidt Lab and are now lifelong friends.

I would like to thank my parents, Ishani Acharya and Jagat Acharya, whose support and love has never waivered and helped keep me greatly as I worked toward completing this degree.

Lastly, I would like to acknowledge contributions towards the work in this dissertation. Chapter 2 is a version of Acharya, S., Jiang, C., Kuo, C., Nararian, R., Li, K., Ma, A., Siegal, B., Schmidt, J. Improved measurement of proteins using a solid state nanopore coupled with a hydrogel. *In preparation for publication* (2019). All authors except J. Schmidt contributed to data collection and analysis. Chapter 3 is a version of Acharya, S., Jiang, C., Schmidt, J. The mechanism of capture and study of dwell time for hydrogel facilitated detection of proteins. *In preparation for publication* (2019). C. Jiang contributed to data collection. Chapter 4 is a version of Acharya, S., Kuo, C., Schmidt, J. Hidden Markov models and Canny edge detection for parsing noisy nanopore measurement data. *In preparation for publication* (2019). C. Kuo aided in algorithm development and testing. In the above manuscripts, J. Schmidt was the principle investigator.

Vita

2012 B.S. University of California, Los Angeles

Bioengineering

Publications

Acharya, S., Lu, B., Edwards, S., Toh, C., Petersen, A., Yong, C., Lyu, P., Huang, A. and Schmidt, J., 2018. Disruption of artificial lipid bilayers in the presence of transition metal oxide and rare earth metal oxide nanoparticles. *Journal of Physics D: Applied Physics*, *52*(4), p.044002.

Runas, K.A., Acharya, S.J., Schmidt, J.J. and Malmstadt, N., 2016. Addition of cleaved tail fragments during lipid oxidation stabilizes membrane permeability behavior. *Langmuir*, *32*(3), pp.779-786.

Vijayvergiya, V., Acharya, S., Wilson, S.P. and Schmidt, J.J., 2015. Measurement of Ensemble TRPV1 Ion Channel Currents Using Droplet Bilayers. *PloS one*, *10*(10), p.e0141366.

Acharya, S., Edwards, S. and Schmidt, J., 2015. Research highlights: nanopore protein detection and analysis. *Lab on a Chip*, *15*(17), pp.3424-3427.

Vijayvergiya, V., Acharya, S., Poulos, J. and Schmidt, J., 2015. Single channel and ensemble hERG conductance measured in droplet bilayers. *Biomedical microdevices*, *17*(1), pp.1-7.

Ferl, G.Z., O'Connor, J.P., Parker, G.J., Carano, R.A., Acharya, S.J., Jayson, G.C. and Port, R.E., 2015. Mixed-effects modeling of clinical DCE-MRI data: application to colorectal liver metastases treated with bevacizumab. *Journal of Magnetic Resonance Imaging*, *41*(1), pp.132-141.

Acharya, S.A., Portman, A., Salazar, C.S. and Schmidt, J.J., 2013. Hydrogel-stabilized droplet bilayers for high speed solution exchange. *Scientific reports*, *3*, p.3139

Chapter 1

Introduction

Nanopores are powerful sensors for single molecule measurements and have gained significant attention for a wide variety of biosensing applications. Their high sensitivity at the nanometer scale coupled with direct electrical measurement has motivated a range of studies, from small molecule detection to nucleic acid sequencing. Proteins in native conformation are a highly interesting target analyte given their medical relevance but are difficult to measure with nanopores due to the short duration of time they spend inside the nanopore. This chapter provides a background on nanopores with an emphasis on the measurement of proteins in native conformation and contextualizes the significance of the work shown in this dissertation, the use of a hydrogel to increase the residence time of proteins in nanopores.

1.1 Nanopores as single-molecule sensors

1.1.1 Principles of nanopore sensing

Nanopores are smaller scaled relatives to coulter counters,[1] which are used for the counting and characterization of micron and submicron sized particles. In the case of both nanopores and coulter counters, a small hole in an insulating material joins two conductive solutions, typically aqueous salt buffers. A voltage is applied across the pore, and the resulting measured current has a magnitude related to the resistance of the pore, which in turn is related the pore's length and diameter. The effective resistance of the pore can be calculated with the equation for the resistance of a cylinder with an adjustment to account for electric field convergence at the ends of the

nanopore. This access resistance was found by Hall to be equal to $\rho/4r_{pore}$,[2] where ρ is the resistivity of the electrolyte and r_{pore} is the radius of the pore, giving a total pore resistance of

$$R = R_{pore} + 2R_{access} = \frac{\rho l_{pore}}{\pi {r_{pore}}^2} + \frac{\rho}{2r_{pore}}$$

where l_{pore} is the length of the pore. The equation shows that longer and smaller radius pores will have greater resistance and that the relative effect of access resistance is more significant when the pore radius is similar to or greater than the length of the pore.

When a non-conductive particle enters the pore (Fig 1.1), the effective resistance of the pore increases, reducing the measured current. Extension of Maxwell's derivation, which found the effective resistivity of a suspension of insulating spheres,[3] has led to an expression for the change in current when a particle enters the pore,[4–7]

$$\Delta I = -\frac{V_{app}\delta\gamma}{\rho\left(l_{pore} + 0.8d_{pore}\right)^2} S(d_{part}, d_{pore})$$

where V_{app} is the applied voltage across the pore, δ is the volume of the particle, γ is the shape factor of the particle[8–10] (1.5 for spheres), d_{pore} is the diameter of the pore, and $S(d_{part},d_{pore})$ is a correction factor that is relevant as the diameter of the particle approaches the diameter of the pore. The equation shows that the change in current is directly related to the volume and shape of the translocating particle. The equation also shows, due to the dependence of ΔI on pore length and diameter, that detection of nanometer-scale particles (single molecules) becomes more feasible if measured with a nanometer-scale pore.

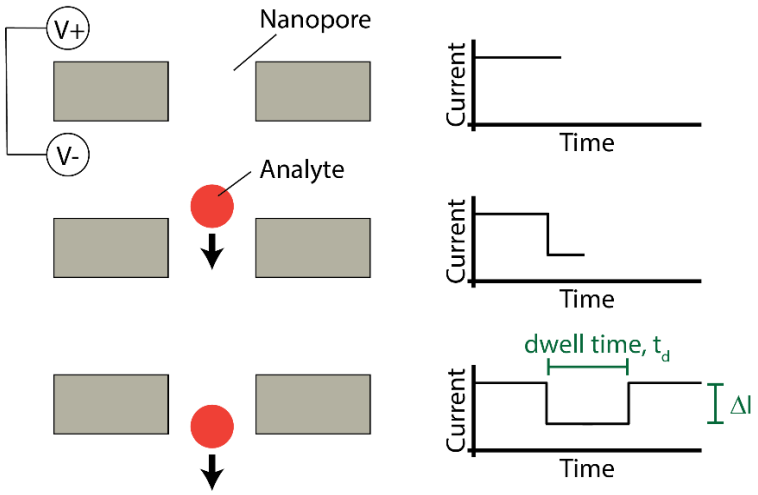


Fig 1.1 In typical experiments, a nanopore separates two electrolyte solutions. A voltage is applied across the pore and the resultant current, dependent on the size of the pore and the conductivity of electrolyte, is measured (top). As a particle enters the pore, the current reduces according to the size of the particle. Once the particle exits the pore, the current returns to the open-pore current level. In many studies, a translocation event is characterized by the amplitude (ΔI) of the blockade current as well as the dwell time of the particle in the pore, given by the duration of the blockade current.

Blockade currents are also often quantified by their duration, indicating the dwell time of the particle inside the pore. The dwell time of the particle is affected by diffusion, electrophoresis if the particle is charged, and electroosmotic flow if the pore is charged. A probability distribution of dwell times accounting only for diffusion and electrophoresis can derived from the Fokker-Planck equation,[11–13]

$$P_{dwell}(t) = \frac{l_p}{t\sqrt{4\pi D t}} \exp\left(-\frac{\left(l_{pore} - F_E \mu_e t\right)^2}{4Dt}\right)$$

where D is the diffusivity of the particle, F_E is the electrophoretic force on a charged particle by the electric field, and μ_e is the electrophoretic mobility of the particle. The equation can be adjusted

to account for electroosmotic flow by scaling the $F_E \mu_e$ term, but otherwise does not account for other effects on the dwell time, such as particle-pore interaction. This model is discussed in greater detail in Chapter 3 for measurement of proteins,

1.1.2 Biological nanopores and solid state nanopores

Most nanopore studies use pores that can be categorized as biological or solid-state nanopores. The first nanopores used for particle sensing were ion channels reconstituted into artificial lipid bilayers. Study of ion channels in biological and artificial lipid membranes by measurement of the ionic current passing through them was established prior to the first nanopore measurements.[14,15] However, a lipid membrane, which is electrically insulating, and an ion channel that could remain open for extended times could be exploited as a nanoscale coulter counter. α-Hemolysin, a pore forming, cytotoxic agent produced by Staphylococcus aureus was first used as a nanopore in 1996 by Kasianowicz *et al.* to measure single molecules of single stranded DNA and RNA.[16] Since then, α-Hemolysin as well as other ion channels have been exploited for nanopore sensing experiments.

Biological pores have several characteristics that make them attractive as pore sensors. Ion channels have small diameters (α-Hemolysin has sensing region diameter of 1.4 nm[17]), and these dimensions are highly reproducible because they are determined by the folding structure of the proteins. Biological pores can also be engineered with molecular biology techniques, such as mutagenesis of the amino acid sequence, to tailor sensing capability for specific applications.[18] Conversely, the fixed size of ion channels can also be regarded as a drawback as sensing larger molecules requires larger pores, and the measurement signal will be largest when the size of the pore nearly matches the size of the particle. More importantly, biological pores, and specifically

the lipid membranes that contain the channels, have limited stability and longevity, especially regarding external factors such as measurement solution, mechanical stress, temperature, and application of large voltages. As a result, there has been much work done to improve the stability of lipid bilayers, specifically addressing mechanical stress and longevity.[19–27]

The drawbacks of biological pores are addressed by solid-state pores, which have controllable pore dimensions and significantly greater robustness when compared to biological pores. Nanopores formed in thin, Silicon Nitride (SiN) membranes are commonly used alternatives to biological nanopores. Formation of a solid-state nanopore was first demonstrated in 2001 by Li *et al.* who formed a nanopore in a SiN membrane with a custom Argon ion beam tool.[28] Subsequent work has shown the fabrication of nanopores in similar ways using standard focused ion beam (FIB) and transmission electron microscope (TEM) apparatus.[29–31] More recently, controlled dielectric breakdown was shown to be a more accessible method to form nanopores in thin SiN membranes, with requirement only for a voltage supply and current monitoring apparatus to form controlled diameter nanopores.[32–35]

1.1.3 Single molecule measurements with nanopores

Nanopores have been used to study a wide variety of molecules, though the primary focus of most research in the field has been sequencing nucleic acids with nanopores. From the first paper using a nanopore to detect nucleic acids,[16] nanopore-based nucleic acid sequencing has been a prominent goal as nanopore sequencing would involve low cost, direct, single molecule measurement of nucleic acid sequence requiring no amplification and low sample volume. For the majority of the past two decades, sequencing with nanopores was primarily hindered by the rapid translocation of nucleic acids through nanopores, which did not allow the individual bases to reside

in the sensing region of the nanopore long enough to produce resolved, discriminable current amplitudes.[36,37] Although many different approaches were taken to address the translocation speed of DNA through a nanopore,[38–43] an effective method was not discovered until 2012 when it was shown that a DNA polymerase could be used to control the rate of DNA translocation through a nanopore. Since then, Oxford Nanopore Technologies has produced commercial, biological nanopore-based sequencers, and a large number of studies have emerged the company's commercialized devices.[44–50]

In addition to nucleic acid sequencing, nanopores have been used a range of other applications. A number of studies have performed sequence specific detection of nucleic acids,[51–55] detection of epigenetic markers,[56–60] and the study of DNA structures such as hairpins, dumbells, and knots.[61–65] Nanopores have also been used for the direct and indirect measurement of other types of molecules, such as non-nucleic acid polymers,[66,67] proteins,[6,68–75] and small molecules and ions.[76–84] The breadth of target analytes and the diversity in mechanisms of detection exemplify the value and capabilities of nanopore sensors.

1.2 Measurement of native conformation proteins with nanopores

Proteins are highly interesting targets for detection and measurement due to their relevance in medicine in biology, and in particular, proteins in native conformation are well suited to being detected by solid-state nanopores. Like nucleic acids, there has been much interest in the detection and study of proteins with nanopores, and it is an ongoing research area that is drawing significant interest.

1.2.1 Value in measurement of native conformation proteins

Proteins have a diverse array of functions, including providing structure, facilitating biochemical reactions, aiding in immune response, signaling, and transport. Proteins are related to nearly all biological function, and conversely, learning about proteins or protein expression can inform on biological function. Presence, absence, or mutation of proteins from cell to organism level has been shown to indicate a number of diseases, making them excellent biomarkers.[84-94] In relation to this, proteome analysis is of significant research interest for its relevance to detection of disease and biological research.[96–105] Proteins are also common drug targets, and sensors that can inform on protein properties and protein interactions can benefit drug development and discovery.[106,107]

The relevance of proteins in medicine, biology, and research makes them highly valuable targets for biosensors. Nanopores are particularly interesting sensors for these applications because they can probe proteins at the single-molecule level. As a result, sample quantity can be kept very low. This advantage is especially important for single-cell proteomics, which is an exciting and developing field in medical and biological research.[108–110] Nanopores also directly probe information on protein structure and physical characteristics, which is relevant for the study of proteins in native conformation.

Unlike nucleic acids which are most interesting for their primary sequence of bases (and epigenetic modifications), the secondary, tertiary, and quaternary structure of proteins are unique to each protein and inform significantly on protein function. It may be possible to measure the primary amino acid sequence of a protein with a nanopore, but the difficulty in discriminating between the twenty unique amino acids is daunting in comparison to discrimination between four nucleic acid bases. Nivala *et al.* used a method similar to that used in nanopore nucleic acid

sequencing to investigate the primary structure of proteins, finding that measured currents did not inform on the sequence directly, but current patterns could be matched to those of previously measured protein domains.[111,112] Kolomogorov translocated denatured protein through subnanometer diameter pores and found that measured currents correlated to the size of amino acid quadramers, though the authors suggested, similarly to Nivala's work, that the blockade current spectra were best analyzed by matching to previously generated libraries.

In contrast to identifying proteins by primary sequence, measurement of folded proteins is equally or more informative than probing the primary amino acid sequence of the protein. Solid-state nanopores can measure the physical characteristics of large single molecules and make excellent sensors for proteins in folded, native conformations. An additional benefit to measuring proteins in native conformations is that protein folding is essential for protein function. As a result, nanopore measurement of native proteins can potentially probe both physical characteristics of proteins as well as protein activity and interaction with substrates.

1.2.2 Measurements of native proteins with solid state nanopores

Solid state nanopores have been used to measure native proteins numerous times previously. A wide variety of proteins in native conformation have been directly measured through plain solid state nanopores, including silicon nitride, hafnium oxide, and glass capillary nanopores.[6,68,71,75,113–116] Proteins have also had facilitated detection using other mechanisms for capture, such as DNA carriers and aptamer binding sites.[70,72,117,118] Protein denaturation and unfolding due to chemical, electrostatic, or other means has also been shown to be measurable with nanopores.[12,119–125] Measurements have also shown current blockades affected by protein-protein interaction and protein interaction with other substrates.[6,126,127] Although the above list is

not exhaustive, it serves to demonstrate the breadth of studies measuring proteins with solid state nanopores.

1.3 Inaccurate measurements of native proteins due to limited residence time

Although sensing of proteins is highly interesting in nanopore research, conventional measurements deviate from theoretical values significantly. The lack of congruency between conventional nanopore measurements and theory has hindered progress in measurement of native proteins.

1.3.1 Inadequate residence time of proteins in nanopores

Most previous measurements of native proteins deviated from theoretical expectations, with models significantly failing to predict measured protein amplitudes, dwell times, and event frequencies.[114] The Smoluchowski equation can be used to predict theoretical event rates for particle translocation through nanopores based on diffusion flux,

$$J = 2\pi cDr_{part}$$

where c is the concentration of the particle, D is the diffusivity of the particle, and r_{part} is the radius of the particle. Plesa *et al*. summarized results from five previous studies as well as their own work, including 12 different proteins in total, and showed that for these experiments, measurements of native proteins with solid-state nanopores never exhibited event translocation rates that met or exceeded the theoretical rate (Fig 1.2).[114] This finding is particularly astonishing given that nucleic acid measurements all exhibited observed event rates greater than the theoretical rate, while some protein measurements fell 3-5 orders of magnitude below the theoretical event rate. In most experimentation, the applied voltage should increase the rate of translocating proteins due

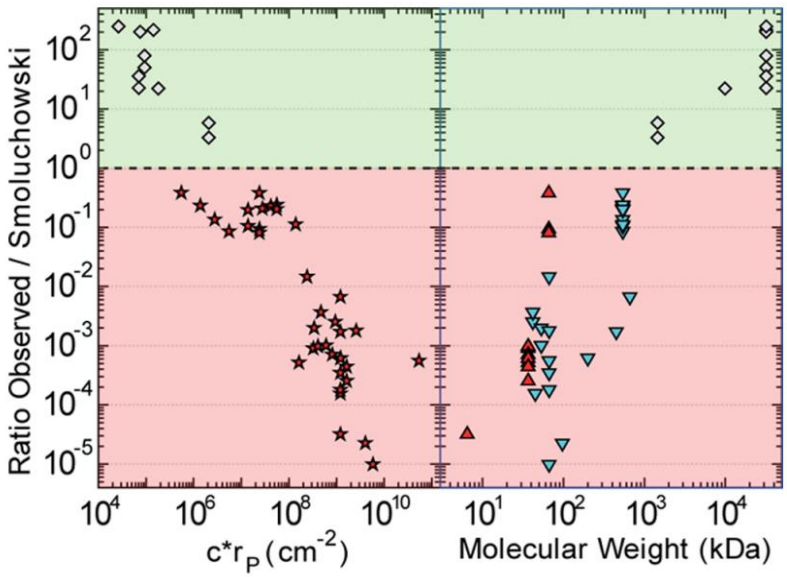


Fig 1.2 Plesa *et al.* summarized the measured event rate of native proteins through solid state nanopores from their own data and five previous studies and compared this event rate to a theoretical, diffusion-based event rate calculated from the Smoluchowski equation. The ratio of the observed rate to the theoretical rate is plotted above for these experiments as well as for translocation of double stranded DNA (dsDNA) against the product of the concentration of protein and the radius of the nanopore (red stars – proteins, white diamonds –dsDNA). Ratios of observed rates to theoretical rates were all less than one for proteins (in some cases three to five orders of magnitude below one), while dsDNA measurements all had ratios greater than one. The same ratios are plotted against the molecular weight of the molecules, showing some correlation between the size of the protein and the ratio of observed rate to theoretical rate, indicating the measured event rates deviate more dramatically for smaller molecules (red triangles – positively charged protein, green triangles – negatively charged proteins, white diamonds – dsDNA). Adapted with permission from Fast Translocation of Proteins through Solid State Nanopores.[114] Copyright 2013 American Chemical Society.

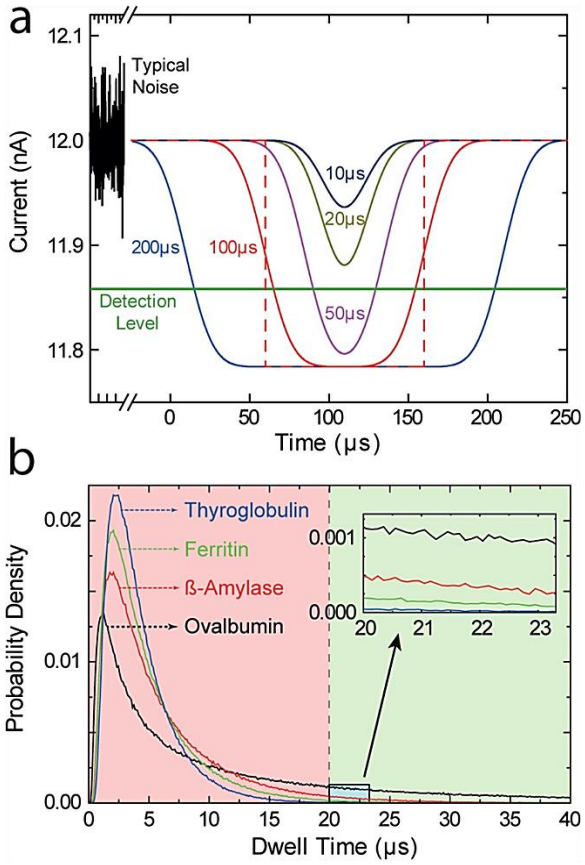


Fig 1.3 (A) The effect of a 10 kHz Gaussian low pass filter on events of durations from 10-200 µs is shown. Idealized events before filter all had the same amplitude (an example of a 100 µs duration event before filtering is shown as a dashed red line). For events shorter than 66 µs (twice the rise time of the filter), the observed current amplitude is less than the actual amplitude of the event. In the example shown, events with durations of 20 µs and below have amplitudes that are attenuated to the point that they fall below the detection threshold and cannot be distinguished from normal baseline noise. These events would fully evade detection. (B) Simulation of the dwell time distribution for four proteins shows that almost all translocation events for native proteins will have durations of 1-10 µs. For a given detection threshold of 20 µs, which is an appropriate value for many nanopore experiments, only a small fraction of events have durations long enough to be detected, and a smaller fraction will last long enough to produce accurate amplitudes due to filtering. Adapted with permission from Fast Translocation of Proteins through Solid State Nanopores.[114] Copyright 2013 American Chemical Society.

to electrophoresis, and therefore, the theoretical rates used for reference here, which only account for diffusion, are still conservative estimates of the event frequency. A preliminary clue from this analysis can be found in the relationship between the size of proteins and the ratio of the observed event rate to the theoretical event rate, showing that measurement of larger proteins exhibits event rates that are closer to the theoretically predicted rates.

A well supported theory proposed by Plesa *et al.* is that proteins in native conformation translocate too quickly through nanopore to be resolved by standard measurement apparatus.[114,128] The Axopatch 200B amplifier is commonly used in many nanopore experiments and has a measurement bandwidth of approximately 52 kHz.[129] Experiments using this measurement apparatus and otherwise also typically include low pass filters to reduce high frequency capacitive noise. Although these filters can help signal-to-noise ratio, they can also reduce the amplitude of short duration events. Fig 1.3A shows the effects of a 10 kHz low pass Gaussian filter on events ranging from 10-200 μs, demonstrating that events shorter than 66 μs (twice the rise time of the filter) have attenuated observed amplitudes and increased observed dwell times. For the example given, events with durations of 20 μs or less have amplitudes that are reduced sufficiently that current blockades cannot be discriminated from the baseline noise, and as a result, these events evade detection entirely. A dwell time distribution model based on the Fokker-Planck equation, as introduced earlier in Chapter 1, shows that most proteins that translocate through nanopores are likely to reside in the pore for 1-10 μs, and that is highly unlikely that a protein resides in the pore for longer than the 20 μs detection limit (Fig 1.3B) Of those events that are detected, most are likely short enough in duration to be affected by the low pass filter, resulting in reduced measured amplitudes. The congruency of the evidence leads to a simple conclusion that proteins translocate too quickly through nanopores to be accurately resolved in conventional measurement.

The effects of this fast translocation significantly hinder further work on native protein measurement with nanopores. The size of proteins cannot be measured due to misleading attenuated amplitudes. The observed event rate does not correlate linearly with the concentration of the protein.[114] Observed dwell time distributions are only a small fraction of the full distribution, which is predominantly below the detection threshold. These three metrics are integral to nanopore sensing and all deviate misleadingly from models in conventional sensing of native proteins.

1.3.2 Slowing down nucleic acid translocation through nanopores

The issue of fast translocation of particles through nanopores is not unique to proteins, and in fact, was the primary obstacle to nanopore nucleic acid sequencing. Although nucleic acids could be reliably detected as they translocated through nanopores, the time that individual bases resided in the sensing region of nanopores was too short to leave resolvable and distinct current blockade amplitudes. Approaches used to slow down nucleic acids may be insightful when addressing the fast translocation of proteins. A wide variety of methods were attempted to slow down the translocation of nucleic acids. Decreasing applied voltage, reducing the temperature, or increasing the viscosity of the measurement solution all increased the residence time of DNA translocating through a nanopore, but these methods also reduced the amplitude of the measurement signal, negating the benefits of long dwell time.[130] Changing the salt to Lithium Chloride was shown to reduce the speed of translocating DNA by an order of magnitude, likely because Lithium cations can bind more closely to DNA than conventionally used Potassium and Sodium ions, reducing the effective charge of DNA.[131] This method may also be effective for proteins, but the effect is likely much less dramatic for proteins because proteins are typically significantly less charged than DNA, which has one charge per monomer, and have three

dimensional, globular structure, providing fewer binding sites for counter ions in comparison to one dimensional, linear nucleic acids. In addition to these fundamental methods to slow nucleic acids down, a number of other methods were demonstrated to reduce the speed of translocating nucleic acids,[43,132–138] and these methods have varying degrees of relevance to the detection of native proteins.

The method that first enabled nucleic acid sequencing was the use of a DNA polymerase to control the speed of translocation through a nanopore.[139] Using the enzyme's kinetics to control the rate at which the DNA ratcheted through the nanopore, nanopore measurements could achieve single base resolution. A similar method was implemented for protein measurement, coupling a protein to an unfoldase enzyme, but the method produced arbitrary current signatures that needed to be matched to previous measurements to identify protein domains.[111,112] The method may be effective at identifying proteins given a large enough library of previously measured protein domains, but does not inform on the protein's physical properties and inherently cannot be used for the study of proteins in native conformation.

1.4 Addressing the rapid translocation speed of native proteins

The short dwell times of proteins hinders further progress in using nanopores as tools for the measurement of native proteins. A method to resolve nanopore measurements of native proteins may open avenues towards new research. Previous work has addressed the fast translocation of proteins, and the improved measurements have led to exciting results such as the determination of protein shape, dipole moment, and rotational diffusion constant through nanopore measurements. However, these techniques to improve protein measurement have significant drawbacks in

implementation, and a need still exists for an effective and accessible method to address the fast translocation of proteins.

1.4.1 Tethering proteins to lipid bilayer coatings on nanopores

Yusko *et al.* demonstrated an innovate method to reduce the speed of translocating proteins through nanopores using a coating on the nanopore.[72] By coating the nanopore in a lipid bilayer and tethering proteins with long chemical linkages to lipids in the bilayer, the movement of proteins in free solution was limited by the viscosity of the lipid bilayer rather than the viscosity of the bulk solution. Because the mobility of lipids in a bilayer is much lower than that of proteins in solution, the speed of translocating proteins is significantly reduced as well. Using this technique, proteins can translocate slowly enough through the nanopore to be resolved with standard measurement apparatus. As a second benefit, proteins are less likely to interact with the nanopore surface, reducing the likelihood that the proteins irreversibly clog the nanopore by adsorption. Adsorption of proteins onto solid-state nanopore surfaces is a well studied phenomenon, and is a frequent mode of failure for native protein measurements.[126,127]

In a follow up study, Yusko *et al.* also showed that the increased dwell time of translocation events allowed for measurement of a protein's approximate ellipsoidal shape, dipole moment, and rotational diffusion constant.[140] In most studies, a detected particle is treated as spherical when calculating the volume from measured currents. However, because some proteins are non-spherical, blockade currents will change as a result of the orientation of the protein. Given adequate measurement bandwidth and long enough residence times, the amplitude distributions of protein current blockades can indicate a protein's ellipsoidal shape, its dipole moment, and rotational

diffusion constant. The result is particularly significant because it is the first study to show shape and orientation dependent measurements at the single molecule level using the coulter principle.

1.4.2 High bandwidth measurement of proteins

An alternative to reducing the speed of translocating proteins is increasing the measurement bandwidth. Larkin *et al.* used a high bandwidth amplifier to measure proteins through Silicon Nitride and Halfnium Oxide nanopores at acquisition rates over 4 MHz.[113] The study also exploited the use of nanopores with diameters only slightly larger than the dimensions of proteins to increase the interaction between the protein and pore and therefore, the dwell time. With the high bandwidth amplifier and tailored-diameter nanopores, measurements of proteins showed amplitudes that matched well with models and full dwell time distributions that were not cut off by the detection threshold.

In a very recent study, it was shown that by using a low capacitance solid-state nanopore chip to reduce capacitive noise, measurement bandwidth could be increased sufficiently to also allow determination of protein ellipsoidal shape and dipole moments.[141] This method also showed that the rotational diffusion constants of free protein were higher than that of proteins tethered to lipids by long molecular chains. The higher-bandwidth measurement used in this study led to measurement of tens to hundreds of events greater than 150 μs in duration, the cutoff suggested by the authors needed to probe shape dependent information. This study is of particular relevance to the work presented in this dissertation as it is the first study to demonstrate accurate measurement of protein shape and volume with a label free technique. The study is discussed further in Chapter 2.

1.4.3 Drawbacks to previous methods and use of a hydrogel to increase residence time

Though previous methods of addressing the short dwell time of proteins are effective and can enable accurate measurement of proteins, they also have significant limitations. Tethering proteins to lipid bilayers requires prior knowledge of the target protein and incorporation of lipids with specific linkages into the bilayer. As a result, the method only has value when studying known proteins, and has limited value in studying unknown protein solutions. Measurement with a high bandwidth amplifier addresses this issue, but these amplifiers are not commonly used in nanopore research and are an expensive solution. With standard measurement apparatus and a low capacitance measurement setup, resolved protein measurement is possible, but protein shape and dipole moment can be determined only from a small fraction of events with the longest dwell time. Longer events and a greater number of events would improve the capability of nanopore sensors for the measurement of protein shape and dipole moment. Furthermore, because asymmetrical proteins will generate current blockades dependent on their orientation, current amplitude distributions of protein translocations can be broad, and distinguishing proteins in a mixture is difficult without many long duration events. This is discussed further in Chapter 2.

As a result of these barriers to implementation and use, a label-free method to extend the duration that proteins reside in nanopores is highly desirable as a more practical solution to the fast translocation of proteins through nanopores. In the following chapters, it is shown that a hydrogel placed at the distal side of a nanopore, a method previously used to extend nucleic acid dwell times,[137,138,142,143] increases the dwell time of proteins sufficiently to be resolved by standard measurement apparatus. The method is inexpensive, amenable to almost all measurement apparatus and hardware, and is easy to implement while being effective enough to produce current blockades that can distinguish proteins in mixture and indicate protein ellipsoidal shape. In the

following chapters, hydrogel facilitated detection is compared to conventional detection and explored for native protein sensing. The mechanism of capture is also investigated using theoretical models and simulations. Lastly, a novel method to analyze nanopore measurements, especially low signal-to-noise, long dwell time measurements typical in hydrogel facilitated detection, is presented. This analysis method combines the use of the Canny edge detection method[144] and Hidden Markov Models[145] to parse nanopore measurements and identify current blockades. These chapters introduce the use hydrogels with solid-state nanopores as a method to resolve the issues of short dwell times in the measurement of proteins with solid-state nanopores.

1. Coulter, W. H. Means for counting particles suspended in a fluid. (1953).

2. Hall, J. E. Access Resistance of a Small Circular Pore. *J. Gen. Physiol.* **66**, 531–532 (1975).

3. Maxwell, J. C. *A Treatise on Electricity and Magnetism*. (Clarendon Press, 1904).

4. Gregg, E. C. & Steidley, K. D. Electrical counting and sizing of mammalian cells in suspension. *Biophys. J.* **5**, 393–405 (1965).

5. DeBlois, R. W. & Bean, C. P. Counting and sizing of submicron particles by the resistive pulse technique. *Rev. Sci. Instrum.* **41**, 909–916 (1970).

6. Han, A. *et al.* Label-Free Detection of Single Protein Molecules and Protein−Protein Interactions Using Synthetic Nanopores. *Anal. Chem.* **80**, 4651–4658 (2008).

7. Takashi Ito, Li Sun, and & Crooks*, R. M. Simultaneous Determination of the Size and Surface Charge of Individual Nanoparticles Using a Carbon Nanotube-Based Coulter Counter. (2003). doi:10.1021/AC034072V

8. Golibersuch, D. C. Observation of Aspherical Particle Rotation in Poiseuille Flow via the Resistance Pulse Technique: I. Application to Human Erythrocytes. *Biophys. J.* **13**, 265–280 (1973).

9. Hurley, J. Sizing Particles with a Coulter Counter. *Biophys. J.* **10**, 74–79 (1970).

10. Grover, N. B., Naaman, J., Ben-Sasson, S., Doljanski, F. & Nadav, E. Electrical Sizing of Particles in Suspensions: II. Experiments with Rigid Spheres. *Biophys. J.* **9**, 1415–1425 (1969).

11. Cox, D. R. (David R. & Miller, H. D. (Hilton D. *The Theory of Stochastic Processes*. (Methuen, 1965).

12. Talaga, D. S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* **131**, 9287–9297 (2009).

13. Ling, D. Y. & Ling, X. S. On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrödinger's first-passage-time theory. *J. Phys. Condens. Matter* **25**, 375102 (2013).

14. Hille, B. *Ionic Channels of Excitable Membranes*. (Sinauer, 2001).

15. Neher, E. & Sakmann, B. Single-channel currents recorded from membrane of denervated

frog muscle fibres. *Nature* **260**, 799–802 (1976).

16. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. **93**, 13770–13773 (1996).

17. Song, L. *et al.* Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore. *Science (80-.).* **274**, 1859–1865 (1996).

18. Bayley, H., Braha, O., Cheley, S. & Gu, L.-Q. Engineered Nanopores. in *Nanobiotechnology: Concepty Applications and Perspectives* (eds. Niemeyer, C. M. & Mirkin, C. A.) 93–110 (Wiley, 2004).

19. Kawano, R., Osaki, T., Sasaki, H. & Takeuchi, S. A Polymer-Based Nanopore-Integrated Microfluidic Device for Generating Stable Bilayer Lipid Membranes. *Small* **6**, 2100–2104 (2010).

20. Han, X. *et al.* Nanopore Arrays for Stable and Functional Free-Standing Lipid Bilayers. *Adv. Mater.* **19**, 4466–4470 (2007).

21. and, J. W. S. & Gu*, L. Q. Stochastic Sensing on a Modular Chip Containing a Single-Ion Channel. (2007). doi:10.1021/AC0614285

22. Malmstadt, N., Jeon, T.-J. & Schmidt, J. J. Long-Lived Planar Lipid Bilayer Membranes Anchored to an In Situ Polymerized Hydrogel. *Adv. Mater.* **20**, 84–89 (2008).

23. Knoll, W. *et al.* Functional tethered lipid bilayers. *Rev. Mol. Biotechnol.* **74**, 137–158 (2000).

24. Köper, I. Insulating tethered bilayer lipid membranes to study membrane proteins. *Mol. Biosyst.* **3**, 651 (2007).

25. Costello, R. F., Peterson, I. ., Heptinstall, J. & Walton, D. J. Improved gel-protected bilayers. *Biosens. Bioelectron.* **14**, 265–271 (1999).

26. Ide, T. & Yanagida, T. An Artificial Lipid Bilayer Formed on an Agarose-Coated Glass for Simultaneous Electrical and Optical Measurement of Single Ion Channels. *Biochem. Biophys. Res. Commun.* **265**, 595–599 (1999).

27. Xiao-feng Kang, *,†, Stephen Cheley, †,§, Allison C. Rice-Ficht, † and & Hagan Bayley*, ‡. A Storable Encapsulated Bilayer Chip Containing a Single Protein Nanopore. (2007). doi:10.1021/JA068654G

28. Li, J. *et al.* Ion-beam sculpting at nanometre length scales. *Nature* **412**, 166–169 (2001).

29. Cai, Q., Ledden, B., Krueger, E., Golovchenko, J. A. & Li, J. Nanopore sculpting with noble gas ions. *J. Appl. Phys.* **100**, 024914 (2006).

30. Venkatesan, B. M. *et al.* Highly Sensitive, Mechanically Stable Nanopore Sensors for DNA Analysis. *Adv. Mater.* **21**, 2771–2776 (2009).

31. Storm, A. J., Chen, J. H., Ling, X. S., Zandbergen, H. W. & Dekker, C. Fabrication of solid-state nanopores with single-nanometre precision. *Nat. Mater.* **2**, 537–540 (2003).

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

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

34. Yanagi, I., Hamamura, H., Akahori, R. & Takeda, K. I. Two-step breakdown of a SiN membrane for nanopore fabrication: Formation of thin portion and penetration. *Sci. Rep.* **8**, 10129 (2018).

35. Yanagi, I., Akahori, R., Hatano, T. & Takeda, K. Fabricating nanopores with diameters of sub-1 nm to 3 nm using multilevel pulse-voltage injection. *Sci. Rep.* **4**, 5000 (2015).

36. Storm, A. J. *et al.* Fast DNA translocation through a solid-state nanopore. *Nano Lett.* **5**, 1193–1197 (2005).

37. Branton, D. *et al.* The potential and challenges of nanopore sequencing. in *Nanoscience and Technology* **26**, 1146–1153 (Co-Published with Macmillan Publishers Ltd, UK, 2008).

38. Schneider, G. F. *et al.* DNA Translocation through Graphene Nanopores. *Nano Lett.* **10**, 3163–3167 (2010).

39. Derrington, I. M. *et al.* Nanopore DNA sequencing with MspA. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16060–5 (2010).

40. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).

41. Rincon-Restrepo, M., Mikhailova, E., Bayley, H. & Maglia, G. Controlled Translocation of Individual DNA Molecules through Protein Nanopores with Engineered Molecular Brakes. *Nano Lett.* **11**, 746–750 (2011).

42. Shekar, S. *et al.* Measurement of DNA Translocation Dynamics in a Solid-State Nanopore at 100 ns Temporal Resolution. *Nano Lett.* **16**, 4483–4489 (2016).

43. Zhou, D., Zhao, Y., Tian, E. & Wang, D. Slowing Down DNA Translocation Speed thorough a Nanopore by a Nanofibre Meshed Layer. *Biophys. J.* **114**, 180a (2018).

44. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).

45. Loman, N. J. & Watson, M. Successful test launch for nanopore sequencing. *Nat. Methods* **12**, 303–304 (2015).

46. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics. Proteomics Bioinformatics* **14**, 265–279 (2016).

47. Ashton, P. M. *et al.* MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* **33**, 296–300 (2015).

48. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015).

49. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).

50. Eisenstein, M. Oxford Nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.* **30**, 295 (2012).

51. Liu, N. *et al.* Two-Way Nanopore Sensing of Sequence-Specific Oligonucleotides and Small-Molecule Targets in Complex Matrices Using Integrated DNA Supersandwich Structures. *Angew. Chemie Int. Ed.* **52**, 2007–2011 (2013).

52. Wang, Y., Zheng, D., Tan, Q., Wang, M. X. & Gu, L.-Q. Nanopore-based detection of circulating microRNAs in lung cancer patients. *Nat. Nanotechnol.* **6**, 668–674 (2011).

53. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol. 2010 511* **5**, 807 (2010).

54. Singer, A. *et al.* Nanopore Based Sequence Specific Detection of Duplex DNA for Genomic Profiling. *Nano Lett.* **10**, 738–742 (2010).

55. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.* **19**, 636–639 (2001).

56. Wallace, E. V. B. *et al.* Identification of epigenetic DNA modifications with a protein nanopore. *Chem. Commun. (Camb).* **46**, 8195–7 (2010).

57. Shim, J. *et al.* Detection and quantification of methylation in DNA using solid-state nanopores. *Sci. Rep.* **3**, 1389 (2013).

58. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci.* **110**, 18904 LP – 18909 (2013).

59. Ahmed, T., Haraldsen, J. T., Zhu, J.-X. & Balatsky, A. V. Next-Generation Epigenetic Detection Technique: Identifying Methylated Cytosine Using Graphene Nanopore. *J. Phys. Chem. Lett.* **5**, 2601–2607 (2014).

60. Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores Discriminate among Five C5-Cytosine Variants in DNA. *J. Am. Chem. Soc.* **136**, 16582–16587 (2014).

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

62. Comer, J., Dimitrov, V., Zhao, Q., Timp, G. & Aksimentiev, A. Microscopic Mechanics of Hairpin DNA Translocation through Synthetic Nanopores. *Biophys. J.* **96**, 593–608 (2009).

63. Mathé, J., Visram, H., Viasnoff, V., Rabin, Y. & Meller, A. Nanopore Unzipping of Individual DNA Hairpin Molecules. *Biophys. J.* **87**, 3205–3212 (2004).

64. Vercoutere, W. *et al.* Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat. Biotechnol.* **19**, 248–252 (2001).

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

66. Rodrigues, C. G., Machado, D. C., Chevtchenko, S. F. & Krasilnikov, O. V. Mechanism of KCl Enhancement in Detection of Nonionic Polymers by Nanopore Sensors. *Biophys. J.* **95**, 5186–5192 (2008).

67. Baaken, G., Ankri, N., Schuler, A.-K., Rühe, J. & Behrends, J. C. Nanopore-based single-molecule mass spectrometry on a lipid membrane microarray. *ACS Nano* **5**, 8080–8088 (2011).

68. Fologea, D., Ledden, B., McNabb, D. S. & Li, J. Electrical characterization of protein molecules by a solid-state nanopore. *Appl. Phys. Lett.* **91**, 5–8 (2007).

69. Hornblower, B. *et al.* Single-molecule analysis of DNA-protein complexes using nanopores. *Nat. Methods* **4**, 315–317 (2007).

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

71. Firnkes, M., Pedone, D., Knezevic, J., Döblinger, M. & Rant, U. Electrically facilitated translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* **10**, 2162–2167 (2010).

72. Yusko, E. C. *et al.* Controlling protein translocation through nanopores with bio-inspired fluid walls. *Nat. Nanotechnol.* **6**, 253–260 (2011).

73. Rotem, D., Jayasinghe, L., Salichou, M. & Bayley, H. Protein Detection by Nanopores Equipped with Aptamers. *J. Am. Chem. Soc.* **134**, 2781–2787 (2012).

74. Wei, R., Gatterdam, V., Wieneke, R., Tampé, R. & Rant, U. Stochastic sensing of proteins with receptor-modified solid-state nanopores. *Nat. Nanotechnol.* **7**, 257–263 (2012).

75. Li, W. *et al.* Single Protein Molecule Detection by Glass Nanopores. *ACS Nano* **7**, 4129–4134 (2013).

76. Gu, L.-Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686–690 (1999).

77. Braha, O. *et al.* Simultaneous stochastic sensing of divalent metal ions. *Nat. Biotechnol.* **18**, 1005–1007 (2000).

78. Bayley, H., Braha, O. & Gu, L.-Q. Stochastic Sensing with Protein Pores. *Adv. Mater.* **12**, 139–142 (2000).

79. Gu, L.-Q., Cheley, S. & Bayley, H. Prolonged Residence Time of a Noncovalent Molecular Adapter, β-Cyclodextrin, within the Lumen of Mutant α-Hemolysin Pores. *J. Gen. Physiol.* **118**, 481 LP – 494 (2001).

80. Heins, E. A., Siwy, Z. S., Baker, L. A. & Martin, C. R. Detecting Single Porphyrin Molecules in a Conically Shaped Synthetic Nanopore. *Nano Lett.* **5**, 1824–1829 (2005).

81. Arnaut, V., Langecker, M. & Simmel, F. C. Nanopore Force Spectroscopy of Aptamer–Ligand Complexes. *Biophys. J.* **105**, 1199–1207 (2013).

82. Yang, C. *et al.* Highly Sensitive Simultaneous Detection of Lead(II) and Barium(II) with G-Quadruplex DNA in α-Hemolysin Nanopore. *Anal. Chem.* **85**, 7302–7307 (2013).

83. Kawano, R. *et al.* Rapid detection of a cocaine-binding aptamer using biological nanopores on a chip. *J. Am. Chem. Soc.* **133**, 8474–8477 (2011).

84. Wang, G., Wang, L., Han, Y., Zhou, S. & Guan, X. Nanopore detection of copper ions using a polyhistidine probe. *Biosens. Bioelectron.* **53**, 453–458 (2014).

85. Wang, Q. *et al.* Mutant proteins as cancer-specific biomarkers. *Proc. Natl. Acad. Sci.* **108**, 2444 LP – 2449 (2011).

86. Cai, X. D., Golde, T. E. & Younkin, S. G. Release of excess amyloid beta protein from a mutant amyloid beta protein precursor. *Science (80-.).* **259**, 514 LP – 516 (1993).

87. Surinova, S. *et al.* On the Development of Plasma Protein Biomarkers. *J. Proteome Res.* **10**, 5–16 (2011).

88. Wild, E. J. *et al.* Quantification of mutant huntingtin protein in cerebrospinal fluid from Huntington's disease patients. *J. Clin. Invest.* **125**, 1979–1986 (2015).

89. Welsh, J. B. *et al.* Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc. Natl. Acad. Sci.* **100**, 3410 LP – 3415 (2003).

90. Hewitt, S. M., Dear, J. & Star, R. A. Discovery of Protein Biomarkers for Renal Diseases. *J. Am. Soc. Nephrol.* **15**, 1677 LP – 1689 (2004).

91. Joensson, H. N. *et al.* Detection and Analysis of Low-Abundance Cell-Surface Biomarkers Using Enzymatic Amplification in Microfluidic Droplets. *Angew. Chemie Int. Ed.* **48**, 2518–2521 (2009).

92. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).

93. Chikkaveeraiah, B. V., Bhirde, A. A., Morgan, N. Y., Eden, H. S. & Chen, X. Electrochemical Immunosensors for Detection of Cancer Protein Biomarkers. *ACS Nano* **6**, 6546–6561 (2012).

94. Blennow, K. Cerebrospinal Fluid Protein Biomarkers for Alzheimer's Disease. *NeuroRX* **1**, 213–225 (2004).

95. Theodorescu, D. *et al.* Discovery and validation of new protein biomarkers for urothelial

cancer: a prospective analysis. *Lancet Oncol.* **7**, 230–240 (2006).

96. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science (80-.).* **356**, eaal3321 (2017).

97. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci.* **97**, 9390 LP – 9395 (2000).

98. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).

99. Sharma, S. *et al.* A Deep Proteome Analysis Identifies the Complete Secretome as the Functional Unit of Human Cardiac Progenitor Cells. *Circ. Res.* **120**, 816–834 (2017).

100. Hu, S., Loo, J. A. & Wong, D. T. Human body fluid proteome analysis. *Proteomics* **6**, 6326–6353 (2006).

101. Cánovas, F. M. *et al.* Plant proteome analysis. *Proteomics* **4**, 285–298 (2004).

102. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell* **138**, 795–806 (2009).

103. Patton, W. F. Detection technologies in proteome analysis. *J. Chromatogr. B* **771**, 3–31 (2002).

104. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247 (2001).

105. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

106. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat. Rev. Drug Discov.* **5**, 310–321 (2006).

107. Karlsson, R. *et al.* Biosensor Analysis of Drug–Target Interactions: Direct and Competitive Binding Assays for Investigation of Interactions between Thrombin and Thrombin Inhibitors. *Anal. Biochem.* **278**, 1–13 (2000).

108. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

109. Wang, D. & Bodovitz, S. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290 (2010).

110. Irish, J. M., Kotecha, N. & Nolan, G. P. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat. Rev. Cancer* **6**, 146–155 (2006).

111. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* **8**, 12365–12375 (2014).

112. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–50 (2013).

113. Larkin, J., Henley, R. Y., Muthukumar, M., Rosenstein, J. K. & Wanunu, M. High-bandwidth protein analysis using solid-state nanopores. *Biophys. J.* **106**, 696–704 (2014).

114. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).

115. Steinbock, L. J. *et al.* Probing the size of proteins with glass nanopores. *Nanoscale* **6**, 14380–14387 (2014).

116. Japrung, D. *et al.* Single-Molecule Studies of Intrinsically Disordered Proteins Using Solid-State Nanopores. *Anal. Chem.* **85**, 2449–2456 (2013).

117. Rotem, D., Jayasinghe, L., Salichou, M. & Bayley, H. Protein Detection by Nanopores Equipped with Aptamers. *J. Am. Chem. Soc.* **134**, 2781–2787 (2012).

118. Ding, S., Gao, C. & Gu, L.-Q. Capturing Single Molecules of Immunoglobulin and Ricin with an Aptamer-Encoded Glass Nanopore. *Anal. Chem.* **81**, 6649–6655 (2009).

119. Li, J., Fologea, D., Rollings, R. & Ledden, B. Characterization of Protein Unfolding with Solid-state Nanopores. *Protein Pept. Lett.* **21**, 256–265 (2014).

120. Freedman, K. J., Haq, S. R., Edel, J. B., Jemth, P. & Kim, M. J. Single molecule unfolding and stretching of protein domains inside a solid-state nanopore by electric field. *Sci. Rep.* **3**, 1638 (2013).

121. Oukhaled, A. *et al.* Dynamics of completely unfolded and native proteins through solid-state nanopores as a function of electric driving force. *ACS Nano* **5**, 3628–3638 (2011).

122. Pastoriza-Gallego, M. *et al.* Evidence of Unfolded Protein Translocation through a Protein Nanopore. *ACS Nano* **8**, 11350–11360 (2014).

123. Oukhaled, G. *et al.* Unfolding of Proteins and Long Transient Conformations Detected by Single Nanopore Recording. *Phys. Rev. Lett.* **98**, 158101 (2007).

124. Cressiot, B. *et al.* Protein Transport through a Narrow Solid-State Nanopore at High Voltage: Experiments and Theory. *ACS Nano* **6**, 6236–6243 (2012).

125. Freedman, K. J. *et al.* Chemical, Thermal, and Electric Field Induced Unfolding of Single Protein Molecules Studied Using Nanopores. *Anal. Chem.* **83**, 5137–5144 (2011).

126. Balme, S. *et al.* Influence of adsorption on proteins and amyloid detection by silicon nitride nanopore. *Langmuir* **32**, 8916–8925 (2016).

127. Niedzwiecki, D. J. & Movileanu, L. Monitoring Protein Adsorption with Solid-state Nanopores. *J. Vis. Exp.* e3560 (2011). doi:10.3791/3560

128. Pedone, D., Firnkes, M. & Rant, U. Data Analysis of Translocation Events in Nanopore Experiments. *Anal. Chem.* **81**, 9689–9694 (2009).

129. Uram, J. D., Ke, K. & Mayer, M. Noise and Bandwidth of Current Recordings from Submicrometer Pores and Nanopores. *ACS Nano* **2**, 857–872 (2008).

130. Daniel Fologea, †, James Uplinger, †, Brian Thomas, †, David S. McNabb, ‡ and & Jiali Li*, †. Slowing DNA Translocation in a Solid-State Nanopore. (2005). doi:10.1021/NL051063O

131. Kowalczyk, S. W., Wells, D. B., Aksimentiev, A. & Dekker, C. Slowing down DNA Translocation through a Nanopore in Lithium Chloride. *Nano Lett.* **12**, 1038–1044 (2012).

132. Krishnakumar, P. *et al.* Slowing DNA Translocation through a Nanopore Using a Functionalized Electrode. *ACS Nano* **7**, 10319–10326 (2013).

133. Zhang, H. *et al.* Slowing Down DNA Translocation Through Solid-State Nanopores by Pressure. *Small* **9**, 4112–4117 (2013).

134. Larkin, J. *et al.* Slow DNA Transport through Nanopores in Hafnium Oxide Membranes. *ACS Nano* **7**, 10121–10128 (2013).

135. Melnikov, D. V, Leburton, J.-P. & Gracheva, M. E. Slowing down and stretching DNA with an electrically tunable nanopore in a p–n semiconductor membrane. *Nanotechnology* **23**, 255501 (2012).

136. Hyun, C., Kaur, H., Rollings, R., Xiao, M. & Li, J. Threading Immobilized DNA Molecules

through a Solid-State Nanopore at >100 μs per Base Rate. *ACS Nano* **7**, 5892–5900 (2013).

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

138. Al Sulaiman, D., Cadinu, P., Ivanov, A. P., Edel, J. B. & Ladame, S. Chemically Modified Hydrogel-Filled Nanopores: A Tunable Platform for Single-Molecule Sensing. *Nano Lett.* **18**, 6084–6093 (2018).

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

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

141. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* acsnano.8b09555 (2019). doi:10.1021/acsnano.8b09555

142. Tang, Z. *et al.* Gel mesh as "brake" to slow down DNA translocation through solid-state nanopores. *Nanoscale* **7**, 13207–13214 (2015).

143. Squires, A. H., Hersey, J. S., Grinstaff, M. W. & Meller, A. A Nanopore–Nanofiber Mesh Biosensor To Control DNA Translocation. *J. Am. Chem. Soc.* **135**, 16304–16307 (2013).

144. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).

145. Baum, L. E. & Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Stat.* **37**, 1554–1563 (1966).

Chapter 2

Improved measurement of proteins using a solid state nanopore coupled with a hydrogel

2.1 Introduction

The most significant obstacle to measurement of native proteins is their short dwell time in nanopores,[1] which causes incorrect measurement of amplitude, dwell time distribution, and frequency of translocation events when using typical measurement apparatus. Addressing the short dwell time of proteins in nanopores is the first step towards the use of nanopores for applied native protein sensing. This chapter explores the use of a hydrogel placed on the distal side of a nanopore as a method to increase the dwell time of measured proteins and enable accurate measurement of proteins.

Because an extremely small fraction of events will have residence times that are long enough to be resolved, the simplest solution would be extension of experiment durations such that enough of these rare, long duration events are captured to produce adequate statistics. However, solid state nanopore measurements are often terminated due to irreversible clogging of the nanopore, which is more likely as more translocation events occur or as the experiment duration increases. This is particularly problematic with native proteins, as a large number of translocation events occur without being resolved. Protein detection is additionally more difficult at lower concentrations, where the detected event frequency may drop low enough such that non-specific event frequency due to noise or contaminant may interfere with identification of protein events.

Previous nanopore measurements of proteins, especially small to medium sized proteins, often used high protein concentrations, between 0.1-10 μM to observe detection frequencies of 1-10 Hz, which would be needed to ensure that the majority of events are due to the translocation of protein and not non-specific contaminants.[1] These issues with protein detection make it less likely that a nanopore experiment will last long enough before failure to provide accurate statistics. As the needs for applications develop, this limitation will become additionally constricting. For instance, previous work has shown that higher bandwidth measurement can enable the measurement of size, shape, and dipole moment of freely translocating proteins, but the number of candidate events in this study was in the tens to hundreds for a given experiment, limiting the statistics of measurement and the chance to measure multiple proteins in a single experiment.[2] This study also relied on high protein concentrations of 1-10 μM to measure event frequencies of 1-5 Hz, and were not able to produce useful measurements at concentrations below 100 nM. Although measurement bandwidth could be further increased with improved measurement hardware, a method to extend the residence times of nanopores would also address this issue.

This chapter explores an approach to extending the residence time of proteins within the nanopore by interfacing the nanopore on one side with a hydrogel. Previously, hydrogels have been interfaced with nanopores to improve the sensing of nucleic acids, increasing the dwell time by multiple orders of magnitude.[3–6] However, hydrogels have not previously been used in conjunction with solid state nanopores to improve the detection of proteins. Here, it is shown that a hydrogel dramatically increases protein event rates, improves accuracy of measured blockade currents, is label free, and enables effective protein detection for typical measurement bandwidths and nanopores that are not size-specific to the target protein. The hydrogel sterically interferes with protein transport through the nanopore while leaving the electrical measurement substantially

unchanged. We show detection of immunoglobulin G1 (IgG), bovine serum albumin (BSA), and streptavidin at a 10nM concentration with event frequencies orders of magnitude higher than previous reports[1] and also clear discrimination of signals resulting from individual proteins in a mixture. Analysis of individual translocation events may also indicate effects on the detected currents arising from protein shape, as shown previously,[2,7] but without the need for protein-specific labels.

2.2 Results and Discussion

2.2.1 Interfacing hydrogels with nanopores and potential effect on protein detection

We hypothesized that a hydrogel present on the distal side of a nanopore would sterically interact with translocating native proteins to increase their residence time within the nanopore (Fig. 2.1). Hydrogels have previously been used to reduce the speed of nucleic acids translocating through a nanopore,[3–6] with the relationship between mesh size of the hydrogel and the radius of gyration of the nucleic acids directly relating to the resultant dwell time distributions. It was also shown that dwell time distributions could be tuned by changes to the hydrogel's pore size or chemistry.[6]

In studies measuring nucleic acids with hydrogel interfaced nanopores, results indicated that the molecule fully translocated through to the *trans* chamber, but the hydrogel reduced the electrophoretic mobility of the molecule and therefore, the speed of the molecule through the nanopore.[3–6] A similar mechanism may apply to protein translocation (Fig 2.1B). However, given protein tertiary and quaternary structure and the smaller hydrogel pore size used here (3.1 nm), it is also possible that the hydrogel acts as a barrier to complete translocation of proteins through the nanopore. Instead, upon capture, the protein escapes by diffusing against the net

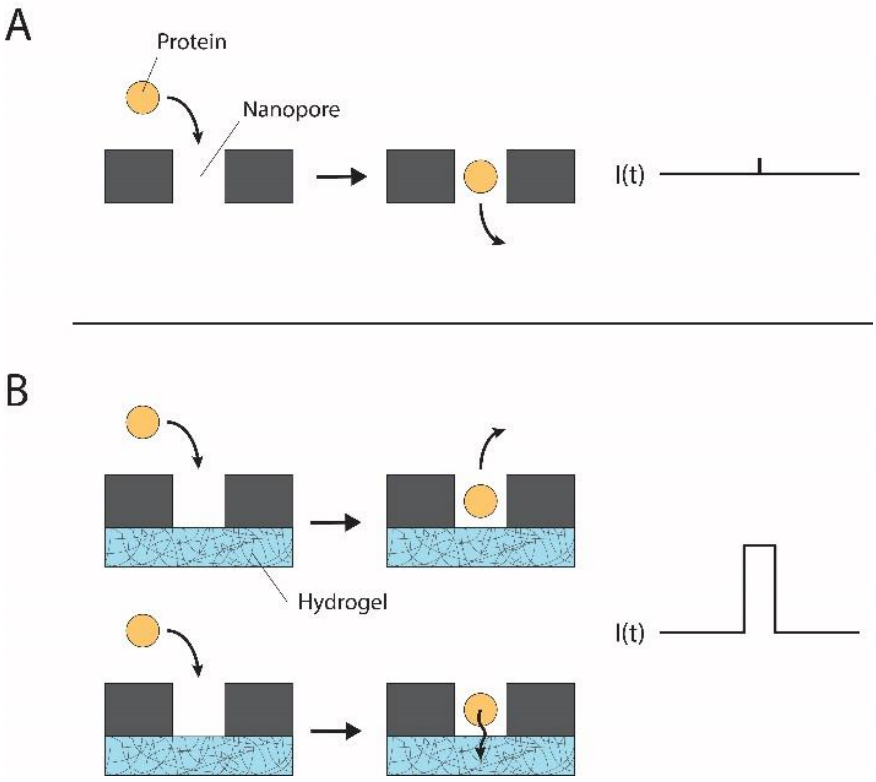


Fig 2.1 In typical nanopore sensing experiments, proteins translocate through the nanopore too quickly to be resolved or accurately measured. Translocation speed as a result of electrophoresis, electroosmotic flow, and diffusion result in dwell times for typical proteins that are predominantly under 10 µs, too short for typical measurement apparatus to resolve at 10-100 kHz bandwidth. As a result, current blockade events are typically very short duration with amplitudes that are misleadingly small. (A). With a hydrogel present on the distal side of the nanopore, proteins are unable to complete translocation into the *trans* compartment of the chamber and are forced to exit the pore to the *cis* side by diffusion (B, top), or the hydrogel decreases the mobility of the protein as it translocates (B, bottom). In either case, the presence of a hydrogel increases the residence time of the protein in the nanopore, increasing the probability that the residence time is sufficiently long to allow detection and accurate resolution of the blockade amplitude of the translocation event.

electrophoretic/electroosmotic drag forces back into the *cis* chamber (Fig 2.1B). Escape by diffusion against electrophoretic force would yield longer residence times than in traditional nanopore sensing, where electrophoresis contributes towards protein escape. In either scenario, we expect the dwell time of translocating proteins to increase. These mechanisms are further explored in Chapter 3.

2.2.2 Hydrogels on the distal side of the nanopore greatly improve protein detection

To explore the hypothesized steric interaction between translocating proteins and the hydrogel, we sought to form a hydrogel with a pore size that was close to or smaller than the size of typical protein analytes. We chose PEG-DMA hydrogels since they are commercially available, have mesh sizes that can be tuned by the PEG chain length, and PEG is a neutrally charged polymer known to minimize protein adsorption and interaction.[8,9] In the work described here, we chose a PEG (1000 MW)-DMA hydrogel which we measured by a method previously shown by Canal and Peppas[10] to have a mean pore size of 3.1 nm, a size that is small enough to sterically interact with many protein species.

Our experiments utilized two aqueous chambers filled with 2M KCl, 10mM Tris-HCl, pH 10 buffer separated by a 15 nm thick SiN membrane. The *trans* chamber also contained PEG(1000MW)-DMA, which was chemically crosslinked. Following hydrogel formation, a nanopore was formed in the SiN membrane using the dielectric breakdown method.[11] After addition of protein to the *cis* side and application of a negative electric potential difference to the nanopore (*cis* minus *trans*), conductance measurements showed events much greater in number and amplitude compared to measurements performed with no hydrogel present (Fig 2.2).

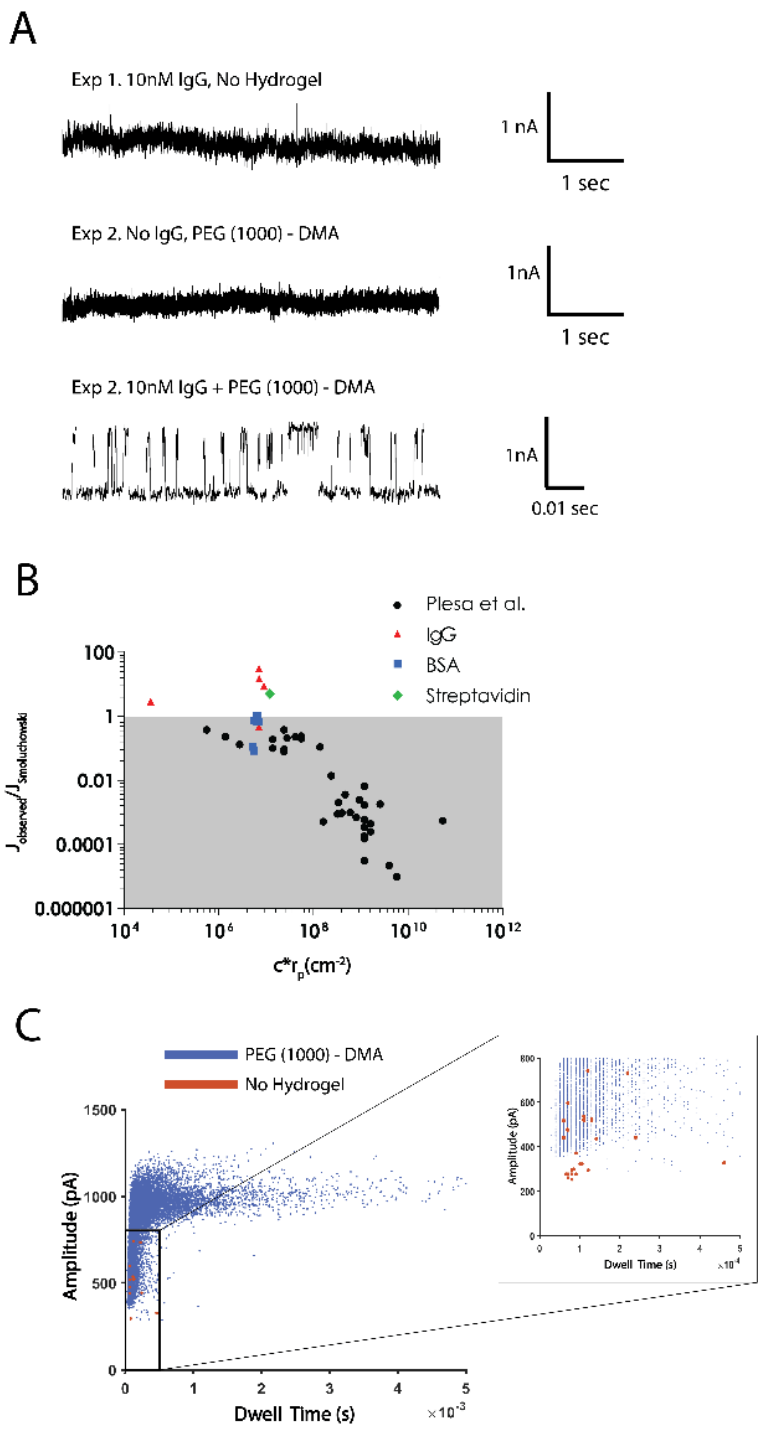


Fig 2.3 Enhanced measurement of proteins with solid state nanopores interfaced with a hydrogel. (A) Example current measurement at -60 mV with 10 nM IgG added to the *cis* side of a ~24 nm diameter nanopore in 2 M KCl, pH 10 buffer on both sides (top). We observed an event detection rate of 0.49 Hz, and a most probable maximum event amplitude of 538±120 pA. Using a second

nanopore of similar size (~24 nm), we repeated the experiment in identical conditions, but after polymerizing a PEG (1000 MW)-DMA hydrogel in the *trans* compartment (bottom; the control hydrogel experiment without protein is shown in the middle trace). With the hydrogel present, the event detection rate increased to 418 Hz with a most probable maximum amplitude of 963±470 pA. (B) Plesa *et al.* calculated the ratio of the rates of experimentally observed translocation ($J_{observed}$) to diffusion-dependent translocation estimated with the Smoluchowski equation ($J_{Smoluchowski}$) from 5 previous published reports and their own collected data of protein detection with synthetic nanopores.[1] These ratios are plotted (black dots) against the product of the protein concentration and the nanopore diameter in each experiment. We have added data from 14 hydrogel-nanopore experiments measuring IgG (red triangles), BSA (blue squares), and streptavidin (green diamond) to this plot. (C) A scatter plot of the dwell time and amplitude of each translocation event detected from 25 seconds of the hydrogel (blue, n=11241) and no-hydrogel control (red, n=24) data is shown in Fig 2A, clearly demonstrating the increased detection frequency, dwell time, and amplitude of protein translocation events when a hydrogel is present on the *trans* side of the nanopore.

Fig. 2.2A (top trace) shows results from a representative experiment using a *cis-trans* potential difference of -60 mV with 10 nM IgG added to the *cis* side of a ~24 nm nanopore in 2 M KCl (pH 10) buffer, with no hydrogel present. We observed an event detection rate of 0.49 Hz, with a most probable blockade current amplitude of 538±120 pA. Using a second nanopore of similar size (~24 nm), we repeated the experiment in identical conditions, but after polymerizing a hydrogel in the *trans* side compartment. With the hydrogel present, the event detection rate increased to 418 Hz with a most probable current amplitude of 963±470 pA (Fig. 2.2A, bottom trace). Current blockade amplitudes of 963 pA would be expected for a spherical molecule with a hydrodynamic diameter of 10.8 nm given our experimental conditions. This value agrees well

previously experimentally measured hydrodynamic diameters for IgG of ~11 nm.[12,13] No events were detected in the presence of the hydrogel when no protein was added (Fig. 2.2A, middle).

The low event frequency of our hydrogel-free control experiment is consistent with the low event rates observed by other groups measuring native, folded proteins with synthetic nanopores. Plesa *et al.* calculated the ratio of the rates of experimentally observed translocation ($J_{observed}$) to diffusion-dependent translocation estimated with the Smoluchowski equation ($J_{Smoluchowski}$) from 5 previous published reports and their own collected data of protein detection with synthetic nanopores.[1] These ratios are plotted in Fig. 2.2B (black dots) against the product of the protein concentration and the nanopore diameter in each experiment. Plesa *et al.* found that for these previous reports, the event rate never met or exceeded the theoretical rate and in most cases was orders of magnitude less.

With a hydrogel present on the distal side of the nanopore, a greater number of translocation events are detected. We have added data from 14 hydrogel-nanopore experiments measuring IgG (red triangles), BSA (blue squares), and Streptavidin (green diamonds) to this plot. For these experiments, 7 exceed the theoretical event rate and of the remaining 7 data points, 6 are within an order of magnitude of the theoretical rate. All 3 proteins have negative charge at pH10 with isoelectric points of ~8.5,[14,15] 4.5-5.0,[16,17] and 5.0-6.0[18] for IgG, BSA, and Streptavidin respectively. Example traces of hydrogel facilitated measurement of BSA and streptavidin are shown in Fig 2.3.

Using the Smoluchowski rate equation to estimate rate of diffusion of particles to the nanopore, we calculated an expected free diffusion (ignoring electrophoresis) translocation rate for 10 nM IgG of approximately 14 events/s for a 24 nm diameter nanopore, much greater than the 0.49 events/sec we saw without the hydrogel. The presence of the hydrogel causes the event

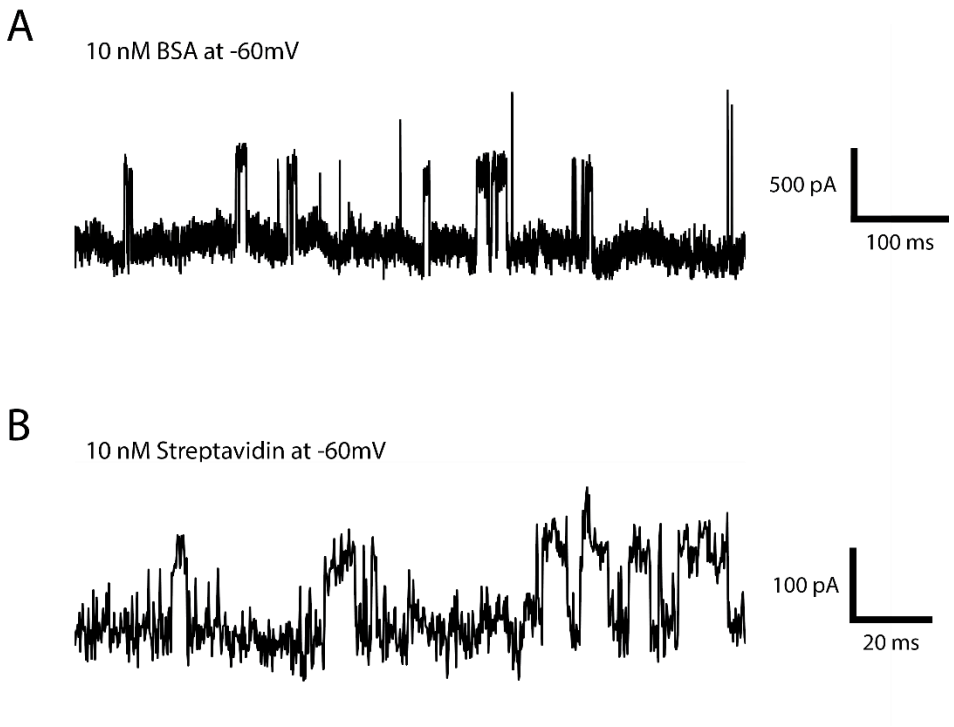


Fig 2.3 Example traces are shown for measurement of BSA (A) and streptavidin (B) with a hydrogel interfaced nanopore. Both measurements were conducted in 2M KCl, pH 10 buffer at -60mV at a protein concentration of 10 nM. The measurements demonstrate the high event frequency observed when measuring protein with hydrogel interfaced nanopores. Data from these measurements are included as data points in Fig 2.2B.

frequency to increase to almost 1000 times the rate of the conventional experiment without a hydrogel and ~17 times greater than the rate expected from diffusion, which we attribute to the electrophoretic enhancement of the capture rate due to IgG's negative charge at pH 10 and the negative applied voltage.

We characterized the detected events by their maximum amplitude and dwell time; a scatter plot of the amplitude vs. dwell time for events captured in 25 seconds of continuous acquisition (Fig. 2.2C) further demonstrates the effect of the hydrogel on the measured reduction in current

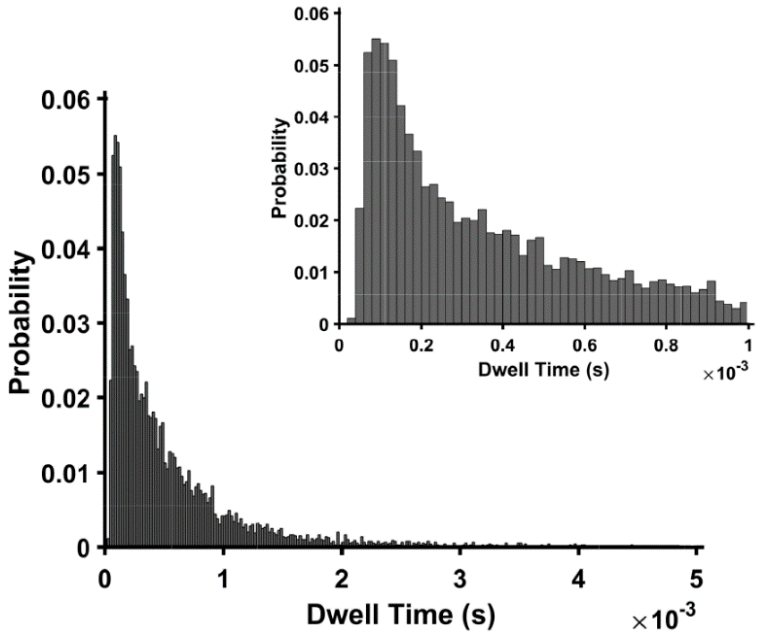


Fig 2.4 Dwell time distribution of data in Figure 2.2C of 10nM IgG detection with a hydrogel interfaced nanopore (n=11241). With a hydrogel on the distal side of the nanopore, event durations have a most probably residence time of approximately 100 µs with an exponentially decreasing residence time probability to the millisecond time scale. Although the 100 µs peak is near the detection threshold at our measurement bandwidth, the significant fraction of events with durations greater than 100 µs is dramatically different from conventional measurement, where almost all detected translocation events would have residence times between 20-50 µs. Inset shows dwell time distribution from 0 to 1 ms.

and its duration. Without a hydrogel present, the number of events is greatly diminished and the majority of the measured dwell times was between 20-50 µs, with a most probable blockade amplitude of 538±120. With the hydrogel present, it is immediately apparent that a much great number of events were detected, but additionally, the population of events extends in amplitude and dwell time. The dwell time distribution is broader than traditional experiments, with dwell times extending to the millisecond time scale (the dwell time distribution is shown in Fig 2.4). The

wide distribution of dwell times is most likely related to the shift of the dwell time distribution to longer times in the presence of a hydrogel. For traditional protein measurement, theoretical models predict dwell time distributions with peaks an order of magnitude below the detection threshold of our measurement apparatus.[1,19] Here, we expect that the presence of a hydrogel has shifted the true dwell time probability peak closer to and possibly above the detection threshold, allowing us to measure a greater fraction of translocation events. Given the large number of events detected with a hydrogel (n=11241 for Fig 2.2C) and the expected increase in residence time due to the presence of the hydrogel, it is also not surprising that some events have residence times that extend to the millisecond time scale.

Previous work coupling hydrogels to nanopores formed by dielectric breakdown showed that yield of experiments was lowered by the presence of the hydrogel, and that nanopores would irreversibly clog after hundreds to thousands of nucleic acid detection events.[3] We found that the yield of useful nanopores was more significantly affected by batch-to-batch variation in SiN windows and the size (typically >20nm) of the nanopore, which is most likely related to the formation of pinholes and additional pores during the dielectric breakdown process.[20] In agreement with previous work,[3] effective measurements were usually associated with nanopores exhibiting linear current to voltage curves. These nanopores were typically lower noise and showed stable open pore baseline currents, although, on average, we found that pores interfaced with PEG-DMA hydrogels exhibited greater noise than pores without hydrogels (Figure 2.5). Measurements were conducted at 100kHz acquisition with a 10kHz hardware filter to aid in noise reduction which is a common measurement condition for solid state nanopore measurements, and we found the amplitude detection threshold for conventional, gel-free experiments was typically ~80-100pA and for gel facilitated measurements it was ~150-200 pA. Experiments typically lasted 2-6 hours before

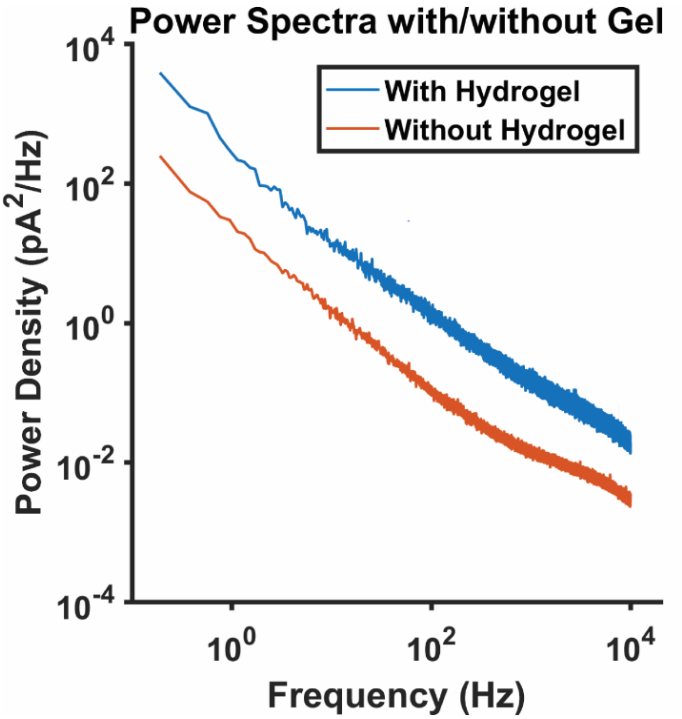


Fig 2.5 Measurements of nanopores with hydrogels typically exhibit more noise than nanopores without hydrogels. The mean power spectra of 5 experiments in which a hydrogel was present on the trans side of the chamber, and 5 experiments with no hydrogel are shown. All experiments were conducted in 1 M KCl, pH 10 at -120 mV. We hypothesize that the increased noise is due to motion of polymer chains immediately adjacent to the nanopore which can cause changes to pore conductance.

becoming irreversibly clogged. During this time, tens of thousands of events could be collected as needed. The high frequency of events and long duration of measurement also suggest that proteins are most likely not able to penetrate the hydrogel as we would expect the mobility of ~5-10nm proteins in a 3.1 nm mesh size gel would be extremely low if proteins penetrate the gel, and as a result, the hydrogel would rapidly clog as molecules accumulate in the mesh.

2.2.3 Measurement of a two protein mixture

The short dwell times of translocating proteins in conventional nanopore measurement can deter measurement of a protein mixture because incorrectly resolved amplitudes hinders event discrimination. The extended residence times of proteins in hydrogel-facilitated nanopore measurements addresses this shortcoming, potentially enabling accurate measurement of solutions containing multiple particle species. To probe the use of hydrogel-backed nanopores for this application, we measured a solution containing a mixture of IgG and BSA. BSA, a 65 kD protein having a hydrodynamic diameter of 6.7 nm,[7] would be expected to block less current than IgG in similarly sized nanopores. In a similar experiment to that described above (Fig. 2.2), we measured 10 nM BSA using a hydrogel-backed, 22 nm diameter nanopore with a -70 mV applied potential (Fig. 2.6, blue data). We observed current blockages predominantly in the range of 200-800 pA (full distribution shown in Fig. 2.7).

We then exchanged the buffer containing BSA for one containing 10 nM IgG, obtaining much larger blockage currents (Fig. 2.6, orange data), with amplitudes in the range of 250-2000 pA. Finally, to this IgG-containing solution, we added a small volume of concentrated BSA to the *cis* chamber to a final concentration of 10 nM without significantly diluting the concentration of IgG (Fig. 2.6, green data). With the BSA present, we saw a dramatic increase in the frequency of events with amplitudes less than 800 pA, while still maintaining measurement of events >1500 pA.

A scatter plot of the dwell times and amplitudes of each event shows two separate sets of data points (Fig. 2.6A) from the measurements of IgG and BSA individually. Data points from the measurement of both proteins overlapped with the distributions of each individual protein, although they were significantly denser in the BSA region of the plot. The scatter plot also shows

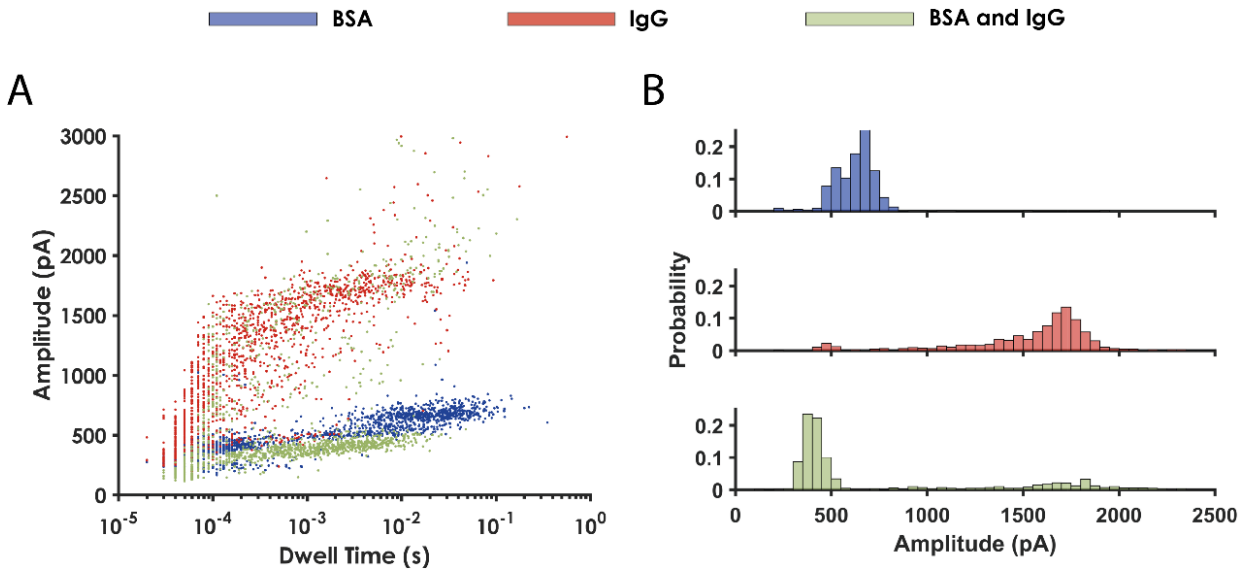


Fig 2.6 Simultaneous measurement of BSA and IgG with a 22 nm nanopore having a PEG (1000)-DMA hydrogel on the *trans* side. Data shown in blue were taken with 10 nM BSA in the *cis* chamber (90 seconds of acquisition); data shown in red were taken when the BSA solution was replaced with 10 nM IgG (100 seconds of acquisition); data shown in green were taken when and BSA were IgG both present at 10 nM (200 seconds of acquisition). (A) Each event was characterized by current amplitude and duration and plotted in a scatter plot. (B) The current data from (A) for all events with duration longer than 400 µs is histogrammed, showing amplitude distributions that clearly correlate with the presence of IgG and/or BSA. All measurements were made with 2 M KCl, pH 10 buffer at -70 mV.

the effect of the 10 kHz low pass filter, with measured current amplitudes significantly decreasing for dwell times below 10^{-4} sec.

The scatter plot also indicates additional narrowing of the distribution of measured currents for dwell times greater than 200 µs, which could be due to protein rotation. This was previously studied by Yusko *et al.*, who found that non-spherically-symmetric proteins blocked varying amounts of current in a nanopore depending on their orientation, and that, for IgG, events with dwell times over 400 µs were sufficiently long for the protein to permute through all possible

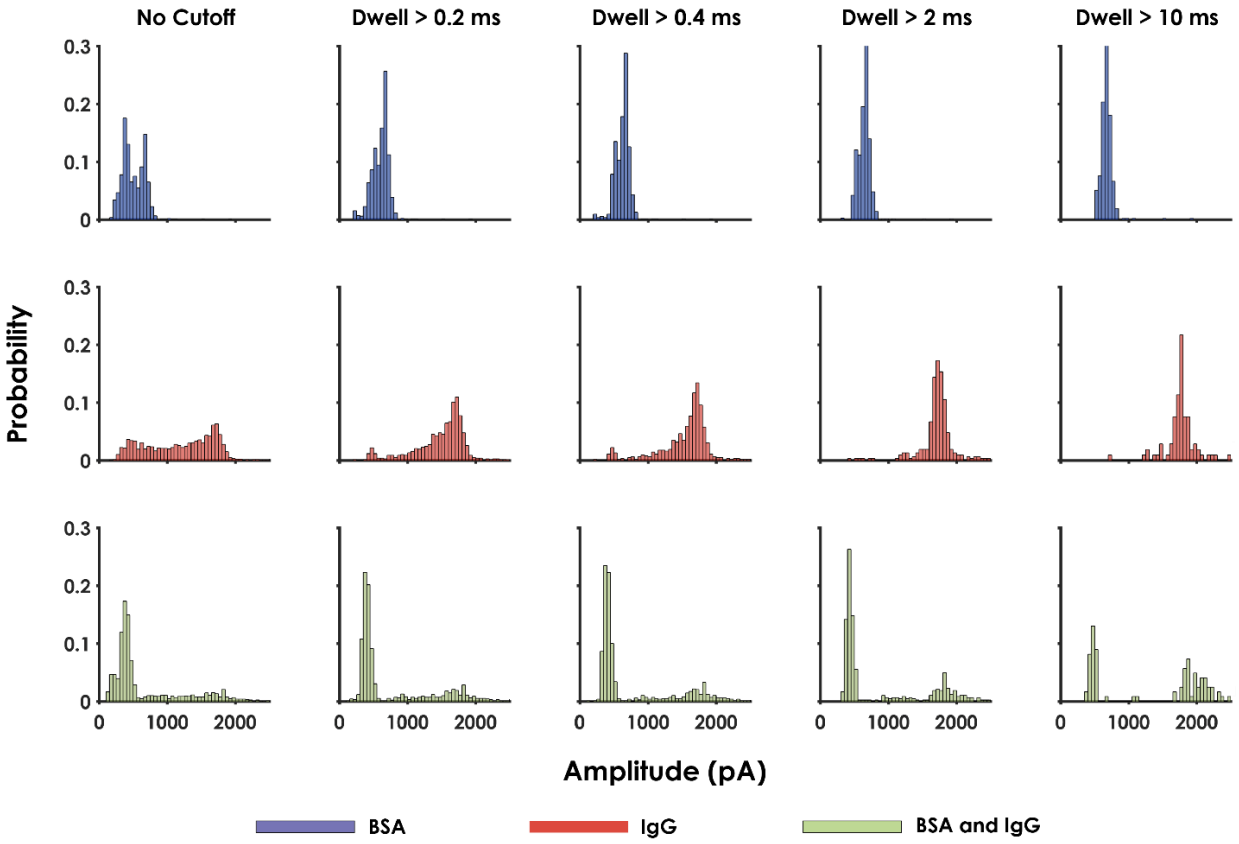


Fig 2.7 Amplitude histograms for data shown in Fig 2.6 with different dwell time cutoffs show the importance of long dwell time measurements for separating proteins in mixture. 10nM BSA (blue), IgG (red), and a mixture of the two proteins (green) were measured sequentially with a hydrogel backed, 22 nm diameter nanopore at -70mV in 2M KCl pH 10 buffer. The left-most column shows a histogram of the maximum current amplitude of each event for these measurements. Histograms of both separate BSA (n=1725) and IgG (n=1343) measurements show unique histograms with distributions that span different current ranges, but the distributions overlap significantly. The histogram of the mixture (n=1428) is unique compared to histograms of the separately measured proteins, but the distribution overlap makes it impossible to confidently discern BSA from IgG. Removing all events with dwell times less than 0.2 ms (2nd column from the left) should leave events that are unattenuated by the 10 kHz low pass filter used during data acquisition, and the related histograms show a greater difference between separately measured BSA (n=991) and IgG (n=774). Furthermore, the histogram of the mixture (n=1022) now exhibits a more bimodal distribution. Yusko *et al.* found that IgG events longer than 0.4 ms (middle column) were long enough to allow

measurement of shape dependent effects from single events.[7] Removing events shorter than 0.4 ms yields histograms that show more separation and stronger bimodal characteristics in the mixture as compared to the previously generated histograms (BSA: n=883; IgG: n=626; mixture: n=890). Removing events at increasingly longer dwell times (2 right-most columns) creates additional separation between modes in the histogram of the protein mixture (2 ms cutoff, BSA: n=789; IgG: n=313; mixture: n=487). With hydrogel facilitated detection, we were able to remove all events shorter than 10 ms (right-most column) and still produce distinct and clear distributions for BSA and IgG separately (BSA: n=516; IgG: n=106) and in mixture (n=123).

orientations. As a result, the recorded maximum current amplitude most likely corresponds to the protein orientation that produces the maximum blockade current. In Figure 2.6B, we have selected all events from Figure 2.6A with dwell times greater than 400 μs in order to reduce orientation effects of the current amplitude distributions. Figure 2.6B shows the current histogram for BSA alone ranging from 400-800 pA, and a histogram for IgG alone ranging primarily between 900-1950 pA; the histogram for the mixture of the two proteins shows two separated distributions with ranges of 300-550 pA and 1300-2000 pA. The separated distribution allows clear association of events with BSA and IgG and demonstrates that events longer than 400 μs can be assigned as BSA or IgG detection events with high confidence. Histograms of the currents shown in Figure 2.6A for different dwell time cutoffs are shown in Figure 2.7.

As discussed above, the measured blockade current is proportional to the protein volume and is also affected by the protein orientation. Fitting the histograms in Figure 3B to Gaussian distributions yields a most probable amplitude of 644±17 pA for BSA alone, 1690±15 pA for IgG alone, and 401±6 pA and 1732±42 pA for BSA and IgG mixture, respectively. The ratio of the IgG and BSA most probable amplitudes is 2.62 (95% confidence: 2.53 - 2.72) for the proteins

measured separately and 4.32 (95% confidence: 4.15 – 4.48) when they were measured in a mixture. The change in amplitude ratio between separated protein and mixture appears to be related to the reduction in amplitude of BSA capture events, though we cannot speculate on the reason why the amplitude reduced for BSA events and remained the same for IgG when in mixture. It is possible that the presence of a second protein affected the capture of BSA, or that the protein-pore or protein-gel interaction changed over the duration of the experiment. We also have also measured amplitude ratios of ~2.6 for the proteins in mixture in other experiments. The ratios are comparable to previously reported volumes of BSA and IgG approximated by dynamic light scattering of 143 and 391 nm^3 (ratio of 2.73),[7,21] approximate volumes calculated by the hydrodynamic radii of the proteins, 3.35 nm and 5.5 nm (volume ratio of 4.3), and volume measurements determined by other experimental means of BSA and IgG of 109 and 347 nm^3 respectively (ratio of 3.18).[21,22] The agreement with the literature-reported volumes indicates that the hydrogel enables accurate measurement of protein blockade amplitudes and simultaneous measurement of a protein mixture.

Although not demonstrated, we expect that previous work addressing the rapid translocation of proteins may also enable simultaneous measurement and discriminations of proteins in a mixture, but measurement of proteins with a hydrogel would be the simplest and most efficient method. It is worth noting that a high-bandwidth amplifier would provide accurately resolved blockade currents, but because rotational motion induces wide blockade amplitude distributions, discrimination of multiplexed proteins using amplitudes would still require dwell times long enough to ensure that the protein had permuted through all possible orientations. As a result, a technique to increase the residence time of proteins or another method of analysis, possibly including intraevent analysis, would still be necessary in order to identify proteins in a mixture

even if a high-bandwidth measurement could resolve very short duration events. Here, we rejected all events under 400 μs duration and still retained over 883 (52% of captured events), 626 (47%), and 890 (62%) events when measuring BSA, IgG, and a mixture of the two respectively.

Longer measurement times or a method to increase the number of events with sufficiently long dwell times may allow for measurement of more than two proteins in a mixture. Although, here, we have shown simultaneous measurement of two relatively different sized proteins, future work may require discrimination of a larger set of proteins in mixture or two proteins that are more similar in size. Given the separation in amplitude distributions that is shown in these histograms, measurement of a mixture of 3 or 4 proteins should be possible without modification to the protocol given that the proteins are differ in size. For this experiment, we were able to produce useful histograms even after removal of events shorter than 10 ms (Figure 2.7), with these histograms showing stronger separation in the amplitude distributions for IgG and BSA in mixture. At dwell time cutoff points that are this long or longer, maximum amplitude may be a less effective measurement than average amplitude. As dwell times get longer, maximum amplitudes will typically increase due to random noise fluctuation. Average amplitude may be a stronger metric because these durations are multiple times longer than the time required for proteins to permute through all rotations, and as the durations get longer, the average amplitude distribution should narrow.

2.2.4 Shape information from hydrogel enhanced measurements

Recently, Yusko *et al.* found that increased residence times of proteins within nanopores can enable inference of a protein's 3D ellipsoidal shape, rotational diffusion coefficient, and dipole moment.[7] The study drew upon DeBlois and Bean's work showing that ionic current through a

submicron diameter pore was affected both by the size of a particle located inside of the pore and, in the case of asymmetrical particles, the orientation of the particle inside the pore.[23] Their measurements utilized proteins tethered to a lipid bilayer coated on the nanopore surface, increasing the protein's residence time within the pore and enabling accurate measurement of current blockade amplitudes and analysis of current fluctuations within a blockage event. They used histograms of each event's maximum current to determine protein size and shape. In a more recent study, the same group demonstrated that measurements of protein shape and dipole moment were possible with free protein in solution, using low capacitance measurement setups to increase the bandwidth of the measurement.[2] In this work, intraevent currents of events with durations over 150 μs, the cutoff suggested by the work to be long enough to allow protein permutation through all orientations, were analyzed. The group also used the maximum amplitude of individual events longer than 30 μs to determine protein shape and dipole moment.

With hydrogel facilitated measurements, we observe asymmetrical current amplitude distributions that we hypothesized correlate to the shape of the protein. To explore this hypothesis, we measured spherical 10 nm gold nanoparticles and IgG, creating maximum amplitude histograms from the data for each (Figure 2.9A and 2.9B, respectively). The maximum currents measured with gold nanoparticles were normally distributed, as expected for spherical particles with ellipsoidal semi-axis ratios equal to 1.[23] In contrast, the maximum currents measured with IgG were asymmetrically distributed, consistent with IgG's aspherical shape.[7,23] For data in Fig. 2.9B, the asymmetrical current distribution can be described by a model developed previously for Coulter counting[23–25] and then reapplied to protein measurement with nanopores by Yusko *et al*.[7] The model predicts varying current blockade amplitudes dependent on the orientation of the nanopore with orientations where the ellipsoidal semi-axes are parallel to direction of translocation

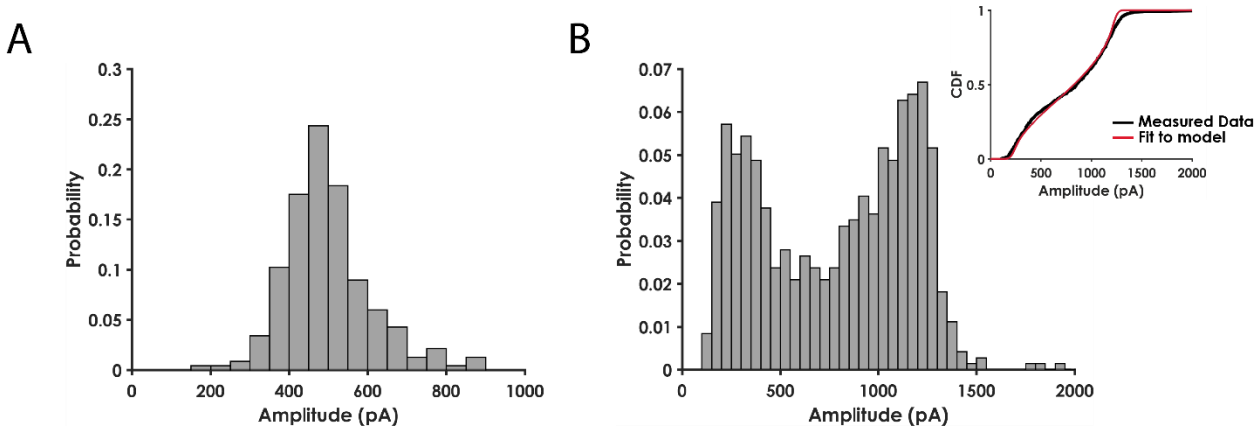


Fig 2.9 Particle shape dependence of measured blockade currents. Histograms of the maximum change in current of each blockade event for (A) 10 nm gold nanospheres measured at -40 mV (n=234) and (B) IgG measured at -50 mV (n=717) (B). The distribution in (A) is unimodal, symmetric, and gaussian, which would be expected for a spherical particle. The distribution in (B) is bi-modal, as may be expected for a non-spherical protein.[7] The cumulative distribution was fit to an orientation-dependent current blockade model (B, inset) with a lower peak of ~250 pA and upper peak of ~1210 pA.

are more probable. Here, the cumulative distribution of blockade amplitudes in Fig. 2.9B was fit to the model as described by Yusko to determine a ΔI_{min} of 250 pA and a ΔI_{max} of 1210 pA, (R^2 = 0.992, errors in fitted parameters are <1pA) (Fig. 2.9B, inset) and the corresponding protein shape factors. The measured currents predict an ellipsoidal semi-axis ratio, a/b, of 0.14, agreeing well with measurements found by Yusko.[7]

The fitted distribution predicted a dipole moment of only 90±32 D, significantly less than dipole moments measured and predicted by Yusko around 850 D, and more importantly, the height of the ΔI_{max} peak is larger than the ΔI_{min} peak, predicting a dipole moment that aligns with the wrong ellipsoidal semiaxis.[7] The incorrect prediction is likely due to two causes. Firstly, the increased dwell time distributions of hydrogel facilitated detection are broad, with non-negligible

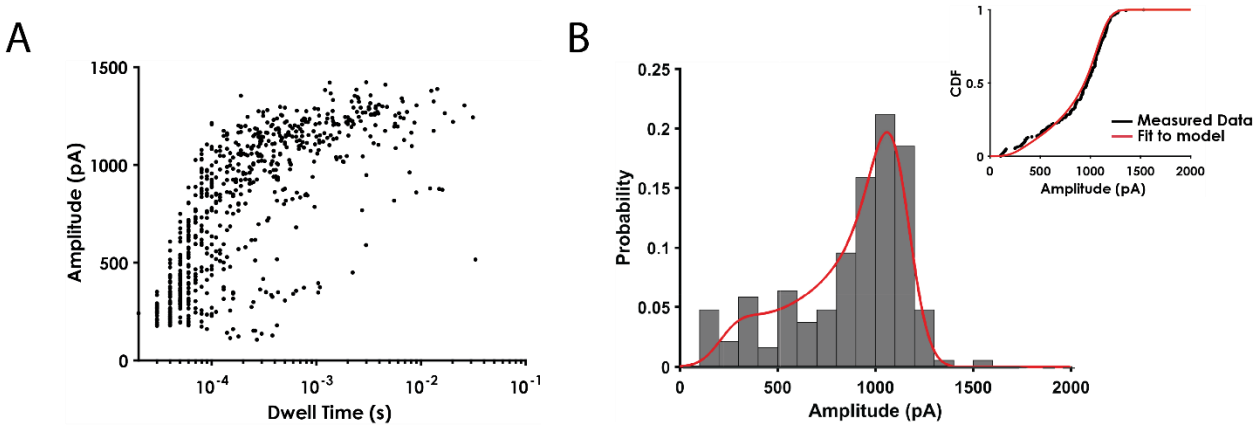


Fig 2.10 Data in Fig 2.9 shows measurement of IgG in 2M KCl, pH10 at -50 mV with a hydrogel back nanopore produces current blockades with orientation dependent maximum amplitudes. However, the data only predicts ΔI_{min} and ΔI_{max} accurately and fails to predict the magnitude and orientation of the dipole moment. A possible cause for this misprediction may be interaction between the protein and hydrogel, changing orientation preference for the protein. A second cause may be the wide dwell time distribution because longer dwell times will bias the distribution towards ΔI_{max}. This is more clearly demonstrated by plotting a scatter plot of dwell times against maximum amplitude for all data points (A). For dwell times longer than 1 ms, almost all data points are near or greater than 1000 pA. To address any filtering effects and to reduce the dwell time bias in the data, the same analysis and fitting was performed on data points with dwell times greater than 100 µs and less than 400 µs (shown in Fig S4B, n=189). A histogram of this subset shows an asymmetrical distribution, and fitting the data similarly to data shown in Fig 4 gives a ΔI_{min} of 238pA and ΔI_{max} of 1127pA (R^2 = 0.995), predicting an ellipsoidal semi-axis ratio of 0.15 which agrees well with previous work.[7] However, the bias towards ΔI_{max} has increased in this subset of data which may indicate that dwell times under 100 µs are needed for determination of dipole moments or that protein-hydrogel interaction significantly affects the orientational preference of a protein inside a nanopore. Further work would be required to elucidate the exact steric effect of the hydrogel and if determination of the dipole moment is feasible with hydrogel backed nanopores.

probability up to the millisecond time scale. As dwell times get longer, it is more likely that the protein will have, at some point during capture, aligned in the maximum blockade orientation, biasing maximum amplitude distributions towards ΔI_{max}. Yusko *et al.* tethered proteins to a lipid bilayer to increase the dwell time of translocation events, but they report that the majority of translocation events producing their maximum amplitude distributions were 50-100 μs. These durations were short enough to produce accurate distributions to predict the dipole moment of the protein. In contrast, only 20.2% of events in Fig 2.9B are 50-100 μs in duration, and 50.2% of events are longer than 150 μs, which is the time previously found to allow the protein to permute through all orientations (scatter plots of data in Fig 2.9B are shown in Fig 2.10).[2] As a result of long dwell times, the maximum amplitude distributions produced by hydrogel facilitated measurements are biased towards the ΔI_{max} peak. Fig 2.10 shows the maximum amplitude histogram for events that are 100-400 μs in duration, showing that the bias towards ΔI_{max} is increased when events shorter than 100 μs are removed, leading to a more significant miscalculation of the dipole moment.

In addition to the bias caused by longer dwell times, it is also likely that there is some hydrogel-protein interaction or that longer dwell times lead to greater chance of protein-pore interaction. Given that we hypothesize that the long dwell times observed with a hydrogel backed nanopore are due to steric effects, we also expect that this interaction will also affect protein rotation, possibly biasing the orientation probabilities and therefore leading to incorrect dipole moment predictions. Houghtaling *et al.* also showed that shape estimations could be affected by protein interaction with pore surfaces, and used lipids to prevent this interaction.[2] It is likely in our experiments, especially with the long durations that proteins reside inside nanopores, that many events involve some interaction of the protein with the pore surface.

The discussion here serves to show that hydrogel facilitated measurement of proteins can inform on protein shape, but further work is needed to fully explore these measurements with hydrogel interfaced nanopores. The recent work by Houghtaling *et al.* explores measurement of the shape and dipole moment of freely rotating proteins with nanopores, and offers insight on further work.[2] The study capitalized on a low capacitance measurement setup for higher measurement bandwidth and lipid coated nanopores to enable their shape and dipole moment measurements. The study found that proteins in free solution rotate more rapidly than when tethered to lipids and that low pass filtering can attenuate the predicted length-to-diameter ratios of proteins, specifically for intraevent analysis. A second, relevant finding in the study was that proteins located near pore surfaces could cause significant deviations in the measured protein shape, and that lipid coatings could prevent this undesirable behavior. The study also concluded that a greater number of candidate events and longer residence times can lead to more accurate analysis. Interfacing hydrogels with nanopores is amenable with low capacitance setups and lipid coatings, and coupling these techniques may lead to improved measurement of protein shape and dipole moment due to the increased number and duration of candidate events and the higher available measurement bandwidth.

2.3 Conclusion

All proteins and nanoparticles used in this study were negatively charged in pH 10 buffers and were larger than the mean PEG 1000-DMA hydrogel mesh size. We saw very few to no events when a positive potential was applied (*cis* > *trans*), and a large number of events with a negative applied potential. In negative applied potentials, negative particles experience an electric force through the pore from *cis* to *trans*, and a drag force within the pore due to electroosmotic flow

from the negatively charged SiN nanopore that is directed from the *trans* chamber to the *cis* chamber. Since we saw events only at negative potentials, we concluded that the electrophoretic force was greater than the electroosmotic drag for these particles.

The materials and technique described here are especially attractive because they are easy to implement with conventional nanopores, measurement apparatus, and a wide range of protein analytes. As the technique is label-free, the identity of the protein analyte does not need to be known for the measurement to be successful. As implemented here, the only modification from typical solid state nanopore measurements is the exchange of the aqueous solution on one side of the nanopore with hydrogel precursor solution. Once polymerized, we found that allowing at least 12 hours for the hydrogel to swell yielded best results. The presence of the hydrogel typically resulted in increased baseline noise during measurement (Fig 2.5), but the measured protein events were still clearly identifiable from the baseline. In this study, we used nanopores formed using dielectric breakdown primarily because the technique allows convenient formation and growth of the nanopore.[11,26,27] Dielectric breakdown can also occur after the polymerization and swelling of the hydrogel, reducing the chance that the hydrogel expands into the nanopore and prevents proteins from entering.

Protein translocation through conventional synthetic nanopores occurs over short timescales, resulting in a large number of missed events and incorrectly resolved blockade currents. Here we show that the presence of a hydrogel on the distal side of the nanopore can sufficiently increase the residence time of proteins introduced to the proximal side to enable orders of magnitude increased detection rates and improved resolution of current blockage amplitudes. We used this system for the label-free detection of BSA, IgG, streptavidin, and gold nanoparticles.

We also showed identification of proteins measured from a mixture, and event analysis exhibited indications of particle shape.

Future work may include further investigation of the interaction between the hydrogel and protein. This kind of investigation may help in interpretation of currents for better identification of protein shape, dipole moment, and rotational diffusion constant given that protein-hydrogel interaction may affect rotation of the protein. In an effort to improve measurement of protein shape and dipole moment, using higher measurement bandwidth and lipid coatings on the nanopore may address issues in identifying the dipole moment of a protein using a hydrogel interfaced nanopore. Previous work has also shown that changing the mesh size and chemistry of a hydrogel allowed tuning of the dwell time distribution of translocating nucleic acids.[6] A similar approach may be applicable to protein sensing, especially given the diversity in physical protein characteristics such as charge, size, and shape.

The hydrogels used in this work are inexpensive, commercially available, easy to implement in existing measurement apparatus, and applicable to a wide range of nanopore substrates and unknown, un-labeled protein targets. In future work, we will explore effects of hydrogel mesh size on the detectable range of proteins, and the effect of gel type of capture rate and dwell time. Further experimentation is also required to understand the mechanism by which the hydrogel affects the residence time of particles within the nanopore.

2.4 Materials and Methods

All materials were purchased from Sigma-Aldrich unless otherwise noted and used as received unless otherwise noted. Immunoglobulin G 1 (IgG) and bovine serum albumin (BSA)

were solubilized in 50% glycerol, 50% 5 mM NaCl, pH 7, and streptavidin was solubilized in 5 mM NaCl, pH 7. All proteins were stored at -20°C at 40-50 μM.

2.4.1 Nanopore fabrication

A thin layer of PDMS was painted and cured onto the silicon nitride face of commercially available 15 nm (Norcada) and 15 nm (TedPella) thick TEM windows, exposing only the area of the window, to reduce capacitive noise.[28] The TEM windows were plasma treated for 20-30 seconds at 30 W and promptly mounted in a home-built Teflon fluidic chamber. The chambers were filled with ethanol and placed under vacuum to evacuate any trapped air bubbles in the chamber. The ethanol was replaced with 2 mL measurement buffer on both halves of the chamber and capacitance and resistance were measured to ensure proper wetting of the membrane. The measurement buffer was removed from the silicon nitride-facing half of the chamber and the hydrogel was polymerized as described below. For all steps where the chamber was filled with measurement buffer or ethanol, a silver chloride wire bridged the two halves of the chamber to minimize pinhole formation from charge imbalance across the two halves of the chamber.[20]

Nanopores were formed by controlled dielectric breakdown through LabVIEW with a NI USB-6251 (National Instruments), a simple current to voltage converting op-amp circuit with a OPA111 opamp, and silver chloride electrodes. For all measurements, the silicon nitride face of the window was grounded. A 100-200mV voltage was applied while monitoring currents, which remain <20pA before a nanopore has formed. A 100ms -4-7V voltage pulse was applied at 2Hz until currents increased above 20pA, which indicated formation of a nanopore. Further pulses of ±3-5V were

applied to enlarge the pore until conductance indicated the nanopore was approximately the target diameter, typically 22-30 nm.

2.4.2 Hydrogel preparation

PEG (1000 MW) Dimethacrylate (PEG-DMA) was dissolved in measurement buffer, 2M KCl, 10mM Tris-HCl, pH10 or pH12, at 10% w/v. This gel precursor solution was vortexed thoroughly to ensure complete dissolution and then sonicated under vacuum for at least 10 minutes to degas the solution. Gel precursor solution was mixed with the APS solution and tetramethylethylenediamine (TEMED) so that final concentrations of APS and TEMED were both 10mM. 1.4mL of the mixture was then pipetted into the side of the chamber facing the silicon nitride side of the TEM window and allowed to polymerize for 10 minutes. After polymerization, 600µL of measurement buffer was pipetted into the hydrogel half of the chamber to balance the volumes of solution on either side of the chamber and to allow swelling of the hydrogel. Hydrogels were allowed to swell for at least 12 hours before further protocol.

Gel mesh sizes were determined by a method as described previously.[10] Briefly, gels were prepared exactly as described above. Masses were measured after polymerization, again after allowing at least 24 hours for swelling, and again after lyophilization. Weights after lyophilization were adjusted for expected salt content. For reference, swollen equilibrium polymer volume fractions were 0.093 ± 0.009. All mass measurements were conducted in centrifuge tubes to mimic the nanopore fluidic chambers.

2.4.3 Measurement of nanopores

Meaurement of nanopores was conducted using an Axopatch 200B amplifier, Digidata 1440B or 1322A, and data was acquired in Clampex at 100 kHz with a 10 kHz hardware low pass filter. The side of the chamber adjacent to the silicon nitride face of the TEM window, and also containing the hydrogel, was grounded for all measurements. Voltages between -100 mV and -30 mV were held across the nanopore to determine baseline current characteristics of the nanopore. These control measurements were performed for at least 5 minutes at each relevant voltage. Subsequently, protein or 10 nm diameter, PVP or Citrate functionalized gold nanoparticles (NanoComposix) were pipetted to the side of the chamber facing the silicon side of the TEM window. For gold nanoparticles, if volumes of analyte solution exceeded 10 µL, the same volume of measurement buffer was pipetted to the opposite side of the chamber to ensure minimal change in hydrostatic pressure across the nanopore. Voltages applied during measurement of open pore baseline were repeated, with at least 30 minutes of total measurement time following the addition of analyte to allow capture rate to reach steady-state. If events lasted for longer than 10 seconds, the zap button on the Axopatch (+1.3V for 50ms) was used to electrophoretically force the analyte out of the pore to reduce risk of clogging. If no events were observed by 30 minutes, the experiment was terminated.

2.4.4 Data analysis

All data was analyzed in MATLAB. For all traces, traces were unfiltered or otherwise unprocessed after acquisition. Event starts and ends were marked by points where currents deviated and returned from the baseline, so events where a second analyte entered the pore were not processed uniquely from single analyte events, though events with second level step changes, which may be

caused by a second analyte entering the pore, were not common in our measurements. In some experiments, analyte residence times were long enough that second level events were common, but these experiments were not included in this work as second analyte events may have differing characteristics from first analyte events. Because the long observed dwell times can result in misleading deviations from the Smoluchowski rate equation due to reduced event frequency, the inverse of the average interevent time ($1/t_{off}$) was calculated and reported as an equivalent to the event frequency that is independent of the dwell time. Where most probable amplitudes are reported, histograms of event amplitudes were fit to a Gaussian distribution, and the mean was reported. In the case of simultaneous measurement of two proteins, the data was fit to a bimodal distribution. Hydrodynamic diameters were calculated from these most probable amplitudes using models previously used in pore sensing by DeBlois and Bean[23] and later by Yusko *et al.*[7] assuming spherical particle shape. Protein shape factors were calculated by fitting to models as described by Yusko *et al.* without assuming shape of the protein.[7]

1. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13,** 658–663 (2013).

2. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* acsnano.8b09555 (2019). doi:10.1021/acsnano.8b09555

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

4. Squires, A. H., Hersey, J. S., Grinstaff, M. W. & Meller, A. A Nanopore–Nanofiber Mesh Biosensor To Control DNA Translocation. *J. Am. Chem. Soc.* **135,** 16304–16307 (2013).

5. Tang, Z. *et al.* Gel mesh as "brake" to slow down DNA translocation through solid-state nanopores. *Nanoscale* **7,** 13207–13214 (2015).

6. Al Sulaiman, D., Cadinu, P., Ivanov, A. P., Edel, J. B. & Ladame, S. Chemically Modified Hydrogel-Filled Nanopores: A Tunable Platform for Single-Molecule Sensing. *Nano Lett.* **18,** 6084–6093 (2018).

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

8. Bernhard, C. *et al.* Repelling and ordering: The influence of poly(ethylene glycol) on protein adsorption. *Phys. Chem. Chem. Phys.* **19,** 28182–28188 (2017).

9. Benjamin, C. J. *et al.* Nonfouling NTA-PEG-Based TEM Grid Coatings for Selective Capture of Histidine-Tagged Protein Targets from Cell Lysates. *Langmuir* **32,** 551–559 (2016).

10. Canal, T. & Peppas, N. A. Correlation between mesh size and equilibrium degree of swelling of polymeric networks. *J. Biomed. Mater. Res.* **23,** 1183–1193 (1989).

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

12. Jøssang, T., Feder, J. & Rosenqvist, E. Photon correlation spectroscopy of human IgG. *J. Protein Chem.* **7,** 165–171 (1988).

13. Bermudez, O. & Forciniti, D. Aggregation and denaturation of antibodies: A capillary electrophoresis, dynamic light scattering, and aqueous two-phase partitioning study. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **807,** 17–24 (2004).

14. Mathes, J. & Friess, W. Influence of pH and ionic strength on IgG adsorption to vials. *Eur. J. Pharm. Biopharm.* **78,** 239–247 (2011).

15. Khawli, L. A. *et al.* Charge variants in IgG1. *MAbs* **2,** 613–624 (2010).

16. Valiño, V. *et al.* Accurate determination of key surface properties that determine the efficient separation of bovine milk BSA and LF proteins. *Sep. Purif. Technol.* **135,** 145–157 (2014).

17. Yadav, S., Shire, S. J. & Kalonia, D. S. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **28,** 1973–1983 (2011).

18. Hendrickson, W. A. *et al.* Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2190–4 (1989).

19. Talaga, D. S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* **131,** 9287–9297 (2009).

20. Matsui, K., Yanagi, I., Goto, Y. & Takeda, K. I. Prevention of dielectric breakdown of nanopore membranes by charge neutralization. *Sci. Rep.* **5,** 17819 (2015).

21. Singh, M., Chand, H. & Gupta, K. C. The studies of density, apparent molar volume, and viscosity of bovine serum albumin, egg albumin, and lysozyme in aqueous and RbI, CsI, and DTAB aqueous solutions at 303.15 K. *Chem. Biodivers.* **2,** 809–824 (2005).

22. Schneider, S. W., Lärmer, J., Henderson, R. M. & Oberleithner, H. Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflugers Arch. Eur. J. Physiol.* **435,** 362–367 (1998).

23. DeBlois, R. W. & Bean, C. P. Counting and sizing of submicron particles by the resistive pulse technique. *Rev. Sci. Instrum.* **41,** 909–916 (1970).

24. Hurley, J. Sizing Particles with a Coulter Counter. *Biophys. J.* **10,** 74–79 (1970).

25. Grover, N. B., Naaman, J., Ben-Sasson, S., Doljanski, F. & Nadav, E. Electrical Sizing of Particles in Suspensions: II. Experiments with Rigid Spheres. *Biophys. J.* **9,** 1415–1425 (1969).

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

27. Yanagi, I., Akahori, R., Hatano, T. & Takeda, K. Fabricating nanopores with diameters of sub-1 nm to 3 nm using multilevel pulse-voltage injection. *Sci. Rep.* **4,** 5000 (2015).

28. Tabard-Cossa, V., Trivedi, D., Wiggin, M., Jetha, N. N. & Marziali, A. Noise analysis and reduction in solid-state nanopores. *Nanotechnology* **18,** 305505 (2007).

Chapter 3

The mechanism of capture and study of dwell time for hydrogel facilitated detection of proteins

3.1 Introduction

The presence of a hydrogel on the distal side of a nanopore has been shown in the previous chapter to address many of the issues in protein detection related to short dwell times. Although the presence of a hydrogel on the distal side of a nanopore is effective at increasing the dwell time of proteins in nanopores, the exact mechanism by which it does so is still unknown. Elucidating and understanding this mechanism has significant impact on future work. The most immediate benefit is better interpretation of results, specifically in relation to dwell times of detected events. Dwell times may be able to offer additional insight into the physical characteristics of proteins or possibly be used in helping identify proteins in mixture. Additionally, understanding the mechanism of capture could enable tailoring of dwell times for specific applications and possibly enable new types of measurements altogether.

The increase in dwell time is most likely a result of steric interaction between the hydrogel and protein. It is unlikely that there is a electrostatic or chemical interaction between the PEG-DMA hydrogel and protein as PEG minimizes protein adsorption and interaction[1,2] and is neutrally charged. Previous work has shown that the relationship between the mesh size of the hydrogel and the radius of gyration of the translocating nucleic acids had significant impact on the dwell times, further supporting that increased durations are a steric effect.[3–6] In these studies, nucleic acids were

expected to fully complete translocation and residence times increased because the mobility of the nucleic acid reduced in the presence of the gel. These studies showed that reducing the mesh size of the hydrogel could increase the average dwell time of translocating nucleic acids.[3,6]

However, the PEG-DMA hydrogel used previously had a mesh size of 3.1 nm, smaller than the size of all proteins measured previously. Unlike nucleic acids, the tertiary and quaternary structures of native proteins make it unlikely that the molecule unfolds to fit into smaller hydrogel pores. As a result, the hydrogel may completely prevent translocation of proteins into the *trans* compartment. Conversely, it is also possible that the hydrogel mesh may have enough variance to have pores large enough or PEG chains that are flexible enough to allow protein penetration.

Here, we show explore these two possible mechanisms, one in which the protein penetrates the hydrogel and the electrophoretic mobility is reduced and one in which the protein cannot penetrate the hydrogel and must diffuse back into the *cis* chamber. A partial analytical model based on the Fokker-Planck equation is presented and COMSOL simulations are used to predict dwell time behavior for both mechanisms. We show that for proteins that we have measured, immunoglobulin G 1 (IgG) and bovine serum albumin (BSA), the dwell time behavior suggests that the proteins cannot penetrate the PEG-DMA hydrogel. As a result, dwell times increase exponentially with voltage and the dwell time distribution is much more sensitive to the diffusivity and electrophoretic mobility of the protein. We show both behaviors experimentally with IgG and BSA, showing that the dwell time is more sensitive to protein diffusivity and electrophoretic mobility when compared to conventional nanopore sensing.

3.2 Results and Discussion

3.2.1 Analytical models for dwell time distribution

The residence time distributions of proteins detected by nanopores can be derived from the Fokker-Planck equation, describing 1-D diffusion under force,

$$\frac{\partial P(x,t)}{\partial t} = F_E \mu_e \frac{\partial P(x,t)}{\partial x} + D \frac{\partial^2 P(x,t)}{\partial x^2}$$

where, F_E is the force of the electric field, μ_e is the electrophoretic mobility of the protein, and D is the diffusivity of the protein. The equation has been solved for numerous conditions including the first passage time of charged particles through nanopores.[7,8] Here, we have reproduced the solution and expand on it in relation to the effect a hydrogel may have on the translocation of proteins.

For equations shown here, the origin is the boundary of the nanopore to the *trans* compartment. The initial condition is defined as $P(x, t=0) = \delta(x-l_p)$, where $x=l_p$ is the *cis* entrance of the nanopore. A solution for this initial condition with boundary conditions as $P(x=\pm\infty,t) = 0$ is shown here:

$$P(x,t) = \frac{\exp\left(-\frac{\left(x - l_p - F_E \mu_e t\right)^2}{4Dt}\right)}{\sqrt{4\pi D t}}$$

In typical nanopore measurements, the protein is unlikely to move against the force produced by an electric field and therefore irreversibly translocates through the nanopore into the *trans* compartment. As a result, for nanopore measurements, an absorbing boundary condition can be defined at the *trans* boundary of the nanopore, $P(x = 0,t) = 0$. An image source at $x=-l_p$ provides a method to satisfy the boundary condition, resulting in the solution,

$$P(x,t) = \sqrt{\frac{1}{4\pi Dt}}\left[\exp\left(-\frac{\left(x - l_p - F_E\mu_e t\right)^2}{4Dt}\right) - \exp\left(\frac{F_E\mu_e l_p}{D} - \frac{\left(x + l_p + F_E\mu_e t\right)^2}{4Dt}\right)\right]$$

and a dwell time distribution,

$$P_{dwell}(t) = \frac{\partial}{\partial t}\left(1 - \int_0^\infty P(x,t)\,dx\right) = \frac{l_p}{t\sqrt{4\pi Dt}}\exp\left(-\frac{\left(l_p - F_E\mu_e t\right)^2}{4Dt}\right)$$

These results were previously shown by Talaga *et al.* (a correction to the original result was produced by Ling and Ling[8]) and used to model the dwell time distribution of tranlocating proteins in typical nanopore measurements (Fig 3.1B).[7] A PEG(1000 MW) DMA hydrogel placed on the *trans* side of a nanopore improves measurement of proteins with nanopores by increasing the residence time of detected proteins. PEG is a neutrally charged, unreactive polymer known to be nonfouling,[1,2] and therefore, the increased residence time in PEG-DMA facilitated measurements is most likely not due to a chemical or electrostatic mechanism. Instead, we suggest two possible steric mechanisms that explain the increased residence time of proteins when a hydrogel is on the *trans* side of the nanopore. In one case, the hydrogel acts only as a steric hinderance to the protein, allowing the protein to penetrate the hydrogel and fully translocate from the *cis* side to the *trans* side of the nanopore (Fig 3.1C). The hydrogel effectively reduces the mobility of the protein, and the above solution would be applicable though D and μ_E would be reduced relative to the mobility of a protein in a typical measurement without a hydrogel.

Alternatively, the hydrogel may act as an impenetrable steric barrier to translocation. The protein will be electrophoretically driven into the pore, but will not be allowed to exit the nanopore through the distal side due to the presence of the hydrogel (Fig 3.1D). In this case, the protein can only exit the pore through the *cis* opening of the nanopore, diffusing upstream against the

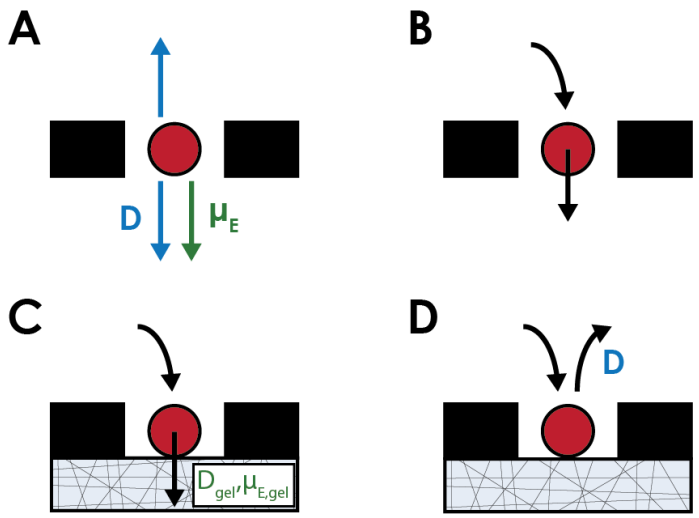


Fig 3.1 Proteins detected by nanopores have movement dictated by diffusion, dictated by the protein's diffusivity, D, electrophoresis, determined by the applied voltage and the protein's electrophoretic mobility, μ_E, and electroosmotic flow, determined by the applied voltage and the nanopore surface charge. Here, to simplify the model, electroosmotic flow is not included, though effects of electroosmotic flow can be estimated in the model by subtracting the electroosmotic flow speed from the electrophoretic speed to yield the net convection. An equivalent effect is achieved by linearly attenuating the electrophoretic terms such as μ_E or the applied voltage. (A) Diffusion results in random forward or backward movement of the protein, leading to variance in the time the protein resides in the pore. Electrophoresis will push the protein towards the *trans* chamber with a constant drift velocity. (B) In traditional nanopore measurements, a detected protein is highly likely to complete translocation due to the electrophoretic force. (C) With a hydrogel at the distal side of the nanopore, it is possible that the protein can still penetrate the hydrogel, though the electrophoretic and diffusive terms may be reduced in the presence of the hydrogel, resulting in a slower translocation. (D) The protein may also not be able to penetrate the hydrogel, and as a result, can only exit the nanopore by diffusion against the direction of electrophoretic force, also resulting in longer dwell times.

electrophoretic field. As a result, a no flux boundary condition at the *trans* boundary of the nanopore,

$$\frac{dP(x=0,t)}{dx}=0$$

accounts for the impenetrable hydrogel barrier. The boundary condition, P(x=∞,t) = 0, is still valid. The solution for a rebounding boundary condition has been previously solved,[9]

$$P(x,t)=\sqrt{\frac{1}{4\pi Dt}}\left\{\exp\left(-\frac{\left(x-l_p-F_E\mu_e t\right)^2}{4Dt}\right)+\exp\left(-\frac{4l_pF_E\mu_e t-\left(x+l_p+F_E\mu_e t\right)^2}{4Dt}\right)\right.$$
$$\left.+\frac{F_E\mu_e}{2D}\exp\left(\frac{2F_E\mu_e x}{4Dt}\right)\left[1-\mathrm{erf}\left(\frac{x+l_p+F_E\mu_e t}{\sqrt{2Dt}}\right)\right]\right\}$$

though an analytical solution for the dwell time distribution cannot be solved, and instead a computational approach must be taken to predict dwell time distributions for a rebounding wall boundary condition.

3.2.2 COMSOL models for dwell time distribution

Because an analytical solution for the dwell time distribution with a rebounding wall at the *trans* compartment boundary is not available, a COMSOL model provides a way to determine a theoretical dwell time distribution for all cases, including when particles are not able to complete translocation into the *trans* compartment. The finite element model tracks diffusion of particles under electrophoretic force, calculating the particle's probability density over time (P(x,t)). The 1-D model includes two regions. As shown in Fig 3.2A., a 15nm long region from x=0 to x=L represents the nanopore, where x=0 is the *cis* side boundary of the nanopore and x=L is the *trans* boundary of the nanopore. For all COMSOL simulations here, L=15nm to match the nominal length of nanopores used to collect experimental data in this article. A second region extends left

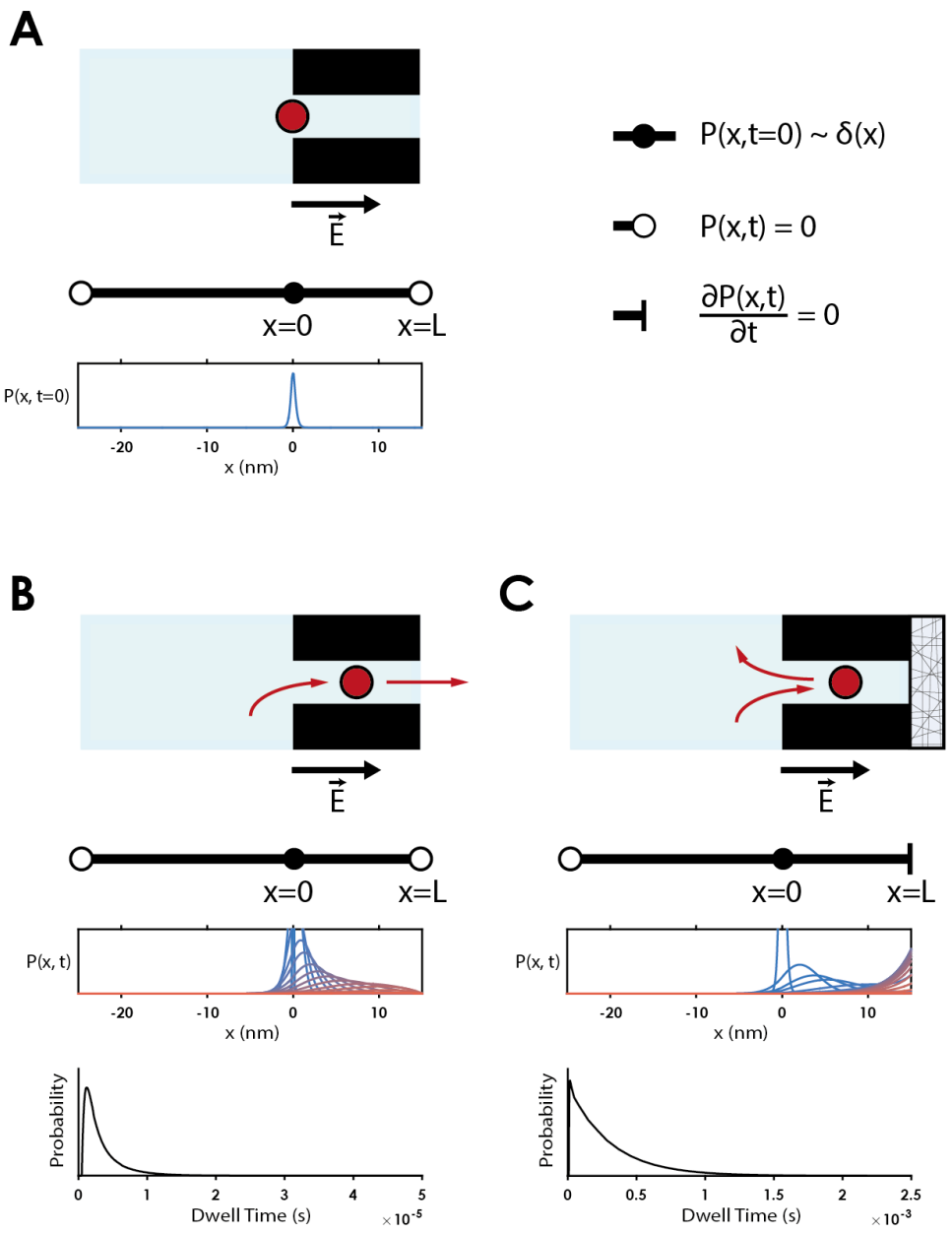


Fig 3.2 COMSOL was used to simulate the movement of proteins through a nanopore as a result of diffusion and electrophoresis. The probability density distribution was calculated over a space including the nanopore and a region on the *cis* side that approximates the extent of the electric field, and over times up to a point when the probability was 0 over the calculated space (steady state solution). (A) In all simulations, a narrow gaussian pulse at the *cis* entrance of the nanopore was used to approximate a delta function initial condition. The probability was also fixed at 0 at the far left boundary of the *cis* region to represent the boundary condition at x=-∞. (B) For simulations of

traditional nanopore experiments with an absorbing boundary at the *trans* boundary, the probability at the boundary was fixed at 0. (C) For simulations of hydrogel facilitated measurement where the protein cannot penetrate the hydrogel, a no-flux boundary condition was defined at the *trans* chamber boundary. Probability densities for a simulation of a protein with a diffusivity of $3*10^{-11}$ m^2/s and electrophoretic mobility of $3*10^{-9}$ $m^2/(V*s)$ with an applied voltage of 50mV (the electric field is calculated from this value accounting for pore length and access resistance) are shown in the upper plots of panel B and C, with colors from blue to red showing 20 sequential, logarithmically spaced, time points from 0s to steady state (~50 µs for the absorbing boundary condition and ~3ms for the rebounding boundary condition). The corresponding predicted dwell time probability distribution is shown below these plots. For conventional nanopore sensing, the majority of proteins are expected to have dwell times under 10 µs, agreeing well with previous experimental and theoretical results.[7,10,11] For protein sensing with an impenetrable hydrogel, the dwell time distribution extends to times two orders of magnitude larger.

from x=0 to x=-d_p (pore diameter) to account for access resistance and to keep space finite for the boundary condition at x=-∞. All simulations shown here assume that this second region extends between -25nm and 0nm. An electric field exists over the entire space, with magnitude according to an applied voltage divided by a length and corrected for access resistance according to equations by Hall.[12] The electric field persisting over the full length of the *cis* side region works as a compromise between Hall's assumption that the access resistance acts within a semispherical volume outside the entrance of the pore[12] and another report by Hyun *et al*. which found that the effects of access resistance persist as far as 5 pore diameters away from the entrance of a nanopore.[13] In lieu of a true delta function at x=0, a normal distribution with a standard deviation of 0.025 nm acts as the initial condition (Fig 3.2A, top).

It is worth noting, that for COMSOL models, we have opted to not simulate electroosmotic flow, which would oppose and attenuate the electrophoretic velocity in our experimental conditions, in order to simplify the simulations. Given the presence of the hydrogel, it is difficult to guess at the flow profile as it moves from flow resistant solution on the hydrogel side to free solution on the *cis* side. In the case that hydrogel does not interrupt electroosmotic flow entirely, experimental results have shown that that electroosmotic flow should contribute to the movement of proteins through nanopores,[14] but the magnitude of electroosmotic flow is expected to change based on pore shape[15–17] and decrease at higher salt concentration.[18] Additionally, variances experimental parameters, such as pore length, which affects the strength of the electric field, and pore diameter, which may encourage protein-pore interaction,[10,19,20] make one-to-one translation of simulation to experimental results unlikely. Because of these obfuscating factors, estimating the electroosmotic flow is nearly impossible and it must be measured experimentally to be quantified. If measured, the electrophoretic and electroosmotic terms can be added to describe the net convective movement of the protein. In this case, the dwell times distribution would be expected to increase for absorbing boundary condition solutions and reduce for rebounding wall solutions as electrophoresis and electroosmotic flow oppose each other, effectively attenuating or reversing the convection due to electrophoresis alone.

We replicated the same boundary conditions as used in the analytical solution. An absorbing boundary at the *trans* boundary of the nanopore simulates traditional nanopore sensing measurements where the particle translocates fully to the *trans* compartment. Fig 3.2B shows the COMSOL predicted probability density function over time (with line colors from blue to red representing forward progress in time) for a particle with a diffusivity of $3*10^{-11}$ m^2/s and electrophoretic mobility of $3*10^{-9}$ $m^2/(V*s)$. As expected, the distribution functions show a

gaussian shape that broadens over time due to diffusion and a mean that moves in the direction of electrophoretic force. Over the course of the simulation, the area under the probability distribution curve decreases as a result of the two absorbing boundary conditions, with much greater flux at the *trans* boundary. Similarly to the analytical solution shown above, we can use the COMSOL output to produce a theoretical dwell time distribution by taking the integral of the probability density function inside the nanopore to find the probability that particles are inside the nanopore. 1 minus this value is the probability that the particle is not inside the pore, which is the cumulative distribution function of the dwell time. The derivative of this function is a dwell time probability, which is shown in the bottom most plot of Fig 3.2B. The predicted dwell time distribution agrees well with the analytical solution, including solutions shown previously,[7,11] and predicts that the majority translocating proteins will not reside in the nanopore long enough to be resolved accurately by standard measurement apparatus acquiring signal at 10-100kHz bandwidth.

Adding a hydrogel may have two possible effects. In the case that the protein can penetrate the hydrogel and the effective electrophoretic mobility and diffusivity decreases, the dwell time distribution should shift towards longer dwell times as the protein moves more slowly than previously. The effect on the electrophoretic mobility would be similar to adjusting voltage or making the protein less charged or neutral by titrating the pH of measurement solution as shown before.[14] The reduced electrophoretic mobility and diffusivity may explain the increased dwell times and capture rate of proteins that we have previously observed when measuring IgG and BSA with a hydrogel.

It is also possible that the hydrogel entirely prevents translocation. Using the COMSOL model, we also investigated the effects of a rebounding wall (no flux boundary condition) at the *trans* boundary of the nanopore. The probability distribution and dwell time distribution are shown

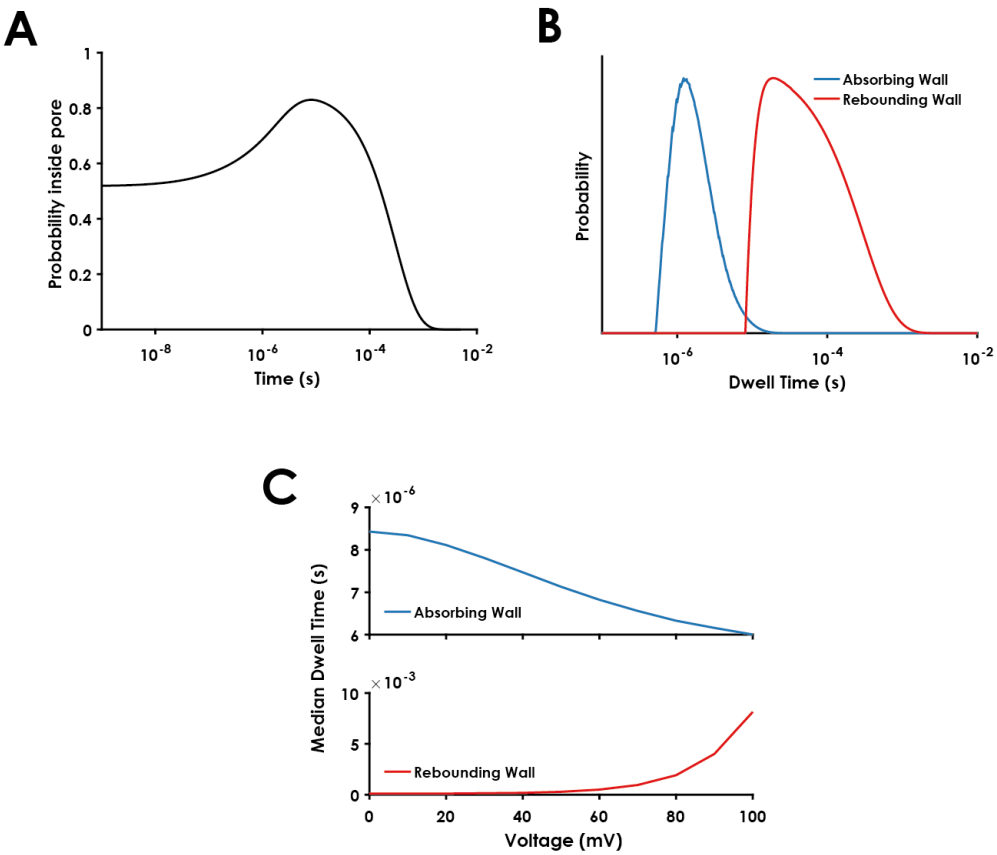


Fig 3.3 (A) For simulations with a rebounding boundary condition at the trans chamber boundary, the probability that the protein is inside the pore has a delay before reaching a maximum. This rising time leads to negative calculated dwell time probabilities, which we address by setting these probabilities to 0. This deviation does not significantly affect the dwell time distribution as the dwell times that are set to 0 are 1-3 orders of magnitude below the relevant dwell time scales for a given simulation. (B) Dwell time distributions for the two different boundary conditions are shown on the same plot to highlight the potential difference that can arise from the presence of a hydrogel at the trans boundary of the nanopore. Dwell time distributions shown here are for proteins with the same diffusivity and electrophoretic mobility as those in Fig 3.2B and 3.2C (D = 3*10-11 m2/s and μ_E = 3*10-9 m2/(V*s)) at an applied voltage of -50mV. (C) Varying the voltage has significantly different effects on the dwell time distribution depending on the boundary condition. Here, medians are plotted against voltage as a metric to quantify the dwell time distribution. For an absorbing boundary condition, the median dwell time decreases linearly with respect to voltage, as is expected

since the applied voltage helps to drive the protein through the nanopore. Conversely, for a rebounding boundary condition, the median dwell time increases exponentially with respect to voltage because increasing voltage acts to trap the protein inside the nanopore.

in Fig 3.2C. With a rebounding boundary condition at the *trans* boundary, the probability density accumulates against the *trans* boundary as a result of electrophoretic force. Once the peak of the gaussian distribution reaches the *trans* boundary, the probability density is maximum at the boundary with an exponential decay away from the boundary. Over time, diffusion results in a broadening of the distribution and leads to flux at the left most boundary on the *cis* side. With a rebounding wall at the *trans* boundary, the dwell time distribution increases by two orders of magnitude for the same protein and applied voltage. The difference is significant and shows that the improved detection of proteins with a hydrogel may also be caused by the hydrogel preventing translocation of proteins into the *trans* compartment.

For rebounding wall simulations, there is a time delay before the probability that the particle is inside the pore reaches a maximum (Fig 3.3A). As a result, the calculated dwell time probability would be negative for all times before this time point. For the absorbing wall, because the *trans* boundary is a sink, the delay until reaching this maximum is very short in duration, about 10^{-7} s for particles with a diffusivity of $3*10^{-11}$ m^2/s and electrophoretic mobility of $3*10^{-9}$ $m^2/(V*s)$ at an applied voltage of -50mV. In the case of a rebounding wall, the delay is approximately 10^{-5} s for the same type of particle at the same voltage. To account for the delay before the probability of particles inside the pore reaches a maximum, these negative dwell times are set to 0. Though this leads to some error in the predicted dwell time probabilities, the portion of the dwell time distribution that is changed is typically 1 to 3 orders of magnitude smaller than the remaining dwell time distribution, and we expect that this error does not significantly alter the

predicted dwell time distributions.. The dwell time distributions for the absorbing and rebounding conditions are shown more clearly on a log scale in Fig 3.3B, showing the effect of the delay correction, with the shortest non-zero probable dwell times being different based on the boundary condition. The two distributions also show the increased dwell times as a result of the rebounding boundary condition.

3.2.3 Relationship between voltage and dwell time

The two mechanisms both result in longer residence times of the protein in the sensing region of the nanopore, but in the former's case, the protein escapes the pore in the direction of the electric field, and in the latter, the protein escapes against the direction of the electric field. As a result, if the protein can penetrate the hydrogel, the dwell time of events should decrease as the applied voltage increases. Conversely, if the hydrogel is impenetrable, then the dwell time of events should increase as the voltage increases. The analytical and COMSOL models both show this behavior. The median dwell time predicted by COMSOL is plotted against applied voltage in Fig 3.3C for the absorbing and rebounding boundary conditions. The median was chosen to quantify the dwell time distribution as it is a metric that does not assume an underlying distribution and is easily interpretable on a cumulative distribution function as the 50 percentile value. For an absorbing boundary condition, increasing voltage correlates with a decreasing median dwell time, though increasing the voltage from 0 to 100 mV decreases the median dwell by only a fraction of the median dwell time at 0mV. Conversely, with a rebounding condition, the median dwell time increases exponentially with voltage. In this case, doubling the voltage results in an order of magnitude increase in the median dwell time.

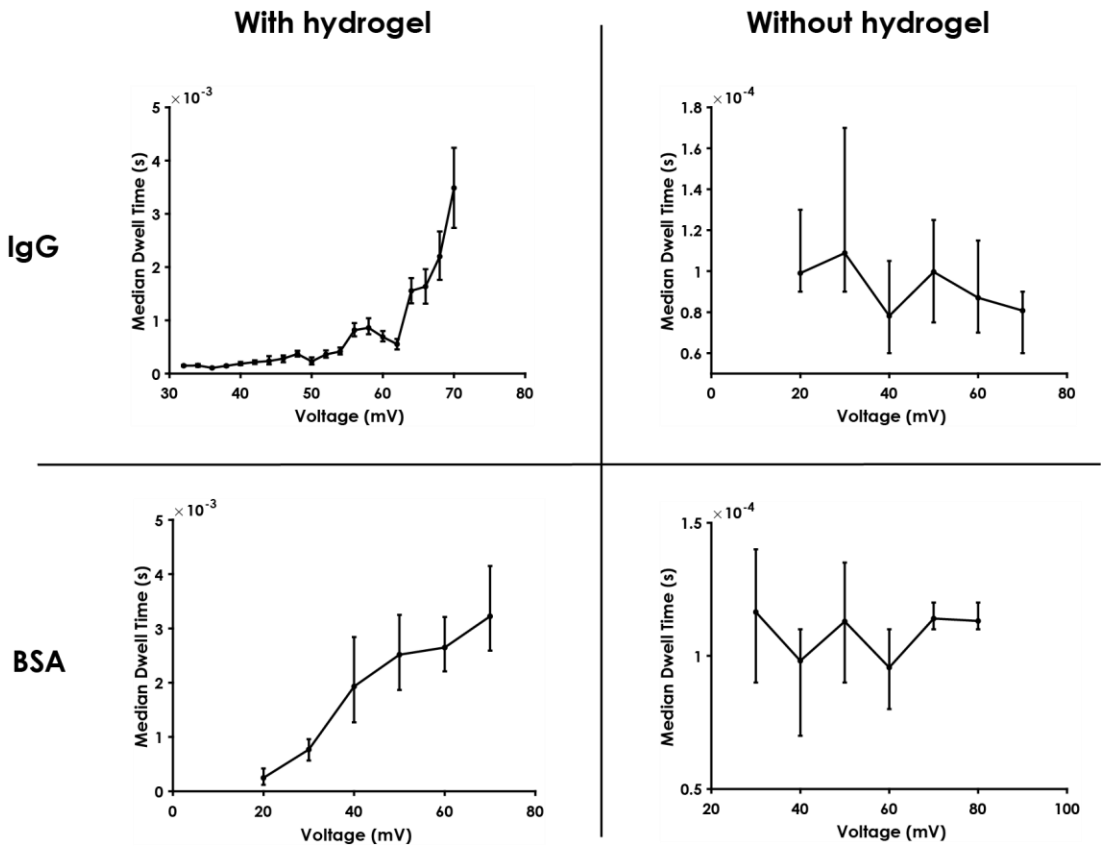


Fig 3.4 IgG and BSA were measured on two pores each (4 pores total), one with a PEG-DMA hydrogel in the *trans* compartment and one with traditional measurement conditions (without a hydrogel). IgG was measured with a 24nm diameter nanopore facilitated with a hydrogel at 2mV increments from 32mV to 70mV (top left). The median dwell time of the protein inside the nanopore increased exponentially with respect to voltage. Similarly, BSA was measured with 23nm diameter nanopore with a hydrogel in the *trans* compartment from 20 to 70mV in 10mV increments, and median dwell times were also shown to increase with respect to voltage. These results support the rebounding wall mechanism, suggesting that IgG and BSA cannot penetrate the hydrogel. In contrast, when measured without a hydrogel, both IgG (23nm diameter nanopore, 20 to 70mV by 10mV increments) and BSA (17nm diameter nanopore, 30 to 90mV by 10mV increments) exhibited no clear trend in dwell time with respect to voltage. Median dwell times for these experiments were around 100 µs, less than an order of magnitude away from the detection cutoff of 20 µs and near the timescale cutoff of the 10kHz low pass filter applied during measurement. The lack of a trend

between voltage and median dwell time when no hydrogel is present is most likely because most translocating proteins reside in the nanopore too briefly to be detected.

We measured IgG and BSA with each two nanopores, with and without a hydrogel. The median dwell times for these measurements are plotted against voltage in Fig 3.4. Without a hydrogel, there is no measureable trend between median dwell time and voltage, but we expect that the median dwell time should decrease with increasing voltage. The deviation from theory has been addressed previously by Plesa *et al.*, suggesting that due to the limitations of the measurement apparatus, only the tail end of the distribution is observed in nanopore experiments.[11] The majority of the distribution is below the 20 μs detection threshold because the proteins reside in the pore too briefly to be detected. As a result, only the longest events in a given distribution are detected, and of those that are detected, the majority are at or near the detection threshold. For both IgG and BSA measurements without a hydrogel, the median dwell time is within an order of magnitude of the detection threshold. With a hydrogel, experimental data and the COMSOL simulations suggest that a greater fraction of events are captured, and a greater fraction of the dwell time distribution lies above the detection threshold. In contrast, for both IgG and BSA measured with a hydrogel, there is a noticeable positive trend between median dwell time and voltage.

This relationship between median dwell time and voltage with hydrogels matches the model predicted behavior with a rebounding wall boundary condition at the *trans* boundary, suggesting that neither IgG nor BSA can penetrate the hydrogel. In particular, the IgG measurement shown in Fig 3.4 shows an exponential increase in median dwell time with respect to voltage that closely matches the exponential trend predicted by the COMSOL model. The rebounding wall mechanism is particularly interesting because of the relationship between voltage

and dwell times. In traditional nanopore sensing, including nucleic acid sensing with biological and solid state nanopores, high voltages lead to larger event amplitudes and better signal-to-noise, but lower voltages lead to longer residence times which allow looser bandwidth requirements for accurate measurement of the analyte. The opposing effects of these two behaviors makes addressing measurement problems with nanopores difficult. However, in the case of hydrogel facilitated detection of native proteins, it is possible to have detection where increasing voltage yields larger amplitudes and longer dwell times. This unique behavior allows for experimental opportunities previously not possible, specifically in regard to long duration, single molecule measurement. Long duration events should result in reduced amplitude error from noise, which should in turn lead to fewer needed events to produce adequate statistics. This advantage coupled with the increased detection frequency when measuring with hydrogel backed nanopores may lead to significantly increased measurement throughput, reducing times that proteins need to be measured before having enough data for statistical significance.

3.2.4 Increased dwell time sensitivity to diffusivity and electrophoretic mobility

In a traditional nanopore measurement, motion of the protein is affected by diffusion, electrophoresis, and electroosmostic flow, and modeling and experimental data shows that once proteins are inside the nanopore, the probability that the protein completes translocation to the *trans* chamber is much higher than its return to the *cis* chamber for charged proteins.[7,11,14] However, in hydrogel facilitated detection where the hydrogel is impermeable, the electrophoretic force on the protein contributes to the protein capture and diffusion is the only mechanism of escape. As a result of their competitive contribution, changes to the electropohoretic mobility and

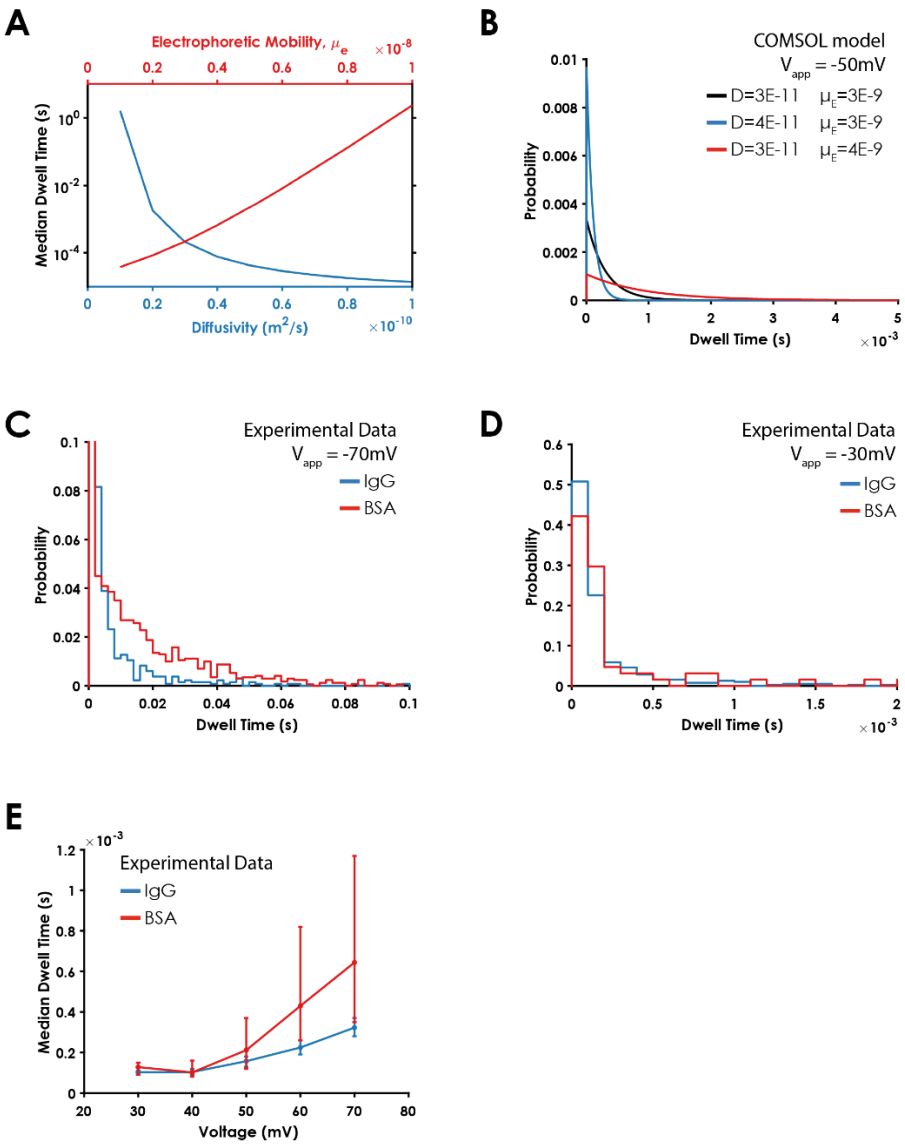


Fig 3.5 Because experimental evidence supports that for IgG and BSA, the protein cannot penetrate the hydrogel, the relationship between protein diffusivity and electrophoretic mobility and the residence time is significantly different than for traditional nanopore sensing. (A) A plot of the median dwell time (log scale) against diffusivity (μ_E fixed at of $3*10^{-9}$ $m^2/(V*s)$), as simulated in COMSOL, shows that as diffusivity increases, the median dwell time decreases exponentially. Conversely, as electrophoretic mobility increases (D fixed at $3*10^{-11}$ m^2/s), the median dwell time increases exponentially. (B) The results suggest that relatively small changes in either of these protein characteristics may lead to observable changes in dwell time. Plotting the COMSOL predicted dwell time probabilities for a protein with a diffusivity of $3*10^{-11}$ m^2/s and electrophoretic

mobility of $3*10^{-9}$ $m^2/(V*s)$ next to dwell time probabilities for a similar protein with a diffusivity of $4*10^{-11}$ m^2/s and another with electrophoretic mobility of $4*10^{-9}$ $m^2/(V*s)$ shows the effect on dwell time of these relatively small changes in physical properties (all at -50 mV applied potential). (C) IgG and BSA were measured sequentially on the same, 23nm diameter nanopore with an applied voltage of 70 mV in 2M KCl, pH10 buffer. Dwell time histograms show that BSA was more likely to have longer events than IgG, which agrees well with models because, although BSA is smaller than IgG (larger diffusivity), it has a much greater charge at pH10 than IgG (higher electrophoretic mobility). (D) At 30mV, the difference in dwell time distribution was significantly attenuated due to the relatively lower contribution of electrophoresis. (E) Plotting the median dwell times for this experiment at voltage between 30 to 70 mV in 10 mV increments highlights these results, showing that at lower voltages, IgG and BSA have similar median dwell times, though at higher voltages, the median dwell time for BSA is increasingly greater than IgG, most likely due to the larger electrophoretic mobility of BSA. Error bars show the 95% confidence interval for each data point as calculated by bootstrapping. The increased width of error bars for BSA is not a result of the number of data points collected as there is no clear correlation between number of points and width of error. Instead, the error bars are more likely related to the shape of the dwell time distribution, which is broader for BSA.

diffusivity of a protein will have relatively greater effect on the residence time of the protein in the nanopore as compared to traditional sensing.

Simulated results in COMSOL show that the dwell time of proteins in the nanopore is affected exponentially by the diffusivity or the electrophoretic mobility (Fig 3.5A). Doubling the electrophoretic mobility of a protein results in an order of magnitude increase in the median dwell time, and extrapolating the plot shows that, as expected, a protein with no charge will still have non-zero dwell time with motion entirely dependent on diffusion. The diffusivity of the protein has an exponential relationship with the log of the median dwell time, showing that, especially for

smaller proteins, the dwell time will change significantly for proteins of different sizes. The exponential relationship with the log of the median dwell time is intuitive because a translocating particle cannot have 0 diffusivity, and therefore, there should be no y-axis intercept. The y-axis is expected to act as an asymptote for the median dwell time.

A simulated dwell time distribution is shown in Fig 3.5B for a particle with a diffusivity of $3*10^{-11}$ m^2/s and electrophoretic mobility of $3*10^{-9}$ $m^2/(V*s)$ at an applied potential of -50mV. Simulated dwell time distributions are also shown for two more particles with the same parameters, though one with a diffusivity $4*10^{-11}$ m^2/s and another with electrophoretic mobility of $4*10^{-9}$ $m^2/(V*s)$. Increasing the diffusivity of a protein while keeping electrophoretic mobility constant (for example by reducing the size while adjusting protein charge to compensate) leads to a dwell time distribution that has increased probability at lower dwell times. Increasing electrophoretic mobility (for example by increasing the charge) results in a dwell time distribution that has lower probability at lower dwell times and extends to longer dwell times. The changes in modeled dwell time distribution due to varying electrophoretic mobility and diffusivity reflect the fact that, under the rebounding wall mechanism, an increased electrophoretic mobility elongates protein residence time while an increased diffusivity reduce overall residence time, though the shape of the dwell time distribution remains unchanged. This latter point is significant as it shows that dwell times cannot be used alone to identify proteins in a mixture, unlike amplitudes which we previously used to identify IgG and BSA in mixture. The dwell time distributions for each protein overlap too significantly to associate given events with particular proteins by dwell time alone. However, given the expected sensitivity of the dwell time distribution to protein characteristics, it is possible that dwell time can be used in conjunction with amplitude to make a higher confidence analysis of protein mixtures.

We measured IgG and BSA sequentially on the same 23nm diameter nanopore to investigate if dwell time distributions between the two proteins would be significantly different due to their differing physical characteristics. As was shown previously, in traditional sensing experiments, the majority of translocation events evade capture, and as a result dwell time distributions do not typically offer any useful information about the protein.[11] At pH 10, we expect IgG to have a diffusivity of $4.2*10^{-11}$ m^2/s and electrophoretic mobility of $2.5*10^{-9}$ $m^2/(V*s)$ and BSA to have a diffusivity of $5.52*10^{-11}$ m^2/s and electrophoretic mobility of $4.48*10^{-8}$ $m^2/(V*s)$, calculated as shown previously[11] from measured protein diameters and charges.[21–26] Dwell time distributions for the two proteins with a 70mV applied potential are shown in Fig 3.5C, and BSA has a higher probability of longer dwell time events as compared to IgG. Fig 3.5D shows data from the same experiment, but at 30mV, and the dwell time distributions are no longer distinguishable. The results agree well with predictions from the model because we expect electrophoretic contribution to scale with voltage, but the diffusive contribution should remain constant regardless of voltage. Although BSA is smaller than IgG and therefore has a larger diffusivity, the difference in charge at pH 10 results in a much greater difference in electrophoretic mobility. At 70 mV, the difference in electrophoretic mobility results in an observable different in dwell time due to the increased electrophoretic mobility of BSA. However, at 30mV, the electrophoretic contribution diminishes, and the effect of BSA's greater electrophoretic mobility is also reduced, yielding distributions that are very similar.

Fig 3.5E shows the median dwell times for both proteins from 30 to 70 mV with 10 mV steps, further showing the effect of the difference in electrophoretic mobility. Both proteins show increasing dwell time with respect to voltage, agreeing with previous results shown in Fig 3.4. However, the rate at which the median dwell times increases in greater for BSA due to the protein's

greater electrophoretic mobility, which results in median dwell times that are greater for BSA than IgG at larger voltages. Another significant difference is the width of the error bars, which are the 95% confidence intervals determined from bootstrapping the median. The larger error bars for the BSA data points are not a result of a smaller number of detected events. The event count is highest at 70mV for both proteins and BSA has a greater number of events than IgG at that voltage, and a comparable number of events at the remaining voltages (with an order of magnitude). The greater variance in median dwell time is most likely related to BSA's broader dwell time distribution, which in turn may be related to BSA's greater diffusivity in comparison to IgG.

3.2.5 Protein-pore interaction and effect on dwell time

The experimental data suggests that proteins cannot penetrate the hydrogel, and therefore, the presented models that have a rebounding boundary condition at the *trans* boundary of the nanopore apply to this measurement system. However, these models only account for diffusion and electrophoresis and, if needed, can be adapted to account for electroosmotic flow. A significant omission from these models is the effect of protein-pore interaction, which has been shown previously to have a significant effect on observed dwell times.[10,19,20] To study this behavior with hydrogel facilitated detection of proteins, we measured 10nM IgG with a 23 nm diameter nanopore with a PEG-DMA hydrogel in the *trans* compartment at -60 mV. Initially, the measurement buffer was 2M KCl, pH 9.5, and measurements were conducted until several hundred current blockades were observed. The buffer in the *cis* chamber was changed to 2M KCl, pH 7 with 10nM IgG, and data was further collected until several hundred blockade currents were observed.

The empirical cumulative distribution function of dwell times for both the pH 9.5 and pH 7 conditions are plotted in Fig 3.6. The distribution functions show that at pH 7, dwell times are

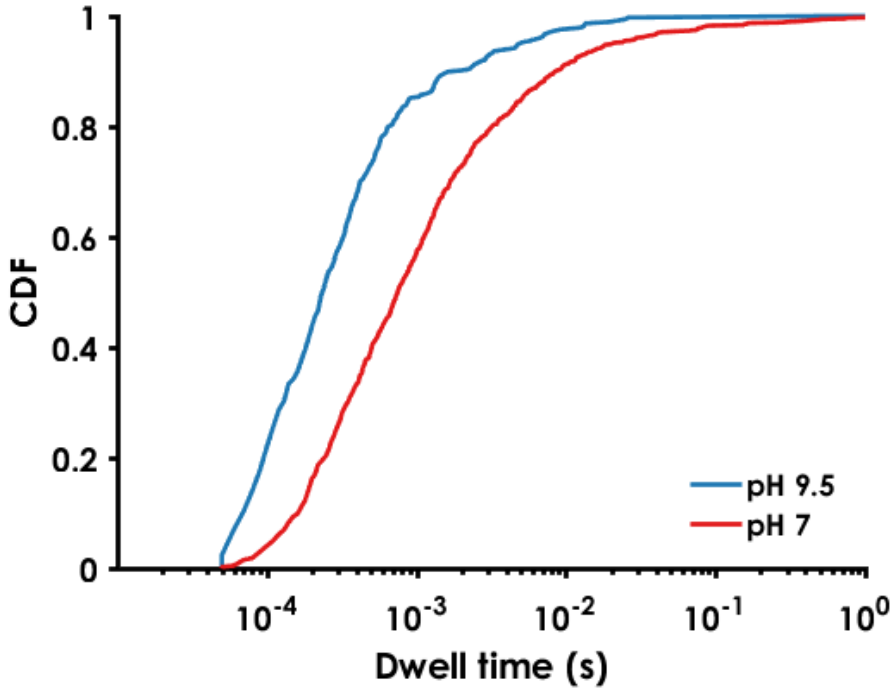


Fig 3.6 Measurement of IgG with a hydrogel backed nanopore at pH 9.5 and pH 7 shows that IgG typically has longer dwell times at pH 7 than at pH 9.5. IgG was measured with a 23 nm diameter, hydrogel backed, nanopore at -60 mV in 2M KCl, 10mM Tris-HCl, pH 9.5 buffer and subsequently in 2M KCl, 10mM Tris-HCl, pH 7 buffer. Empirical cumulative distribution functions are shown for dwell times at the pH 9.5 condition (blue) and pH 7 condition (red), showing that dwell times for IgG at pH 7 are typically 3-4 times longer than for IgG at pH 9.5. The longer dwell time at pH 7 is most likely a result of increased protein-pore interaction at that pH as compared to pH 9.5.

more likely to be longer than at pH 9.5, with dwell times from the pH 7 data likely 3-4 times larger than dwell times from the pH 9.5 data. This result directly contradicts the theoretical predictions. Because IgG has an isoelectric point of ~7-8,[24,27] IgG is expected to be negatively charged at pH 9.5 and neutral to slightly positively charged at pH 7. As a result, at pH 9.5, IgG should have a greater electrophoretic mobility than at pH 7, and the model predicts that the residence time of IgG should be greater at pH 9.5.

However, the negatively charged SiN surface (pI ~ 5)[14] would repel IgG at pH 9.5 and would have very low electrostatic interaction with IgG at pH 7. We expect that the differences in dwell time distribution between the pH 9.5 and pH 7 condition are a result of the increased protein-pore interaction at pH 7. The result shows that protein-pore interactions can have significant impacts on residence time, and that the model presented here is not a fully robust model. This model can act as a theoretical guide to interpretation of molecule behavior with hydrogel facilitated detection. It is possible that the model would have greater accuracy in predicting dwell times for lipid coated nanopores, where protein-pore interaction is greatly reduced.[28,29]

3.3 Conclusion

This chapter explores the possibility of two mechanisms for the increase of protein residence time in nanopores when interfaced with a hydrogel - one in which the mobility of proteins is reduced due to the presence of the gel, and one where the gel prevents translocation of proteins and forces proteins to diffuse back into the *cis* chamber. Measurement of IgG and BSA showed that both proteins had increasing dwell times at increasing voltages, suggesting that neither protein can penetrate the PEG-DMA hydrogel. This finding is particularly interesting because it provides an opportunity in nanopore sensing that increasing the applied voltage can lead to larger dwell times and larger amplitudes. In conventional sensing, changes to voltage lead to a tradeoff between amplitude and dwell time. Hydrogel facilitated detection, specifically with solid-state nanopores, may benefit from high voltage measurement as a single molecule can be measured for very long duration without sacrificing signal-to-noise.

It is important to note that the model presented here only acts as a guideline for hydrogel facilitated measurements. Measurements of neutral and negatively charged IgG showed that

protein-pore interaction can have a significant effect on residence time. Future work studying lipid coated pores interfaced with hydrogels may allow the theoretical models to be more accurate, as protein-pore interaction would be significantly diminished. Small protein measurement has yet to be explored as well, and it is likely that proteins smaller than the hydrogel's pore size will penetrate the hydrogel. In this case, we would expect that the conventional model with an absorbing boundary condition at the *trans* boundary would apply, with the protein's mobility being reduced due to the presence of the gel. Further work on the relationship between the size of the protein and the hydrogel mesh size will help in predicting and interpreting measurements.

3.4 Materials and Methods

All materials were purchased from Sigma-Aldrich unless otherwise noted and used as received unless otherwise noted. Immunoglobulin G 1 (IgG) and bovine serum albumin (BSA) were dissolved in 50% glycerol, 50% 5 mM NaCl, pH 7 and stored at -20°C at 40-50 µM.

3.4.1 COMSOL model

All COMSOL simulations were run with COMSOL 5.3. The model consisted of a 1-D geometry with two regions, one from 0 to 15 nm representing the pore, and one from 0 to -25nm representing the *cis* side of the nanopore where an electric field is present. All physics were handled by the transport of diluted species module in COMSOL, which tracks the concentration distribution of particles under diffusion and convection. To translate the results from the transport of diluted species, the computed concentration distributions were normalized such that the area under the curve at t=0 was equal to 1. The convection velocity over the full computation space was equal to $\mu_E*V/(l_p+l_A)$, where μ_E is the electrophoretic mobility of the protein, V is the applied voltage, l_p is

the pore length, and l_A is the *cis* side region representing the access resistance. The concentration at the x=-25nm was fixed to 0 mol/m^3 to represent an absorbing wall. At x=15nm, either the concentration was fixed to 0 mol/m^3 for absorbing wall boundary condition simulations or set to be a no flux boundary for reflecting wall boundary condition simulations. A gaussian pulse with a standard deviation of 0.025 nm was centered at x=0 as the initial condition.

The mesh nodes were located at 0.05nm intervals over the entire space. For simulations that reached steady state in less than 1 ms, computed time values were set linearly from 0 to 1ms with 10^{-8} s intervals. For simulations that reached steady state at times between 1 ms and 100 s, computed time values were set logarithmically from 10^{-8} s to an appropriately long time value, with 500 to 700 times computed per decade. In occasional cases where steady state was reached at times greater than 100 s, the same logarithmic spacing was used as described, though only 300 time points were computed per decade. Two domain probes were used in COMSOL to calculate the integral of the concentration in the nanopore region and over the full space. These values along with the concentration distribution at the given space and time points were saved.

The values were imported into MATLAB 2016a to determine dwell time distributions and medians. To determine the dwell time distribution and medians, the integrals of concentration in the nanopore were first normalized to the maximum value of the area under the curve of the initial condition. The normalization turns the integrals of concentration into the probability over time that the particle is inside the pore. 1 minus this probability gave the cumulative distribution function of the probability that the particle was not inside the pore, and can be interpreted as the cumulative distribution function of the dwell time. The median was the interpolated dwell time for which the

distribution function equaled 0.5. The derivative of this distribution function was the dwell time probability distribution, and values of the dwell time distribution at the shortest dwell times that were negative were set to 0.

3.4.2 Formation of Nanopores

Experimental measurements were conducted as described previously. Briefly, commercially available 15nm thick silicon nitride TEM windows (TedPella, Norcada) were coated with PDMS on the silicon nitride face of the chip, leaving a small area of approximately 100 x 100 μm, to reduce capacitive noise. The chip was mounted in a teflon chamber which was then filled with 2mL 2M KCl, 10mM Tris-HCl, pH 9.5 on both sides of the chip. A LabView setup including an NI USB-6251, an op-amp circuit to double the applied voltage from the LabView DAQ, and a current to voltage op-amp circuit was first used to apply voltages from -500mV to 500mV while measuring the resultant currents to verify that these currents were <20 pA, ensuring that the window was intact and that there was no pinhole formation.[30] For hydrogel facilitated measurements, a solution of degassed 10% polyethylene glycol (1000 MW) dimethacrylate (PEG-DMA) (PolySciences, Inc.) in 2M KCl, 10mM Tris-HCl, pH 9.5, 10mM ammonium persulfate in water, and 10mM tetramethylethylenediamine were mixed briefly, and 1.4 mL of this solution was used to replace the buffer in the measurement chamber adjacent to the silicon nitride face of the chip (*trans* side of the chamber). The hydrogel solution was allowed to polymerize for 10 minutes before 600 μL of measurement buffer was added to the same side of the chamber, ensuring that the hydrogel did not dry over the course of the experiment.

Nanopores were formed by the dielectric breakdown method.[31] Using the same LabView controlled measurement setup, a voltage of -5 to -7 V was applied until breakdown was observed, indicated by a sudden increase in conductance. 100ms pulses of 3 to 6V were applied to enlarge the pore until the conductance indicated that the pore had reached the desired size, typically 20-30nm in diameter.

In all measurements, the silicon nitride face of the nanopore was referenced as ground. A short Ag/AgCl wire also bridged the two halves of the measurement chamber at all times that measurements were not being taken to prevent pinhole formation.[30]

3.4.3 Measurement of Protein

All nanopore measurement of proteins was conducted with an Axopatch 200B amplifier, Digidata 1440B or 1322A, and the Clampex software suite at 100kHz and a 10kHz hardware lowpass filter. Measurements were conducted in 2M KCl, 10mM Tris-HCl, pH 9.5 or pH 7. Upon formation and growth of the nanopore, baseline currents were measured at measurement voltages to ensure that no non-specific events were present and that baseline currents were stable for at least 10 minutes. Applied voltage protocols included a 5 second segment where the measurement voltage was held, typically between -20 and -70 mV, and a 2 second segment at +180 mV to electrophoretically eject and proteins and reduce the chance of irreversible clogging of the nanopore. The sequence was repeated, cyclically applying different measurement voltages during each iteration. After the 10 minutes of baseline measurements, immunoglobulin G (IgG) or bovine serum albumin (BSA) was added to the *cis* chamber such that the final concentration of protein was 10nM. In experiments were multiple proteins were measured sequentially with the same nanopore, the solution in the *cis*

chamber was replaced with measurement buffer between protein additions to ensure that no residual events were detected from the previous measured protein.

All data was analyzed with MATLAB 2016a, with in-house scripts (described in detail in Chapter 4) that identify events and record their amplitudes and dwell times. Error bars for median vs voltage plots were generated by bootstrapping the dwell time distributions, with n=10000, and taking the 5% and 95% quantiles of the median distribution as the lower and upper bounds of the error bars.

1. Benjamin, C. J. *et al.* Nonfouling NTA-PEG-Based TEM Grid Coatings for Selective Capture of Histidine-Tagged Protein Targets from Cell Lysates. *Langmuir* **32**, 551–559 (2016).

2. Bernhard, C. *et al.* Repelling and ordering: The influence of poly(ethylene glycol) on protein adsorption. *Phys. Chem. Chem. Phys.* **19**, 28182–28188 (2017).

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

4. Squires, A. H., Hersey, J. S., Grinstaff, M. W. & Meller, A. A Nanopore–Nanofiber Mesh Biosensor To Control DNA Translocation. *J. Am. Chem. Soc.* **135**, 16304–16307 (2013).

5. Tang, Z. *et al.* Gel mesh as "brake" to slow down DNA translocation through solid-state nanopores. *Nanoscale* **7**, 13207–13214 (2015).

6. Al Sulaiman, D., Cadinu, P., Ivanov, A. P., Edel, J. B. & Ladame, S. Chemically Modified Hydrogel-Filled Nanopores: A Tunable Platform for Single-Molecule Sensing. *Nano Lett.* **18**, 6084–6093 (2018).

7. Talaga, D. S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* **131**, 9287–9297 (2009).

8. Ling, D. Y. & Ling, X. S. On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrödinger's first-passage-time theory. *J. Phys. Condens. Matter* **25**, 375102 (2013).

9. Cox, D. R. (David R. & Miller, H. D. (Hilton D. *The Theory of Stochastic Processes*. (Methuen, 1965).

10. Larkin, J., Henley, R. Y., Muthukumar, M., Rosenstein, J. K. & Wanunu, M. High-bandwidth protein analysis using solid-state nanopores. *Biophys. J.* **106**, 696–704 (2014).

11. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).

12. Hall, J. E. Access Resistance of a Small Circular Pore. *J. Gen. Physiol.* **66**, 531–532 (1975).

13. Hyun, C., Rollings, R. & Li, J. Probing Access Resistance of Solid-State Nanopores with a Scanning-Probe Microscope Tip. *Small* **8**, 385–392 (2012).

14. Firnkes, M., Pedone, D., Knezevic, J., Döblinger, M. & Rant, U. Electrically facilitated

translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* **10**, 2162–2167 (2010).

15. Ai, Y., Zhang, M., Joo, S. W., Cheney, M. A. & Qian, S. Effects of electroosmotic flow on ionic current rectification in conical nanopores. *J. Phys. Chem. C* **114**, 3883–3890 (2010).

16. Lin, D.-H., Lin, C.-Y., Tseng, S. & Hsu, J.-P. Influence of electroosmotic flow on the ionic current rectification in a pH-regulated, conical nanopore. *Nanoscale* **7**, 14023–14031 (2015).

17. Laohakunakorn, N. & Keyser, U. F. Electroosmotic flow rectification in conical nanopores. *Nanotechnology* **26**, 275202 (2015).

18. van der Heyden, F. H. J., Stein, D. & Dekker, C. Streaming Currents in a Single Nanofluidic Channel. *Phys. Rev. Lett.* **95**, 116104 (2005).

19. Balme, S. *et al.* Influence of adsorption on proteins and amyloid detection by silicon nitride nanopore. *Langmuir* **32**, 8916–8925 (2016).

20. Niedzwiecki, D. J. & Movileanu, L. Monitoring Protein Adsorption with Solid-state Nanopores. *J. Vis. Exp.* e3560 (2011). doi:10.3791/3560

21. Sung, J. H., Chun, M.-S. & Choi, H. J. On the behavior of electrokinetic streaming potential during protein filtration with fully and partially retentive nanopores. *J. Colloid Interface Sci.* **264**, 195–202 (2003).

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

23. Chun, M.-S., Il Cho, H. & Song, I. K. Electrokinetic behavior of membrane zeta potential during the filtration of colloidal suspensions. *Desalination* **148**, 363–368 (2002).

24. Mathes, J. & Friess, W. Influence of pH and ionic strength on IgG adsorption to vials. *Eur. J. Pharm. Biopharm.* **78**, 239–247 (2011).

25. Yadav, S., Shire, S. J. & Kalonia, D. S. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **28**, 1973–1983 (2011).

26. Fologea, D., Ledden, B., McNabb, D. S. & Li, J. Electrical characterization of protein molecules by a solid-state nanopore. *Appl. Phys. Lett.* **91**, 5–8 (2007).

27. Danielsson, Å., Ljunglöf, A. & Lindblom, H. One-step purification of monoclonal IgG

antibodies from mouse ascites: An evaluation of different adsorption techniques using high performance liquid chromatography. *J. Immunol. Methods* **115**, 79–88 (1988).

28. Yusko, E. C. *et al.* Controlling protein translocation through nanopores with bio-inspired fluid walls. *Nat. Nanotechnol.* **6**, 253–260 (2011).

29. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* acsnano.8b09555 (2019). doi:10.1021/acsnano.8b09555

30. Matsui, K., Yanagi, I., Goto, Y. & Takeda, K. I. Prevention of dielectric breakdown of nanopore membranes by charge neutralization. *Sci. Rep.* **5**, 17819 (2015).

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

Chapter 4

Nanopore measurement data analysis method based on Hidden Markov models and Canny edge detection

4.1 Introduction

Noise in nanopore measurements results in variance in the measured blockade amplitudes, and as a result, most studies rely on the measurement of a large population of events to accurately quantify their data. Analysis of a large population of events is even more essential for shape dependent measurements and dwell time analysis (such as those shown in Chapters 2 and 3) where the entire distribution of event characteristics is important, rather than just the mean. To produce adequate statistics, typically hundreds to tens of thousands of events must be acquired and quantified. Parsing nanopore measurements to identify and quantify these events is almost always performed computationally as analysis by hand is impractical given the large number of blockade events found in most experiments.

A variety of computational techniques are used to analyze nanopore measurements. Most studies use analysis methods that involve a threshold, set at a fixed distance from the open pore baseline, to identify current blockades (Fig 4.1).[1–4] Thresholding algorithms are excellent methods of data analysis for nanopore measurement as they are typically simple, effective, and computationally fast. However, these methods inherently cannot analyze data where the signal-to-noise ratio (SNR) is not much greater than 1, as a satisfactory threshold will not exist to separate blockade currents from the baseline if the data is too noisy (Fig 4.1B). Furthermore, thresholding

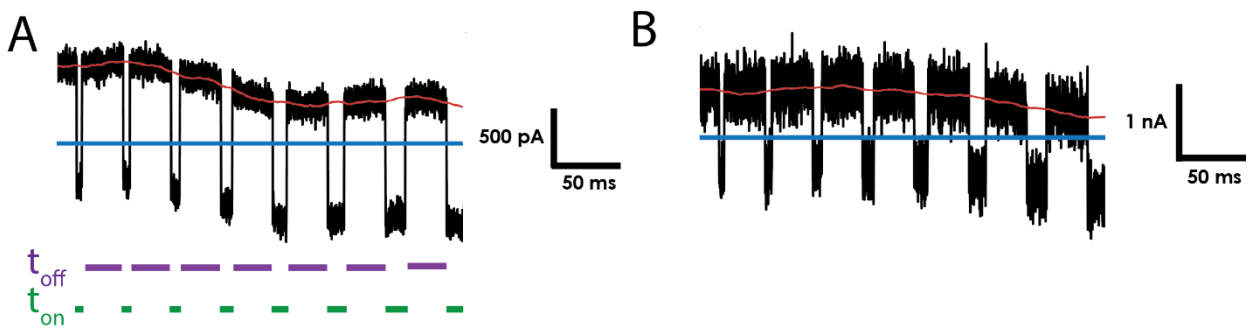


Fig 4.1 (A) Simulated data that includes capacitive noise (high frequency noise that raises the noise floor) and flicker noise (low frequency noise that appears as baseline fluctuations) is shown, with the underlying, open pore baseline, without high frequency noise, shown in red. The time that the pore is open, t_{off}, and the time that the pore is occupied,t_{on}, are indicated below the data in purple and green marks respectively. The effect of flicker noise can be clearly seen as the fluctuations in the red line. A baseline without flicker noise would be perfectly flat. In data with large signal amplitude compared to noise, a threshold can be used to identify events. Here, a suitable threshold is shown in blue, and all data below the threshold would be considered an event. Analysis with a threshold on this data would identify events without error (B) A simulated, trace is shown with similar flicker noise but significantly increased capacitive noise (event amplitudes and durations are the same as in A). Threshold analysis is not possible on this data as noise spikes frequently cross the threshold, causing the incorrect identification of too many events. Furthermore, flicker noise in this trace causes the baseline to fall to a value near the threshold near the end of the trace. This would lead to dramatic over detection of events as the threshold crosses the noise floor of the baseline. This is demonstrated further in Fig 4.2.

methods rely on accurate tracking of the open pore baseline, which can fluctuate as a result of low frequency (flicker) noise[5,6] or drift if the pore size is unstable. Flicker noise exhibits 1/f spectral density and in nanopore measurement, has most significant effect on the signal in the 1-1000 Hz range. In measurement, significant flicker noise results in fluctuating baselines and higher noise

floors. Incorrect baselines can lead to incorrect thresholds, which in turn can lead to significant error in identification of events. Fig 4.1A and B use simulated data to show the effects of flicker noise and capacitive noise when analyzing data with thresholding. Methods to account for baseline fluctuation used in conjunction with thresholding compose more robust analysis methods than thresholding alone.

The analysis methods used in Clampfit (Molecular Devices)[7] and Transalyzer, a MATLAB script published by Plesa *et al.*,[2] both account for baseline fluctuations or drift over the course of measurement. Both methods adopt some form of a moving average to estimate the baseline, however, these averaging methods alone can fail to estimate the baseline accurately in certain data sets. In order for baselines the be estimated accurately by averaging, the time that the pore is open (t_{off}) must be much greater than the time that the pore is occupied (t_{on}), or the estimated baseline will deviate towards the blockade current level. Similarly, in a moving average calculation, if the baseline shifts significantly within the window, a single average value will incorrectly approximate the baseline over the window. As a result, data sets that have lower SNR and those that have long duration events relative to the time scale of the baseline fluctuations can cause averaging techniques to poorly estimate the baseline. Both Clampfit and Transalyzer employ additional methods to account for baseline fluctuation and long duration events when analyzing data with significant flicker noise or baseline drift. However, these additional baseline estimation techniques still rely on thresholding, and as mentioned previously, thresholding methods can fail to accurately parse noisy data sets. As a result, Clampfit and Transalyzer can analyze data with either long duration events or significant flicker noise, but struggle to analyze data with both long duration events and significant flicker noise. In some cases, this poor baseline estimation leads to incorrect quantification of individual event amplitudes or dwell times. In more severe cases, the poor

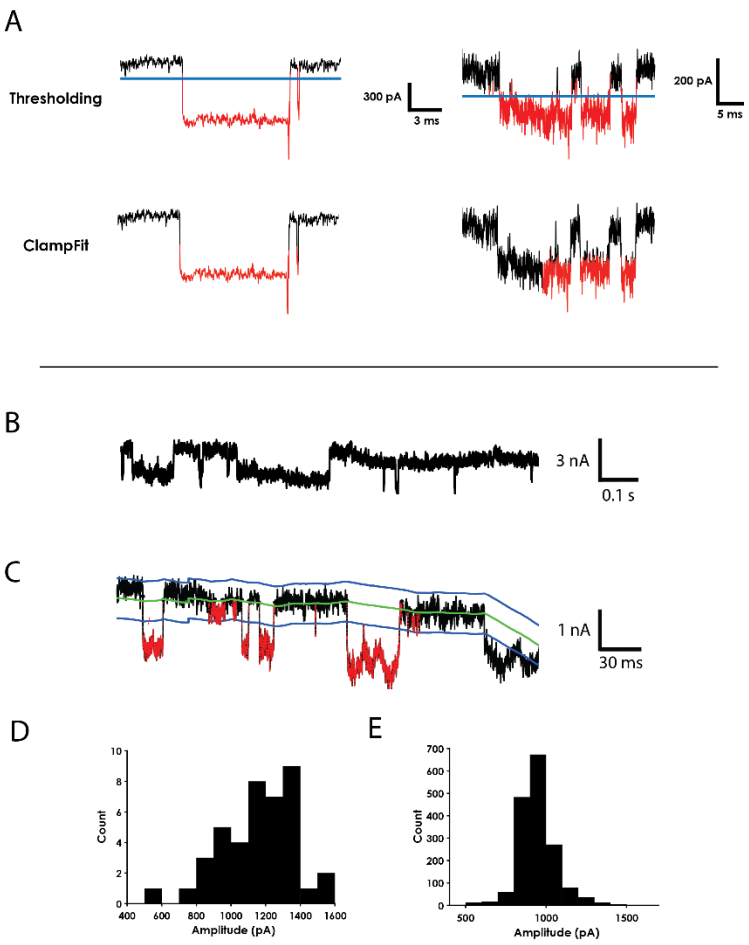


Fig 4.2 Threshold based methods are the most common approaches to analysis of nanopore measurements. (A) Analysis is shown for a simple thresholding analysis (top) and for Clampfit (bottom) which also uses thresholding, but includes baseline and event level adjustment (blue is the threshold, black is identified baseline, red is identified as an event). Data in the left column is of 10nM IgG measured with a nanopore interfaced with a PEG-DMA hydrogel in 2M KCl, pH 10 at -50mV (data is normalized to positive currents for consistency throughout this chapter). This data segment shows successful analysis by both methods. Data on the right is measurement of 10nM BSA with a hydrogel interfaced nanopore in 2M KCl, pH 10 at -20mV (data is normalized to positive currents for consistency throughout this chapter). For this data segment, both methods incorrectly considered noise during the event as open-pore baseline current, leading to no correctly identified events and many false-positive detections. (B) A short segment of IgG measured with a PEG-DMA backed nanopore at -70mV in 2M KCl, ph 9.5 is shown. This data is used here and later in this

chapter as an example measurement that shows clear events but has significant capacitive and flicker noise. (C) A segment of data from the same experiment as the data in (B) is shown and analyzed by Transalyzer, which first uses an iterative method using a moving average, threshold, and event subtraction to estimate the baseline (green line). Events are identified in a second thresholding step (blue line is the threshold). The red dots indicate data points that were assigned to events. Transalyzer produces a mostly accurate approximation of the baseline, but small errors in the baseline estimate coupled with noise lead to many incorrect events. (D) ~7 seconds of data from this experiment were analyzed by hand as a control analysis method, yielding 43 detected events. A histogram of the average amplitude of these events is shown, with a most probable amplitude of ~1350 pA and a non-normal distribution (E). Analysis of this same segment of data by thresholding yields 1649 events. The histogram of the average amplitudes is shown, with a most probable amplitude near 950 pA and a normal distribution.

baseline estimation can lead to significant under- or over-detection of events. For data where the event amplitudes are similar to or less than the noise, all threshold-based methods (such as Clampfit and Transalyzer) cannot discern events accurately without low pass filtering, though filtering can reduce amplitudes of short duration events making them indistinguishable from noise.[8,9] Example analysis of experimental data is shown in Fig 4.2, showing successful and unsuccessful analysis with thresholding methods.

In addition to thresholding methods, other approaches to parsing nanopore measurements have been demonstrated.[3,4] In particular, Hidden Markov models (HMMs) are well suited for nanopore analysis, being frequently used for intraevent analysis of nucleic acid sequencing data.[10–12] HMMs have also been shown previously for data analysis of high-bandwidth measurements where SNR is low.[13] This method was very effective for parsing noisy measurements, but it was

demonstrated on measurements with biological channels, where baseline fluctuations are very low. In solid-state nanopore measurement, baseline fluctuations can be significant in relation to event amplitude, and HMMs will fail to identify events accurately in data sets with significant baseline fluctuations. HMM analysis of nanopore measurements is discussed in greater detail later in this chapter.

We have developed a program that improves the parsing of nanopore data by addressing issues in thresholding and HMM-based analysis methods. Our method uses unsupervised HMM and a highly modified Canny edge detection[14] together to estimate a baseline and accurately identify event locations in noisy data traces with short and long duration events. HMM and Canny edge detection work together to addresses weaknesses in each technique to yield a robust analysis algorithm. Here, we discuss the implementation of HMM and Canny edge detection separately, and also show a data analysis algorithm that exploits both methods. We demonstrate these analysis methods on measurements of protein with nanopores interfaced with polyethylene glycol (1000MW) dimethacrylate (PEG-DMA) hydrogels. These measurements typically exhibit more noise than conventional solid-state nanopore measurements and event durations range from tens of microseconds to over ten milliseconds, making these measurements difficult to analyze by previous analysis methods. The algorithm presented here accurately parses these data sets, demonstrating the robustness of the method.

4.2 Results and Discussion

4.2.1 Common methods of nanopore data analysis

Most commonly used methods of nanopore data analysis rely on thresholding, defining events by regions where the measured current exceeds a fixed value above the baseline. As

discussed previously, thresholding methods can fail when analyzing noisy data sets. Below, we use Clampfit[7] and Transalyzer,[2] both using thresholding in conjunction with other methods, as examples to demonstrate limitations of threshold analysis (Fig 4.2).

Clampfit, a data analysis software included in the pClamp (Molecular Devices) software suite used commonly in nanopore research, identifies events by comparing currents to user defined values for the baseline and the event amplitude, marking the location of events by a threshold placed halfway between these two current values.[7] Clampfit handles drifting baselines and flicker noise by updating the defined baseline and event levels while parsing through the trace, setting the baseline level as the average current between any two events. This method leads to accurate detection of events if the baseline does not shift significantly between events but is not accurate if events are spaced significantly apart or there is significant flicker noise. Also, in our experience, for events that are particularly short (< 4 data points), Clampfit does not accurately mark the start and end of events, leading to incorrect recorded dwell times and amplitudes. Lastly, as this method relies on a threshold to identify events, event amplitudes that are near to the noise floor of the baseline can lead to poor event identification.

Plesa *et al.* have also developed an effective tool in MATLAB to analyze nanopore data, Transalyzer.[2] Transalyzer calculates a moving average to approximate baseline currents. A simple moving average method to determine baselines can have significant error if the time that the pore is occupied (t_{on}) is similar to or larger than the time that the pore is free (t_{off}). To address this, Transalyzer calculates the baseline is multiple steps. The first step involves a simple moving average to provide an initial guess at the baseline and the location of events. In subsequent steps, the regions that are previously determined to be events are replaced by the value of the calculated baseline at the start of the event, effectively removing the events from the current trace. This

method is very effective for baselines that do not drift significantly. However, when flicker noise is significant, the moving average window must be shortened to ensure that the baseline is correctly approximated, but if an event is longer than the moving average window, the baseline will not be correctly calculated and Transalyzer will miss the event. Therefore, Transalzyer will fail to correctly locate events if the baseline exhibits significant flicker noise or if the event durations are long enough to exceed the moving average window length. As with all thresholding algorithms, Transalyzer will also fail to correctly identify events if the event amplitudes near the noise floor of the baseline.

Two example data traces analyzed by Clampfit and Transalyzer are shown in Fig 4.2A. For traces where the baseline has relatively low noise and event amplitudes clearly exceed the noise floor, both methods can effectively identify blockade events and their boundaries. If the SNR decreases, flicker noise increases, or t_{on} to t_{off} ratio increases, these methods are unable to accurately identify event locations from the baseline, resulting in analysis that can have significant error. Fig 4.2B shows a short segment of data from a measurement of IgG with a hydrogel backed nanopore at -70mV in 2M KCl, pH 9.5. This data set is used throughout this chapter as an example data set that shows clear events, but the presence of significant flicker noise and capacitive noise leads to poor analysis with most methods. A segment of this data from this experiment, analyzed by Transalyzer, is shown in Fig 4.2C. Transalyzer's method of subtracting events to estimate the baseline is mostly effective, but because the method uses thresholding and a moving average, small deviations in the baseline estimation and significant noise in the data lead to a significant number of errors in detection. Approximately 7 seconds of data from this experiment was analyzed by hand as a benchmark analysis method and also analyzed by thresholding, setting the threshold at 3.5 the standard deviation of the baseline currents. The threshold was determined iteratively to produce

the least error. The 7 second data set contained 43 events, but when analyzed by thresholding, 1649 events were detected with only 2 events correctly identified. This massive over-detection is a result of the high flicker noise and low SNR, which results in frequent misidentification of noise as events, regardless of where the threshold is set. Histograms of the average amplitude as found by both analysis methods are shown in Fig 4.2D and E, demonstrating the significant error caused by the poor analysis with the thresholding method.

The threshold-based methods of nanopore data analysis have been very effective for many types of solid state nanopore data. For example, nucleic acid measurements are often analyzed successfully by thresholding because these events are typically short in duration relative to the time in between events, and the blockade amplitudes can far exceed the capacitive noise of solid state nanopores. We have recently demonstrated that a hydrogel placed on the distal side of a nanopore can significantly increase the duration of protein translocation events, from ~20-40 μs to 40-10000 μs (as shown in Chapters 2 and 3). The presence of the hydrogel also increases the noise as well, and as a result, thresholding methods are not effective for identification of these long duration protein events (as shown previously in Fig 4.2). Proteins are also able to induce varying current blockades within a translocation event due to shape-dependent effects, as shown previously,[15,16] which can additionally make it difficult to accurately identify single events with a threshold based analysis method. These orientation changes coupled with baselines that can fluctuate significantly make Clampfit ineffective at analyzing this data. The large t_{on} relative to t_{off} makes it difficult for Transalyzer to correctly identify the baseline, even through iterative baseline recalculation. As mentioned previously, both methods cannot successfully analyze measurements where the SNR is near or less than 1.

4.2.2 Hidden Markov modeling

We address this problem by utilizing Hidden Markov Models (HMMs), which are well suited for nanopore measurements[13,17] as these measurements can be considered Markov processes. In a Markov process, the probability of the value of any given data point is dependent only on the data point immediately preceding it. Data from nanopore measurements does not exhibit any "memory" and therefore can be considered a Markov chain; the value of the next data point depends only on whether, at the previous data point, a particle was inside the pore and the chance that a particle leaves/enters the pore. Markov models can be defined by the possible states, the emission probabilities of those states, and the probabilities that, given a data point, the subsequent data point remains in the same state or switches states (transition probabilities).

In the case of typical nanopore measurement, the system has two states, open pore (baseline) and occupied pore (event), where the emissions are the measured currents (Fig 4.3A). Each state, baseline and event, has unique emissions with different means and distributions (measured current distributions) that describe the noise and characteristics of those states. For example, for a perfectly flat baseline with capacitive noise in nanopore measurement, the emission probability distribution for this state would be a gaussian distribution with the mean as the mean of the baseline and the standard deviation of the distribution set dependent on the noise characteristics. Similarly, an event will have its own unique emission distribution with a mean and possibly standard deviation different from the baseline emission distribution. The t_{on} and t_{off} characteristics of nanopore data can be translated to transmission probabilities, where, for example, a data set with a low t_{off} would have a higher transition probability from baseline to event state and a lower transition probability of remaining in the baseline state.

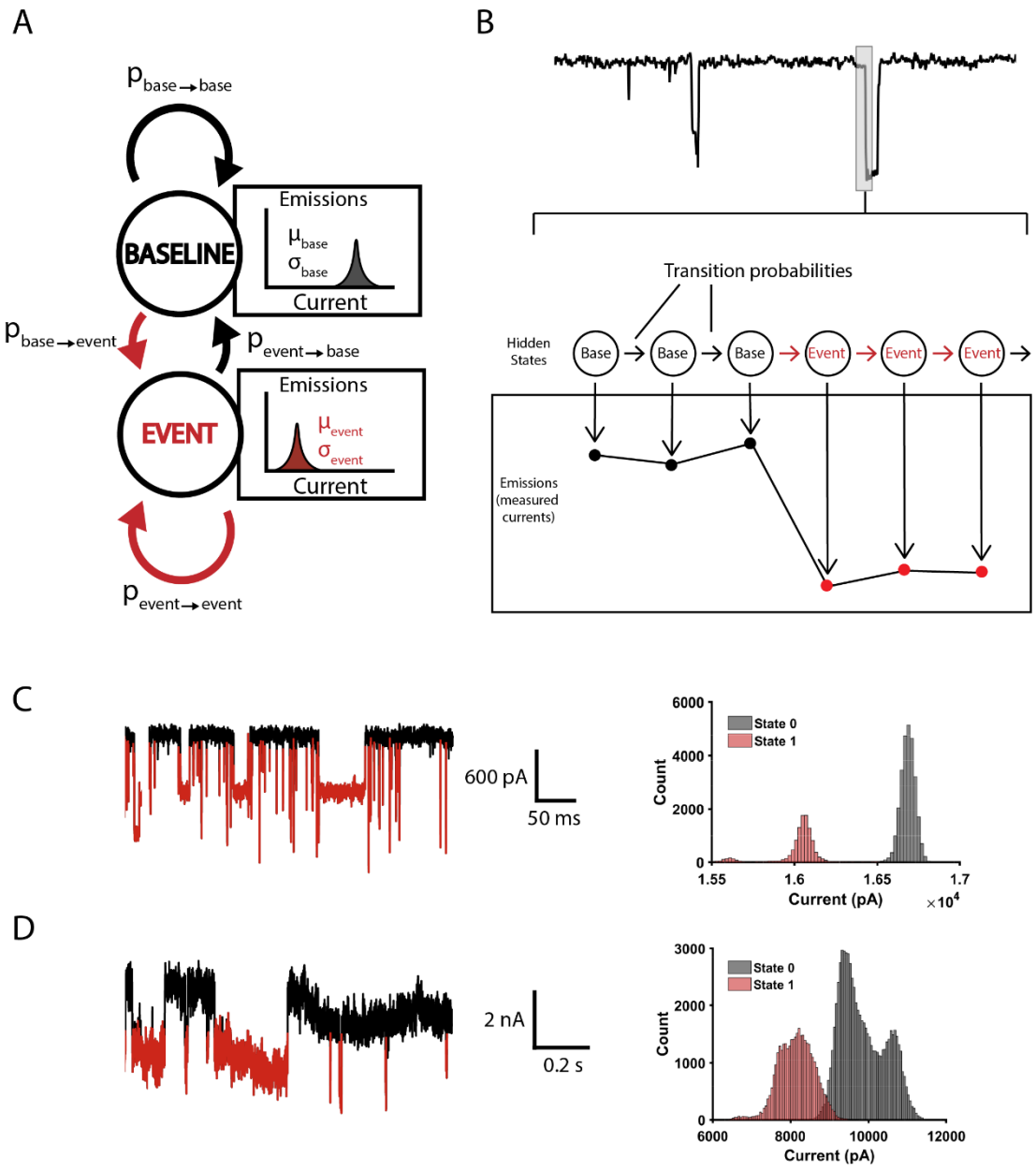


Fig 4.3 HMMs are well suited for analysis of nanopore measurements. (A) HMMs can categorize nanopore data into two states, baseline and event, where both states have unique emission (measured current) distributions with unique means (μ) and variances (σ). Specifically, because captured particles will reduce current, the emission spectra of the event state will have a reduced mean compared to the baseline state. The probability that for any data point, the subsequent data point switches states can be given by the transition probabilities (p). (B) An unsupervised HMM using an

expectation maximization algorithm can take just the measured currents (emissions) and train the HMM parameters (μ and σ for each state, transition probabilities) and also infer the hidden states of the data (typically baseline and event states, as shown in panel A). (C) HMMs were used to analyze two measurements of IgG with hydrogel interfaced nanopore. Analysis of a data set where the baseline is relative low noise and flat compared to event amplitudes shows that the baseline (black) and events (red) were identified well without significant number of false-positive or missed event detections (measurement at -45 mV). Inset shows the emission distributions for the two states. (D) Analysis of a nosier data segment, with particularly significant flicker noise and longer dwell time events, shows poor identification of baseline and events, and HMM incorrectly labels regions within the two longest events as baseline. These errors lead to a high number of false positive events and therefore, significantly affect measurement statistics. (Same data as shown in Fig 4.2B for comparison)

Although nanopore measurements can be described well by Markov models, the states are not directly observable and only the emissions (measured currents) can be observed. HMMs and their associated algorithms can help solve problems associated with observable emissions with underlying states (Fig 4.3B. Specifically, unsupervised HMMs can use expectation maximization algorithms to take the observed emissions (the measured currents) and find the best fit parameters for a Markov model (the emission spectra of the states, transition probabilities) and provide a guess for the underlying sequence of hidden states (identify which data points are baseline currents and which data points are event currents). Here, the HMMs used assumed Gaussian emission spectra, where each state has a unique mean and standard deviation. Because event currents are inherently lower in magnitude than the baseline, the HMM will typically converge on a solution separating the baseline and event currents due to differences in their means. It is worth noting that blockade

currents will not always exhibit Gaussian emission spectra, for instance when measuring the shape of proteins,[15,16] but in our experience, this assumption does not interfere effective data analysis.

An additional advantage of HMMs as compared to thresholding methods is that Markov models account for both the amplitude of the events (the emission distributions) as well as the time during and in between events (transition probabilities). As a result, analysis using HMMs reduce the chance of identifying random noise as events, while still being able to identify large, short dwell time current blockades as events. In contrast, threshold-based methods only look at the amplitude of the data to identify events.

We opted to use the hmmlearn package in python because functions have already been developed for unsupervised training of hidden markov models using the Baum-Welch algorithm.[18] This unsupervised training only requires the data to be analyzed and the number of states. Initial guesses for a set of solvable parameters - the mean of both states, the variance of both states, and the transition probabilities - can be provided or default guesses can be used. The learning algorithm starts with the given initial guess for these parameters and iteratively adjusts the parameters until convergence is reached. In our experience, typical nanopore measurements and hydrogel facilitated measurements do not require optimized initial guesses for these parameters to yield converged and accurate analysis. Because efficient algorithms for solving parameters are well known and nanopore data is well suited for this method, implementation of HMMs to nanopore data is simple and computationally fast.

Figure 4.3C shows event detection by HMMs on hydrogel facilitated detection of IgG, with analysis assigning current regions as baseline and event without error. The data set is optimal for HMM as the events are clearly distinguishable from baseline and the baseline does not significantly fluctuate. The inset shows the emission distributions of the two states of the data (all

points histograms). This data set is a trivially analyzed by HMM because the baseline distribution is Gaussian and the emission spectra of both states are strong separated.

Because HMMs assume that the emissions distributions are the same for the entire duration of the data trace, HMMs fail when baseline drift or flicker noise has an amplitude similar to or larger than the event amplitudes. In these cases, the true mean of the emission distribution for the open pore and the occupied pore state is changing over the course of the experiment, but HMMs do not account for any change in emission spectra over time. Given data that exhibits significant baseline change, HMMs will converge on a set of state guesses that do not correctly identify the baseline and event regions. Fig 4.3D shows an example trace where HMM was not able to correctly identify the event and baseline regions of the data trace. Although the majority of the trace is assigned accurately, the errors in the two longest events of this trace result in the fragmenting of those long events into many shorter events. As a result, the statistics in analysis are significantly affected. The emission spectra shown in the inset of Fig 4.3C overlap significantly, and notably, the spectra meant to be associated with the baseline is non-Gaussian as a result of the significant flicker noise and partial inclusion of event data in this state. This further indicates that HMM was not able to converge on a satisfactory guess at the hidden state sequence.

4.2.3 Canny edge detection and baseline reconstruction

The issues HMM suffers with high variance baselines can be addressed if the baseline can be accurately estimated through a second computational method. Once an accurate baseline is resolved, this baseline can be subtracted from the raw data, leaving a baseline adjusted data set that is suited for HMM. The fluctuating baselines make thresholding methods ineffective, so a new method of calculating event locations and the baseline is needed. Because translocation events are

most easily visually identified by obvious, sudden shifts in current, we opted to utilize an edge detection algorithm to identify events.

Canny edge detection has been an established and effective algorithm for edge detection in images and is still effective for 1-D data analysis.[14] This method typically occurs in five steps: 1) filter out noise by convolving the data with a Gaussian filter (smoothed data) 2) convolve the data with a second filter that has the shape of the first derivative of a Gaussian in order to generate signal where there are edges (edge plot) 3) set all data points that are not local maxima to 0 (non-maximal suppression plot) 4) Use a threshold to identify which edges to keep 5) remove low intensity edges that are not connected to high intensity edges. For 1-D data, the last step does not apply.

This edge detection method can fail to adequately identify edges in certain types of data and has been modified previously to suit different applications.[19–24] Our method also modifies the algorithm to increase the quality of analysis for nanopore measurement data because the core algorithm does not adequately handle the varying noise, amplitude, and dwell times of nanopore current blockade events.

For consistency, we first sign-correct the data so that all currents are positive, and events are facing downwards (toward zero) relative to the baseline (Fig 4.4A). The raw data is then convolved with a Gaussian kernal with a fixed standard deviation (σ) (Fig 4.4B), reducing noise in the trace. We then subject the filtered data to another convolution, this time with the first derivative of the Gaussian operator with the same σ (Fig 4.4C), which is referred to as an edge plot from here forward. The convolution produces local maxima in the edge plot corresponding to the

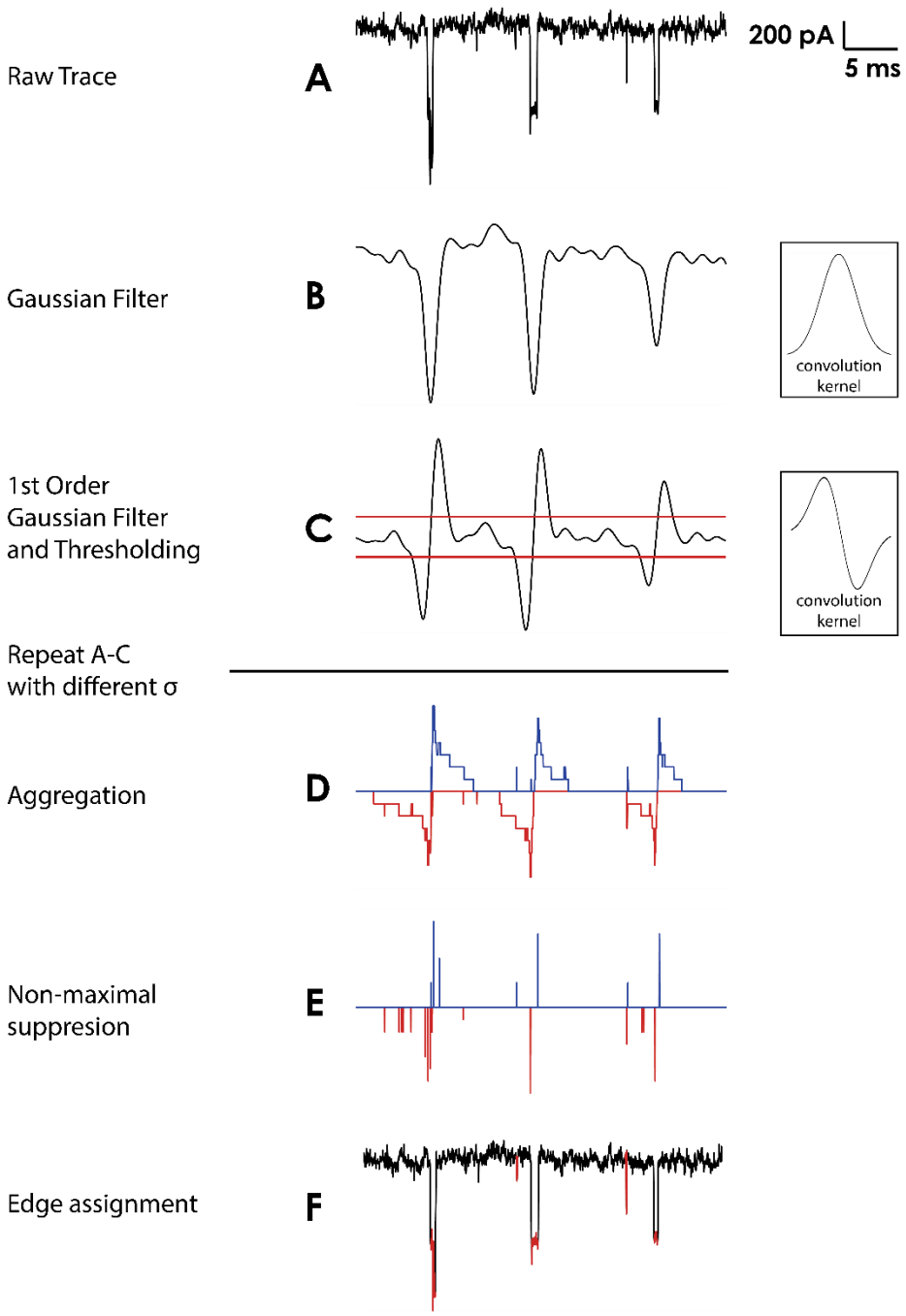


Fig 4.4 A modified Canny edge detection method can be used to analyze nanopore measurements and identify events. (A) Example measurement used here is IgG measured with a hydrogel interfaced nanopore in 2M KCl, pH 9.5 at -45mV. Data is sign corrected to positive currents for consistency. (B) The data trace is first convolved with a Gaussian kernel to reduce noise (kernel shape shown in

box on right). (C) The filtered trace is then convolved with a first derivative Gaussian kernel (shape shown in box to the right) to extract the gradient information of the data. Larger values in this step indicate the presence of edges. A threshold is calculated from a segment of event-free data. Values in the post-convolution trace that are less than this threshold are set to 0, and the remaining values exceeding the threshold are set to 1. (D) The previous steps are repeated, but for convolutions kernels of varying σ. The resultant traces are added point-by-point to produce an edge plot that has contribution from the convolution steps with different σ kernels. (E) All local non-maximum values are set to 0, and if a local maximum is over a plateau, the data point corresponding to the steepest slope in the raw data is identified as the local maximum. (F) A final thresholding step identifies the strongest edges, and events are identified by taking pairs of rising and falling edges that leave and return to values near the baseline.

locations of edges, with the sign (+/-) of each peak indicating the edge's direction ("rising" or "falling", i.e. the sign of the derivative with respect to time). Identical convolutions are also performed on a segment of data where no events are present. The maximum value from the edge plot of this control data segment is used as a threshold. All points in the edge plot below this threshold and all points that are not local maxima are set to 0, and the remaining points, which should all correlate to the location of detected edges, are set to 1, producing a binary edge plot indicating the locations of detected edges.

Different values of σ for the convolution filter result in increased or decreased contrast of certain edges dependent on the number of data points between two edges (Fig 4.5). For example, convolution filters with a large σ (200 data points or greater) will obscure edges associated with events shorter than 4 data points or particularly high frequency event translocations, while a smaller σ (10 data points or less) will not have significant contrast for signal of edges that are associated with long duration events and will introduce more false-positives. To ensure that all

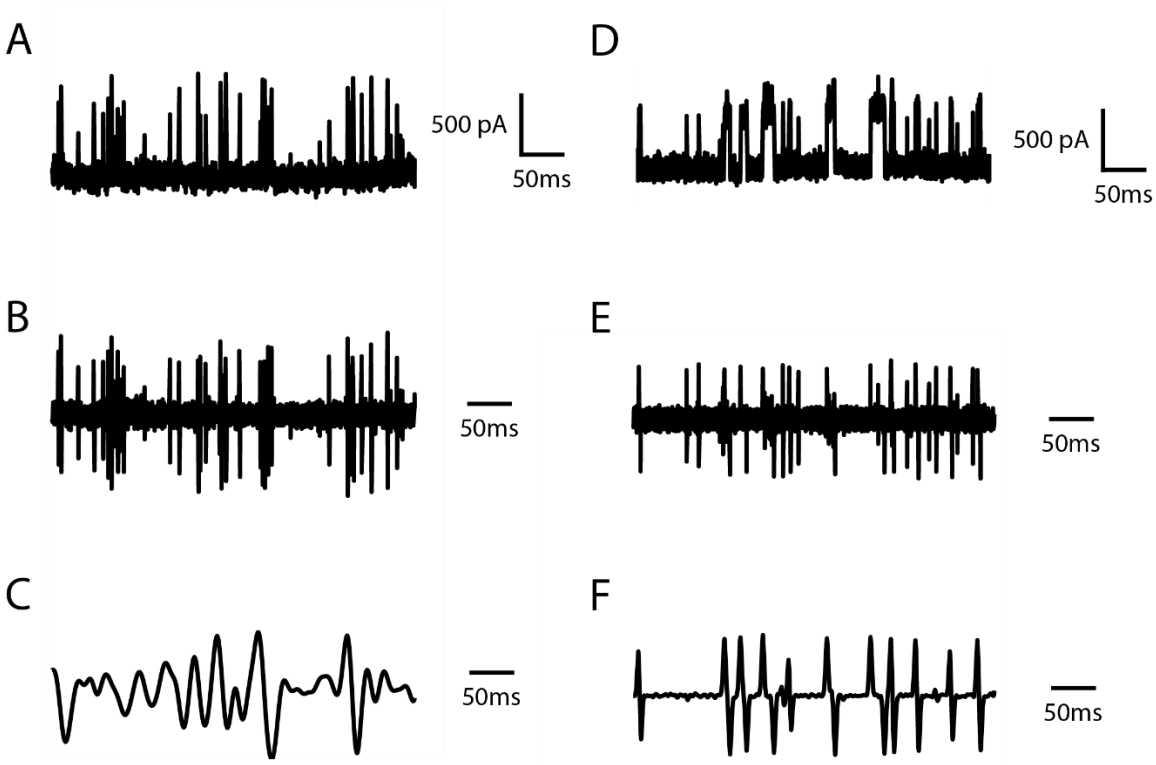


Fig 4.5 The edge contrast in the edge plot depends significantly on the width of the convolution filters used to generate the edge plot. (A) A measurement of IgG is shown with relatively short duration events. (B) Deriving the edge plot with convolution kernels with a σ of 2 data points shows clear signal for the location of the edges. (C) Convolving the same data with filters with a σ of 200 data points produces a trace with no identifiable edge locations and low intensity at the edges. (D) A different measurement of IgG is shown where event durations are typically longer. (E) The edge plot of this data when convolved with kernels with a σ of 2 data points show clear signal for the event edges. (F) The same data can be convolved with filters with a σ of 200 data points, and the signal for edges is very prominent compared to background, especially for longer duration events. Large σ convolution filtering can show edges in extremely noisy data if event durations are sufficiently long.

types of events are accurately identified, a data trace is convolved multiple times with filters of varying σ from 1 to 500 data points. The binary edge plots for each σ are summed up point-by-point, and the resulting summed edge plot should show significant signal for edges regardless of the event and interevent durations (Fig 4.4D).

This summed edge plot then undergoes non-maximal suppression (NMS) and thresholding to determine the location of event edges (Fig 4.4E). In cases where there is a plateau at the local maximum, the first derivative is taken in the region, and the edge is assigned to the point with the highest derivative as we expect the center of the edge to have the steepest slope. A thresholding step is conducted on the NMS plot to retain the strongest edges. In our experience, thresholds of 3 to 4 (three to four of the various σ convolutions agree on the presence of an edge) usually have the least number of false-positive events while still detecting nearly all edges in typical measurements. As a final check to ensure the elimination of false positive edges, the amplitude of each edge is calculated; edges with amplitudes below a minimum threshold, set at 2.8 times the standard deviation of the event-free data (an approximation of the noise level of the nanopore measurement) are removed.

This method can analyze traces with high signal-to-noise ratio successfully without any additional processing, but noisier measurements may yield unacceptable error rates. As a result, we use this modified Canny edge detection to produce a high-quality estimate of the baseline currents. By identifying the locations of blockade currents and their associated edges, events can be subtracted from the data trace, leaving an event-free baseline estimation. Because an undetected event during this event subtraction has a higher penalty than a false-positive event detection, we do not threshold in the last step of our modified method when estimating baselines. Instead, all

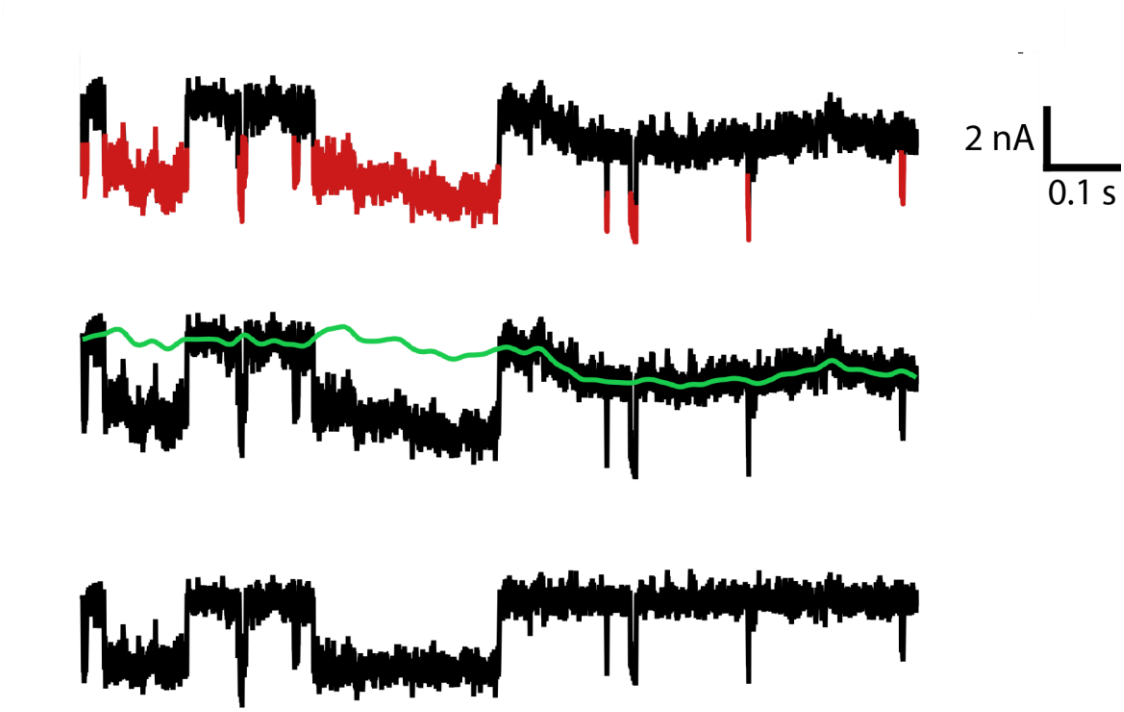


Fig 4.6 Data of IgG measured with a PEG-DMA backed nanopore in 2M KCl, ph 9.5 at -70 mV is shown. Analysis of this segment of data with the modified Canny edge detection algorithm shows accurate analysis of all events (top, events highlighted in red). Here, no threshold was used in the final step, and all edges were accepted to ensure that no events are missed. Data during events are subtracted by the average height of the leading and trailing edges of each event. The resultant trace is low pass filtered, yielding an estimate of the baseline that does not include capacitive noise but retains low frequency noise (middle, green line is the baseline estimation). The estimated baseline can be subtracted from the original trace to produce a trace that has a perfectly flat baseline and retains all events and their characteristics (bottom).

edges that produce non-zero signal after the NMS step are accepted. This ensures that all current blockades are captured.

After edge detection, events are then defined by two criteria (Fig 4.4F). Events must have both a leading edge (decrease in current) and trailing edge (increase in current) that are in opposite

directions, and the trailing edge must return to a current value that at least within the noise (2.8 times the standard deviation) of the baseline before the leading edge. Once events are identified by assigning edge pairs according to these conditions, each event region is subtracted by the average amplitude of each event's leading and trailing edges, effectively removing events from the experimental data. The trace is the low pass filtered at 25 Hz to produce a baseline that retains the flicker noise of the original data set but not the high frequency noise characteristics. As a final step, the entire baseline calculation method is repeated on the newly calculated baseline to ensure that any previously unidentified events or events in which two conductance steps are present (such as when two analytes occupy the pore simultaneously but have unique entrance or exit times) are removed in the second iteration. Figure 4.6 shows baseline estimation and subtraction on experimental data.

The result of the edge detection-based baseline is a significant improvement over other baseline reconstruction approaches that rely on averaging methods because the edge detection-based baseline only makes adjustments at event locations. As a result, baselines can be successfully estimated for measurements with significant flicker noise and long duration events.

4.2.4 Hidden Markov modeling with edge detected baseline correction

Subtracting the baseline predicted by the edge detection method produces good HMM candidate data for measurements that would otherwise fail with HMM. Figure 4.7A shows the same trace in Fig 4.2B and Fig 4.3D that was not successfully analyzed by thresholding or HMM alone (Fig 4.7B reproduces the HMM analysis for comparison). Analysis with these methods led to a falsely high number of detected events. Fig 4.7C shows analysis of the same segment of data using only edge detection. Edge detection also has an unacceptable error rate, with a large number

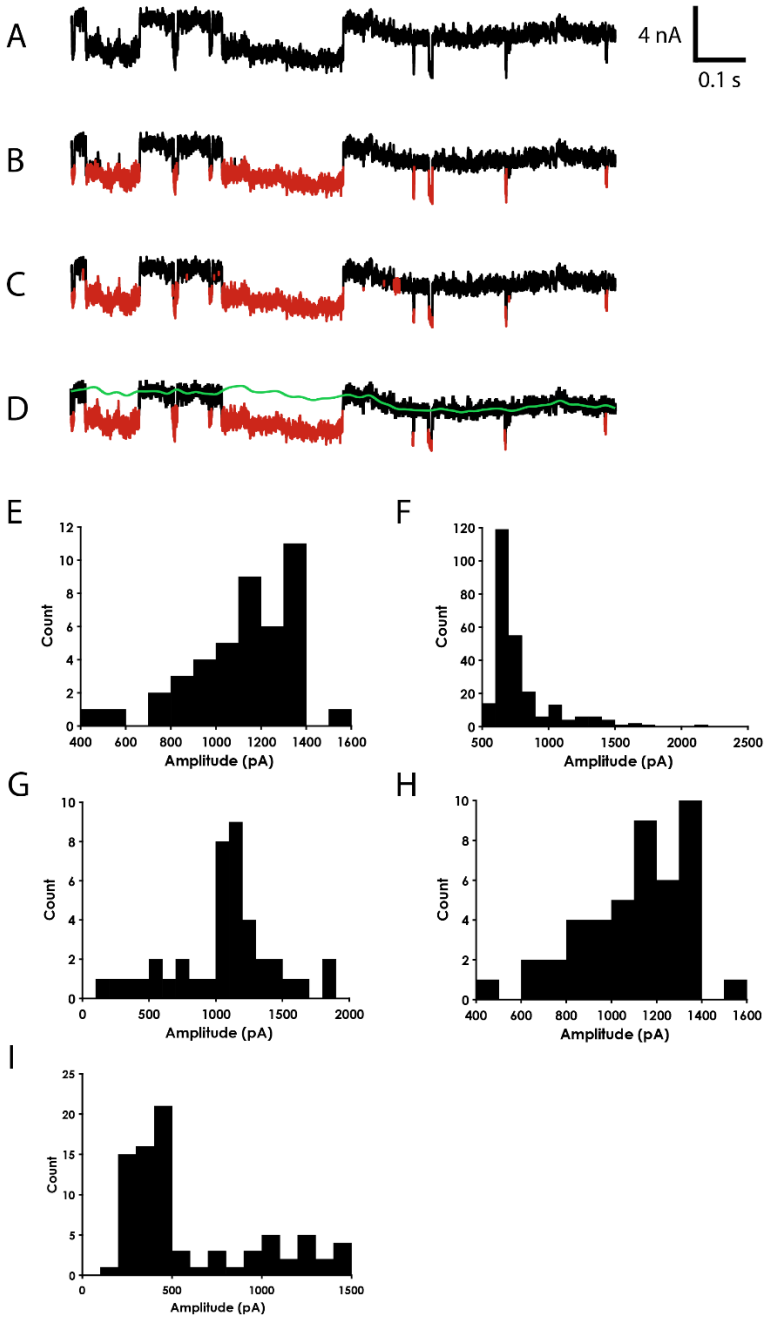


Fig 4.7 HMM can be used with the modified Canny edge detection algorithm to analyze data traces with significant noise. (A) A measurement of IgG with a hydrogel backed nanopore is shown (same data is shown in Fig 4.2B) (B) Analysis of this data with HMM inaccurately identifies event regions as baseline (black regions are identified as baseline, red regions are identified as events). (C) Analysis of this data with edge detection also fails to accurately identify all events, with a

significant number of identified false positive events in the baseline. (D) Events detected by edge detection can be removed, and the resulting trace can be low pass filtered to produce a good estimate of the baseline (green line). Subtracting this baseline from the original data can remove effects of flicker noise while retaining all events in the data (shown in Fig 4.6). HMM can analyze this data with minimal error. (E) A ~7 second segment of data from the same experiment as the data shown in Fig 4.5A was analyzed by hand to ensure accuracy, and a histogram of the average amplitude of events is shown. 43 events were identified. (F) Analysis of the same data with HMM produces an average amplitude distribution that has a lower most probable amplitude and is significantly different in shape and due to a significant false-positive event rate. (G) Edge detection alone is more accurate than HMM or thresholding at analysis of this data with 40 detected events, but 10 of these events were incorrectly assigned, and the histogram shows significant difference from the true histogram. (H) Estimating the baseline with the modified Canny edge detection algorithm and then analyzing the data with HMM is the most accurate method and identifies 44 events, 42 of which are accurately assigned. The histogram of amplitudes identified with this method closely matches the histogram in (E). (I) For comparison, the data was also analyzed using Transalyzer[2] which also estimates the baseline through event subtraction before analysis, though the method uses thresholding methods for all event detection. Transalyzer detected 82 events, of which only 25 were accurate, and the histogram does not match the true amplitude histogram. The method outperforms HMM alone for data with significant flicker noise, but is still not accurate enough for analysis of data that exhibits significant flicker noise along with capacitive noise and long duration events.

of false positive events generated by baseline noise. The baseline can be estimated by subtracting the events detected by edge detection, and this baseline can be subtracted from the original trace. The baseline adjusted data retains all events but does not have any flicker noise, and HMM can accurately analyze this data without any error (Fig 4.7D).

The need for accurate analysis is highlighted by the amplitude distributions of events detected by these methods. For reference, Fig 4.7E shows the average amplitude distribution of a ~7 second segment of data from the same experiment as the data shown in Fig 4.7A which was analyzed by hand to ensure accuracy (the same segment analyzed in Figs 4.2B-E and 4.3D). The histogram shows a most probable average amplitude of ~1350pA and an asymmetrical distribution. The same segment of data was analyzed, in 7 separate, one second long segments with HMM alone (segmenting the data like this increased accuracy), detecting 253 events and yielding the histogram shown in Fig 4.7F, where amplitudes were calculated as the average amplitude during an event subtracted by the mean current of the baseline state in the segment. The histogram shows a most probable average amplitude of ~650pA, significantly different than the true average amplitude. Using edge detection alone, with edge thresholds optimized for accuracy, 40 events were detected, though only 30 of these events accurately identified the bounds of the event. The histogram for this analysis is shown in Fig 4.7G with event amplitudes calculated as the difference of the average event current and the baseline current preceding the event. The amplitudes have a most expected value of ~1250pA, but the shape of the distribution is noticeably different from the true distribution. Fig 4.7H shows a histogram from the analysis of this data using the hybrid Canny edge detection and HMM method which identified 44 events, of which 42 were accurately analyzed. As a result, the histogram shows agreement with the true histogram (Fig 4.7E). Amplitudes were calculated similarly to the edge detection analysis. It is worth noting that this data set was selected as an example as it exhibits significant flicker noise, low SNR, and long duration events, and as a result, none of the methods were fully accurate, though the 95.5% accuracy of the baseline adjusted HMM provides adequate statistics when interpreting the data. In measurements that exhibit fewer of these difficult measurement characteristics, the methods here

Analysis Method	Correctly Identified Events	Incorrectly Identified Events	% Correct
Manual Analysis	40-43	-	-
Thresholding	2	1647	0.1%
Transalyzer[2]	25	57	30.5%
HMM	34	219	13.4%
Edge Detection	30	10	75.0%
Baseline Adj. HMM	42	2	95.5%

Table 4.1 Comparison of analysis methods. Table shows number of correctly identified events, incorrectly identified events, and the percentage of correctly identified events from all events for a ~7 second segment of data measuring IgG with a hydrogel backed nanopore at -70 mV in 2M KCl, pH 9.5. The data segment exhibits significant flicker noise, low SNR, and has long duration (> 1 ms) events. (Example traces shown in Fig 4.1B and 4.5A. Histograms shown in Fig 4.1C and Fig 5E-H).

would all most likely increase in accuracy. The method presented here using modified Canny edge detection and HMM is robust enough to accurately analyze a wider range of measurements without modification. A summary of the analysis methods on this data, including thresholding, is shown in Table 4.1.

In our experience, this method, which exploits both HMM and a modified Canny edge detection algorithm, is robust enough to analyze most nanopore measurements. However, there are two cases where the method fails to adequately analyze data. If the edge detection algorithm cannot accurately estimate the baseline, HMM will also fail to accurately identify events. Typically, baseline miscalculation occurs over the region of a single event or a few events, and we address

this by correcting the estimated baseline by hand. HMM can also fail to accurately parse data if event frequency is very low or if event duration is extremely long. In these cases, the HMM's transition probabilities converge on incorrect values, regardless of initial guesses. This failure mode is particularly catastrophic, as HMM will detect a large number of false positive events in a low frequency data trace, entirely masking the true event statistics. For data sets that have low event frequency and long dwell times, down-sampling the data can resolve this issue. However, we can estimate event duration and event frequency by edge detection to determine if HMM can successfully parse the data. Since edge detection is already performed on the data trace for baseline estimation, this calculation does not add significant additional computation time. In cases where the mean t_{off} or t_{on} is over 3 seconds at an acquisition rate of 100 kHz, which is a conservative threshold, we use edge detection as the only analysis method and omit HMM. Figure 4.7 summarizes the full data analysis workflow.

This workflow ensures that all traces can be analyzed by the method introduced in this chapter. For most data, the algorithm can successfully parse data sets and identify events without error. In especially noisy data, the algorithm typically outperforms other methods, but can still miss or incorrectly identify some events. In these cases, manual supervision of the analysis, usually adjustment of poor baseline estimations around single events, is needed to ensure high quality analysis. Although this manual supervision increases time commitment for analysis, we find that for these difficult data sets, the method presented here is still the most efficient method for analysis with minimal error.

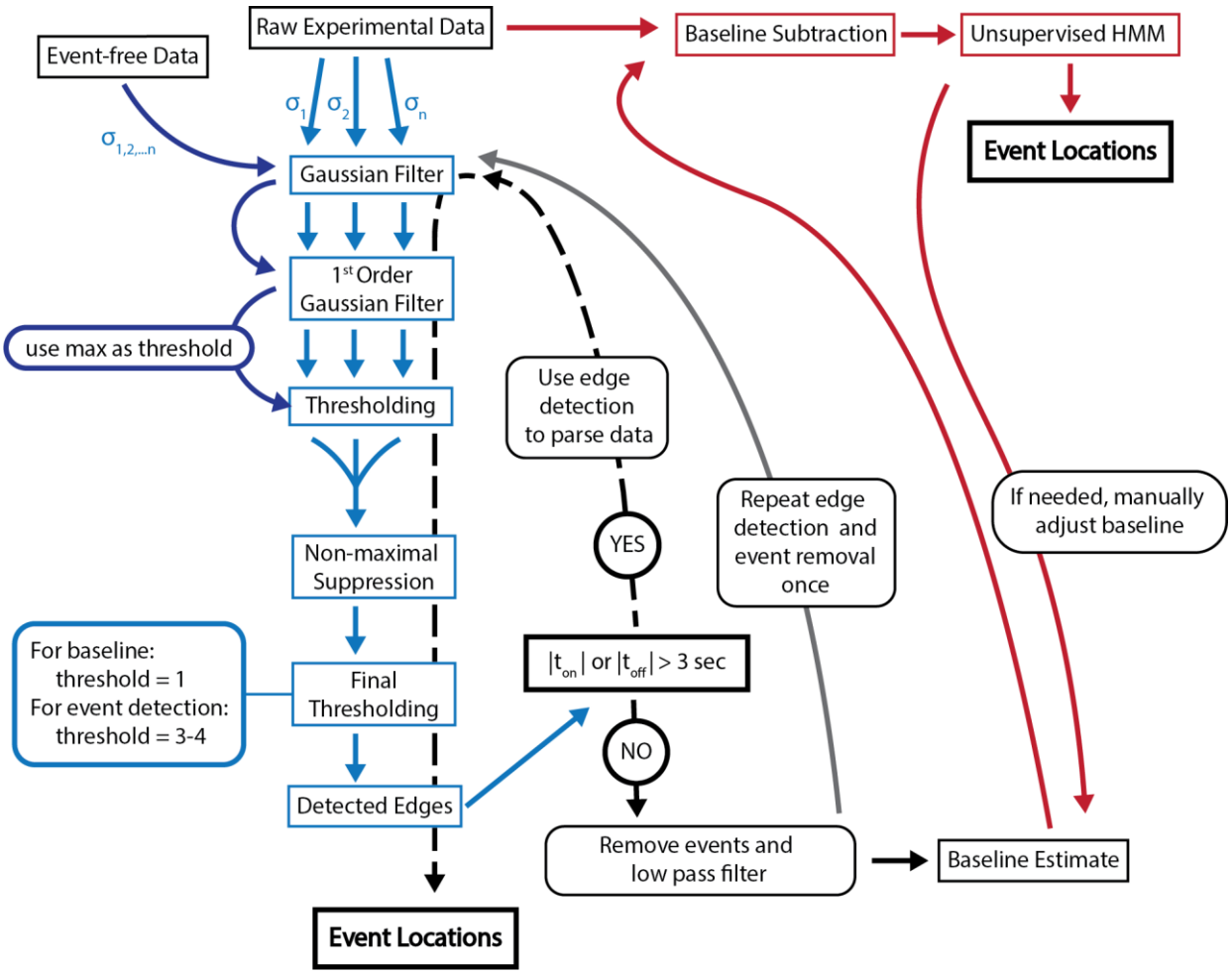


Fig 4.8 A summary of the data processing workflow is shown. Edge detection steps are shown in blue and HMM related steps are shown in red. Data first undergoes the modified edge detection described in this chapter, using a user provided event-free data set to determine the threshold of the edge plot. The final threshold is set at 1 to ensure all detected edges are retained and ensuring that no events are missed. Using this analysis as an approximation, the average t_{on} and t_{off} times are estimated, and if either is over 3 seconds, edge detection is used to identify events, though the final threshold is set at 3 or 4 in this iteration to reduce the number of false positive events. In cases where the average t_{on} and t_{off} times are below 3 seconds, the identified events are subtracted from the raw trace, and the data is filtered to produce a baseline estimate. To ensure all events are removed, the entire edge detection process is repeated, and additional detected events are also

subtracted from the trace. The baseline estimate can then be subtracted from the raw data, and HMM can be used to identify the event locations. If needed, the baseline estimate can be manually adjusted, and the baseline adjusted data can be reanalyzed by HMM without repeating the entire workflow.

4.3 Conclusion

HMM, edge detection, and thresholding methods are all capable of accurate analysis of lower noise nanopore measurements with no other facilitated methods. However, for traces with low signal-to-noise ratio, thresholding can fail to accurately identify events. Here, we capitalize on HMMs and a modified Canny edge detection algorithm to make a more robust data analysis method, especially for noisier measurements. The methods address weaknesses that either method would have when used alone. HMM is only effective when the baseline fluctuation is smaller than event amplitudes. Edge detection, as discussed earlier, misidentifies events on noisy data with enough frequency to cause deviations in event statistics. However, because edge detection can still yield a reliable event-free baseline, the method can be used with HMM to produce high accuracy event identification.

Although thresholding-based methods can be effective for many nanopore measurements, the method shown in this chapter is effective on a greater range of data sets. For example, hydrogel facilitated measurements of proteins, as shown in Chapter 2 and 3, can exhibit high event frequencies and long dwell times (t_{on}/t_{off} can be near or greater than 1), and, because the measurements are performed with solid state nanopores, flicker noise and capacitive noise. We have found these measurements to be difficult to analyze with other methods, and therefore used these data sets as benchmarks for this work. For measurements of nucleic acids and more recently, proteins, studies are using high bandwidth measurement to enable new measurements.[16,25,26] High

bandwidth measurements may also require a more robust analysis method as these measurements are plagued by increased capacitive noise at higher frequencies.

Further work would include exploring how HMM handles multi-level current blockades, as may be seen from multiple particles occupying the pore simultaneously. A potential advantage of HMM may be the ability to separate current blockade levels within an event without additional processing. The method here also requires a control, event-free data trace to determine appropriate thresholds, but many methods that use Canny edge detection determine these thresholds by processing the raw data.[21–24,27] Including a method to determine the best thresholds from raw data would increase convenience of analysis and may reduce issues in analysis when poor event-free traces are provided.

4.4 Materials and Methods

4.4.1 Data analysis

All analysis was conducted using MATLAB 2016a and Python 3.6. HMM training was performed using the hmmlearn package in Python with the GaussianHMM model. Training was conducted for two states for all data, and all initialization parameters for the HMM model were left to default except for transition probabilities, which were set to 10^{-20} for transition between states and to $1-10^{-20}$ for probability to remain in state. Adjusting initialization parameters, including transition probabilities, did not significantly affect the learned parameters, and we did not have to adjust initializations for each data set.

The data analysis method discussed in this paper was integrated into a graphical user interface (GUI) which could accept data, perform the analysis, adjust estimated baselines, plot detected

events, and export event statistics. The HMM functions were called from MATLAB using built in support for Python. Modified Canny edge detection was conducted entirely in MATLAB.

4.4.2 Data collection

Measurements were taken as described previously in Chapter 2. Briefly, 15nm thick silicon nitride TEM windows (Norcada, Tedpella) were coated in PDMS to reduce capacitive noise[28] and mounted in Teflon fluidic chambers. The chamber was filled on both sides with ethanol and placed under vacuum to remove any air bubbles and to ensure wetting of the TED window. The ethanol was then replaced with 2M KCl, pH 9.5 on the *cis* half of the chamber and a PEG (1000)-DMA hydrogel in the same buffer on the *trans* side. Gel was polymerized with 10mM ammonium persulfate and 10mM tetramethylethylenediamine. After polymerization, additional buffer was added to the side of the hydrogel to ensure that the hydrogel does not dehydrate during the course of the experiment. Nanopores were formed by dielectric breakdown[29–31] and grown to a ~25nm diameter. Nanopores were subsequently measured with an Axopatch 200B (Molecular Devices) amplifier and Digidata 1440B digitizer (Molecular Devices) to ensure that currents were stable and free of any contaminant. Protein was then added to the side of the nanopore opposite of the hydrogel and measurements were taken until at least several hundred events were acquired. Measurements were taken at voltages between -70 and -20 mV typically.

1. Gu, Z., Ying, Y.-L., Cao, C., He, P. & Long, Y.-T. Accurate Data Process for Nanopore Analysis. *Anal. Chem.* **87**, 907–913 (2015).

2. Plesa, C. & Dekker, C. Data analysis methods for solid-state nanopores. *Nanotechnology* **26**, 084003 (2015).

3. Raillon, C., Granjon, P., Graf, M., Steinbock, L. J. & Radenovic, A. Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale* **4**, 4916 (2012).

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

5. Smeets, R. M. M., Dekker, N. H. & Dekker, C. Low-frequency noise in solid-state nanopores. *Nanotechnology* **20**, 095501 (2009).

6. Smeets, R. M. M., Keyser, U. F., Dekker, N. H. & Dekker, C. Noise in solid-state nanopores. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 417–21 (2008).

7. Dempster, J. *Computer analysis of electrophysiological signals*. (Academic Press, 1993).

8. Pedone, D., Firnkes, M. & Rant, U. Data Analysis of Translocation Events in Nanopore Experiments. *Anal. Chem.* **81**, 9689–9694 (2009).

9. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).

10. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015).

11. Rand, A. C. *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).

12. Timp, W., Comer, J. & Aksimentiev, A. DNA Base-Calling from a Nanopore Using a Viterbi Algorithm. *Biophys. J.* **102**, L37–L39 (2012).

13. Zhang, J. *et al.* High-bandwidth nanopore data analysis by using a modified hidden Markov model. *Nanoscale* **9**, 3458–3465 (2017).

14. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).

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

16. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* acsnano.8b09555 (2019). doi:10.1021/acsnano.8b09555

17. Baum, L. E. & Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Stat.* **37**, 1554–1563 (1966).

18. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Stat. Soc. Ser. B* **39**, 1–22 (1977).

19. Liu, H. & Jezek, K. C. Automated extraction of coastline from satellite imagery by integrating Canny edge detection and locally adaptive thresholding methods. *Int. J. Remote Sens.* **25**, 937–958 (2004).

20. Ding, L. & Goshtasby, A. On the Canny edge detector. *Pattern Recognit.* **34**, 721–725 (2001).

21. Biswas, R. & Sil, J. An Improved Canny Edge Detection Algorithm Based on Type-2 Fuzzy Sets. *Procedia Technol.* **4**, 820–824 (2012).

22. Rong, W., Li, Z., Zhang, W. & Sun, L. An improved Canny edge detection algorithm. in *2014 IEEE International Conference on Mechatronics and Automation* 577–582 (2014). doi:10.1109/ICMA.2014.6885761

23. Chang, S. H., Gong, L., Li, M., Hu, X. & Yan, J. Small retinal vessel extraction using modified Canny edge detection. in *2008 International Conference on Audio, Language and Image Processing* 1255–1259 (2008). doi:10.1109/ICALIP.2008.4590140

24. Nikolic, M., Tuba, E. & Tuba, M. Edge detection in medical ultrasound images using adjusted Canny edge detection algorithm. in *2016 24th Telecommunications Forum (TELFOR)* 1–4 (2016). doi:10.1109/TELFOR.2016.7818878

25. Rosenstein, J. K., Wanunu, M., Merchant, C. A., Drndic, M. & Shepard, K. L. Integrated nanopore sensing platform with sub-microsecond temporal resolution. *Nat. Methods* **9**, 487–492 (2012).

26. Larkin, J., Henley, R. Y., Muthukumar, M., Rosenstein, J. K. & Wanunu, M. High-bandwidth protein analysis using solid-state nanopores. *Biophys. J.* **106**, 696–704 (2014).

27. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **9**, 62–66 (1979).

28. Tabard-Cossa, V., Trivedi, D., Wiggin, M., Jetha, N. N. & Marziali, A. Noise analysis and reduction in solid-state nanopores. *Nanotechnology* **18**, 305505 (2007).

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

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

31. Yanagi, I., Hamamura, H., Akahori, R. & Takeda, K. I. Two-step breakdown of a SiN membrane for nanopore fabrication: Formation of thin portion and penetration. *Sci. Rep.* **8**, 10129 (2018).

Chapter 5

Future Directions

5.1 Summary

Nanopores are powerful single molecule sensors, with the ability to directly probe physical characteristics of molecules through electrical measurement. Proteins are particularly interesting targets because of their relevance in medicine and biology, and in contrast to nucleic acids, which have been extensively studied with nanopores, their folding structure informs greatly on their identity and function. Nanopores, in theory, can probe the size, shape, volume, charge, and dipole moment of proteins in native conformation,[1,2] making them highly capable tools to measure native proteins. However, theoretical and experimental work has shown that proteins translocate through pores too rapidly to be resolved accurately. Here, it is shown that a hydrogel placed on the distal side of a nanopore can sufficiently increase the residence time of proteins to be resolved accurately by typical measurement apparatus. Measurement of IgG, BSA, Streptavidin showed event frequencies that exceeded diffusion-based theoretical predictions at concentrations multiple orders of magnitude below those used typical protein studies. Dwell time distributions extended to significantly greater times, with events extending to tens of milliseconds in duration. Event amplitudes agreed with the volume of proteins as calculated from current blockade models.[3–5] As a result of the enhanced detection with a hydrogel, IgG and BSA could be measured simultaneously in mixture and current blockades informed on the particle shape, though the dipole moment was incorrectly measured.

A model was presented to describe the capture mechanism and dwell time of protein measurement with hydrogel backed nanopores. It was shown that, for IgG and BSA, the median dwell time of measured events increased with increasing voltage, suggesting that neither protein could penetrate the hydrogel. The result is particularly interesting because the mechanism allows for increased dwell time and increased blockade amplitude at increasing voltages, in contrast to conventional sensing. Because the hydrogel is impermeable, the residence time was also shown to be much more sensitive to the diffusivity and electrophoretic mobility of the protein.

In Chapter 4, a novel method to analyze nanopore measurements was presented using hidden Markov models[6] and a modified Canny edge detection algorithm.[7] The method addresses difficulty in analyzing data that shows significant flicker noise, capacitive noise, or has long duration events at high frequency, such as data produced by hydrogel facilitated detection of protein. The presented method uses two steps: 1) a modified Canny edge detection to estimate a baseline and remove flicker noise from the baseline, and 2) an unsupervised hidden Markov model to analyze the baseline adjusted traces. The method is robust enough to analyze a large variety of data, including noisy traces or traces with long dwell times which cannot be analyzed by other computational methods.

5.2 Future Work

5.2.1 Exploring limits of hydrogel facilitated detection of proteins

Subsequent work should investigate the limits of hydrogel facilitated detection. Here, aside from the addition of the hydrogel, our measurement protocols were similar to those used in previous studies. In previous measurement, the range of measurable conditions was significantly limited by the difficulty in measuring proteins. For example, concentration of protein and the size

of measurable proteins had limited ranges defined by the need to produce detectable signals above background. With hydrogel facilitated detection, results already indicate that the range of measurable conditions is significantly increased and that the work shown so far has not approached protocol limits. Exploring these protocol limits as well as other measurement parameters, such as applied voltage, will define the scope of application for hydrogel facilitated detection of native proteins.

In our work, we measured protein at a concentration of 10 nM. While this concentration is already 1-4 orders of magnitude below protein concentrations used in other studies,[1,2,8,9] our high measured event rates suggest that protein concentration can still be reduced further. The concentration detection limit is one of the most important parameters that defines the scope of application for a sensor. For instance, a wide variety of relevant biomarkers circulate at the picomolar concentration range, such as vascular endothelial growth factor, whose expression levels can indicate rheumatoid arthritis and certain cancers,[10–12] and human epidermal growth factor receptor 2 (HER2/neu), which can indicate for breast cancer.[13,14] Reduced concentration detection limits are also generally important for research or clinical applications where sample volume is limited.

Future work should also investigate size limits of detection. To our knowledge, there are no known reports measuring small proteins (<40kD) using standard measurement apparatus. Although Larkin *et al.* have measured smaller proteins in the 10-30 kD size range, their measurements specifically relied on use of a high-bandwidth amplifier not commonly used in most nanopore studies and nanopores with diameters that were specific to the size of the measured proteins. The long dwell times of blockade events in hydrogel facilitated detection may make measurement of smaller proteins possible without tailoring the size of the nanopore specifically to

target proteins. Additionally, hydrogel facilitated detection has the unique property of increased amplitude and dwell time with increased voltage (for proteins measured thus far). Exploiting this behavior may reduce both the size and concentration detection thresholds. Similarly to concentration limits, exploring the size limit of hydrogel facilitated detection informs on use-cases in the future.

In addition to the size detection limit, the size resolution of the technique should be investigated to determine the capability of discriminating proteins in mixture. In our measurements of BSA and IgG in mixture, we could reject events with dwell times as long as 10 ms and still retain enough events to produce adequate statistics. These distributions suggest that many more proteins could be measured in mixture, and the smallest discernable size difference would be much smaller than the ~70kD difference between BSA and IgG. Exploring and maximizing this size resolution is highly relevant for clinical and research applications where purified protein is not available. In most practical situations, the solution to be measured will contain a mixture of proteins, and either the solution must undergo sample preparation to separate the proteins or the nanopore must be able to discriminate amongst the mixture. In a case where nanopores can discriminate amongst a diverse mixture of proteins with low error, simultaneous measurement is preferred to sample preparation as it significantly increases throughput and reduces labor. Additionally, it has been shown that detecting a panel of biomarkers can indicate for disease with significantly higher confidence than a single biomarker.[15–18] Either separately or simultaneously, the capability of a nanopore sensor to detect and quantify a suite of biomarkers would make the sensor significantly more attractive for diagnostic applications. In general, one of the primary goals of nanopore sensing is discrimination and identification of proteins in mixture. The work shown

here is one of the first studies to show label-free discrimination amongst a mixture of proteins with a nanopore, and further investigation is warranted given the promise of the technique.

Lastly, because IgG and BSA cannot penetrate the hydrogel, increasing voltages can yield larger signal amplitude and longer dwell times. This was briefly mentioned previously as a method to potentially decrease the size and concentration limits of detection. This behavior also poses a unique opportunity in nanopore measurement to get extremely long, single-molecule measurements with large current blockade amplitude, which may also significantly improve the size resolution. It is possible that these measurements may lead to clogging of the nanopore given the extreme residence time of the protein in the nanopore. Coating the nanopore with a lipid bilayer may prevent this clogging[2,8] and enable interesting, long duration measurements of a single protein.

Increasing the voltage to increase dwell time can also have drawbacks in relation to throughput, especially for event population studies (as opposed to single-event/intraevent analysis). For applications where the highest frequency of events is desired, increasing voltage can reduce the time in between events for charged particles because electrophoresis enhances capture rate. However, as shown by measurements of IgG and BSA, some proteins will also reside in the pore for longer times, potentially reducing the frequency of events if these durations are significant relative to the time in between events. As a result, there will be an optimal measurement voltage for throughput, balancing the time between events and the time during an event. Investigating this hypothesis for various proteins, nanopore diameters, and nanopore lengths may provide insight on best protocols for measurement and may further reduce the concentration detection threshold and increase the dynamic range of measurement.

5.2.2 Novel directions in protein detection with hydrogel backed nanopores

Hydrogel facilitated detection is also amenable with other techniques to improve nanopore detection. Lipid coatings on the surface of the nanopore have been shown to prevent clogging and protein-pore interactions, which are still issues in hydrogel facilitated detection. These coatings coupled with hydrogels can facilitate long duration, single molecule measurement as mentioned previously. It was also shown recently that measurement of a protein's dipole moment is inaccurate in bare pores but can be measured accurately in lipid coated pores due to edge effects as the protein approaches the pore wall. This finding may explain the incorrect dipole measurements of proteins measured with hydrogel interfaced nanopores, and lipid coating a hydrogel backed nanopore may lead to accurate measurements of the protein dipole moment.

The same study demonstrated that protein shape, volume, and dipole moment could be measured from individual blockade events that are at least 150 μs in duration. In contrast to shape and volume measurements shown in this dissertation which relies on population analysis, this method of analysis which analyzes each event separately is highly attractive because it may allow for measurement of shape and dipole moment of proteins in a mixture. As a result, individual event analysis may allow for discrimination of similarly size proteins in a mixture if the proteins have differing shapes. However, this previous study was limited to analysis of only tens to hundreds of blockade events due to rapid translocation of proteins, and protein concentrations were in the 1-10 μM range. Their results indicated that having a greater number of events, and more importantly, longer duration events, led to more accurate measurement of protein shape and dipole moment. Coupling a hydrogel to a lipid coated pore would produce significantly more events and longer events without issues related to protein-pore interaction and protein-pore proximity effects. As a result, this joint technique may yield more precise measurements of protein shape and dipole

moment from individual events. As a note, in our experience, coating nanopores formed by dielectric breakdown with lipids is difficult with very low yield, and therefore, the proposed method of exploiting lipid coatings and hydrogels would require nanopores formed by ion or electron beam. We do not anticipate this to be an issue as our results indicate that interfacing hydrogels with ion beam drilled nanopores is still effective at increasing the residence time and observed frequency of captured proteins.

In previous measuring nucleic acid with hydrogel interfaced nanopores, it was shown that changing the characteristics of the hydrogel such as charge and pore size can change the residence times of nucleic acids.[19] The same technique could be used to tune dwell times of translocating proteins or add partial selectivity to protein measurements with nanopores. Thus far, we have used amplitudes to identify proteins, though the results shown in Chapter 3 suggest that dwell time distributions can change significantly between proteins based on charge and size. The differences in dwell time were not significant enough between IgG and BSA to allow discrimination between the proteins based on dwell time. However, as an example, incorporation of a charged hydrogel, and possibly reducing the pH until it is near the isoelectric point of one of the proteins, may lead to significant enough differences in dwell time that it can be used to identify proteins. Even with some overlap in dwell time, it may be possible to use both dwell time and amplitude in conjunction to increase confidence in identifying proteins in mixture.

Adjusting the mesh size of the hydrogel may also allow for selectivity of analyte. Thus far, we have only measured proteins with dimensions larger than the mesh size of the hydrogel, but we expect that for smaller molecules and polymer chains such as nucleic acids,[19–22] the molecule will penetrate the hydrogel. At higher applied voltages, larger proteins would be expected to increase in dwell time, but smaller molecules and polymer chains would reduce in dwell time. This

separation in dwell time may be distinguishable in mixture. Adjusting the mesh size of the hydrogel would determine the size threshold of molecules that increase or decrease in dwell time at larger voltages, possibly creating a greater measured separation between two molecules that would otherwise be difficult to separate.

An example of exploiting dwell time behavior is the measurement of streptavidin or avidin with biotin terminated nucleic acids. In a mixture of these molecules, the avidin molecules would be expected to bind with the nucleic acids, though some free protein and polymer will remain in solution. Measurement of this mixture would be difficult without a hydrogel present, as the radius of gyration of nucleic acids can approach the size of proteins. However, if measured near the isoelectric point of the avidin and at high voltage (>80 mV), the free nucleic acids would penetrate the hydrogel and have short dwell times. The uncharged free protein would not be able to penetrate the hydrogel but would have short dwell times as there would be no electrophoretic force enhancing capture. However, avidin bound to the nucleic acids would have very long dwell times as the protein would not be able to penetrate the hydrogel and the nucleic acid would add a large negative charge to the particle, significantly increasing the electrophoretic force on the molecule. A measurement of this mixture of molecules would include short duration events, which could be the result of free nucleic acid, free protein, or rarely a nucleic acid bound protein, and very long duration events (most likely greater than 1 second) that would almost all be captured avidin with bound nucleic acid. Exploiting the relationship between the mesh size of the hydrogel and the size and type of molecules may allow for similar novel measurements with other molecules.

The example of measuring avidins and nucleic acids in mixture demonstrates a method to exploit hydrogel backed nanopores for improved labeled measurement of proteins and other types of analytes. Tethering nucleic acids to analytes by biotin linkages, antibodies, or as aptamers can

allow for long duration, direct measurement of analytes. This type of mechanism has been exploited multiple times previously with biological pores for detection of single molecules,[23–26] though in these measurements, the nucleic acid tether was measured and the target of interest remained outside the sensing zone of the nanopore. The same method was exploited for the sub-picoampere resolution discrimination of nucleotides.[27–30] Hydrogel facilitated detection can enable the same mechanism of capture but now with direct measurement of the bound target rather than the nucleic acid tether. As a result, high resolution and long duration measurement of single molecules may be feasible. In previous cases where the nucleic acid was in the sensing region, discriminating between multiple targets required engineering nucleic acid strands with reliably, identifiable markers. However, with hydrogel facilitated capture, the target is measured directly, and additional tagging would be unnecessary if the target particles are different in size.

5.2.3 Development of data analysis methods

In Chapter 4, a method to analyze data was presented that uses a modified Canny edge detection algorithm and hidden Markov models (HMMs) to locate events in noisy data. HMMs can be further developed to improve event detection and to parse additional information from nanopores. In the HMMs used for data analysis, emission spectra from both states are assumed to be normally distributed. This assumption has not interfered with the quality of analysis, but is not ideal for measurement of asymmetrical particles, such as proteins. In the case of these particles, the emission spectra of the blockade state can be non-normal, as modeled previously.[1] Modifying the hmmlearn package so that one state is fit to a normal distribution (the baseline state) and one state is fit to the orientation-dependent current blockade model (the event state) should increase accuracy of analysis for asymmetrical particles. Furthermore, because the unsupervised HMM fits

the data to these emission models, it doubly acts as a curve fitting algorithm to determine ΔI_{max}, ΔI_{min}, and the dipole moment during analysis.

In our implementation of HMM, we also always solved for 2 available states, but HMM can be expanded to solve for additional states as needed. This can have relevance in parsing data where more than one particle can exist in the nanopore. In these situations, HMM could identify the baseline, the first level blockade, and any additional blockade levels from each other to make separated analysis of levels possible. A more exciting implementation for multistate HMM is the analysis of individual event data from a mixture of proteins. In a mixture of proteins, each detected event should have a current distribution that is unique to a given protein. HMM may be capable of separating events from baseline and attributing individual events to particular proteins in mixture without knowing the identity of those proteins before analysis. In situations where the number of proteins in mixture is unknown, likelihood-based methods such as the Akaike information criterion can be used to guess at the number of different detected proteins.[31] Although experimental methods and data for this analysis has yet to be shown, HMM promises to be an excellent tool for analysis of this type of data.

5.3 Concluding remarks

This work presented in this dissertation introduces interfacing hydrogels with nanopores to address the fast translocation of proteins in conventional measurement. The issue of short dwell times obstructs further work on protein measurement with nanopores. Unlike other methods to address this issue, hydrogel interfaced nanopores are amenable to almost all measurement apparatuses, easy to implement, inexpensive, label-free, and highly sensitive. Although the technique has only

begun to be explored, these characteristics are highly desirable as research on protein measurement approaches more ambitious applications, such as biomarker detection and single-cell proteomics.

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

2. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* acsnano.8b09555 (2019). doi:10.1021/acsnano.8b09555

3. Maxwell, J. C. *A Treatise on Electricity and Magnetism*. (Clarendon Press, 1904).

4. DeBlois, R. W. & Bean, C. P. Counting and sizing of submicron particles by the resistive pulse technique. *Rev. Sci. Instrum.* **41**, 909–916 (1970).

5. Gregg, E. C. & Steidley, K. D. Electrical counting and sizing of mammalian cells in suspension. *Biophys. J.* **5**, 393–405 (1965).

6. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Stat. Soc. Ser. B* **39**, 1–22 (1977).

7. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).

8. Yusko, E. C. *et al.* Controlling protein translocation through nanopores with bio-inspired fluid walls. *Nat. Nanotechnol.* **6**, 253–260 (2011).

9. Plesa, C. *et al.* Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).

10. Kuramochi, H. *et al.* Vascular Endothelial Growth Factor Messenger RNA Expression Level Is Preserved in Liver Metastases Compared with Corresponding Primary Colorectal Cancer. *Clin. Cancer Res.* **12**, 29 LP – 33 (2006).

11. Raben, D. & Ryan, A. Vascular-Targeting Agents and Radiation Therapy in Lung Cancer: Where Do We Stand in 2005? *Clin. Lung Cancer* **7**, 175–179 (2005).

12. Nakahara, H. *et al.* Anti–interleukin-6 receptor antibody therapy reduces vascular endothelial growth factor production in rheumatoid arthritis. *Arthritis Rheum.* **48**, 1521–1529 (2003).

13. Etzioni, R. *et al.* The case for early detection. *Nat. Rev. Cancer* **3**, 243–252 (2003).

14. Ross, J. S. *et al.* The HER-2 Receptor and Breast Cancer: Ten Years of Targeted Anti–HER-2 Therapy and Personalized Medicine. *Oncologist* **14**, 320–368 (2009).

15. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).

16. Kulasingam, V. & Diamandis, E. P. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat. Clin. Pract. Oncol.* **5**, 588 (2008).

17. Tothill, I. E. Biosensors for cancer markers diagnosis. *Semin. Cell Dev. Biol.* **20**, 55–62 (2009).

18. Ebert, M. P. A., Korc, M., Malfertheiner, P. & Röcken, C. Advances, Challenges, and Limitations in Serum-Proteome-Based Cancer Diagnosis. *J. Proteome Res.* **5**, 19–25 (2006).

19. Al Sulaiman, D., Cadinu, P., Ivanov, A. P., Edel, J. B. & Ladame, S. Chemically Modified Hydrogel-Filled Nanopores: A Tunable Platform for Single-Molecule Sensing. *Nano Lett.* **18**, 6084–6093 (2018).

20. Tang, Z. *et al.* Gel mesh as "brake" to slow down DNA translocation through solid-state nanopores. *Nanoscale* **7**, 13207–13214 (2015).

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

22. Zhou, D., Zhao, Y., Tian, E. & Wang, D. Slowing Down DNA Translocation Speed thorough a Nanopore by a Nanofibre Meshed Layer. *Biophys. J.* **114**, 180a (2018).

23. Wang, H., Song, Z., Zhang, H. & Chen, S. Single-molecule analysis of lead (II)-binding aptamer conformational changes in an α-hemolysin nanopore, and sensitive detection of lead (II). *Microchim. Acta* **183**, 1003–1010 (2016).

24. Kawano, R. *et al.* Rapid detection of a cocaine-binding aptamer using biological nanopores on a chip. *J. Am. Chem. Soc.* **133**, 8474–8477 (2011).

25. Ding, S., Gao, C. & Gu, L.-Q. Capturing Single Molecules of Immunoglobulin and Ricin with an Aptamer-Encoded Glass Nanopore. *Anal. Chem.* **81**, 6649–6655 (2009).

26. Rotem, D., Jayasinghe, L., Salichou, M. & Bayley, H. Protein Detection by Nanopores Equipped with Aptamers. *J. Am. Chem. Soc.* **134**, 2781–2787 (2012).

27. Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M. & Gundlach, J. H. Nucleotide Discrimination with DNA Immobilized in the MspA Nanopore. *PLoS One* **6**, e25723 (2011).

28. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci.* **106**, 7702 LP – 7707 (2009).

29. Purnell, R. F., Mehta, K. K. & Schmidt, J. J. Nucleotide Identification and Orientation Discrimination of DNA Homopolymers Immobilized in a Protein Nanopore. *Nano Lett.* **8**, 3029–3034 (2008).

30. Purnell, R. F. & Schmidt, J. J. Discrimination of Single Base Substitutions in a DNA Strand Immobilized in a Biological Nanopore. *ACS Nano* **3**, 2533–2538 (2009).

31. Akaike, H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60**, 255–265 (1973).